Modeling Pattern Importance in Chopin's Mazurkas

Tom Collins, Robin Laney, Alistair Willis, &
Paul H. Garthwaite
The Open University, Milton Keynes, United Kingdom

This study relates various quantifiable characteristics of a musical pattern to subjective assessments of a pattern's salience. Via score analysis and listening, twelve music undergraduates examined excerpts taken from Chopin's mazurkas. They were instructed to rate already discovered patterns, giving high ratings to patterns that they thought were noticeable and/or important. Each undergraduate rated thirty specified patterns and ninety patterns were examined in total. Twenty-nine quantifiable attributes (some novel but most proposed previously) were determined for each pattern, such as the number of notes a pattern contained. A model useful for relating participants' ratings to the attributes was determined using variable selection and cross-validation. Individual participants were much poorer than the model at predicting the consensus ratings of other participants. While the favored model contains only three variables, many variables were identified as having some predictive value if considered in isolation. Implications for music psychology, analysis, and information retrieval are discussed.

Received February 12, 2010, accepted September 8, 2010.

Key words: pattern discovery, rating, music analysis, linear regression model, translation

It seems uncontroversial to suggest that *repetition* plays a central role in our perception of musical structure. In Schenker's words (1906/1973): "Only by repetition can a series of tones be characterized as something definite. Only repetition can demarcate a series of tones and its purpose. Repetition thus is the basis of music as an art" (p. 5).

Schenker's discussion presents several kinds of repetition. One of his examples, shown in Figure 1, exemplifies what we will call *exact* repetition. Also cited is the excerpt shown in Figure 2. We prefer to call the latter *repetition with interpolation*, since there are differences between original statement and repetition: mainly the *interpolated* notes F5, E5, D5, C5, and B4 in the second half of bar 3. The excerpt in Figure 3 is another example of *repetition with interpolation*, but this time there is a larger number of interpolated notes. For the sake of clarity, those notes in the original statement that are repeated have black noteheads, as do the repeated notes themselves. There will be further cause to consider how the amount of interpolation affects perception of musical structure. In Figure 3, the durations of some of the notes differ between original statement and repetition. For instance, the C3 eighth-note in bar 1 becomes a C3 quarter-note in bar 25. This relaxation of the requirement for exact repetition of duration seems to be in keeping with Schenker's examples.

An aspect of repetition overlooked by the analysis in Figure 3 is that some of the notes among the original statement have their onsets shifted relative to others in the repetition. For instance, the downbeat notes of bars 2 and 4 in the first violin are shifted from the downbeats of bars 26 and 28 due to the triplet variation technique, whereas the corresponding notes in the cello remain on the downbeat. Repetitions involving shifted notes are excluded from our study—a shortcoming no doubt, but one that is avoided largely in what follows by choosing a corpus of music in which occurrences of such nonrigid repetition are relatively rare.

Schenker (1906/1973) observes that "not only the melody but the other elements of music as well (e.g., rhythm, harmony, etc.) may contribute to the associative effect of more or less exact repetition" (p. 7). In this spirit, we propose three further types of repetition:

Transposed real. The excerpt shown in Figure 4 is an example of transposed repetition, sometimes referred to as a *real sequence*. If each note in the original statement is transposed up three semitones, this gives the notes contained in the repetition. The exactness of the semitonic transposition is what defines a real sequence.

Transposed tonal. Tonal sequences, on the other hand, are usually defined as adjusted real sequences, where adjustments are made in order to remain in key. For instance, the first pair of brackets in Figure 5 indicates a tonal sequence. Most notes in the original statement—such as A5 in the first violin—are transposed down three semitones to give notes contained in the repetition, but

Music Perception volume 28, issue 4, pp. 387–414, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2011.28.4.387

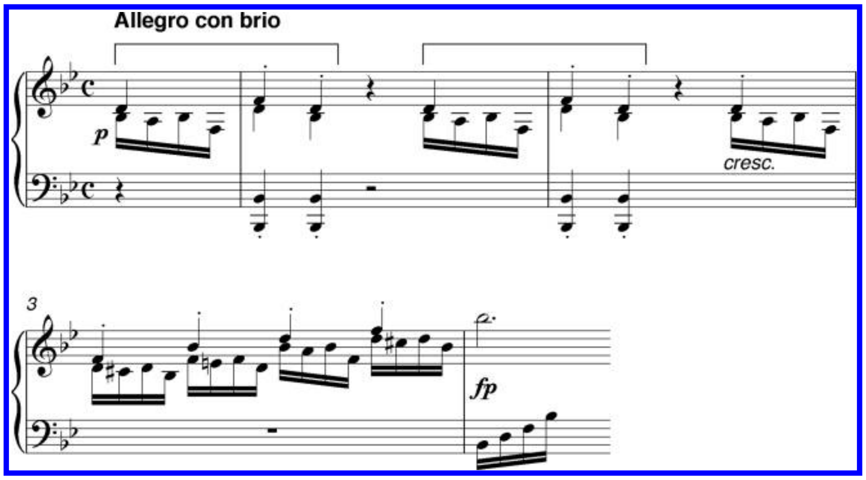

FIGURE 1. Bars 1–4 from the first movement of the Piano Sonata No. 11 in B♭ major Op. 22 by Ludwig van Beethoven (1770–1827). The brackets are Schenker's (1906/1973, p. 6).

FIGURE 2. Bars 1–4 from the first movement of the Piano Sonata No. 8 in A minor K. 310 by Wolfgang Amadeus Mozart (1756–1791). The brackets are Schenker's (1906/1973, p. 5).

FIGURE 3. Bars 1–4 and 25–28 from the fourth movement of the Octet in F major D. 803 by Franz Schubert (1797–1828). The brackets indicate an instance of *repetition with interpolation*. For the sake of clarity, those notes in the original statement that are repeated have black noteheads, as do the repeated notes themselves.

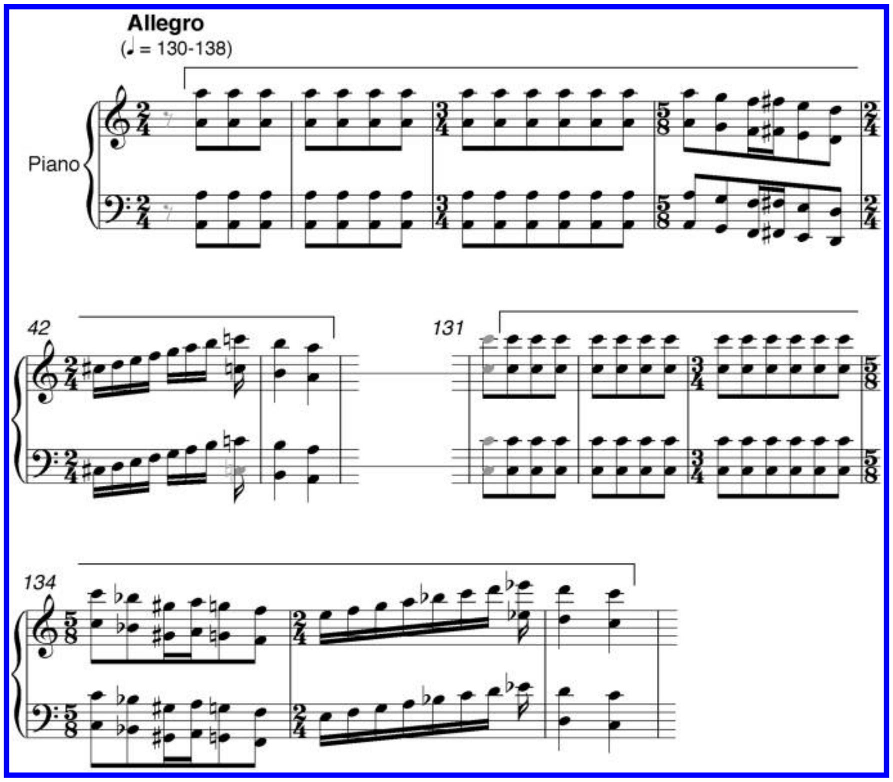

FIGURE 4. Bars 38–43 and 131–136 from the first movement of the Piano Concerto No. 1 by Béla Bartók (1881–1945). The brackets indicate an instance of a real sequence. Black noteheads help to show which notes are involved. © Copyright 1927 by Boosey & Hawkes Inc. Copyright Renewed. Reprinted by permission.

FIGURE 5. Bars 18–24 from the Allemande of the Chamber Sonata in B minor Op. 2, No.8 by Arcangelo Corelli (1653–1713). The first pair of brackets indicates a tonal sequence. The next three brackets indicate another. Black noteheads help to show which notes are involved, and numbers below the continuo part are figured bass notation.

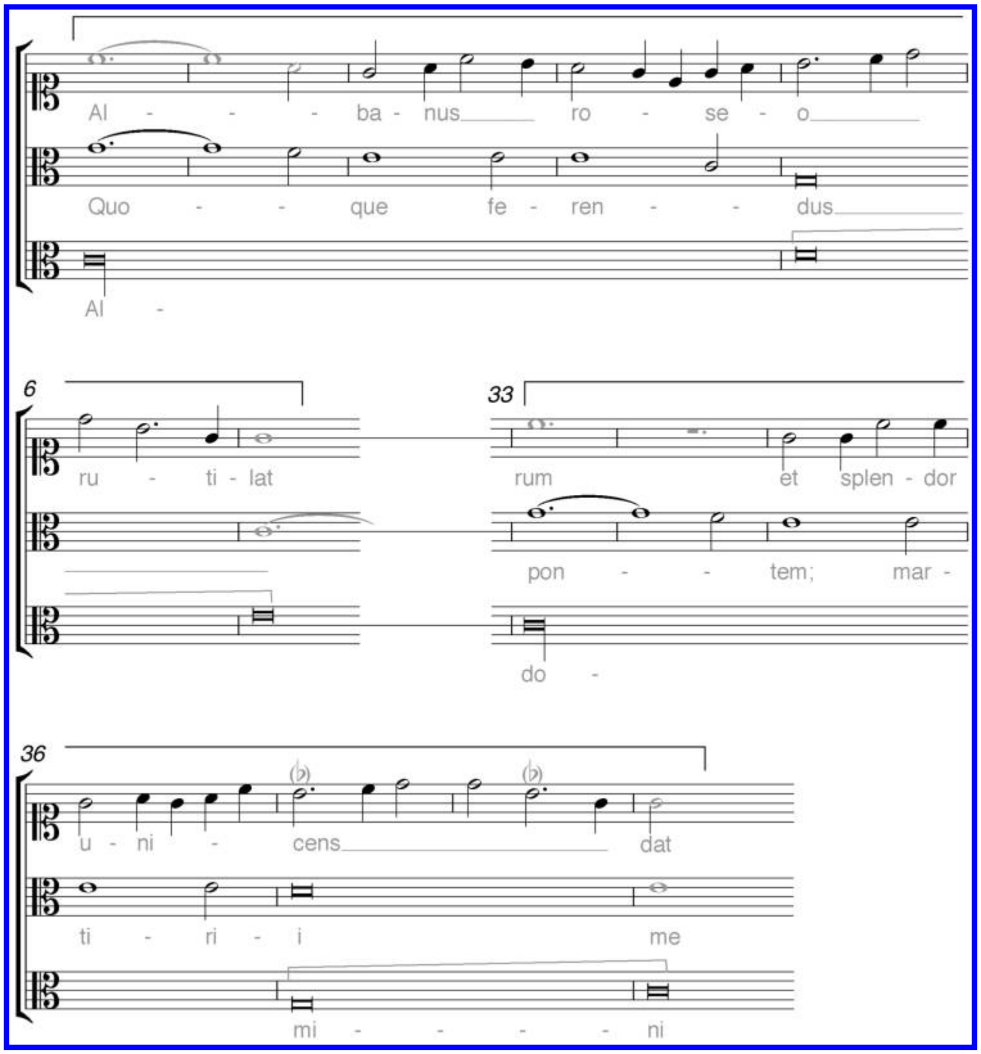

FIGURE 6. Excerpts from 'Albanus roseo rutilat' by John Dunstaple (*c* 1390–1453). The brackets indicate an instance of durational repetition. For the sake of clarity, those notes in the original statement whose durations are repeated have black noteheads, as do the repeated notes themselves.

some—such as the F♯5 in the second violin—must be transposed down four semitones to remain in key.

Durational. This type of repetition in a score of a piece can be obscured in performance: a staccato quarter note might be perceived as an eighth note, and notes are often sustained to thicken the texture. However, we argue that trying to discover durational repetition is still worthwhile. It can underpin the *composition* of entire pieces, even if it is lost in *performance*. For instance, the excerpt in Figure 6 constitutes durational repetition. It is taken from an isorhythmic motet, a defining feature of which is "the periodic repetition or recurrence of rhythmic configurations, often with changing melodic content" (Bent, 2001, p. 618). Durational repetition is also used in later musical periods, often in much more obvious ways.

The class of *repetition* could be broadened further. However, Schenker's own examples of repetition involving "other elements of music" arguably stretch the term *repetition* too far, to the point where *imitation* might be more appropriate. For example, in Figure 7 a motif spanning an octave (G3 to G4) is bracketed. The next bracketed occurrence spans a compound minor third (G3 to B♭4), so arguably is more accurately described as an *imitation* rather than a *repetition* of the original statement. It certainly does not correspond to exact repetition, repetition with interpolation, real or tonal transposition. The bracketed motif *does* correspond to durational repetition—three consecutive quarter notes—but the lack of further bracketed instances suggests it was not this common durational pattern but a particular pitch profile that Schenker had in mind. That said, we concur with

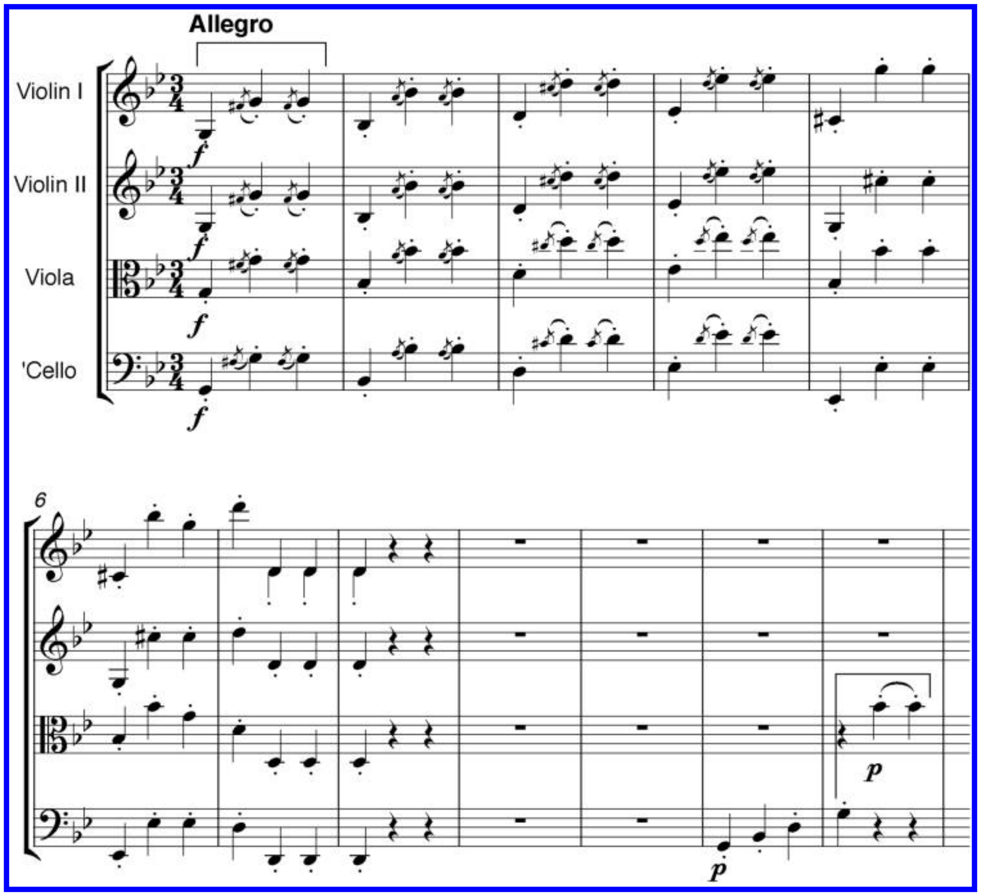

FIGURE 7. Bars 1–12 from the first movement of the String Quartet in G minor, 'The Horseman,' Op. 74, No. 3 by Joseph Haydn (1732–1809). The brackets are Schenker's (1906/1973, p. 8).

Schenker's point of view that repetitive/imitative material can move voices (in this case from the cello to the viola). It could be inferred from recent work on separating musical textures into perceptually valid voices and streams (Cambouropoulos, 2008) that repetitive material ought to remain within the same voice, but to stipulate as much at this stage seems peremptory.

For the purposes of this paper, the five types of repetition outlined above (exact, with interpolation, transposed real, transposed tonal, and durational) are labelled the *proto-analytical class* of repetition types. They can be thought of as the basic constituents of a proper analytical method, but an analytical method consisting of these repetition types alone is plainly insufficient—hence *proto*. Our proto-analytical class is oblivious to some basic concepts; for instance scale, triad, and octave equivalence, not to mention concepts that comprise more sophisticated analytical methods such as Schenkerian theory (Forte & Gilbert, 1982) or Ockelford's (2005) zygonic theory. Even in terms of handling repetition, there are occasions where our proto-analytical class is not wholly adequate, such as in Figure 3. A more positive characteristic of the class is its adherence to the principle of "repetition as creator of form" (Schenker, 1906/1973, p. 9), meaning that it does not distinguish between small- and large-scale repetitions. There is no need to agonize over definitions of motif, theme, section, and then go shoehorning repeated material into one category or the other. Rather, the proto-analytical class can be used to identify various instances of repetition in a piece, and then the analyst can categorize these instances according to small- and large-scale considerations if desired.

The Aim Of Our Experiment

While it is legitimate to distinguish as Cross (1998) does between *music analysis* (a largely conscious and voluntary process undertaken by experts) and what we might call *ordinary listening* (mostly unconscious and involuntary—what tends to be studied in music psychology), both analysis and ordinary listening involve the discovery of patterns, and there is clearly some common ground between them.

Our aim is to explore how this pattern discovery process works. We asked analysis students to rate already-discovered patterns, according to which patterns they would give priority to mentioning in an analysis essay. Attributes of these patterns and the excerpts in which they occur were quantified, and inferences made of the form

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p, \quad (1)$$

where y is the rating given to a pattern, $x_1, x_2, \ldots, x_p$ are attributes of the pattern and excerpt, and $\alpha, \beta_1, \beta_2, \ldots, \beta_p$ are coefficients to be estimated from the data. We are not suggesting for a moment that music analysts operate according to a formula such as (1), or that music analysis would benefit from doing so. Rather, we are claiming that some aspect of the pattern discovery process could be modeled by a weighted sum of pattern attributes. It is hoped that testing this claim will shed some light on how both ordinary listening and expert analysis work, and might therefore be of interest to music psychologists.

A second field where this work can have an impact is music information retrieval (MIR). Indeed, many of the pattern attributes $x_1, x_2, \ldots, x_p$ from (1) that are considered below come from this domain. In MIR there is an unsolved problem of how to *order* (and even *discard* some of) the output of a pattern discovery system. The flow chart in Figure 8 depicts a framework for a pattern *discovery* system: algorithms that have been or could be cast within this framework are proposed by Meredith, Lemström, and Wiggins (2002), Forth and Wiggins (2009), Conklin and Bergeron (2008), Cambouropoulos (2006), and Lartillot (2004). The flow chart in Figure 9 shows a framework for the more common task of pattern *matching*: algorithms cast within this framework abound, as robust pattern matching systems are something of a holy grail in MIR (an overview is given by Downie, 2003, and a specific example is found in Doraisamy & Rüger, 2003). The two flow charts should be contrasted because in pattern matching, there is an obvious way of ordering the output matches: *rate them by relevance or proximity to the original query*, using an appropriate relevance metric. With pattern discovery, it seems less obvious how the analogous step should work. Suppose that an algorithm has discovered hundreds of patterns within a piece of music. Now these must be presented to the user, but in what order? Unlike with pattern matching, there is no original query to compare with discovered patterns.[1]

[1]While we have contrasted pattern discovery with pattern matching, cross-pollination between the two can still occur. For example, Meredith, Lemström, and Wiggins (2003) use the same underlying concepts to formulate a pattern discovery algorithm called SIATEC and a pattern matching algorithm called SIAMESE.

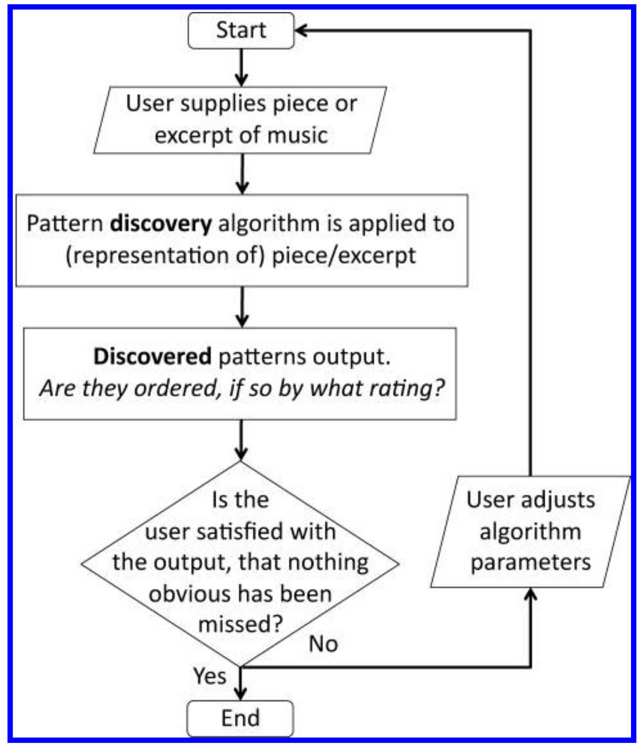

FIGURE 8. Flow chart depicting a framework for a pattern discovery system.

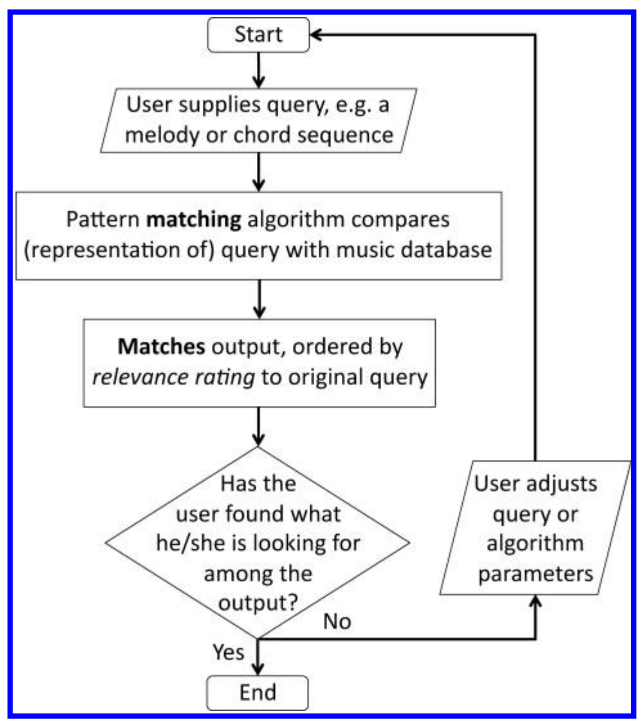

FIGURE 9. Flow chart depicting a framework for a pattern matching system.

Researchers have addressed this unsolved problem by defining various concepts and formulae. Some of these will be presented in the Method section, some are deferred to Appendix A, and we introduce a few now. To our knowledge, none of these formulae were derived empirically, and only two (Eerola & North, 2000) have been validated empirically. Hence, statistically derived models of the form (1) would constitute a methodological improvement.

Meredith, Lemström, and Wiggins (2003) define the concepts of *coverage*, *compactness*, and *compression ratio* and combine them in a multiplicative formula. Forth and Wiggins (2009) also combine them multiplicatively. It is claimed that these measures help to identify "perceptually salient patterns" (Forth & Wiggins, 2009, p. 1) that would be "considered musically interesting by an analyst or expert listener" (Meredith et al., 2003, p. 7). Conklin and Bergeron (2008) put forward a formula for the *interest* of a discovered pattern, and Conklin and Anagnostopoulou (2001) define something similar called a pattern's *score*. Both of these formulae are based on the concept of the number of times one *expects* to hear/see a given pattern in a piece or excerpt. There is an analogy to be made here with bioinformatics, in terms of the expected number of occurrences of a subsequence in a DNA string (Ewens & Grant, 2001). There is also the related concept of a *distinctive* pattern (Conklin, 2008; Huron, 2001a). Cambouropoulos (2006) defines a formula for the *prominence* of a discovered pattern: "The patterns that score highest should be the most significant" (p. 254). Only fifteen patterns are discovered in the example provided by Lartillot (2004), so it hardly seems necessary to rate them. Consequently no formula is suggested, which is a shame since this is the only research that claims explicitly to be founded on "modeling of listening strategies" (Lartillot, 2004, p. 53).

In summary, we focus on five kinds of repetition that we label collectively as the proto-analytical class. In our experiment, analysis students were asked to rate already discovered patterns, according to which patterns they would give priority to mentioning in a music analysis essay. The model in (1) gives the general form of inference to be drawn from these ratings. The primary contribution of our experiment is that it tests the conjecture that some aspect of the pattern discovery process can be modeled by a weighted sum of pattern attributes. As such, it should shed some light on how both ordinary listening and expert analysis work, and therefore be of interest to music psychologists. The work is also relevant to MIR and the unsolved problem of how to arrange the output of a pattern discovery system.

Method

Participants and Instructions

Music undergraduates (7 males and 5 females) from the University of Cambridge were paid £10 for an hour for their time, during which they were asked to rate already discovered patterns.[2] Participants were returning for their second or third year (mean age = 20.83 years, *SD* = 0.95) and had attended music analytical lectures and written music analytical essays as part of their studies. The instructions began by alluding to these essays, and the preparatory work of identifying recurring patterns—the restatement of material, the appearance of themes, motifs, gestures. Participants received the following instructions for the main task:

In the following exercise such recurring patterns *have been* identified and will be presented for you to *rate* according to how noticeable and/or important you think they are.

- High ratings should be given to the most noticeable and/or important patterns. Even if they might be 'obvious', these are the kind of patterns that deserve at least a mention in a standard analytical essay.
- Low ratings should be given to patterns that are difficult to see or hear and are of little musical importance. One would struggle to justify mentioning them in an essay.
- Middling ratings apply to any other patterns—quite important but not that noticeable, or vice versa. Something will be lacking in such patterns that prevents them receiving the highest ratings, yet they are more readily perceived than low-rating patterns.

As there is considerable variety in the terminology used to qualify ratings, participants were invited to rate patterns according to what *they* would mention in—or omit from—an analysis essay. The term *noticeable and/or important* covers as much of the terminological variety as possible, but arguably it is not as meaningful to participants as the reference to writing an analysis essay.

Participants were asked to rate patterns on a scale of 1 to 10 (least to most noticeable and/or important), giving their ratings to one decimal place. The decimal place was helpful for distinguishing between two patterns if both received the same integer rating initially. The instructions also indicated that a noticeable pattern was not

[2] See http://users.mct.open.ac.uk/tec69/supporting-material.html for a copy of the instructions for participants. Other supporting material is available at this address, including the design matrix, box and whisker plots of explanatory variables, and all patterns chosen for the study.

FIGURE 10. Bars 1–20 from the Mazurka in G♯ minor Op. 33, No. 1 by Frédéric Chopin (1810–1849). Occurrences of pattern *A* are indicated by black noteheads. Dynamic and other expressive markings have been removed from this and subsequent figures to aid clarity.

necessarily an important pattern and vice versa. A darker font for pattern noteheads than for nonpattern material was used to identify the patterns to participants, as in Figure 10. Participants had access to a digital piano and a recording (Biret, 1990) of each excerpt throughout. This arrangement was intended to be *typical of the environment* in which an undergraduate begins analyzing a piece of music. Participants were able to ask questions of clarification at any point, they were able to revise ratings, and were assured that they were not taking a test. They were encouraged to form responses on the basis of their *musicality* and not by concocting some formula.

A balanced incomplete block design was used ($v = 9$, $b = 12$, $r = 4$, $k = 3$, $\lambda = 1$). This means that $v = 9$ excerpts of music were prepared, and a different combination of $k = 3$ excerpts was given to each of the $b = 12$ participants, such that each excerpt appeared in exactly $r = 4$ combinations, and each pair of excerpts appeared in exactly $\lambda = 1$ of the combinations (Mathon & Rosa, 1996). Ten patterns per excerpt were presented, so that each participant had $3 \times 10 = 30$ patterns to rate in total. The order of presentation of excerpts and the order of patterns within excerpts were randomized to allow for any ordering effects. Immediately prior to this task, each participant completed the same short warm up task, rating five already discovered patterns. The warm up task was intended to help participants to familiarize themselves with the format of presentation, answer sheet, and the rating scale. It also gave them an opportunity to ask questions. The format had been tested in a pilot study and adjusted accordingly for ease of understanding and use.

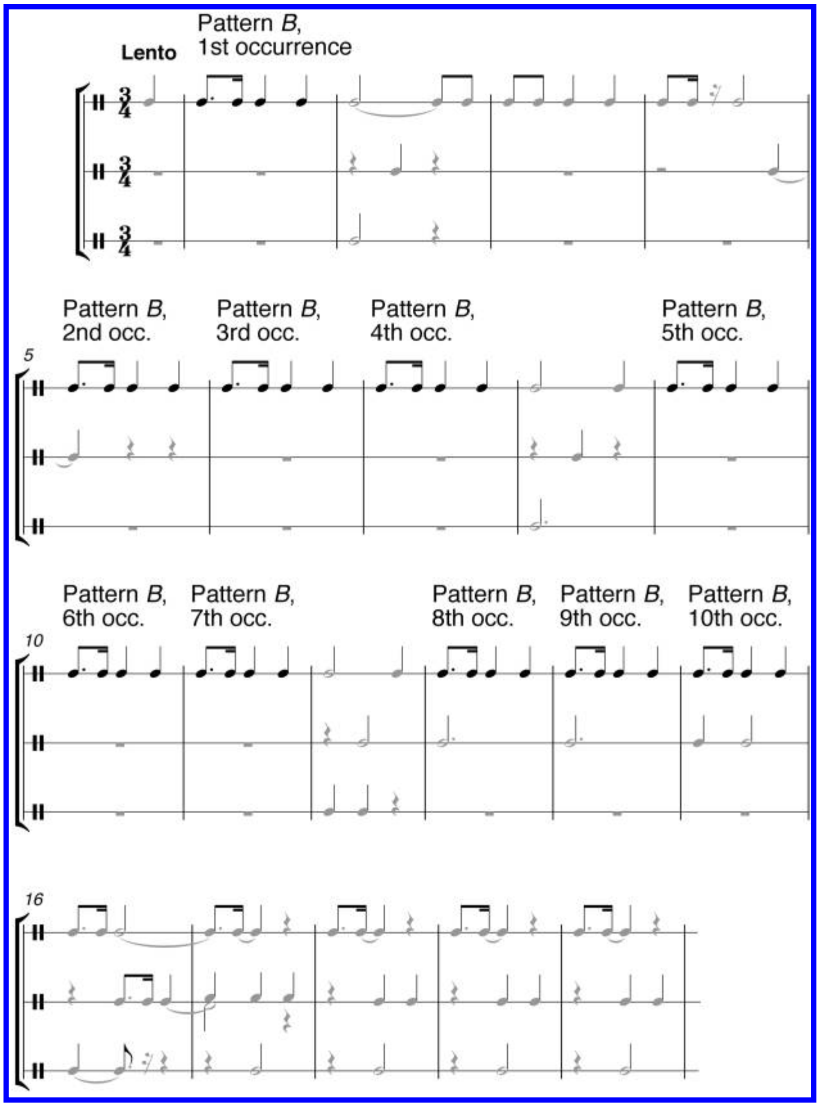

FIGURE 11. A rhythmic representation of bars 1–20 from the Mazurka in G♯ minor Op. 33, No. 1 by Chopin. Occurrences of pattern *B* are indicated by black noteheads.

Selection of Excerpts and Patterns

In any study such as this, the selection of stimuli influences the results. Our excerpts were selected from Paderewski's (1953) edition of mazurkas by Chopin, using a different mazurka for each excerpt.[3] The mazurka is a "Polish folk dance from the Mazovia region [where Chopin spent his childhood]. . . . In his [fifty plus] examples the dance became a highly stylized piece for the fashionable salon of the 19th century" (Downes, 2001, p. 189). With an eye on appropriate material for the participants in our study, music from the Western classical tradition was chosen, though students may well not have met the music before. One of the selected mazurkas (Op. 7, No. 5) was short enough to be presented in its entirety, but for other mazurkas a substantial section was chosen, not always from the beginning. Relatively speaking, Chopin's mazurkas are texturally and stylistically homogeneous, but still rich enough to contain examples of all the types of repetition from our proto-analytical class.

Approximately half of the discovered patterns were selected by the first author, such as patterns *A* and *B* in Figures 10 and 11. The remaining patterns were chosen

[3] Op. 7, No. 5 bars 1–20; Op. 24, No. 1 bars 17–32; Op. 24, No. 3 bars 1–24; Op. 30, No. 1 bars 17–36; Op. 33, No. 1 bars 1–20; Op. 33, No. 4 bars 1–24; Op. 50, No. 1 bars 25–48; Op. 56, No. 2 bars 45–68; Op. 67, No. 3 bars 1–16.

FIGURE 12. Bars 1–20 from the Mazurka in G♯ Minor Op. 33, No. 1 by Chopin. The first occurrence of pattern *C* contains three notes, as indicated by the arrows and black noteheads.

randomly from the output of Meredith et al.'s (2002) structural inference algorithm for translational equivalence classes (SIATEC) when applied to each excerpt, such as pattern *C* in Figure 12. Participants were not told of the composer or the source of the discoveries. This method of selection (half handpicked and half chosen at random from a large set) was used because we wanted to elicit a full range of judgements, whereas an entirely handpicked set of stimuli might all be relatively noticeable and/or important. On the other hand, Cook (1987) claims that he "can 'hear' the most preposterous analytical relationships if [he] choose[s] to" (p. 57). We felt that the inclusion of some preposterous patterns—for example, pattern *C*—was necessary to see what kind of ratings they received from participants. When handpicking five of the ten patterns for an excerpt, we tried first to select noticeable/important motifs such as pattern *A*. Second, we tried to select longer sections that support Schenker's notion of repetition as "creator of form," such as (approximately) bars 5–8 in Figure 10, repeating at bars 9–12. Third, an attempt was made to represent each type of repetition from our proto-analytical class. Finally, on occasion a pattern was chosen (nonrandomly) from the SIATEC output for its resemblance to a handpicked pattern. Thus, we tried to avoid participants realizing that half of the patterns were handpicked and half discovered algorithmically.

SIATEC was used in preference to other algorithms because the patterns that it returns are consistent with the proto-analytical class. Furthermore, while some of its results "correspond to the patterns involved in perceptually significant repetitions" (Meredith et al., 2002, p. 331), the sheer number of output patterns per excerpt means that at least some fall under the heading of Cook's (1987) "preposterous analytical relationships." Alternative pattern discovery algorithms that might have been used were mentioned in relation to Figure 8. These, as well as other candidates (Chiu, Shan, Huang & Li, 2009; Knopke & Jürgensen, 2009; Meek & Birmingham, 2003) are either not as consistent with the proto-analytical class as SIATEC, or make musical assumptions that do not apply to the textures of Chopin's mazurkas.

The following is an outline of the mechanics of SIATEC; fuller descriptions are given in Meredith et al. (2002) and Collins, Thurlow, Laney, Willis, and Garthwaite (2010). First, each note in an excerpt of music is converted to a point in multidimensional space, as demonstrated in Figure 13A and Figure 13B. Meredith et al. (2002) call the set of points a *dataset* and employ two representations of pitch (*MIDI note number* corresponding to real transposition, and *morphetic pitch number* corresponding to tonal transposition) as well as ontime and duration. Therefore the dimension of the space is at least two (ontime and one of the pitch representations or duration) and at most four. Second, so-called *translational patterns* that exist in the dataset are calculated, subject to a maximality condition that curbs the computational complexity. Third, for each calculated translational pattern, the excerpt (set of points) is checked for extra occurrences. For instance in Figure 13C, the set P might have been discovered due to its translational relationship with Q, so this third step accounts for any further translations of P that occur, namely R.

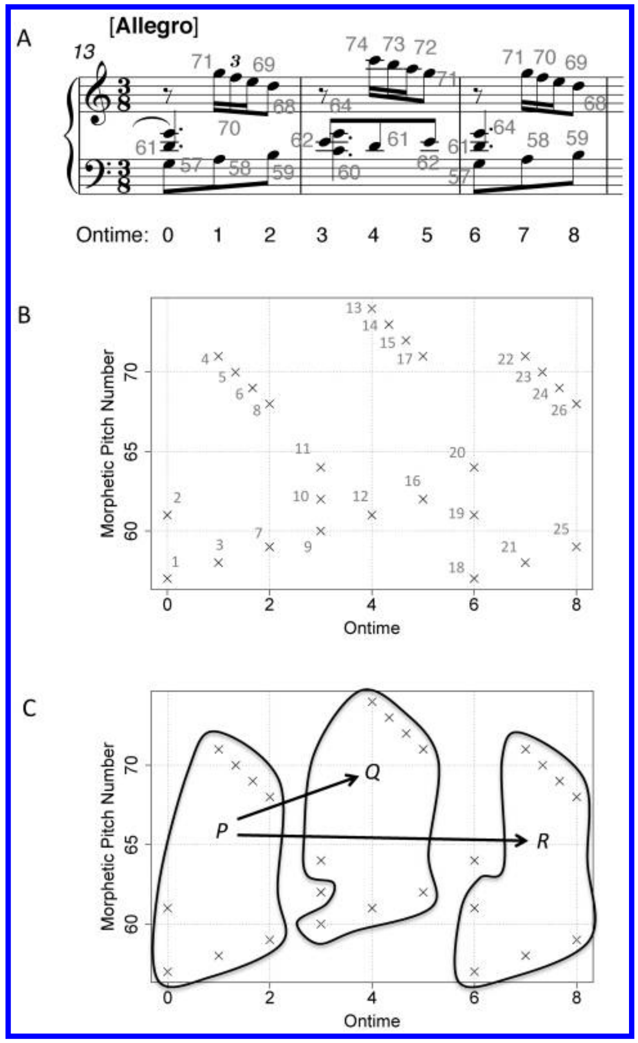

FIGURE 13. (A) Bars 13–15 from the Sonata in C major L. 3 by Domenico Scarlatti (1685–1757), annotated with morphetic pitch numbers and ontimes; (B) each note from the excerpt is converted to a pair consisting of an ontime and a morphetic pitch number. Morphetic pitch number is plotted against ontime, and points are labelled in lexicographical order $\mathbf{d}_1$ to $\mathbf{d}_{26}$; (C) the same plot as above, with three ringed patterns, *P*, *Q*, and *R*. Arrows indicate that both *Q* and *R* are translations of *P*.

Explanatory Variables

We consider *linear regression models* for rating discovered patterns in music, as in (1). The ratings given to patterns form the response variable: the explanatory variables quantify attributes of a pattern and the excerpt in which it appears. Other common methods, such as principal component analysis or a support vector machine, do not address our specific suggestion that a formula such as (1) could be involved at some stage of the pattern discovery process. It should be recalled that *linear* means *linear in the coefficients*. That is, linear models can contain explanatory variables that are quite complex, nonlinear functions of simpler variables, and this is true of some of the pattern attributes considered below. Forward selection, backward elimination, and cross-validation were used to select which explanatory variables should be used in the regression models. Eighteen of the twenty-nine explanatory variables included in the regression are formulae from existing work, and eleven variables are our suggestions. Occasionally an existing formula had to be adapted, if originally it was defined only for melodic material. Below is a list introducing the explanatory variables that emerged as being of most importance in this study. More details of these variables can be found in the Appendix, along with definitions of the remaining explanatory variables. The models that were fitted also included factors for participants and excerpts, to allow for fixed differences between participants in their ratings.

Cardinality is the number of notes contained in one occurrence of a pattern.

Occurrences refers to the number of times that a pattern occurs in an excerpt.

Coverage: The *coverage* of a pattern in a dataset is "the number of datapoints in the dataset that are members of occurrences of the pattern" (Meredith et al., 2003, p. 7). Recall that a *dataset* is the set of all datapoints representing an excerpt of music. If no occurrences of a pattern overlap (in the sense of sharing notes) then the *coverage* of a pattern is the product of its *cardinality* and *occurrences*.

Compactness: Meredith et al. (2003) define the *compactness* of a pattern in a dataset to be "the ratio of the number of points in the pattern to the total number of points in the dataset that occur within the region spanned by the pattern within a particular representation" (p. 8). There are several plausible definitions of region. We employ two and use whichever results in the maximum compactness.

Compression ratio is equal to *coverage* divided by the sum of *cardinality* and the number of nonzero translators (*occurrences* minus 1). It is the amount of compression that "can be achieved by representing the set of points covered by all occurrences of the pattern by specifying simply one occurrence of the pattern and all the vectors by which the pattern can be translated" (Meredith et al., 2003, p. 8).

Expected occurrences: Conklin and Bergeron (2008) give a formula for calculating the expected number of occurrences of a pattern in a dataset. The intuition is that patterns less likely to arise by chance—because they involve less common pitches or rhythms—should be more noticeable. The calculation of *expected occurrences* involves the *empirical distribution* (the relative frequency of occurrence of pitches and/or other musical events in an excerpt), and requires adapting from *melodic* material with no overlapping patterns allowed, to *textures* where overlapping patterns and patterns *with interpolation* are allowed. Models based on relative frequency of occurrence are liable to criticism for being oversimple,[4] but we prefer to include a variable *expected occurrences* in the regression, and then assess its credentials. From the empirical distribution, it is possible to calculate the likelihood of the event that a given pattern occurs. Multiplying this likelihood by the number of places in which the pattern can occur gives the *expected occurrences*.

[4]For instance, in relation to one of his models, Temperley (2007) observes that such a "proposal may seem wholly implausible as a model of the compositional process. But it is not intended as a model of the compositional process, only as a model of how listeners might represent the compositional process for . . . [a particular] purpose" (p. 83).

Interest: The interest of a pattern in a dataset is defined to be "the ratio of observed to expected counts," the rationale being that "large differences between observed and expected counts indicate potentially interesting patterns" (Conklin & Bergeron, 2008, p. 64).

Score: In earlier work, Conklin and Anagnostopoulou (2001) formulated essentially the same concept but in a different way, calling it the *score* of a pattern. This is the squared difference between observed and expected occurrences, divided by expected occurrences.

Rhythm only: If a pattern consists of rhythms only, then the variable *rhythm only* takes the value 1, and 0 otherwise. The intuition is that rhythm-only patterns are less noticeable than patterns that involve pitch.

Transposed repetition: The repetition of a pattern may be at the same pitch as the first occurrence, or transposed. The variable *transposed repetitions* counts the number of transposed repetitions of a pattern. Patterns with a high number of transposed repetitions could highlight real or tonal sequences, and these are likely to be noticeable/important.

Results

Model Selection

The explanatory variables to include in the regression were chosen by forward selection and also by backward elimination. A .05 significance level was used as the cutoff criteria for entering/removing variables. Forward selection begins with a model consisting of the participant variables (denoted by $\text{par}_2, \text{par}_3, \ldots, \text{par}_{12}$). These are protected from removal during model selection, as a blocking factor should generally be retained in models. The first step is to include each of the pattern attributes in this model, individually, and determine which of these attributes most reduces the residual sum of squares (RSS). The results of these individual fittings are shown in Table 1. It can be seen that *compactness* most reduces the RSS, as its value for r^2 is greatest. The coefficient for *compactness* is significant at the .05 level, so now we are considering a model that consists of the participant variables and *compactness*. The second step is similar to the first: take the new model, and include each of the remaining pattern attributes individually. To determine which attribute most reduces the RSS, a table similar in *format* to Table 1 can be constructed. The *order* of

TABLE 1. Individual fittings for the first step of forward selection.

Variable	Coefficient	*SE*	*t* value	*p* value	r^2
Compactness	6.12	0.26	23.87	$< 2.0 \times 10^{-16}$	.63
Rhythmic density	1.68	0.08	21.61	$< 2.0 \times 10^{-16}$	.58
Expected occurrences	–0.07	3.7×10^{-3}	–19.73	$< 2.0 \times 10^{-16}$	.54
Coverage	0.04	2.6×10^{-3}	14.14	$< 2.0 \times 10^{-16}$	.38
ThreeCs	9.03	0.71	12.80	$< 2.0 \times 10^{-16}$	.33
Rhythmic variability	6.56	0.56	11.72	$< 2.0 \times 10^{-16}$	.30
Signed pitch range	0.12	0.01	11.18	$< 2.0 \times 10^{-16}$	.28
Prominence	7.3×10^{-3}	6.5×10^{-4}	11.16	$< 2.0 \times 10^{-16}$	.28
Cadential	3.50	0.32	10.89	$< 2.0 \times 10^{-16}$	.27
Cardinality	0.06	6.3×10^{-3}	9.96	$< 2.0 \times 10^{-16}$	.24
Compression ratio	1.68	0.18	9.10	$< 2.0 \times 10^{-16}$	.21
Alt. prominence	0.07	7.6×10^{-3}	8.61	2.7×10^{-16}	.19
Phrasal	2.17	0.26	8.20	4.6×10^{-15}	.18
Small intervals	0.18	0.02	8.15	6.6×10^{-15}	.18
Max. pitch center	–0.17	0.02	–6.94	1.9×10^{-11}	.14
Score	–0.02	3.5×10^{-3}	–6.48	3.1×10^{-10}	.13
M.C. card. × occ.	–0.02	2.6×10^{-3}	–6.31	8.3×10^{-10}	.12
Chromatic	0.39	0.06	6.07	3.3×10^{-9}	.11
Metric syncopation	2.59	0.51	5.13	4.9×10^{-7}	.09
Transposed repetition	–0.29	0.07	–4.21	3.3×10^{-5}	.07
Unsigned pitch range	0.07	0.02	3.35	9.0×10^{-4}	.05
Interest	0.02	7.6×10^{-3}	2.85	4.7×10^{-3}	.04
Intervallic leaps	0.09	0.04	2.15	.03	.04
Tempo fluctuation	0.31	0.28	1.09	.28	.02
Occurrences	–0.05	0.05	–0.90	.37	.02
Rhythm only	0.29	0.44	0.66	.51	.02
Unsigned dyn. level	0.02	0.06	0.43	.67	.02
Signed dyn. level	0.03	0.07	0.42	.68	.02
Geom. mean likelihd.	0.23	1.50	0.16	.88	.02

Note: Each row in this table represents an individual fitting. For example the first row contains the results of fitting a model including the participant (block) variables *and compactness*. The standard error (*SE*) relates to the width of the confidence interval about the coefficient estimate. The ratio of explained sum of squares to total sum of squares is given by r^2, also known as the coefficient of determination.

attributes in that table may be completely different, however, due to the effect of including *compactness*. For instance, there is no guarantee that *rhythmic density* will most reduce the RSS—in fact, *expected occurrences* is the next attribute to be appended. Variables continue to be appended in this fashion while the corresponding coefficients are significant at the .05 level. The resulting *forward model* is

$$\begin{aligned} \text{rating} = 4.79 &+ 0.01 \cdot \text{par}_2 - 1.19 \cdot \text{par}_3 - 0.96 \cdot \text{par}_4 \\ &- 0.60 \cdot \text{par}_5 - 1.18 \cdot \text{par}_6 - 0.90 \cdot \text{par}_7 - 0.27 \cdot \text{par}_8 \\ &- 0.70 \cdot \text{par}_9 - 0.62 \cdot \text{par}_{10} - 0.49 \cdot \text{par}_{11} \\ &+ 0.73 \cdot \text{par}_{12} + 3.42 \cdot \text{compactness} \\ &- 0.04 \cdot \text{expected\_occurrences} \\ &+ 0.65 \cdot \text{compression\_ratio}, \end{aligned} \tag{2}$$

with test statistic $F(14, 345) = 59.12$, $p < 2.2 \times 10^{-16}$, and $s = 1.67$ as the error standard deviation.

Backward elimination works in an analogous fashion. It begins with a full model, consisting of variables for participants, excerpts, and pattern attributes. At each step the variable whose exclusion least increases the RSS is removed. Variables continue to be removed in this way while the corresponding coefficients are not significant at the .05 level. The *backward model* that resulted was

$$\begin{aligned} \text{rating} = 3.88 &- 0.13 \cdot \text{par}_2 - 1.28 \cdot \text{par}_3 - 0.88 \cdot \text{par}_4 \\ &- 0.74 \cdot \text{par}_5 - 1.34 \cdot \text{par}_6 - 0.97 \cdot \text{par}_7 - 0.35 \cdot \text{par}_8 \\ &- 0.78 \cdot \text{par}_9 - 0.70 \cdot \text{par}_{10} - 0.59 \cdot \text{par}_{11} + 0.68 \cdot \text{par}_{12} \\ &+ 0.07 \cdot \text{cardinality} + 0.89 \cdot \text{occurrences} \\ &- 0.04 \cdot \text{coverage} + 3.39 \cdot \text{compactness} \\ &+ 1.48 \cdot \text{compression\_ratio} \\ &- 0.53 \cdot \text{expected\_occurrences} - 0.99 \cdot \text{interest} \\ &+ 0.50 \cdot \text{score} + 0.94 \cdot \text{rhythm\_only} \\ &+ 0.15 \cdot \text{transposed\_repetitions}, \end{aligned} \tag{3}$$

with test statistic $F(21, 338) = 42.01$, $p < 2.2 \times 10^{-16}$, and $s = 1.64$ as the error standard deviation. For the forward model, $r^2 = .71$, meaning that this model explains 71% of the variation in the ratings. For the backward model, $r^2 = .72$. Hence, both models explain a substantial proportion of the variation in ratings and the difference in the amount they explain is minimal. The forward model is more parsimonious than the backward model—it is built on just three explanatory variables (apart from the between-participant factor) while the backward model uses ten.

The sign of some of the coefficients in the backward model (3) is concerning. For example, *coverage* and *interest* have negative coefficients but, by definition, these variables would be expected to contribute positively towards a pattern being rated as noticeable/important. In defense of the backward model, it should be recalled that some variables are constituents of other variables. For example, *occurrences* is a constituent of *coverage* and *interest*. Hence it is an oversimplification to say that the backward model contains counterintuitive coefficients without examining the overall contribution of variables such as *occurrences*.

Partitioning the design matrix according to the nine mazurka excerpts, we performed nine-fold cross-validation for the forward and backward models, comparing their mean prediction for each pattern rating with the observed mean rating. That is, we kept one mazurka to one side, estimated regression parameters with data from the other eight mazurkas, and used the resulting regression models to predict mean ratings for patterns in the mazurka kept to one side. This process was repeated for each mazurka in turn. The result is that, on average, the forward model's mean predictions are much closer to the observed mean ratings ($MSE = 0.96$) than are the backward model's mean predictions ($MSE = 2.37$). Therefore the forward model outperforms the backward model and there is evidence that the forward model in (2) gives better predictions than the backward model in (3).[5]

For the forward model, in Figure 14 the mean predictions from the cross-validation are plotted against the observed mean ratings. Figure 15 is the analogous plot for the backward model. There are acceptable straight-line fits in each plot and, in the main, there is little to choose between the two models. However, it can be seen from Figure 15 that one of the backward model's mean predictions is particularly large (the point at approximately (9, 20) in the split plot). This poor prediction is the reason that the backward model was out-performed by the forward model in cross-validation, so this item was investigated further. Figure 16 is a plot of ratings for patterns 1–20

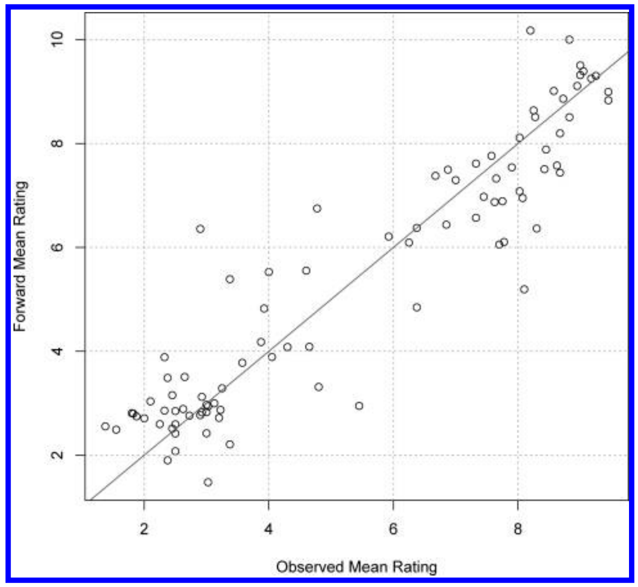

FIGURE 14. A plot of the forward model's mean prediction against the observed mean prediction for each of the ninety patterns.

(from the first two excerpts). For a given pattern, the four participant ratings are plotted as dots, joined by a line to give an indication of the range of the response. If fewer than four dots are visible, then this is due to coincident ratings. The observed mean rating—the mean of the four participant ratings—is plotted as a cross, the forward model's mean prediction is plotted as an asterisk, and the backward model's mean prediction as a diamond.

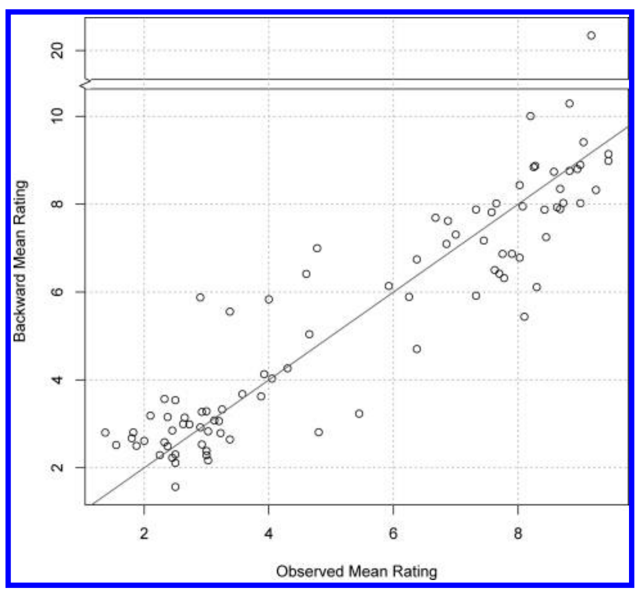

FIGURE 15. A plot of the backward model's mean prediction against the observed mean prediction for each of the ninety patterns.

[5]Plots made to check model assumptions and to check for outliers did not lead to any model revisions or any outlying data being removed.

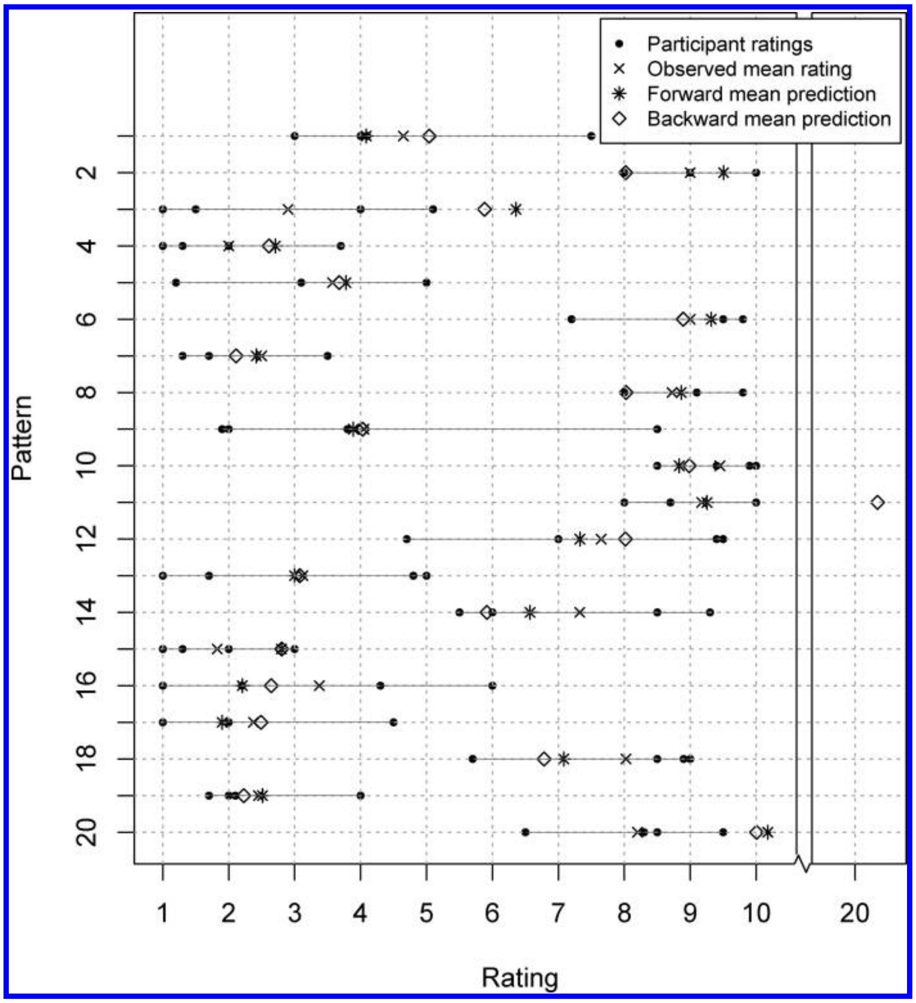

FIGURE 16. Observed and predicted ratings for patterns 1–20 (from the first two excerpts). If fewer than four dots (participant ratings) are visible per pattern then this is due to coincident ratings.

The backward model's poor prediction is for pattern eleven. This pattern has a higher *score* variable (= 396.01) than any other pattern and this is the cause of the large predicted rating. The forward model does not suffer from the same waywardness and, moreover, this will typically be the case—the forward model contains several fewer parameters than the backward model, making it more robust.

An aim of this study was to address an unsolved problem in MIR, by producing a formula for predicting ratings that could be applied to *unseen* excerpts/pieces of music. To this end, the forward model in (2) was adapted so as not to include $\text{par}_2, \text{par}_3, \ldots, \text{par}_{12}$, which relate to the individual participants and only apply in this study. Specifically the mean of the relevant coefficients, $(0.00 + 0.01 - 1.19 + \ldots + 0.73)/12 = -0.52$, was added to the constant term 4.79, changing it to 4.28. So our formula for rating the extent to which a discovered musical pattern is noticeable and/or important is

$$\begin{aligned} \text{rating} = 4.28 &+ 3.42 \cdot \text{compactness} \\ &- 0.04 \cdot \text{expected\_occurrences} \\ &+ 0.65 \cdot \text{compression\_ratio}. \end{aligned} \tag{4}$$

Predictive Value of Individual Variables

Several of the explanatory variables that are in neither our forward nor backward models have been proposed by others as useful for predicting the salience of a pattern. The fact that a variable was not in these models does not imply it has no predictive ability. It is just that correlations between the explanatory variables mean that adding more variables to a model does not significantly improve the predictions of ratings after the model already contains certain variables. Other researchers may wish to construct other models or devise other explanatory variables, so further information about the predictive ability of variables is important. Table 1 is useful in this respect, as it shows the results of individual fittings. The rows are ordered by r^2, the proportion of variability in the data that is explained by the model. The line just below *intervallic leaps* indicates a cut-off point in this table. Above this line, there is evidence to suggest that participants used a particular variable to form their ratings ($p < .05$). For all but six of the explanatory variables, there is evidence that the variable was useful for predicting participants' ratings. However, *maximum pitch center* has a negative coefficient, which is contrary to the intuition given in the Appendix, and the same is true of *score*, *transposed repetitions*, and the interaction *mean-centered cardinality* × *occurrences*. Hence a cut-off point between $r^2 = .18$ and $r^2 = .14$ may give a better distinction between the variables that seem useful for predicting salience and those that do not.

Predictive Ability of Participants Compared With the Formula

The present work has developed a model for evaluating the salience of a pattern. It accounted for just over 70% of the variation in participants' ratings, which looks useful, but begs the question of whether it is easy to predict their ratings. Is the model useful or could a person give ratings effortlessly and more effectively? One approach to examining this question is to determine if participants in the experiment could predict the ratings that other participants gave. Does the formula that we have proposed in (4) give predictions that are closer to the *consensus* than any one music undergraduate can get? For instance, the first participant rated patterns from three excerpts. Each pattern in each of these excerpts was also rated by three *other* participants. For each pattern we call the mean of these *other* ratings a *consensus*. Now, on average, are the first participant's ratings or the formula's predictions closer to this consensus? Accuracy was evaluated by calculating the mean squared error $\sum(\text{observed\_value} - \text{prediction})^2/N$, where N is the number of observations (we ignore patterns that the first participant did not examine). It turns out that the formula in (4) outperforms the first participant in terms of mean squared error (MSE). Analogous consensus tests for participants 2–12 found that the formula in (4) outperformed every participant. The MSE for the participants and the formula are given in Table 2.

The table shows that ratings from participant 9 were closest to the consensus, but even this participant had a substantially larger MSE than our formula. In predicting the consensus rating, the MSE of a participant was between 37% and 461% larger than the model and, on average, it was more than 200% larger. The extent to which the formula in (4) improved on participants' predictions surprised the authors. Judging by these results, it would be much better to use the formula to rate the salience of a pattern than to use ratings of any one of the participants.

TABLE 2. Consensus Test Results.

Participant	Participant MSE	Formula MSE
1	4.34	0.87
2	2.65	1.22
3	2.40	1.13
4	2.97	1.15
5	2.68	0.67
6	4.23	1.47
7	3.09	1.17
8	4.00	0.71
9	1.64	1.20
10	3.57	1.78
11	3.89	1.28
12	7.69	1.37
Mean	3.60	1.17

Note: Mean squared error (MSE) when one participant's ratings are used to estimate the consensus rating given by other participants, and when the formula in (4) is used to predict the consensus.

Discussion

Conclusions

The primary aim of the work reported here was to investigate the claim that a formula such as (1) could model some aspect of the pattern discovery process. The formula in (4) was derived empirically using a linear model (2) that emerged as the stronger performer on cross-validation, and accounted for just over 70% of the variability in the participants' responses.[6] We are cautious about drawing general conclusions from the results of one participant study, but can say that the above results do nothing to undermine the claim. A secondary aim of this paper was to address an unsolved problem in MIR, of arranging the output of a pattern discovery system. For this purpose also, the formula in (4) was derived for rating patterns as relatively noticeable and/or important

[6]Interested in estimating the maximum value of r^2 that could be achieved for this dataset, a model was fitted consisting of eleven participant indicator variables and 89 pattern indicator variables. For this model $r^2 = .80$, $s = 1.60$.

based on variables that quantify attributes of a pattern and the piece of music in which it appears. We hope that MIR researchers will find the formula useful, especially when arranging the output of their pattern discovery systems. Our review of existing work suggests that, up to this point, researchers have proposed formulae for rating discovered patterns with little foundation of empirical evidence. This paper seems to be the first to adopt an empirical method in the context of rating discovered patterns. The value of $r^2 = .71$ from the forward model in (2) is the proportion of variability in the ratings that is explained by the forward model, and it is greater (of course) than any of the r^2 values given in Table 1: hence, the empirical method leads to a formula that offers a better explanation of the ratings than any of the proposed formulae do individually.

The results of the consensus tests suggest that the formula in (4) can be used with confidence, as it is better at predicting the consensus rating than any of the human participants. There are further advantages to using the formula in (4). First, it can be used to filter or screen a large amount of data in a way that a human cannot. Second, the formula's rating for a certain pattern in a given piece is not subject to change, whereas a human may become tired or alter their preferences over time. We hope that this work will act as a springboard for other researchers wishing to build their own models. The model put forward in (2) is appealing due to its performance on cross-validation and also because of its parsimony, but there is the potential to test other models. In this respect, Table 1 gives some idea of which existing variables might lead to plausible alternative models.

In the introduction it was suggested that this paper would be of interest to those working in music psychology. Can any more general conclusions be drawn that are pertinent to this field? First, forward selection, backward elimination, and cross-validation could be valuable for testing hypotheses in other areas of music perception. Second, one can *imagine* situating each of the twenty-nine explanatory variables included in the regression on a line; its position on that line determined by how likely it is that an average music undergraduate is familiar with the variable's meaning. For example, consider *intervallic leaps* and *compression ratio.* Most music undergraduates would be able to furnish a definition of *intervallic leaps*, but even if the definition of *compression ratio* were given, few would acknowledge it as musically relevant. It is surprising and telling that the variables that appear in the formula in (4)—*compactness*, *expected occurrences*, and *compression ratio*—are *not* those that we associate as being particularly familiar to music undergraduates. Perhaps these variables do not have much currency in music psychology and music analysis because they are relatively recent, or because their definitions (intuitive or mathematical) are somewhat *unmusical*. However, we have found evidence for their perceptual validity—in that they have emerged as predictors for participant ratings—and therefore they deserve a more prominent place in music psychology and music analysis. In particular, it would be worth attempting to situate the concepts of *compactness*, *expected occurrences*, and *compression ratio* in relation to music Gestalt principles, such as in voice-leading (Huron, 2001b) and stream segregation (Bregman, 1990).

Now that a rating formula (4) has been proposed, it would perhaps be helpful to discuss some of its components. This is done with reference to Figures 17 and 18, and Table 3, which contains ratings and attributes for the patterns shown in these figures. Pattern *E* (Figure 17) is a real sequence with three occurrences. The same figure contains pattern *F*, a predominantly scalic motif that passes between right and left hands in an imitative

FIGURE 17. Bars 45–68 from the Mazurka in C major Op. 56, No. 2 by Chopin. Occurrences of patterns *E* and *F* are indicated by black noteheads. Boxes demarcate the first two occurrences of pattern *F*.

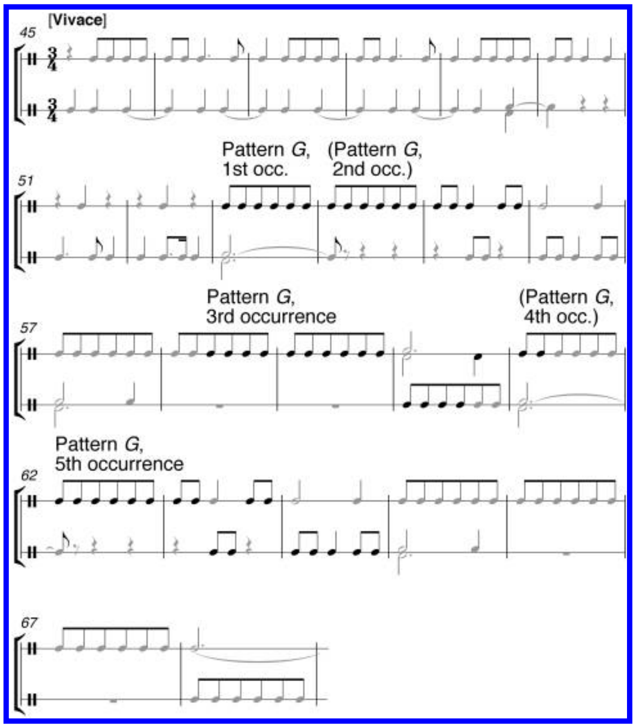

FIGURE 18. A rhythmic representation of bars 45–68 from the Mazurka in C major Op. 56, No. 2 by Chopin. Due to overlapping notes, two occurrences (second and fourth) of pattern *G* are not shown.

fashion, and has eight occurrences overall. Lastly, pattern *G* (Figure 18) consists of the durations of pattern *F*, with an extra first note. From Table 3 it can be seen that patterns *E* and *F* are similar in terms of *cardinality* (the number of notes contained in one occurrence), as well as in terms of the significant components *compactness* and *expected occurrences*. They differ in how they *account* for the excerpt. For instance, a listener that comprehends pattern *F* as consisting of 16 notes that repeat 8 – 1 = 7 times, is able to encode approximately sixteen bars of music using 16 + 7 = 23 pieces of information. On the other hand, a listener that does not comprehend pattern *F* just hears 16 × 8 = 128 notes. That is, they must try to encode the same sixteen bars of music using 128 pieces of information. The parsimony of comprehending pattern *F* is quantified by the *compression ratio* 128/23 ≈ 5.57. Pattern *F* accounts for approximately twice as many bars of music as does pattern *E*, which accordingly has a lower *compression ratio* of 45/(15 + 3 – 1) ≈ 2.65. So pattern *F* is rated higher than *E* (10.1 versus 7.9) by the formula in (4).

It has been suggested that listeners use parsimonious encodings where available to aid memorization of note sequences (Deutsch, 1980). Although we did not ask participants to memorize any passages, we too suggest that the availability of parsimonious encodings is intimately linked to the perception of musical structure. Comprehending a pattern and its occurrences may also confirm or undermine what the listener perceives as the established meter. For example, pattern *B* in Figure 11 lasts for three beats and often repeats immediately, confirming the prevailing triple meter of the mazurka. However, pattern *H* in Figure 19 lasts for two beats and repeats immediately twice. Hence, the listener might hear three bars in duple meter, rather than two bars in triple meter (as written), thus undermining the prevailing triple meter of the mazurka.

Like patterns *E* and *F*, patterns *F* and *G* (Figure 18) have similar *cardinalities*. Although pattern *G* fares worse than *F* in terms of the significant components *compactness* and *compression ratio*, the most striking difference is in their *expected occurrences*. One way of interpreting

TABLE 3. Ratings by Participants 3, 6, 7, and 11 of Patterns *E*, *F*, *G*, and *I*.

Ratings and attributes	Pattern *E*	Pattern *F*	Pattern *G*	Pattern *I*
Participant 3	8.0	10.0	6.0	9.6
Participant 6	8.5	8.5	5.0	9.5
Participant 7	10.0	9.6	1.0	7.5
Participant 11	8.0	7.2	4.0	6.5
Observed mean	8.6	8.8	4.0	8.3
Rating formula	7.9	10.1	5.8	8.9
Compactness	.94	1	.81	.99
Expected occurrences	33.96	32.95	83.33	0.72
Compression ratio	2.65	5.57	3.00	1.98
Cardinality	15	16	17	80
Occurrences	3	8	5	2
Coverage	45	128	63	160

Note: The observed means, ratings according to the formula in (4), and various attributes are also given.

FIGURE 19. Bars 40–52 from the Mazurka in C♯ minor Op. 41, No. 1 by Chopin. Occurrences of pattern *H* are indicated by black noteheads.

these values is to say that a pattern like *G*, consisting of fourteen consecutive eighth notes, a quarter note, and two more eighth notes, is not that unexpected in the context of a mazurka. In fact, it could be said that pattern *G* is 83.33/32.95 ≈ 2.5 times more likely to occur than pattern *F*, which is essentially the same pattern but with a *specific* pitch configuration. Recalling that *expected occurrences* in (4) has a *negative* coefficient, the high value for pattern *G* contributes –0.04 × 83.33 ≈ –3 to the rating, whereas the lower value for pattern *F* contributes only –0.04 × 32.95 ≈ –1. Hence, the *expected occurrences* component accounts for approximately half of the difference between the rating for pattern *G* of 5.8, and that of 10.1 for pattern *F*.

Pattern *I* (Figure 20 and Table 3) and pattern *J* (Figure 21) highlight two ways in which the formula in (4) might be improved. All but one of the participants, as well as the rating formula, agree that pattern *F* (formula rating 10.1) should be rated above pattern *I* (formula rating 8.9). Pattern *I* defines a small section, with an original statement at bars 53–60 being repeated immediately at bars 61–68. However, if instead the listener hears bars 53–68 as consisting of eight occurrences of pattern *F*, then the two occurrences of pattern *I* are more or less *implied*. Therefore, pattern *F rather than* pattern *I* would be mentioned in an analysis essay, so the rating of 8.9 for *I* is too high. Augmenting the rating formula in (4) to adjust for these kind of *implications* could lead to improved performance. For an instance of high participant ratings (9.1, 7.8, 8.3, and 8.0), being at odds with a lower formula rating (6.3), we turn to pattern *J* (Figure 21). First, the formula's performance could be improved if the concept of octave equivalence was incorporated in the representation: each occurrence of pattern *J* is followed immediately by the pitch classes G and B. Participants may have heard these other notes as part of the pattern and this could have inflated the ratings. Second, the formula's performance could be improved if the concept of harmonic function was incorporated—a more ambitious aim. One of the reasons why pattern *J* receives a relatively low formula rating is because there are three non-pattern notes among the first occurrence at bar 29, reducing the *compactness*. The chord on beat 3 of bar 29 is an augmented sixth chord (German or French depending on whether the E♭5 or D5 is counted), whereas the chord on beat 3 of bars 30–32 is a diminished seventh or dominant chord (again depending on whether the E♭5 or D5 is counted) above a G3 pedal. Hence, the different chords, augmented sixth versus dominant, make legitimate the omission of non-pattern notes in the left hand of bar 29. However, the *harmonic function* in each case is similar—moving towards G major (or G dominant seventh)—so in this sense the omitted notes *are* part of a more abstract pattern.

What does it mean if a variable appears below the line of statistical significance in Table 1? It *could* mean that

FIGURE 20. Bars 45–68 from the Mazurka in C major Op. 56, No. 2 by Chopin. Occurrences of pattern *I* are indicated by black noteheads.

the concept giving rise to this variable is music-perceptually and analytically obsolete. More likely, however, it means that we have failed to capture the concept adequately in the variable's definition. For instance the *signed dynamic level* of a pattern is calculated by summing over scores given to dynamic markings, but perhaps it would be better captured by analyzing the amplitude of waveform segments. Another possibility is that the variable *does* capture the concept, but that participants did not apply this concept in a *consistent manner* when forming ratings. Although the forward model in (2) does reveal an underlying consistent explanation ($r^2 = .71$), there is still considerable leeway. This leeway is perhaps where the music analyst makes his or her mark, by interpreting a piece of music in a novel way, yet within the realm of feasibility. To reiterate a point from the introduction, we are not suggesting that music analysts should use a formula, just that the process of rating musical patterns as more or less noticeable and/or important reflects the practice of deciding what to mention and what not to mention in a music analytical essay. A pedagogical outcome of this paper is that the

FIGURE 21. Bars 17–36 from the Mazurka in C minor Op. 30, No. 1 by Chopin. Occurrences of pattern *J* are indicated by black noteheads.

results could form part of a tool to help students with the discovery of patterns in music, so fostering the "desire to encounter a piece of music more closely, to submit to it at length, and to be deeply engaged by it, in the hope of thereby understanding more fully how it makes its effect" (Pople, 2004, p. 127).

Future Work

An outstanding issue to address is whether the formula in (4) can be applied to translational patterns in mazurkas other than those included in the participant study. Does the formula scale up to longer excerpts/entire pieces? And does it generalize to music by other composers, for different instrumental forces, from different periods, genres, etc? Answering the second question is beyond the scope of this paper, but a tentative answer to the scaling question follows. Box and whisker plots of the absolute errors between observed mean ratings and forward mean ratings against excerpt length are shown in Figure 22. Two of the nine excerpts are 16 bars long, three are 20 bars long, and four are 24 bars long. Any trends in these box and whisker plots—for instance if the median (thick black) line increased with excerpt length—might suggest that the formula in (4) does not scale up to longer excerpts. Looking at the plots, there is no evidence to suggest that the forward model suffers from scaling problems: neither the median nor the interquartile ranges appear to be a function of excerpt length; whereas there are outlying values for the 16- and 20-bar excerpts (points more than 1.5 times the interquartile range from the box), the errors for the 24-bar excerpts contain no outliers.

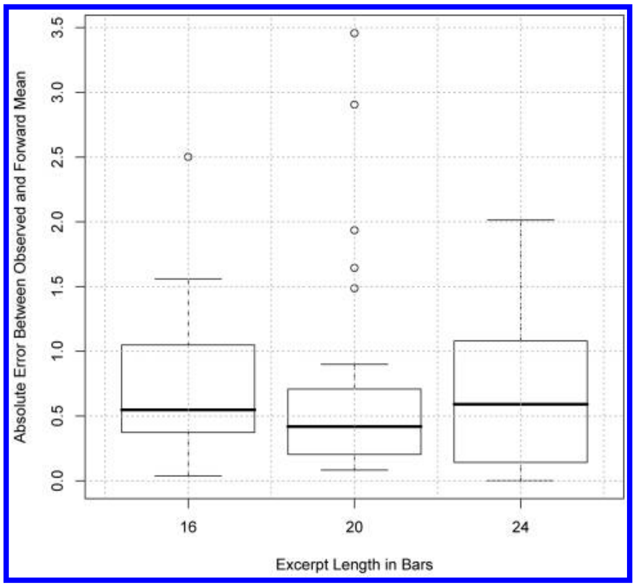

FIGURE 22. Box and whisker plots to explore the relationship between model performance and excerpt length. For each of the ninety patterns investigated, the absolute error between the observed mean rating and forward mean rating is calculated. This data is then split into three categories depending on the length of excerpt in which a pattern occurs.

There are several worthwhile directions in which this research could be taken. First, the participants in the study described above were twelve music undergraduates. But music listeners—expert and nonexpert alike—might be able to rate discovered patterns. With music undergraduates, it was possible to assume a substantial amount of knowledge and expertise. Music undergraduates at the University of Cambridge prepare for exams in which they analyze and compose whole pieces of music without recourse to recordings or a means of playing through passages. Therefore, it was not deemed necessary to isolate patterns aurally for participants. Further, it did not surprise or concern the authors that none of the participants made substantial use of the digital piano, and that two participants did not want to listen to the recordings of excerpts. It could be said that a particular performance of a piece can have an *undue* influence on the perception of musical structure. On the other hand, such an approach seems to neglect that music is heard primarily rather than seen, and this reliance on the score has been criticized before, and labelled *scriptism* (Cook, 1994). In short, with a greater number of participants and considerable amendments to the design, a similar trial could be conducted with nonexpert listeners. Second, a previous participant study (Tan, Spackman, & Peaslee, 2006) investigated how listeners' judgements of music were affected by repeated exposure, by conducting a trial with the *same* participants on two occasions, separated by two days. Neither repeated exposure nor time were considered as factors in our study, yet there is much anecdotal evidence to suggest that comprehension of a piece varies with exposure, and in particular that a listener discovers new patterns in a piece over time. This acknowledgement could be cause for concern, for how can the performance of a pattern discovery system be evaluated by comparison with a human benchmark performance, if this benchmark is not an absolute but instead depends on exposure or time? The definition of a human benchmark merits further attention, although different definitions may be appropriate to different situations.

Third, it is possible to argue that *different* occurrences of the *same* pattern ought to be rated individually. With reference to Figure 10, arguably the first occurrence of pattern *A* is more noticeable than the second occurrence. The first occurrence is at the excerpt's very beginning, isolated to some extent, whereas the second occurrence dovetails with preceding and proceeding phrases. In the study described above, participants were asked to give a pattern one *overall* rating, taking all occurrences of the pattern into account. Both the issues of ratings affecting one another and of pattern occurrences being rated individually merit further investigation. Finally, aspects of our analysis have focused on *mean* ratings. However, there was marked disagreement between participant ratings over some patterns. For example pattern 9 in Figure 16 received ratings with a standard deviation of 3.09, whereas the standard deviation of ratings for pattern 10, say, was only 0.99. Although the mean rating for pattern 9 is lower than that for pattern 10, some might argue that pattern 9 is the more important of the two: it polarized the participants for some reason. Identifying factors that cause participant polarization is another worthy topic for future work.

Author Note

We are grateful to David Temperley and three anonymous reviewers for very helpful comments on earlier versions of this manuscript. The authors also wish to thank Jeremy Thurlow and the Warden and Fellows of Robinson College, Cambridge, for hosting the participant study, as well as the music students who took part. We are grateful to Alec Frank-Gemmill and Rachael Ibbott for their advice on the pilot study. This paper also benefited from helpful discussions with David Meredith and Jamie Forth.

Correspondence concerning this article should be addressed to Tom Collins, Department of Computing, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK. E-MAIL: t.e.collins@open.ac.uk

Appendix

Definitions of the Explanatory Variables Included in the Regressions

An intuitive definition is given for each variable unless one is given in the paper. A mathematical definition is then given where appropriate. We adopt the notation from Meredith et al. (2002). That is, a *dataset D* (containing *datapoints* or just *points*) is a representation of an excerpt of music. A *translational pattern* P_1 is a subset of D, for which there exist $m - 1$ translations in D, written $P_2, P_3, \ldots, P_m$. Meredith et al. (2002) call $\{P_1, P_2, \ldots, P_m\}$ the *translational equivalence class* of P_1 in D, or $TEC(P_1, D)$ for short.

Cardinality is the number of notes contained in one occurrence of a pattern. We write $|P|$ for the cardinality of the pattern, or sometimes l.

Occurrences refers to the number of times that a pattern occurs in an excerpt. This will be denoted by $|TEC(P, D)|$ or simply m.

Mean-centered cardinality × occurrences: The intuition behind this variable is that a pattern containing many notes and/or occurring many times is likely to be relatively noticeable/important. It is common practice to use mean-centered values when forming such interactions, so if P^* is the mean of the cardinalities of all discovered patterns and m^* is the mean of the occurrences of all discovered patterns, then the *mean-centered cardinality × occurrences* for a pattern P in a dataset D is given by

$$\text{mc\_cardinality} \times \text{occurrences}(P, D) = (P - P^*)(m - m^*). \quad (5)$$

Coverage: The *coverage* of a pattern in a dataset is "the number of datapoints in the dataset that are members of occurrences of the pattern" (Meredith et al., 2003, p. 7).

$$\text{coverage}(P, D) = |\cup_{Q \in TEC(P, D)} Q|. \quad (6)$$

The variables *coverage* and *mean-centered cardinality × occurrences* are likely to be highly positively correlated as, before mean-centering, the latter is an upper bound for the former.

Compactness: For a pattern $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_l\}$ in a dataset $D = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n\}$, one definition of *compactness* is

$$\text{compactness}(P, D) = |P| \,/\, |\{\mathbf{d}_i \in D : \mathbf{p}_1 \leq \mathbf{d}_i \leq \mathbf{p}_l\}|, \quad (7)$$

where the lexicographic ordering '$\leq$' is defined by Meredith et al. (2002, p. 331). A second definition is used for patterns that appear in one stave only. If the pattern P occurs in the top stave only, say, then D in (7) should be replaced by $D^{\uparrow}$, the set of all datapoints occurring in the top stave.

Compression ratio: The mathematical definition is

$$\text{compression\_ratio}(P, D) = \text{coverage}(P, D) \,/\, (|P| + |TEC(P, D)| - 1). \quad (8)$$

ThreeCs: The variables *coverage*, *compactness* and *compression ratio* are combined by Meredith et al. (2003) so as to order discovered patterns, but the nature of this combination is not made explicit. The combination takes the form

$$\text{coverage}^a \cdot \text{compactness}^b \cdot \text{compression\_ratio}^c, \quad (9)$$

where a, b, and c are parameters to be chosen by the user (D. Meredith, personal communication, August 12, 2009). Stipulating these parameters a priori is not always appropriate, so Forth and Wiggins (2009) offer some useful alternatives. Forth and Wiggins (2009), having calculated the *coverage*, *compactness,* and *compression ratio* for all the discovered patterns in a dataset, normalize the values for each variable independently and linearly to [0, 1]. One of their proposed combinations, which we refer to as the *threeCs* variable, is again multiplicative, with $a = b = c = 1$.[7]

Expected occurrences: Let Z be a random variable giving the number of times the pattern $P \subseteq D$ or one of its translations occur over the excerpt (dataset), and let $\mathbb{E}(Z)$ be the expectation of Z.

Given a dataset D, suppose $\mathbf{d}_1^*, \mathbf{d}_2^*, \ldots, \mathbf{d}_n^*$ is a list consisting of each datapoint $\mathbf{d}_i \in D$, but stripped of its ontime value. Then remove repetitions from this list to form the set $D' = \{\mathbf{d}_1', \mathbf{d}_2', \ldots, \mathbf{d}_{n'}'\}$. Each $\mathbf{d}_i' \in D'$ has a *relative frequency of occurrence* in the list $\mathbf{d}_1^*, \mathbf{d}_2^*, \ldots, \mathbf{d}_n^*$, which is recorded in the probability vector $\pi = (\pi_1, \pi_2, \ldots, \pi_{n'})$. It is supposed that the music can be modeled by independent and identically distributed random variables $X_1, X_2, \ldots$, each having the distribution given in π. This is what is meant by a "zero-order model" (Conklin & Bergeron, 2008, p. 64). The probability of seeing the pattern $P \subseteq D$ is the probability of the event

$$A_1 = \{X_1 = \mathbf{p}_1' = \mathbf{d}_{i(1)}', X_2 = \mathbf{p}_2' = \mathbf{d}_{i(2)}', \ldots, X_l = \mathbf{p}_l' = \mathbf{d}_{i(l)}'\}, \quad (10)$$

so that

$$\mathbb{P}(A_1) = \textstyle\prod_{j = 1, 2, \ldots, l} \pi_{i(j)}. \quad (11)$$

There may also be translations $P_2, P_3, \ldots, P_M$ of P with corresponding events $A_2, A_3, \ldots, A_M$ that have nonzero probability. The probability of seeing the pattern $P \subseteq D$ *or one of its translations* is the probability of the event $A = A_1 \cup A_2 \cup \ldots \cup A_M$. These events are mutually exclusive, so

$$\mathbb{P}(A) = \textstyle\sum_{i = 1, 2, \ldots, M} \mathbb{P}(A_i). \quad (12)$$

Let Y be an indicator variable for the event A, and Z be the random variable for the number of times A happens across a dataset consisting of n datapoints. It follows that $Z = Y_1 + Y_2 + \ldots + Y_K$, where K is yet to be determined. It is $\mathbb{E}(Z)$, the expected number of times the event A happens across a dataset, that is required:

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(\textstyle\sum_{i = 1, 2, \ldots, K} Y_i) & (13)\\ &= \textstyle\sum_{i = 1, 2, \ldots, K} \mathbb{E}(Y_i) & (14)\\ &= K\mathbb{E}(Y) & (15)\\ &= K\mathbb{P}(A), & (16) \end{aligned}$$

[7]It should be noted that Forth and Wiggins (2009) actually use an altered version of *coverage* as the first term in their product.

where (14) follows by linearity of expectation, (15) follows by $Y_1, Y_2, \ldots, Y_K$ being identically distributed, and (16) follows as Y is an indicator variable for the event A.

If it is only nonoverlapping patterns with no *interpolation* that are allowed, then $K = n/l$, or $\lfloor n/l \rfloor$, as this is the number of nonoverlapping patterns of size l with no gaps that can fit in a dataset of size n. If overlaps are allowed, $K = n - l + 1$. If any amount of interpolation is permitted then K could be as large as binom(n, l), the number of ways of choosing l objects from n. This makes the last definition of K too lenient: it leads to counting instances of patterns that would not be heard. For example suppose $P = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ and n is large. Even if $X_1 = \mathbf{p}_1'$, $X_2 = \mathbf{p}_2'$, and $X_n = \mathbf{p}_3'$, then this should not count as an instance of P, as the gap between X_2 and X_n is too large. It seems reasonable to limit the amount of interpolation according to the *span* of the original pattern, denoted $s(P, D)$, given by the denominator of (7). So K should be the number of ways of choosing l objects from n such that these objects do not exceed the span:

$$K = \text{binom}(s(P, D), l) + [n - s(P, D)]\cdot\text{binom}(s(P, D) - 1, l - 1). \quad (17)$$

It is common to transform likelihoods and expectations, as the untransformed values can have a large range—in the case of the *expected occurrences* variable, 10^{-183} to 10^{6}. Temperley (2009) uses a log transform, whereas we map $[10^{-183}, 10^{6}]$ to $[10^{-2}, 10^{2}]$ using a power law:

$$\text{expected\_occurrences}(P, D) = \|\mathbb{E}(Z)\| = a\mathbb{E}(Z)^b, \quad (18)$$

where $a \approx 71.12$ and $b \approx 0.02$ are constants determined by the chosen interval $[10^{-2}, 10^{2}]$. The reason for not using a log transform is that *expected occurrences* is derived from a nonnegative random variable. Therefore, it too should be nonnegative, but $\log x < 0$ for $x < 1$. It would be inappropriate to include negative values in subsequent variables (see *interest* and *score*), hence the use of a power law instead. We acknowledge that $[10^{-2}, 10^{2}]$ is an arbitrary—and therefore somewhat unsatisfactory—choice of interval. A second, less arbitrary solution was also considered (see *geometric mean likelihood* below), but this variable did not emerge as significant.

Geometric mean likelihood is a slight variant on (11):

$$\text{geom\_mean\_likelihood}(P, D) = \left(\prod_{j=1,2,\ldots,l} \pi_{i(j)} \right)^{(1/l)}, \quad (19)$$

where $\pi_{i(1)}, \pi_{i(2)}, \ldots, \pi_{i(l)}$ are the individual probabilities defined a few lines before (10).

Interest: The mathematical definition (Conklin & Bergeron, 2008) is

$$\text{interest}(P, D) = \text{occurrences}(P, D) / \text{expected\_occurrences}(P, D). \quad (20)$$

Score: The score of a pattern in a dataset is the squared difference between observed and expected occurrences, divided by expected occurrences (Conklin & Anagnostopoulou, 2001).

$$\text{score}(P, D) = (|TEC(P, D)| - \|\mathbb{E}(Z)\|)^2 / \|\mathbb{E}(Z)\|. \quad (21)$$

This variable bears a very close resemblance to Pearson's statistic (Davison, 2003), and it seems (from Conklin, 2008) that *interest* may be based on the related concept of likelihood ratio. Though relevant, Conklin's (2008) work on the discovery of distinctive patterns via use of an *anticorpus* is not yet explained fully enough to be included in the regression. The same observation applies to Temperley's (2009) Bayesian derivation of the probability that a given pattern occurs.

Prominence (also 'selection function') is the name given to a formula by Cambouropoulos (2006), involving the variables *cardinality*, *occurrences*, and *coverage*.

$$\text{prominence}(P, D) = l^a \cdot m^b \cdot 10^{c(\text{coverage}(P,D) - ml) / \text{coverage}(P,D)}, \quad (22)$$

where $a = 1$, $b = 2$, and $c = 3$.

Alternative prominence: There are now several explanatory variables (*mean-centered cardinality* × *occurrences*, *coverage*, and *prominence*) that involve a product of *cardinality* and *occurrences*. In (22) the variable *occurrences* is squared (the m^b term) but it could be argued that *cardinality* is the dominant factor, and that this variable ought to be squared instead. This is the intuition behind the variable *alternative prominence*:

$$\begin{aligned} \text{alt\_prominence}(P, D) &= \text{occurrences} \cdot (\text{cardinality} - 1)^2 / \text{dataset\_cardinality} \quad &(23) \\ &= m(|P| - 1)^2 / n. &(24) \end{aligned}$$

Maximum pitch center: The previous variables have not really taken into account the *musical* attributes of a pattern, and whether they might make it noticeable/important. Pearce and Wiggins (2007) quantified specific musical attributes in order to tease out deficiencies in their generated chorale melodies. The next ten variables are adapted from their work. *Pitch center* is defined as "the absolute distance, in semitones, of the mean pitch of a [pattern] . . . from the mean pitch of the dataset" (Pearce & Wiggins 2007, p. 78; see also von Hippel, 2000). By taking the maximum *pitch center* over all occurrences of a pattern, we hope to isolate either unusually high, or unusually low occurrences. Denoting the mean MIDI note number of a subset $Q \subseteq D$ by y_Q^*,

$$\text{max\_pitch\_center}(P, D) = \max\{|y_Q^* - y_D^*| : Q \in TEC(P, D)\}. \quad (25)$$

Signed pitch range is defined as the "distance, in semitones, of the pitch range of a [pattern] . . . from the mean pitch range of [other discovered patterns from the same dataset]" (Pearce & Wiggins 2007, p. 78; see also von Hippel, 2000). This variable is based on the suggestion that the larger a pattern's pitch range, the more noticeable it is. Denoting the range in semitones of a pattern P in a dataset D by $r(P)$, and the other discovered patterns in D by $P_1, P_2, \ldots, P_M$, with ranges $r(P_1), r(P_2), \ldots, r(P_M)$,

$$\text{signed\_pitch\_range}(P, D) = r(P) - M^{-1}\sum_{i=1,2,\ldots,M} r(P_i). \quad (26)$$

Unsigned pitch range is defined as the absolute value of the *signed pitch range* of a pattern. The previous variable took account of patterns with unusually large pitch ranges, based on findings that such patterns are more noticeable/important. But one could argue instead that unusually large *or* small pitch ranges give rise to noticeable patterns. Including the variable *unsigned pitch range* in the regression allows this argument to be considered as well.

Small intervals: This variable counts the number of small intervals (less than two steps on the stave) that are present in the melody line of a pattern. The intuition is that scalic, static, or stepwise melodies may be rated as more noticeable/important. A *top-line* definition of melody is applied: at each of the pattern's distinct ontimes there will be at least one datapoint present. At this ontime the melody takes the value of the highest pitch present. If a melody consists of the morphetic pitch (Meredith et al., 2002) numbers $y_1, y_2, \ldots, y_l$, then

$$\text{small\_intervals}(P, D) = |\{y_i : |y_i - y_{i-1}| < 2, i = 2, 3, \ldots, l\}|. \quad (27)$$

Intervallic leaps: This variable counts the number of intervallic leaps (greater than two steps on the stave) that are present in the melody line of a pattern, the intuition being that leaping melodies may be rated as more noticeable/important. The same *top-line* rule as above is applied. If a melody consists of the morphetic pitch numbers $y_1, y_2, \ldots, y_l$, then

$$\text{intervallic\_leaps}(P, D) = |\{y_i : |y_i - y_{i-1}| > 2, i = 2, 3, \ldots, l\}|. \quad (28)$$

Chromatic: The variable *chromatic* is the maximum number of non-key notes present, taken over all occurrences of a pattern. A particularly chromatic pattern is likely to be more noticeable—and rated higher therefore—than one that remains entirely in key.

Cadential: If a pattern contains a cadential figure, the variable *cadential* takes the value 1, and 0 otherwise. Cadences are often mentioned in music-analytical essays, as they help to segment long passages. Therefore, if a pattern contains a cadential figure it may be rated as noticeable/important. An algorithmic definition of *cadence* was not used, as it is more convenient to assign values by hand.

Phrasal: This variable takes high values for patterns that coincide with phrase marks. Take an occurrence of a pattern: its first note may coincide with the opening of a phrase mark and its last note may coincide with the closing of a phrase mark. A point is scored for each so that, per occurrence, a score of 0, 1, or 2 is possible. These scores are then averaged across all occurrences of a pattern.

Rhythmic density is defined as "the mean number of events per tactus beat" (Pearce & Wiggins 2007, p. 78; see also Eerola & North, 2000). It is likely that this variable and *compactness* will be highly positively correlated, but being a specific musical property, *rhythmic density* may be a more suitable variable. In a pattern $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_l\}$, let $\mathbf{p}_i$ have ontime x_i, $i = 1, 2, \ldots, l$. The tactus beats are then the integers from $a = \lfloor x_1 \rfloor$ to $b = \lfloor x_l \rfloor$, assuming that beats coincide with integer ontimes and that the bottom number in the time signature does not change over the course of the pattern. The rhythmic density of the pattern at beat $c \in [a, b]$, denoted $\rho(P, c)$, is given by

$$\rho(P, c) = |\{\mathbf{p}_i \in P : \lfloor x_i \rfloor = c\}| \quad (29)$$

so that

$$\text{rhythmic\_density}(P) = (b - a + 1)^{-1} \Sigma_{c \in [a, b]}\, \rho(P, c). \quad (30)$$

Rhythmic variability is defined as "the degree of change in note duration (*i.e.*, the standard deviation of the log of the event durations)" (Pearce & Wiggins 2007, p. 78; see also Eerola & North, 2000). While it has been suggested that patterns with much rhythmic variation are more difficult to perceive (Eerola & North, 2000), such patterns could actually be more distinctive/important. For a pattern $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_l\}$, denote the durations of each datapoint by $z_1, z_2, \ldots, z_l$. Then

$$\text{rhythmic\_variability}(P) = \text{sqrt}[l^{-1} \Sigma_{i = 1, 2, \ldots, l} (\log z_i - \log z^*)^2], \quad (31)$$

where $\log z^*$ is the mean of the log durations.

Signed dynamic level: The last four variables attempt to cover musical aspects that have not been taken into account already. Participants listened to recordings of excerpts and had access to the entire score of each piece (Paderewski, 1953). So consciously or unconsciously, participants may rate as more noticeable patterns that are marked and performed louder. The variable *signed dynamic level* involves mapping dynamic levels to numbers, and summing over the dynamic levels that apply to a pattern. Where occurrences of a pattern have different *signed dynamic levels*, the maximum value is taken. The mapping is given by

$$\boldsymbol{pp} \to -3, \boldsymbol{sp} \to -2, \boldsymbol{p} \to -1, \textit{mezza voce} \to 0, \boldsymbol{f} \to 1, \boldsymbol{sf} \to 2, \boldsymbol{ff} \to 3. \quad (32)$$

Chopin seems not to have distinguished between ***mp*** and ***mf***, preferring the term *mezzo voce*. Crescendi and diminuendi are mapped to 1/2 and –1/2 respectively. The terms *sotto voce* and *dolce* are mapped to –1 and 0 respectively.

Unsigned dynamic level: One could argue that it is naïve to assume that louder patterns are more noticeable/important. Could not a pattern that was played suddenly very softly be just as noticeable/important? The variable *unsigned dynamic level* takes the absolute values of the mapping used in the previous variable. In all other respects it is the same, and should assume large values for patterns that contain extreme dynamics, one way or the other.

Tempo fluctuation: Given fluctuations in tempo are used by composers and performers to emphasize aspects of the music, it seems reasonable to assume that patterns containing a pause mark, accelerando, ritardando, or rubato would be more noticeable/important than those that did not. The *tempo fluctuation* of a pattern is defined to be the number of tempo directions that apply to the pattern. Where occurrences of a pattern have different *tempo fluctuation* values, the maximum is taken.

Metric syncopation: The term *hemiola* applies to a scenario in which six beats are arranged as three groups of two, contrary to the prevailing arrangement of two groups of three, as with pattern *H* in Figure 19. *Hemiola* is a little too specific to be a variable in itself, so we define *metric syncopation* to apply to a scenario in which the prevailing arrangement of beats is contradicted. From the point of view of rating patterns, a pattern that contains a metric syncopation is likely to be noticeable/important. If a pattern contains a metric syncopation, then the variable *metric syncopation* takes the value 1, and 0 otherwise.

References

Bent, M. (2001). Isorhythm. In S. Sadie & J. Tyrrell (Eds.), *The new Grove dictionary of music and musicians* (2nd ed., Vol. 12, pp. 618–623). London, UK: Macmillan.

Biret, I. (1990). *Chopin mazurkas* (Volumes 1–2). Munich: Naxos.

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

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: A computational approach. *Music Perception, 23*, 249–267.

Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception, 26*, 75–94.

Chiu, S.-C., Shan, M.-K., Huang, J.-L., & Li, H.-F. (2009). Mining polyphonic repeating patterns from music data using bit-string based approaches. In R. Radhakrishnan & R. Yan (Eds.), *Proceedings of the IEEE International Conference on Multimedia and Expo* (pp. 1170–1173), New York: IEEE.

Collins, T., Thurlow, J., Laney, R., Willis, A., & Garthwaite, P. (2010). A comparative evaluation of algorithms for discovering translational patterns in Baroque keyboard works. In J. S. Downie & R. Veltkamp (Eds.), *Proceedings of the International Symposium on Music Information Retrieval* (pp. 3–8), Utrecht: International Society for Music Information Retrieval.

Conklin, D. (2008). Discovery of distinctive patterns in music. In *Proceedings of the International Workshop on Machine Learning and Music* (2 pages). Helsinki: University of Helsinki. Retrieved July 14, 2009 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.158.4152

Conklin, D., & Anagnostopoulou, C. (2001). Representation and discovery of multiple viewpoint patterns. In A. Schloss, R. Dannenberg, & P. Driessen (Eds.), *Proceedings of the International Computer Music Conference* (pp. 479–485). Havana: International Computer Music Association.

Conklin, D., & Bergeron, M. (2008). Feature set patterns in music. *Computer Music Journal*, *32*, 60–70.

Cook, N. (1987). *A guide to musical analysis*. Oxford, UK: Oxford University Press.

Cook, N. (1994). Perception: A perspective from music theory. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 64–95). Oxford, UK: Oxford University Press.

Cross, I. (1998). Music analysis and music perception. *Music Analysis, 17*, 3–20.

Davison, A. (2003). *Statistical models*. Cambridge, UK: Cambridge University Press.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics, 28*, 381–389.

Doraisamy, S., & Rüger, S. (2003). Robust polyphonic music retrieval with *N*-grams. *Journal of Intelligent Information Systems, 21*, 53–70.

Downes, S. (2001). Mazurka. In S. Sadie & J. Tyrrell (Eds.), *The new Grove dictionary of music and musicians* (2nd ed., Vol. 16, pp. 189–190). London, UK: Macmillan.

Downie, J. S. (2003). Music information retrieval. In B. Cronin (Ed.), *Annual review of information science and technology 37* (pp. 295–340). Medford, NJ: Information Today.

Eerola, T., & North, A. (2000). Expectancy-based model of melodic complexity. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.), *Proceedings of the International Conference on Music Perception and Cognition*. Keele, UK: Keele University. Retrieved June 12, 2010 from http://users.jyu.fi/~ptee/publications/3_2000.pdf

Ewens, W., & Grant, G. (2001). *Statistical methods in bioinformatics: An introduction*. New York: Springer-Verlag.

Forte, A., & Gilbert, S. E. (1982). *Introduction to Schenkerian analysis*. New York: W. W. Norton and Company.

Forth, J., & Wiggins, G. A. (2009). An approach for identifying salient repetition in multidimensional representations of polyphonic music. In J. Chan, J. Daykin, & M. Sohel Rahman (Eds.), *London algorithmics 2008: Theory and practice* (pp. 44–58). London, UK: College Publications.

Huron, D. (2001a). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19*, 1–64.

Huron, D. (2001b). What is a musical feature? Forte's analysis of Brahms's opus 51, no.1, revisited. *Music Theory Online*, *7*. Retrieved July 14, 2009 from http://mto.societymusictheory.org/issues/mto.01.7.4/mto.01.7.4.huron_frames.html

Knopke, I., & Jürgensen, F. (2009). A system for identifying common melodic phrases in the masses of Palestrina. *Journal of New Music Research, 38*, 171–181.

Lartillot, O. (2004). A musical pattern discovery system founded on a modeling of listening strategies. *Computer Music Journal*, *28*, 53–67.

Mathon, R., & Rosa, A. (1996). 2-(v, k, λ) designs of small order. In C. Colbourn & J. Dinitz (Eds.), *The CRC handbook of combinatorial designs* (pp. 3–41). Boca Raton, FL: CRC Press.

Meek, C., & Birmingham, W. P. (2003). Automatic thematic extractor. *Journal of Intelligent Information Systems, 21*, 9–33.

Meredith, D., Lemström, K., & Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research, 31*, 321–345.

Meredith, D., Lemström, K., & Wiggins, G. A. (2003). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. In *Cambridge*

Music Processing Colloquium (11 pages), Cambridge, UK: University of Cambridge. Retrieved May 12, 2009 from http://www.titanmusic.com/papers/public/cmpc2003.pdf

OCKELFORD, A. (2005). *Repetition in music: Theoretical and metatheoretical perspectives*. Aldershot: Ashgate.

PADEREWSKI, I. (Ed.). (1953). *Fryderyk Chopin: Complete works* (Volume 10). Warsaw: Instytut Fryderyka Chopina.

PEARCE, M., & WIGGINS, G. A. (2007). Evaluating cognitive models of composition. In A. Cardoso & G. Wiggins (Eds.), *Proceedings of the Fourth International Joint Workshop on Computational Creativity* (pp. 73–80). London, UK: Goldsmiths, University of London.

POPLE, A. (2004). Modeling musical structure. In E. Clarke & N. Cook (Eds.), *Empirical musicology: Aims, methods, prospects* (pp. 127–156). Oxford: Oxford University Press.

SCHENKER, H. (1973). *Harmony* (O. Jonas, Ed., E. Borgese, Trans.). Cambridge, MA: MIT Press. (Original work published 1906)

TAN, S.-L., SPACKMAN, M., & PEASLEE, C. (2006). The effects of repeated exposure on liking and judgements of musical unity of intact and patchwork compositions. *Music Perception, 23*, 407–421.

TEMPERLEY, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

TEMPERLEY, D. (2009). A unified probabilistic model for polyphonic music analysis. *Journal of New Music Research, 38*, 3–18.

VON HIPPEL, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception, 17*, 315–327.