

Distinctive Patterns in the First Movement of Brahms' String Quartet in C Minor

Darrell Conklin

Department of Computer Science and Artificial Intelligence

Universidad del País Vasco

San Sebastián, Spain

IKERBASQUE: Basque Foundation for Science

Bilbao, Spain

conklin@ikerbasque.org

Abstract

Music analysis can be seen from the perspective of data mining, aiming to discover succinct patterns that are distinctive: occurring with significantly higher probability in an analysis piece as compared to an anticorpus. In this paper a computational inductive method for maximally general distinctive pattern discovery in music is developed. The method is applied to the first movement of Brahms' string quartet No. 1 in C Minor, with the first movements of string quartets Nos. 2 and 3 used as an anticorpus. The set of patterns discovered represent most of the structures independently proposed by Forte (1983), including a proposed musical cryptogram, and three new motives are also discovered. The results show that pattern discovery can be a powerful tool for computational music analysis.

keywords: pattern discovery, motivic analysis, anticorpus, Brahms' string quartet opus 51 No. 1 in C minor, computational music analysis, computational musicology, data mining

This is a preprint of an article whose final and definitive form has been published in the Journal of Mathematics and Music, Vol. 4, No. 2, July 2010, 85-92. copyright Taylor & Francis. DOI: 10.1080/17459737.2010.515421. The final and definitive version is at <http://dx.doi.org/10.1080/17459737.2010.515421>

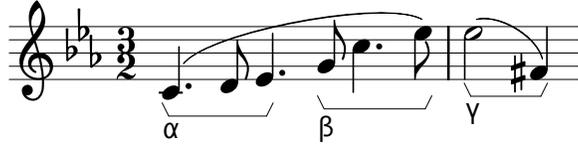


Figure 1: The opening theme of Brahms’ string quartet in C minor, with the Forte α ($[+2, +1]$), β ($[+5, +3]$), and γ ($[-9]$) motives.

1 Introduction

Brahms’ String Quartet in C minor, opus 51 no 1, published in 1873, is a work exhibiting extensive motivic compositional structure. In his seminal motivic analysis of the first movement of the work, Forte [1, page 476] proposed 11 motive groups which are repeated and combined throughout the movement. Motive groups are succinctly expressed by Forte as intervals or interval dyads along with their transformations. For example, the alpha group comprises the prime form interval dyad $[+2, +1]$ (α), its inversion $[-2, -1]$ ($\bar{\alpha}$), its retrograde $[-1, -2]$ (α'), and its retrograde inversion $[+1, +2]$ ($\bar{\alpha}'$). In this paper the terms *motive* or *pattern* will be used interchangeably to refer to a particular member of a Forte motive group. A written form (e.g., beta, gamma) will be used to refer to a motive group; following Forte, Greek letters (e.g., β , γ) will be used to refer to particular patterns within a group.

Figure 1 illustrates three prominent Forte motives (the prime forms of alpha, beta, and gamma) of the movement, indeed the three which together comprise the principal theme of the movement. Immediately apparent is the micro-structural nature of these motives; three are needed just to build the principal theme. Indeed, Forte refers to motives as “atomic constituents” and “[miniscule] motivic particles”.

It is evident that Forte identified motive groups according to their repetition and transformation throughout the movement, and according to their succinctness or generality. In his seminal computational analysis of the first movement, Huron [2] argued that the mere presence of a pattern does not justify its motivic status, and suggested that motives should

be somehow *distinctive*, that is, over-represented in the work as compared to a comparison corpus. Focusing specifically on the alpha motive group, and finding that it was not distinctive (using the first movements of Brahms' string quartets opus 51 No. 2 and opus 67 as a comparison corpus, and using a χ^2 test for statistical significance), he considered a series of refinements, noting that distinctiveness was achieved only when its prime form α is linked with rhythmic features.

The procedure employed by Huron to determine a distinctive refinement to the alpha prime form motive can be characterized as a *deductive* analysis: a piece or corpus is computationally queried with known or postulated patterns to determine their instances. An *inductive* analysis, by contrast, is one that in principle presents the possible interesting queries without their prior knowledge. To elaborate further, for a deductive analysis, we are given a corpus \mathcal{D} and pattern P , and are interested in all instances x of the pattern, resulting in the answer set $\{x : \mathcal{D} \models P(x)\}$. In an inductive analysis, a pattern is not specified and we are interested in all patterns P satisfying an interestingness predicate I , resulting in the answer set $\{P : \mathcal{D} \models I(P)\}$. Interestingness may be a conjunction of conditions such as frequency of occurrence, distinctiveness in a particular piece or group of pieces, succinctness or generality, pattern length, textual allusions, and dispersal throughout a piece.

The hypothesis of this paper is that a computational inductive analysis of the Brahms string quartet in C minor, opus 51 No. 1, first movement, can reveal the Forte motive groups. To explore the hypothesis of this paper, a general inductive method is presented. This method employs an anticorpus and discovers patterns that are both succinct and distinctive. In the exploration of this hypothesis the prior analysis by Forte provides a unique opportunity to measure the quality of the results.

The results of this paper validate the main hypothesis; with two expected exceptions, all of Forte's motive groups can be found using a pattern discovery algorithm. Regarding the alpha group, in agreement with Huron [2] only the prime form is found to be distinctive,

and then only when specialized or refined. In contrast to Huron, who considers only *vertical* refinements to the α motive components, here it is found that *horizontal* or sequential specializations of motives can also reveal distinctive refinements. An additional interesting result is that the α motive is not ranked the highest, and several other motives (σ , β , $\bar{\epsilon}$, and $\bar{\beta}'$) are found to be more distinctive.

2 Methods

This section will outline the important terminology needed to describe the inductive method; patterns, pattern subsumption, pattern frequency, pattern distinctiveness, and finally the concept of a maximally general distinctive pattern. Concepts of sequential patterns are standard in data mining [3], and a procedure to perform structural music analysis with discovered patterns was described by Ruwet in the 1960s [4]. These early ideas influenced later computational methods for pattern analysis of single pieces [5, 6] and large corpora [7].

A *pattern* is a sequence of event features (for this paper, simply melodic intervals). Patterns can be viewed as unary predicates: functions from pieces to boolean values. A piece x *instantiates* a pattern P , written $P(x)$, if the pattern occurs (possibly multiple times) in the sequence: if the components of the pattern are instantiated by successive events in the sequence. For example, the pattern $[+3, 0]$ has two components and occurs at position 5 of the Brahms theme in Figure 1.

A pattern P *subsumes* (is more general than) a pattern Q if all instances of Q are also instances of P : if $\forall x Q(x) \Rightarrow P(x)$ is valid (true in all possible corpora). A sound and complete algorithm to determine subsumption is to simply search for a contiguous mapping from the components of P to the components of Q . For example, the pattern $[+3]$ subsumes the pattern $[+3, +1]$, which in turn subsumes, for example, $[+2, +3, +1]$. This type of subsumption may be called sequential or *horizontal*, in contrast to a *vertical*

subsumption which would specialize individual pattern components (e.g., requiring that the +3 interval is linked with a particular duration).

In music sequential patterns can be counted in one of two ways depending on the analysis task: as the number of pieces that contain the pattern (for large corpus analysis); or as the total number of occurrences of the pattern (for single piece analysis, as in the present study). The *total count* of a pattern is the total number occurrences of the pattern in the corpus (which may be a single piece), not counting overlapping occurrences. A *frequent pattern* occurs with a minimum specified total count in the piece. Frequency alone is inadequate for ranking discovered patterns, especially with a pattern language that allows very general and abstract patterns to be expressed. For example, the empty pattern is maximally frequent, but is completely uninteresting. Therefore, patterns must somehow be scored or ranked, in a way that balances their specificity and frequency.

To rank discovered patterns it is possible to partition a collection of pieces into an explicit anticorpus (denoted \ominus), contrasting it with the analysis corpus (denoted \oplus). A *distinctive pattern* is one that is frequent and sufficiently over-represented in the corpus as compared to the anticorpus. An intuitive way to measure this is according to the relative empirical probabilities of a pattern in the corpus and anticorpus, equivalently as the ratio of its count to its expected count:

$$\Delta(P) \stackrel{\text{def}}{=} \frac{p(P|\oplus)}{p(P|\ominus)} = \frac{c^\oplus(P)}{p(P|\ominus) \times n^\oplus}. \quad (1)$$

In the definition above, $c^\oplus(P)$ is the total count of a pattern P in the corpus, n^\oplus is the number of events in the corpus, and $p(P|\oplus)$ and $p(P|\ominus)$ are the empirical probabilities of the pattern P given the corpus and anticorpus, respectively (see Glossary of Notation in Table 1). Patterns with $\Delta(P) > 1$ occur more frequently than expected, and patterns are ranked from high to low values of $\Delta(P)$ (if $p(P|\ominus) = 0$, $\Delta(P)$ is defined as ∞).

This study uses the first movements of three Brahms string quartets in Humdrum kern

Table 1: Glossary of Notation.

Notation	Meaning
$c^\oplus(P), c^\ominus(P)$	total count of pattern P in a corpus/anticorpus
n^\oplus, n^\ominus	number of events in a corpus/anticorpus
$p(P \oplus)$	empirical probability of a pattern P given the corpus: $c^\oplus(P)/n^\oplus$
$p(P \ominus)$	empirical probability of a pattern P given the anticorpus: $c^\ominus(P)/n^\ominus$
$\Delta(P)$	distinctiveness of a pattern P

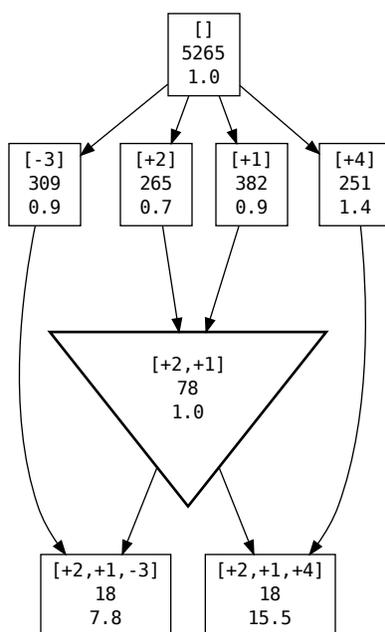


Figure 2: The α ($[+2, +1]$) pattern, with a surrounding subsumption space. Each node contains a pattern, its total count, and its distinctiveness.

format [8]. Several cycles of error correction were performed, using the reference text *Brahms Quartet for 2 Violins, Viola and Violoncello, C minor, Op. 51/1*, Edition Eulenburg Ltd., London, No. 240. For all three string quartets, only one instance of the repeated exposition, using the second ending, was taken. Melodic intervals were derived for each instrumental part. Only one note of a multiple stop was used for melodic interval derivation. Rests in the score were ignored, that is, melodic intervals interrupted by rests were also computed. For this study rhythmic aspects were also ignored; they are not necessary to describe the Forte motive set, but they could also be considered by this pattern discovery method [7].

There may be excessively many distinctive patterns in a corpus: this is because a pattern subsumed by (more specific than) a distinctive pattern will also tend to be distinctive. A pattern P is called a *maximally general distinctive pattern* (MGDP) if it is distinctive (with at least a specified minimum $\Delta(P)$), and if there does not exist a more general (subsuming) pattern that is also distinctive. Thus the set of all MGDP is the top border of the virtual pattern subsumption taxonomy such that all patterns above the border are not distinctive.

Figure 2 provides an illustration of the concepts outlined above, drawn directly from actual results with the Brahms corpus and anticorpus (see Results). Arrows represent pattern subsumption (which is a transitive relation), and nodes represent patterns. Each node is also annotated with its total count $c^{\oplus}(P)$ (total count in all four instrumental parts of the first movement of opus 51 No. 1) and its distinctiveness $\Delta(P)$. The pattern space presented in Figure 2 revolves around the alpha prime form α ($[+2, +1]$) in the inverted triangle. It is not distinctive, with an $\Delta(P)$ of 1.0 (occurring with equal probability in the corpus and anticorpus). Different MGDP sets result from different minimum values of $\Delta(P)$. For example, with $\Delta(P) \geq 1.4$ the patterns $[+2, +1, -3]$ and $[+4]$ would form the MGDP set. The pattern $[+2, +1, +4]$, though distinctive, would not be included as it is subsumed by the distinctive pattern $[+4]$. However, with $\Delta(P) \geq 3$, it would indeed be in the MGDP set which then becomes $[+2, +1, -3]$ and $[+2, +1, +4]$. Note that although the subsumption taxonomy

of patterns and their total counts remains unchanged, the MGDP set varies according to the choice of the minimum value for distinctiveness.

To find the MGDP set, a depth-first search of a subsumption space iteratively refines patterns at each search node to make them more specific. At each refinement step, two instance lists (one for the corpus, one for the anticorpus) are locally updated efficiently using hash map representations. The search space is structured so that no pattern need be evaluated more than once. Given that maximally general distinctive patterns are sought, a branch in the search tree may be immediately terminated with success if the minimal specified pattern interest is achieved by the pattern at a search node. This is safe due to the antimonotonicity of the MGDP property: if P is an MGDP, and P subsumes Q , then Q cannot possibly be an MGDP.

3 Results

The pattern discovery method described in Methods was applied to the first movement of the Brahms string quartet in C minor, opus 51 No. 1 ($n^{\oplus} = 5265$), using the first movements of the string quartets Nos. 2 and 3 as the anticorpus ($n^{\ominus} = 9092$). A pattern P was considered interesting if it occurs at least once in each part (violin 1, violin 2, viola, and cello), and is distinctive (with $c^{\oplus}(P) \geq 10$ and $\Delta(P) \geq 3$). It was found that much lower values of $\Delta(P)$ led to lower precision (when the Forte set was taken as reference), whereas with much higher values patterns tended to be overly specific.

To evaluate the method, it is possible to consider the proportion of Forte’s motive groups that are present in discovered patterns. Of the original 11 Forte motive groups, two are not considered in this evaluation: group omega, a descending perfect fifth, is included by Forte as it outlines the Schenkerian G to C background. Motive group delta, outlining an interval of a perfect fourth, occurs both in the middleground and foreground. Both omega and delta

Table 2: Distinctive patterns, corresponding to Forte motive groups, discovered in the first movement of Brahms' String Quartet in C Minor.

Pattern	Instance	$\Delta(P)$	$c^\oplus(P)$	Forte motive
[+3, +3, +2]		∞	10	σ [+3, +3]
[+5, +3, 0]		31.1	18	β [+5, +3]
[+2, +2, +3]		22.4	26	$\bar{\epsilon}$ [+2, +2]
[+2, +3, +5]		17.3	20	$\bar{\beta}'$ [+3, +5]
[+2, +1, +4]		15.5	18	α [+2, +1]
[+4, -9]		10.4	12	γ [-9]
[0, +12]		9.3	113	$\bar{\theta}$ [+12]
[-12, 0]		8.0	111	θ [-12]
[+2, +1, -3]		7.8	18	α [+2, +1]
[-1, 0, +1]		7.1	62	$\bar{\lambda}, \lambda$ [-1], [+1]
[0, -12]		6.4	114	θ [-12]
[+3, +5, +4]		5.2	30	$\bar{\beta}'$ [+3, +5]
[+12, 0]		5.0	114	$\bar{\theta}$ [+12]
[+4, 0, +3]		4.1	12	$\bar{\phi}'$ [+4, +3]
[-8, 0]		3.8	20	χ [-8]
[+1, 0, -1]		$\frac{3.1}{9}$	45	$\lambda, \bar{\lambda}$ [+1], [-1]

have instances that are not contiguous on the music surface, and cannot be expected to be discovered by this inductive method.

Exactly 19 patterns were discovered and, with the noted exception of omega and delta, all of the Forte motive groups were found to be present within discovered patterns. Table 2 lists 16 of the 19 patterns, sorted from high to low distinctiveness. All patterns are highly significant according to χ^2 tests (data not shown). For each pattern an instance, taken from the earliest occurrence within violin 1, is presented. The bar number of the first note of the instance is provided. Rhythmic values are not notated as they are not relevant for the instantiation of a pattern. The pattern distinctiveness $\Delta(P)$ and the pattern total count $c^\oplus(P)$ are presented. In the final column is a reference to the Forte motive labelling.

The patterns in Table 2 can be discussed with reference to the Forte motive group that they represent or elaborate:

sigma is considered by Forte as the musical cryptogram revealing the “secret emblem of the work”. In a prominent instance the prime form of σ ($[+3,+3]$) occurs with pitch classes (A, C, Es), and c and es also are the key areas of the first and second subject of the exposition. Both allude to the name Clara Schumann, C as in Clara, and Es as in Schumann [1, page 499]. The σ motive is found within the pattern $[+3,+3,+2]$. This pattern is found to be maximally distinctive, occurring 10 times in the corpus and not once in the anticorpus;

beta is found in three patterns, in the prime form β ($[+5,+3,0]$) and twice in retrograde inverted $\overline{\beta'}$ forms ($[+2,+3,+5]$ and $[+3,+5,+4]$);

alpha in its prime form α is found in two extended patterns ($[+2,+1,+4]$ and $[+2,+1,-3]$), with the original Forte α motive followed either by a major third upwards or minor third downwards;

epsilon in its prime form ϵ ($[-2,-2]$) a descending major third with a passing tone, is found

in inverted (ascending) form $\bar{\epsilon}$ in the distinctive pattern $[+2, +2, +3]$;

theta occurs within four patterns, twice in prime form θ (within $[0, -12]$ and $[-12, 0]$) and twice in inverted form $\bar{\theta}$ (within $[12, 0]$ and $[0, 12]$). In fact, the extended pattern forms a complete distinctive motive group with a prime form $[-12, 0]$, and its three transformations $\overline{[-12, 0]} = [12, 0]$, $[-12, 0]' = [0, 12]$, $\overline{[-12, 0]}' = [0, -12]$;

gamma in its prime form γ ($[-9]$) is found within the pattern $[+4, -9]$;

lambda is found in the patterns $[-1, 0, +1]$ and $[+1, 0, -1]$. Thus in these patterns the prime form λ ($[+1]$) always appears in connection with its inverted form $\bar{\lambda}$ ($[-1]$);

chi in its prime form a descending minor sixth, noted as “not an especially significant or prominent motive” by Forte, is indeed distinctive and is found within the pattern $[-8, 0]$;

phi in its retrograde inversion form $\bar{\phi}'$ ($[+4, +3]$), outlining a major triad, is found in an elaborated pattern $[+4, 0, +3]$ (the third of the major triad being repeated).

Three distinctive patterns not corresponding to Forte motive groups are present within the 19 discovered patterns (Table 3). The pattern $[-1, +4]$ often occurs as a linker between the presentation of the principal theme of the movement and its modified presentation. It appears prominently in pitch doubled form in bar 22, and again carried by only violin 1 in bar 163. The pattern $[0, +6]$ appears in all four parts, in two clusters in the movement, exactly corresponding to the extension of the second subject. The pattern $[+3, -3]$ appears prominently in the beginning of the development section (bar 84), with alternating thirds carried for several bars in the first and second violin parts.

Table 3: Distinctive patterns, not corresponding to Forte motive groups, discovered in the first movement of Brahms’ String Quartet in C Minor.

Pattern	Instance	$\Delta(P)$	$c^\oplus(P)$
$[-1, +4]$		13.0	15
$[0, +6]$		3.9	18
$[+3, -3]$		3.3	62

4 Discussion

This paper has presented the results of an inductive analysis of the first movement of Brahms’ String Quartet in C Minor. In this study the prior analysis by Forte has provided a unique and invaluable control to evaluate the performance of pattern discovery. The results can be summarized as follows:

- The Forte motive groups, with the expected exception of delta and omega, all occur in the discovered pattern set, though all in a slightly more specialized form. This result vindicates both Forte (in that nearly all of his proposed motives are present in distinctive patterns) and Huron (in that the Forte motives alone, particularly the α pattern, are too general to be distinctive).
- The α pattern, though arguably the most salient due to its prominence in the opening theme, is not the most distinctive pattern in the results. Patterns containing forms of sigma, beta, and epsilon are more distinctive.
- In agreement with Huron, the alpha motive group is not distinctive in its prime form α but only when refined. However, while Huron performed a vertical specialization of α components, here it is found that two horizontal, sequential, refinements of the

α motive are distinctive. None of the three transformed forms of α appear within distinctive patterns.

- The pattern $[-12, 0]$ is distinctive in all of its four transformations, thus it might have been a good alternative choice for the theta prime form. Similarly, the pattern $[-1, 0, +1]$, which contains both forms of the lambda group, is also present in its retrograde inversion form, and might have been considered as a prime form for a more specific lambda group.
- In no other cases are all members of a motive group distinctive. This result recalls Huron’s pertinent criticism that “set theory implies an equivalent or enhanced status for the retrograde, inversion, and retrograde inversion forms of the feature that are often not borne out by the work itself” [2, page 24].
- Novel motives may be proposed, for example the motive $[-1, +4]$ which serves as a linker between the presentation of the theme and its modified form.

Though Forte identified motives not due to their distinctiveness but rather for their explanatory value and succinctness, this study has shown that with slight refinements most of them are distinctive when the other two string quartets by Brahms are used as an anticorpus. The idea that a pattern should be distinctive, but also maximally general, appears to be a powerful concept for the suggestion of motives during computational music analysis.

Therefore the hypothesis of this paper, that an inductive analysis of the Brahms can reveal the Forte motive groups, has been validated. Though only a single piece has been studied here, one may hypothesize that high distinctiveness usually implies a motivic status. The converse, that all motives are distinctive, is certainly not valid. Motives in music need not always be frequent or prevalent. In fact, even within the Forte set there are two motives that could not be distinctive at the music surface, and motives might have an explanatory and intertextual value not explained solely by the notion of distinctiveness.

The computational method presented here, employing the general notions of corpus, anticorpus, and maximally general distinctive patterns, can be used for many music analysis tasks. This automated procedure that can suggest to an analyst a possible motive set will be invaluable for future studies of many other music corpora.

References

- [1] A. Forte, *Motivic design and structural levels in the first movement of Brahms's String Quartet in C minor*, *The Musical Quarterly* 69 (1983), pp. 471–502.
- [2] D. Huron, *What is a musical feature? Forte's analysis of Brahms's Opus 51, No. 1, revisited*, *Music Theory Online* 7 (2001).
- [3] J.M. Adamo *Data Mining for Association Rules and Sequential Patterns*, Springer, 2001.
- [4] N. Ruwet, *Méthodes d'analyse en musicologie*, *Revue belge de musicologie* 20 (1966), pp. 65–90.
- [5] E. Cambouropoulos, *Musical parallelism and melodic segmentation: a computational approach*, *Music Perception* 23 (2006), pp. 249–269.
- [6] O. Lartillot, *A musical pattern discovery system founded on a modeling of listening strategies*, *Computer Music Journal* 28 (2004), pp. 53–67.
- [7] D. Conklin, *Discovery of distinctive patterns in music*, *Intelligent Data Analysis* 14 (2010), pp. 547–545.
- [8] <http://csml.som.ohio-state.edu/HumdrumDatabases/classical/Romantic/Brahms>, accessed November 13, 2007.