David Temperley, *Music and Probability*. Cambridge, Massachusetts: MIT Press, 2007, ISBN-13: 978-0-262-20166-7 (hardcover) $40.00.

PROBABILITY theory and Bayesian reasoning have been used widely and successfully for modelling perception and cognition in recent times (e.g. Chater et al., 2006, and Chater and Manning, 2006) but, for some reason, application of these approaches in musical research has been rather sporadic, even though many scholars since Leonard Meyer (1957) have noted that many concepts in music might lend themselves naturally to a formulation in probability theory. So, David Temperley is right to say, in the introduction to his new book, that "the time is ripe, then, for a reconsideration of music and probability" (p. 1).

With *Music and Probability,* Temperley sets out to fulfill two main tasks: to give an introduction to Bayesian reasoning and probabilistic models; and to present actual research that demonstrates the strengths of this approach when employed for modelling musical phenomena. For the latter, the book not only covers his own probabilistic models but also summarises very concisely and thoroughly related work by other researchers (this is one of his strengths as a writer). At the very beginning, three basic principles are proposed: that "perception is an inferential, multileveled, uncertain process"; that "our knowledge of probabilities comes, in large part, from regularities in the environment"; and that "producers of communication are sensitive to, and affected by, its probabilistic nature" (p. 3).

These axiomatic statements provide the motivation for Temperley's engagement with probabilistic models, and lead into an introduction to probability theory sufficient to provide an intuitive understanding of the concepts required to understand later material. Conditional probability is introduced and followed through to Bayes' rule, used throughout in the book, most often in the form:

$$P \text{ (structure|surface)} \propto P \text{ (surface|structure)} \, P \text{ (structure)} \quad \text{(p. 11),}$$

where P (structure|surface) denotes the probability that a given musical surface has a certain structure, and where 'surface' and 'structure' refer respectively to some directly observable factor in musical data (e.g. notes in a bar) and some underlying structure to be inferred (e.g. the tonality of that bar).

With an example borrowed from speech recognition, Temperley demonstrates that one of the advantages of Bayesian reasoning is that inference from probabilistic models can be informed by prior domain knowledge. It is unfortunate that this example is somewhat atypical of the remainder of the book, being a rare case where a non-uniform prior distribution is used and the full posterior distribution is calculated.

Other probabilistic concepts are discussed in the introductory chapters, most importantly entropy and cross-entropy in the context of model-fitting. The discussion of the need to be careful about assumptions made, is well put, and leads into a brief introduction to finite-state and hidden Markov models.

The introductory section concludes with a survey of early work on probability and music, with particular reference to the observations from Cohen (1962) about the need to model listeners' background appropriately. Temperley's own criticism of previous work employing probabilistic models is that it modelled the stream of surface events in music directly, rather than building models of underlying structure, which he argues is closer to the listening process.

Temperley's Models

In chapters 3, 4, and 6, Temperley presents his three main probabilistic models. The application and discussion of these models covers the major part of remainder of the book. Chapter 3 introduces the rhythm model whose purpose is to infer a metrical grid (the structure) from a monodic sequence of note onsets (the surface). In comparison to other Bayesian approaches to this problem (e.g., Raphael, 2002; Cemgil et al., 2000), Temperley is concerned with inferring a complete metrical grid (rather than a score position relative to the bar) using a simple model with as few parameters as possible. Following the Bayesian approach, he selects the most probable grid given the onset pattern of a concrete rhythm sequence. This, in turn, requires a model that estimates the likelihood of an onset pattern given a metrical grid and the prior probability of a metrical grid; the most probable grid is then the one that maximises the product of the likelihood and the prior over all possible grids. The grids considered consist of three

hierarchical levels: the tactus, a higher level and a lower level. In order to infer the prior probabilities of metrical grids, Temperley introduces a number of parameters that control factors such as the relative probability of a duple or triple upper level, the initial tactus interval and so on. He estimates these parameters either empirically from the Essen Folk Song Collection (EFSC) or he assigns reasonable values himself. A dynamic programming solution is presented to reduce the computational complexity of maximising the product of the likelihood and the prior over all possible grids. In contrast to the overall tone of this book and to the extremely clear description of dynamic programming in his previous book (Temperley, 2001), the description provided here is rather dense and technical.

The rhythm model is tested on Temperley's own performances of 65 melodies from the EFSC in a quantisation and note alignment task, using the proportion of note onsets correctly aligned with beats at the correct level as an evaluation metric. In comparison to his Melisma model (Temperley, 2001), the Bayesian model performs well on level 1 beats (supra-tactus) but somewhat less well on levels 2 (tactus) and 3 (sub-tactus). Further applications of the rhythm model are described in chapter 5. In an error detection task using 10 randomly distorted versions for each of the same 65 melodies from the EFSC, the model assigns the undistorted folk melody a higher probability than the distorted version in 82% of cases. The modelling of rhythmic expectations is also discussed, briefly and in qualitative terms, as a potential additional application of the rhythm model.

In chapter 4, Temperley presents his pitch model whose purpose is to infer a musical key (the structure) from a monodic sequence of pitches (the surface). The chapter begins with a clear review of related research but fails to specifically motivate the need for a new model, probabilistic or otherwise. As with the rhythm model, the pitch model is based on a set of assumptions about how surfaces are generated from structures. These assumptions are embodied in distributions (whose parameters are estimated from the EFSC) governing the range of a melody (range profile), the pitch proximity of successive notes (proximity profile) and the conditional probability of chromatic scale degrees given a major or minor key (the key profile). The generative process involves selecting a key and central pitch (from distributions estimated from the EFSC) and then sampling successive pitches from an RPK profile (obtained by taking the product of the Range, Proximity and Key profiles). Therefore, the probability of a given pitch in a melody is the normalised product of its probability in the range profile (given the central pitch), its probability in the proximity profile (given the previous pitch) and its probability in the key profile (given the selected key). This model can be used to estimate the probability of a pitch sequence given a key; the key that maximises the likelihood of a given pitch sequence is selected. The pitch model is tested in several ways. First, its performance as a model of key finding is examined on the same 65 melodies from the EFSC used in chapter 3 and the 48 fugue subjects of Bach's *Well Tempered Clavier*. The predicted key was the same as the annotated key in 87.7% and 83.3% of cases respectively. This performance appears to be comparable to other existing key-finding algorithms.

Temperley goes on to use the pitch model to predict expectations in terms of the conditional probability of a pitch given a preceding sequence of pitches. Using parameters estimated from the EFSC, this model achieves a reasonable fit to a set of data reflecting human pitch expectations in single interval contexts (Cuddy & Lunny, 1995). By hand-optimising the parameters to the data, however, Temperley was able to achieve a better fit than existing rule-based models (Cuddy & Lunny, 1995; Schellenberg, 1997), arguing that this optimisation is analogous to the least squares fitting used in multiple regression analysis of these models. Finally, on an error detection test (comparable to that used for the rhythm model), the pitch model achieved a level of 88.2% accuracy in spotting randomly distorted folk melodies.

To find the key and possible key modulations in polyphonic pieces Temperley proposes a procedure that differs slightly from these pitch and rhythm models. In contrast to the pitch and rhythm models, the polyphonic key-finding model discards sequential information, dealing only with unordered collections of note occurrences (represented as scale degrees) relative to the key of a passage of music. He derives estimates of the scale degree frequencies from a corpus of common-practice pieces (Kostka & Payne, 1995) reflecting early 19th century music which come with detailed annotations of key and modulations. He then uses a generative model that calculates the product of the probabilities of scale degree sets given a key and modulation within a sequence of keys. Again, dynamic programming is used to find the sequence of keys that maximises this probability for a given sequence of scale degree sets. When evaluating the accuracy of the model, it appears to be comparable to the Krumhansl-Schmuckler model (Krumhansl, 1990) using the right scale degree profile and it is very similar to his own non-probabilistic model from 2001. In fact, as he mentioned in the original publication of the polyphonic key finding model (Temperley, 2004), the new probabilistic model is structurally very similar to the earlier rule-based model.

And given the slightly better performance of the rule-based model, one is again left wondering about the benefits of the Bayesian approach.

Other than the task of finding the key for a passage of polyphonic music, Temperley explores using the model to address questions that are somewhat more interesting from a musicological or music-theoretic perspective. With references to the literature, he discusses the relations between keys, the measurement of tonalness and tonal ambiguity, differences between major and minor, and harmonic ambiguities in common practice music and music from the classical period.

The Larger Picture

In the last two chapters, Temperley takes a step away from the applications of the individual models and discusses various topics in music theory and music perception. It is one of the virtues of this book that it tries to pull several strands of thinking together, which come from areas such as empirical musicology, music theory, perception and cognition, as well as music information retrieval. But once again the reader might struggle to find a convincing connection between the problems addressed and the necessity of the probabilistic approach in doing so.

In Chapter 9, Temperley considers broader issues, making the important point that Bayesian model comparison allows the comparative evaluation of different models of the same musical surface, giving examples using tempo and Schenkerian analysis. In the latter case, he explores how 'Schenkerian theory' might be evaluated as a theory, starting from a hypothetical probabilistic, generative model of the theory and then using cross entropy to assess the probability of the theory given musical examples. Implicit in this is a comparison with other generative theories, such as his own discussed earlier in the book. A caution here seems necessary – this comparison is only valid where the models aim to explain the same musical surface and nothing else. Whatever Schenkerian analysis claims for itself (and an exploration of its diverse churches and their philosophies is well beyond the scope of both the book and this review; see Cook, 1987, chapters 2 and 6), it never claims to account fully for all the surface pitches or rhythms on the page: since the elaborative processes considered are simply a restatement of musical theories of counterpoint, and since the background structures permit significant insertion of 'implied' pitches, a generative Schenkerian model is unlikely to perform better than other music-theoretic models and could easily perform worse. Moreover, since Temperley quite rightly suggests an important supplementary evaluation of models based on a quantitative version of Occam's razor, the theory is almost certain to fail. Should one throw Schenkerian theory away if this happens? Schenkerians who agree with the foregoing reasoning might feel justified in arguing that the method "aims to omit inessentials and to highlight important relationships" (Cook, 1987, p. 28). Thus, whilst Temperley's explanation of how an automated analysis system might operate is a useful one as illustration of another application of Bayesian reasoning, it is unclear how Schenkerian theory could be evaluated purely in terms of notated surface pitches and rhythms.

In the final chapter, Temperley steps from a purely probabilistic world to a more directly cognitive one, introducing communicative pressure, which relates to the third of the basic principles stated in the introduction and is probably the most important cognitive concept of the entire book. It is defined as a general principle for enabling successful communication through agreement between sender and receiver about the constraints for retrieving a message (structure) from a communicative signal (surface). Adopting the old sender-channel-receiver model from communication theory for the musical context, he exemplifies this principle by discussing several phenomena from different musical styles, among them voice-leading rules in polyphonic art music, the trade-off between expressive tempo [1] and syncopation in common-practice art music vs. popular music, and the trade-off between chord inversions and extensions in common-practice music and jazz. The essence of communicative pressure is that a composer has to put strong constraints on one musical dimension if he wants to make expressive use of another related dimension (e.g. you can't have a high level of syncopation and tempo variation at the same time without blurring the rhythmic structure).

There are at least two points of concern regarding the empirical support for this concept and regarding the concept itself. Addressing the former, Temperley continues to argue empirically as he did in previous chapters, but the empirical evidence he uses becomes quite sparse and misleading at times. So his style of argument approximates the convince-by-example approach widely used in music theory, but disguised with empirical terminology. It is generally accepted, for example, that the use of expressive tempo in Western Classical music increased throughout the 19th century, but not only is this very hard, if

not impossible, to evaluate, but it also makes it difficult to take seriously the implication that communicative pressure requires maximal communication between sender and receiver. Temperley's observation of an increase in regularity in the right-hand patterns of piano music is intended to explain the increase of expressive timing between the 18th and 19th century, but is fallacious and based on a comparison between different genres of music (something that he partly addresses in a footnote).

The discussion of voice-leading is similarly awkward, failing to discuss or acknowledge the points at which the 'rules' are not applied with the same rigour, even though the effects are audible (for example, and considering only Western music: very early polyphonic choral music, e.g. Perotin's famous *Sederunt Principes;* much instrumental music, especially for chord-playing instruments; and popular musics) or the points at which they are applied where they would not be perceptible (such as Tallis's *Spem in Alium* which contains no parallel fifths even though, with up to 40 independent voices at a time, it is unlikely that they would be noticed)

For the assessment of the occurrences of chord inversion and extensions the common practice repertoire is compared to 50 jazz pieces as laid down in the lead-sheet format of the *New Real Book.* More inversions are observed in the common practice pieces but more chord extensions in the jazz pieces. It seems surprisingly naive to assume that *Real Book* lead-sheets reflect actual compositions or performances. They are clearly meant as indications of the significant melodic and harmonic constituents of a piece of popular music, but for most standard jazz pieces (see Tagg, 2003) there is no indication whatsoever of which note a bass player is supposed to play for a certain chord or what the spacing of a piano chord should be. [2] This would in fact be contrary to the very core principles of jazz. In the particular book that he uses, the arrangements are generally slightly simplified for reading ease and only show bass progressions in a small number of cases. If taken at all as an example repertoire, one would have to transcribe jazz performances to observe the interplay between actual bass notes and chord extensions. Even so, Temperley's core observation, that there is ambiguity between chords in jazz is significant here. Often, the only difference between two chord labels is the choice of bass note; in such a situation, the identity of the chord is defined by the bass and, as such, inversion is impossible, since it would change the chord label.

The second point of concern relates to the simplistic communication model which forms the basis of the communicative pressure principle. Firstly, there is a long history of fierce debate about whether music can be considered a language with the aim of communicating messages; the strong opposition to this claim in the literature isn't even mentioned (see e.g., Agawu, 1999; Samson, 1999; Storr, 1992). Secondly, even if music can be considered a language, would the musical structure be the message? This again is a venerable and controversial topic in music theory dating back to Hanslick (1854). Given the debate about communication and meaning in music, it is highly controversial whether these concepts can be tied to notes on a score. Even if they can, is it really the key and the metre that a composer wants to convey to a listener? In short, the theoretical foundations of the concept of communicative pressure in music have yet to be laid.

Another point that seems problematic about the concept of communicative pressure is that, unlike David Huron in his recent book (Huron, 2006), Temperley makes no clear distinction between musical models of perception and production (or even between perception and transcription). In chapter 9, for example, he considers differences between compositional styles in terms of different parameterisations of the pitch and rhythm models. This assumption that producers and perceivers share essentially the same cognitive representations and machinery for processing music is expanded in chapter 10 as the basis for the concept of communicative pressure. Here, the same generative systems are clearly being used as models of both perception and composition, in spite of the fact that Temperley states (of the rhythm model) that (p. 31): "This generative model – like those presented in later chapters – is not intended as a model of the creative process, only as a model of how listeners might model that process for the purposes of meter perception." However, it seems likely that the cognitive mechanisms underlying the perception and production of music do differ (Huron, 2006; Sadakata et al., 2006) and that their comparison in a specific domain could reveal interesting insights.

Discussion

It is not only the last two chapters that exhibit a weak approach towards model evaluation and hypothesis testing. In chapter 3, for example, the fact that Temperley performs the melodies used to test the rhythm model is a potential source of bias. One also wonders why the same 65 folk melodies are used in all tests of the pitch and the rhythm models: more reliable estimates of generalisation performance could be

obtained by examining other training and test sets. Furthermore, when comparing the performance of other models, significance values for differences are not given. [3] We learn, for example, that three polyphonic key finding models achieve accuracies of 84.9%, 86.3% and 80.4% respectively but have no idea of whether these performance differences are really significant. Finally, the comparison of the proposed models to existing models is often incomplete (e.g., chapters 3 and 4) or absent, as in chapter 8 which fails to acknowledge a venerable tradition of research that has explored in detail the use of information-theoretic measures (including cross-entropy) for evaluating, comparing and selecting predictive models of musical structure (see e.g., Cohen, 1962; Conklin & Witten, 1995; Pearce & Wiggins, 2004).

As just one example of the problems arising from evaluation methodology, it is worth looking at chapter 7, where Temperley tests a hypothesis supporting his concept of communicative pressure discussed in chapter 10. The hypothesis states that the pitch-class content of a V-I cadence in the major is substantially more 'tonally ambiguous' than either a V7-I cadence or a V-I cadence in the minor. At points where the composer is attempting to unambiguously communicate tonality (such as, by Temperley's argument, the final point of a movement), V-I transitions in the major should be disfavoured. Table 7.6 illustrates the results from his empirical evaluation of final cadences of piano sonatas from Haydn, Mozart, and Beethoven, and he makes the claim (p.134) that for all three composers, it can be seen that the proportion of V-I cadences in minor-key movements is much higher than in major-key movements; in all three corpora, this difference is reported as being statistically significant according to a x^2 test and this is taken to be a successful prediction of his model of tonal ambiguity in compositional practice. Temperley presents the data from his empirical evaluation in sufficient detail that readers can perform their own analyses from their own perspectives (effectively, with different prior probabilities), such that any Bayesian analyst can model this experiment and generate likelihoods and posterior distributions himself. Noting first that the x^2 statistic and p-value for the Mozart corpus are incorrect in table 7.6 [4], we agree that the raw data tabulated are certainly suggestive [5] and merit further examination. Looking into the empirical procedures in greater detail, then, we consider the method by which Temperley classifies the final cadence into these categories. [6] The method includes a strict set of criteria for accepting the final cadence of a movement into the test set, and indeed many sonata movements have been eliminated from consideration. [7] This is problematic, as it is likely that the effect is to remove a disproportionate number of tonally-ambiguous (in Temperley's pitch-class sense) final sections, which evidently has an impact on what the analysis is purporting to demonstrate. Another danger of the overly simple procedure for identifying final cadences and their classification into V7-I or V-I is that it ignores (possibly varied) repetitions and simply identifies the final cadence wrongly. [8]

In conclusion with respect to this single example, we take the view that (aside from the error in the table of results) there are several major methodological problems in acquiring the data analysed, which are in some respects representative of the evaluation methodology that Temperley applies at several points in his book.

Leaving aside the evaluation methodology, another aspect of the book that is worth discussing is Temperley's treatment of cognitive and perceptual processes. Although one of his aims is to "to uncover the mental processes and representations involved in musical behaviour" (p. 5), he makes little attempt to compare the behaviour of the various models proposed to human behaviour. Instead, music analysis is generally used as a proxy for music perception. This approach avoids the need to run time-consuming experiments on listeners and is relatively common in research on music cognition. However, while musicological analyses will often accord with perceived musical structures, the correspondence cannot always be assumed. For example, good performance of the pitch model in predicting the keys annotated by musicologists does not necessarily indicate a good model of human key perception. Listeners might perform much worse than the model with respect to the annotated keys; reproducing the failures would be just as important in evaluating a cognitive model as reproducing the successes. One of the few places that Temperley does use behavioural data from human listeners is in evaluating the pitch model against the experimental data reported by Cuddy and Lunny (1995). However, this is not a very demanding test of a model of melodic expectations since the judgements were elicited in the context of a single isolated interval. It is more interesting to ask whether such a model, without substantial reconfiguration, continues to predict the expectations of listeners elicited in a range of melodic contexts differing in length, style and structural setting (Pearce & Wiggins, 2006).

In summary, Temperley is right to assert that there is a pressing need for a textbook on Bayesian reasoning and probabilistic modelling in music. However, it is not clear that this book completely satisfies this need. The introduction to probabilistic modelling from a Bayesian perspective seems all too brief at

times and the interested reader will have to resort to a technical textbook for more thorough explanations. Furthermore, this book is unlikely to convince a non-Bayesian audience of the specific advantages of this approach to music cognition. The performance of his Bayesian models on specific tasks is not clearly superior to existing models, and the claims of improved rigour and objectivity, frequently made throughout the book, are at odds with the methodological shortcomings in empirical evaluation and theoretical reasoning.

Marcus Pearce, Daniel Müllensiefen, David Lewis, Christophe Rhodes
Goldsmiths College, University of London

NOTES

[1] Temperley here uses the unfortunate term rubato, clearly meaning the later usage of this word, as unstructured, expressive rhythmic variation. The earlier usage, which co-existed with this throughout the nineteenth century, would suggest a strict tempo with timing variations only below the bar level. Since these two meanings would have different implications for this discussion, we choose to avoid the term altogether.

[2] Notable exceptions include, for example, pieces using slash chords where the (often non-triadic) bass note is part of the harmonic sonority and has to be specified separately, as in many funk- and rock-jazz pieces. Slash chords are in any case a harmonic phenomenon that clearly differs from inversions in classical harmony theory.

[3] While a Bayesian statistician might not wish to perform a significance test in the usual frequentist style, there are nevertheless ways of summarising the belief that the performance of two experiments has yielded a different result, for instance by examining the variability of the results of the experiments or the posterior distributions of any inferred parameters.

[4] Consider the Mozart corpus and estimating from the tabulated results, the null hypothesis is $p(V\text{-}I) = 7/24$ and $p(V7\text{-}I) = 17/24$. This then gives expected observation counts of

	Major	Minor
V-I	6.125	0.875
V7-I	14.875	2.125

and a x^2 statistic, calculated by $\sum \frac{(O_i - E_i)^2}{E_i}$, of 2.33, corresponding to $p > .05$ from a x^2 distribution with d.f.=1 – or a less significant statistic if Yates' continuity correction is applied. The other values given in table 7.6 are reproducible following the above method (without Yates' correction).

[5] In the Mozart case, they can be no more than suggestive – with only three pieces of data in the Minor category, only very extreme values for the proportion are excluded; the evidence for the Haydn and Beethoven cases is stronger, as there are more data included.

[6] Working on a 'tactus' (defined by the time-signature) granularity, Temperley identifies the 'final tonic span' as the longest continuous region from the end containing scale degrees ˆ1, ˆ3 and ˆ5 only; then the 'final dominant span' as the longest continuous region abutting the final tonic span containing no scale degrees but ˆ5, ˆ7, ˆ2 and ˆ4, containing at least one ˆ5 and ˆ7, and with ˆ5 as the lowest pitch. If either of these final spans cannot be found, then the movement is discarded from consideration; otherwise, the final cadence is classified as V7 - I if the final dominant span contains a ˆ4 and V- I otherwise.

[7] While Temperley counts 24 final cadences from Mozart sonata movements, he disregards another 30 – and the proportion eliminated is even greater for both Haydn and Beethoven, where 50 and 60 movements respectively have been excluded. In addition to these exclusions, early Haydn sonatas are removed from consideration, on the basis that they have thin two-part textures and often do not use 'complete' harmonies at cadences.

[8] Consider the end of Mozart's Sonata in C minor, K457, first movement as displayed in the Figure 1 below. Temperley's prescription for finding the final cadence will identify regions labelled V_T and I_T, and will count this movement as having a V-I final cadence. Note that the bar in which V_T is located is an analogue of the bar before, where the pitch set for the analogous part of the bar is the V7 pitch class set, which suggests that the underlying harmony (considering the last crotchet of the two bars in aggregate) is suggestive of V7. Two other things to note are that although the time signature of this sonata movement is 4/4, the pulse is minim-based rather than crotchet – and using a minim-based segmentation rather than a crotchet-based one would have led to this movement being excluded. More fundamentally still, though, it might validly be argued that the moment of tonal closure is not the final cadence but eight bars before it (marked V_X, with the last ten bars of the movement being an elaboration on the tonic). Taking that view, the cadence is clearly V7-I. Indeed, as Schoenberg (1967, p. 185) observes "Since many movements have no codas, it is evident that the coda must be considered as an extrinsic addition. The assumption that it serves to establish the tonality is hardly justified; it could scarcely compensate for failure to establish the tonality in the previous sections."

Fig. 1: Final bars of W.A. Mozart's Sonata in C minor, K457, first movement. The final cadence as identified by Temperley's procedure is labeled V_T and I_T. An alternative candidate for the final cadence is marked by V_X and I_X in the score.

REFERENCES

Agawu, K. (1999). The challenge of semiotics. In: *Rethinking Music*. N. Cook & M. Everist (Eds.). Oxford: Oxford University Press, pp. 138-160.

Chater, N., Tenenbaum, J.B. & Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *TRENDS in Cognitive Sciences*, Vol. 10, No. 7, pp. 287-291.

Chater, N. & Manning, C.D. (2006). Probabilistic models of language processing and acquisition. *TRENDS in Cognitive Sciences*, Vol. 10, No. 7, pp. 335-344.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, Vol. 7, No. 2, pp. 137–163.

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, Vol. 24, No. 1, pp. 51–73.

Cook, N. (1987). *A Guide to Musical Analysis*. Oxford: Oxford University Press.

Cuddy, L. L. & Lunny, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgements of continuity. *Perception & Psychophysics*, Vol. 57, No. 4, pp. 451–462.

Hanslick, E. (1854). *Vom_Musikalisch-Schönen: Ein Beitrag zur Revision der Aestethik der Tonkunst*. Leipzig: Rudolph Weigel.

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

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Meyer, L. B. (1967). Meaning in music and information theory. In: *Music, the Arts, and Ideas*. Chicago: University of Chicago Press, pp. 5–21.

Pearce, M. T. & Wiggins, G. A. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, Vol. 33, No. 4, pp. 367–385.

Pearce, M. T. & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, Vol. 23, No. 5, pp. 377–405.

Sadakata, M., Desain, P.W.M., & Honing, H.J. (2006). The Bayesian way to relate rhythm perception and production. *Music Perception,* Vol. 23, No. 3, pp. 267-286.

Samson, J. (1999). Analysis in context. In: *Rethinking Music*. N. Cook & M. Everist (Eds.). Oxford: Oxford University Press, pp. 35-54.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, Vol. 14, No. 3, pp. 295–318.

Storr, A (1992). *Music and the Mind*. New York: Ballantine Books.

Tagg, P. (2003). Leadsheet. In: *Continuum Encyclopedia of Popular Music of the World*. J. Shepherd, D. Horn & D. Laing (Eds.). New York: Continuum.

Schoenberg, A. (1970). *Fundamentals of Musical Composition* (G. Strong and L. Stein, ed.). Faber and Faber, London. (Original work published 1967)

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.