TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 4 (2012) 625–652

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2012.01214.x

Auditory Expectation: The Information Dynamics of Music Perception and Cognition

Marcus T. Pearce and Geraint A. Wiggins

Centre for Cognition, Computation and Culture, Goldsmiths, University of London
Centre for Digital Music, Queen Mary, University of London

Received 30 September 2010; received in revision form 23 June 2011; accepted 25 July 2011

Abstract

Following in a psychological and musicological tradition beginning with Leonard Meyer, and continuing through David Huron, we present a functional, cognitive account of the phenomenon of expectation in music, grounded in computational, probabilistic modeling. We summarize a range of evidence for this approach, from psychology, neuroscience, musicology, linguistics, and creativity studies, and argue that simulating expectation is an important part of understanding a broad range of human faculties, in music and beyond.

Keywords: Expectation; Probabilistic modeling; Prediction; Musical melody; Pitch; Segmentation; Aesthetics; Creativity

1. Introduction

> Once a musical style has become part of the habit responses of composers, performers, and practiced listeners, it may be regarded as a complex system of probabilities … Out of such internalized probability systems arise the expectations—the tendencies—upon which musical meaning is built. (Meyer, 1957, p. 414)

The ability to anticipate the future is a fundamental property of the human brain (Dennett, 1991). Expectations play a role in a multitude of cognitive processes from sensory perception, through learning and memory, to motor responses and emotion generation. Accurate expectations allow organisms to respond to environmental events faster and more appropriately and to identify incomplete or ambiguous perceptual input. To deal appropriately with

Correspondence should be sent to Marcus T. Pearce, School of Electronic Engineering and Computer Science, Queen Mary, University of London, E1 4NS, UK. E-mail: marcus.pearce@eecs.qmul.ac.uk

changes in the environment, expectations must be grounded in processes of learning and memory. Because of the important implications of accurate expectations for survival, expectations are thought to be closely related to emotion and reward-related neural circuits. This paper is about the role that cognitive processes of expectation play in music cognition.

In his seminal book, *Emotion and Meaning in Music*, Meyer (1956) aimed to link musical structure with the communication of emotion and meaning without appealing to referential semantics. The link Meyer identified was the way in which certain musical structures create perceptual expectations for forthcoming musical structures. By manipulating these implications, a composer may communicate emotions ranging from pleasure when expectations are satisfied, to disappointment when they are violated, frustration when they are delayed, or tension when implications are ambiguous. Meyer (1957), quoted above, expressed the cognitive process of musical expectation as a mechanism of learning and generating conditional probabilities, linking musical meaning with information-theoretic processing of musical structure.

Meyer's approach has been developed in three ways: Musicologists like Narmour (1990, 1992) have elaborated its musical aspects; cognitive scientists have studied computational models of perceptual expectations in music; and behavioral and neural processes involved in musical expectation have been empirically investigated. From a psychological perspective, musical expectations have been found to influence recognition memory for music (Schmuckler, 1997), the production of music (Carlsen, 1981; Schmuckler, 1990; Thompson, Cuddy, & Plaus, 1997), the perception of music (Cuddy & Lunny, 1995; Krumhansl, 1995; Schellenberg, 1996; Schmuckler, 1989), the transcription of music (Unyk & Carlsen, 1987), and emotional responses to music (Steinbeis, Koelsch, & Sloboda, 2006). While most empirical research has examined the influence of melodic pitch structure, expectations in music have also been examined in relation to rhythmic and metrical structure (Jones, 1987; Jones & Boltz, 1989; Large & Jones, 1999) as well as harmonic structure (Bharucha, 1987; Schmuckler, 1989; Steinbeis et al., 2006; Tillmann, Bharucha, & Bigand, 2000; Tillmann, Bigand, & Pineau, 1998). Many of Meyer's proposals about the relationship between expectation and emotion in music remain to be tested empirically (Juslin & Västfjäll, 2008) and it is only recently that information theory has been used to investigate expectations in any of these areas.

In this paper, we present our perspective on Meyer's idea and its implications, aiming for an over-arching theory, grounded in evolutionary process and contextualized within a larger, explicitly layered model of cognition. The core idea is that of information transmission via musical structure during the listening experience, in context of knowledge shared between producer and listener. We explore the cognitive processes involved in this transmission and their relationship with more general processes in human cognition. We focus on the cognitive processes that generate expectations for how a musical sequence will continue in the future: What will be the properties (pitch, timing, etc.) of the next musical event? (see also Tillmann, 2011). To discuss the effect of context in music cognition, one needs also an account of how that contextual knowledge is acquired: We use online implicit learning (see Rohrmeier & Rebuschat, 2011, for a review) and place information theory at the core of cognitive processing.

Our approach is firmly based in computational modeling, and therefore we develop our exposition around a successful model of musical pitch expectation, which simulates implicit learning and generates predictions from what is learned. The information-dynamic properties of this model are then shown to predict structural segmentation of musical melody by listeners. Thus, one theory is shown to predict two different aspects of a perceptual phenomenon. Further work demonstrates a relationship between information content, which can be consciously reported, and neural behavior during listening, suggesting a direct link between information dynamics, auditory expectation, and the experience of musical listening. We conclude with several more speculative sections, which cover preliminary research on the potential contribution of expectation to aesthetics and creativity; the aim here is to identify fruitful research topics for the short- and medium-term future.

Overall, our aim is to argue for the paramount importance of auditory expectation in the experience of music, and to propose credible cognitive mechanisms by which such experience may be generated, while also setting out the next steps in this research program. In doing so, we summarize experimental results from existing published studies.

2. Learning, memory, and expectation for music

2.1. Evolutionary context

To begin, we ground our argument in an evolutionary context by asking what expectations are for. We avoid the debate about music's evolutionary selection pressure (Cross, 2007; Fitch, 2006; Justus & Hutsler, 2005; McDermott & Hauser, 2005; Pinker, 1995; Wallin, Merker, & Brown, 1999), but the cognitive processes and models we propose should at least be consistent with evolutionary theory. Whether these functions are adapted or exapted does not matter to the current work.

We assume that cognitive mechanisms underlying musical expectation are specific instances of those supporting general auditory expectation. Cognitive processes of top-down expectation confer several potential advantages on an organism. By anticipating what is likely to appear in a given context, an organism can reduce orienting responses (Zajonc, 1968; Huron, 2006), identify incomplete, noisy or ambiguous stimuli (Summerfield & Egner, 2009), and prepare faster and more appropriate responses (Schultz, Dayan, & Montague, 1997).

Failures of expectation can be fatal, so organisms should be motivated to expect as accurately as possible, with two consequences. First, the life-preserving advantage of avoiding failure entails that successful organisms must *pre-emptively* experience non-fatal penalties for predictive failures, and rewards for predictive successes, through negative and positive emotions, respectively (Huron, 2006).

Second, in complex, changing auditory environments, organisms that adapt their expectations to experience are favored. This is why our account is based on learning. Instead of innate representational rules, we invoke innate, general-purpose learning mechanisms, imposing architectural, not representational, constraints on cognitive development (Elman et al., 1996). Given exposure to appropriate stimuli, these learning mechanisms acquire domain-specific representations and behavior. We regard these learning mechanisms as general-purpose processes in auditory cognition, and not specific to music. Eight-month-old infants (Saffran, Johnson, Aslin, & Newport 1999) and non-human primates (Hauser, Aslin, & Newport, 2001) exhibit learning of statistical associations between auditory events. We ask: What mechanism enables learning?

Therefore, we seek a mechanism for generating expectations, which learns through experience with neither oracular top-down assistance nor prior music-theoretical knowledge.

We also consider the evolutionary status of the auditory features over which musical expectations operate. From our theoretical perspective, we need a perceptual dimension of pitch, which behaves mathematically as a linearly ordered abelian (or commutative) group[1] (Wiggins et al., 1989), and a time dimension, with the same basic mathematical behavior. Fundamental pitch features in human music (e.g., octave equivalence: Greenwood, 1996) are shared by non-human species, so can be assumed as pre-extant. Similarly, we assume the ability to perceive repeating time periods as a given (Large, Almonte, & Velasco, 2010). We believe that both these faculties are exapted for music, as organisms exhibiting them evolved long before music was exhibited by humans, although the human capacity for consistent, deliberate rhythmic entrainment does seem to be unique (Patel, Iversen, Bregman, & Schulz, 2009; Schachner, Brady, Pepperberg, & Hauser, 2009). Other musical dimensions, (e.g., timbre, dynamics) are compatible with our approach but remain to be investigated within it.

2.2. Background: Information theory

Hartley (1928) began research in information theory, although the first significant developments arrived in Claude Shannon's seminal mathematical theory of communication (Shannon, 1948). This work inspired interest in information theory throughout the 1950s, in fields ranging from psychology (e.g., Attneave, 1959) to linguistics (e.g., Shannon, 1951). Particularly relevant here is the portion of Shannon's theory capturing discrete noiseless systems and their representation as stochastic Markov sources, the use of *n*-grams to estimate the statistical structure of the source and the development of *entropy* as a quantitative measure of the predictability of the source.

An *n-gram model* (of order $n-1$) computes the conditional probability of an element e_i at index $i \in \{n,\ldots,j\}$ in a sequence e_1^j of length j, over an alphabet, E, given the preceding $n-1$ elements, e_{i-n}^{i-1}:

$$p(e_i|e_{i-n}^{i-1}) = \frac{count(e_{i-n}^{i})}{count(e_{i-n}^{i-1})} \tag{1}$$

where e_m^n is the contiguous subsequence (substring) of sequence e between elements m and n, e_m is the element at index m of the sequence e, and $count(x)$ is the number of times that x appears in some training corpus of sequences.

Given an n-gram model of order $n - 1$, the degree to which an event appearing in a given context in a melody is unexpected can be defined as the *information content* (MacKay, 2003), $h(e_i|e_{i-n}^{i-1})$, of the event given the context:

$$h(e_i|e_{i-n}^{i-1}) = \log_2 \frac{1}{p(e_i|e_{i-n}^{i-1})}. \tag{2}$$

The information content can be interpreted as the contextual unexpectedness or surprisal associated with an event. The contextual uncertainty of the model's expectations in a given melodic context can be defined as the *entropy* (or average information content) of the predictive context itself (Shannon, 1948):

$$H(e_{i-n}^{i-1}) = \sum_{e \in E} p(e_i|e_{i-n}^{i-1}) h(e_i|e_{i-n}^{i-1}). \tag{3}$$

More sophisticated information-theoretic measures (e.g., *predictive information*: Abdallah & Plumbley, 2009) exist but are not considered here as they have yet to be applied to music cognition.

2.3. Information-theoretic models of music

Information theory was applied to music in 1955 (Cohen, 1962) and used throughout the 1950s and 1960s to analyze music (Cohen, 1962; Hiller & Bean, 1966; Hiller & Fuller, 1967; Meyer, 1957; Youngblood, 1958) and to compose (e.g., Ames, 1987, 1989; Brooks Jr., Hopkins, Neumann, & Wright, 1957; Hiller, 1970; Hiller & Isaacson, 1959; Pinkerton, 1956).

These early studies ran into difficulties (Cohen, 1962). The first is the estimation of probabilities from the samples of music (Cohen, 1962). A distribution estimated from a sample of music is supposed to accurately reflect a listener's perception of that sample. However, a listener's perception (e.g., of the first note) cannot be influenced by music she has not yet heard (e.g., the last note), so her knowledge and expectation changes with each new note (Meyer, 1957). To address this, Coons and Kraehenbuehl calculated dynamic measures of information (predictive failure) in a sequence (Coons & Kraehenbueh, l958; Kraehenbuehl & Coons, 1959). However, it remains unclear whether the method could be implemented and generalized beyond their simple examples. Furthermore, the method still fails to model the listener's prior experience of music (Cohen, 1962). Second, the early studies are generally limited to low, fixed-order probability estimates and therefore do not take full statistical advantage of musical structure. Third, except for Hiller and Fuller (1967), the music representations were exclusively simple representations of pitch (Cohen, 1962), ignoring other musical dimensions. Even Hiller and Fuller (1967) considered each dimension separately, as they had no way of combining the derived information.

Information-theory lost favor in psychology in the late 1950s and early 1960s during the ''cognitive revolution'' that ended behaviorism (Miller, 2003). This was because of objective inadequacies of basic Markov chains as models of psychological representations, particularly for language (Chomsky, 1957); it may also have been due to limitations in corpus size and the processing power of contemporaneous computers. The knowledge engineering approach dominated cognitive science until the 1980s, when renewed interest in connectionism (Rumelhart & McClelland, 1986) revitalized work on learning and the statistical structure of the environment.

These trends in cognitive science affected research on music. Connectionist models became popular in the late 1980s (Bharucha, 1987; Desain & Honing, 1989; Todd, 1988). However, with a few isolated exceptions (e.g., Baffioni, Guerra, & Lalli, 1984; Coffman, 1992; Knopoff & Hutchinson, 1981, 1983; Snyder, 1990), it was not until the mid-1990s that information theory and statistical methods were again applied to music (Conklin & Witten, 1995; Dubnov, Assayag, & El-Yaniv, 1998; Hall & Smith, 1996; Ponsford, Wiggins, & Mellish, 1999), as Darrell Conklin's sophisticated statistical models of musical structure (Conklin & Witten, 1995) addressed many of the early limitations.

3. IDyOM: A cognitive model of musical expectation

3.1. Introduction: Locating the model

By way of explaining our approach to the study of information dynamics and the associated experience of expectation, we now present an overview of the Information Dynamics of Music (IDyOM) model of musical melody processing. As a caveat: The use of the word ''model'' is problematic here, as it is the only appropriate term to use for the whole of the IDyOM theory-and-system, which is a model of a process, but also for some of its components, which are (Markov) models of data.

This work is motivated by empirical evidence of implicit learning of statistical regularities in musical melody (Oram & Cuddy, 1995; Saffran, Aslin, & Newport, 1996; Saffran et al., 1999). In particular, Krumhansl, Louhivuori, Toiviainen, Järvinen, and Eerola (1999) presented evidence for the influence of higher order distributions in melodic learning. Ponsford et al. (1999) used third- and fourth-order models to capture implicit learning of harmony, and evaluated them against musicological judgements. So there is evidence that broadly the same kind of model can capture at least two different aspects of music (melody and harmony) but also that they predict the expectations of untrained listeners as well as specialist theoreticians. The aim, then, was to construct a computational system embodying these theories and to subject them to rigorous testing.

Fig. 1 locates the abstract architecture of our model in a bird's eye view of music cognition. We have supplied a mechanism for learning enabling this overall structure (Pearce, 2005), and we hypothesize that it approximates the human mechanism at the level illustrated. We aim to understand the relationship between auditory stimuli (bottom of Fig. 1) and musical experience (top of Fig. 1). The results in the rest of this section are summarized from other more detailed publications, citations of which are given throughout.

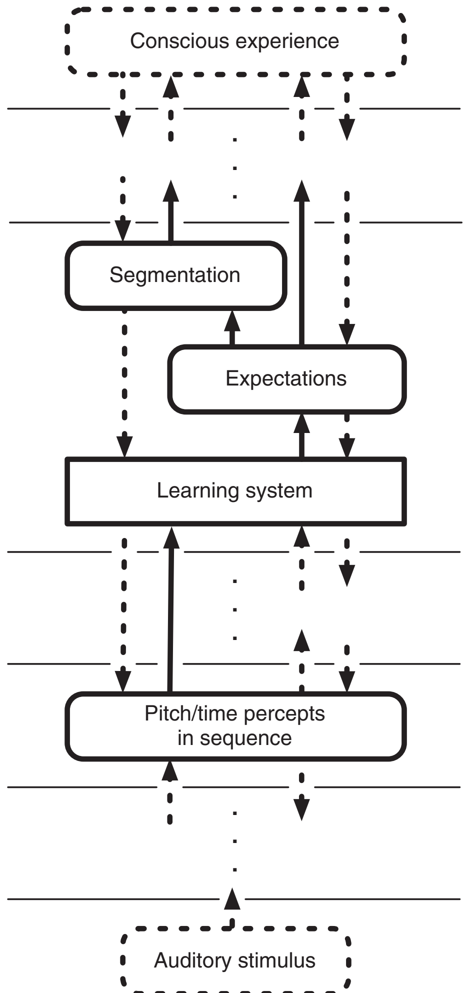

Fig. 1. An abstract layered map, locating our model in a larger cognitive system. The various layers, which are delineated by horizontal lines, and some of which are elided by …, contain processes (in squared boxes) and phenomena (in rounded boxes). These are connected by information flow, denoted by arrows. Solid lines denote processes, phenomena, and information flows that are explicitly represented in our model, while dotted ones indicate those that we believe to exist but that are not modeled, either because they are outside of the scope of the present work (such as emotional response to music) or because our strict bottom-up hypothesis forbids it for the present (such as expectation feedback into basic audio perception). Below the bottom perceptual/cognitive layer lies the physical auditory stimulus.

For methodological clarity, we work strictly bottom-up, requiring that phenomena (e.g., segmentation) arise from learning alone, and that learning be unsupervised—that is, the system is given *no* information about the outputs required. Learning also applies elsewhere—for example, in the lower level process of pitch categorization underlying formation of note-event percepts, which we presuppose here.

3.2. Outline

The core of IDyOM is a model of human melodic pitch prediction (Pearce, 2005) that builds on music informatics (Conklin & Witten, 1995), data compression (Bunton, 1997; Cleary & Teahan, 1997), and statistical language modeling (Manning & Schütze, 1999). It learns unsupervised, simulating implicit learning by exposure alone, without training, so it is *strongly bottom-up* (Cairns, Shillcock, Chater, & Levy, 1997). It uses Markov models or *n*-grams (Section 2.2; Manning & Schütze, 1999, ch. 9).

As IDyOM encounters the musical corpus from which it learns, it creates a compact representation of the data (Pearce, 2005), facilitating matching of new note sequences against previously encountered ones. Basic Markov modeling (Manning & Schütze, 1999, ch. 9) is extended in two ways.

First, the model is of variable order, incorporating an *interpolated smoothing strategy* to allow the predictions of *n*-gram models of all possible orders to contribute probability mass to each predicted distribution (Cleary & Witten, 1984), and an *escape strategy* admitting distributions including previously unseen symbols (Cleary & Witten, 1984; Moffat, 1990). The combination of available methods used in IDyOM is the most effective for musical melody (Pearce & Wiggins, 2004). The back-off strategy, PPM* (Cleary & Teahan, 1997), first tries the longest possible context and works down to nothing, summing probabilities until the context is empty, each weighted proportionally to the number of back-off steps required to reach it. IDyOM's escape method is Method C of Moffat (1990).

Second, the model is multidimensional, in two ways. First, following Conklin and Witten (1995), the system is configured with two functionally identical models, one for long-term (LTM), which is exposed to an entire corpus (modeling a listener's learned experience and supplying the context for information theoretic analysis) and the other for short-term (STM), which is exposed only to the current melody (modeling current listening).[2] Each model produces a distribution predicting each note as the melody proceeds, and the two distributions may be combined to give a final output (Fig. 3), weighted by the Shannon (1948) entropy of the distribution (more information weighs more heavily; Conklin & Witten, 1995; Pearce, Conklin, & Wiggins, 2005). There are five configurations: Each model alone (STM, LTM), two models together (BOTH), where the LTM is fixed and does not learn from the current stimulus data, and LTM+ and BOTH+, where the LTM does learn as the stimulus proceeds. LTM+, BOTH, and BOTH+ are serious candidates as models of human music cognition; STM and LTM alone are included for completeness, although both can tell us about musical structure (Potter, Wiggins, & Pearce, 2007). The second multidimensional aspect is *within* each model, where there are multiple distributions derived from multiple features of the data, as detailed in Fig. 2 and the next section (Conklin & Witten, 1995). These are combined using the same weighting strategy to give the overall output distribution for each model (Pearce, 2005; Pearce et al., 2005).

It is crucial that the model is *never* given the answers that it is expected to produce, nor is it optimized with reference to those answers. Thus, its predictions are in a sense epiphenomenal, and this is the strongest reason for proposing IDyOM, and the strong statistical view in

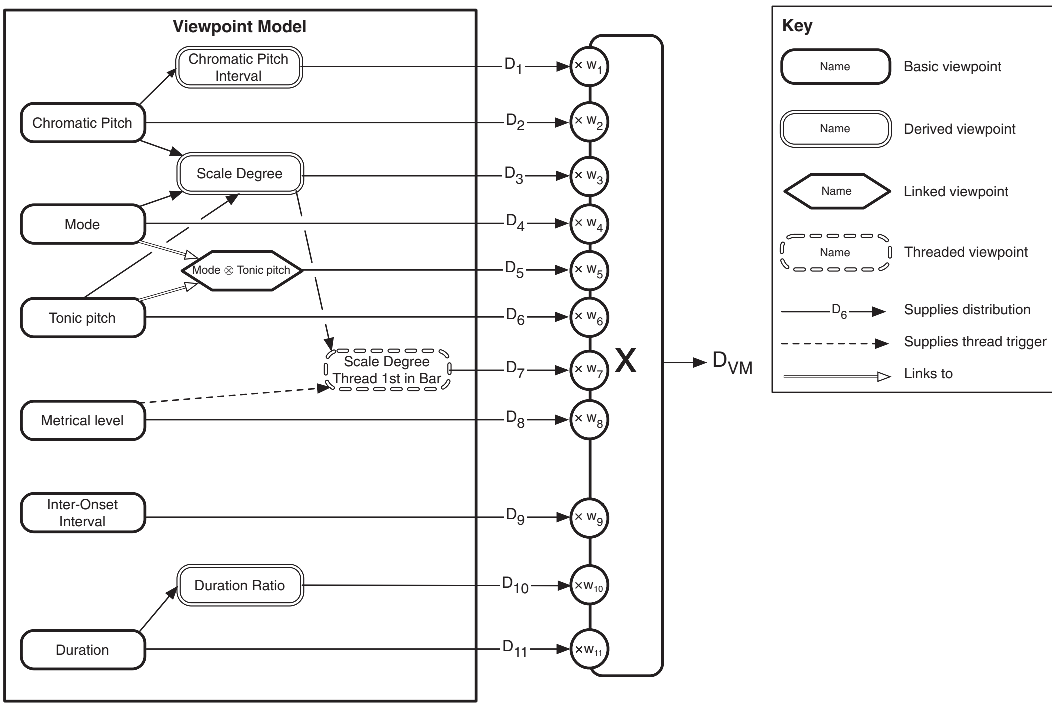

Fig. 2. Schematic diagram of the viewpoint models, showing a subset of available viewpoints. D_i are distributions across the alphabets of viewpoints, w_i are the entropic weights introduced in Section 3.3, and D_{VM} is the overall distribution derived from the combined viewpoints.

general, as a veridical mechanistic model of music cognition at this level of abstraction: It does what it is required to do *without being told how*.

3.3. Data representation

IDyOM operates at the level of abstraction described above: Its inputs are note percepts described in terms of pitch and time. These dimensions, however, engender multiple features of each note, derived from pitch or time or both. Added to these percept representations is an explicit representation of sequence in time: Sequence is the fundamental unit of representation.

IDyOM uses a uniform view of these features of data sequences (Conklin & Witten, 1995). Given a sequence of percepts, we define functions, *viewpoints*, that accept initial subsequences of a sequence and select a specific dimension of the percepts in that sequence. For example, there is a viewpoint function that selects values of pitch from melodic data; given a sequence of pitches, it returns the pitch of the final note. However, it is most often convenient to think of viewpoints as sequences of these values.

The model starts from *basic* viewpoints, literal selections of note features as presented to the system, including[3] pitch, notestarttime, duration, and mode. Further viewpoints are

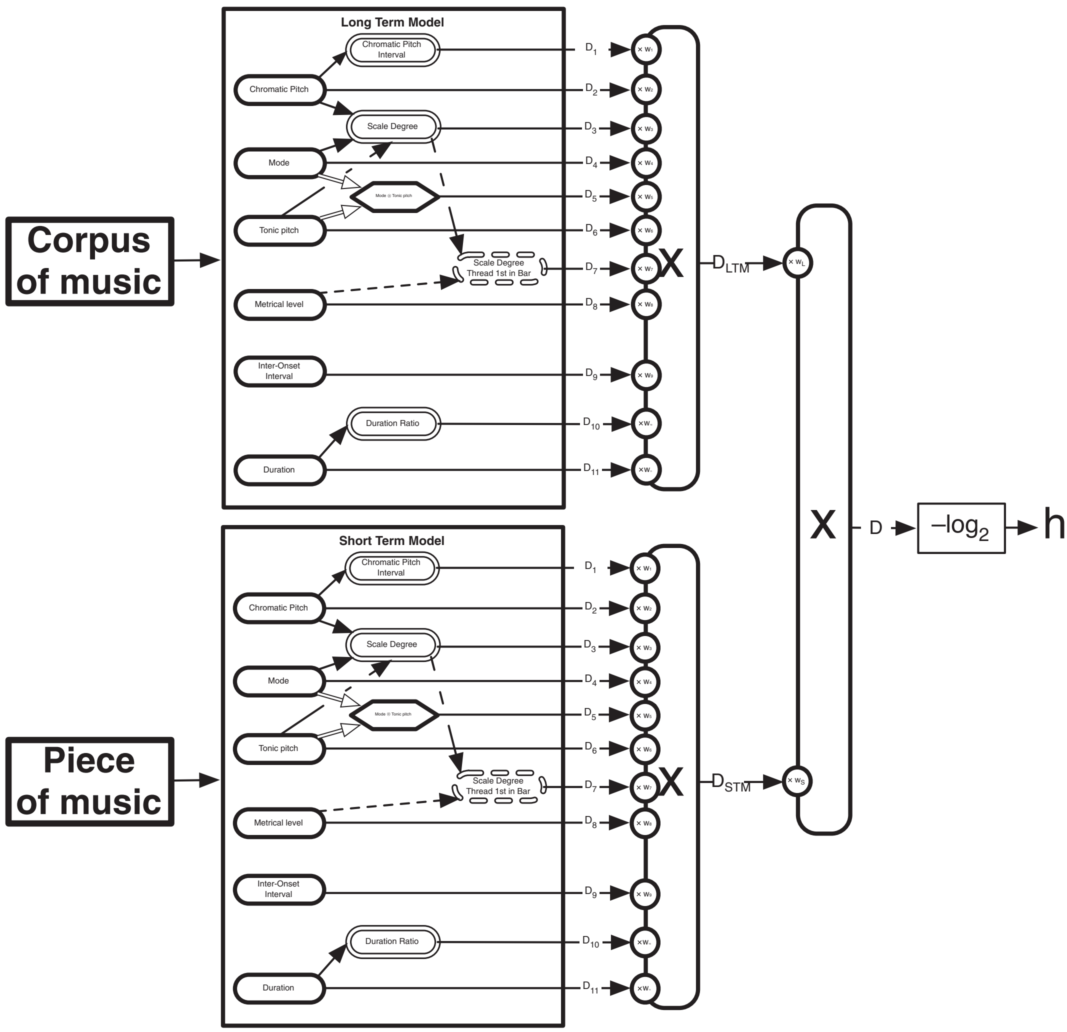

Fig. 3. Schematic diagram of combined IDyOM short-term and long-term models.

derived, such as pitch interval (the distance between two pitches). Two viewpoints may be *linked* ($A \otimes B$, where A and B are the source viewpoints), creating a compound whose alphabet is the cross-product of those of the two extant viewpoints. Finally, *threaded* viewpoints select elements of a sequence, depending on an external predicate: for example, selecting the scale degree of the first note in each bar of a melody, if metrical information is given (see Fig. 3).

Each of these data-feature models is carefully considered in music-perceptual, musicological, and mathematical terms (Wiggins et al., 1989), in some cases using feedback from musical expert participants (Pearce & Wiggins, 2007). Each viewpoint models a percept, which is expressed and used in music theory and hence there is clear, careful motivation for each feature.[4]

Having said this, it is important to understand that we are not predisposing the key feature of the system, its operation over sequences of percept features, in any hard-coded or rule-based way. These features are merely the properties of the data, psychologically grounded at a level of abstraction *below* the level of interest of the current study, that are *made available* for prediction; thus, their use does not contradict our claims of domain-generality and methodological neutrality at the level of interest of sequence processing. How those properties arise is not our focus of interest in the current presentation, but it will be the object of future work. The system itself selects which of the available representations is actually *used*, as described in the next section.

3.4. Viewpoint selection

The learning system is enhanced by an optimization step, based on the hypothesis that brains compress information, and that they do so efficiently. The optimization works by choosing the representation of the musical features from a pre-defined repertoire of music-theoretically valid representations, here defined by the set of viewpoints used in a model. For example, imagine two pitch viewpoints (representations of pitch) are available, one in absolute terms and the other in terms of the difference (*interval*, in musical terms) between successive notes. The system chooses the relative representation and discards the absolute one, because the relative representation allows the music to be represented independently of musical key, and this requires fewer symbols (by a factor of 12). There is evidence that humans may go through a similar process as exposure to music increases: Infants demonstrate absolute pitch, but the vast majority quickly learn relative pitch, and this becomes the dominant percept (Saffran & Griepentrog, 2001). Nevertheless, there is also evidence that people who develop relative pitch retain their absolute perception at a non-conscious level (Levitin, 1994; Schellenberg & Trehub, 2003).

Again, it is important to emphasize that no *training*, nor programmer intervention, with respect to or in favor of the solutions being sought, is involved here: Using a hill-climbing search method applied over the set of all viewpoints present (Pearce, 2005), the system objectively picks the set of viewpoints that encodes the data in a model with the lowest possible average information content[5] ($\bar{h}$). Thus, the data itself determines the selection of the viewpoints best able to represent it efficiently; a level playing field for prediction is provided by the fact that each viewpoint distribution is converted into a basic one before comparison: Thus, $\bar{h}$ is computed from the pitch distribution of each model. The selection approach is a brute force simulation of a more subtle process proposed in cognitive theories such as that of Gärdenfors (2000), which allow for the re-representation of conceptual spaces in response to newly learned data: In Gärdenfors' terms, viewpoints are *quality dimensions*, which can be rendered redundant by new, alternative, learned additions to the representational ontology, and therefore forgotten, or at least de-emphasized. A general mechanism by which this may take place in our statistical model is a focus of our current research, beyond the scope of the current paper.

3.5. Shortcomings of the model

This model is the first stage of an extended research program of cognitive modeling. In this context, it is important that we note its shortcomings as well as its successes and potentials. We do so at this point to make a clear distinction between the issues, which are outstanding for IDyOM as a model, and those which are relevant to the discourse on expectation presented in the next sections.

First, the model is currently limited to monodic melodic music, which is only one aspect of the massively multidimensional range of music available; while our focus on melody is perceptually, musicologically, and methodologically defensible, the other aspects need to be considered in due course. Elsewhere, we have begun to study the modeling of musical harmony (Whorley, Pearce, & Wiggins, 2008; Whorley, Wiggins, Rhodes, & Pearce, 2010), following on from the early efforts of Ponsford et al. (1999), and to extend IDyOM's coverage beyond music, looking at the possibility of language processing using the same technology (Wiggins, 2011b), given evidence of shared neural and cognitive mechanisms involved in processing complex sequential regularities in both domains (Tillmann, 2011).

Second, and more fundamentally, the memory model used here is inadequate: The model exhibits total recall and its memory never fails. This may be why it outperforms humans in some implicit learning tasks (see Rohrmeier & Rebuschat, 2011). There is work to do on the statistical memory mechanism (currently based on exact literal recording and matching by identity) to model human associative memory more closely. Options include pruning the leaves of the tree (e.g., Ron, Singer, & Tishby, 1996) or neural networks (e.g., Mozer, 1994), but we refer these possibilities to future work.

Third, as explained above, the viewpoints used in the system are chosen from music theory and must be implemented by hand. This is useful for the purposes of research, because we are able to interpret, to some extent, what the model is doing by looking at the viewpoints it selects. For example, the viewpoint scaledegree $\otimes$ pitchinterval encodes aspects of tonal listening (Lerdahl, 2001), and this viewpoint consistently emerges from the compression of tonal music databases. However, a purer system would be capable of constructing its own viewpoints (based on established perceptual principles) and choosing new ones, which lead to more compact models, akin to methods such as deep learning (Hinton & Salakhutdinov, 2006). This could be posited as a model of perceptual learning, in which new quality dimensions (Gärdenfors, 2000) are created in the perceiver's representation as they are required. This would greatly increase the power of the system, because it would be able to determine its own representation, by reflection.

4. Pitch expectation

Our approach invokes a cognitive learning process through which expectations contribute to accurate predictions about the auditory environment. Here, we study pitch expectations in melody, where evidence exists for learning. Melodic pitch expectations vary between musical styles (Krumhansl et al., 2000) and cultures (Carlsen, 1981; Castellano, Bharucha, &

Krumhansl, 1984; Eerola, 2004; Kessler, Hansen, & Shepard, 1984; Krumhansl et al., l999), throughout development (Schellenberg, Adachi, Purdy, & McKinnon, 2002) and across degrees of musical training and familiarity (Krumhansl et al., 2000; Pearce, Herrojo Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

The most influential theory of melodic pitch expectation, the Implication Realization (IR) theory (Narmour, 1990, 1992), proposes that expectations are governed in part by a few innate rules as well as by top-down influences; Schellenberg (1997) provides cognitive-scientific support. However, these rules would be unnecessary if the aspects of expectation they cover can be learned through exposure to music. The original purpose of the IDyOM model was to simulate human melodic pitch expectations and investigate whether they can be accounted for entirely by statistical learning (Pearce, 2005; Pearce & Wiggins, 2006).

Pearce and Wiggins (2006) tested this by exposing IDyOM's LTM to a corpus of 903 tonal folk melodies and comparing the predictions made by the BOTH+ model during simulated listening with the expectations of human listeners elicited in previous studies: using single-interval contexts (Cuddy & Lunny, 1995); using longer contexts from British folk songs (Schellenberg, 1996); and for each note in two chorale melodies (Manzara, Witten & James, 1992). Table 1 shows the results and a comparison with the two-factor IR model of Schellenberg (1997).[6] IDyOM generates the most accurate predictions ofpitch expectation in the literature to date, especially incomplex melodic contexts.

In these studies, melodies were paused to allow listeners to respond. However, this tends to elicit expectations related to closure (Aarden, 2003; Toiviainen & Krumhansl, 2003). Using a visual cue to elicit expectations without pausing the melody, Pearce et al. (2010) verified that IDyOM's predictions correlate well with human pitch expectations to notes in English hymns as indicated both by ratings ($r^2 = .78$, $p < .01$) and response times ($r^2 = .56$, $p < .01$). Again, the IDyOM model predicted the listeners' expectations better than the two-factor IR model.

Cognitive neuroscientific studies of musical expectations have tended to focus on EEG and MEG, which have far superior temporal resolution to other methods such as fMRI. ERP research has identified characteristic neural responses, in particular, an early anterior negativity peaking at around 180 ms post-stimulus, to violations of harmonic expectation in real musical excerpts (Steinbeis et al., 2006). There is evidence that the amplitude of this

Table 1

Results from IDyOM prediction experiments (Pearce & Wiggins, 2006; Pearce et al., 2010)

Data From	Stimuli	Schellenberg's (1997) Model (r^2)	IDyOM (r^2)
Cuddy and Lunny (1995)	Single intervals	.68	.72
Schellenberg (1996)	British folksongs	.75	.83*
Manzara et al. (1992)	German chorales	.13	.63*
Pearce et al. (2010)	English hymns	.73	.78

Note. In cases indicated by *, IDyOM significantly outperforms its nearest competitor on this task ($p < .01$). Data from Pearce and Wiggins (2006) © 2006 by the Regents of the University of California, published by University of California press; data reprinted from Pearce et al. (2010) © 2010, with permission from Elsevier.

component is related to the long-term digram probability of the chord occuring (Kim, Kim, & Chung, 2011; Loui, Wu, Wessel, & Knight, 2009). Violations of melodic expectation appear to produce early anterior responses with a slightly earlier latency (Koelsch & Jentschke, 2010) but only when they break tonal rules (Miranda & Ullman, 2007). In an EEG study of listeners to hymn melodies, Pearce et al. (2010) examined oscillatory and phase responses to notes with high information content as predicted by IDyOM. The results indicated that violations of melodic expectation increase phase synchrony across a wide network of sensor locations and generate characteristic patterns of beta-band activation in superior parietal lobule (see Fig. 4), which have previously been associated with tasks involving auditory–motor interaction, suggesting that violations of expectation may stimulate networks linking perception with action.

5. From expectation to structure

Grouping and boundary perception are core functions in many areas of cognitive science, such as natural language processing (e.g., speech segmentation and word discovery, Brent, 1999a,b; Jusczyk, 1997), motor learning (e.g., identifying behavioral episodes, Reynolds, Zacks, & Braver, 2007), memory storage and retrieval (e.g., chunking, Kurby, & Zacks, 2007), and visual perception (e.g., analyzing spatial organization, Marr, 1982). The *segmentation* of a sequence of musical notes into contiguous groups occurring sequentially in time (e.g., motifs, phrases etc.) is one of the central processes in music cognition (Lerdahl & Jackendoff, 1983).

Narmour (1990) proposed that grouping boundaries are perceived where expectations are weak: No particularly strong expectations are generated beyond the boundary. Saffran et al. (1999) have demonstrated empirically that infants and adults spontaneously perceive grouping boundaries in tone and syllable sequences at points where first-order probabilities are low (i.e., expectation is violated). Furthermore, word-boundaries in English text and infant-directed speech can be identified with some success using algorithms that segment before unexpected events (Brent, 1999b; Cohen, Adams, & Heeringa, 2007; Elman, 1990) and in uncertain contexts (Cohen et al., 2007).

Therefore, we hypothesize that musical grouping boundaries are perceived before events for which the unexpectedness of the outcome (h) and the uncertainty of the prediction (H) are high. We tested this in two experiments using the IDyOM model (trained on 907 Western tonal melodies; Pearce, 2005) to predict perceived grouping boundaries at peaks in the information content profile for a melody.

The first study (Pearce, Müllensiefen & Wiggins, 2010a) concerned phrase boundaries annotated by a musicologist in 1,705 Germanic folk songs from the Essen Database (Schaffrath, 1995). IDyOM predicted the annotated boundaries with precision .76 and recall .50, so $F_1 = .58$. The second (Pearce, Müllensiefen & Wiggins, 2010b) examined the boundary perceptions of 25 listeners to 15 unfamiliar popular melodies. Here, IDyOM predicted the listener's boundaries with mean precision .57 and recall .73, so $F_1 = .64$. These results are summarized in Table 2.

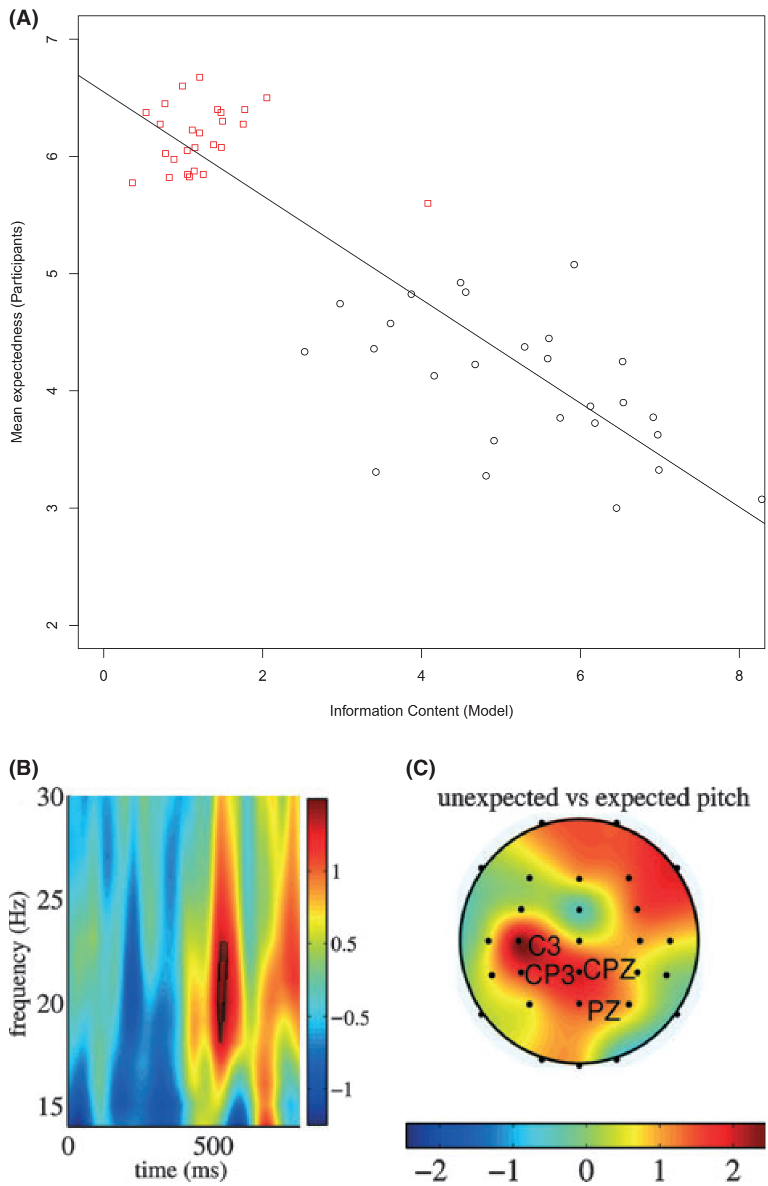

Fig. 4. Summary of results of Pearce et al. (2010) showing the three-way connection between model prediction, behavioral data, and neurophysiological responses. (A) The correlation between the mean expectedness ratings of the listeners for each probed note (ordinate) and the information content of IDyOM (abscissa). The notes were divided into two groups: high information content (black circles) and low information content (red squares). (B) Spectrogram showing differences in spectral power between high and low-information content notes in the beta band (14–30 Hz) over peristimulus time with regions of significant difference, indicated by the permutation test, identified by the black contour. (C) Topography of the difference power at 18–23 Hz over the time window 500–550 ms. Reprinted from Pearce et al. (2010) © 2010, with permission from Elsevier.

Table 2
Summary of results presented by Pearce et al. (2010a,b)

Model	1705 Folk Songs			15 Pop Songs		
	Precision	Recall	F1	Precision	Recall	F1
Grouper	0.71	0.62	0.66	0.86	0.82	0.83
LBDM	0.70	0.60	0.63	0.79	0.81	0.78
IDyOM	0.76	0.50	0.58	0.57	0.73	0.64
GPR2a	0.99	0.45	0.58	0.70	0.54	0.58
GPR2b	0.47	0.42	0.39	0.47	0.45	0.43
GPR3a	0.29	0.46	0.35	0.26	0.43	0.30
GPR3d	0.66	0.22	0.31	0.17	0.11	0.11
PMI	0.16	0.32	0.21	0.24	0.49	0.31
TP	0.17	0.19	0.17	0.25	0.45	0.31
Always	0.13	1.0	0.22	0.13	1.0	0.23
Never	0.0	0.0	0.0	0.0	0.0	0.0

Note. The segmentation models are Grouper (Temperley, 2001), Local Boundary Detection Model (Cambouropoulos, 2001), the Grouping Preference Rules (GPRs) of GTTM (Lerdahl & Jackendoff, 1983), simple statistical models based on transition probabilities (TP) and pointwise mutual information (PMI) (Saffran et al., l999; Brent, 1999a) and two baseline models, which predict boundaries for every note (Always) and for no notes (Never). Data from Pearce et al. (2010a) reproduced by permission of Pion Limited, London, UK; data from Pearce et al. (2010a) reproduced with kind permission of Springer Science+Business Media © 2010.

These results are better than simple first-order statistical models and broadly comparable to those of hand-crafted rule-based grouping models. Although they fall short of the best rule-based models, IDyOM does predict boundaries not captured by those models. Given that the model learns unsupervised and was neither optimized for segmentation nor given information about grouping, this constitutes a very pure test of the hypothesis that perceived grouping structure arises from expectation violation.

We have also investigated whether IDyOM can segment speech signals (qua phoneme sequences); preliminary evidence suggests that it can, and that the extensions to Markov Modeling detailed above improve performance here too (Wiggins, 2011b). This adds further evidence to our claim that we are modeling at a rather general level, and that the model is consistent with evolutionary likelihood, because deployment of a mechanism in multiple areas both simplifies the hypothetical system, thus making evolution more likely, and increases the evolutionary advantage the mechanism conveys.

6. From expectation to experience

Looking now to the future, we consider how the current state of our research fulfils our aim: Explicating the conscious experience of music. We have explained how expectation can be simulated by the IDyOM model, using unsupervised analytical methods, and not as a trained outcome (Pearce & Wiggins, 2006). Furthermore, the time-variant signal so produced can be analyzed to predict perceptual segmentation in both music and language

(Pearce et al., 2010b,a; Wiggins, 2011a). Also, the model reliably predicts specific neural activity associated with unexpectedness (Pearce et al., 2010).

The key points are that IDyOM's predictions correspond reliably with specific detectable neural activity, and that experimental participants experience the corresponding effect as a conscious feeling of expectedness. Therefore, we hypothesize that IDyOM is a veridical, although approximate, abstract simulation of the *actual cognitive processes* involved in these phenomena. Furthermore, we hypothesize that the neural activity predicted is either the cause or the result (we aim to discover which) of the associated reported experience. Thus, the model is directly predicting aspects of what is experienced. This strong claim demands further verification, of course, and we are engaged on such a program.

7. From expectation to aesthetics

People value music primarily for the emotions it generates (Juslin & Laukka, 2004). Meyer (1956) linked the emotional experience of music with musical structure via the listener's expectations, which create patterns of tension and resolution that generate affective states differing in arousal and valence. Thus, he viewed violated expectation as inherently negatively valenced, indicating predictive failure:

> if our expectations are continually mistaken or inhibited, then doubt and uncertainty … will result. … the mind rejects and reacts against such uncomfortable states and … looks forward to a return to the certainty of regularity and clarity. (Meyer, 1956, p. 27)

In an evolutionary framework (Section 2.1) of probabilistic modeling, expected events *should* engender pleasure, as they indicate a successful domain model. Unexpected events, however, indicate predictive failure, which should be penalized, affectively, to stimulate further learning and improve the model. However, in music, this raises a conundrum: How can unexpected events be pleasurable per se?

Huron (2006) examines the relationship between musical expectations and aesthetic pleasure, identifying several cognitive processes involved both in generating expectations about a forthcoming event and generating response to it when it occurs. He identifies three kinds of response to an event: A *prediction response*, evaluating the extent to which it conforms to expectations; the *reaction response*, a fast, automatic, subcortical response to its nature; and an *appraisal response*, a more leisurely, cortically mediated process of consideration and assessment yielding positive and negative reinforcement associated with the outcome.

Huron describes the prediction effect whereby positive emotions resulting (via the prediction response) from anticipatory success are misattributed to the stimulus itself, leading to a preference for predictable events. Conversely, the stress resulting from surprising events, indicating maladaptive anticipatory failure, has two main effects. First, it activates one of three fast, conservative responses: fight, flight, or freeze (depending on the perceived severity of the threat and degree of control over the outcome). Second, it informs the cognitive system about the predictive utility of competing potential representations of the environment. Just as we select viewpoints for IDyOM based on prediction performance (see Section 3.4), Huron proposes that neural representations yielding accurate predictions are strengthened and reused, while those that do not atrophy.

So how can surprise be enjoyable, even when associated with negative emotion, due to the prediction effect? Huron's answer invokes emotional *contrastive valence* between the different expectation responses. An event that is welcome but unexpected induces a negative prediction response that increases the positive limbic effect of the reaction or appraisal responses. Thus, even events that are merely innocuous, but unexpected, can generate positive emotions.

Expectation also engenders physiological effects. Unexpected chords produce greater physiological arousal (skin conductance) than expected chords (Koelsch, Kilches, Steinbeis, & Schelinski, 2008; Steinbeis et al., 2006). Huron (2006) suggests that contrastive valence produces three kinds of pleasurable physiological response: *awe*, *laughter*, and *frisson*. Here we focus on frisson (also called *chills* or *shivers*). Chills are a frequent response to music (Panksepp, 13; Sloboda, 1991), usually experienced as pleasurable (Goldstein, 1980), involving increased subjective emotion and physiological arousal (Grewe, Kopiez, & Altenmüller, 2009). They tend to be associated with unexpected harmonies, sudden dynamic or textural changes, or other new elements in the music (Grewe, Nagel, Kopiez, & Altenmüller, 2007; Sloboda, 1991). Familiarity is also a significant influence on chills (Grewe et al., 2009). In a PET study, Blood and Zatorre (2001) found that the intensity of chills correlated positively with regional cerebral blood flow (rCBF) in brain regions related to reward (e.g., left ventral striatum and orbito-frontal cortex) and negatively with rCBF in regions involved in processing negative emotions (e.g., bilateral amygdale). Recently, Salimpoor, Benovoy, Larcher, Dagher, and Zatorre (2011) have shown that chills are associated with striatal dopamine relase and activation in the nucleus accumbens, while the caudate nucleus was activated during anticipation of a passage of music inducing chills. In another line of research, Biederman and Vessel (2006) propose that aesthetic pleasure is bound to perceptual learning, due to an increasing density of mu-opioid receptors in the ventral visual stream from primary to association cortex. Consistent with this theory, the frequency of chills to music was diminished in some participants treated with naloxone, a specific endomorphin antagonist (Goldstein, 1980).

On a more (literally) anecdotal level, there is everyday evidence of the effects of expectation violation in jokes (Ritchie, 2003). The violation can be of various kinds, the most obvious being semantic violations in puns, where an expectation is set up and then violated by use of a double meaning. For example,

> There are two fish in a tank. One says to the other, ''How on earth do we drive this thing?''

Here, very strong expectation is set up that the tank in question is a fish tank, and so the revelation that it is actually a (military) vehicle is highly unexpected, and, in some listeners, causes laughter. The chain of events leading to that particular somatic reaction is discussed by Huron (2006, Ch. 14) along with other more subtle, musical kinds of humor.

Theoretical proposals in experimental aesthetics predict that subjective stimulus complexity should show an inverted U-shaped relationship with liking (Berlyne, 1974). This explicates both the *mere exposure effect* (Zajonc, 1968), where preference increases with increasing exposure, and the *boredom effect* (Cantor, 1968), where preference decreases with increasing exposure, by positing different initial levels of subjective stimulus complexity. Indeed, an inverted U-shaped relationship exists between subjective complexity and liking in music perception (North & Hargreaves, 1995). We suggest that the relationship between complexity and liking is very plausibly mediated by predictability, measurable by mean information content. We suggest that intermediate degrees of predictability are preferred with very predictable and very unpredictable music (with respect to prior knowledge) both being disliked. This is consistent with the proposals of Biederman and Vessel (2006) as stimuli exhibiting these extremes of unpredictability afford reduced opportunities for learning. In this account, the learning stimulated by moderate degrees of expectation violation would be pleasurable per se.

8. From expectation to creativity

Expectation and its cognitive mechanisms may also be important in understanding creativity. Here, we distinguish between the large-scale, conscious, planned creativity of the orchestral arranger and the flash of inspiration, as, for example, a fragment of melody appears in the mind of any musical person: We refer to the latter, which forms the seed of musical creation. As our approach is partly derived from linguistics methods, and it has shown basic capacity for linguistic prediction (Wiggins, 2011a), we hypothesize that a common mechanism like it may underpin both domains. Plotkin (1998) argues that creativity is necessary in everyday language (and not only in creative writing), to construct sequences of words. We suggest that a common method of generation, also, may relate the two domains.

Expectation-based models can generate structures in both domains, using common statistical sampling-optimization methods (Conklin, 2003; Conklin & Witten, 1995). Acceptable musical structures can be found, given a good enough representation (Pearce & Wiggins, 2007; Ponsford et al., 1999). In the context of the literature on creativity, the IDyOM model may be thought of as supplying an implicit definition of a *conceptual space* (Boden, 2004; Wiggins, 2006a), while the sampling method used for generating from it constitutes the *traversal strategy* in the Creative Systems Framework of Wiggins (2006a,b). What is missing is the corresponding *evaluation* function that chooses *high-quality* artistic structures, and this is an open research topic, partly because current models tend to be incomplete representations of the phenomena they capture but also because quality criteria are subjective and context-dependent. What is more, following Berlyne (Section 7) generation using naïve objective functions (e.g., maximum likelihood) is unlikely to be aesthetically successful unless models include self-description: The ''ebb and flow'' of musical expectation is aesthetically important, and not just any will do. One approach might include the expectation generated by the music as part of the learned model, introducing reflection—a primary component of consciousness (Shanahan, 2010).

Most important, from the current perspective, is a common mechanism underlying perception and generation, and happening more or less continuously, in the ways implied by Plotkin (1998) and explicitly suggested by Shanahan (2010), albeit with different ancillary cognition. The alternative, complex paired mechanisms admitting two tightly coupled phenomena such as generation and perception of perceptual sequences, is much less convincing, in evolutionary terms. The point, then, is that the mechanism needed to manage expectation in a perceptual domain may also serve as a mechanism underpinning (but not completely accounting for) creativity in that domain; these modeling attempts demonstrate how this can happen. Fig. 5 shows a hymn tune harmonized by a system, which has learned to harmonize by mere exposure, based on the methods used by IDyOM (Whorley, Wiggins, Rhodes, & Pearce, 2010; Whorley et al., 2008).

9. Expectations for the future

Quantifying aspects of musical experiences (e.g., expectation and segmentation) in information-theoretic terms yields a formal mathematical model of the cognitive processes generating these experiences. A computational approach ensures that all design assumptions are explicit (Johnson-Laird, 1983; Longuet-Higgins, 1981) and allows the responses of the model to a stimulus set to be compared quantitatively to the empirically determined

Fig. 5. The score of a hymn tune/harmonization performed by Raymond Whorley's creative system (Whorley et al., 2008, 2010), which uses extended versions of the techniques presented here. The tune is a French church melody, from *Chants Ordinaires de l'Office Divin* (Paris, 1881); it is reprinted as Hymn No. 33, *Grafton*, in the 1993 edition of the English Hymnal. The harmonization is produced by Whorley's (unassisted) creative system.

responses of human listeners to the same stimuli (Newell & Simon, 1976; Simon & Kaplan, 1989).

Our emphasis on probabilistic learning has three primary advantages over rule-based models of music cognition. First, it provides an explicit account of acquisition of the cognitive processes that we study, potentially allowing prediction of behavior change through development (e.g., Schellenberg et al., 2002) and across cultures (e.g., Eerola, 2004). Second, the models generalize naturally to cognitive processing in other sequential domains such as language, visual sequencing, or motor planning, allowing us to posit a domain-general learning mechanism, instances of which can become specialized to a particular domain through exposure to examples from it (Elman et al., 1996). Finally, probabilistic models of perceptual processes such as expectation and segmentation have a more natural neurobiological interpretation than static domain-specific rules in terms of current theories of predictive coding in neural processing of perceptual stimuli (Barlow, 1959; Friston, 2005; Smith & Lewicki, 2006).

We believe that probabilistic processes underlie expectation, which, in turn underlies a substantial proportion of human experience. The study of music in this context is extremely valuable, because it is simultaneously intrinsically complex but almost free of extrinsic reference. We believe that the study of music using the methods outlined here will produce significant advances in cognitive science in the immediate future.

Notes

1. A mathematical construct that can be summarized as describing a line of discrete values, with an addition operation, an identity element (zero), an inverse function (negative), and an ordering relation $\leq$ which is antisymmetric, transitive, and total. The integers with addition form such a group. This level of abstraction also admits non-Western notions of pitch, so long as they are organized in a way corresponding broadly with scales.
2. The distinction between our LTM and STM is related to the distinction made by Bharucha (1987) between *schematic* and *veridical* expectations—although the STM generates expectations from the structure of the *current piece*, as opposed to predicting from a *memory* of its literal structure.
3. Some viewpoints are named differently here from the actual implementation and from previous presentations. They are in 1-to-1 correspondence, and the names here are more musically informative.
4. Music theory is arguably the most formally developed example of a folk psychology currently extant, being based on extensive and careful study of the aural constructs used in a particular musical culture (Western art music), and their associated semiotic connotations, in terms of their usage in that culture. A point sometimes missed in the interdisciplinary music literature is that the constructs of music theory almost always correspond with perceptual principles identifiable in general auditory psychology. For example, the musical concept of *melody* relies on *auditory streaming* (Bregman,

1990) of sequences of pitched events (Wiggins, Harris & Smaill, 1989), and artistic attempts deliberately to create alternative notions of melody, which break these constraints, such as Schoenberg's *tonfarbenmelodie* (Schoenberg, 1974), have met with less than complete success. Western music notation often encodes these musical properties (in particular, the overarching construct of *tonality*) implicitly.
5. Manning and Schütze (1999) and Conklin (1990) call this same quantity ‘‘cross-entropy’’; we find the current terminology more accurately descriptive.
6. r^2 estimates the proportion of variance in the participants' responses accounted for by the model.

Acknowledgments

EPSRC awarded grants GR/S82220/01 and EP/H01294X/1, which supported both authors. We gratefully acknowledge the influence of Ray Whorley (in Computing), Keith Potter (in Music), Daniel Müllensiefen, Selina Kapasi, Joydeep Bhattacharya, and Job Lindsen (in Psychology) at Goldsmiths; Maria Herrojo-Ruiz (at Charité, Universitätsmedizin Berlin); and Samer Abdallah and Mark Plumbley in Electrical Engineering at Queen Mary, University of London.

References

Aarden, B. (2003). *Dynamic melodic expectancy*. PhD thesis, Columbus, OH: Ohio State University.

Abdallah, S. A., & Plumbley, M. D. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, *21*(2), 89–117.

Ames, C. (1987). Automated composition in retrospect: 1956–1986. *Leonardo*, *20*(2), 169–185.

Ames, C. (1989). The Markov process as a compositional model: A survey and tutorial. *Leonardo 22*(2), 175–187.

Attneave, F. (1959). *Applications of information theory to psychology*. New York: Holt.

Baffioni, C., Guerra, F., & Lalli, L. (1984). The theory of stochastic processes and dynamical systems as a basis for models of musical structures. In M. Baroni & L. Callegari (Eds.), *Musical grammars and computer analysis* (pp. 317–324). Florence: Leo S. Olschki.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In Her Majesty's Stationery Office (Ed.), *Proceedings of a symposium on the mechanisation of thought processes*, volume 2 (pp. 537–559). London: National Physical Laboratory, Teddington. Her Majesty's Stationery Office.

Berlyne, D. E. (1974). The new experimental aesthetics. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps towards an objective psychology of aesthetic appreciation* (pp. 1–25). Washington, DC: Hemisphere Publishing Co.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*(1), 1–30.

Biederman, I., & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist*, *94*, 247–253.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *9*(20), 11818–11823.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). London, UK: Routledge.

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

Brent, M. R. (1999a). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, *34*(1–3), 71–105.

Brent, M. R. (1999b). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Science*, *3*, 294–301.

Brooks Jr., F. P., Hopkins, A. L., Neumann, P. G., & Wright, W. V. (1957). An experiment in musical composition. *IRE Transactions on Electronic Computers*, *EC-6*(1), 175–182.

Bunton, S. (1997). Semantically motivated improvements for PPM variants. *The Computer Journal*, *40*(2/3), 76–93.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111–153.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In N. Bernadini (Ed.), *Proceedings of the international computer music conference* (pp. 17–22). San Francisco: ICMA.

Cantor, G. N. (1968). Children's ''like-dislike'' ratings of familiarized and unfamiliarized visual stimuli. *Journal of Experimental Child Psychology*, *6*, 651–657.

Carlsen, J. C. (1981) Some factors which influence melodic expectancy. *Psychomusicology*, *1*(1), 12–29.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*(3), 394–412.

Chomsky, N. (1957) *Syntactic structures*. The Hague, The Netherlands: Mouton.

Cleary, J. G., & Teahan, W. J. (1997). Unbounded length contexts for PPM. *The Computer Journal*, *40*(23), 67–75.

Cleary, J. G., & Witten, I. H. (1984). Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications*, *32*(4), 396–402.

Coffman, D. D. (1992). Measuring musical originality using information theory. *Psychology of Music*, *20*, 154–161.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, *7*(2), 137–163.

Cohen, P. R., Adams, N., & Heeringa, B. (2007). Voting experts: An unsupervised algorithm for segmenting sequences. *Intelligent Data Analysis*, *11*(6), 607–625.

Conklin, D. (1990) *Prediction and entropy of music*. Master's thesis, Canada: Department of Computer Science, University of Calgary.

Conklin, D. (2003). Music generation from statistical models. In M. Lee (Ed.), *Proceedings of the AISB 2003 symposium on artificial intelligence and creativity in the arts and sciences* (pp. 30–35). Brighton, UK: SSAISB.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Coons, E., & Kraehenbuehl, D. (1958). Information as a measure of structure in music. *Journal of Music Theory*, *2*(2), 127–161.

Cross, I. (2007). Music and cognitive evolution. In R. Dunbar & L. Barrett (Eds.), *OUP handbook of evolutionary psychology* (pp. 649–667). Oxford, UK: Oxford University Press.

Cuddy, L. L., & Lunny, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgements of continuity. *Perception and Psychophysics*, *57*(4), 451–462.

Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Desain, P., & Honing, H. (1989). The quantisation of musical time: A connectionist approach. *Computer Music Journal*, *13*(3), 56–66. Also in *Music and Connectionism*, 1991, P. M. Todd and D. G. Loy (Eds.), Cambridge, MA: MIT Press.

Dubnov, S., Assayag, G., & El-Yaniv, R. (1998). Universal classification applied to musical sequences. In M. Simoni (Ed.), *Proceedings of the 1998 International Computer Music Conference* (pp. 332–340). San Francisco: ICMA.

Eerola, T. (2004). Data-driven influences on melodic expectancy: Continuations in North Sami Yoiks rated by South African traditional healers. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the eighth international conference of music perception and cognition* (pp. 83–87). Adelaide, Australia: Causal Productions.

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

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*, 173–215.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*, 815–836.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology*, *8*, 126–129.

Greenwood, D. D. (1996). Comparing octaves, frequency ranges, and cochlear-map curvature across species. *Hearing Research*, *94*(1–2), 157–162.

Grewe, O., Kopiez, R., & Altenmüller, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, *27*(1), 61–74.

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*, 297–314.

Hall, M., & Smith, L. (1996). A computer model of blues music and its evaluation. *Journal of the Acoustical Society of America*, *100*(2), 1163–1167.

Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, *7*, 535–563.

Hauser, M. D., Aslin, E. L., & Newport, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning incotton-top tamarins. *Cognition*, *78*, B53–B64.

Hiller, L. (1970). Music composed with computers — a historical survey. In H. B. Lincoln (Ed.), *The computer and music* (pp. 42–96). Cornell, NY: Cornell University Press.

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 in Webern's Symphonie, Op. 21. *Journal of Music Theory*, *11*(1), 60–115.

Hiller, L., & Isaacson, L. (1959). *Experimental music*. New York: McGraw–Hill.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

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

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception and Psychophysics*, *41*(6), 621–634.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*, 217–238.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception*, *23*(1), 1–27.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception*, *2*(2), 131–165.

Kim, S. G., Kim, J. S., & Chung, C. K. (2011). The effect of conditional probability of chord progression on brain response: An MEG study. *PLoS ONE*, *6*, e17337.

Knopoff, L., & Hutchinson, W. (1981). Information theory for musical continua. *Journal of Music Theory*, *25*, 17–44.

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, *27*, 75–97.

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

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS One*, *3*(7), e2631.

Kraehenbuehl, D., & Coons, E. (1959). Information as a measure of the experience of music. *Journal of Aesthetics and Art Criticism*, *17*(4), 510–522.

Krumhansl, C. L. (1995). Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft*, *3*(2), 211–250.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual hymns: Convergence of statistical, behavioural and computational approaches. *Music Perception*, *17*(2), 151–195.

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.

Kurby, C. A., & Zacks, J. M. (2007). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79.

Large, E. W., Almonte, F., & Velasco, M. (2010). A canonical model for gradient frequency neural networks. *Physica D. 239*, 905–911.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review*, *106*, 119–159.

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford, UK: Oxford University Press.

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of Tonal music*. Cambridge, MA: MIT Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, *56*(4), 927–935.

Longuet-Higgins, H. C. (1981). Artificial intelligence — a new theoretical psychology? *Cognition*, *10*(1–3), 197–200.

Loui, P., Wu, E. H., Wessel, D. L., & Knight, R. T. (2009). A generalized mechanism for perception of pitch patterns. *Journal of Neuroscience*, *29*(2), 454–459.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the entropy of music: An experiment with Bach chorale melodies. *Leonardo*, *2*(1), 81–88.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness and evolution. *Music Perception*, *23*(1), 29–59.

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

Meyer, L. B. (1957). Meaning in music and information theory. *Journal of Aesthetics and Art Criticism*, *15*(4), 412–424.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage*, *38*, 331–345.

Moffat, A. (1990). Implementing the PPM data compression scheme. *IEEE Transactions on Communications*, *38*(11), 1917–1921.

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

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

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realisation model*. Chicago: University of Chicago Press.

Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity and liking for popular music. *Psychomusicology*, *14*, 77–93.

Oram, N. & Cuddy, L.L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*(2), 103–118.

Panksepp, J. (1995). The emotional sources of ''chills'' induced by music. *Music Perception*, *13*, 171–207.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology: CB*, *19*(10), 827–830.

Pearce, M.T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, London, UK: Department of Computing, City University.

Pearce, M. T., Conklin, D., & Wiggins, G. A. (2005). Methods for combining statistical models of music. In U. K. Wiil (Ed.), *Computer music modelling and retrieval* (pp. 295–312). Berlin: Springer.

Pearce, M. T., Herrojo Ruiz, M., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage*, *50*(1), 303–314.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010a). Melodic grouping in music information retrieval: New methods and applications. In Z. Ras & A. Wieczorkowska (Eds.), *Advances in music information retrieval, volume 274 of studies in computational intelligence* (pp. 364–388). Berlin/Heidelberg: Springer.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010b). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, *9*, 1367–1391.

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. T., & Wiggins, G. A. (2007). Evaluating cognitive models of musical composition. In A. Cardoso & G. A. Wiggins (Eds.), *Proceedings of the 4th international joint workshop on computational creativity* (pp. 73–80). London: Goldsmiths, University of London.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

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

Plotkin, H. (1998). *Evolution in mind*. Cambridge, MA: Harvard University Press.

Ponsford, D., Wiggins, G. A., & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research*, *28*(2), 150–177.

Potter, K., Wiggins, G. A., & Pearce, M. T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicæ Scientiæ*, *11*(2), 295–322.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643.

Ritchie, G. (2003). *The linguistic analysis of jokes*. London/New York: Routledge.

Rohrmeier, M., & Rebuschat, P. (in press). Implicit learning of music: What do we know today? *Topics in Cognitive Science*.

Ron, D., Singer, Y., & Tishby, N. (1996). The power of amnesia: Learning probabilistic automata with variable memory length. *Machine Learning*, *25*(2–3), 117–149.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Exploration in the microstructure of cognition*, volumes 1 and 2. Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*, 257–262.

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current biology: CB*, *19*(10), 831–836.

Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide [computer database]*. Menlo Park, CA: CCARH.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realisation model. *Cognition*, *58*(1), 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, *14*(3), 295–318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511–537.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*(3), 262–266.

Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, *7*(2), 109–150.

Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicology*, *9*(2), 122–147.

Schmuckler, M. A. (1997). Expectancy effects in memory for melodies. *Canadian Journal of Experimental Psychology*, *51*(4), 292–305.

Schoenberg, A. (1974). *Letters*. London: Faber. Edited by Erwin Stein. Translated from the original German by Eithne Wilkins and Ernst Kaiser.

Schultz, W., Dayan, P., & Montague, R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Shanahan, M. (2010). *Embodiment and the inner life: Cognition and consciousness in the space of possible minds*. Oxford, UK: Oxford University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. and 623–656.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell System Technical Journal*, *30*, 50–64.

Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 1–47). Cambridge, MA: MIT Press.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*, 110–120.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*, 978–982.

Snyder, J. L. (1990). Entropy as a measure of music style: The influence of a priori assumptions. *Music Theory Spectrum*, *12*(1), 121–160.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (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.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception and Psychophysics*, *59*(7), 1069–1076.

Tillmann, B. (2011). Music and Language Perception: Expectations, structural integration and cognitive sequencing. *Topics in Cognitive Science*, doi: 10.1111/j.1756-8765.2012.01209.x.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of music: A self-organizing approach. *Psychological Review*, *107*, 885–913.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception*, *16*, 99–118.

Todd, P. M. (1988). A sequential neural network design for musical applications. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 connectionist models summer school* (pp. 76–84). San Mateo, CA: Morgan Kaufmann.

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

Unyk, A. M., & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, *7*(1), 3–23.

Wallin, N. L., Merker, B., & Brown, S. (Eds.) (1999). *The origins of music*. Cambridge, MA: MIT Press.

Whorley, R. P., Pearce, M. T., & Wiggins, G. A. (2008). Computational modelling of the cognition of harmonic movement. In K. Miyazaki, M. Adachi, Y. Hiraga, Y. Nakajima and M. Tsuzaki (Eds.), *Abstracts of the 10th international conference on music perception and cognition* (p. 212). Japan: Sapporo.

Whorley, R., Wiggins, G. A., Rhodes, C., & Pearce, M. (2010). Development of techniques for the computational modelling of harmony. In D. Ventura, et al. (Eds.), *Proceedings of the first international conference on computational creativity* (pp. 11–15). University of Lisbon: Portugal.

Wiggins, G. A. (2006a). A preliminary framework for description, analysis and comparison of creative systems. *Journal of Knowledge Based Systems*, *19*(7), 449–458.

Wiggins, G. A. (2006b). Searching for computational creativity. *New Generation Computing*, *24*(3), 209–222.

Wiggins, G. A. (2011a). Computer models of (music) cognition. In P. Rebuschat, M. Rohrmeier, I. Cross, & J. Hawkins (Eds.), *Language and music as cognitive systems* (pp. 169–188). Oxford, UK: Oxford University Press. In press.

Wiggins, G. A. (2011b). ''I let the music speak'': cross-domain application of a cognitive model of musical learning. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition*. Amsterdam, The Netherlands: Mouton De Gruyter.

Wiggins, G. A., Harris, M., & Smaill, A. (1989). Representing music for analysis and composition. In M. Balaban, K. Ebcioğlu, O. Laske, C. Lischka, & L. Soriso (Eds.), *Proceedings of the second workshop on AI and music* (pp. 63–71). Menlo Park, CA: AAAI.

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

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, Monograph Supplement*, *9*, 1–27.