A spectral clustering method for musical motifs classification

Alberto Pinto

Quintade Research
Viale San Gimignano, 4
20146 Milano (Italy)
apinto@quintade.org
apinto@ccrma.stanford.edu

Abstract. In recent years, spectral clustering methods are getting more and more attention in many fields of investigation for analysis and classification tasks. Nevertheless, no applications to symbolic music have been provided yet.
Here we present a method for motif classification based on spectral clustering of music scores that can be exploited, for instance, in automatic or computer-assisted music analysis. Scores are represented through a network-graph of segments and then ranked depending on their centrality within the network itself, which can be measured through the components of the leading eigenvector associated to the Laplacian of the graph. Moreover, segments with higher centrality are more likely to be relevant for music summarization.
An experimental musicological analysis has been performed on J.S.Bach's 2-part Inventions to prove the effectiveness of the method.

Keywords: spectral clustering, graph, centrality

1 Introduction

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. 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 handmade 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.

Some proposed automatic methods 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.

9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)
19-22 June 2012, Queen Mary University of London

Notwithstanding the conspicuousness of the literature, current approaches seem to rely just on repetitions [1] [2] [3], assigning higher scores to recurring equivalent melodic and harmonic patterns [4]. Recently reported approaches to melodic clustering based on motivic topologies [5], graph distance [6] [7] and paradigmatic analysis [8] have been used to select relevant subsequences among highly repeated ones by heuristic criteria [9] [10].

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. Spectral ranking on this eigenvector allows to better identify different sections within a piece through the partitioning of the score into clusters of similar melodies.

2 Related approaches

Lartillot [11] [12] 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 [13], 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 [14]. He also developed methods to recognize variations of filling and thinning (through note insertion and deletion) into the original melody. Cambouropoulos and Widmer [15] 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.

2.1 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 rhythmical 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 [16]. 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.

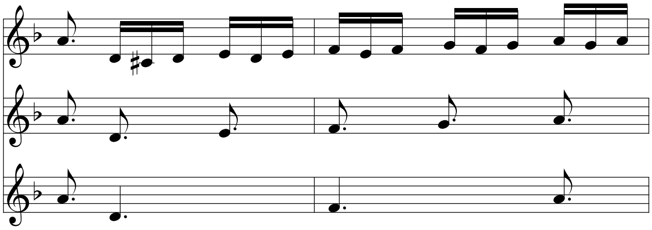

Fig. 1. J.S. Bach, BWV 1080: Score reductions.

A very interesting reductionistic approach to music analysis has been attempted by Fred Lerdahl and Ray Jackendoff. Lerdahl and Jakendoff [17] 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.

2.2 Topological approaches

Mazzola and Buteau [18] 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 repetition paradigm.

The method proposed by Adiloglu, Noll and Obermayer in [10] 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 rhythmical 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 $M = (M_i)_{i=1,\dots,N}$ 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 [14] 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 [10] 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

Our point of view can be synthesized in the following points:

1. consider a music piece as a network graph of segments,
2. take into account both melodic and rhythmical structures of segments
3. 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. Thus, a music piece can be looked at like a complete graph K_n. In graph theory, a complete graph is a simple graph where an edge connects every pair of distinct vertices. The complete graph on n vertices has $n(n-1)/2$ edges and is a regular graph of degree $n-1$. In this representation, score segments correspond to graph nodes and the similarity between couples of segments correspond to edge weights.

3.1 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

$$d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$$

between pairs of segments and turns this set into a (possibly metric) space $(\mathcal{S}, d)$. A natural choice for point sets of a metric space is the Hausdorff metric [19] but any other distance discovered to be useful in music perception, like EMD/PTD [20], can be assumed as well.

Here we assume d to be:

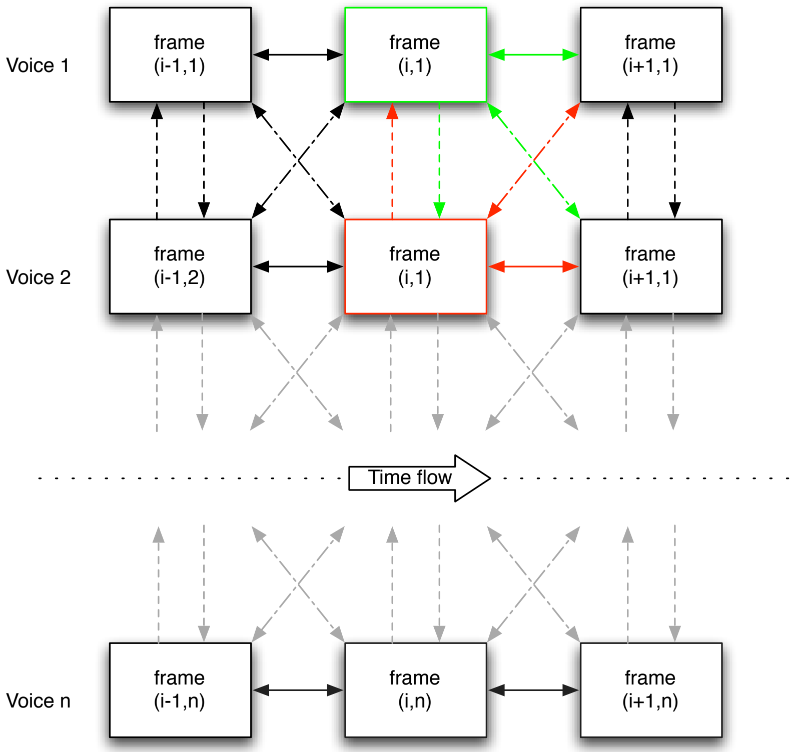

Fig. 2. A representation of the (first-order) network of frames.

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

As a matter of fact, most musically relevant perceptual distances do not satisfy all metric axioms [20]. 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:

$$d_1(s_1, s_2) = \sqrt{\sum_{|s|} |[s_1]_{12} - [s_2]_{12}|^2} \tag{1}$$

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

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 Adiloglu (2006).

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 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 respect to transpositions of the two voices so that, for instance, it makes also d_1 invariant respect to transpositions, and not just to octave shifts.

By exploiting those distance concepts, it is possible to endow 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.

3.2 The algorithm

Let $d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ denote a distance function on $\mathcal{S}$, like those defined in Section 3.1, 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

$$w_{i,j} = \exp(-\frac{-d^2(x_i, x_j)}{2\sigma^2}) \qquad (3)$$

   The parameter σ can be chosen experimentally, a possible choice is the standard deviation of the similarity values within the considered network graph (this has been our choice in the experimental part);
3. Form the Laplacian matrix $L = D^{-1/2}WD^{-1/2}$, where D is the diagonal matrix whose (i, i) element is the sum of W's i-th row
4. Compute the leading eigenvector $x = [x_i]$ of L and rank each segment s_i according to the component x_i of x.
5. Perform a k-means algorithm on the leading eigenvector to cluster the segments.

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 [10], 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 [21] and [22].

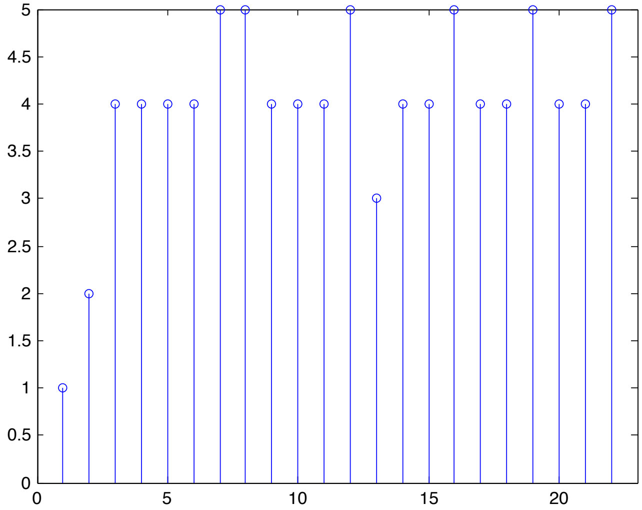

Fig. 3. Clustered bars in BWV 772 according to k-means performed on the leading eigenvector of the laplacian matrix.

When compared to musicological analysis [10] [21] [22] it is evident that the centrality-based model outperforms the repetition-based model, providing also more significative 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.

Model	Precision (%)
Repetition	43
d_1	77
d_2	95

Table 1. Precision results for the three models applied to J. S. Bach's Inventions.

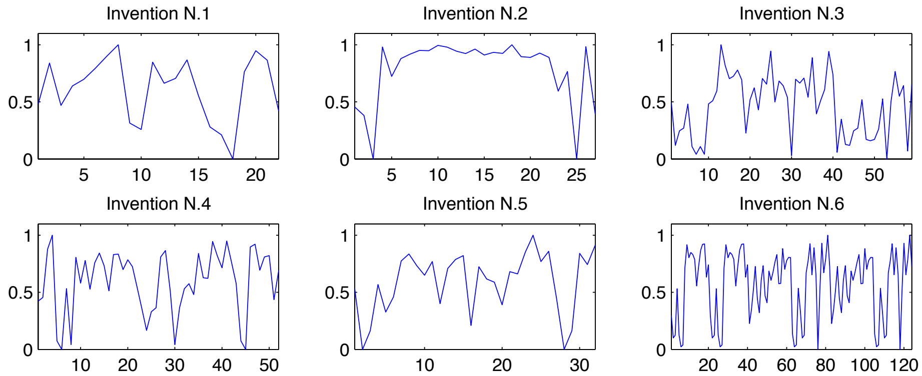

Fig. 4. Centrality values plotted against bar numbers for the first 6 J.S.Bach's Two-Part Inventions.

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 Figures 4 and 5 the components of the main eigenvector for each invention, representing the degree of centrality of each bar within the network graph, 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 3 reports the results for bar spectral clustering in the case of BWV 772 according to k-means, with k=5, performed on the leading eigenvector of the laplacian matrix. It is evident how the main theme which appears in the first two bars is identified in the first two clusters.

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. Spectral is performed in order to classify all the bar segments. Bars with higher centrality grouped into the same cluster 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.

Further investigations deal, for instance, with the relationships between particular mathematical entities (e.g. spectra) and particular musical issues (e.g. genre, authorship).

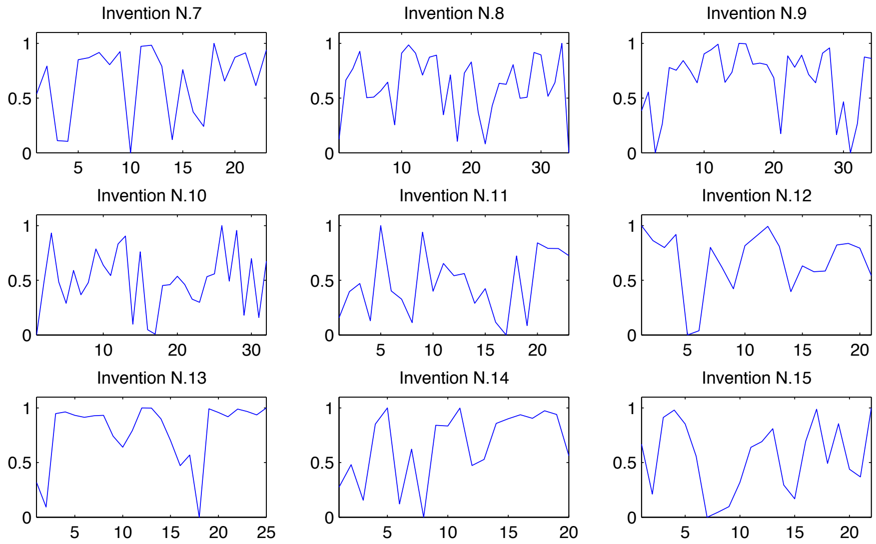

Fig. 5. Centrality values plotted against bar numbers for the last 9 J.S.Bach's Two-Part Inventions.

Second, one could investigate how different metrics d relate to different concepts of melodic and harmonic similarity and how this is related to cluster stability. In this context, the inverse problem of finding metrics d induced by a priori eigenvectors (coming from a hand-made musicological 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 segmentation derived from spectral clustering.

References

1. Pienimaki, A.: Indexing Music Databases Using Automatic Extraction of Frequent Phrases. Proceedings of the International Conference on Music Information Retrieval (2002) 25–30
2. Cambouropoulos, E., Crochemore, M., Iliopoulos, C., Mouchard, L., Pinzon, Y.: Algorithms for computing approximate repetitions in musical sequences. International Journal of Computer Mathematics **79**(11) (2002) 1135–1148
3. Livingstone, S., Palmer, C., Schubert, E.: Emotional response to musical repetition. (2011)
4. Crawford, T., Iliopoulos, C., Raman, R.: String Matching Techniques for Musical Similarity and Melodic Recognition. Computing in Musicology **11** (1998) 73–100
5. Mazzola, G., Müller, S.: The Topos of Music: Geometric Logic of Concepts, Theory, and Performance. Birkhäuser (2002)
6. Pinto, A.: Mining music graphs through immanantal polynomials. In: In Proceedings of the 6th International Workshop on Mining and Learning with Graphs. (2008)

7. Pinto, A.: Multi-model music content description and retrieval using IEEE 1599 XML standard. Journal of Multimedia **4**(1) (2009) 30
8. Nestke, A.: Paradigmatic Motivic Analysis. Perspectives in Mathematical and Computational Music Theory, Osnabrück Series on Music and Computation (2004) 343–365
9. Lartillot, O., Saint-James, E.: Automating Motivic Analysis through the Application of Perceptual Rules. Music Query: Methods, Strategies, and User Studies (Computing in Musicology) **13** (2004)
10. Adiloglu, K., Noll, T., Obermayer, K.: A paradigmatic approach to extract the melodic structure of a musical piece. Journal of New Music Research **35**(3) (2006) 221–236
11. Lartillot, O.: Discovering musical patterns through perceptive heuristics. Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003) (2003) 89–96
12. Lartillot, O.: A musical pattern discovery system founded on a modeling of listening strategies. Comput. Music J. **28**(3) (2004) 53–67
13. Cambouropoulos, E.: Extracting'Significant'Patterns from Musical Strings: Some Interesting Problems. Presente aux London String Days **2000** (2000)
14. Cambouropoulos, E.: Musical pattern extraction for melodic segmentation. Proceedings of the ESCOM Conference 2003 (2003)
15. Cambouropoulos, E., Widmer, G.: Automated motivic analysis via melodic clustering. Journal of New Music Research **29**(4) (2000) 303–318
16. Selfridge-Field, E.: Towards a Measure of Cognitive Distance in Melodic Similarity. Computing in Musicology **13** (2004) 93–111
17. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge, Massachusetts (1996)
18. Buteau, C., Mazzola, G.: From Contour Similarity to Motivic Topologies. Musicae Scientiae **4**(2) (2000) 125–149
19. Di Lorenzo, P., Di Maio, G.: The Hausdorff Metric in the Melody Space: A New Approach to Melodic Similarity. In: Ninth International Conference on Music Perception and Cognition. (2006)
20. Typke, R., Wiering, F., Veltkamp., R.C.: Transportation distances and human perception of melodic similarity. Musicae Scientiae, Discussion Forum 4A, 2007 (special issue on similarity perception in listening to music), p. 153-182.
21. Derr, E.: The Two-Part Inventions: Bach's Composers' Vademecum. Music Theory Spectrum **3** (1981) 26–48
22. Williams, P.: JS Bach. Cambridge University Press