International Journal of Psychophysiology 83 (2012) 164–175

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Review

Predictive information processing in music cognition. A critical review

Martin A. Rohrmeier *, Stefan Koelsch

Cluster Languages of Emotion, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 6 August 2011
Received in revised form 27 December 2011
Accepted 28 December 2011
Available online 12 January 2012

Keywords:
Prediction
Expectation
Expectancy
Music cognition
Computational modelling
n-gram model
Probabilistic model
Simple recurrent network
Neuroimaging
ERAN
MMN

ABSTRACT

Expectation and prediction constitute central mechanisms in the perception and cognition of music, which have been explored in theoretical and empirical accounts. We review the scope and limits of theoretical accounts of musical prediction with respect to feature-based and temporal prediction. While the concept of prediction is unproblematic for basic single-stream features such as melody, it is not straight-forward for polyphonic structures or higher-order features such as formal predictions. Behavioural results based on explicit and implicit (priming) paradigms provide evidence of priming in various domains that may reflect predictive behaviour. Computational learning models, including symbolic (fragment-based), probabilistic/graphical, or connectionist approaches, provide well-specified predictive models of specific features and feature combinations. While models match some experimental results, full-fledged music prediction cannot yet be modelled. Neuroscientific results regarding the early right-anterior negativity (ERAN) and mismatch negativity (MMN) reflect expectancy violations on different levels of processing complexity, and provide some neural evidence for different predictive mechanisms. At present, the combinations of neural and computational modelling methodologies are at early stages and require further research.

1. Introduction

"A mind is fundamentally an anticipator, an expectation-generator." (Dennett, 1996: 57). It is well-understood that constant predictive activity is indispensable and vital for survival: it affords for the interaction with the environments' complexity as well as the alignment of internal structure with sensory input for the construction and update of the cognitive world model. Expectancy plays an evolutionarily established role in all forms of cognition, thus also representing one cornerstone for music cognition.

Predictive information processing is fundamental to music in three ways. (1) Prediction and expectancy incorporate the essence of the dynamics of musical temporality. Further they make the experience of local or large-scale goal-directed processes in music possible (based on, e.g., melodic, harmonic or modal features; Schenker, 1935; Schoenberg, 1978; Narmour, 1990). (2) Predictive processing constitutes a major process involved in musical interaction and synchronisation (Keller, 2008; Keller and Koch, 2008; Bharucha et al., 2006; Cross, 2003; Large, 2010b). (3) Finally, processes of expectancy and prediction are understood to be linked with specific emotional and aesthetic musical effects (Meyer, 1956; Huron, 2006; Steinbeis et al., 2006; Koelsch, 2010).

This article focuses on structural aspects of predictive information processing in music. We will begin with a brief music theoretical discussion of the concept of musical prediction and its peculiarities. We will then outline some of the main empirical findings on musical expectation and prediction from behavioural, computational and neuroscientific approaches and discuss their potential for converging evidence as well as open questions. With respect to the terminology, we will follow the suggested distinctions by Bubic et al. (2010) as well as Marcus Pearce (personal communication). We use *prediction* as a general term for the overall process of future-directed information processing, *expectation* as the representation of what is predicted to occur (including a probability distribution and not necessarily a single event), *prospection* as the prediction of events in the more distant future. Further we use *expectancy* as a general state or process of expecting something to happen. With respect to musical events, we use *anticipation* in accordance with the music theoretical term as an occurrence of an event earlier than its expected occurrence (in contrast to the use suggested by Bubic et al., 2010).

2. Predictable information within the music

Once we conceptualise music (in its physical form) as one or more concurrent temporal streams of auditory events, primarily two fundamental, categorically different forms of musical play with expectancy are possible: the play with *what* to expect and the play with *when* to expect an event (as well as unspecific expectancies *that* something is

* Corresponding author.
E-mail addresses: mrohrmeier@cantab.net (M.A. Rohrmeier), koelsch@cbs.mpg.de (S. Koelsch).

0167-8760/$ – see front matter
doi:10.1016/j.ijpsycho.2011.12.010

going to happen). The *what* involves the expectancy of a particular event out of a set of possible events of the same kind (even though the expectancy or prediction of events of different parameters or features may, and do, interact). The *when* involves the matching of the present structure with metrical structures, rhythmic or rhythmic-metrical templates that can be extrapolated into the future. Such kinds of play with expectancy are reflected in compositional devices such as anticipation, suspension, delay/retardation, deceptive or evaded cadences, or applied dominants (e.g. the second chord of "Yesterday", "All of Me", or Oscar Peterson's "Hymn to Freedom"). However, the formalisation of the features that we predict while listening to music entails a number of conceptual caveats that will be briefly outlined here.

2.1. Structural and temporal prediction

With regard to the "What"-aspect, the music cognition literature mainly discusses local prediction on the level of melody, phrase structure, harmony, or potentially key or formal structure. The listener is, for instance, assumed to form expectancies with respect to the next note(s) or with respect to the boundary of a phrase. While melodic, phrase or formal predictions bear some cross-cultural validity, features of harmony or key are specific to Western tonal music. Enculturated Western listeners presumably form expectations with respect to upcoming chords or harmonic structure from a small lexicon of tonal chords. Highly experienced listeners may even expect a typical change to the key of the dominant or the relative minor within large scale form of, say, a sonata schema. Similarly, the formal structure of a piece may trigger formal expectancies or prospections for the expert listener (such as the prospection of a development or recapitulation section in a sonata, or the descending section in *Maqam Bayyati*, Marcus, 2003). These higher-order expectancies, however, are to some extent different from lower-order predictions (e.g. note to note) since it may not be well-defined how higher-order structures map to the musical surface and when such expectancies are actually fulfilled or not (for instance, when would the expectation of a formal section be accounted as fulfilled? with the first note, bar, in the middle of the first phrase? see also below). Accordingly, a stream-of-features view of musical prediction presumes that parallel, ongoing predictions are constantly formed for subsequent events with respect to multiple musical features (such as different melodic lines, harmony, key, phrase level, etc.). In this context, one core precondition for the emergence and formation of musical expectancies is that the structure is built from a discrete small number of available elements (e.g. number of scale tones, chords, keys, number of formal parts). In contrast to continuous features, such restricted, small alphabets of structural elements on different levels facilitate combinatoriality and predictability.

With regard to the "When"-question, both rhythmic and metric predictions are formed. The inference of metrical structure on different levels is assumed to create cyclical weighted peaks of attention at strong beats (Large and Palmer, 2002) that correspond with expectancies of structurally weighted events. Similarly, acquired or online-learned rhythmic patterns imply predictions about their continuation. Temporal and structural predictions interact, because structural predictions also often involve temporal features and vice versa. An expected note can become unexpected when it happens early (anticipations) or late (delay/retardation). Conversely, when other ornamental notes are inserted on a lower metrical level prior to the expected note occurring at the expected time, one may not necessarily regard this as an expectancy violation (and rather as an insertion). Such aspects of ornamental events, structurally fundamental and less fundamental events upon which predictive processes may operate are theorised by theories of musical syntax (e.g., Schenker, 1935; Lerdahl and Jackendoff, 1983; Steedman, 1996; Rohrmeier, 2007a; 2011). Accordingly, predictive processes may not necessarily operate on the flat stream of notes or other features, but in interaction with metrical/rhythmical attention as well as potentially different levels of reduction based on syntactic parse trees derived from a syntax model. Prediction thus involves combined structural and temporal expectations.

Beyond the particular framework of Western tonal music theory, however, other forms of prediction (or at least surprise) are plausible that are not frequently theorised: for instance, the strong effect of abrupt breaks or changes of soundscape (as found in contemporary pop music), orchestration, or timbre (cf. Sloboda et al., 2001; Grewe et al., 2007) reveals indirectly that cognitively based non-symbolic expectations of, at least, continuity of the auditory scene or continuity of these features were at work, that accord with theories of ecological listening and involve the appropriation of other cognitive auditory mechanisms (Clarke, 2005). It is difficult, however, to conceptualise such forms of ongoing predictive soundscape integration under the same umbrella term as structural, feature-based prediction (because it is unclear which predictions except for continuity are formed).

2.2. Implications of syntactic models for musical prediction

Some complexities arise when prediction and its research methodologies are matched with syntactic theories of music. In syntactic accounts of music, that employ recursive structures or are based on context-free grammars (Lerdahl and Jackendoff, 1983; Steedman, 1996; Tojo et al., 2006; Rohrmeier, 2007a, 2011), the prediction of future events becomes complex compared with models based on simpler finite-state grammars or Markovian models (see Chomsky (1956), for a discussion of the difference between context-free, finite-state and Markov models; see also below). Long-distance dependencies may affect the probability profiles of expected events on top of their immediately preceding context or may govern the prediction in cases of local closures. The formalism of context-free models generally licenses insertions within event sequences, which may prepare intermediary goals or prolong an event. This entails that predictive contexts may not necessarily be locally connected, but may affect local predictions. Similarly, schema-based theories of music (e.g. Gjerdingen, 2007; Byros, 2009) involve larger-scale predictions based on schemata that may not necessarily operate on a note-to-note level. These music theoretical accounts motivate a view that musical prediction may not necessarily operate from one event to the next (see also Neuwirth, 2008). This view is not in line with cognitive accounts based on priming or computational predictive models (below) which make predictions for local element- by- element transitions. While a cognitive perspective clearly motivates the view that prediction is formed and updated constantly, the theoretical accounts presented here make predictions with respect to higher order musical predictions (beyond the note-to-note level).

2.3. Sources of prediction

The formation of musical predictions for different features involves complex interactions between acquired style-specific syntactic or schematic knowledge (e.g., harmonic rules or patterns of chord progressions), sensory and low-level predictions (e.g., predictions based on metre, timbre, and texture), veridical knowledge of the present piece as well as non-sensory structures acquired during a piece through processes of online-learning (e.g., predictions based on previous occurrence of the same motive; see Fig. 1 for an overview). Style-specific structural knowledge may involve syntactic structures or schemata. Sensory or low-level predictions may stem from Gestalt perception or 'data-driven' perception (Bharucha and Stoeckig, 1987). Veridical source of prediction refers to the knowledge acquired by prior exposure to the same piece (Bharucha and Stoeckig, 1987; Eerola, 2003), while online-learned structures such as motives, statistical structures or probability profiles, for instance, refer to

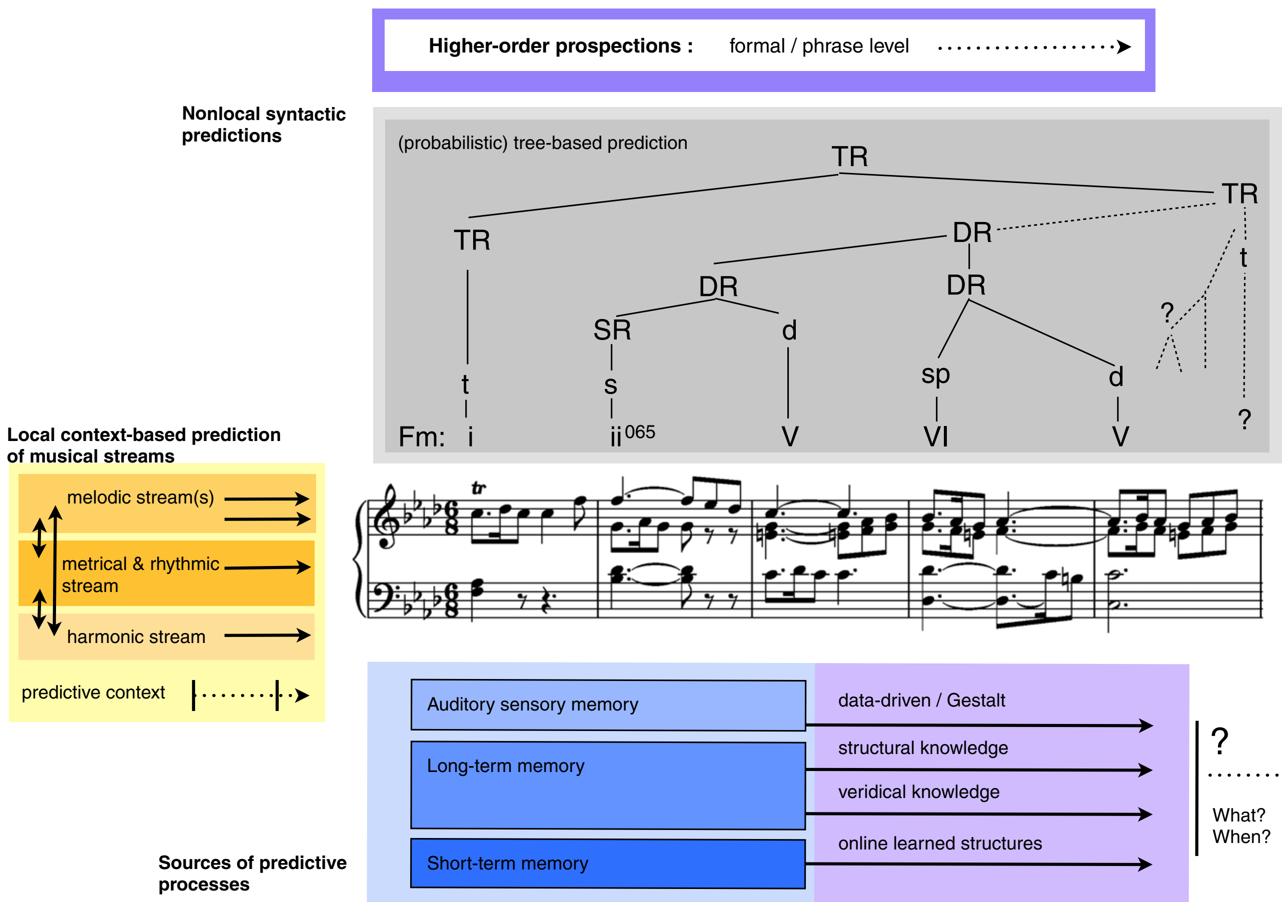

Fig. 1. Different sources of musical prediction.

knowledge acquired during listening to a piece through online-learning (see also Rohrmeier et al., 2011; Rohrmeier, 2009). These different sources for the formation of expectancies and predictions may be in mutual conflict.

2.4. Caveats

As noted above, the prediction of a melodic note or a chord may involve a clearly distinct next element, the prediction of formal or key structure is to some extent less well-defined because it does not necessarily involve the prediction of the very next musical event and because it does not necessarily project to a specific unambiguous starting point or event. A passage in the dominant key may raise the expectancy of the return to the tonic key, but this form of expectancy is temporally flexible. Within this perspective, the prediction of harmony is, to some extent, comparably soft as well. Although chords form a relatively small lexicon that affords for stream-based sequential prediction, the fact that the component tones can be flexibly spread within a temporal window and include non-chord tones (cf. Bach's famous C-minor or C-Major preludes in the *Well-tempered Clavier I*), requires a number of theoretical as well as cognitive assumptions (cf. Temperley and Sleator, 1999; Temperley, 2001; Rohrmeier, 2005), such as stable harmonic rhythm, to render them as predictive units or to potentially conceive of them as structures similar to local pitch profiles with particular root.

The outlined stream-of-feature understanding of predictive perception entails some caveats and open questions: There are forms of music, such as African, (Middle-)Eastern, Western Avantgarde or Baroque music, which employ complex and non-aligned concurrence of features, such as polyrhythms, polymetre or complex polyphony, which involve multiple non-aligned rhythms, metres or melodic patterns (cf. Agawu, 1995; London, 2004; Temperley, 2001). While the assumption of parallel streams of attentive prediction for e.g. melodic, harmonic and key structure is plausible, there is no coherent picture from music-psychological research concerning the interaction of attention and prediction within complex polyphonic streams, or the choice of attentive strategies when there are several concurrent streams in a musical piece.

Fig. 2 illustrates some of the difficulties regarding the concept of prediction with respect to complex polyphonic music as marked by the arrows. The example illustrates issues with respect to stream-segregation and stream-based prediction. Additional, temporary, interrupted or paused voice streams, melodic leaps and crossing voices and changes of texture (e.g., measures 2, 3, 4, 10) make prediction based on independent melodic voice streams difficult to apply consistently. In addition, potentially simultaneous presence of binary, ternary or quaternary division of the underlying quarter note beat (here briefly in measure 2) adds complexities with respect to the temporal prediction (if the voices should be treated in interaction). With respect to harmony, tones of predicted chords may not appear at the same time. Musical texture provides a challenge to a coherent prediction concept. The transition from measure 1 to measure 2 involves a change from monophonic melody to (mostly) four-part polyphony. The predictable chord in measure 9 contains five voices, whereas the preceding context contains four voices. The texture of measure 10 differs from the previous context (though it is still based on four voices as most of this excerpt). All cases in this example shall not argue against the plausibility of prediction per se, but they should highlight that the prediction of harmony, independent voices, the interaction between rhythmic, melodic and harmonic structure is

Fig. 2. Excerpt from Liszt's B-Minor Sonata (1854), mm. 404–413. The example illustrates some of the intricacies regarding the concept of prediction with respect to complex polyphonic music.

not straight-forward coherent and requires very flexible general specifications in order to be transferable to complex musical cases. Predictive processes may not operate on the entire polyphonic context but depend on the listener attention to particular melodic or harmonic streams.

In summary, caveats concerning the definition of musical prediction and expectancy involve that: (a) event expectations do not necessarily refer to the physically next event, (b) temporal and structural expectations, as well as attention interact, (c) some complex feature expectations may not directly correlate with precise time points or note events, (d) the extension of prediction for multiple complex feature streams is potentially ill-defined.

3. Behavioural findings

The majority of cognitive research has focused on melodic and to some extent on harmonic and rhythmic prediction, while the prediction of polyphonic, formal, timbral, or sound features has found little attention. Musical prediction is largely assumed to be governed by sensory (or data-driven/Gestalt-based) prediction, syntactic/schematic, veridical and online-learned predictive processes. Except for sensory or Gestalt-based accounts of cognition (such as Narmour's bottom-up principles, Narmour, 1990,, see below)little, if barely any, musical competence involved in prediction is assumed to be innate in music cognition. Accordingly, music prediction is understood to be largely governed by implicit knowledge that people acquire incidentally during interactions with music (with respect to syntactic, schematic or veridic knowledge, Rohrmeier, 2010; Rohrmeier et al., 2011; Tillmann, 2005) or with the present performance. With respect to the differences between these forms of predictive knowledge, research on the types of structures that can be acquired implicitly is crucial (see below, cf. Pothos, 2007).

In a behavioural setting, a series of probe-tone studies as well as the work by Schmuckler (1989), Manzara et al. (1992) and Witten et al. (1994) provided major results regarding musical feature prediction. Tone profiles (unigram distributions) as modelled by early computational approaches (e.g. Youngblood, 1958; Budge, 1943; Ames, 1989) were found to constitute basic tonal forms of prediction for neutral tonal probe contexts (Krumhansl and Kessler, 1982; Krumhansl and Keil, 1982; Krumhansl, 1990). Cuddy and Lunney (1995) found that Narmour's intervallic principles were overriding tone profiles in continuation tasks of short sequences. In a probe-chord context, Schmuckler (1989) employed different contexts of varying complexity from a Schumann lied and found participants' behaviour to largely accord with Piston's table of usual root progressions (Piston, 1948), which may be regarded as a handcrafted predictive bigram model of harmony (see below). Employing a predictive betting paradigm, Manzara et al. (1992) and Witten et al. (1994) obtained behavioural data of predictive distributions and entropy for chorale melodies and compared these with predictions from computational models of melodic prediction. A number of cross-cultural melodic prediction experiments further disentangled some culture-specific from general data-driven prediction principles (Eerola, 2003; Eerola et al., 2009; Krumhansl et al., 1999, 2000). Many of these behavioural results could be modelled by computational models (see below). However, these studies did not use methodologies to test for automatic or implicit knowledge, which makes it difficult to disentangle the contribution of explicit higher-order cognitive processes from implicit, automatic processes.

Musical priming studies give ample evidence of implicit local knowledge that governs the prediction of musical structure in musicians as well as nonmusicians (Tillmann, 2005). While participants, particularly nonmusicians, may not be able to explicitly name or describe an expected event, they are found to exhibit considerable musical knowledge in implicit tasks (Bigand and Poulin-Charronnat, 2006). In this context, priming studies give evidence for knowledge-driven expectations and expectancy-driven response preparation and facilitation in musical processing. Musical priming has been mainly studied for harmonic sequences. Adapted from linguistic semantic studies (Neely, 1976), Bharucha and Stoeckig (1986, 1987) and Tekman and Bharucha (1998) applied the priming paradigm with short harmonic chord-target contexts, and found evidence for subjects' sensitivity to in-key and out-of-key targets. Using the same paradigm, Tillmann and Bharucha (2002) found that harmonic relatedness and temporal synchronicity play a comparably important role. Interestingly, priming in music appears to be more strongly governed by tonal relatedness than immediate repetition (Bigand et al., 2005). Bigand and Pineau (1997) and Bigand et al. (1999) used larger harmonic contexts for the priming task. Their data suggests that musical expectations are not only governed by local contexts (direct chord-to-chord relationships), but also higher tonal key relationships and temporal aspects. For comprehensive overviews of the current state of priming in music and language see Tillmann (2005) and Tillmann and Bharucha (2002).

4. Computational models

Predictive computational models of music cognition provide a link between theoretical, behavioural and neural accounts of musical prediction. Because *all* assumptions have to be made explicit in order to be implementable in a computational model, such models make the empirical testing of overt and hidden assumptions concerning *knowledge representation*, *mechanisms* and *modules of information processing*, drawn from these individual disciplines, possible. Moreover, they allow to explore their implications and hypothesised predictions for novel contexts (Wiggins, 2011). While there are many models for different practical aspects of music information retrieval, this review focuses specifically on cognitive, predictive models of music.

The symbolic nature of higher-order (Western) musical structure and notation affords for a representation based on a discrete, small alphabet of symbols, which facilitates cognitive expectancy formation and has in turn enabled and triggered the development of various successful symbolic computational models of music prediction. Operating on a small (melodic or harmonic) alphabet such models abstract from features such as small deviations in pitch intonation or timbre, and few predictive modelling approaches operate on the complex signal level representation (see Wiggins, 2011, for a discussion about the validity and independence of levels of representation).

Most current computational models of music cognition are based on *learning* and *inference* methods rather than (earlier) "knowledge engineering" and hand-crafted processes and representations. The advantage of powerful learning algorithms over hard-wired processes in cognitive modelling corresponds with the understanding that adaptive organisms, that learn and adapt to regularities from the environment, are more flexible in complex environments. In correspondence with cognitive results on the implicitly acquired nature of musical competence (cf. Rohrmeier, 2010; Tillmann, 2005), present computational models are machine-learning models that acquire their knowledge from training on a large musical corpus. The discussed learning models differ with respect to the distinction whether they are *supervised* or *unsupervised*. Supervised models are trained using prespecified pairs of inputs and desired outputs, while unsupervised models are merely trained by input sequences only and derive their output from the statistical structure of the input. In this respect unsupervised models are more adequate as cognitive models than supervised models.

Computational models of music prediction have been created on different levels of representation (Marr, 1982). Most frequent and successful modelling approaches have been symbolic fragment and n-gram models, graphical and Hidden Markov Models (modelling at level of algorithmic or statistical computation), and connectionist models (modelling the implementation level of neural connectivity). The problem setting of music prediction shared by the different modelling approaches is to compute the probability distribution $p(e_i|e_1^{i-1})$ of an event e_i, which may as well represent temporal information, given the sequence of past events e_1^{i-1}. Here, e_i denotes the ith event in the sequence and e_a^b denotes the subsequence from index a to b.[1]

4.1. Hand-crafted and early statistical models

There are a number of music theoretical, hand-crafted models of musical expectancy. We count them as computational models because they constitute pen-and-paper models (or simple empirical models) of musical prediction that are sufficiently well-specified to be unambiguous, computationally implementable and yield testable predictions (Wiggins, 2011). Some of these hand-crafted theoretical models employ the form of simple statistical models.

For instance, Piston (1948) described a table of usual root progressions, which lists common progressions between scale degree roots and intuitively estimated frequencies (e.g. II is followed by V, sometimes by IV, VI, less often by I, III). This table of root progressions constitutes a simple case of a model which derives a prediction of the subsequent root based on the context of one (the current) root. A corpus analysis of Bach's chorales found that Piston's progression table is largely in agreement with a statistical bigram model of pitch class set transitions (Rohrmeier, 2005). Another analysis of harmonic structure was undertaken by Eberlein (1994) with a small, manual corpus analysis of chord progressions.

A similar case of hand-crafted predictive models constitutes the melodic principles of Narmour's Implication-Realization model (Narmour, 1990). Narmour's theory (Narmour, 1990, 1999) formalises melodic expectation proposing in part a static, innate top down model. It proposed a set of rules characterising how melodic *implicative* intervals are followed by specific *realised* intervals. The theory has been simplified and quantified (Krumhansl, 1995; Schellenberg, 1997) based on five principles, such as *registral direction* (small intervals are continued in the same direction, large intervals in the different direction) or *registral return* (the realised interval tends to return to the same pitch or neighbouring pitches, when it changes the direction of the implicative interval). Accordingly, similar to Piston's table, this theory constitutes a hand-crafted model that specifies a prediction of an interval (realised interval) from a given context (implicative interval). This model, as Piston's model, can be implemented and computationally evaluated (see below).

Similarly, models of element distributions constitute simple statistical models of music and musical prediction. Pitch profiles, such as the ones suggested theoretically by Lerdahl (2001) or empirically by Krumhansl (1990) and Krumhansl and Kessler (1982), make predictions about expected frequencies of pitch occurrences given a key context (e.g. in C major). This model can be conceptualised as a pitch prediction model with a zero-length context, which affords for a generalisation by fragment models (as argued below). Similar generalisations hold for early statistical approaches, such as Budge's (1943) analysis of chord frequencies in tonal pieces, and early computational approaches (see Ames, 1989) or Youngblood (1958), who analysed tone frequencies and (first-order) pitch class transitions in songs by Schubert, Mendelssohn and Schumann, as well as in Gregorian chant.

4.2. Fragment and Markovian models

Most of symbolic, fragment-based models intend to model cognitive processes and their operations on data in a statistical way independently of their neural implementation. In accordance with chunk-based theories of implicit learning (Servan-Schreiber and Anderson, 1990; Perruchet and Pacteau, 1990; Pothos, 2007), fragment models are symbolic models that implement the assumption that learning is based on extracting, storing, and combining small contiguous fragments (also referred to as chunks or n-grams) of sequences for prediction or recognition of sequences. A variety of different computational models, like n-gram or prediction by partial matching (Ames, 1989; Manning and Schütze, 1999; Pearce and Wiggins, 2004), multiple-viewpoint (Conklin and Witten, 1995), Competitive Chunker (Servan-Schreiber and Anderson, 1990), PARSER (Perruchet and Vinter, 1998), or IDyOM (Pearce, 2005; Pearce and Wiggins, in press) share and implement similar ideas of unsupervised fragment acquisition. They differ mainly in the ways how fragments are represented, combined, matched or weighted. The application of statistical fragment models reaches back to the early 50s and 60s (e.g. Cohen, 1962; Youngblood, 1958; Hiller and Bean, 1966; Hiller and Fuller, 1967; Pinkerton, 1956; Ames, 1987, 1989). In the context of musical prediction, predominantly n-gram and derivative models were applied while Competitive Chunker or PARSER remained unexplored.

4.2.1. N-gram models

Most simply, n-gram models learn information from sequences by chopping them into short fragments up to a size of n (referred to as

[1] Note that this definition does not include predictions or prospections beyond the immediately subsequent element.

"n-grams") and compute predictions for a given context based on the statistics of the set of stored n-grams, which contain a part or all of the given context as well as its continuation. For instance, a 3-gram model for melody would pick from the tone sequence 'A C E G C E A C E G' the following set of 3-grams (with number of occurrence): $\{ACE:2, CEG:2, EGC:1, GCE:1, CEA:1, EAC:1\}$. It's prediction for the context 'C E', for instance, would be 'G' or 'A' with respective probabilities of 2/3 and 1/3. Hence a bigram model, that acquires tone pairs, results to be identical to a simple Markovian tone transition matrix. The fundamental assumption behind this and other fragment-based approaches is the Markov assumption (1) that the probability $p(e_i|e_1^{i-1})$ of the current event e_i (given the context of the entire past sequence e_1^{i-1}) is well-approximated (or equals) the probability $p(e_i|e_{i-n+1}^{i-1})$ given the past n events in the sequence. In the example, the Markov assumption for 3-gram models would state that contexts longer than 2 items are not respected in the prediction, i.e. the prediction 'G' with $p = 2/3$ after 'CE' is assumed not to be affected by any note before 'CE'. Hence the Markov assumption methodically excludes the impact of long-distance dependencies (longer than the context length) on the prediction of the current event.

$$p(e_i|e_1^{i-1}) \approx p(e_i|e_{i-n+1}^{i-1}) \quad (1)$$

The computation of the actual prediction of the current event is based on the probability distribution that arises from the continuations of all encountered fragments that match the current predictive context (Eq. (2); in analogy with the explanation above).

$$p(e_i|e_{i-n}^{i-1}) = \frac{count\left(e_{i-n}^{i}\right)}{count\left(e_{i-n}^{i-1}\right)} \quad (2)$$

This basic predictive model has to be amended in order to deal with cases when very few or no instances of the current context were observed. Secondly, because predictions based on different context-lengths may differ to a larger extent, methods of combining predictions from various contexts exist. A number of smoothing methods have been proposed in order to achieve good prediction that balance between too little specificity (avoid problems of very rare or unseen contexts in which the numerator would be 0) and too much specificity (avoiding overfitting with too precise, piece-specific knowledge in which the fraction approaches 1 when the set in the denominator is too small or singleton). Zero-escape methods approximate rare or unseen cases with smoothed probabilities from other rare cases. Smoothing methods combine ("blend") weighted predictive probability distributions from different context lengths in order to balance different degrees of specificity. An extensive comparison of different methods found Witten–Bell smoothing (Witten and Bell, 1991) to work best for the case of melodic prediction (Pearce and Wiggins, 2004). They further found that melodic prediction was not much improved beyond 3-gram models involving short context-lengths of 2 sequence units.

N-gram models have been successfully applied for modelling the prediction of melody (Pearce and Wiggins, 2006) as well as harmony (Ponsford et al., 1999; Whorley et al., 2010). N-gram models entail a generalisation that hand-crafted models such as pitch profiles (Krumhansl, 1990), for instance, are revealed to be a special case of unigram models (with a zero-length context, i.e. predicting based on the overall pitch distribution without context).Similarly, Piston's (1948) table of usual root progressions is a case of a hand-crafted bigram model of root progressions. A similar generalisation holds for Narmour's melodic principles. Narmour's principles constitute hand-crafted intervallic bigram models of melody. Generally, Narmour's, 1990 static top-down model of melodic prediction, that he assumed to be innate and based on general Gestalt principles, has been suggested to be subsumed by n-gram learning (Pearce and Wiggins, 2006).

The *multiple viewpoint idea* (Conklin and Witten, 1995) expands n-gram methods to modelling the interaction of information of parallel streams of basic and (precoded) high-level musical features. It constitutes the core of some predictive cognition approaches (e.g. Reis, 1999) and the heart of the IDyOM model (Pearce, 2005; Pearce and Wiggins, in press). The core idea of this approach lies in utilising redundancy and dependencies between features for the prediction of the present symbol: for instance, to utilise duration or metrical position to improve the prediction of pitch class. It creates a combined n-gram prediction for different sets of features and feature combinations and projects the feature space down to the feature(s) to be predicted. This way the predictive power of different sets of features for the prediction of another specific feature could be compared. However, the selection of the most powerful set of features cannot be easily derived within this paradigm. Optimisation and search techniques (such as genetic algorithms) have to be applied in order to find strongly predictive parameter combinations.

Another powerful step, suggested in combination with multiple-viewpoint models (Conklin and Witten, 1995), combines two memory components, one that involves the knowledge from the present piece (*short-term model*) and one that involves the acquired knowledge from the whole corpus (*long-term model*). Such a combination has proven powerful to model the overlap between long-term acquired (melodic) patterns and specific repetitive and frequent patterns that are particular to the current piece and picked up during implicit online-learning. The heart of the powerful IDyOM model (Pearce, 2005; Pearce and Wiggins, in press) is constituted by the multiple-viewpoint technique, combined with optimised ways of smoothing, viewpoint combination as well as the blending of long-term and short-term predictions.

N-gram and multiple-viewpoint methods have proven powerful for melodic prediction (Conklin and Witten, 1995; Pearce and Wiggins, 2004, 2006), key region prediction (Rohrmeier, 2007b), melodic and vertical voice-leading patterns (Conklin and Anagnostopoulou, 2001; Conklin, 2002, 2010) as well as in initial results for harmonic prediction (Ponsford et al., 1999; Whorley et al., 2010). As cognitive model, IDyOM has further been evaluated with a series of experimental results. It has been shown to be efficient for the prediction of melodic phrase-boundaries (Pearce et al., 2010a; see also Neuhaus et al., 2006) or melodic prediction (Pearce and Wiggins, 2006; compare also Manzara et al., 1992; Witten et al., 1994) without requiring hand-crafted knowledge engineering. In addition it has been shown to predict electroencephalographic (EEG) data for melodic prediction (Pearce et al., 2010b).

The knowledge representation of the Competitive Chunker model (Servan-Schreiber and Anderson, 1990) exceeds n-gram models with higher-order integration of smaller chunks into hierarchically organised larger chunks, that are closer to syntactic theoretical descriptions of tonal music. Yet they have not been extensively been applied to music. In contrast to n-gram models or Competitive Chunker (as well as Hidden Markov Models, or neural network models), the PARSER model (Perruchet and Vinter, 1998) involves explicit forgetting and interference between fragments. In a model comparison based on a range of implicit learning experiments, n-gram models were found to generally outperform humans with respect to learning performance,whereas Competitive Chunker was close to the human range and PARSER below it (Rohrmeier, 2010).

4.2.2. Hidden Markov Models

Hidden Markov Models (HMM) are well-known graphical models which extend the notion of a Markov model to a higher-order model of sequence generation or prediction (Rabiner, 1989). A Markov transition matrix, which is equivalent to a 2-gram model, models the transition probabilities between single events in the alphabet and creates predictions based on this matrix. Hidden Markov Models in contrast extend this idea by assuming a Markov transition matrix, i.e. a single discrete random variable, that does not describe transitions between surface symbols but between hidden deep structure states. The hidden states in turn emit surface symbols by individual associated probability distributions over the alphabet of surface symbols. Identical hidden states for subsequent events of a sequence entail that the surface symbols are drawn from the same distribution while changes in hidden states imply different emission distributions. HMMs have been very efficiently applied in various domains which involve temporal sequence prediction, recognition or processing such as speech or gesture recognition, or bioinformatics.

One major difference to n-gram models is that HMMs take the entire sequence of past events into account for prediction and thus do not ground on the Markov assumption (see Eq. (1)). While unbounded n-gram models are theoretically possible though practically implausible due to data sparsity and the precise context matching, the nature of the generating predictions based on the forward algorithm makes it possible to model the probabilistic impact of local and nonlocal dependencies in a different way.

HMMs were not yet frequently employed for cognitive modelling, however, they were applied for problems of functional or key analysis, harmonisation or audio alignment (e.g. Raphael and Stoddard, 2004; Raphael, 2010; Allan and Williams, 2005). Paiement et al. (2009) employed a specific type of input–output HMMs for the modelling of melodic prediction based on harmonic structure. Another extension of the HMM model is provided by the idea of Dynamic Bayesian Networks (DBN, Murphy, 2002). In analogy to the multiple viewpoint approach, such temporal graphical models may take advantage of dependencies and redundancies between different musical features such as chord, duration, or mode. Unpublished data by Rohrmeier & Graepel suggests that the predictive power of DBN models of harmony improves when other features (e.g. mode) are taken into account and that some of these DBNs slightly outperform multiple-viewpoint n-gram models.

4.3. Connectionist models

Neural network models operate on a different level of abstraction and representation inspired by the connectivity, firing and growth dynamics of assemblies of biological neurons. The contribution of early connectionist models can be conceived of as a way to give a proof-of-concept demonstration that a particular framework was able to capture some of the complex higher-order features of human cognition such as error tolerance, or fuzzy knowledge (Hopfield, 1982; McClelland, 2009). Connectionist models were employed for modelling aspects of music production or perception (Bharucha and Stoeckig, 1987; Bharucha and Todd, 1989; Griffith and Todd, 1999; Leman, 1997; Stevens and Latimer, 1992, 1997; Mozer, 1994; Gang et al., 1998; Franklin, 2006; Maxwell et al., 2009) and applied to explain or predict features of predictive behaviour based on supervised or semi-supervised learning. Due to space limitations, this section will only focus on discussing central findings and select core studies from the large set of methodologically diverse studies.

Many connectionist approaches are, like the probabilistic models discussed above, grounded in symbolic representations of music and encode musical sequences symbolically by assigning single input neurons to single symbols from the discrete musical alphabet to be modelled. There are, however, a number of approaches to model subsymbolic processes of music perception (e.g. Large, 2010a, 2011; Leman, 1997; Toiviainen, 1996). Early connectionist models of music, such as MUSACT (Bharucha and Krumhansl, 1983; Bharucha and Stoeckig, 1987) had Western features and representations (the 12 chromatic pitches, the 24 diatonic major and minor chords, the 24 keys), hard-wired in their architecture, and were, less surprisingly, successful in predicting some features of tonal perception and prediction (Bharucha and Stoeckig (1987); Bharucha (1998); Bharucha et al. (2011); see Wiggins (2011), for a discussion). However, Bharucha et al. (2011) outlined specific surprising cases of the behaviour of his model that matched experimental data. Tillmann et al. (2000, 2001) built a Self-Organising Map (Kohonen, 1995) which they found to match with some major experimental results of tonal cognition, such as chord relations (Krumhansl et al., 1982; Bharucha and Krumhansl, 1983), key relations (Cuddy and Thompson, 1992; Krumhansl and Kessler, 1982) and tone relations (Krumhansl, 1979; Dowling, 1978). This model was, however, criticised by Wiggins (2011) for its inadequacy as cognitive model. However, their study constitutes a proof of concept with respect to the plausibility of unsupervised learning of the above features of tonal music by self-organisation.

One central issue in the connectionist modelling of predictive musical information processing constitutes the temporality of the musical stream. N-gram models capture time indirectly through the storage of sequence fragments and their size, as well as potential note duration features. Plain feed-forward networks, however, do not capture time since the processing of one sequence element in the feed-forward process does not affect the processing of the next element. Two cognitive approaches to solve this problem are the Simple Recurrent Network and buffer models (other approaches, for instance are spiking neural networks, Gerstner and Kistler, 2002; Ichishita and Fujii, 2007). Elman's,(1990) architecture of the Simple Recurrent Network (SRN) extends 3-layer feed-forward networks with the capacities for the processing of temporal sequences by storing a copy of the directly preceding hidden layer activations in a context layer that was connected to the output layer in the same way as the hidden layer. The buffer model (Boucher and Dienes, 2003) in turn endows a simple feed-forward network with a larger input layer of a number of simultaneous input symbols that does not only encode the present but also past sequence events. The model then shifts a sequence stepwise through the input layer so that sequence events except anchor positions have been moved through all positions of the input layer.

Mozer (1994) created an amended model specialised for the case of music which operates on prefigured, multilevel musical representations such as pitch height, chroma and relationships based on the cycle of fifths, and relative duration. The network was trained to predict the next musical event from the given input representation. The study showed that this specialised neural network outperformed bigram models, but was unable to capture and generalise to overarching structure (comparable with the findings of the Bayesian model comparison by Rohrmeier, 2010). Simple Recurrent Networks have been known to be able to learn transition probabilities and structures of the complexity of finite-state grammars (cf. Cleeremans, 1993; Cleeremans and Dienes, 2008). Hence they match the complexities required for local melodic prediction/perception of the levels of pitch profiles as well as Narmour's (1990) principles. They have been found to converge towards efficient sparse representations of melodic structure (Agres et al., 2009). However, their learning curves have been found to be less steep than n-gram models and in performance they are largely outperformed by n-gram models (cf. Rohrmeier, 2009, 2010). On the other hand, Cartling (2008) provided some evidence that SRNs are able to perform predictions for simple (language like) context-free sequences, a finding which extends and partially revises Elman's, 1991 earlier results (see also Christiansen and Chater, 1999).

For the prediction of temporal structure, Large and Kolen (1999) have developed dynamic, cognitively motivated oscillator models which are able to adapt to metrical structure rapidly after a number of initial expressively timed events. Large and Palmer (2002) as well as Large and Kolen (1999) further showed that models of combined oscillators further predicted different metrical weights as implied by metrical models as in Lerdahl and Jackendoff (1983) in terms differently strong peaks of attention. However, Collins (2006) criticised its slow adaptation to tempo changes as well as the implied automaticity of beat tracking, which neglected the involved higher-order cognitive processes.

4.4. Modelling limitations

Most computational models provide and instantiate the link of learning or adaptive behaviour with prediction by exemplifying that different forms of statistical learning from exposure suffice to explain a variety of aspects of predictive behaviour. While most models are applied to cases of symbolic harmonic or melodic prediction, and are to some extent successful in predicting human behaviour, there are yet no overarching model architectures that integrate different complex musical structures or representations. More understanding is required with respect to how emergent features such as polyphony, harmony or form are modelled from basic features in probabilistic or connectionist methodologies.

While there is evidence that the syntax of (at least) Western tonal music involves nonlinear, hierarchical long-distance dependencies similar to language (Lerdahl and Jackendoff, 1983; Steedman, 1984, 1996; Rohrmeier, 2007a, 2011), most cognitive or predictive models of music do not capture such features. Though fragment models are very successful, nonlocal prediction contradicts the fundamental Markov assumption (Eq. (1)) and therefore creates a mismatch with structural models of music (as above). The indirect covering of nonlocal features through sufficiently large fragments is implausible based on the Zipf-distribution of musical fragments (Zanette, 2006; Rohrmeier, 2005; Rohrmeier and Cross, 2008) because they would require an unrealistic mass of observations from a corpus (see the discussion in Rohrmeier, 2011). However some sensitivity to non-local dependencies may be achieved by SRNs or HMMs. Yet still, the relationship between such models and algorithmic parsers, that would be required for a full processing of context-free complexity and the complex formation of prediction, remains open. While Jackendoff (1991) or Temperley (2001) discuss some cases of ambiguity and revision which involve a parsing process, the integration or such processes within neural frameworks require more computational as well as behavioural research.

5. Neuroscientific evidence

From a neuroscience perspective, a body of research explores neural responses to music-syntactic expectancy violations. In this respect the early right-anterior negativity (ERAN, see below for explanation) was taken as an electrophysiological reflection of such expectancy violations (Koelsch et al., 2000). While these findings strongly relate to prediction, so far the actual contribution of predictive processes to the generation of the ERAN has not been isolated (similar to the lack of evidence for predictive processing in the case of the mismatch negativity, MMN, but see also contributions on the MMN in this special issue).

5.1. Harmonic and melodic prediction

Neurophysiological studies using EEG and MEG showed that music-syntactically irregular chord functions elicit brain potentials with negative polarity that are maximal at around 150–350 ms after the onset of an irregular chord, and have a frontal/fronto-temporal scalp distribution, often with right-hemispheric weighting. In experiments with isochronous, repetitive stimulation (that is, in experiments, in which participants know when and which regular/irregular events may occur), this effect is maximal at around 150–200 ms over right anterior electrodes, and denoted as early right anterior negativity, or ERAN (for a review, see Koelsch, 2009). In experiments in which the position of irregular chords (Patel et al., 1998; Koelsch and Mulder, 2002) or improbable tones of melodies (Pearce et al., 2010b; Brattico et al., 2006) within a sequence is not known (and thus unpredictable), the negativity often has a longer latency, and a more posterior (centro-temporal) distribution (also referred to as right anterior-temporal negativity, or RATN; Patel et al., 1998). The ERAN elicited by irregular tones of melodies usually has a shorter peak latency than the ERAN elicited by irregular chord functions (e.g., Pearce et al., 2010b; Koelsch and Jentschke, 2010). Studies on violations of musical contour are not discussed here (for such studies see, e.g., Trainor et al., 2002; Fujioka et al., 2004; Schiavetto et al., 1999).

5.1.1. Local vs. hierarchical processing

As outlined above, theoretical accounts characterise music by local dependencies as well as some overarching context-free syntactic principles. So far, music psychological studies have hardly addressed this issue, and in most studies the music-syntactic violations that were introduced to investigate music-syntactic processing represented both local and hierarchical violations. For example, Neapolitan chords following a dominant that were used in some studies as a music-syntactically incorrect chord function at the end of a harmonic sequence (e.g., Koelsch, 2000; Koelsch et al., 2005; Leino et al., 2007) represent both a local violation (a Neapolitan chord does usually not follow a dominant) and a hierarchical violation (a harmonic sequence never ends on a Neapolitan chord). Thus, from such studies it is not clear whether the ERAN reflects local or hierarchical processing. However, it is likely that the ERAN (or a subcomponent of the ERAN) reflects at least in part hierarchical processing: E.g., in a study by Koelsch et al. (2007) an ERAN was elicited by a secondary dominant following the dominant at the end of a harmonic sequence. Locally, this chord transition does not represent a clear-cut violation (because in a different key context these chords can function as tonic–dominant progression, which is a normal chord progression), thus pointing to contributions of hierarchical processing to the ERAN (because this chord progression is only irregular in a wider context of harmonic functions within the chord sequence, in which the final chords function as dominant–secondary dominant progression, and not as tonic–dominant progression). Nevertheless, regular chord sequences consisted of tonic–subdominant–supertonic–dominant–tonic progressions, and it is still possible, that this chord progression was represented as harmonic schema or fragment in the brains of listeners, and that thus processing of a violation based on an n-gram model or another fragment model(and not hierarchical processing) elicited the ERAN. Thus, future experiments are required that directly aim at disentangling neural correlates of local and hierarchical music-syntactic processing (and the predictive processes that are possibly involved).

It is interesting to note that functional neuroimaging studies using such chord sequence paradigms (e.g.,Maess et al., 2001; Koelsch et al., 2002; Tillmann et al., 2003; Koelsch et al., 2005; Garza Villareal et al., 2011) suggest that music-syntactic processing involves the pars opercularis of the inferior frontal gyrus corresponding to Brodmann area (BA) 44 bilaterally, but with right-hemispheric weighting (for an fMRI study using melodies see Janata et al., 2002a). It seems likely that such involvement of (inferior) BA 44 (probably area 44v according to Amunts et al., 2010) in music-syntactic processing is due to the hierarchical processing of syntactic structure: This part of Broca's area appears to be involved in the hierarchical processing of syntax in language (Friederici et al., 2006; Makuuchi et al., 2009), the hierarchical processing of action sequences (e.g. Koechlin and Jubault, 2006; Fazio et al., 2009), and possibly also in the processing of hierarchically organised mathematical formulas and termini (although activation in the latter study cannot clearly be assigned to BA 44 or BA 45, Friedrich and Friederici, 2009). Note that these findings suggest that at least some cognitive operations of music-syntactic and language-syntactic processing (and neural populations mediating such operations) overlap, and are shared with the syntactic processing of actions, mathematical formulas, and other structures based on long-distance dependencies involving hierarchical organisation (phrase-structure grammar).

However, it appears that inferior BA 44 is not the only structure involved in music-syntactic processing: additional structures include the superior part of the pars opercularis (Koelsch et al., 2002), the anterior portion of the superior temporal gyrus (STG) (Koelsch et al., 2002, 2005), and ventral premotor cortex (PMCv; Janata et al., 2002b; Koelsch et al., 2002, 2005; Parsons, 2001). The PMCv (but not BA 44) appears to be important for the processing of musical structure of finite-state complexity. With regard to language, Friederici (2004) reported that activation foci of functional neuroimaging studies on the processing of long-distance hierarchies and transformations are located in the posterior IFG (with the mean of the coordinates reported in that article being located in the inferior pars opercularis), whereas activation foci of functional neuroimaging studies on the processing of local structural violations are located in the PMCv (see also Friederici et al., 2006; Makuuchi et al., 2009; Opitz and Kotz, in press). Moreover, patients with lesion in the PMCv show disruption of the processing of finite state, but not phrase-structure grammar (Opitz and Kotz, in press). This points to the involvement of PMC in the processing of local dependencies. However, whether the involvement of PMC in music-syntactic processing is due to the processing of local dependencies, or the generation/modification/updating of predictions, or both, remains to be investigated:

Activations of PMCv have been reported in a variety of functional imaging studies on auditory processing using musical stimuli, linguistic stimuli, auditory oddball paradigms, pitch discrimination tasks and serial prediction tasks, underlining the importance of these structures for the sequencing of structural information, the recognition of structure, and the prediction of sequential information (Janata and Grafton, 2003; Schubotz, 2007).

As mentioned above, in the above-mentioned experiments that used chord sequence paradigms to investigate the processing of harmonic structure, the music-syntactic processing of the chord functions probably involved processing of both local dependencies and phrase-structure grammar (involving long-distance dependencies). With regard to the processing of phrase-structure grammar, syntactic parsing requires the establishment and continuous update of the harmonic relation between a chord function and the context of preceding chord functions with respect to a maintained harmonic context model which is continuously refined or revised based on incoming information. In addition, we assume that the harmonic model is maintained in working memory in order to be updated when encountering new harmonic information, as well as when dealing with long-distance harmonic dependencies (or local harmonic insertions within a larger context, as it is possible within harmonic phrase-structure, Rohrmeier, 2011). The revision of harmonic phrase-structure required by irregular chord functions is more complex than for regular chord functions, and this difference in complexity is presumably reflected in a stronger activation of (inferior) BA 44 in response to irregular chords. On the other hand, the local transition probability from the penultimate to the final chord (finite-state complexity) is lower, e.g., for a dominant–supertonic progression than for a dominant–tonic progression (compare the empirical results by Rohrmeier and Cross, 2008), and the computation of the (less predicted) lower-probability progression is presumably reflected in a stronger activation of PMCv in response to irregular chords. The stronger activation of both BA 44 and PMCv appears to correlate with the perception of a music-syntactically irregular chord as “unexpected”.

Finally, it appears that the auditory cortex can already form predictions based on simple repetition (of an event or an event pattern), as evidenced by the mismatch negativity (MMN, see also contributions on the MMN in this issue): Although the MMN also receives contributions from the frontal cortex (that presumably contribute to prediction formation), data from patients with lesions of (dorsolateral) frontal cortex, and data from anaesthetised individuals (Koelsch et al., 2006), suggest that the MMN can be elicited without contributions from the frontal cortex (Koelsch, 2009). That is, it appears that the auditory cortex can establish a model of acoustic regularities, form predictions according to these regularities, and compare whether new acoustic input matches (or mismatches) with this input. However, it is difficult to prove that MMN potentials are due to predictive processes, and not merely to retrospective processing (for investigations on predictive processes in the auditory domain see also Bendixen et al., 2009; Minati et al., 2010).

6. Combining converging evidence: insights and open questions

As in other domains of computational modelling, music prediction is best modelled based on forms of statistical learning and massive prior exposure. Analogously, neural predictive information processing may be assumed to be mostly grounded on the availability of a powerful learning mechanism and memory capacities (in terms of an online and long-term pattern storage as well as a buffer during processing). Based on traditional accounts of memory (Baddeley, 1995, 1999), three memory components are presumably involved in music prediction: (a) the auditory sensory memory which acts as a buffer, which is stable with respect to attention and lasts only for a short time (in the range of seconds). It presumably forms local predictions (see Schroeger et al., this volume) that would be to a larger extent compatible with the research on priming discussed above. It would account for local predictions such as pitch repetitions, as well as auditory Gestalt formation that is potentially compatible with accounts such as Narmour's (1990) melodic principles. (b) Working memory that is active with respect to short-term context storage (also required for complex pattern matching with respect to larger, potentially syntactic or schematic musical structures stored in long term memory), manipulation of structural information, maintenance and update of the current contextual mental model of the musical structure (such as the stepwise completion of a phrase). Particularly, the integration of long-distance dependencies into structural contexts and event prediction requires working memory. (c) Long-term memory with respect to the retrieval and long-term storage of musical style-specific schemata/patterns or rules.

With respect to the computational modelling perspectives discussed here, however, such a conventional account needs to be amended in order to be able to capture the phenomenon of *online learning*, which relates to the constant adaptive activity of the brain to update predictions, interpretations and representations on-the-fly. During the course of a single piece, specific individual patterns may occur repeatedly and manipulate the listener patterns of expectation for this particular piece. For instance, Schumann's lied “Am leuchtenden Sommermorgen” repeatedly uses a German sixth chord revision of a (0,2,6,9) pitch class set until it reverses the normally preferred dominant seventh hearing of it as a surprise (towards the end of the piece). Such musical effects are common across various styles (particularly also Minimal Music). They require an active process of structural online learning, which has been integrated into the model architecture of IDyOM (Conklin and Witten, 1995; Pearce and Wiggins, 2006; Potter et al., 2007). Particularly, they distinguish between a long-term and a short-term model — however, the analogy of the short-term model with short-term memory does not exactly hold: while short-term memory is assumed to last only for a limited amount of time (unless the memorised information is activity maintained), the time-span of online learning during a musical piece is significantly larger. Hence, this form of online learning performance suggests a memory performance that partially undermines a strict common three-fold distinction of memory components or requires a new module. While computational models require the exact specification of the interaction and the weighting of different predictive (memory) components, this interaction has not much been explored or discussed in the neuroscientific literature.

We aimed to point out a number of caveats with respect to (musical) processing: While prediction is assumed to be ubiquitous for all features of musical structure, there are some difficulties with respect to its definition in terms of polyphonic music, feature interaction/integration and higher-order musical features.

While there are theoretical and some computational grounds to assume the processing of long-distance dependencies for music prediction, there is not much research yet that disentangles long-distance and local prediction methodologically and provides evidence for supporting or falsifying predictive effects of long-distance dependencies (for example, it is not yet clear to which extent the ERAN reflects local or hierarchical processing, or whether local and hierarchical processing is reflected in subcomponents of the ERAN). Such differences may touch upon the difference between context-free and finite-state processing as portrayed above. Similarly, musical expectancies might concern the prediction of an event that is about to appear (but does not necessarily need to be the subsequent event in the sequence). Such forms of non-subsequent event expectancies constitute additional challenges to experimental ERP paradigms or computational modelling.

Theoretical and cognitive reasoning render the assumed constant predictive activity of the mind very plausible and are supported by computational models that predict human experimental behaviour. It is difficult, however, to provide definite neural results that are unambiguously related to prediction (or prediction violations), and not merely to some reactive processes to the deviant stimulus. While most ERP experimental paradigms measure the responses to different forms of unexpected musical events, it is not absolutely certain whether these responses are due to predictive behaviour or to reactive or revision processes. If one would measure, for instance, skin conductance or heart rate with respect to a particular unexpected event, one would tend to relate observed measures as reactive and not predictive processes, although the result constitutes the same type of data as ERP responses. Even though the predictive processing explanation of ERP results is highly likely, it is difficult to design straight-forward experimental paradigms to disentangle it from reactive or revision processing.

The most promising perspective to elucidate the processes and representations involved in predictive music processing appears to be the combination of neuroscientific methodologies with computational modelling. This forces to make psychological "pen-and-paper-models" explicit (cf. Wiggins, 2011). The study by Pearce et al. (2010b) provides an important methodological link between predictive computational modelling and neurophysiological data by showing that probabilistic measures (high or low) of the melodic notes (as obtained by the computational model) were related to the ERAN amplitude. Similarly, another major research contribution by Janata et al. (2002a) and Janata (2009) identified neural correlates for the temporal dynamics of tonal structure as modelled by Toiviainen and Krumhansl (2003).

This promising line of research calls for future research with respect to (a) other musical structures (e.g. harmony), (b) expansion with respect to the information theoretical predictors (e.g. towards continuous probability representations, or entropy-based measures), or (c) further neuroscientific methods (e.g. fMRI). Further computational research that explores the relationship between syntactic descriptive models and cognitive models is required in order to yield experimentally testable predictions concerning this intricate range of questions.

Acknowledgements

We would like to thank Moritz Lehne, Renzo Torrecuso and two reviewers for many helpful comments on the manuscript.

References

Agawu, K., 1995. African Rhythm. Cambridge University Press, Cambridge, A Northern Ewe perspective.
Agres, K.R., DeLong, J.E., Spivey, M., 2009. The sparsity of simple recurrent networks in musical structure learning. In: van Rijn, N.T.H. (Ed.), Proceedings of the 31th Annual Conference of the Cognitive Science Society, pp. 3099–3104.
Allan, M., Williams, C.K.I., 2005. Harmonising Chorales by Probabilistic Inference. Advances in Neural Information Processing Systems, Vol. 17.
Ames, C., 1987. Automated composition in retrospect: 1956–1986. Leonardo 20 (2), 169–185.
Ames, C., 1989. Themarkov process as a compositional model. A survey and tutorial. Leonardo 22 (2), 175–187.
Amunts, K., Lenzen, M., Friederici, A., Schleicher, A., Morosan, P., Palomero-Gallagher, N., Zilles, K., 2010. Broca's region: novel organizational principles and multiple receptor mapping. PLoS Biology 8 (9), e1000489.
Baddeley, A.D., 1995. Working Memory. Clarendon Press, Oxford.
Baddeley, A.D., 1999. Essentials of Human Memory. Psychology Press, Hove.
Bendixen, A., Schröger, E., Winkler, I., 2009. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. The Journal of Neuroscience 29 (26), 8447–8451.
Bharucha, J.J., 1998. Neural nets, temporal composites and tonality. In: Deutsch, D. (Ed.), The Psychology of Music. Academic Press, New York.
Bharucha, J., Krumhansl, C., 1983. The representation of harmonic structure in music: hierarchies of stability as a function of context. Cognition 13, 63–102.
Bharucha, J.J., Stoeckig, K., 1986. Reaction time and musical expectancy: priming of chords. Journal of Experimental Psychology. Human Perception and Performance 12, 403–410.
Bharucha, J.J., Stoeckig, K., 1987. Priming of chords: spreading activation or overlapping frequency spectra? Perception & Psychophysics 41, 519–524.
Bharucha, J.J., Todd, P.M., 1989. Modeling the perception of tonal structure with neural nets. Computer Music Journal 13 (4), 44–53.
Bharucha, J.J., Curtis, M., Paroo, K., 2006. Varieties of musical experience. Cognition 100, 131–172.
Bharucha, J., Curtis, M., Paroo, K., 2011. Musical communication as alignment of non-propositional brain states. In: Rebuschat, P., Rohrmeier, M., Hawkins, J., Cross, I. (Eds.), Music and Language as Cognitive Systems. Oxford University Press, Oxford, pp. 139–155.
Bigand, E., Pineau, M., 1997. Global context effects on musical expectancy. Perception & Psychophysics 59, 1098–1107.
Bigand, E., Poulin-Charronnat, B., 2006. Are we "experienced listeners"? a review of the musical capacities that do not depend on formal musical training. Cognition 100, 100–130.
Bigand, E., Madurell, F., Tillmann, B., Pineau, M., 1999. Effect of global structure and temporal organization on chord processing. Journalof Experimental Psychology. Human Perception and Performance 25 (1), 184–197.
Bigand, E., Tillmann, B., Poulin-Charronnat, B., Manderlier, D., 2005. Repetition priming: is music special? The Quarterly Journal of Experimental Psychology 58A (8), 1347–1375.
Boucher, L., Dienes, Z., 2003. Two ways of learning associations. Cognitive Science 27, 807–842 Jan.
Brattico, E., Tervaniemi, M., Naatanen, R., Peretz, I., 2006. Musical scale properties are automatically processed in the human auditory cortex. Brain Research 1117 (1), 162–174.
Bubic, A., Cramon, D.Y.V., Schubotz, R.I., 2010. Prediction, cognition and the brain. Frontiers in Human Neuroscience 4.
Budge, H., 1943. A Study of Chord Frequencies. Teacher College.
Byros, V., 2009. Towards an "archaeology" of hearing: schemata and eighteenth-century consciousness. Musica Humana 1 (2), 235–306.
Cartling, B., 2008. On the implicit acquisition of a context-free grammar by a simple recurrent neural network. Neurocomputing 71, 1527–1537.
Chomsky, N., 1956. Three models for the description of language. IRE Transactions on Information Theory 2, 113–124.
Christiansen, M., Chater, N., 1999. Toward a connectionist model of recursion in human linguistic performance. Cognitive Science 23, 157–205.
Clarke, E., 2005. Ways of Listening. An Ecological Approach to the Perception of Musical Meaning. Oxford University Press, New York.
Cleeremans, A., 1993. Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing. MIT Press, Cambridge, MA.
Cleeremans, A., Dienes, Z., 2008. Computational models of implicit learning. In: Sun, R. (Ed.), Cambridge Handbook of Computational Psychology. Cambridge University Press, pp. 396–421.
Cohen, J.E., 1962. Information theory and music. Behavioural Science 7 (2), 137–163.
Collins, N., 2006. Towards autonomous agents for live computer music: Realtime machine listening and interactive music systems. Ph.D. thesis, Centre for Science and Music, Faculty of Music, University of Cambridge.
Conklin, D., 2002. Representation and discovery of vertical patterns in music. In: Anagnostopoulou, C., Ferrand, M., Smaill, A. (Eds.), Proceedings of the Second International Conference of Music and Artificial Intelligence, pp. 32–42.
Conklin, D., 2010. Discovery of distinctive patterns in music. Intelligent Data Analysis 14 (5), 547–554.
Conklin, D., Anagnostopoulou, C., 2001. Representation and discovery of multiple viewpoint patterns. Proceedings of the International Computer Music Conference. International Computer Music Association, Havana, Cuba, pp. 479–485.
Conklin, D., Witten, I., 1995. Multiple viewpoint systems for music prediction. Journal of New Music Research 24 (1), 51–73.
Cross, I., 2003. Music and biocultural evolution. In: Clayton, M., Herbert, T., Middleton, R. (Eds.), The Cultural Study of Music: A Critical Introduction. Routledge, pp. 19–30.

Cuddy, L.L., Lunney, C.A., 1995. Expectancies generated by melodic intervals: Perceptual judgements of continuity. Perception & Psychophysics 57, 451–462.

Cuddy, L.L., Thompson, W.F., 1992. Asymmetry of perceived key movement in chorale sequences: converging evidence from a probe-tone analysis. Psychological Research 54 (2), 51–59.

Dennett, D., 1996. Kinds of Minds. Basic Books, New York.

Dowling, W.J., 1978. Scale and contour: two components of a theory of memory for melodies. Psychological Review 85, 341–354.

Eberlein, R., 1994. Die Entstehung der tonalen Klangsyntax (The Emergence of Tonal-Harmonic Syntax). Peter Lang, Frankfurt.

Eerola, T., 2003. The dynamics of musical expectancy. cross-cultural and statistical approaches to melodic expectations. Ph.D. thesis, University of Jyväskylä.

Eerola, T., Louhivuori, J., Lebaka, E., 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.

Elman, J.L., 1990. Finding structure in time. Cognitive Science 14 (2), 179–211.

Elman, J., 1991. Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning 7, 195–225.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A., Pozzo, T., Calzolari, F., Granieri, E., Fadiga, L., 2009. Encoding of human action in Broca's area. Brain 132 (7), 1980.

Franklin, J.A., 2006. Recurrent neural networks for music computation. INFORMS Journal on Computing 18 (3), 321–338.

Friederici, A., 2004. Processing local transitions versus long-distance syntactichierarchies. Trends in Cognitive Sciences 8 (6), 245–247.

Friederici, A., Bahlmann, J., Heim, S., Schubotz, R., Anwander, A., 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proceedings of the National Academy of Sciences 103 (7), 2458.

Friedrich, R., Friederici, A., 2009. Mathematical logic in the human brain: syntax. PLoS One 4 (5).

Fujioka, T., Trainor, L., Ross, B., Kakigi, R., Pantev, C., 2004. Musical training enhances automatic encoding of melodic contour and interval structure. Journal of Cognitive Neuroscience 16 (6), 1010–1021.

Gang, D., Lehmann, D., Wagner, N., 1998. Tuning neural network for harmonizing melodies in real-time. International Computer Music Conference (ICMC98).

Garza Villareal, E., Brattico, E., Leino, S., Østergaard, L., Vuust, P., 2011. Distinct neural responses to chord violations: A multiple source analysis study. Brain Research 1389, 103–114.

Gerstner, W., Kistler, W., 2002. Spiking Neuron Models. Cambridge University Press.

Gjerdingen, R., 2007. Music in the Galant Style. Oxford University Press.

Grewe, O., Nagel, F., Kopiez, R., Altenmüller, E., 2007. Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. Music Perception 24 (3), 297–314.

Griffith, N., Todd, P.M. (Eds.), 1999. Musical Networks: Parallel Distributed Perception and Performance. MIT Press, Cambridge, MA.

Hiller, L., Bean, C., 1966. Information theory analyses of four sonata expositions. Journal of Music Theory 10 (1), 96–137.

Hiller, L., Fuller, R., 1967. Structure and information structure and information in Weberns Symphonie, op. 21. Journal of Music Theory 11 (1), 60–115.

Hopfield, J., 1982. Neural networks and physical systems with emergent collective computational abilities. National Academy of Science 79, 2554–2558.

Huron, D., 2006. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge, Massachusetts.

Ichishita, T., Fujii, R.H., 2007. Performance evaluation of a temporal sequence learning spiking neural network. Seventh International Conference on Computer and Information Technology (CIT 2007), pp. 616–620.

Jackendoff, R., 1991. Musical parsing and musical affect. Music Perception 9, 199–230.

Janata, P., 2009. The neural architecture of music-evoked autobiographical memories. Cerebral Cortex 19 (11), 2579.

Janata, P., Grafton, S.T., 2003. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. Nature Neuroscience 6 (7), 682–687.

Janata, P., Birk, J., Van Horn, J., Leman, M., Tillmann, B., Bharucha, J., 2002a. The cortical topography of tonal structures underlying Western music. Science 298 (5601), 2167.

Janata, P., Tillmann, B., Bharucha, J., 2002b. Listening to polyphonic music recruits domain-general attention and working memory circuits. Cognitive, Affective, & Behavioral Neuroscience 2 (2), 121.

Keller, P.E., 2008. Joint action in music performance. In: Morganti, F., Carassa, A., Riva, G. (Eds.), Enacting Intersubjectivity: A Cognitive and Social Perspective to the Study of Interactions. IOS Press, Amsterdam, pp. 205–221.

Keller, P., Koch, I., 2008. Action planning in sequential skills: relations to music performance. The Quarterly Journal of Experimental Psychology 61, 275–291.

Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. Neuron 50 (6), 963–974.

Koelsch, S., 2000. Brain and Music — A Contribution to the Investigation of Central Auditory Processing with a New Electrophysiological Approach. Risse, Leipzig.

Koelsch, S., 2009. Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. Psychophysiology 46 (1), 179–190.

Koelsch, S., 2010. Towards a neural basis of music-evoked emotions. Trends in Cognitive Sciences 14 (3), 131–137.

Koelsch, S., Jentschke, S., 2010. Differences in electric brain responses to melodies and chords. Journal of Cognitive Neuroscience 22 (10), 2251–2262.

Koelsch, S., Mulder, J., 2002. Electric brain responses to inappropriate harmonies during listening to expressive music. Clinical Neurophysiology 113 (6), 862–869.

Koelsch, S., Gunter, T.C., Friederici, A.D., Schröger, E., 2000. Brain indices of music processing: "non-musicians" are musical. Journal of Cognitive Neuroscience 12 (3), 520–541.

Koelsch, S., Gunter, T.C., von Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: a cortical "language-network" serves the processing of music. NeuroImage 17, 956–966.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., Schlaug, G., 2005. Adults and children processing music: an fMRI study. NeuroImage 25 (4), 1068–1076.

Koelsch, S., Heinke, W., Sammler, D., Olthoff, D., 2006. Auditory processing during deep propofol sedation and recovery from unconsciousness. Clinical Neurophysiology 117 (8), 1746–1759.

Koelsch, S., Jentschke, S., Sammler, D., Mietchen, D., 2007. Untangling syntactic and sensory processing: an ERP study of music perception. Psychophysiology 44 (3), 476–490.

Kohonen, T., 1995. Self-Organizing Maps. Springer-Verlag, Berlin.

Krumhansl, C.L., 1979. The psychological representation of musical pitch in a tonal context. Cognitive Psychology 11, 346–374.

Krumhansl, C., 1990. Cognitive Foundations of Musical Pitch. Oxford University Press, USA.

Krumhansl, C.L., 1995. Effects of musical context on similarity and expectancy. Systematische Musikwissenschaft 3, 211–250.

Krumhansl, C.L., Keil, F.C., 1982. Acquisition of the hierarchy of tonal functions in music. Memory and Cognition 10, 243–251.

Krumhansl, C.L., Kessler, E.J., 1982. Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. Psychological Review 89, 334–368.

Krumhansl, C.L., Bharucha, J., Castellano, M., 1982. Key distance effects on perceived harmonic structure in music. Perception & Psychophysics 32 (2), 96–108.

Krumhansl, C.L., Louhivuori, J., Toiviainen, P., Järvinen, T., Eerola, T., 1999. Melodic expectation in Finnish spiritual folk hymns: convergence of statistical, behavioral, and computational approaches. Music Perception 17, 151–196.

Krumhansl, C.L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., Louhivuori, J., 2000. Cross-cultural music cognition: cognitive methodology applied to north sami yoiks. Cognition 76 (1), 13–58.

Large, E.W., 2010a. Dynamics of musical tonality. In: Huys, R., Jirsa, V.K. (Eds.), Nonlinear Dynamics in Human Behavior. Springer-Verlag, Berlin, pp. 193–211.

Large, E.W., 2010b. Neurodynamics of music. In: Jones, M.R., Fay, R.R., Popper, A.N. (Eds.), Springer Handbook of Auditory Research. : Music Perception, 36. Springer, New York, pp. 201–231.

Large, E.W., 2011. Musical tonality, neural resonance and hebbian learning. In: Agon, C., Amiot, E., Andreatta, M., Assayag, G., Bresson, J., Mandereau, J. (Eds.), Mathematics and Computation in Music. Springer-Verlag, Berlin, pp. 115–125.

Large, E.W., Kolen, J.F., 1999. Resonance and the perception of musical meter. In: Griffith, N., Todd, P. (Eds.), Musical Networks: Parallel Distributed Perception and Performance. MIT Press, Cambridge, MA, pp. 279–312.

Large, E.W., Palmer, C., 2002. Perceiving temporal regularity in music. Cognitive Science 26, 1–37.

Leino, S., Brattico, E., Tervaniemi, M., Vuust, P., 2007. Representation of harmony rules in the human brain: further evidence from event-related potentials. Brain Research 1142, 169–177.

Leman, M. (Ed.), 1997. Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology. Springer, Berlin and Heidelberg.

Lerdahl, F., 2001. Tonal Pitch Space. Oxford University Press, New York.

Lerdahl, F., Jackendoff, R., 1983. A Generative Theory of Tonal Music. MIT Press, Cambridge, MA.

London, J., 2004. Hearing in Time. Oxford University Press.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in the area of Broca: an MEG-study. Nature Neuroscience 4 (5), 540–545.

Makuuchi, M., Bahlmann, J., Anwander, A., Friederici, A., 2009. Segregating the core computational faculty of human language from working memory. Proceedings of the National Academy of Sciences 106 (20), 8362.

Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA.

Manzara, L.C., Witten, I.H., James, M., 1992. On the entropy of music: an experiment with Bach chorale melodies. Leonardo 2, 81–88.

Marcus, S., 2003. The Garland encyclopedia of world music. The Middle East. Routledge, New York, Ch. The Eastern Arab System of Melodic Modes: A Case Study of Maqam Bayyati, pp. 33–44.

Marr, D., 1982. Vision: A Computational Approach. Freeman & Co., San Francisco.

Maxwell, J.B., Pasquier, P., Eigenfeldt, A., 2009. Hierarchical sequential memory for music: a cognitive model. International Society for Music Information Retrieval Conference (ISMIR), pp. 429–434.

McClelland, J.L., 2009. The place of modeling in cognitive science. Topics in Cognitive ScienTopics in Cognitive Science 1 (1), 11–38.

Meyer, L., 1956. Emotion and Meaning in Music. University of Chicago Press, London.

Minati, L., Salvatoni, L., Rosazza, C., Pietrocini, E., Visani, E., Panzica, F., Scaioli, V., Avanzini, G., Franceschetti, S., 2010. Event-related potential (erp) markers of melodic processing: the n2 component is modulated by structural complexity, not by melodic 'meaningfulness'. Brain Research Bulletin 83 (1–2), 23–28.

Mozer, M.C., 1994. Neural network music composition by prediction: exploring the benefits of psychoacoustic constraints and multiscale processing. Connection Science 6 (2–3), 247–280.

Murphy, K., 2002. Dynamic bayesian networks: Representation, inference and learning. Ph.D. thesis, UC Berkeley.

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

Narmour, E., 1999. Hierarchical expectation and musical style, In: Deutsch, D. (Ed.), The Psychology of Music, 2nd Edition. Academic Press, San Diego, California, pp. 441–472.

Neely, J., 1976. Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory processes. Memory & Cognition 4 (5), 648–654.

Neuhaus, C., Knösche, T., Friederici, A., 2006. Effects of musical expertise and boundary markers on phrase perception in music. Journal of Cognitive Neuroscience 18 (3), 472–493.

Neuwirth, M., 2008. Das konzept der expektanz in der musikalischen analyse: Möglichkeiten und probleme einer kognitive orientierten musikanalyse. In: Sprick, J.P., Bahr, R., von Troschke, M. (Eds.), Musiktheorie in Kontext. Weidler, Hamburg, pp. 557–573.

Opitz, B., Kotz, S., in press. Ventral Premotor Cortex Lesions disrupt learning of sequential grammatical structures. Cortex. http://dx.doi.org/10.1016/j.cortex.2011.02.013.

Paiement, J.-F., Bengio, S., Eck, D., 2009. Probabilistic models for melodic prediction. Artificial Intelligence 173 (14), 1266–1274.

Parsons, L., 2001. Exploring the functional neuroanatomy of music performance, perception, and comprehension. Annals of the New York Academy of Sciences 930, 211–231.

Patel, A., Gibson, E., Ratner, J., Besson, M., Holcomb, P., 1998. Processing syntactic relations in language and music: an event-related potential study. Journal of Cognitive Neuroscience 10 (6), 717–733.

Pearce, M., 2005. The construction and evaluation of statistical models of melodic structure in music perception and composition. Ph.D. thesis, City University, London.

Pearce, M.T., Wiggins, G.A., 2004. Improved methods for statistical modelling of monophonic music. Journal of New Music Research 33 (4), 367–385.

Pearce, M.T., Wiggins, G.A., 2006. Expectation in melody: the influence of context and learning. Music Perception 23 (5), 377–405.

Pearce, M., Wiggins, G., in press. Auditory expectation: The information dynamics of music perception and cognition. Topics in Cognitive Science.

Pearce, M.T., Müllensiefen, D., Wiggins, G.A., 2010a. The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. Perception 39, 1367–1391.

Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., Bhattacharya, J., 2010b. Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. NeuroImage 50, 302–313.

Perruchet, P., Pacteau, C., 1990. Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. Journal of Experimental Psychology. General 119 (3), 264–275.

Perruchet, P., Vinter, A., 1998. Parser: a model of word segmentation. Journal of Memory and Language 39, 246–263.

Pinkerton, R.C., 1956. Information theory and melody. Scientific American 194 (2), 77–86.

Piston, W., 1948. Harmony. W.W.Norton & Company, New York.

Ponsford, D., Wiggins, G., Mellish, C., 1999. Statistical learning of harmonic movement. Journal of New Music Research 28, 150–177.

Pothos, E.M., 2007. Theories of artificial grammar learning. Psychological Bulletin 133 (2), 227–244.

Potter, K., Wiggins, G.A., Pearce, M.T., 2007. Towards greater objectivity in music theory: information-dynamic analysis of minimalist music. Musicae Scientiae 11 (2), 295–322.

Rabiner, L., 1989. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 257–286 Jan.

Raphael, C., 2010. Music plus one and machine learning. Machine Learning, Proceedings of the Twenty-Seventh International Conference (ICML 2010).

Raphael, C., Stoddard, J., 2004. Functional analysis using probabilistic models. Computer Music Journal 28 (3), 45–52.

Reis, B. Y., 1999. Simulating music learning with autonomous listening agents: Entropy, ambiguity and context. Ph.D. thesis, Computer Laboratory, University of Cambridge, UK., computer Laboratory.

Rohrmeier, M., 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.

Rohrmeier, M., 2007a. A generative grammar approach to diatonic harmonic structure. In: Georgaki, Kouroupetroglou, A. (Eds.), Proceedings of the 4th Sound and Music Computing Conference, pp. 97–100.

Rohrmeier, M., 2007b. Modelling dynamics of key induction in harmony progressions. In: Spyridis, Georgaki, Kouroupetroglou, Anagnostopoulou (Eds.), Proceedings of the 4th Sound and Music Computing Conference, pp. 82–89.

Rohrmeier, M., 2009. Learning on the fly. Computational analyses of an unsupervised online-learning effect in artificial grammar learning. Proceedings of the 9th International Conference on Cognitive Modelling.

Rohrmeier, M., 2010. Implicit learning of musical structure: Experimental and computational modelling approaches. Ph.D. thesis, University of Cambridge.

Rohrmeier, M., 2011. Towards a generative syntax of tonal harmony. Journal of Mathematics and Music 5 (1), 35–53.

Rohrmeier, M., Cross, I., 2008. Statistical properties of harmony in Bach's chorales. Proceedings of the 10th International Conference on Music Perception and Cognition, pp. 619–627.

Rohrmeier, M., Rebuschat, P., Cross, I., 2011. Incidental and online learning of melodic structure. Consciousness and Cognition 20, 214–222.

Schellenberg, G., 1997. Simplifying the implication-realization model of melodic expectancy. Music Perception 14, 295–318.

Schenker, H., 1935. Der Freie Satz. Neue musikalische Theorien und Phantasien. Margada, Liège, Belgium.

Schiavetto, A., Cortese, F., Alain, C., 1999. Global and local processing of musical sequences: an event-related brain potential study. Neuroreport 10 (12), 2467.

Schmuckler, M., 1989. Expectation and music: investigation of melodic and harmonic processes. Music Perception 7, 109–150.

Schoenberg, A., 1978. Theory of Harmony. University of California Press, Berkeley. r. E. Carter, trans. (Originally published, 1922.).

Schubotz, R., 2007. Prediction of external events with our motor system: towards a new framework. Trends in Cognitive Sciences 11 (5), 211–218.

Servan-Schreiber, D., Anderson, J., 1990. Learning artificial grammars with competitive chunking. Journal of Experimental Psychology: Learning, Memory, and Cognition 16, 592–608.

Sloboda, J., O Neill, S., Ivaldi, A., 2001. Functions of music in everyday life: an exploratory study using the experience sampling method. Musicae Scientiae 5 (1), 9–32.

Steedman, M., 1984. A generative grammar for jazz chord sequences. Music Perception 2 (1), 52–77.

Steedman, M.J., 1996. The blues and the abstract truth: music and mental models. In: Garnham, A., Oakhill, J. (Eds.), Mental Models in Cognitive Science. Erlbaum, Mahwah, NJ, pp. 305–318.

Steinbeis, N., Koelsch, S., Sloboda, J., 2006. The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. Journal of Cognitive Neuroscience 18 (8), 1380–1393.

Stevens, C., Latimer, C., 1992. A comparison of connectionist models of music recognition and human performance. Minds and Machines 2, 379–400.

Stevens, C., Latimer, C., 1997. Music recognition: an illustrative application of a connectionist model. Psychology of Music 25, 161–185.

Tekman, H., Bharucha, J.J., 1998. Implicit knowledge versus psychoacoustic similarity in priming of chords. Journal of Experimental Psychology. Human Perception and Performance 24, 252–260.

Temperley, D., 2001. The Cognition of Basic Musical Structures. MIT Press, Cambridge, MA.

Temperley, D., Sleator, D., 1999. Modeling meter and harmony: a preference rule approach. Computer Music Journal 15 (1), 10–27.

Tillmann, B., 2005. Implicit investigations of tonal knowledge in nonmusician listeners. Annals of the New York Academy of Science 1060, 100–110.

Tillmann, B., Bharucha, J.J., 2002. Effect of harmonic relatedness on the detection of temporal asynchronies. Perception & Psychophysics 64 (4), 640–649.

Tillmann, B., Bharucha, J.J., Bigand, E., 2000. Implicit learning of tonality: a self-organizing approach. Psychological Review 107 (4), 885–913.

Tillmann, B., Bharucha, J.J., Bigand, E., 2001. Implicit learning of regularities in western tonal music by self-organisation. In: French, R.M., Sougne, J.P. (Eds.), Connectionist Models of Learning, Development and Evolution. Springer, Heidelberg.

Tillmann, B., Janata, P., Bharucha, J., 2003. Activation of the inferior frontal cortex in musical priming. Cognitive Brain Research 16 (2), 145–161.

Toiviainen, P., 1996. Optimizing auditory images and distance metrics for self-organizing timbre maps. Journal of New Music Research 25 (1), 1–30.

Toiviainen, P., Krumhansl, C.L., 2003. Measuring and modeling real-time responses to music: the dynamics of tonality induction. Perception 32 (6), 741–766.

Tojo, S., Oka, Y., Nishida, M., 2006. Analysis of chord progression by HPSG. Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications. ACTA Press, Anaheim, CA, USA, pp. 305–310. Innsbruck, Austria.

Trainor, L., McDonald, K., Alain, C., 2002. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. Journal of Cognitive Neuroscience 14 (3), 430–442.

Whorley, R., Wiggins, G. A., Rhodes, C. S., Pearce, M. T., 2010. Development of techniques for the computational modelling of harmony. In: et al., V. (Ed.), Proceedings of the International Conference on Computational Creativity. Lisbon.

Wiggins, G., 2011. Computational models of music. In: Rebuschat, P., Rohrmeier, M., Cross, I., Hawkins, J. (Eds.), Music and Language as Cognitive Systems. Oxford University Press, Oxford, pp. 169–188.

Witten, I.H., Bell, T., 1991. The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression. IEEE Transactions on Information Theory 37 (4), 1085–1094.

Witten, I.H., Manzara, L.C., Conklin, D., 1994. Comparing human and computational models of music prediction. Computer Music Journal 18 (1), 70–80.

Youngblood, J.E., 1958. Style as information. Journal of Music Theory 2, 24–35.

Zanette, D.H., 2006. Zipf's law and the creation of musical context. Musicae Scientiae 10, 3–18.