- **David Temperley, *The Cognition of Basic Musical Structures* Cambridge, MA: MIT Press, 2001, 404 pp. ISBN 0-262-20134-8.**

REVIEWER:
David Meredith
Department of Computing, City University, London.

ADDRESS FOR CORRESPONDENCE:
David Meredith,
The Barn,
Middle Broad Drove,
Tydd St. Giles,
Wisbech,
CAMBS.,
PE13 5PA,
UK.

Introduction

In *The Cognition of Basic Musical Structures*, David Temperley presents a computational theory of music cognition that is deeply influenced by Lerdahl and Jackendoff's (1983) *A Generative Theory of Tonal Music* (henceforth, *GTTM*). Like Lerdahl and Jackendoff, Temperley attempts to explain the cognition of common-practice music by means of a system that generates structural descriptions from musical 'surfaces'. As in *GTTM*, the hypothesis underlying Temperley's theory is that the analysis it generates for a passage of common-practice music correctly describes certain aspects of how the passage is interpreted by listeners who are experienced in the idiom.

Like *GTTM*, Temperley's theory consists of a number of *preference rule systems*, each containing *well-formedness rules* that define a class of structural descriptions and *preference rules* that specify an optimal structural description for a given input. Temperley presents preference rule systems for six aspects of musical structure: metre, phrasing, counterpoint, harmony, key and pitch spelling.

In collaboration with Daniel Sleator, Temperley has implemented most of his theory as computer programs. Finding the best analysis satisfying a given set of preference rules is an *optimisation problem* that can be solved using a technique called *dynamic programming* (Bellman, 1957; Cormen et al., 1990, Chapter 16). Each of Temperley's six models is implemented using the dynamic programming technique, the systems for harmony and pitch spelling being combined into a single program.

Temperley evaluated each of his six models using objective tests. For example, he tested his metre program on a corpus of 46 excerpts from a theory workbook by Kostka and Payne (1995), comparing the output of the program with the scores of the excerpts.

The input representation

Temperley's systems take 'piano-roll' representations as input (Temperley, 2001, pp. 9–12). Such representations give the pitch of each note (expressed by its MIDI number) and its onset- and offset-time (in milliseconds). Piano-roll representations can automatically be derived from MIDI files. If they are derived from MIDI files created using notation software, then they will be 'metronomic' representations in which the onset-times and durations are strictly proportional to their notated values. Piano-roll representations can also be derived from performances on MIDI instruments, in which case they will be 'expressive' representations in which the timing is not strictly regular because of slight imperfections and expressive tempo changes. Temperley's theory works with both 'metronomic' and 'expressive' piano-roll representations (Temperley, 2001, p. 12).

As Temperley (2001, p. 361, note 2) observes, the programs implementing his theory can be used as the basis of a tool for computing a staff-notation score from a performance on MIDI instruments. If this had been his primary motivation, then using piano-roll representations as input would certainly have been appropriate.

However, Temperley's goal was to construct a computational model of how 'we extract basic kinds of musical information…from music *as we hear it*' [my emphasis] (Temperley, 2001, p. ix), and using piano-roll representations as input to such a model may be inappropriate, as I shall now explain.

Temperley (2001, p. 5) claims that one can model a cognitive process at the computational level 'without worrying about how it might be instantiated neurologically'. However, in order to be feasible, a computational model of musical listening must be *consistent* with the neurology, physiology, psychophysics and psychology underlying musical listening. In particular, the information used by the model must be either present in the input to that part of the auditory system that performs the process being modelled, or it must be derivable from this input using a procedure that could feasibly be performed by that part of the auditory system. I believe that some of the information used by Temperley's preference rule systems for metre, phrasing, harmony and key is not available to those parts of the auditory system that extract these types of information.

A piano-roll representation of a musical passage specifies the notes that need to be played in order to perform the passage. For most listeners, even singing all the notes in a single close position chord is harder than tapping in time with a passage or identifying phrase boundaries. In general, identifying the notes that have been played is harder even than singing the root of a chord or the tonic of a short passage. Scheirer (2000, p. 70) goes so far as to claim that 'there are no mental entities that correspond to events happening in the music being heard'.

Temperley claims that his theory models the cognitive processes involved in tasks such as

- tapping in time to a passage;
- identifying phrase boundaries;
- identifying the root of a chord; and
- identifying the tonic of a passage.

These are all tasks that many listeners can perform successfully as they listen to a passage for the first time or immediately after hearing it once. Specifically, they are all tasks that can be performed successfully *without* first identifying all the notes that have been played. Indeed, even expert transcribers often have to listen many times to a complex passage in order to identify successfully the notes that have been played. This suggests that those parts of the auditory system responsible for identifying metre, phrasing, harmonic roots and key tonics during listening neither require nor have access to detailed information about the notes that have been played. Yet this information is explicitly encoded in the piano-roll representations used as input to Temperley's models of metre, phrasing, harmony and key.

Metrical structure

Most of the rules in Temperley's model of metre are borrowed from the metre component of *GTTM* (see Table 1). However, unlike most previous approaches, Temperley's metre model can process 'expressive' polyphonic input representations. Temperley's system first quantizes the event onsets and offsets in the input representation to the nearest 'pip' (35 ms). This improves the running time of the metre program by dividing the number of possible instants at which onsets and offsets can occur by 35.

Temperley's MPR 1 (Temperley, 2001, p. 32) states that strong beats should preferably be aligned with event-onsets and Temperley proposes that 'the more event-onsets at a time point, the better a beat location it is' (Temperley, 2001, p. 33). As Temperley points out, in a human performance the notes of a chord are seldom played precisely simultaneously leading to what he calls 'smudged chords' in which the event-onsets may differ by up to 20 ms (Temperley, 2001, p. 40). This means that some event-onsets that are intended to be simultaneous are not. Consequently, the number of event-onsets at a given time point in an 'expressive' input representation may not be a good indicator of how strong a beat it is. Temperley claims that quantizing the 'expressive' input into pips, causes the onset-times of the events in a smudged chord to be made equal. However, this would seem to be a rather hit-and-miss strategy, given that if a pip is 35ms long, then two events that are 20 ms apart are more likely to be quantized to *different* pips than to the same one!

I argued above that identifying all the note events in a passage is not a prerequisite for being able to tap in time to it. In particular, listening to a chord and then identifying the number of notes played is a hard task. Whereas most listeners can tap in time with a passage at various metrical levels, only highly trained transcribers can reliably identify the number of note events that occur simultaneously at each point in the passage *as they hear it*. Temperley's proposal that a listener's perception of the metrical strength of a time point is influenced by the number of event-onsets that occur at that time point is therefore not feasible.

Lerdahl and Jackendoff's (1983, pp. 72, 347) MWFR 4 specifies that beats at the tactus level and above must be 'equally spaced'. Clearly, this only applies if the input is a 'metronomic' representation. Temperley enables his model to handle 'expressive' representations by re-expressing Lerdahl and Jackendoff's MWFR 4 as a preference rule (MPR 3), stating that beats at each level should be 'maximally evenly spaced' (Temperley, 2001, p. 35). Despite this modification, his program produced rather better results with 'metronomic' input representations than with 'expressive' ones (Temperley, 2001, p. 45).

Although Temperley carried out objective tests to evaluate his theory of metre, he did not compare its performance with that of other models. He also does not mention any relevant work on metre published since 1996, ignoring, for example, important recent work on audio beat-tracking and tempo-tracking (Dixon, 1999, 2000; Dixon and Cambouropoulos, 2000; Scheirer, 1998, 2000).

Melodic phrase structure

Temperley's theory of melodic phrase structure (Temperley, 2001, Chapter 3) is the least well-developed and least successful of his models. It only works on monophonic music with a regular metric structure which must be explicitly supplied as input to the system. This phrase-structure model consists of just three rules, none of which take pitch structure into account.

The first of the phrase structure preference rules, PSPR 1 (Temperley, 2001, pp. 68, 358), is a modification of Lerdahl and Jackendoff's (1983, pp. 45, 344) GPR 2. It is a simple variation on the Gestalt principle of proximity.

The second rule, PSPR 2 (Temperley, 2001, pp. 69, 358), is a simplistic and *ad hoc* rule that claims listeners prefer phrases to contain about 8 notes. In practice, the average number of notes per phrase depends on the style, genre and instrumentation of the music. For example, as Temperley (2001, p. 82) himself points out, in instrumental music, phrases routinely contain many more than 8 notes. Even in the sample of simple folk songs that Temperley used to test his implementation, the average length of phrases was 9.2 notes, which meant that the model performed best on this particular corpus when PSPR 2 was modified to favour 10-note phrases (Temperley, 2001, p. 74). This illustrates just how *ad hoc* PSPR 2 is.

PSPR 3 (Temperley, 2001, pp. 70, 358) is a half-hearted attempt to model the effect of parallelism on grouping. Temperley acknowledges that identifying parallelism (i.e., significant repetitions) is an important factor in grouping structure (Temperley, 2001, p. 69), metrical structure (Temperley, 2001, pp. 49–51) and contrapuntal structure (Temperley, 2001, p. 113). However, he considers it to be a 'huge and complex problem' and so does not attempt to solve it in his preference rule systems, even though many of the mistakes made by these systems seem to be due to their failure to take parallelism into account.

In fact, in recent years, considerable progress has been made in constructing computer programs for discovering parallelism not only in monophonic music (e.g., Rolland, 1999) but, more recently, in polyphonic music also (Meredith et al., 2001, 2002). Indeed, Cambouropoulos (1998a,b, 2001b) has presented a sophisticated computational model of grouping and parallelism in monophonic music which performs considerably better than Temperley's simplistic model on a range of different musical styles. One wonders why Temperley omitted to mention this closely related work. Also, as with his model of metre, Temperley did not compare the performance of his melodic phrase structure model with that of other computational models of phrasing and grouping structure.

Contrapuntal structure

Temperley's model of contrapuntal structure is based on auditory stream segregation (Bregman, 1990). His CPR 1 (Temperley, 2001, p. 100) states that large pitch intervals should be avoided within streams. This clearly derives from the principle, discovered by van Noorden (1975), that listeners cannot hear a sequence of alternating tones as belonging to a single stream when the frequency difference between the tones is greater than a certain value known as the 'temporal coherence boundary' (TCB) (Bregman, 1990, pp. 59–60). However, the TCB increases approximately linearly with the duration of the tones in the sequence. This suggests that the likelihood of two consecutive tones being perceived to be in the same stream depends not only on the pitch interval between them but also their durations. In Temperley's model, each interval incurs a penalty that is proportional to its size. One wonders whether the system's performance could be improved by considering the durations of the tones involved in each interval. Also, Ortmann (1926) (cited by Bregman, 1990, p. 462) showed that, although the frequency of occurrence of intervals between consecutive tones in melodies is roughly inversely proportional to interval size, harmonic intervals occur more often than predicted by this simple rule.

One wonders whether the performance of Temperley's system could be improved by making the penalty imposed on each interval depend on its frequency of occurrence within melodies.

Temperley's counterpoint model accounts well for the effects of repetition, rate of alternation and frequency separation on the perception of streams in two-tone alternating sequences (Temperley, 2001, pp. 106–107; see also Bregman, 1990, pp. 58–65, 128–130). He also claims that listeners have a 'perceptual abhorrence … for having multiple simultaneous notes in a single stream' (Temperley, 2001, pp. 107–108) and his second contrapuntal well-formedness rule (CWFR 2) (Temperley, 2001, p. 98) ensures that this never happens in the analyses generated by his system.

However, it seems to me that musical streams often contain 'sound objects', each one *perceived* as a single percept but *notated and produced* using two or more notes. Consider, for example, a sequence of close position chords such as those in the left hand in Figure 1 or a theme played in octaves such as that in the right-hand in Figure 2. Thus, while it may be true that musical streams must contain no more than one 'unified percept' at any given instant, it is surely not true that each of these percepts must be a single note.

Temperley (2001, p. 366) suggests that 'polyphonic' streams like those in Figures 1 and 2 are actually 'higher-level' streams and he claims he is concerned only with the 'lowest-level' streams, in which no more than one note occurs at a time. However, his stated goal is to model music cognition not construct a music transcription system, so his concern should have been with what listeners *hear*. And while anyone who understands music notation might be able to *see* a multi-level, hierarchical organization of note-streams in a score, most of us are incapable of *hearing* such complexities. Thus, Temperley's claim that he is concerned only with streams that contain no more than one note at a time seems to conflict with his goal of explaining how listeners extract information from music as they hear it.

Pitch spelling

Temperley distinguishes between *neutral-pitch-classes* (NPCs), which are the usual pitch classes as used in pitch class set theory, and *tonal-pitch-classes* (TPCs) which are identified using note names with the octave designation omitted (i.e., G sharp, A flat, etc.). The basic idea underlying Temperley's pitch spelling system is that nearby events in the music have TPCs that are as close together as possible on the 'line of fifths'—an infinite unidimensional space in which each TPC is separated from the next by a perfect fifth. Thus, Temperley claims we prefer to spell the upper note in the final chord in Figure 3 as C sharp because it is nearer to the 'center-of-gravity' (COG) of the preceding notes when they are represented on the line of fifths (Temperley, 2001, p. 120).

According to harmonic theory (see, for example, Piston, 1978), the pitch name of a note is determined by its perceived function within the ambient key. Thus, in Figure 3 *there is no preferred spelling* for the upper tone in the final chord because the correct spelling depends upon how the fragment is continued. If it is continued as shown in Figure 4a, then the correct spelling would be C sharp; but if it is continued as shown in Figure 4b, then the correct spelling would be D flat. It may be that the continuation in Figure 4a is more expected than that in Figure 4b, but the way a tone is spelt is not determined by expectation, it is determined by its tonal function.

Temperley and Sleator implemented the pitch-spelling and harmonic structure models together in a single program in order to incorporate the 'Harmonic Feedback Rule' (TPR 3) (Temperley, 2001, p. 131) which causes the system to prefer TPC representations that result in 'good' harmonic representations. Temperley therefore does acknowledge that harmonic structure is a factor in pitch spelling. However, he emphasizes that 'one of the main claims' of his model 'is that spelling can be accomplished without relying on "top-down" key information' (Temperley, 2001, p. 126). Since one can only determine if a note is spelt correctly by analysing the harmonic and key structure of its context, it follows that any pitch-spelling method that does not employ harmony and key information must be just a 'rule-of-thumb' that may give the right results surprisingly often, but not for the right reasons and not in all cases. Temperley's claim that 'recognizing spelling distinctions at the level of pitches provides useful input in harmonic and key analysis' (Temperley, 2001, p. 122) therefore seems to be the reverse of what happens in perception.

It is not always easy, however, to determine the correct spelling of a note from the harmonic and key structure of its context. For example, as Piston (1978, p. 390) observes, the tenor E flat in the third and fourth bars of Figure 5 should be spelt as a D sharp if one perceives the harmonic progression here to be $^{+}\mathrm{II}^{2}$–I as shown. But spelling the soprano E flat in the fourth bar as D sharp would result in a strange melodic line.

As with his metre and phrasing models, Temperley fails to compare the performance of his pitch spelling program with that of other systems. He also again omits to mention certain important closely-related work—in particular, the pitch-spelling algorithm developed by Cambouropoulos (1996, 1998b, 2000). This is especially unfortunate because, like Temperley's system, Cambouropoulos's algorithm derives pitch-spelling without (explicitly) considering harmonic and key information. Also, Cambouropoulos (2001a) shows that his system performs better than Temperley's on a corpus of Mozart sonatas containing over 40000 notes.

Harmonic structure and key structure

Temperley divides the problem of harmonic analysis into two subproblems: root-finding and key-finding. His preference rule system for harmonic structure performs the first of these tasks, his key-structure system performs the second.

As I see it, the only problematic feature of Temperley's harmonic structure system is the fact that it requires TPC information as input (see discussion above). Even this problem is mitigated by the 'Harmonic Feedback Rule' (TPR 3) (Temperley, 2001, p. 131). Two nice features of this harmonic structure model are its ability to deal with unaccompanied melodies (Temperley, 2001, pp. 159–160) and the fact that it takes tempo into account (Temperley, 2001, pp. 161–162). The latter feature models the fact that some tones that are perceived as chord tones at slow tempos may be considered ornamental at faster tempos.

Temperley's key-finding algorithm is based on that of Krumhansl and Schmuckler (Krumhansl 1990, Chapter 4). This algorithm predicts that the perceived key of a segment is the one whose tonal hierarchy most closely correlates with an 'input vector' giving, for each pitch class, the sum of the durations of tones with that pitch class within the segment. As an example, Temperley presents a table (Temperley, 2001, pp. 175, Table 7.1) giving the correlation between the input vector for the first bar of 'Yankee-Doodle' (Figure 6) and the tonal hierarchy of each of the major and minor keys. Unfortunately, the values in Temperley's table are incorrect—the correct values are shown in Table 2. As can be seen, the tonal hierarchy (or key-profile) of G major correlates most strongly with the input vector of the 'Yankee-Doodle' extract, so the algorithm correctly predicts the perceived key of this segment.

Temperley modifies the Krumhansl-Schmuckler algorithm in three ways. First, he modifies the tonal hierarchies by increasing the difference between the diatonic and chromatic degrees, decreasing the relative value for the flattened seventh degree and increasing the value for the leading note by a relatively large amount (Temperley, 2001, p. 180). Second, he uses 'flat' input vectors in which the value for each pitch class is 1 or 0, indicating whether or not the pitch class is present in the segment (Temperley, 2001, p. 182). Third, he adapts the tonal hierarchies for use with TPC rather than NPC input, assigning the lowest values to the flattened seventh and all TPCs more than 5 steps away from the tonic or dominant on the line of fifths (Temperley, 2001, pp. 183–187). As discussed above, using TPC information to determine key seems to be the reverse of what happens in perception—one generally needs to take key into consideration in order to determine whether or not a note-spelling is correct.

Temperley tested two versions of the algorithm, one taking TPC information as input as described above, the other using just NPC information. He also compared the performance of his algorithm with that of algorithms described by Longuet-Higgins and Steedman (1971), Vos and van Geenen (1996) and Holtzman (1977) (Temperley, 2001, pp. 191–198). Temperley found that his algorithm was not as successful at finding the main keys of the fugue subjects in Bach's *Das Wohltemperirte Klavier* as the algorithms of Holtzman and Longuet-Higgins and Steedman. However, as Temperley (2001, p. 193) points out, these algorithms are rather limited, being incapable of dealing with modulations and polyphonic music. Interestingly, he found that the NPC version of his algorithm performed nearly as well as the version that was provided with TPC information (Temperley, 2001, p. 198).

Revision, ambiguity and expectation

Temperley argues that the preference rule approach can be used to model ambiguity as well as certain 'diachronic' aspects of musical listening. He points out that, because they use dynamic programming, each of his programs analyses the input in a 'left-to-right' manner, keeping track of a set of 'best-so-far' analyses as it proceeds. Typically, at each step in this process one of these 'best-so-far' analyses scores highest, in which case, Temperley proposes, the degree of ambiguity perceived at that step will be low. If, however, there are two or more highest-scoring analyses at a given step, then the perceived ambiguity will be greater (Temperley, 2001, p. 219). Also, the highest-scoring analysis for a step might not be contained within the highest-scoring analysis for some later step and this, Temperley suggests, models the phenomenon of 'revision', where a listener reinterprets an event in the light of a later one (Temperley, 2001, pp. 206–7). Temperley, (2001, p. 232) also proposes that 'the more expected events' are 'those permitting a more "high-scoring" analysis'.

Temperley provides various examples of musical expectation (Temperley, 2001, pp. 231–235), ambiguity (Temperley, 2001, pp. 219–230) and revision (Temperley, 2001, p. 214) and discusses how his models might account for them. In particular, he gives examples of 'tonal' revision in which the key of a segment is reinterpreted in the light of later evidence. He provides two examples of output from a version of his key program adapted to give a 'running analysis' of a passage. The second of these examples is an analysis of 16 bars from a Chopin Mazurka (Op. 67, No. 2) (Temperley, 2001, p. 217). The tonality of this passage oscillates between G minor and B flat major. However, although the program correctly identifies the modulations, it predicts that each one causes the listener to reinterpret everything that has gone before as being in the new key. I find this particular example unconvincing. However, his use of preference rules to account for expectation, ambiguity and revision is, in general, elegant and compelling.

Application of the theory to rock music and African music

Temperley shows that, with small modifications, some of his models can be applied to styles other than common practice music. Although he does not implement these modified rule systems as computer programs, he provides in-depth discussions of how his metre, harmony and key models can be applied to rock (Temperley, 2001, Chapter 9), and how his metre and phrasing models can be applied to African music (Temperley, 2001, Chapter 10).

Generally, authoritative scores exist neither for rock songs nor African music. This makes it more difficult to evaluate theories of structure in these musical styles (Temperley, 2001, p. 238).

Temperley's model of metric structure states that event-onsets and stressed syllables of text should preferably be aligned with strong beats (Temperley, 2001, pp. 32, 51). However, in rock music, stressed syllables and melodic events very often occur on weak beats (Temperley, 2001, pp.240–242). Temperley (2001, p. 243) therefore proposes a 'Syncopation Shift Rule' which is applied to the 'surface' input representation in order to yield a 'deep representation' in which each syncopated melodic event has been shifted forward onto the next strong beat. This deep representation then forms the input to his metre model. It seems possible that this simple modification might significantly improve the performance of his metre model on rock music. However, he does not explain how one is supposed to decide *which* metrically weak melodic events are syncopations.

Temperley (2001, p. 255) also proposes that the performance of his harmony and key models on rock music could be significantly improved by using deep representations in which the syncopations have been removed. However, as Temperley points out, rock music uses modality in a way that common practice music does not. In particular, the Ionian, Dorian, Mixolydian and Aeolian modes are commonly used—sometimes in the same song (Temperley, 2001, p. 259). For a given tonic, these modes overlap each other on the line of fifths so that if they are all combined, one gets what Temperley calls a 'supermode', spanning 10 positions on the line of fifths (Temperley, 2001, p. 260). Temperley proposes that this 'supermode' be used to construct a new 'tonal hierarchy' adapted for key-finding in rock music (Temperley, 2001, p. 262).

Temperley argues that the same fundamental cognitive structures underlie rhythm perception in both African and Western music (Temperley, 2001, p. 289) and claims that, with only minor modifications, his metre and grouping models could account for African rhythm. For example, to Africans, the 'regularity rule' (MPR 3, Temperley, 2001, p. 35) is more important and the 'accent rules' (MPRs 1, 2, 7 and 8, Temperley, 2001, pp. 32, 51) are less important than they are to Westerners. A Western listener thus tends to shift the metrical structure to get a good match with phenomenal accents, whereas an African listener maintains a strictly regular tactus even if it entails more syncopation (Temperley, 2001, p. 278). An interesting difference between grouping in African and Western music seems to be that Africans prefer to align strong beats with the *ends* of phrases (Temperley, 2001, p. 288), whereas, in common practice music, strong beats tend to occur near the beginnings of phrases (MPR 4, Temperley, 2001, p. 38).

Temperley emphasizes that most of the evidence for rhythm perception in African music comes from observations and transcriptions made by Western ethnomusicologists (Temperley, 2001, p. 266), whose use of bar-lines and slurs is often inconsistent (Temperley, 2001, pp. 270, 286). This makes it difficult to evaluate any theory of African rhythm.

Style, composition and performance

Temperley, (2001, Chapter 11) argues (convincingly) that the preference rule approach can be used to explain aspects of musical style, composition and performance. When one of his computational models is used to analyse a passage, it returns not only the 'most preferred' analysis but also a numerical score indicating how well this analysis satisfies the preference rules (Temperley, 2001, p. 293). If the score achieved by a passage is very low, this indicates that the passage cannot be interpreted in terms of the preference rule system. For example, Temperley's key-structure model would return a low-scoring analysis for a piece of 12-tone serial music (Temperley, 2001, p. 297). On the other hand, a passage that achieves a very high score when analysed by one of Temperley's models, would, in general, be unacceptably boring (Temperley, 2001, p. 298). Temperley proposes that between these extremes lies a 'range of acceptability' within which pieces are considered to be in the 'style' implicitly defined by the preference rule system (Temperley, 2001, p. 314). Passages achieving scores at the high end of this range will be perceived to be 'typical', while those that score at the lower end will be perceived to be more 'tense or daring'. In this way, preference rule systems can be used to characterize both musical styles and musical 'tension'.

Temperley also proposes that preference rules can be viewed as constraints on compositional practice, compositional choices being guided by the composer's goal to produce a piece that 'optimally' satisfies the preference rules governing the musical style in which he or she is working (Temperley, 2001, p. 298). As Temperley explains (Temperley, 2001, pp. 320-321), the preference rule approach also sheds light on aspects of performance. For example, one could hypothesize that performers play events that begin on strong beats longer and louder (Sloboda, 1983; Palmer and Kelly, 1992; Drake and Palmer, 1993) because, according to Temperley's MPRs 2 and 7, listeners prefer to hear stronger beats on longer, louder events. Also, according to his PSPR 1 (Temperley, 2001, p. 68), listeners prefer to place phrase boundaries between events separated by longer gaps. This suggests that performers slow down at the ends of phrases because this lengthens the gap in which the phrase boundary should be heard without disrupting the metric structure.

Motivic structure

In the final chapter, Temperley discusses how the low-level aspects of musical structure modelled by his systems affect higher-level structure. In particular, he considers motivic structure which he defines to be the 'network of segments in a piece heard as similar or related' (Temperley, 2001, p. 326). His ideas on this topic are based on those of Deutsch and Feroe (1981). Like them, he stresses the importance in motivic analysis of identifying appropriate 'alphabets' and claims that not just motivic analysis but harmonic and tonal analysis can be viewed as 'a search for a parsimonious encoding of musical input' (Temperley, 2001, p. 328). He supplements the ideas of Deutsch and Feroe, however, by highlighting the importance of metre in motivic structure. As he points out, two motivically related patterns are perceived to be related even when they are not played with metronomic precision, suggesting that the onset-times and durations of the events in the patterns are represented mentally as 'integer multiples of a low-level beat' (Temperley, 2001, p. 331). His discussion of motivic structure culminates with the following hypothesis:

> for two segments to be recognized as motivically related, they must be (1) related in pitch pattern (in the way specified by Deutsch [and Feroe]), (2) identical in rhythm (that is, they must have the same series of quantized inter-onset intervals), and (3) metrically parallel.
> (Temperley, 2001, p. 333)

Temperley points out that this hypothesis is only intended to apply to 'motivic relations that are recognized in a direct, automatic way' (Temperley, 2001, p 380). For example, it does not account for melodic diminutions (Forte and Gilbert, 1982, Chapter 1). However, anyone who has taught music analysis will vouch for the fact that some individuals automatically recognize relationships that are not at all obvious to others. Another weakness in Temperley's discussion is his failure to review the closely related work of Cambouropoulos and his co-workers (Cambouropoulos, 1998a,b; Cambouropoulos and Smaill, 1995; Cambouropoulos et al., 1999a,b,c, 2000).

Conclusions

This book is a worthy heir to *GTTM* that avoids some of the failings of its ancestor through computer implementation and objective, quantitative testing. For the most part, both the breadth and depth of coverage in the book is very impressive. Temperley convincingly argues that the preference rule approach can be used not only to explain aspects of musical listening, but also features of musical style perception, composition and performance. He also makes a praiseworthy effort to apply his theory to musical styles other than common practice music.

However, the book does suffer from certain weaknesses. First, there is some question as to whether the input representation Temperley employs is strictly appropriate for his purposes. Second, he fails to review some important and highly relevant recent work. Third, the notion that pitch spelling is used to determine harmony and key seems to be the reverse of what happens in perception. Fourth, his melodic phrase structure model is underdeveloped and *ad hoc*. Finally, he does not compare the performance of his models of metre, phrasing, counterpoint, pitch-spelling and harmony with that of other systems.

Nevertheless, I believe this book will (deservedly) become a standard text in computational music cognition and analysis.

References

Richard Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Cambouropoulos, E. (1996). A general pitch interval representation: Theory and applications. *Journal of New Music Research*, **25**, 231–251.

Cambouropoulos, E. (1998a). Musical parallelism and melodic segmentation. *Proceedings of the XII Colloquium of Musical Informatics, Gorizia, Italy*.

Cambouropoulos, E. (1998b). Towards a General Computational Theory of Musical Structure. PhD thesis, University of Edinburgh.

Cambouropoulos, E. (2000). From MIDI to Traditional Musical Notation. *Proceedings of the AAAI 2000 Workshop on Artificial Intelligence and Music, 17th National Conf. on Artificial Intelligence (AAAI'2000)*, Austin, TX.

Cambouropoulos, E. (2001a). Automatic Pitch Spelling: From Numbers to Sharps and Flats. *VIII Brazilian Symposium on Computer Music (SBC{\&}M 2001)*.

Cambouropoulos, E. (2001b). Melodic Cue Abstraction, Similarity, and Category Formation: A Formal Model. *Music Perception*, **18**(3), 347ff.

Cambouropoulos, E. and Smaill, A. (1995). A computational theory for the discovery of parallel melodic passages. *Proceedings of the {XI} Colloquio di Informatica Musicale*. Bologna, Italy.

Cambouropoulos, E., Crochemore, M., Iliopoulos, C. S., Mouchard, L. and Pinzon, Y. J. (1999a). Algorithms for computing approximate repetitions in musical sequences. *Proceedings of the 10th Australasian Workshop on Combinatorial Algorithms (AWOCA'99)*, pp. 129–144, Perth, WA, Australia.

Cambouropoulos, E., Smaill, A. and Widmer, G. (1999b). A clustering algorithm for melodic analysis. *DIDEROT FORUM on Mathematics and Music: Computational and Mathematical Methods in Music*, pp. 21–28, Vienna.

Cambouropoulos, E., Crawford T. and Iliopoulos, C. S. (1999c). Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects. *Proceedings of the AISB'99 Symposium on Musical Creativity*, pp. 42–47, Edinburgh.

Cambouropoulos, E., Crawford T. and Iliopoulos, C. S. (2000). Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects. *Computers and the Humanities*, **34**(4).

Cormen, T. H., Leiserson, C. E. and Rivest, R. L. (1990). *Introduction to Algorithms*. Cambridge, MA: MIT Press.

Deutsch, D. and Feroe, J. (1981). The Internal Representation of Pitch Sequences in Tonal Music. *Psychological Review*, **88**(6), 503—522.

Dixon, S. (1999). A beat tracking system for audio signals. Proceedings of the Diderot Forum on Mathematics and Music: Computational and Mathematical Methods in Music, December 2—4, pp. 101—110, Vienna.

Dixon, S. (2000). A Lightweight Multi-Agent Musical Beat Tracking System. *PRICAI 2000: Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, pp. 778–788, Melbourne, Australia.

Dixon, S. and Cambouropoulos, E. (2000). Beat Tracking with Musical Knowledge. *ECAI 2000: Proceedings of the 14th European Conference on Artificial Intelligence*, pp. 626—630.

Drake, C. and Palmer, C. (1993). Accent structures in music performance. *Music Perception*, **10**, 348–378.

Forte, A. and Gilbert, S. E. (1982). *Introduction to Schenkerian Analysis*. Norton: New York.

Holtzman, S. R. (1977). A program for key determination. *Interface*, **6**, 29–56.

Kostka, S. and Payne, D. (1995). *Workbook for Tonal Harmony*. New York: McGraw-Hill.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. (Oxford Psychology Series, Vol. 17). New York and Oxford: Oxford University Press.

Lerdahl, F. and Jackendoff R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Longuet-Higgins, H. C. and Steedman, M. J. (1971). On interpreting Bach. In B. Meltzer and D. Michie (eds.), *Machine Intelligence* (Vol. 6) (pp. 221–241). Edinburgh, University of Edinburgh Press.

Meredith, D., Wiggins, G. A. and Lemström, K. (2001). Pattern induction and matching in polyphonic music and other multidimensional datasets. *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI2001), July 22-25*, Vol. X, pp. 61—66. Orlando, FL.

Meredith, D., Lemström, K. and Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, to appear.

Ortmann, O. (1926). On the melodic relativity of tones. *Psychological Monographs* **35**(Whole No. 162).

Palmer, C. and Kelly, M. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, **31**, 525–42.

Piston, W. (1978). *Harmony*. London: Gollancz.

Rolland, P.-Y. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, **28**(4), 334–350.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, **103**(1), 588–601.

Scheirer, E. D. (2000). Music Listening Systems. PhD thesis, MIT.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, **35**, 377–96.

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

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.

Vos, P. G. and van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception*, **14**, 185–224.

Temperley		*GTTM*
MWFR 1	=	MWFR 2
MWFR 2	=	MWFR 3
MPR 1	=	MPR 3
MPR 2	=	MPR 5a
MPR 3	≈	MWFR 4
MPR 4	≈	MPR 2
MPR 5	=	MPR 10
MPR 6	≈	MPR 5f
MPR 7	=	MPR 4
MPR 8		–
MPR 9	≈	MPR 1

Table 1. Correspondence between rules in Temperley's theory of metrical structure and those in the metrical structure component of *GTTM* (see Lerdahl and Jackendoff, 1983, pp. 345—347 and Temperley, 2001, pp. 357—358). The symbol '=' indicates that the two rules are the same; the symbol '≈' indicates that Temperley's rule is a modification of the rule in *GTTM*; the symbol '—' indicates that there is no equivalent of Temperley's rule in *GTTM*.

Key	*Score*	*Key*	*Score*
C major	0.274	C minor	-0.013
C sharp major	-0.559	C sharp minor	-0.332
D major	0.543	D minor	0.149
E flat major	-0.130	E flat minor	-0.398
E major	-0.001	E minor	0.447
F major	0.003	F minor	-0.431
F sharp major	-0.381	F sharp minor	0.012
G major	0.777	G minor	0.443
A flat major	-0.487	A flat minor	-0.106
A major	0.177	A minor	0.251
B flat major	-0.146	B flat minor	-0.513
B major	-0.069	B minor	0.491

Table 2. Correction of Temperley's Table 7.1 (Temperley, 2001, p. 175) showing key-profile scores for opening bar of 'Yankee-Doodle' (see Figure 6).

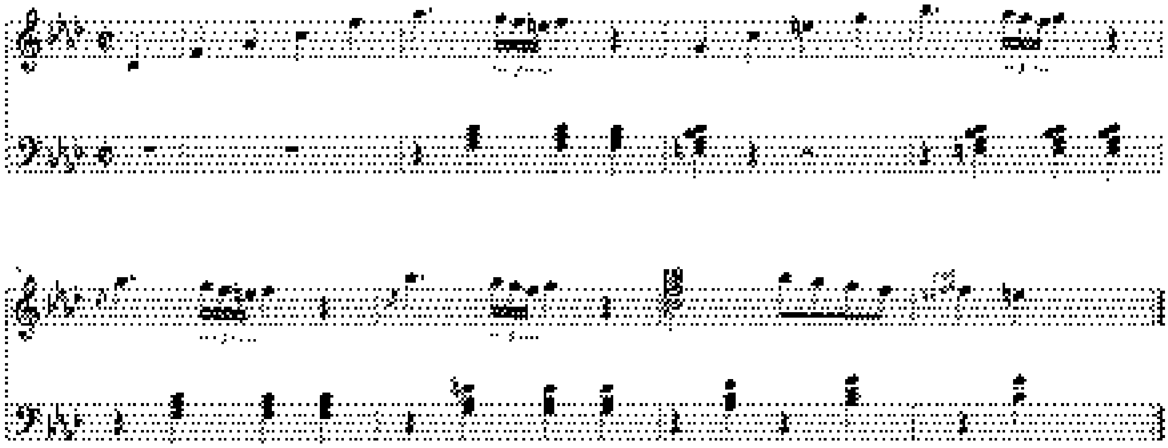

Figure 1. Opening bars of Beethoven's Piano Sonata in F minor, Op.2, No.1, first movement.

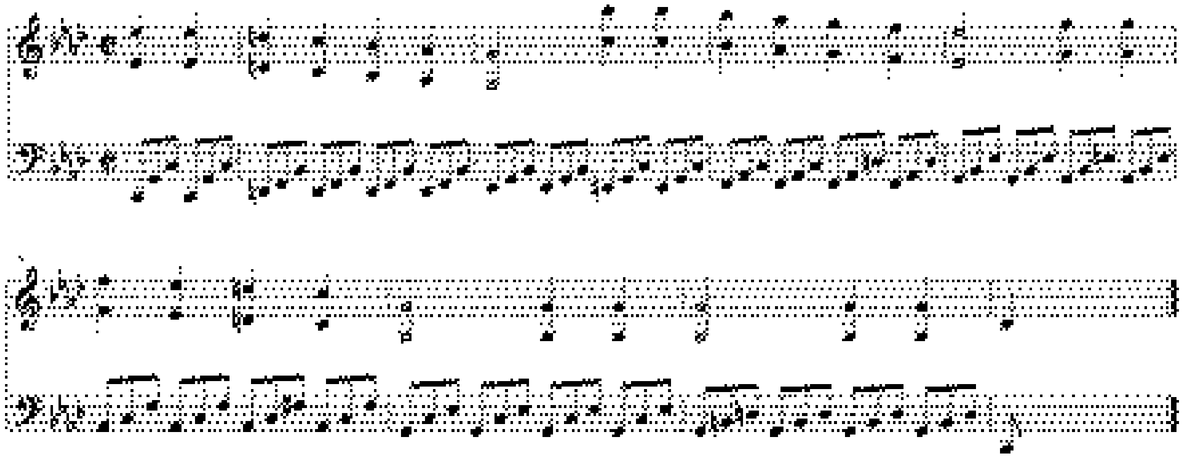

Figure 2. Bars 36–42 of the fourth movement of Beethoven’s Piano Sonata in F minor, Op.2, No.1.

Figure 3. The spelling of the upper note in the final chord cannot be determined. (From Temperley, 2001, p. 121.)

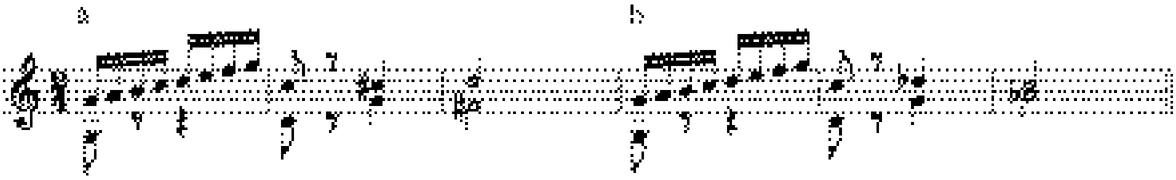

Figure 4. The spelling of the upper note in the last chord in the second bar depends on what follows.

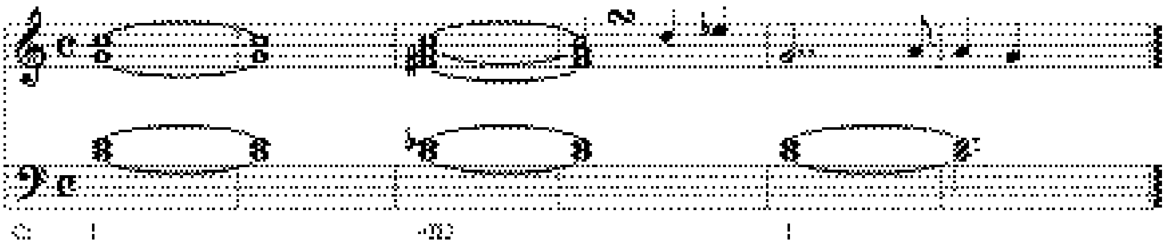

Figure 5. Example of ambiguous pitch spelling. (From Piston, 1978, p. 390.)

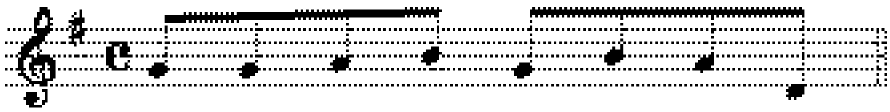

Figure 6. The opening of ‘Yankee-Doodle’.