A BAYESIAN KEY-FINDING MODEL

David Temperley

Eastman School of Music
University of Rochester
26 Gibbs St.
Rochester, NY 14604
dtemp@mrl.esm.rochester.edu

ABSTRACT

This model determines the key of a piece (represented as a MIDI file) using "key-profiles", in combination with Bayesian principles. The model divides the piece into short segments; it then searches for the most probable "key structure", where a key structure is a labeling of each segment with a key. The probability of a key structure is a function of the number of modulations in the structure, and the fit of the key of each segment to the pitch-classes in the segment. The main key for the piece is then defined by the key of the first of segment of this structure.

Keywords: key-finding, Bayesian modeling.

1 INTRODUCTION

In this abstract I present a model for key-finding in polyphonic music, using symbolic (MIDI) input. The model uses "key-profiles"—an ideal pitch-class distribution for each key—in combination with Bayesian principles. It is essentially the model presented in [1] (though with different parameters) and [2].

2 THE MODEL

Given a MIDI file as input, the model begins by creating a list of notes with the on-time, off-time and pitch of each note. It then divides the piece into segments of 1.2 seconds in length. (In [1] and [2], the segments were defined by the metrical structure; in this case, no metrical information is available.)

The model then searches for the optimal "key structure", where a key structure is a labeling of each segment with a key. Here we use Bayesian reasoning. We are confronted by a pattern of pitches—a "surface"—and we want to know the underlying key structure that gave rise to it—the "structure". Bayes' rule tells us that

$$P(\text{structure} \mid \text{surface}) \propto P(\text{surface} \mid \text{structure})\, P(\text{structure}) \tag{1}$$

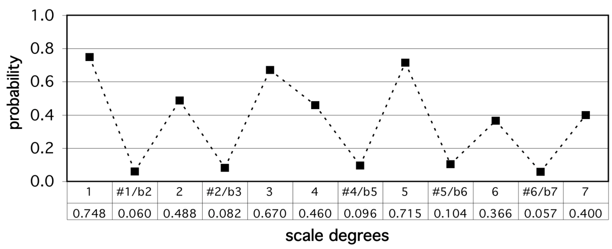
The structure maximizing the right side of this expression will be the one maximizing the left side—thus telling us the most probable key structure given the surface.

The prior probability of a key structure—a set of key labels assigned to segments—is defined as follows:

$$P(\text{structure}) = \prod_{s} M \tag{2}$$

where s indicates segments and M is a modulation score. For the first segment, M = 1/24 for every key. For each subsequent segment, M = .998 if the key of the segment is the same as the previous one; if not, M = .002 / 23. (These parameters were arrived at through trial-and-error testing.)

We now consider the probability of a surface given a structure. We represent the surface simply as a series of pitch-class sets, one in each segment. Let us assume that, in each segment, the composer makes twelve independent decisions as to whether or not to use each pitch-class, depending only on the key of the segment. (Thus we are concerned only with which pitch-classes are present in a segment, disregarding the number of events of that pitch-class in the segment or their duration.) These probabilities are represented in "key-profiles" [3]. The key-profiles used in the model are shown in Figure 1. The profiles show scale-degrees (pitch-classes in relation to a key); for example, in C major, C is 1 and F# is #4. The key-profiles were derived empirically, using a corpus of excerpts taken from the Kostka and Payne music theory textbook [4], in which keys are explicitly marked. The corpus was divided into segments and analyzed to find the proportion of all major-key segments containing each scale-degree; this yields the value for each scale-degree in the major-key profile. For example, the profile shows that .748 of segments contain scale-degree 1. The process was then repeated for minor-key segments.



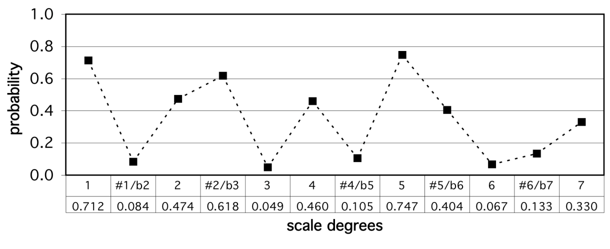


Figure 1. Key-profiles for major keys (above) and minor keys (below).

The key-profiles thus indicate the probability of each pitch-class occurring in a segment of a given key. The probability of a pitch-class *not* occurring in a segment is just 1 minus this value. We can then define the probability of a set of pitch-classes occurring in a segment given a key (we call this a "key-segment score"):

$$\text{key-segment score} = \prod_{p} K_{pc} \prod_{\sim p}(1-K_{pc})) \tag{3}$$

where p refers to pitch-classes present in each segment, $\sim p$ refers to pitch-classes not present, and K_{pc} is the key-profile value. The probability of a series of pitch-class sets, given a series of key labels for those segments, is then the product of the key-segment scores over all segments. This gives us a measure of the probability of the surface given the structure:

$$P(\text{surface} \mid \text{structure}) = \prod_{s} (\prod_{p} K_{pc} \prod_{\sim p}(1-K_{pc})) \tag{4}$$

The probability of a key structure given a surface is then proportional to P(structure) × P(surface | structure), as defined above.

Given this formula, the model then searches for the most probable key structure using dynamic programming. Briefly, we go through the segments from left to right; at each segment S_n, for each key K, we calculate the "best-so-far" key structure ending with that key at S_n; this can be done by adding S_n (with key K) on to each "best-so-far" analysis at segment S_{n-1} (already calculated). At the end of the piece, one key has the best score overall, and that can be traced back to yield the best analysis.

Given the optimal key structure (a labeling of each segment with a key), how then does the model compute the main key of the entire piece (bearing in mind that it is only given a short excerpt from the beginning of the piece)? This is a question I have not considered in previous research. After trying several different methods on the MIREX training data, it seemed that the best method was simply to choose the key of the first segment in the optimal key structure, as defined above. It is important to note that this is *not* the same as simply analyzing the first 1.2-second segment in isolation. Rather, the entire piece is analyzed, with a preference for both minimizing key changes begtween segments and optimizing the fit of each key to the pitches of each segment; the main key is defined by the key of the first segment in the resulting global structure.

One further refinement seemed desirable: In the analysis process described above, the initial segment of the piece is divided into four segments of 0.3 seconds in length; in effect this gives the first 1.2-second segment of the piece extra weight in determining the key.

3 RESULTS

On the MIREX 2005 Evaluation, the Bayesian model described above was tested on its ability to identify the correct key of 1252 excerpts from classical pieces of 10, 20, or 30 seconds in length, represented as midifiles. The model labeled 1127 out of 1252 excerpts correctly. Using a metric in which partial credit was given for incorrect guesses, the system received a score of 91.4%.

On a "B 0" computer (see the MIREX hardware specifications), the system ran the 1252 excerpts in a total time of 91 seconds, or .07 seconds per excerpt.

4 CONNECTIONS WITH OTHER WORK

The idea of using key-profiles for key-finding was first proposed by Krumhansl and Schmuckler [3]. Their model creates an input vector representing the total duration of each pitch-class in a passage, correlates this with key-profiles for each key (derived from experimental data), and chooses the key with the highest correlation. In [5], I proposed a modified version of this model, which divides the piece into short segments, represents each pitch-class as present or absent within each segment, calculates the match between input vector and key-profile as the scalar product of the two vectors, and factors in a modulation penalty for key changes. The model presented in [1] is very similar to this, but the Bayesian reasoning proposed there leads to a slightly different mathematical formula (the one presented above). Finally, in [2], I proposed the empirical "Kostka-Payne" profiles used here.

REFERENCES

[1] D. Temperley. 2002. A Bayesian key-finding model. In C. Anagnostopoulou et al. (eds.), *Music and Artificial Intelligence*. Berlin: Springer.

[2] D. Temperley. Forthcoming. *Music and Probability*.

[3] C. Krumhansl. 1990. *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.

[4] S. Kostka & D. Payne. 1995. *Tonal Harmony.* New York: McGraw Hill.

[5] D. Temperley. 2001. *The Cognition of Basic Musical Structures.* Cambridge: MIT Press.