Musicae Scientiae
Fall 2004, Vol VIII, n° 2, 175-205

Bayesian models of musical structure and cognition

DAVID TEMPERLEY
Eastman School of Music, University of Rochester

• ABSTRACT

This paper explores the application of Bayesian probabilistic modeling to issues of music cognition and music theory. The main concern is with the problem of key-finding: the process of inferring the key from a pattern of notes. The Bayesian perspective leads to a simple, elegant, and highly effective model of this process; the same approach can also be extended to other aspects of music perception, such as metrical structure and melodic structure. Bayesian modeling also relates in interesting ways to a number of other musical issues, including musical tension, ambiguity, expectation, and the quantitative description of styles and stylistic differences.

MUSIC AND PROBABILITY

More than forty years ago, Leonard B. Meyer remarked on the fundamental link between musical style, perception, and probability:

> Once a musical style has become part of the habit responses of composers, performers, and practiced listeners it may be regarded as a complex system of probabilities. [...] Out of such internalized probability systems arise the expectations — the tendencies — upon which musical meaning is built. [... T]he probability relationships embodied in a particular musical style together with the various modes of mental behavior involved in the perception and understanding of the materials of the style constitute the *norms* of the style (1967[1957], pp. 8-9).

To me (and I believe to many others who have read them), these words ring profoundly true; they seem to capture something essential about the nature of music and musical communication. Building on these ideas — towards an understanding of how probabilities shape musical style and perception — would seem to be a natural enterprise for music cognition and music theory.

Perhaps surprisingly, the probabilistic approach to musical modeling has not been very widely explored. It saw a flurry of activity in the late 1950's and 1960's,

and then fell into relative neglect for almost thirty years; very recently, it has enjoyed something of a resurgence. Most of this work, both from the earlier and later periods, has involved the *Markov chain*: a system of probabilities operating between events in a sequence (see, for example, Youngblood, 1958; Cohen, 1962; Hiller and Fuller, 1967; Conklin and Witten, 1995; Alamkan *et al.*, 1999; Ponsford *et al.*, 1999). My concern here is with a rather different approach to probabilistic modeling, sometimes known as Bayesian modeling. The Bayesian approach, I will argue, opens the door to a more musically sophisticated investigation of the probabilistic aspect of music than has been possible before. Whereas Markov models are fundamentally concerned only with relationships between surface elements (for example, the probability of one note or chord following another), the Bayesian approach is inherently concerned with *structure*, and the relationship between structure and surface: the way structures constrain surfaces in composition, and the way surfaces convey structures in perception[1].

In this paper I will explore several ways that the Bayesian perspective might inform and advance the study of music. My main concern will be with "key-finding", the perceptual process of identifying the key of a piece. I will suggest that Bayesian modeling offers an elegant and highly effective approach to this problem — an approach that also applies well to other aspects of music perception, such as metrical structure and melodic structure. I will also argue for the relevance of Bayesian modeling to a number of other musical issues, including musical tension, ambiguity, expectation, and the quantitative description of styles and stylistic differences.

A BRIEF INTRODUCTION TO BAYESIAN MODELING

Communication generally involves the transmission of a message from a producer to a perceiver. As perceivers, we are often given some kind of surface representation of a message (what I will simply call a *surface*); our task is to recover the underlying content that gave rise to it — the information that the sender was trying to convey — which I will simply call a *structure*. The problem is probabilistic in the sense that a single surface might arise from many different structures. We wish to know the structure that is most probable, given a particular surface — in the conventional notation of probability, we need to determine

(1) We should note that it is possible to use Markov models in such a way that does incorporate structure — so-called "hidden Markov models" (HMMs). In an HMM, some elements of the Markov chain are observable while others are hidden; the hidden elements may correspond to meaningful structural entities. (In some cases they do not, and simply serve as a way of weighting relationships between surface elements, as in Ponsford *et al.*, 1999.) Some HMM's are very close in spirit to Bayesian models; examples in the musical domain include Raphael's automatic transcription system (2002) and Bod's model of melodic segmentation (2002). Several other musical studies reflecting, or relating to, the Bayesian perspective will be discussed below.

$$\text{argmax}_{\text{structure}}\ p(\text{structure} \mid \text{surface}) \tag{1}$$

where "$\text{argmax}_{\text{structure}}$" means the value of "structure" that maximizes the expression to the right.

The solution to this problem lies in Bayes' rule, a fundamental theorem of probability. This states that, for any two events A and B, the probability of A given B can be computed from the probability of B given A, as well as the overall probabilities (known as the "prior probabilities") of A and B:

$$p(A \mid B) = \frac{p(B \mid A)\ p(A)}{p(B)} \tag{2}$$

In our terms, for a given surface and a given structure:

$$p(\text{structure} \mid \text{surface}) = \frac{p(\text{surface} \mid \text{structure})\ p(\text{structure})}{p(\text{surface})} \tag{3}$$

To find the structure that maximizes the left side of equation 3, we need only find the structure that maximizes the right side — and this turns out to be easier. Note, first of all, that "p(surface)"— the overall probability of a given surface —will be the same for all values of "structure". This means that it can simply be disregarded. Thus

$$\text{argmax}_{\text{structure}}\ p(\text{structure} \mid \text{surface}) = \text{argmax}_{\text{structure}}\ p(\text{surface} \mid \text{structure})\ p(\text{structure}) \tag{4}$$

Thus, to find the most probable structure given a particular surface, we need to know — for every possible structure — the probability of the surface given the structure, and the prior probability of the structure.

Two other points will be relevant in what follows. The probability of a surface and a structure occurring in combination is

$$p(\text{surface} \ \&\ \text{structure}) = p(\text{surface} \mid \text{structure})\ p(\text{structure}) \tag{5}$$

Note that the expression on the right is exactly what must be computed to find the most probable structure given a surface (equation 4). Also of interest is the prior probability of a surface, which sums the expression in equation 5 over all possible structures:

$$p(\text{surface}) = \sum p(\text{surface} \mid \text{structure})\ p(\text{structure}) \tag{6}$$

The Bayesian approach has proven to be extremely useful in a number of areas of cognitive modeling and information processing. An illustrative example is the problem of speech recognition. In listening to speech, we are given a sequence of

phonetic units — phones — and we need to determine the sequence of words that the speaker intended. In this case, then, the sequence of phones is the surface and the sequence of words is the structure. (Determining the sequence of phones that was spoken is in itself a complex process, but we will not consider that here.) The problem is that a single sequence of phones could result from many different words. Consider the phone sequence [ni], as in "the knights who say 'Ni'", from *Monty Python and the Holy Grail* (this example is taken wholesale from Jurafsky and Martin, 2000). Various words can be pronounced [ni], under certain circumstances: "new", "neat", "need", "knee", and even "the". (This may seem counterintuitive; it is due to the fact that the pronunciation of words can vary greatly depending on context. For example, in the phrase "neat little", the final "t" on "neat" may be omitted, leaving [ni].) However, not all of these words are equally likely to be pronounced [ni]. The probability of the pronunciation [ni] given each word (according to Jurafsky and Martin, based on analysis of a large corpus of spoken text) is as follows:

new	.36
neat	.52
need	.11
knee	1.00
the	0

This, then, is "p(surface | structure)" for each of the five words. (For all other words, p(surface | structure) = 0.) In addition, however, some of the words are more probable than others — the prior probability of each word (according to Jurafsky and Martin) is

new	.001
neat	.00031
need	.00056
knee	.000024
the	.046

This gives us "p(structure)" for each word. Taking the product of the two values for each word gives

new	.00036
neat	.000068
need	.000062
knee	.000024
the	0

The structure maximizing this value for the phone string [ni] — and hence the most probable structure given that surface — is the word "new". Of course, this model could be improved by the addition of other information. Most importantly, the probability of a given word depends greatly on the context: in the context "I scraped my...", we expect "knee" much more than "new". Incorporating information of this kind could give us a much better estimate of the prior probability of each word.

Bayesian modeling is also widely used in syntactic parsing (Charniak, 1996; Manning and Schütze, 2000). The "structure" in this case can be defined as a syntactic tree, down to a set of syntactic categories (noun, verb, etc.). For any structure, the prior probability can be calculated from the combined probability of all the syntactic expansions involved — S (sentence) expanding to NP (noun phrase) + VP (verb phrase), VP expanding to V (verb) + PP (prepositional phrase), etc. The probability of the surface given the structure then depends on the probability of each word given a certain syntactic category: for example, the probability of a noun being "dog" (as opposed to "cat" or "mouse"). In this way we can calculate both "p(structure)" and "p(surface | structure)" for each possible structure, and thus determine the most likely structure given the surface.

A Bayesian model of key-finding

Consider the problem of key-finding: inferring the key of a piece from the notes. Key-finding is a centrally important process in music cognition, one that has been studied quite widely from both experimental and computational perspectives (Longuet-Higgins and Steedman, 1971; Holtzmann, 1977; Bharucha, 1987; Butler, 1989; Krumhansl, 1990; Leman, 1995; Vos and Van Geenen, 1996). As with speech recognition, inferring the "surface" — the notes of a piece — from sound input is in itself a highly complex problem; we will simply assume that the notes have already been identified. We will allow for the possibility of modulations — that is, the key may change from one moment to the next. We will, however, assume a division of the piece into segments — roughly corresponding to measures — such that the key may change from one segment to the next, but not within segments. Thus a single key must be chosen for each segment. (Let us overlook the complication of pivot chords for now; a possible way of incorporating these will be considered later.)

Defined in this way, key-finding reflects a fundamental similarity to the processes described above: the problem is to infer the most probable structure, given a surface. In this case, the structure is a sequence of keys; the surface is a pattern of notes. To solve this problem using the Bayesian method, we need to know — for all possible structures — the probability of the structure itself and the probability of the surface given the structure.

First consider the probability of a structure itself: a labeling of each segment with a key. We will assume that, for the initial segment of a piece, all 24 keys (12 major and 12 minor) are equally probable. For subsequent segments, there is a high

probability of remaining in the same key as the previous segment; switching to another key carries a lower probability. This captures the "inertia" of key: perceptually, we tend to assume that the current key remains in force, unless there is strong evidence to the contrary. (For example, a G major triad in the context of C major will normally be heard as V of C rather than I of G.) Let us assume, for any segment except the first, a probability of .8 of remaining in the same key as the previous segment; this leaves a probability of .2 for moving to one of the other 23 keys (as the probabilities of all possible outcomes must sum to 1), or a probability of .2/23 = .0087 for each key. (We consider all key changes to be equally likely, though this is undoubtedly an oversimplification; I discuss this further below.) The probability of a complete key structure can then be calculated as the product of these probabilities — we will call them "modulation scores" (S_m) — for all segments. For a structure of four segments, C major - C major - C major - G major, the probability will be

$$1/24 \times .8 \times .8 \times .2/23 = .000232 \qquad (7)$$

The next task is to define the probability of a surface given a structure. This problem could be solved in many different ways; I will propose one solution here, and then consider other possibilities later on. Let us suppose, for the moment, that the only information relevant to the key of a segment is the set of pitch-classes that the segment contains. We further assume that, in each segment, the composer makes twelve independent decisions as to whether or not to use each pitch-class[2]. These probabilities can be expressed in a twelve-valued vector — conventionally known as a "key-profile". We could base these key-profiles on actual data as to how often each pitch-class is used in segments of a particular key. Such data is shown in Figure 1 for the Kostka-Payne corpus — a corpus of 46 excerpts from the common-practice repertoire, taken from the workbook accompanying Stefan Kostka and Dorothy Payne's textbook *Tonal Harmony* (1995)[3]. The workbook is accompanied by an instructors' manual containing analyses by the authors, showing harmonic analyses and modulations; thus data could be gathered on pitch-class distribution relative to the *local* key — something that has not been possible in previous studies of this kind

(2) Bear in mind that what is being proposed here is not a model of composition, but of perception. That is, the suggestion is not that composers really do make twelve independent decisions whether or not to include each pitch-class in each segment — only that listeners assume this for the purposes of key-finding.

(3) Segments were defined by metrical units, using the lowest (fastest) metrical level whose beats were at least one second apart. (Tempi were chosen for each excerpt on an informal basis.) See Temperley (2001) for further information about how the corpus was constructed.

The entire corpus contains 896 segments and 9748 notes. However, in cases where a segment (or part of a segment) was analyzed as being in two keys simultaneously (e.g. a pivot chord), it was counted twice, once in each key; for this reason the actual number of segments counted was 955, not 896.

(Youngblood, 1958; Knopoff and Hutchinson, 1984). The data in Figure 1 is collapsed over all major keys and all minor keys, so that the profiles represent pitch-classes relative to keys — scale degrees, essentially. As an example, scale degree 1 (the tonic) occurs in .748 (74.8%) of segments in major keys; scale degree #4, by contrast, occurs in only .096 (9.6%) of segments. The profiles reflect conventional musical wisdom, much as we would expect. In both major and minor profiles, scalar degrees have higher values than chromatic ones, and notes of the tonic triad score higher than other notes of the scale. (One feature of the profiles worth noting is the very strong presence of the harmonic minor scale, that is, the high frequency in minor keys of the lowered sixth and raised seventh compared to the raised sixth and lowered seventh.)[4]

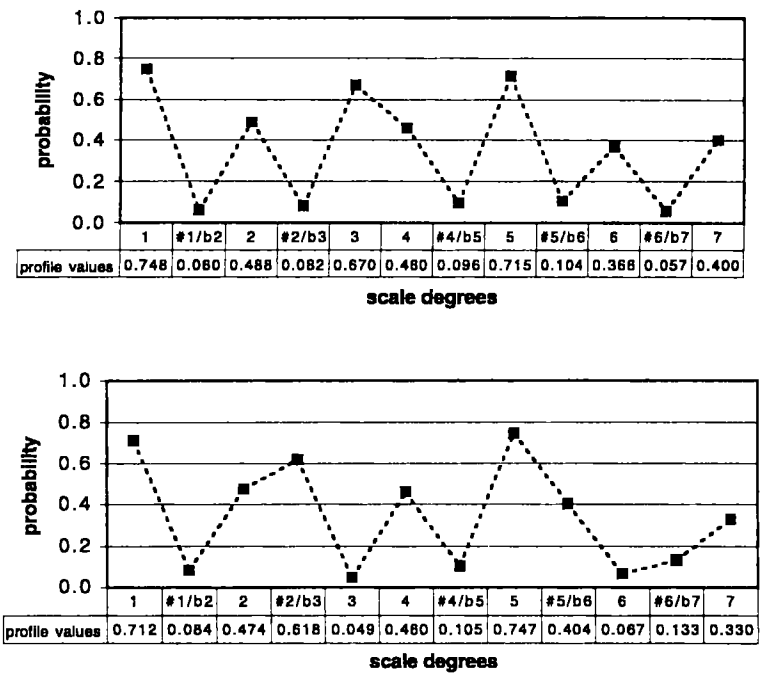


Figure 1.

Key-profiles for a Bayesian key-finding model, for major keys (above) and minor keys (below). The profiles are based on the frequency of occurrence of each scale-degree (relative to the current key) in the Kostka-Payne corpus, for major and minor keys. Profile values indicate the proportion of segments in which each scale-degree occurred.

(4) Another noteworthy feature is that degree 5 is more frequent than degree 1 in minor keys. This is odd and unexpected, and suggests that the sample may not be large enough to represent general practice accurately (though it is possible that this does represent general practice). Repeating this tally with a larger sample would certainly be worthwhile.

The probability of a scale degree *not* occurring in a segment is, of course, 1 minus the score in the profile: for scale degree 1 in major keys, 1–.748 = .252. For a given key, the probability of a certain pitch-class set being used is then given by the product of the key-profile values — we could call these "pc scores" (S_{pc}) — for all pitch-classes present in the segment (p), multiplied by the product of "absent-pc" scores ($S_{\text{-}pc}$) for all pitch-classes not present (-p).

$$\text{key-profile score} = \prod_{p} S_{pc} \prod_{\text{-}p} S_{\text{-}pc} \tag{8}$$

To find the most probable structure given a surface, we need to calculate p(structure) p(surface | structure). This can be calculated, for an entire piece, as the product of the modulation scores (S_m) and the key-profile scores for all segments (s):

$$p(\text{structure})\ p(\text{surface} \mid \text{structure}) = \prod_{s} \left(S_m \prod_{p} S_{pc} \prod_{\text{-}p} S_{\text{-}pc}\right) \tag{9}$$

A standard move in Bayesian modeling is to express such a formula in terms of logarithms. (This is convenient simply because it avoids the tiny numbers that result from multiplying many probabilities together.) Since the function ln x is monotonic, two values of ln x will always have the same ranking of magnitude as the corresponding values of x; if our only aim is to find the maximum value of x, then using ln x instead works just as well. The logarithm for the right side of equation 9 can be expressed as

$$\sum_{s} \left(\ln S_m + \sum_{p} \ln S_{pc} + \sum_{\text{-}p} \ln S_{\text{-}pc}\right) \tag{10}$$

Now the score is a sum of segment scores; each segment score is itself the sum of a modulation score, pc scores for present pc's, and absent-pc scores for absent pc's.

I have said how the probability of key structures can be calculated, but not how the most probable key structure of a piece could actually be found. This is a computationally non-trivial problem. Due to the modulation scores, the probability of a key in one segment depends on the key of the previous segment. Thus, to find the most probable key structure overall, the model must calculate the probabilities of all complete analyses of the entire piece; and the number of these grows exponentially with the number of segments. This problem, and similar problems that arise with Bayesian models in other domains, can be solved using the technique of dynamic programming (see Jurafsky and Martin, 2000; Temperley, 2001). However, this is not our concern for now; let us simply assume that the model considers all possible key structures, calculates their probabilities, and chooses the most probable one.

Before proceeding further, we should examine some objections that might be raised to the model just proposed. The model — construed as a model of human cognition — assumes that key-finding operates primarily on information about the

collection of pitch-classes in use, and is more or less indifferent to the way they are arranged horizontally and vertically. (The model is also insensitive to the repetition of pitch-classes within segments, though not across segments. That is to say, if one pitch-class occurs many times in a passage and another occurs only once, this is likely to affect the model, as the first pitch-class will occur in more segments than the second.) This "statistical" aspect of the model might seem counterintuitive. However, experiments in which the effect of pitch-class distribution is systematically controlled (*e.g.* by using randomly-ordered melodies generated from key-profiles) suggest that distributional information alone can, indeed, be a powerful cue to tonal orientation (Oram and Cuddy, 1995; Smith and Schmuckler, 2000). The role of pitch-class content in key-finding is also demonstrated anecdotally by so-called "pan-diatonic" music, which uses a diatonic scale collection but in non-traditional ways. A case in point is shown in Figure 2, the opening of Stravinsky's *Sonata for Two Pianos*. This excerpt projects an unmistakable sense of an F major tonality in mm. 1-4 modulating to C major in mm. 5-9, despite the general absence of traditional structures of tonal harmony and voice-leading. This suggests that such structures are not necessary to project a sense of tonality; pitch-class distribution alone has surprising power as an indicator of key[5].

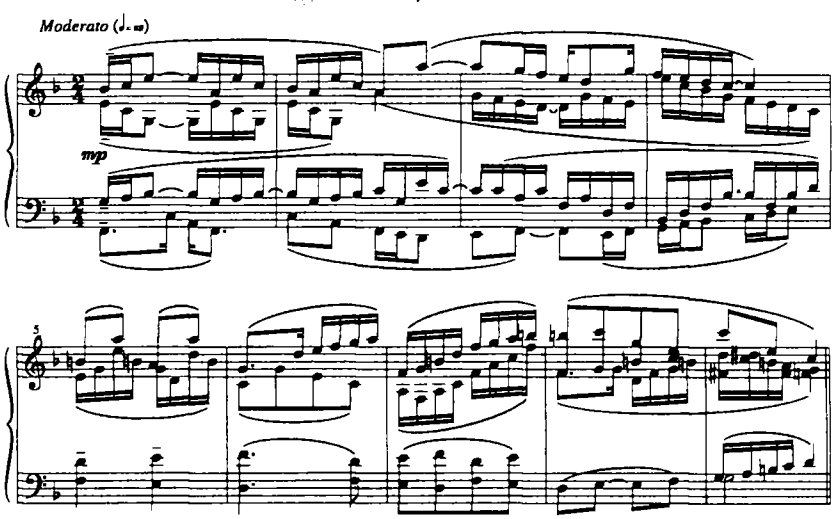


Figure 2. Stravinsky, Sonata for Two Pianos, *I, mm. 1-9.*

(5) In Figure 2, one might point to the F major triad in the bass line in mm. 1-2 as a harmonic cue. However, to explain the sense of modulation to C major in these terms is much more difficult. One finds a C major triad outlined in m. 6; but E minor, G major, and F major triads are also present in mm. 5-7 and at least as prominent. The top voice in m. 9 also outlines a C major triad, but the sense of C major is established well before this.

To experiment with the tonal implications of melodies generated randomly from a key-profile, visit the Melisma Melody Generator at www.link.cs.cmu.edu/melody-generator.

While pitch-class content is unquestionably an important factor in key-finding, it is not the only factor. Certainly, there are some cases where the arrangement of pitches can affect their tonal implications. A simple example is shown in Figure 3, proposed by David Butler (1989), in which the same pitches are arranged in two different ways: Figure 3a strongly implies C major, whereas Figure 3b is more ambiguous. I have suggested elsewhere (Temperley, 2001) that such phenomena point to a role for harmonic structure in key-finding; the progression G7-C implies C major much more strongly than E-F. In practice, however, consideration of harmony appears to be necessary rather rarely. The computational tests presented below provide further evidence as to the role of pitch-class content in key determination, as well as other factors that may be involved.

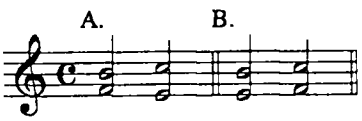


Figure 3. The same pitches arranged differently can have different tonal implications.

One aspect of key is completely neglected by the model presented above: this is its hierarchical aspect. A tonal piece generally has a single main key, but may have secondary key sections within that, and perhaps lower-level tonicizations as well. The model presented here cannot capture this multi-leveled structure, but simply generates a single level of key sections. (It also knows nothing of the conventions of key structure in common-practice music — in particular, the fact that the key established at the beginning of a piece is likely to return at the end.) I will not address this issue here, but leave it as a problem for the future. It is generally assumed that pieces have a basic, intermediate level of key — as indicated by modulations in a Roman numeral analysis; it is this level that concerns us in the present study.

Testing and comparison with other models

The Bayesian key-finding model proposed above has antecedents in two other models of key-finding. The Krumhansl-Schmuckler (hereafter K-S) key-finding algorithm, described most fully in Krumhansl (1990), is based on a set of key-profiles representing the stability or compatibility of each pitch-class relative to each key. (Table 1 shows the model's key-profile values for major and minor keys.) The key-profiles are based on experiments in which subjects were played a musical context such as a cadence or scale, followed by a pitch, and were asked to judge how well the pitch "fit" given the context. A high value for a pitch-class in a given key-profile means that the pitch-class was judged to fit well with that key. Given these profiles, the model judges the key of a piece by generating an "input vector" for the piece; this is, again, a twelve-valued vector, showing the total duration of each pitch-class

in the piece. The correlation value, r, is then calculated between each key-profile vector and the input vector, using the standard correlation formula:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(\sum(x-\bar{x})^2\sum(y-\bar{y})^2)^{1/2}} \qquad (11)$$

where x = input vector values; $\bar{x}$ = the average of the input vector values; y = the key-profile values for a given key; and $\bar{y}$ = the average key-profile value for that key. The key whose profile yields the highest correlation value is the preferred key.

Table 1
Key-profiles for three different key-finding models

Scale degree	K-S model major	K-S model minor	CBMS model major	CBMS model minor	Bayesian model major	Bayesian model minor
1	6.35	6.33	5.0	5.0	.748	.712
♯1/♭2	2.23	2.68	2.0	2.0	.060	.084
2	3.48	3.52	3.5	3.5	.488	.474
♯2/♭3	2.33	5.38	2.0	4.5	.082	.618
3	4.38	2.60	4.5	2.0	.670	.049
4	4.09	3.53	4.0	4.0	.460	.460
♯4/♭5	2.52	2.54	2.0	2.0	.096	.105
5	5.19	4.75	4.5	4.5	.715	.747
♯5/♭6	2.39	3.98	2.0	3.5	.104	.404
6	3.66	2.69	3.5	2.0	.366	.067
♯6/♭7	2.29	3.34	1.5	1.5	.057	.133
7	2.88	3.17	4.0	4.0	.400	.330

In Temperley (2001), I proposed an alternative model of key-finding (which I will call the CBMS model) building on the proposal of Krumhansl and Schmuckler. The original profiles seemed problematic in certain respects — particularly the higher value for the lowered seventh as opposed to the leading-tone in the minor profile — and some adjustments were proposed, leading to improved performance (see Table 1). Another problem with the K-S model was that intensive repetition of a pitch-class within a short time-span seemed to give too much weight to that pitch-class. In Figure 4, for example, the repeated E's cause the K-S model to choose E minor as the key, whereas C major would clearly be a better choice. To address this, I argued that some kind of division of the input into small segments should be assumed, and an input vector calculated for each segment, in which each pitch-class gets a 1 if it is present and 0 if it is not. The match between the input profile and the key-profiles is also calculated in a simpler way: For each key-profile, we take the product of all input vector values with the corresponding key-profile values, and sum these products. Since the input vector values are all 1 or 0, this simply amounts

to adding the key-profile values for the pitch-classes that score 1 in the input vector. The division of the piece into segments also allows the model to handle modulation (which the original K-S model did not)[6]. The CBMS model makes a key judgment for each segment, but imposes a change penalty if the key for one segment differs from the key for the previous segment. These penalties are then combined additively with the key-profile scores to choose the best key for each segment.

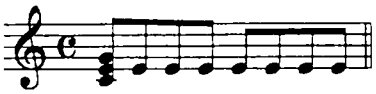


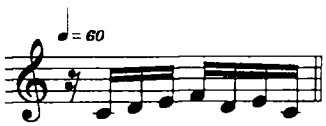
Figure 4.

The three models just presented — the K-S model, the CBMS model, and the Bayesian model — have much in common. All three of them are based on the concept of key-profiles — an ideal pitch-class distribution for a key, to which the actual pitch or pitch-class distribution of a piece is matched. The key-profiles used in the three models are also quite similar, though there are some subtle differences, as can be seen from Table 1. Figure 5 presents a simple musical example, showing how the score for C major would be calculated by all three models[7]. The CBMS model and the Bayesian model are particularly similar. Both models involve a division of the piece into segments; key judgments are made for each segment, choosing the key whose profile best matches the pitch-classes in the segment and also factoring in a penalty for key changes between segments. (In the Bayesian model, this "penalty" is reflected in the fact that probability of remaining in the same key is higher than the probability of modulating.) If we pretend that the key-profile values and modulation penalties from the CBMS model are really logarithms of other numbers, then the two models are virtually identical[8]. There is one significant difference: in the CBMS model, key scores are produced by summing the key-profile scores for the pc's that are present; in the Bayesian model, we also add "absent-pc" scores for pc's that are absent.

(6) There have been other proposals for extensions of the K-S model to handle modulation. Krumhansl (1990) proposes mapping keys onto a four-dimensional space and then tracking the movement of the piece across this space; see also Toiviainen and Krumhansl (2003). Huron and Parncutt (1993) suggest an exponential decay model, in which the input vector is a weighted sum of all previous events (see Temperley, 2001, pp. 198-201, for discussion).

(7) For the K-S model and the Bayesian model, the chosen key on Figure 5 is C major; for the CBMS model, C major and F major are tied for first place.

(8) There are some superficial differences. Since the scores in the Bayesian model are all logarithms of probabilities (numbers between 0 and 1), they will all be negative numbers. Also, the Bayesian model adds modulation scores for all segments, not just modulating segments. These are simply cosmetic differences which could be removed by scaling the values differently in the CBMS model, without changing the results.



Krumhansl-Schmuckler model

input vector = {.5, 0, .5, 0, .5, .25, 0, 0, 0, 0, 0, 0}
key-profile vector for C major = {6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88}
score for C major (correlation value) = 0.622

CBMS model

input vector = {1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0}
key-profile vector for C major = {5.0, 2.0, 3.5, 2.0, 4.5, 4.0, 2.0, 4.5, 2.0, 3.5, 1.5, 4.0}
score for C major = 5.0 + 3.5 + 4.5 + 4.0 = 17.0

Bayesian model

input vector = {1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0}
key-profile vector for C major = {0.748, 0.060, 0.488, 0.082, 0.670, 0.460, 0.096, 0.715, 0.104, 0.366, 0.057, 0.400}
score for C major = ln ((0.748) x (1–0.060) x (0.488) x (1–0.082) x (0.670) x (0.460) x (1–0.096) x (1–0.715) x (1–0.104) x (1–0.366) x (1–0.057) x (1–0.400)) = –4.82

Figure 5.

Sample calculations for three key-finding models. The calculations are for the key of C major, given the input shown below (treated as a single segment). Input vectors and key-profile vectors show the values for the twelve pitch-classes (C, C♯, D, ... B).

The three models were subjected to an empirical test, using the Kostka-Payne corpus discussed earlier. (It may seem questionable to use this corpus for testing, as it was also used for setting the parameters of the Bayesian model; I will return to this issue.) The corpus contains 896 segments and a total of 40 modulations (as indicated by the authors' analyses). The output of the models was compared with the authors' analyses; each model was simply scored on the proportion of segments that were labeled correctly[9]. (It was necessary to modify the K-S model somewhat, since the original model has no mechanism for handling modulations. In this test, the K-S model evaluates each segment independently using the correlation formula, and imposes a change penalty for changes between segments.) With each model, different values of the change penalty were tried, and the value was used that yielded the best

(9) In cases where the correct analyses contained two keys in a segment (e.g. a pivot chord), half a point was given to the model if its judgment corresponded with either "correct" key.

A computer program which implements all three of the key-finding models discussed here is available at www.link.cs.cmu.edu/melisma; the Kostka-Payne corpus is also available there in MIDI format.

performance. Table 2 shows the results. The Bayesian model judged 86.5% of segments correctly, slightly better than the CBMS model (83.8%) and significantly better than the K-S model (67.0%). It seemed likely, however, that some of this difference in performance was due simply to differences between the key-profiles. For this reason, the same test was run using the Kostka-Payne profiles with all three models. This improved the performance of the CBMS program to 86.3% and the K-S model to 80.4%.

Table 2

Results on the Kostka-Payne corpus, for five different key-finding models

Model	Optimal change penalty	Percentage of segments correct
K-S model (using K-S profiles)	2.3	67.0%
CBMS model (using CBMS profiles)	12.0	83.8%
Bayesian model (using Kostka-Payne profiles)	0.998	86.5%
K-S model (using Kostka-Payne profiles)	1.8	80.4%
CBMS model (using Kostka-Payne profiles)	2.5	86.3%

Note: For the K-S and CBMS models, the change penalty represents the penalty assigned to an analysis for each change of key. For the Bayesian model, it represents the probability of remaining in the same key; given a change penalty of 0.998, the probability of changing to any other key is (1–0.998)/23 = 0.00008.

The approach to testing used here — in which the Bayesian key-finding model (and other models) are tested on the Kostka-Payne corpus, using parameters derived from the same corpus — is problematic. Ideally, one would derive the model's parameters from one corpus and then test it on another. This is difficult at present, due to the small amount of encoded data available. (As emphasized earlier, it is important to have data in which local keys are encoded, not merely the global key of the piece.) One could train on part of the Kostka-Payne corpus and test on another part, but this would result in an even smaller database for both training and testing. The magnitude of this problem depends on how representative the scale degree distribution of the Kostka-Payne corpus is of common-practice music generally. If another corpus (call it corpus X) has virtually the same scale degree distribution as the Kostka-Payne corpus, then training the model on corpus X should produce virtually the same results (on any corpus) as training on the Kostka-

Payne corpus. Further accumulation of data will be necessary to resolve this question[10].

Comparison of the output of the Bayesian model with the correct analyses revealed several reasons for the model's errors. The most common source of error was that the model's rate of modulation was either too fast or too slow: in some cases, the model considered something a modulation where the human analysts had only considered it a tonicization, or vice versa. In a few cases, it seemed likely that considering harmonic information would help the model's performance. In particular, the model had trouble with chromatic chords such as augmented sixths, whose tonal implications contradict their pitch content; a typical augmented-sixth chord in C minor contains F♯, which is normally foreign to C minor. Considering pitch-spelling information (*e.g.*, the distinction between A♭ and G♯) might also have helped. With regard to the CBMS model, it was found that introducing pitch-spelling distinctions (so that, for example, E has a higher value in the C major profile than F♭ does) improved performance from 83.8% to 87.4%; incorporating such distinctions into the Bayesian model might well yield a similar improvement. (Using pitch-spelling as input might seem problematic, if we are trying to model cognition; but it is defensible if we suppose that pitch-spellings can be inferred based on context, as proposed in Temperley [2001], and that this information is then available to influence key-finding.) In most cases, the model's errors were the same as those of the CBMS model, whose performance on the Kostka-Payne corpus in discussed at greater length in Temperley (2001).

The Bayesian perspective suggests several other ways that the model might be improved. In the first place, one could set the change penalty systematically, according to the actual number of modulations in the Kostka-Payne corpus. Since the corpus contains 40 modulations and 850 segments (excluding the initial segment of each excerpt), the probability of a modulation should be 40/850 or .047. (By contrast, the value of this parameter found to be optimal through trial-and-error adjustment was .002.) As an experiment, the change penalty was set to reflect this, adding a score of ln(.047/23) for each modulating segment (assuming again that moves to any of the other 23 keys are equally likely), and ln(1–.047) for non-modulating

(10) A further test was done using another corpus, consisting of 10 long (1 minute or more) excerpts from common-practice pieces by a variety of composers, selected and analyzed into key sections by the author. The profiles yielded were qualitatively similar to those of the Kostka-Payne corpus, reflecting the same three-level hierarchy of chromatic, diatonic, and tonic-triad scale degrees. When the profiles of this corpus were used as the parameters for the model, the model labeled 84.9% of segments correct on the Kostka-Payne corpus (marginally lower than the rate of 86.5% obtained using the Kostka-Payne parameters). This suggests that the test reported previously was not greatly biased by the use of the same dataset for training and testing.

In an earlier publication (Temperley, 2002) I reported that the performance of the Bayesian key-finding model (using the Kostka-Payne parameters) on the Kostka-Payne corpus was 77.1%. This was due to an error in the implementation that has now been corrected.

segments. This produced a score of only 81.2% correct — somewhat less than the optimal performance of 86.5%. A second possible improvement would be to modify the assumption that all keys are equally likely. Major keys are certainly more common than minor keys; in the Kostka-Payne corpus, 70.0% of the segments are in major keys. Thus it might be advantageous to give them a higher probability. However, analysis of the Bayesian model's output showed that, even without a special preference for major keys, the model was achieving almost exactly the right proportion of major segments (70.1%). This suggested that adding a preference for major keys was unlikely to improve performance. Finally, one might consider attaching weights to different transitions between keys. When in C major, one is much more likely to move to G major than to (for example) F♯ major. This has not yet been attempted; to construct such a model based on empirical values would require more data, as the Kostka-Payne corpus itself provides only 40 modulations. (Of course, the model could also be made more sophisticated by building in other information about key structures, such as the fact that pieces are likely to end in the same key they began in.)

Before continuing, we should examine one other possible approach to a Bayesian model of key-finding. In the model above, a key-profile is treated, essentially, as 12 independent probability functions indicating the probability of each scale degree occurring in a segment (and, thus, the probability of each pitch-class relative to each key). This approach is not ideal, since it requires a prior segmentation of the piece; there is little reason to think that such a segmentation is involved in human key-finding. An alternative approach — simpler, in some ways — would be to treat each key-profile as a single probability function (so that the 12 values of the profile would sum to 1). This function could then be used to estimate the scale-degree probabilities of an event given a certain key. Events could be treated as independent; the probability of a note sequence given a key would then be given by the product of the key-profile scores for all events — or, in logarithmic terms, the sum of scores for all events. This method resembles the "weighted-input" approach of Krumhansl and Schmuckler's original model, discussed earlier, in which the input vector reflects the number and duration of events of each pitch-class. The problem with this approach has already been noted: it tends to give excessive weight to repeated events. Initial tests of the key-profile model showed significantly better performance when repetitions of a pitch-class within a segment were not counted. Thus it appears that treating the key-profiles as probability functions for independent events is unlikely to work very well. (Intuitively, in Figure 4, the weighted-input approach assumes a generative model in which the composer decides to use C and G once, and then makes eight independent decisions to use E. But a more plausible model is that the composer decides to use certain pitch-classes, and then decides to repeat one of them.) It is possible, however, that a more successful model could be developed based on the "weighted-input" idea. One way would be to assume that a musical surface is generated from a sparser, "reduced" representation of pitches, something

like a middleground representation in a Schenkerian analysis. In such a representation, immediate repetitions of pitches such as those in Figure 4 (and also perhaps octave doublings and the like) would be removed. Possibly, a "weighted-input" model applied to such a reduction would produce better results; it would also avoid the arbitrary segmentation required by the CBMS and Bayesian models. Such an approach would present serious methodological problems, however, since it would require the middleground representation to be derived before key-finding could take place.

While it has definite room for improvement, the Bayesian key-finding model proposed here performs well enough to deserve serious consideration as a model of human key-finding. Of course, having a computational model that performs a process well does not prove that humans perform the process the same way. One way of evaluating a computational cognitive model is by considering its implications and explanatory value with respect to other aspects of cognition, beyond the problem it was originally intended to solve. This will be my aim in the remaining sections of this paper.

Estimating the probability of musical surfaces

In terms of their key-finding performance, the differences between the three models presented earlier are not large (when the same key-profiles are used for all models). In several other respects, however, the Bayesian model has important advantages over both the K-S and CBMS models[11]. For one thing, the Bayesian model provides a very natural way of measuring the probability of actual note patterns — what I have called musical "surfaces". Returning to the earlier presentation of Bayesian theory, the probability of a surface occurring in combination with a structure is p(surface | structure) p(structure) (see equation 5); the total probability of the surface occurring is this quantity summed over all possible structures (equation 6). In terms of the current model, the probability of a certain pitch-class set occurring within a segment in a tonal piece is its probability in combination with a certain key, summed over all keys (k). (Assume that this is the first segment of a piece, so there is no modulation score; each key has a probability of 1/24.)

$$\text{probability of pitch-class set} = \sum_{k} (1/24) \left(\prod_{p} S_{pc} \prod_{\neg p} S_{\neg pc}\right) \qquad (12)$$

Table 3 presents this data for certain well-known pitch-class sets. The table shows, first, the prior probability of each set. It can be seen that, among three-pc sets, the

(11) This section and the following one build on ideas put forth in Temperley (2001) concerning the use of preference rule systems to characterize musical tension, "tonalness", ambiguity, expectation, styles, and stylistic differences. However, the Bayesian framework makes possible a more rational and effective solution to these problems than what was proposed there.

major and minor triad have a higher probability than the diminished triad, which in turn has higher probability than C-C♯-D. No doubt this is because the major and minor triad are highly probable in combination with certain keys, *i.e.* keys in which they are diatonic triads (and the tonic triad in particular). By contrast, the set C-C♯-D is not particularly probable given *any* key, as it will always contain at least one chromatic note. This is made clear by the second column, which shows the probability of each set in combination with its most probable key. Notice that the prior probability figures in Table 3 tell us only the probability of one particular transposition of each set: *e.g.*, C-E-G. The probability of a T_n set-*class* — *e.g.*, major triads in general — would be given by the probability of one form, multiplied by the number of distinct transpositions; this is shown in the rightmost column of Table 3. In the case of the major triad, the total probability is .00173 x 12; in the case of the augmented triad, it is .00079 x 4, since there are only four distinct augmented triads. Among scale collections, the seven-note diatonic scale is much more probable than the six-note whole-tone scale, the eight-note octatonic scale, or the chromatic heptachord [0123456]. These distinctions hold true whether one considers the probability of a single transposition of the set or the probability of all transpositions.

Essentially, the numbers in Table 3 tell us the probability of different pitch-class sets occurring (within a short span of time) in a piece — specifically, a *tonal* piece, a piece using the musical language from which the current key-profiles were generated. To put it a slightly different way, they tell us how characteristic each pitch-class set is of the language of common-practice tonality — how tonal the set is, one might say. Certainly, we are capable of making such judgments as listeners. If we turn on the radio and hear a diatonic scale, we are likely to suspect that the piece is tonal; if we hear [0123456], our estimate of that probability will be significantly less (though one must also take into account the extremely low *prior* probability of hearing a non-tonal piece on the radio!). Of course, this would depend on the way the pitches were arranged; no doubt the set [0123456] could be compositionally realized in such a way as to sound unproblematically tonal. There is more to tonality, and judgments of tonality, than sheer pitch-class content. But the pitch-class content of a passage surely *contributes* to its tonalness, and this aspect of tonality appears to be captured rather well by the Bayesian model[12].

(12) It should be noted that neither the K-S model nor the CBMS model appears to yield judgments of the probability of a surface in any straightforward way. In the Bayesian model, the probability of a segment can be measured (approximately) by the model's score for the highest-scoring key, or (precisely) by the sum of scores for all keys. Neither of these measures appears to be very meaningful for either the K-S or the CBMS model. In the CBMS model, the score for a particular key increases with the number of pitch-classes in the segment (no scores are factored in for pitch-classes not present in the segment); if scores were construed as probabilities, the most probable segment would be the one containing all twelve pitch-classes, which is clearly incorrect. In the K-S model, the problem is that the input vector values — representing the duration of each pitch-

Table 3
Probabilities for certain pitch-class sets as estimated by the Bayesian key-finding model

Pitch-class set	Total ("prior") probability of set	Probability of set combined with most probable key	Number of distinct transpositions of set in T_n set-class	Total probability of T_n set-class
C-E-G (major triad)	0.00173	0.00103 (C)	12	.02080
C-Eb-G (minor triad)	0.00178	0.00098 (Cm)	12	.02137
C-Eb-Gb (dim.triad)	0.00031	0.00004 (Bbm)	12	.00382
C-E-G♯ (aug.triad)	0.00079	0.00019 (Dbm*)	4	.00318
C-C♯-D (012)	0.00022	0.00002 (Gm)	12	.00261
C-D-E-F♯-Ab-Bb (whole-tone set)	0.00001	0.000001 (Gm*)	2	.00002
C-D-E-F-G-A-B (diatonic set)	0.00049	0.00032 (C)	12	.00590
C-D-Eb-F-Gb-Ab-A-B (octatonic set)	0.000004	0.0000007 (Ebm*)	3	.00001
C-C♯-D-Eb-E-F-F♯ (0123456)	0.000006	0.000001 (Db)	12	.00007
total aggregate	0.00000003	0.000000002 (Cm*)	1	.000000003

* Symmetrical sets (such as the augmented triad, whole-tone scale, octatonic scale, and aggregate) yield equal probability judgments for multiple keys; in this case the model makes an arbitrary decision.

One might wonder if this approach could be applied to longer musical passages. To estimate the probability of a note pattern spanning multiple segments, one

class in a passage — are normalized to have a variance of 1. Consider a hypothetical passage, passage A, in which all pitch-classes of the C major scale are used equally often, and no others are used at all; consider also passage B, in which all twelve pitch-classes are used, but pitch-classes within the C major scale are used very slightly (say 1%) more often than chromatic ones. Passage A is clearly more tonal (and more probable) than Passage B, but due to the normalization of the input vector values, the two passages would be treated as equivalent by the K-S model. The K-S model may sometimes yield lower key scores (correlation values) for less tonal pieces; indeed, Krumhansl (1990) found this to be true in an analysis of a Schönberg piece. But this is not a reliable measure, for the reason just stated.

must calculate its probability in combination with all possible analyses of the passage (all ways of combining segment analyses) all summed together[13]. Such calculations do not emerge very naturally out of the framework of the current model, nor does it seem very likely that human listeners perform them. An alternative approach would be to take the probability of a passage in combination with its most likely analysis as representative of the prior probability of the passage. This latter measure might actually be a fairly close approximation to the actual probability, especially if the most likely analysis is far more probable than any other.

With this assumption in mind, consider the kinds of passages that would be judged as probable by the model. A high-scoring passage would be one for which the value p(surface | structure) p(structure) is high, for some structure. For p(surface | structure) to be high, there must be some sequence of keys such that the key of each segment is relatively compatible with the pitches of the segment. For p(structure) to be high, the number of modulations must be relatively low. Consider an excerpt consisting of an alternating pattern of C major and G major triads (each one occupying a segment). In this case, a key analysis which maintains C major throughout allows good compatibility between the key of each segment and the pitches, and also avoids modulations. Now imagine a series of segments consisting of alternating C major and F♯ major triads. In this case, the model would either have to incorporate all of the segments within a single key, such as C major, in which case the F♯ major segments would have a low probability, or it would have to alternate keys at each segment, in which case the probability of the structure would be low. Neither of these analyses would be especially high-scoring, thus the probability of the passage as a whole would be judged as relatively low. A passage which contained many chromatic pitch-class sets — so that no compatible key could be found even for individual segments — would be assigned a low probability also.

A model such as this might yield revealing judgments, not only of the "tonalness" of an entire piece, but of fluctuations in tonalness within pieces. Consider Figure 6, the first movement (excluding the six-measure introduction) of Schumann's *Papillons*. The model's preferred analysis here is to retain D major throughout. The model's judgment of the probability of each segment of the passage (treating measures as segments), in combination with its preferred analysis, is shown in Figure 7 (note that in this case a logarithmic scale is used). Relatively speaking, the first eight measures and last four measures are quite probable; however, the intense chromaticism of the third four-measure phrase leads to much lower probability values. That is to say, the probability of the pitch-class sets in mm. 15-18, given D major, is relatively low, and no other more preferable analysis is available. The model could also have chosen to modulate to A♭ major in mm. 15-16 and then back to D major, but this would have carried low probability as well due to the two modulations in quick succession[14].

(13) This approach is analogous to that taken in computational linguistics. For example, the probability of a certain word sequence can be estimated by summing the probabilities of that

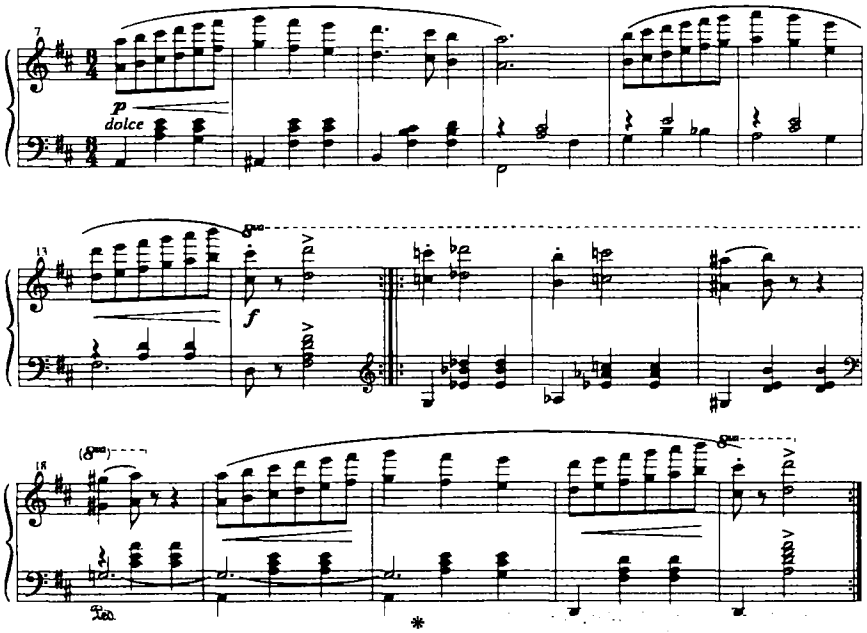


Figure 6. Schumann, Papillons, *I, mm. 7-22.*

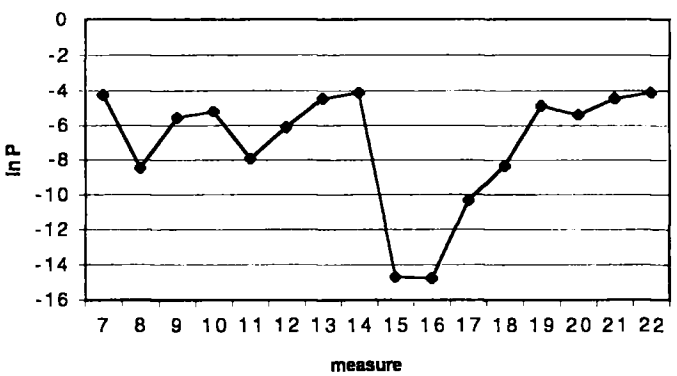


Figure 7.

The model's probabilistic analysis of the first movement of Schumann's Papillons *(the score of which is shown in Figure 6). The graph shows, for each measure, the probability of the pitch-class set in combination with its most likely key analysis.*

(sequence arising from all different possible syntactic structures (Charniak, 1994, p. 75).

(14) Such a model should really take into account that some key transitions are less likely than others, as already suggested with regard to key-finding. This would have the desirable consequence that a move from D major to Ab major would be rated as less probable — hence, more "tense" — than a move from D major to A major.

I submit that these probabilistic scores relate in an interesting way to musical perception and experience — though it is not exactly clear what dimension of musical experience they correspond to most closely. I suggested earlier that they have something to do with judgments of "tonalness": a kind of grammaticality or normality within the common-practice language. They also have something to do with tension: a passage of low probability is likely to have an effect of tension and instability. This may be because it suggests a failure in communication: if the passage (as we analyze it) is improbable, this suggests that we may be misanalyzing it — misunderstanding the structure, or even misperceiving the notes. (It may also indicate that our model is incorrect — the key-profiles we were assuming are not the correct ones.) Of course — to reiterate my earlier caveat — the musical tension of a passage is more than just a simple function of its pitch-class content; all kinds of harmonic, melodic, and rhythmic factors undoubtedly play a role. Yet pitch-class content surely plays *some* role in tension; generally speaking, passages with lots of chromaticism and rapid modulations tend to sound tense and unstable (at least in a common-practice context), no matter how the pitches are arranged. What I am suggesting is that this aspect of the tension of a passage appears to correspond well with its probability, as judged by the current model. We should note also that this proposed correlate to tension is simply an emergent feature of a model that was proposed for quite a different purpose: modeling judgments of key. The tension of a passage, under the current hypothesis, is given by the maximal value of p(surface | structure)p(structure); but this must be computed anyway (if the Bayesian key-finding model is correct), for all possible structures, in order to find the preferred structure.

A further aspect of musical experience that is modeled quite naturally by the current model is ambiguity. In some cases, a passage may be ambiguous with regard to key: two or more keys may be roughly equal in probability. Note that this is a separate issue from the probability of the passage itself. A passage may be highly probable in terms of its pitch-class content, but still somewhat ambiguous with regard to key. (It is also possible in theory for a passage to be highly improbable, yet clear in terms of key; this is more difficult to imagine, since the kinds of phenomena which lead to low-probability surfaces — chromaticism and rapid modulation — also tend to obscure the key.) Consider a passage such as the opening of Chopin's Mazurka Op. 24 No. 2 (Figure 8). The passage is fully diatonic with respect to two keys, C major and G major, and therefore would have a fairly high probability of occurring given either key; thus the probability of the passage itself should be high. However, it is unclear to the model (and also, I would suggest, to the listener) which key is correct. This is reflected simply in the fact that two analyses of the passage — one with C major throughout and the other with G major throughout — are roughly equal in probability. Another frequent site of key ambiguity is pivot chords — segments occurring at the boundary between two key sections, and compatible with either the previous key or the following one. In the current framework, these

would (or at least should) be analyzed as segments in which the previous and following keys are roughly equal in probability.

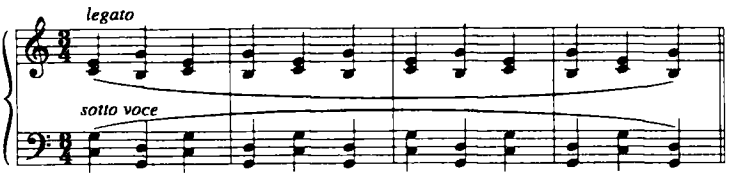


Figure 8. Chopin, Mazurka Op. 24 No. 2, mm. 1-4.

The Bayesian perspective on key perception is also relevant to expectation. I have suggested that the estimation of probabilities of musical surfaces is an important part of musical experience, allowing us to judge the "tonalness" of musical passages and perhaps affecting our understanding of musical tension as well. But in hearing a piece, we are not only analyzing what we have heard but also making predictions of what will happen next. The same probabilistic model that is used to judge the probability of heard music could also be used to generate predictions of future events: the expected event in a melody is, presumably, the one that is judged most probable[15]. In terms of key, the Bayesian model predicts that we would expect a continuation that is highly probable given the current key (we normally expect the current key to be continued, since key changes are improbable) — roughly speaking, one adhering to the current diatonic scale. This seems plausible; in hearing a melody, for example, we generally expect the next note to remain within the currently-established scale, though certainly other kinds of constraints are involved as well. This has also been demonstrated experimentally; Schmuckler (1989) found that expected melodic continuations were highly correlated with pitch stability as predicted by the Krumhansl-Kessler profiles.

Modeling other kinds of musical structure

It was noted above that the Bayesian model of key-finding presented here has a strong resemblance to the model of key-finding put forth in Temperley (2001) — what I have called the CBMS model. The CBMS model was originally presented as a preference rule system. A preference rule system is a model involving a set of criteria or "preference rules"; many analyses are considered, and the one is chosen which best satisfies the rules. Preference rule models have been proposed for a

(15) Here again, there are questions about how exactly this would be quantified. Strictly speaking, the probability of a surface continuation would be given by its probability summed over all possible analyses of the current surface combined with all possible analyses of the prior context, but this may be neither computationally feasible nor psychologically plausible.

variety of aspects of musical perception, including metrical analysis, grouping analysis, pitch reduction, harmonic analysis, and stream segregation (Lerdahl and Jackendoff, 1983, Temperley, 2001). In the case of the CBMS key-finding model, just two rules are involved:

- Key-Profile Rule. Prefer to choose a key for each segment which is compatible with the pitches of the segment (according to the key-profiles).

- Modulation Rule: Prefer to minimize the number of key changes.

It appears, in fact, that preference rule models generally have a close connection to Bayesian models. In the approach of Temperley (2001), preference rule systems are quantified by having each preference rule assign numerical scores to each analysis indicating how "good" it is; these scores are summed, with the highest-scoring analysis overall being the preferred one. This has much in common with the scoring process proposed earlier for the Bayesian key-finding model, where the score for an analysis is the sum of terms representing logarithms of probabilities. This is not to say that the models in Temperley (2001) could be construed, exactly as they are, as Bayesian models; some modifications would be needed in every case. But there is certainly a strong affinity between the two approaches. One insight yielded by a Bayesian view of preference rule systems is that preference rules really fall into two categories. Some rules relate to the probability of a certain structure; we could call these "structure rules". Others relate to the probability of a surface given a structure; we could call these "structure-to-surface rules". In the case of the CBMS key-finding model, the Modulation Rule is a structure rule; the Key-Profile Rule is a structure-to-surface rule.

Another situation where Bayesian modeling appears to apply very naturally is metrical analysis. A preference rule system for metrical analysis was proposed in Temperley (2001) (see also Temperley and Sleator, 1999), building on the earlier model of Lerdahl and Jackendoff (1983). In this case, the structure is a row of beats (or a framework of levels of beats, but we will consider just a single metrical level for now), and the surface is once again a pattern of notes. The model involves three main rules:

- Event Rule: Prefer for beats to coincide with event-onsets.

- Length Rule: Prefer for beats to coincide with longer events.

- Regularity Rule: Prefer for beats to be roughly evenly spaced.

The process of deriving a row of beats involves optimizing over these three rules: choosing the metrical level which aligns beats with as many events as possible, especially long events, while maximizing the regularity of beats. The model in Temperley (2001) evaluates a possible analysis by assigning it scores from each of these three rules and summing these scores. It can be seen how a model of this kind could be reconstrued as a Bayesian model, much as we have reinterpreted the CBMS key-finding model in Bayesian terms. In this case, then, the Regularity Rule is

a structure rule, indicating the probability of structures (more regular structures are more probable); the Event Rule and Length Rule are structure-to-surface rules, indicating the probability of surface patterns given a certain structure (patterns are more probable that align events with beats, especially longer events). Such a model could incorporate multi-leveled metrical structures as well, under the assumption that events on lower-level beats are less probable than events on higher-level beats.

Another Bayesian model of meter-finding — superficially different from the CBMS model, but fundamentally similar — has been proposed by Cemgil *et al.* (2000a, 2000b). In this model, the problem is defined as the recovery of a score from a performance; a score is a representation of events in terms of integer values, essentially indicating their positions in some kind of metrical grid, and a performance is an actual pattern of events in time. The most likely score given a performance is the one maximizing the expression p(score) p(performance | score). In this case, the goodness of the rhythmic pattern itself (the alignment of events with strong beats, etc.) is reflected in p(score); the goodness of the realization of the pattern (*i.e.* the regularity of the actual performed beat) is reflected in p(performance | score). In Cemgil's model, the complexity of a score is judged by its "depth" — essentially, the number of metrical levels required to notate it; a notation requiring sixteenth-notes is deeper, hence less probable, than one requiring only quarter-notes.

A further area where the Bayesian perspective seems relevant is what might be called "melodic structure" — the principles whereby melodies are constructed and perceived. At issue here are not harmonic and tonal principles, but rather matters such as melodic shape, range, and interval size. A good deal of work has been done on the principles governing melodic expectation, leading to quite powerful quantitative models of experimental data (Narmour, 1990; Schellenberg, 1997). Meanwhile, other research has focused on statistical analysis of musical corpora, seeking to identify regularities in the way melodies are constructed (von Hippel and Huron, 2000; von Hippel, 2000). Both in perceptual and compositional research, an important principle that emerges is pitch proximity: the compositional preponderance of, and perceptual expectation for, small melodic intervals as opposed to large ones. As argued in Temperley (2001) (following Bregman [1990] and other psychological research), the principle of pitch proximity is also an important aspect of what I have called "contrapuntal analysis" — the process of grouping notes into contrapuntal lines: we prefer to group notes into lines such that intervals within lines are small. (In this case the surface is, as usual, a pattern of notes; the structure is a pattern of lines implied by those notes.) Here, then, is a situation analogous to what was observed in key-finding and meter-finding: a single probabilistic principle — the preference for small intervals within melodic lines — is reflected in composition (the preponderance of small intervals in musical corpora), perception (the tendency to group notes into lines such that

intervals are small), and expectation (the higher expectation for small intervals as opposed to large ones)[16].

An intriguing prospect suggested by the Bayesian approach is the possibility of quantifying differences in musical structure across styles. One example was suggested in Temperley (2001), relating to metrical structure: the complementary relationship between rubato and syncopation. In general, styles allowing a lot of fluctuation in tempo — such as common-practice music, particularly of the Romantic period — tend to have little syncopation; in styles allowing great syncopation, such as traditional sub-Saharan African music and rock, the pulse tends to be extremely regular. These differences could be quantified in terms of the parameters of a Bayesian model. The tolerance for rubato in a style would be reflected in the degree to which irregular beat patterns were assigned lower probabilities; the tolerance for syncopation would be reflected in the degree to which syncopated note patterns (given a certain beat pattern) were assigned lower probabilities. A Bayesian model of this kind might allow for the quantitative characterization or categorization of actual musical surfaces; if a musical surface was assigned a higher probability estimate by the common-practice model as opposed to the rock model, this would mean it was more likely to be a common-practice piece. It would also raise questions about perception: if different styles reflect different probabilistic parameters (the tolerance for rubato and the tolerance for syncopation), are these differences assimilated by listeners as well, so that African and Western listeners might possess different perceptual models and thus might assign different analyses to the same piece? Finally, this perspective raises the possibility that there may be systematic relationships between different aspects of musical styles. It is surely no coincidence that styles with rubato generally tend to have low syncopation; if a piece had both high rubato and high syncopation, inference of the beat might be impossible. The need to convey certain kinds of information, coupled with the probabilistic nature of the perceptual process, may act as an important constraint on the evolution of musical styles.

Conclusions

In this essay I have pointed to a number of reasons why a Bayesian approach to musical modeling is attractive and promising. The Bayesian approach embraces, or at least connects strongly with, a good deal of work that has already been done within the preference-rule framework towards modeling the perception of key, meter, and other aspects of musical structure, yet it also provides the preference-rule

(16) Also worthy of mention here is Bod's model of melodic segmentation (2002). Bod's model is based on statistical data regarding pitch patterns within phrases and the number of phrases in a melody; this data is used to select the most probable phrase structure given a melody. This could be regarded as a simple Bayesian model, though Bod does not present it in those terms.

approach with a more rational foundation than has been available before. Bayesian models provide the basis for a quantitative and systematic, yet musically informed, investigation into the probabilistic aspect of musical communication — as called for by Meyer in the 1950's. The approach leads to interesting ways of modeling musical tension, tonalness, ambiguity, and expectation, as well as describing stylistic differences (both in music and in listeners). Equally significant, this approach establishes an important connection between music cognition and other branches of cognitive science where Bayesian modeling is used (such as computational linguistics) — a connection which in turn holds out exciting prospects for the sharing of ideas between fields and the identification of cross-domain generalities about cognition[17].

(17) **Address for correspondence:**
David Temperley
Eastman School of Music
26 Gibbs St.
Rochester, NY, 14607, USA
tel.: 585-274-1557
e-mail: dtemp@theory.esm.rochester.edu

• REFERENCES

Alamkan, C., Birmingham, W. P., & Simoni, M. H. (1999). *Stylistic structures: An initial investigation of the stochastic genration of tonal music.* University of Michigan Technical Report CSE-TR-395-99.

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

Bod, R. (2002). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research, 31,* 27-37.

Bregman, A. S. (1990). *Auditory Scene Analysis.* Cambridge, MA: MIT Press.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception, 6,* 219-42.

Cemgil, A. T., Desain, P., & Kappen, B. (2000a). Rhythm quantization for transcription. *Computer Music Journal, 24*(2), 60-76.

Cemgil, A. T., Kappen, B., Desain, P., & Honing, H. (2000b). On Tempo tracking: Tempogram representation and Kalman filtering. *Journal of New Music Research, 29,* 259-73.

Charniak, E. (1996). *Statistical Language Learning.* Cambridge, MA: MIT Press.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science, 7,* 137-63.

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

Hiller, L. A., & Fuller, R. (1967). Structure and information in Webern's Symphonie, Opus 21. *Journal of Music Theory, 11,* 60-115.

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

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12,* 154-71.

Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing.* Upper Saddle River, NJ: Prentice-Hall.

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

Kostka, S., & Payne, D. (1995b). *Workbook for Tonal Harmony.* New York: McGraw–Hill.

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

Leman, M. (1995). *Music and Schema Theory.* Berlin: Springer.

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

Longuet–Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. *Machine Intelligence, 6,* 221–41.

Manning, C. D., & Schütze, H. (2000). *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press.

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

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

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research, 57,* 103-18.

Ponsford, D., Wiggins, G., & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research, 28*, 150-77.

Raphael, C. (2002). Automatic Transcription of Piano Music. In M. Fingerhut (ed.), *Proceedings of the 3rd Annual International Symposium on Music Information Retrieval* (pp. 15-9). Paris : IRCAM-Centre Pompidou.

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

Schmuckler, M. (1989). Expectation and music: Investigation of melodic and harmonic processes. *Music Perception, 7*, 109–50.

Smith, N., & Schmuckler, M. (2000). Pitch-distributional effects on the perception of tonality. In C. Woods, B. B. Luck, R. Brochard, S. A. O'Neil and J. A. Sloboda (eds), *Proceedings of the Sixth International Conference on Music Perception and Cognition* (CD). Keele, UK: Department of Psychology.

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

Temperley, D. (2002). A Bayesian approach to key-finding. In C. Anagnostopoulou, M. Ferrand, and A. Smaill (eds), *Music and Artificial Intelligence* (pp. 195-206). Berlin: Springer.

Temperley, D., & Sleator, D. (1999). Modeling meter and harmony: A preference-rule approach. *Computer Music Journal, 23(1)*, 10-27.

Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception, 32*, 741-66.

von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception, 17*, 315-27.

von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception, 18*, 59–85.

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

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

• Modelos Bayesianos de la estructura y cognición musical

Esta comunicación revisa la aplicación del modelo probabilístico Bayesiano a las cuestiones de cognición y teoría musicales. La cuestión principal tiene que ver con el problema de encontrar la nota principal: el proceso de deducción de la nota principal de un motivo. La perspectiva bayesiana conduce a un modelo de este proceso simple, elegante y altamente efectivo; la misma aproximación se puede aplicar a otros aspectos de la percepción musical, como las estructuras métrica y melódica. Los modelos bayesianos aportan también soluciones interesantes para otras cuestiones musicales, incluyendo tensión musical, ambigüedad, expectación, y las descripciones cuantitativas de estilos y diferencias estilísticas.

• Modelli bayesiani di struttura e cognizione musicale

Il presente articolo studia l'applicazione del modello probabilistico bayesiano a questioni di cognizione e teoria musicale. Esso si concentra sul problema di "trovare la tonalità", vale a dire il processo mediante il quale si stabilisce la tonalità a partire da una sequenza di note. La prospettiva bayesiana conduce ad un modello semplice, elegante ed assai efficace per tale processo; lo stesso modello si può estendere anche ad altri aspetti della percezione musicale, quali la struttura metrica e la struttura melodica. La modellazione bayesiana si associa altresì in modo interessante a molti altri aspetti della musica, ivi comprese tensione, ambiguità, attese musicali, nonché la descrizione quantitativa di stili e differenze stilistiche.

• Modèles bayésiens de la structure et de la cognition musicales

On étudie ici l'application de la modélisation probabiliste de Bayes aux domaines de la cognition et de la théorie musicales. Celle-ci s'avère surtout intéressante dans l'établissement de la tonalité : le processus par lequel la tonalité est inférée à partir d'un pattern de notes. Elle conduit à un modèle simple, élégant et extrêmement efficace de ce processus; la même approche peut être étendue à d'autres aspects de la perception musicale, par exemple les structures métrique et mélodique. Enfin, elle se révèle très intéressante aussi s'agissant de notions comme la tension, l'ambiguïté, l'attente et la description quantitative des styles et des différences stylistiques.

• Bayes-Modellierung musikalischer Struktur und Kognition

In diesem Beitrag werden Wahrscheinlichkeitsmodellierungen nach Bayes auf ihre Anwendbarkeit für Musikkognition und Musiktheorie hin überprüft. Das Hauptinteresse ist die Erkennung von Tonarten, also der Prozess, aus verschiedenen Tönen eine Tonart zu ermitteln. Der bayes'sche Ansatz führt zu einem einfachen,

eleganten und sinnvollen Model dieses Prozesses. Der gleiche Ansatz kann auch auf andere Aspekte der musikalischen Wahrnehmung erweitert werden, wie beispielsweise auf die metrische und die melodische Struktur. Die Modellierung verweist ebenfalls auf andere musikalische Themen wie musikalische Spannung, *Doppeldeutigkeiten, Erwartungen sowie quantitative Beschreibungen von Stilen* und stilistischen Unterschieden.