Relational motif discovery via graph spectral ranking

Alberto Pinto
Università degli Studi di Milano
Dipartimento di Informatica e Comunicazione
Via Comelico, 39/41
I-20135 Milano, Italy
pinto@dico.unimi.it

ABSTRACT
Music summarization aims at finding the most representative parts of a music piece (motifs) that can be exploited for efficient music indexing. Here we present a novel approach for motif discovery in music pieces based on a graph spectral ranking. Scores are segmented into a network graph of music segments and then ranked depending on their centrality. Different poli- and mono-phonic metric concepts can be adopted to compare music segments. Bars with higher centrality are more relevant for music summarization. We present an evaluation on the corpus of J.S.Bach’s 2-part Inventions both in poli- and mono-phonic configuration.

Categories and Subject Descriptors
G.2.2 [Discrete Mathematics]: Graph Theory; H.3.1 [Content Analysis and Indexing]: Indexing methods; I.2.1 [Artificial intelligence]: Applications and Expert Systems; I.5.4 [Pattern recognition]: Applications; J.5 [Arts and Humanities]: Performing Arts—Music

General Terms
Theory

Keywords
relational models, motif discovery, graph spectral ranking

1. INTRODUCTION
Listening to music and perceiving its structure is a relatively easy task for humans, even for listeners without formal musical training. However, building computational models to simulate this process is a hard problem. On the other hand, the problem of automatically identifying relevant characteristic motifs and efficiently store and retrieve the digital content has become an important issue as digital collections are increasing in number and size more or less everywhere.

Some proposed solutions are more focused on tonal music as they exploit the harmonic structures of a piece and voice leading. On the other hand, other methods are more general and do not take into account neither harmony nor rhythm.

Notwithstanding the conspicuousness of the literature, current approaches seem to rely just on repetitions [18] [7], assigning higher scores to recurring equivalent melodic and harmonic patterns [9]. Recently reported approaches to melodic clustering based on motivic topologies [16], graph distance [20] [19] and paradigmatic analysis [17] have been used to select relevant subsequences among highly repeated ones by heuristic criteria [13] [1].

Moreover, the “paradigm of repetition”, in order to be applied, needs by no means a precise definition of “varied repetition”, a concept not easy to define. Of course, it has to include standard music transformation, but it is very difficult to adopt a simple two-valued logic (this is a repetition and this is not) in this context, where a more fuzzy approach seems to better address such a problem.

Here we present a ranking method based on relations instead of repetitions. We show that a distance distribution on a graph of note subsequences induced by music similarity measures generates a ranking real eigenvector whose components reflect the actual relevance of motives. False positives of the repetition paradigm turned out to be less connected nodes of the graph due to their higher degree of dissimilarity with relevant motives.

Our results show how higher indexes of connection, or “centrality”, are more likely to perform better than higher repetition rates in motif discovery, with no additional assumptions on the particular nature of the sequence or the adopted similarity measure.

2. RELATED WORKS
Music segmentation is usually realized through musicological analysis by human experts and, at the moment, automatic segmentation is a difficult task without human intervention. The supposed music themes have often to undergo a hand-made musicological evaluation, aimed at recognizing their expected relevance and completeness of results. As a matter of fact, an automatic process could extract a musical theme which is too long, or too short, or simply irrelevant. That is why a human feedback is still required in order to obtain high-quality results.

We present here an overview of current approaches based on different musical assumptions. We start this Section with a general overview of the literature. Then we introduce harmony related approaches, with a focus to reductionistic ones.
Finally we introduce topology-based models, which share much more similarities than others with our approach.

2.1 General approaches

Lartillot [13] [14] defined a musical pattern discovery system motivated by human listening strategies. Pitch intervals are used together with duration ratios to recognize identical or similar note pairs, which in turn are combined to construct similar patterns. Pattern selection is guided by paradigmatic aspects and overlaps of segments are allowed.

Cambouropoulos [5], on the other hand, proposed methods to divide given musical pieces into mostly non-overlapping segments. A prominence value is calculated for each melody based on the number of exact occurrences of non-overlapping melodies. Prominence values of melodies are used to determine the boundaries of the segments [6]. He also developed methods to recognize variations of filling and thinning (through note insertion and deletion) into the original melody. Cambouropoulos and Widmer [8] proposed methods to construct melodic clusters depending on the melodic and rhythmic features of the given segments. Basically, similarities of these features up to a particular threshold are used to determine the clusters. High computational costs of this method make applications to long pieces difficult.

2.2 Tonal harmony-based approaches

Tonal harmony based approaches exploit particular harmonic patterns (such as tonic-subdominant-dominant-tonic), melodic movements (e.g. sensible-tonic), and some rhythmic punctuation features (pauses, long-duration notes, ...) for a definition of a commonly accepted semantic in many ages and cultures.

These approaches typically lead towards score reductions (see Figure 1), made possible by taking advantage of additional musicological information related to the piece and assigning different level of relevance to the notes of a melody. For example one may choose to assign higher importance to the stressed notes inside a bar [21]. In other words, the goal of comparing two melodic sequences is achieved by reducing musical information into some “primitive types” and comparing the reduced fragments by means of suitable metrics.

![Figure 1: J.S. Bach, BWV 1080: Score reductions.](image)

A very interesting reductionistic approach to music analysis has been attempted by Fred Lerdahl and Ray Jackendoff. Lerdahl and Jackendoff [15] research was oriented towards a formal description of the musical intuitions of a listener who is experienced in a musical idiom. Their purpose was the development of a formal grammar which could be used to analyze any tonal composition.

The study of these mechanisms allows the construction of a grammar able to describe the fundamental rules followed by human mind in the recognition of the underlying structures of a musical piece.

2.3 Topological approaches

Mazzola and Buteau [4] proposed a general theoretical framework for the paradigmatic analysis of the melodic structures. The main idea is that a paradigmatic approach can be turned into a topological approach. They consider not only consecutive tone sequences, but allow any subset of the ambient melody to carry a melodic shape (such as rigid shape, diastematic shape, etc.). The mathematical construction is very complex and, as for the motif selection process, it relies on the repetitions.

The method proposed by Adiloglu, Noll and Obermayer in [1] does not take into account the harmonic structure of a piece and is based just on similarities of melodies and on the concept of similarity neighborhood. Melodies are considered as pure pitch sequences, excluding rests and rhythmic information.

A monophonic piece is considered to be a single melody \( M \), i.e. they reduce the piece to its melodic surface. Similarly, a polyphonic piece is considered to be the list \( M = (M_i)_{i=1,...,N} \) of its voices \( M_i \). The next step is to model a number of different melodic transformations, such as transpositions, inversions and retrogradations and to provide an effective similarity measure based on cross-correlation between melodic fragments that takes into account these transformations. They utilize a mathematical distance measure to recognize melodic similarity and the equivalence classes that makes use of the concept of *neighbourhood* to define a set of similar melodies.

Following the repetition paradigm stated by Cambouropoulos in [6] they define a prominence value to each melody based on the number of occurrences, and on the length of the melody. The only difference is that they allow also melody overlapping. In the end, the significance of a melody \( m \) of length \( n \) within a given piece \( M \) is the normalized cardinality of the similarity neighbourhood set of the given melody. If two melodies appear equal number of times, the longer melody is more significant than the shorter one.

In [1] the complete collection of the Two-part Inventions by J. S. Bach is used to evaluate the method, and this will be also our choice in Section 4.

3. THE MODEL

As stated in Section 2, current methods rely on repetitions. Our point of view can be synthesized in the following points:

1. We consider a music piece as a network graph of segments,
2. we do take into account both melodic and rhythmical structures of segments
3. we do not consider harmony, as it is too much related to tonality.

A single frame may represent, for instance, a bar or a specific voice within a bar like in Fig. 2, but also more general segments of the piece. We do not take into account here the
the model. In fact by using undersized windows we normally get just more detailed results. In our experiments (see Section 4) we decided to adopt a one-bar length window, as we considered metric information relevant to music segmentation, avoiding any form of overlapping.

### 3.2 Metric weights

In this Section we are going to introduce the metric concepts we adopted to calculate similarities between different score windows. The variety of segmentations reflects to a large extent the variety of musical similarity concepts, nevertheless, as stated in Section 4, the model is rather robust respect to metric changes.

In general, we can just require that the set of segments can be endowed with a notion of distance

\[ d : S \times S \rightarrow \mathbb{R} \]

between pairs of segments and turns this set into a (possibly metric) space \((S, d)\). A natural choice for point sets of a metric space is the Hausdorff metric \([11]\) but any other distance discovered to be useful in music perception, like EMD/PTD \([23]\), can be chosen as well.

Here we assume \(d\) to be:

1. real,
2. non-negative,
3. symmetric and
4. such that \(d(s, s) = 0, \forall s \in S\)

As a matter of fact, most musically relevant perceptual distances do not satisfy all metric axioms \([23]\). Therefore no further property, like the identity of indiscernibles or the triangle inequality, is assumed.

Given two segments \(s_1\) and \(s_2\), the metrics we adopted in the experiments are the following:

\[
 d_1(s_1, s_2) = \sqrt{\sum_{|s|} |[s_1]_{12} - [s_2]_{12}|} \tag{1}
\]

\[
 d_2(s_1, s_2) = \sqrt{\sum_{|s|} (s'_1(t) - s'_2(t))^2} \tag{2}
\]

where \(s'\) is the derivative operator on the sequence \(s\), \(|s|\) is the length of \(s\) and \([s]_{12}\) is the sequence \(s\) where each entry has been chosen in the interval \([0, 11]\).

\(d_1\) is a first-order metric that takes into account just octave transpositions of melodies. In fact, pitch classes out of the range \([0, 11]\) are folded back into the same interval, so melodies which differ for one or more octaves belong to the same congruence class modulo 12 semitones. \(d_2\) is a second-order metric that takes into account arbitrary transpositions of a melody. No other assumptions on possible variations have been made, so that an equivalence class of melodies is composed just of transpositions and inversions of the same melody like in \([1]\).

Both distances can be applied to single voice sequences but also to multiple voice sequences, given that a suitable representation has been provided. For instance, in a two voice piece, with voices \(v_1\) and \(v_2\), one can consider the difference vector \(v = v_1 - v_2\) as a good representation of a

---

**Figure 2:** A representation of the (first-order) network of frames.

**Figure 3:** An 8-nodes complete graph.
specific segment, and then apply $d_1$ or $d_2$ to this new object. The advantage of using this differential representation is that it is invariant with respect to transpositions of the two voices so that, for instance, it makes also $d_1$ invariant with respect to transpositions, and not just to octave shifts.

By exploiting those distance concepts, it is possible to encode the edges of the complete graph with metric weights in order to compute the weights of nodes in terms of the main eigenvector, as we are going to show in the following Sections.

3.3 Matrix representation and ranking eigenvector

The adopted algebraic representation of the 'score graph' $K$ is the adjacency matrix $A(K)$. This is a nonnegative matrix as its entries are the distance values between the different segments in which the score has been divided into. Perron-Frobenius theory for nonnegative matrices grants the existence of an eigenvector $x \in \mathbb{R}^n$ with $x \geq 0$ and and $\sum_{i=1}^{n} x_i = 1$, called the Perron vector of $A$ [12].

This result has a natural interpretation in the theory of finite Markov chains, where it is the matrix-theoretic equivalent of the convergence of a finite Markov chain, formulated in terms of the transition matrix of the chain [2].

3.4 The algorithm

Let $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ denote a distance function on $\mathcal{S}$, like those defined in Section 3.2, which assigns each pair of segments $s_i$ and $s_j$ a distance $d(s_i, s_j)$. We can describe the algorithm through the following steps:

1. Form the distance matrix $A = [a_{i,j}]$ such that $a_{i,j} = d(s_i, s_j)$;
2. Form the affinity matrix $W = [w_{i,j}]$ defined by
   \[ w_{i,j} = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right) \]
   The parameter $\sigma$ can be chosen experimentally, a possible choice is the standard deviation of the similarity values within each segment (this has been our choice in the experimental part);
3. Compute the leading eigenvector $x = [x_i]$ of $W$ and rank each segment $s_i$ according to the component $x_i$ of $x$.

4. EXPERIMENTAL RESULTS

In order to evaluate the relevance of the results of the proposed method we need a suitable data collection together with a commonly acceptable ground truth for that collection. Following [1], Johann Sebastian Bach’s Two-part Inventions has been our choice. For this collection, a complete ground truth is provided by musicological analysis and it can be found for example in [10] and [24].

Experiments have been performed with a one-bar long window without overlapping. Two categories of experiments have been performed:

1. Two voices per window, by adopting a differential encoding for each bar $i$
   \[ win_i = alto_i - bass_i \]
2. One voice per window,
   (a) considering either just the alto voice
   \[ win_i = alto_i \]
   (b) or both voices
   \[ win_i = alto_i, win_{i+\text{lengthinbars}} = alto_i \]

4.1 2-voices experiments

Both functional metrics described in Section 3.3 have been evaluated. By performing the experiments, we observed a few variations in the first two ranked levels, and this means that top ranked bars tend to be more “stable” respect to metric changes. Thus we can say that the method is rather robust, as far as these metrics are concerned. In the synthesis reported in Table 2 we considered just the top ranked segments.

When compared to musicological analysis [1] [10] [24] it is evident that the centrality-based model outperforms the repetition-based model, providing also more significant information. Segments with higher rank in the relational model represent always relevant bars of the score, even if they may be different by using different metrics. This means that relevant bars contain a main motif or characterizing sequences. It is not the same for the model based on repetitions: here the relevancy really depends just on the number of repetitions, so it can happen that a trill turns to be more relevant than the rest of the piece just because its repetition rate is higher than that of the other bars.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition</td>
<td>43</td>
</tr>
<tr>
<td>$d_1$</td>
<td>77</td>
</tr>
<tr>
<td>$d_2$</td>
<td>95</td>
</tr>
</tbody>
</table>

Table 1: Precision results for the three models applied to J. S. Bach’s Inventions.
No Catalog Repeated bars $d_1$ $d_2$
1 BWV 772 2, 3, 24, 25 16, 18, 17, 3 16, 18
2 BWV 773 2, 20, 42, 50, 55 54, 41, 19, 48 2, 20, 42, 50, 55
3 BWV 774 2, 20, 42, 50, 55 6, 2, 29, 22 1, 5, 44, 2, 6, 45
4 BWV 775 2, 3, 28, 29 2, 16, 17, 27 2, 28, 3, 29
5 BWV 776 2, 20, 42, 50, 55 50, 41, 19, 48 2, 20, 42, 50, 55
6 BWV 777 5, 25, 42, 63, 84, 105 56, 21, 6, 20 5, 25, 42, 63, 84, 105
7 BWV 778 2, 3, 28, 29 10, 8, 5, 1 7, 4
8 BWV 779 ∅ 1, 6, 20 6, 26
9 BWV 780 1, 2, 3, 29, 30, 31, 32 31, 14, 1, 2 1, 3, 29, 2
10 BWV 781 20, 21, 22, 23, 27, 29, 31 1, 17, 27, 32 1, 27, 29, 31
11 BWV 782 ∅ 17, 4, 14, 6 6, 17
12 BWV 783 ∅ 5, 6, 17, 16 9, 5
13 BWV 784 ∅ 18, 2, 1, 21 8, 1
14 BWV 785 12, 14 8, 6, 7, 16 12, 14, 16
15 BWV 786 ∅ 7, 10, 14, 9 12, 4

Table 2: Experimental results for the repetition paradigm (using both metrics $d_1$ and $d_2$) and the relational paradigm. Gray numbers represents irrelevant bars.

Bar ranking is in principle not affected by the repetition rate of patterns and higher importance is equally given to higher and lower repetition rates. Of course, superpositions of the two methods may happen too. On the other hand, cases exist for which no repetition occurs and, consequently, the repetition paradigm is not applicable in principle, unless defining ad hoc neighborhood concepts for each piece. In these cases, motif centrality can provide significant results.

In Figure 4 the components of the main eigenvector for BWV 773, representing the degree of centrality of each bar, have been plotted against bar numbers. This provides an immediate representation of the “importance” of each bar within the whole piece. Bars with higher values are more likely to contain a main motif of the piece.

Figure 5 shows a two-dimensional projection of the 21-dimensional metric space for BWV 772 obtained through a dimensionality reduction algorithm. From this picture it is evident how the top ranked results occupy the central region of the graph and have darker labels, as the darkness is directly proportional to the correspondent component of the main eigenvector, and thus to the centrality, in the sense of graph theory, of the correspondent segment.

Table 1 presents a synthesis of the results shown in Table 2 in terms of the precision of the three methods.

4.2 1-voice experiments

Another series of experiments has been performed with monophonic windows in order to evaluate the influence of the bass part on the global bar ranking.

First, a network made of just the alto part has been considered. Figure 6 shows the eigenvector profiles for the alto parts, without considering the bass. Second, a network made of both the alto and bass parts has been considered. Figure 7 shows the eigenvector profiles for just the alto part. For both the experiments the $d_2$ metric has been used. By comparing the two plots it is evident how the top ranked results occupy the central region of the graph and have darker labels, as the darkness is directly proportional to the correspondent component of the main eigenvector, and thus to the centrality, in the sense of graph theory, of the correspondent segment.

Table 3 shows the first 5 ranked bars (alto part) with the two methods, together with a list of bars which are common to both methods. Maybe except for BWV 777, the top ranked bars are stable respect to the addition of the bass.
part to the network. This is probably a consequence of the fact that the bass part of the Inventions is highly correlated with the alto part, so that it is easy to identify transpositions and/or inversions of the alto part in the bass. As stated before, the metric we adopted ($d_2$) is a differential one, so it is invariant under transpositions.

5. CONCLUSIONS

We presented an approach for motif discovery in music pieces based on an eigenvector method. Scores are segmented into a network of bars and then ranked depending on their graph centrality. Bars with higher centrality are more likely to be musically relevant and can be exploited for music summarization. Experiments performed on the collection of J.S.Bach’s 2-parts Inventions show the effectiveness of the method.

Besides music information retrieval, we expect this approach to find applications in music theory, perception and visualization. For instance, one could investigate how particular mathematical entities (e.g. spectra) relate to particular musical issues (e.g. genre, authorship).

Second, one could investigate how different metrics $d$ relate to different concepts of melodic and harmonic similarity; in this context, the inverse problem of finding metrics $d$ induced by a priori eigenvectors (coming from a hand-made musico-logical analysis) could provide interesting insights into music similarity perception.

Finally, it is also possible to compare different music pieces from a structural point of view by comparing their associated eigenvectors.

6. ACKNOWLEDGMENTS

The author wishes to acknowledge Professor Goffredo Haus (University of Milan) and LIM (Laboratorio di Informatica Musicale of the University of Milan).

7. REFERENCES

Figure 6: Normalized eigenvector profiles for the 15 Inventions (BWV 772-786) by J. S. Bach with the $\delta_2$ metric. Network graph of bars corresponding to alto part. On the horizontal axis bar numbers have been reported and the corresponding values of centrality can be read on the vertical axis.

Figure 7: Normalized eigenvector profiles for the 15 Inventions (BWV 772-786) by J. S. Bach with the $\delta_2$ metric. Network graph of bars corresponding to both alto and bass parts. On the horizontal axis bar numbers have been reported and the corresponding values of centrality can be read on the vertical axis.
Table 3: Experimental results (first 5 ranked bars) for the alto part (using metric $d_2$) without (alto column) and with (alto + bass column) the bass part in the network.