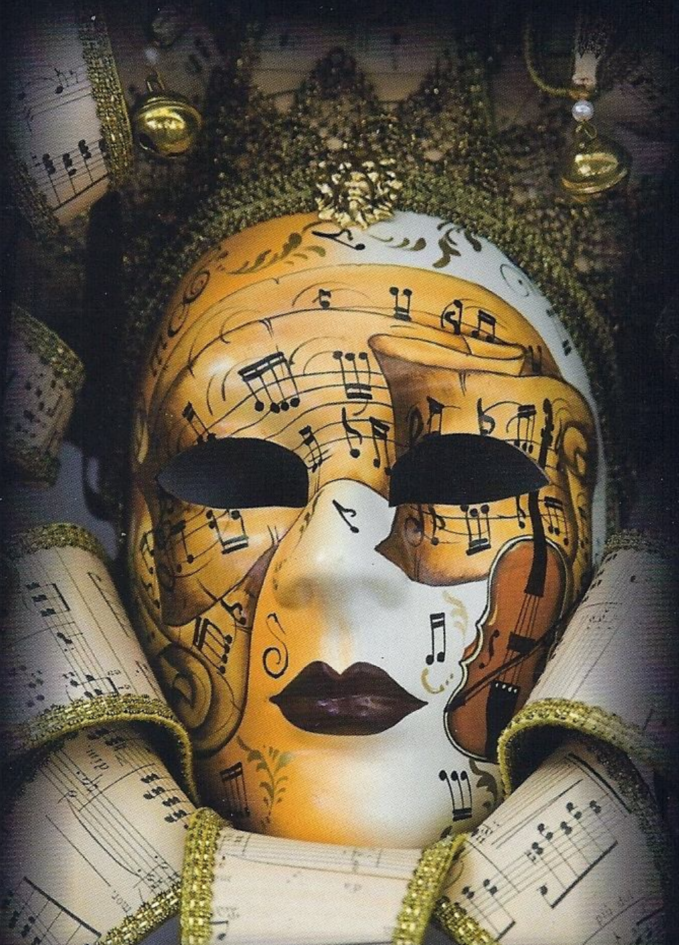
MUSIC SEMIOTICS:
A NETWORK OF SIGNIFICATIONS
In Honour and Memory of Raymond Monelle

ASHGATE



EDITED BY

Esti Sheinberg

But those who bemoan the loss of a rich and dense expression in the works composed after Op. 12 (see Tallián and Ujfalussy) are also right: while Bartók used a rhetorical or narrative type of compositional technique inspired by Beethoven, Liszt and Schumann in the 'Preludio' (and at other moments in Op. 12), he later (from 1926 onwards) abandoned this somewhat romantic construction, preferring a more static, balanced and classical relation, which also poses a more neutral, less 'eloquent' and less expressive relation between the elements or sections of a movement.

In the context of the studies quoted above, the originality of this analysis arises out of the fact that the stressed Hungarian tone, connected with the theme of nature, appears very clearly here – and perhaps for the first time – as the conclusion of a movement (after its very first and peculiarly timid manifestation in the First Quartet Op. 7/I and before its unequivocal presence in the first movement of the Second Quartet of 1915).[38]

When compared with Op. 10/I, the rhetorical expression or narrative structure found in Op. 12/I suggests a richer semantic purport: nature can perfect its song, by completing and concluding it, only after uniting with the Hungarian tone 'inspired', in the 'Preludio', by the passionate song of man.

Moreover, this demonstration suggests, more generally, that a narrative analysis founded on the use of topics deepens our understanding of musical works – beyond analytical generalizations of Bartók's works, that opposed Man to Nature, into far more specific messages, and possibly applicable beyond the specific field of Bartók studies.

[38] See Somfai, 'Béla Bartók Thematic Catalogue' and 'A Characteristic'.

Chapter 11

Semiotic Analysis and Computational Modelling: Two Case Studies on Works by Debussy and Xenakis

Christina Anagnostopoulou and Emilios Cambouropoulos

1. Introduction

Music becomes intelligible to a great extent through its inner self-referential structural relations. Establishment of such relations between new unheard musical passages to those previously heard gives rise to meaningful musical units that unfold in time in a cohesive manner. Repetition, variation and transformation are crucial devices in establishing such relations within a piece of music. Musical entities that are closely linked, sharing musical properties on different parametric levels, can be organized into musical categories or paradigms such as rhythmic and melodic motives, themes and variations, harmonic progression groups, pitch class sets, and various other structural components.

Semiotic analysis at the neutral level is concerned with understanding pieces of music by identifying their constituent structures and by how these are transformed as a musical piece unfolds in time.[1] Paradigmatic analysis is the first stage of semiotic analysis whereby a musical work is segmented and organized into paradigms or categories of meaningful musical units, with segments placed into the same category according to various similarity criteria. The second stage of the process, syntagmatic analysis, involves the description of the temporal distribution and succession of these analytically significant categories.

The two proposed computational models described in this chapter address mainly issues relating to paradigmatic analysis, as this is a prerequisite for syntagmatic analysis. Paradigmatic analysis focuses on aspects of a-temporal structural logical relations. Agawu suggests that 'conventional analysis of tonal structure … has long privileged the chronological manifestation of form; a paradigmatic approach compels us to reckon with a-temporal logical form as

[1] Jean-Jacques Nattiez, *Fondements d'une Sémiologie de la Musique* (Paris, 1975); Jean-Jacques Nattiez, *Music and Discourse: Toward a Semiology of Music*, trans. Carolyne Abbate (Princeton, 1990). Originally in French: *Musicologie générale et sémiologie* (Paris, 1987); see also Nicholas Cook, *A Guide to Musical Analysis* (London, 1987); Raymond Monelle, *Linguistics and Semiotics in Music* (Chur, 1992).

well'.[2] The idea of form without the time dimension is also discussed by Xenakis,[3] and since then has been a vital concept in computational approaches to music analysis.

In the descriptions and discussion below, we chose to focus on the methodology as it relates to music analysis, rather than on algorithmic and implementation details. It will be argued that paradigmatic analysis has so far resisted full formalization that may allow the implementation of sophisticated computational musical analytic systems; this proves to be a continuing challenge for the science behind music analysis, and significant amounts of current research are devoted to this.

2. Semiotic Analysis

Nattiez's attempt to systematize musical analysis makes use of three distinct but closely related levels of music in which analysis may be pursued, as introduced by Molino.[4] First, the neutral level, which is the immanent configurational properties of a musical work; second, the poietic level, which is related to compositional procedures and intentions; and third, the aesthetic level, involving interpretation and perceptual processes.

Despite his strong advocacy of the neutral level, Nattiez seems to acknowledge indirectly the fact that analysis conducted purely at the neutral level is essentially intractable, by stressing the interdependency of the three levels: 'Analysis never stops engineering a dialectical oscillation among the three dimensions of the object. Analysis at the neutral level is dynamic; it displaces itself constantly as the analysis takes place'.[5] In this sense, the neutral level might be considered a methodological device, or a 'methodological artefact' in Laske's words,[6] offering an analytic methodology that forces an analyst to make their own decisions and judgements explicit. Such an approach provides a broad framework for analysis, and points towards a set of explicit representations and methods that may – or may not – lead to pertinent analyses.

[2] Kofi Agawu, *Music as Discourse: Semiotic Adventures in Romantic Music* (Oxford, 2009), p. 167.

[3] Iannis Xenakis, *Formalized Music: Thought and Mathematics in Composition* (New York, 1992).

[4] Nattiez, *Fondements* and *Music and Discourse*; Jean Molino, 'Fait Musical et Sémiologie de la Musique', *Musique en Jeu*, 17 (1975): pp. 37–61.

[5] Nattiez, *Music and Discourse*, p. 32.

[6] Otto Laske, 'Towards a Musicology for the Twentieth Century', *Perspectives of New Music*, 15/2 (1977): p. 221.

Paradigmatic analyses have been primarily applied to melodic surfaces,[7] although Agawu recently attempted to analyse full tonal musical works based on paradigmatic methodology.[8] The two abstract models that we propose are tested on melodic surfaces; however, they could be applied, in principle, to any aspect of a musical surface, for which a symbolic representation can be used.

2.1. Challenging issues in the paradigmatic methodology

There have been several reasons for the unpopularity of the paradigmatic analysis among non-computational approaches; one of them may be the rigour in its application, regarding several issues that need to be addressed. Below we focus on three problem areas that relate, first, to the selection of salient musical parameters for the description of musical entities, second, the hierarchic organization of musical structure, and third, the segmentation of a musical surface. The issue of musical similarity, arguably the most challenging one, is not brought up separately in this section, but discussed throughout the chapter.

Musical parameters

The set of parametric features that is important for classifying musical units of a specific musical work into paradigms is usually defined in an ad hoc manner depending on musical context; each piece of music requires a specially compiled list of features that are deemed relevant for each particular musical context. The paradigmatic methodology does not suggest a general set of features or at least a general strategy as to how such features may be selected.

This, however, is arguably one of the strengths of this particular type of analysis, since it is clear that not only each piece would require a different set of features, but each analyst, too, may choose to focus on different musical properties. This freedom of analytical choice is crucial and stands at the core of the very nature of music analysis. In computational terms, one might envisage a potentially large general set of musical properties and parameters that can be narrowed down to a smaller set of salient parameters within a given context, while allowing the analyst/user to construct new, additional parametric features.

[7] Nicolas Ruwet, 'Méthodes d'analyse en musicologie', *Revue Belge de Musicologie*, 20 (1966): pp. 65–90. Trans. and intro. Mark Everist, *Music Analysis*, 6/1-2 (1987): pp. 3–36; Nattiez, *Fondements*; see also Jean-Jacques Nattiez, 'Varèse's "Density 21.5": A Study in Semiological Analysis', trans. Anna Barry, *Music Analysis*, 1/3 (1982): pp. 243–340; David Lidov, 'Musical Structure and Musical Significance – I' (Toronto, 1980); Elisabeth Morin, *Essai de Stylistique comparée: les variations de William Byrd et John Tomkins sur 'John Come Kiss Me Now'* (Montreal, 1979); Marcelle Guertin, 'Différence et Similitude dans les Préludes pour Piano de Debussy', *Revue de Musique des Universités Canadiennes*, 2 (1981): pp. 56–83.

[8] Agawu, *Music as Discourse*.

Musical hierarchy

The more hierarchically structured the elements of a musical surface are, the harder it is to perform paradigmatic analysis. This is due to the fact that, not only has one to determine a list of features that is relevant for the analysis of the musical surface, but also a set of pertinent reductions of the surface at a number of hierarchical levels, and a list of features that are relevant for the analysis of each reduction. The issue of finding repeated motives in a reduced score is thus quite complicated.[9]

The two approaches presented in this chapter allow the construction of parametric features of musical entities at different hierarchic levels without proposing specific reduction mechanisms.

Segmentation

Paradigmatic analysis, like many other types of formal analysis, relies to a large extent on the segmentation of a musical work – perhaps the most problematic aspect of this method. Segmentation is not merely a prerequisite to paradigmatic analysis but is implicitly affected by the taxonomic process itself. In other words, even though it is plausible that there is a tendency to generate, first, a proto-segmentation based on local discontinuities and, then, to organize segments into paradigms, the reverse relation is also active: repetition, similarity and categorization may affect segmentation itself enabling disambiguation of unclear segmentation points or even suggesting new segmentation boundaries.[10] David Lidov proposes a special type of repetition, referred to as *formative repetition*, 'which refers or directs attention to and marks the material repeated … defining the units of a musical work and establishing their position in a hierarchy of longer and shorter segments'.[11] Segmentation is integrated in the whole paradigmatic process of analysis, and can hardly be thus considered an independent step in the paradigmatic process. It is natural therefore that two different segmentations would produce different results, while a non-acceptable segmentation would ruin an analysis.

9 See Olivier Lartillot, 'Motivic Pattern Extraction in the Symbolic Domain', in Jialie Shen et al. (eds), *Intelligent Music Information Systems: Tools and Methodologies* (Hershey, 2008).

10 For discussions confronting these two directions of segmentational analysis compare Emilios Cambouropoulos and Gerhard Widmer, 'Automated Motivic Analysis via Melodic Clustering', *Journal of New Music Research*, 29/4 (2000): pp. 347–370 and Emilios Cambouropoulos, 'Musical Parallelism and Melodic Segmentation: A Computational Approach', *Music Perception*, 23/3 (2006): pp. 249–269.

11 David Lidov, 'Structure and Function in Musical Repetition', in *Is Language a Music: Writings on Musical Form and Signification* (Bloomington, 2005), p. 30.

2.2. Computational methodology

In his description of discovering paradigms, Ruwet uses the word 'machine' as early as 1966.[12] He envisaged an automatic way in which this can be achieved, and encouraged analysts to be 'machine-like' in the segmentation and comparison of the musical units. Nattiez, however, states: 'It is hard to see how a computer could automatically establish an equivalence which depends on a judgement of similarity transcending concrete resemblances and differences.'[13] Taking this statement as a challenge rather than as a deterrent, a number of researchers, us included, have attempted to develop formal models that can automatically produce segmentations, extract significant patterns and organize segments into meaningful musical categories. The issue of the role of the analyst in this process, together with the possibility of achieving a fully automated analysis, is discussed in the last section of the chapter.

It seems that the paradigmatic methodology has directly or indirectly prompted more research in the domain of computational musicology rather than in more traditional music analysis. Perhaps the main challenge with paradigmatic analysis is one of tractability. Ruwet and Nattiez propose a method for the construction of a 'good' paradigmatic description, a small number of distinct paradigms that cover most of the musical surface, in the course of which appropriate features and segmentations are discovered. However, they do not offer a tractable or explicit method for implementing this, except for the simplest cases of surfaces consisting mostly of exact repetitions, where the search space is sufficiently small. Computational methods can be employed to test out hypotheses and to track down possible routes that lead to acceptable solutions.

Humans are extremely good at sorting out complex musical entities by making quick decisions, intuitively reasonable and sometimes even inspired. However, they have difficulty in exploring such large spaces systematically and in stating explicitly the mechanisms by which certain results are derived. Computers, on the other hand, might have difficulties in such quick common-sense decisions, but are extremely good in performing enormous amounts of calculations and in trying out various search routes in order to find the best solutions that meet given criteria.

In the field of computer science the issue of similarity and classification of objects or concepts has been particularly prominent, since it is at the core of most disciplines, most notably in concept formation and categorization in language. Various researchers have proposed methods for feature selection and for classification and categorization.[14] Often these approaches have been informed by

12 Ruwet, 'Méthodes d'analyse'.

13 Nattiez, 'Varèse's "Density 21.5"', p. 257.

14 George F. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (Boston, 2009).

the area of cognitive psychology, which discusses how humans perform similarity and categorization tasks.[15]

There have been two main approaches in investigating musical similarity, partly reflected by the two approaches described in this chapter. In the first approach, we look for identical patterns in the various parameters, such as pitches, intervals, durations, and other more abstract parameters (e.g. contour, ratios of values). Identity of a pattern in one parameter, such as a succession of interval values, results in musical segments that might be identical or similar. Thus, similarity is captured in terms of identity in the representation of the parameters; the more abstract the parameter, the more distant the segments appear to be. This approach explores the issue of pattern extraction in multi-parametric, hierarchically structured data, and moves towards a paradigmatic analysis, even though it does not produce at this stage a full paradigmatic chart. The second approach starts with a given segmentation (which is a significant step taken for granted), then builds numerous parametric descriptions at various levels of abstraction for each segment, calculates distances between segments in terms of the number of parametric features shared, and finally organizes the segments into paradigms based on a clustering algorithm. During the whole process different parameters gain more weight that reflects how characteristic they are of a certain paradigm. Although identity may exist in the various parameters separately, at the end all parameters are considered and an overall distance is calculated between segments. Both approaches use the score as their starting point (in an appropriate format which can become the input for an algorithm), and not the sound file, as is often the case in musical informatics.

3. Semiotic Analysis as Pattern Discovery: An Example Analysis of Xenakis' *Keren*

This section presents an overview and discusses analytical issues related to the computational analysis of *Keren*, a monophonic piece for trombone solo by Iannis Xenakis. In this analysis, repeated patterns are discovered within musical sequences and a statistical model is used to test their significance. Following the process of syntagmatic analysis, a search is made for additional patterns, in which each element of the pattern is a musical segment rather than a single note. This reveals how the various segments are placed together in time to create larger-scale structures. We discuss the general concept of a pattern, how patterns are discovered, and what this may mean for music analysis. We then bring up the issue of the statistical significance of patterns found by the algorithm.[16]

15 A good review of this subject can be found in Michael W. Eysenck and Mark T. Keane, *Cognitive Psychology: A Student's Handbook*, 4th edn (Philadelphia, 2005).

16 This analysis of *Keren* presents musical aspects of the work that are relevant in the context of the present discussion. For more details, especially on the technical issues, see Christina Anagnostopoulou, Chris Share and Darrell Conklin, 'Xenakis's *Keren* (1986): A

3.1. Pattern discovery in music

Data mining is the field of study in computer science concerned with the discovery of interesting patterns in large databases. Patterns can be important words or word sequences in text databases, significant patterns within long DNA sequences, significant excerpts from films, and so on. In music, patterns usually can be sequences of notes (segments), or sequences of any musical events or structures, such as chords, pitch class sets, any types of segments or any other formal components.

Pattern discovery looks not only for what is repeated in databases, but also for what is considered interesting or significant, according to a set of (usually statistical) criteria. For example, the interval of a major second in a piece or set of pieces might be repeated so often that it becomes trivial, and therefore not significant. An appropriate algorithm should be able to make this distinction, which usually would be achieved by comparing the piece under analysis with other (comparable) pieces. The pattern discovery approach is inductive, in that interesting patterns emerge from the data, using statistical means, and are not known or sought after based on an a priori information (as opposed to pattern recognition).

The general task in musical data mining is shared with that of systematic musicology, namely the discovery of repeated patterns – musical segments, chords, structures, and so on – that stand out in an analysis corpus when related to a comparable group of pieces, thus considered as *significant*. Paradigmatic analysis is the primary type of analysis that makes these discoveries explicit. In this respect, pattern discovery for paradigmatic music analysis purposes seems a natural choice. This topic is further discussed by Lartillot, although he relates it to motivic analysis, emphasizing the salience of a small number of patterns, rather than the discovery of a large pool of repeating patterns in a piece.[17]

Often, interesting and novel facts revealed by pattern discovery, especially in large musical corpora, are detectable only with the use of computational means.

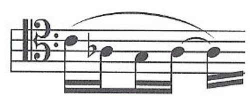
3.2. Knowledge representation

In the computational data ontology, the musical surface of a monophonic piece can be thought of as one long sequence of note units, notated Seq(Note). It can also be thought of a sequence of sequences of notes, which is the segmented piece of music, notated Seq(Seq(Note)). In terms of musical knowledge representation, each note sequence, or segment sequence, can be translated into sequences of parameter values, such as sequences of intervals, contour directions, and so

Computational Semiotic Analysis', in Makis Solomos, Anastasia Georgaki and Giorgos Zervos (eds), *Definitive Proceedings of the International Symposium Iannis Xenakis* (2005), http://cicm.mshparisnord.org/ColloqueXenakis/ (electronic resource, 2006).

17 Lartillot, 'Motivic Pattern'.

on.[18] Figure 11.1a depicts an example of a representation of the first opening notes of *Keren*, with the representation at the note level. Figure 11.1b shows an example of two consecutive musical segments, where the representation is at the segment level.[19]



pitch in MIDI numbers	60	56	55	56
pitch class	0	8	7	8
pitch interval (with previous note)	undef	-4	-1	1
pitch class interval	undef	8	11	1
pitch contour	undef	-	-	+
contour duration	undef	=	=	+

Figure 11.1a The first four notes of *Keren* comprising the opening motif, translated into a number of melodic parameter sequences. Interval and contour viewpoints are undefined for the first element in the sequence

The variety of the chosen musical parameters can be vast: many parameters can be derived from the basic properties of pitch and duration and range from the more concrete to the more abstract ones in the various levels of description.[20] As an example, in Figure 11.1a we depict a contour of durations,

[18] In this chapter the terms *parameter*, *attribute* and *feature* are used interchangeably.

[19] The content of Figures 11.1a, 11.1b and 11.2 is taken from Anagnostopoulou et al., 'Xenakis's *Keren*'.

[20] See examples of parameters in Darrell Conklin and Ian H. Witten, 'Multiple Viewpoint Systems for Music Prediction', *Journal of New Music Research*, 24/1 (1995):

not only pitches. Contour becomes a function that can be applied to any basic parameter. In Figure 11.1b, we use Huron's classifications of melodic shapes as well as a contour or segment lengths, where each segment is compared to the length of its previous one.

set of pitch classes for each segment	[0,7,8]	[0,1,7,8,10]
melodic shape	concave	convex
contour of segment length in notes	undef	+
pitch contour values in segment	[-,+]	[+,-]
segment contour duration	undef	+

Figure 11.1b The first two segments of *Keren*, translated into segmental parametric sequences. The note density contour and the duration contour parameters are undefined for the first segment in the score[21]

3.3. *Pattern discovery in* Keren

There are two components in this analysis of *Keren*. The first part, called simply melodic analysis, looks at the piece as a long sequence of notes, using representations such as the ones in Figure 11.1a, and trying to find repeated and significant patterns of notes. The second part, called segmental analysis, takes as its starting point the segmented score, and looks for interesting patterns of melodic segments, with representations such as those found in Figure 11.1b. The segmentation of the score in this case is taken as given, performed manually following the composer's breath marks on the score.

In both parts the algorithm takes as input the sequences of each parameter in turn and as a first step searches for all repeated patterns within these sequences. For example, in the interval sequence [-1, -3, +3, +4, -1, -3], the pattern [-1, -3] is repeated twice.

pp. 51–73, and in Christina Anagnostopoulou, 'Algorithmic Categorisation in Formal Music Analysis', PhD Thesis, University of Edinburgh, 2001 (supervised by Raymond Monelle).

[21] *Melodic shapes* in Figure 11.1b follows the classification in David Huron, 'The Melodic Arch in Western Folk Songs', *Computing in Musicology*, 10 (1995–1996): pp. 3–23.

However, as pointed above, there might be numerous patterns repeated within a piece, not all of them interesting. A formal way to distinguish between them would be to rank them according to how 'significant' they are, using a statistical framework. Therefore, as a second step, all repeated patterns are evaluated in order to calculate their significance. The way this is carried out is usually by comparison to another piece, or to a set of comparable pieces. In this specific case the comparison piece is a made-up random model piece, using features of *Keren*. The length of a pattern, and its frequency of occurrence compared to the comparison piece, are then taken into account in establishing its significance.

There are two main ways that similarity between two sequences can be captured (see also section 2 above). One is to see what percentage of two musical sequences (e.g. certain notes or intervals) is identical at the surface level ('approximate pattern matching'). The other is to reveal identity at some abstract level of representation. For instance, two musical excerpts are different at the lowest level of the musical surface but they might be identical at some more abstract parametric level (e.g. in their contour or in the pitch pattern between metrically salient notes).[22] In the analysis described in this section we follow this second approach by looking for identities in the abstract representations we choose. The choice of a musical parameter determines the level of abstraction – the more abstract a certain parametric representation is, the further the resulting similarity will be. This allows us to capture distant similarities between patterns, knowing at the same time in which parameter they are identical.

As an example, the repeating sequence [+, -, +, -], [-] of segments was revealed by the algorithm (NB: immediately repeating contour symbols are merged into one symbol, e.g. downward motion [-] in the second segment). This gave rise to observing the following similarity between sequences of two musical segments (see Figure 11.2).

3.4. Discussion

It is clear that patterns found with this approach are not patterns at the musical surface, i.e. of actual notes, but patterns of parameter values, which can be translated into actual notes (or segments). In this respect, the approach described here is not paradigmatic in the sense of Ruwet or Nattiez who in most cases look for patterns of notes. In this way, one can argue that there are various proposed analyses, according to the parameter or parameters chosen. These analyses (or results) are not combined into one final analysis of the musical surface in the way that the approach on the analysis of *Syrinx* described in the next section does – the analysis here is distinctly parametric.

22 See also Emilios Cambouropoulos, Tim Crawford and Costas S. Iliopoulos, 'Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects', *Computers and the Humanities*, 35/1 (2001): pp. 9–21.

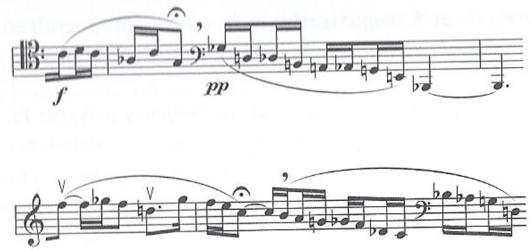


Figure 11.2 Two instances of the segmental pattern [[+,-,+,-],[-]] in *Keren*. Top: segments 3 and 4; bottom: segments 17 and 18, found later on in the piece. Only the new contour directions are captured here

The question whether the statistical significance of certain patterns found by the algorithm necessarily entails musical significance is a debatable one. A statistical measure can take into account certain features that are important in music analysis, such as the length of a pattern (the longer length might point towards something more unique and more interesting), the frequency of occurrence or how over-represented or under-represented it is compared to another piece or set of pieces. However, it would be wrong to assume that this type of mathematical testing can really capture the analytical human judgement. In this respect, the analyst is still an indispensable part of the process, in that he or she needs to evaluate the final results. At the same time, the statistical framework can propose relations and pattern significances that might be hard or impossible to detect otherwise, and in this respect it can be a very important tool for music analysis.

Pattern discovery as a computational technique in music is close to the idea of paradigmatic analysis, in that it can discover segments that are similar in a piece or sets of pieces. On the melodic level, it does not require segmentation of the score. However, the main difference between the two is that in paradigmatic analysis all musical segments are classified into some category, even if they form a category with one instance (whether this is musically interesting or not). In pattern discovery – a type of paradigmatic analysis – only repeated sequences are found, and how much of the musical surface is classified depends on the representation chosen and the strictness of the statistical methodology used.

The repeated segmental patterns found here are considered to be close to a syntagmatic analysis in that we are looking for successions of musical segments and how these are repeated. These successions in turn point towards larger structures in the piece under analysis.

4. Semiotic Analysis as Categorization: Paradigmatic Analysis of Debussy's *Syrinx*

Nattiez provides two paradigmatic analyses of Debussy's *Syrinx* for solo flute; the first consists of segments roughly at the bar level and the second at the beat level.[23]

Below we use a computational model in an attempt to reconstruct Nattiez's second analysis. The goal is to reach results similar to Nattiez's and in this way formalize the process by making his initial criteria explicit. Given a segmentation of a melodic surface, the proposed model initially constructs an appropriate representation for each segment in terms of a number of parameters. These parameters reflect melodic and rhythmic aspects for each surface segment in various levels of abstraction. Then, a clustering algorithm (the *Unscramble* algorithm) is applied for the organization of these segments into 'meaningful' categories or paradigmatic classes. Through a dynamically evolving process the initial set of properties is adjusted so that a satisfactory description is generated. This clustering algorithm automatically determines an appropriate number of clusters, selects the characteristic and defining attributes of each category and allows limited overlapping of clusters.[24]

As has been pointed out above, paradigmatic analysis relies to a large extent on the segmentation of a musical work. The segmentation chosen here is the one provided by Nattiez in his second analysis, which provides a segmentation roughly at the beat level. This is not only because our analysis here focuses on the issue of classification, but also because our goal is to reproduce Nattiez's analysis, therefore using the same given input as a starting point.

4.1. Knowledge representation

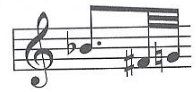
Each segment in the current computational experiment is described by a number of melodic and rhythmic parameters at the musical surface level (grace notes have been omitted in the description of segments). The sets of musical parameters are:

- Various types of pitch intervals: exact number of semitones marked with direction, number of diatonic steps between scale degrees, various types of interval categories (e.g. step, leap), and melodic contour (see Figure 11.3).
- Rhythm: duration in multiples of the smallest unit, inter-onset intervals (time interval between note onsets), and more abstract representations of duration such as describing whether a note is shorter, longer or equal to the preceding note, or duration ratios (ratio between two successive durations).

[23] Nattiez, *Fondements*, pp. 330–354. See also the brief introduction and discussion in Cook, *A Guide*, pp. 151–182, and in Monelle, *Linguistics*, pp. 100–108.

[24] A detailed description of the *Unscramble* algorithm and the approach described here is presented in Cambouropoulos and Widmer, 'Automated Motivic Analysis'.

For instance, the three-note segment depicted in Figure 11.3 can be represented by the parameters shown in the table under it.



Name of Attribute	Musical Parameter: Attribute Explained	Representation for Segment 4
p_sem	pitch: interval in semitones with melodic direction	[-4,+1]
p_ss	pitch: interval in terms of scale degrees with direction	[-3, +1]
p_sl	pitch: step (s) or leap (l) interval with direction	[-l,+s]
p_contour	pitch: melodic direction – contour	[-,+],
r_ioi	rhythm: inter onset interval (multiples of 32nd notes)	[6,1,1]
r_sl	rhythm: relative duration – second note *shorter* (s) than first, and third note *equal* (e) to second note.	[s,e]
r_ratio	rhythm: duration ratio: the second note duration ratio to the first is 1/6, and the ratio of the third to the second note is 1.	[1/6, 1]

Figure 11.3 Seven parameters/attributes for a melodic segment (compare with top segment of column D in Figure 11.4)

The representations used here are comparable to the ones in the previous section on both the melodic and segmental levels, again in various levels of description of the musical surface, from the more concrete to the more abstract ones.

4.2. Musical paradigms

Using these representations, paradigmatic categories are presented as columns of actual melodic segments (exact repetitions of segments have been omitted). Due to space limitations, only the first four main clusters, which account for approximately two-thirds of the segments, are illustrated and discussed. In Figure 11.4 the four clusters (named A, B, C and D) given by Nattiez are presented, as well as some instances of cluster E, also identified by Nattiez.

Figure 11.4 First four main melodic classes (A, B, C, D) given in Nattiez's paradigmatic analysis of Debussy's *Syrinx* (plus class E also presented in the same analysis). The *Unscramble* algorithm renders exactly the same results with the following exceptions: class E is merged with class B (defining attribute: descending melodic steps), the last three members of class C are not placed in this category (see discussion in text), and class D contains one extra segment[25]

[25] See also in Cambouropoulos and Widmer, 'Automated Motivic Analysis'.

The overall clustering produced by *Unscramble* is very similar to the human paradigmatic analysis by Nattiez. More specifically, the main four paradigms produced by *Unscramble* are particularly close to his analysis (represented by the four full columns of Figure 11.4). The algorithm not only discovers these paradigms but also describes them in terms of defining and characteristic attributes. These attributes are the ones that are prominent of the members of a category (a paradigmatic class), compared to the other categories. In computational terms, this is achieved by giving different weights (and therefore different importance) to the various attributes, considering more important those that are shared by many members of the same category, while not shared with members of other categories.

The first paradigmatic class (17 segments including exact repetitions) is identical to Nattiez's class A. Due to the many instances found, this is considered to be the most prominent paradigmatic class of the piece. The algorithm discovers that a defining attribute for this cluster is the *scale-step* attribute: 'step-down, step-up'. The most characteristic attribute is the exact pitch interval attribute: 'minor-2nd-down, major-2nd-up'. In terms of rhythm, the most characteristic attribute, though not as strong, is for *exact inter-onset intervals*: 'dotted-eighth, thirty-second, thirty-second', since it appears 10 times in members of this paradigm and three times in members of other paradigms.

The second paradigm discovered by the algorithm (13 segments) contains all the members of Nattiez's paradigms B and E. The attribute value for this unified paradigm is *descending steps*. This difference occurs because in the current computational experiment there is no distinction between chromatic and diatonic steps at the scale-step abstraction level. For results closer to those achieved by Nattiez, the representation should be refined further to account for chromatic passages.

The third paradigm (14 segments) is the same with Nattiez's class C with the exception of the last three members, which are placed in other categories. The reason is that the last three members contain 'leaps' and that the threshold that was selected by the algorithm as producing the optimal 'goodness' value was somewhat too narrow for them to be included in this category. A richer representation in terms of a broader range of attributes may provide a more relaxed threshold that would allow these segments to be included as well. Defining attribute for this cluster is: 'step-up, step-down'. The most characteristic rhythmic attribute is for the inter-onset interval ratio attribute (not very strong). Inclusion of the grace notes should strengthen this category.

The fourth paradigm (five segments) includes an additional segment (numbered 26). It is not clear why Nattiez did not include this segment in this paradigmatic class – he has actually stretched this segment placing it in between two clusters (the first two notes in one class and the third note in another class). The defining attribute for this cluster is: 'leap-down, step-up'.

Quite different results arise if more initial prominence is given to the rhythmic attributes. For instance, the first segment of cluster C may be placed in cluster A or the two clusters may overlap over this segment.

The *Unscramble* algorithm provides a very useful computational method for studying the task of generating a 'meaningful' paradigmatic analysis of a musical work, as the whole analytic process has to be explicit and consistent. The commonalties and differences between the human and machine analyses may highlight possible additional intuitive musical knowledge that the analyst has used during the analytic process, differences in the importance (weights) of the various parameters, and/or shortcomings of the computational method. Reproducing closely Nattiez's own results shows that the criteria used for the present analysis in terms of attributes and weights, were very close to his. However, determining an adequate initial set of attributes requires further research.

5. Concluding Remarks

In this chapter we investigated some aspects of formalizing semiotic analysis in terms of computational modelling. We discussed related challenging issues, and we presented two models, describing different ways of approaching the issue. The first model is related to pattern discovery, while the second one is related to a classification of segments of the musical surface into paradigmatic classes.

There are two questions that might be brought up at this stage. The first one is, why this type of analysis at all? It is often argued that this type of analysis yields trivial results, perhaps by those who do not believe in the possibility of scientific and formal rigour in music analysis, the way Nattiez suggested. The aims of computational analysis, however, might be slightly different from those of traditional music analysis, in that formalizing a process and creating an analysis based on scientific methods can be an important task per se.

The aim of the analytic procedures offered here is not to reveal or examine the compositional processes that either Debussy or Xenakis might have used, implicitly or explicitly, in their works. Neither is it primarily aimed to detect what a listener might perceive based on auditory principles. Rather, the proposed computational analytic methodology can be viewed as a first step to further develop musical analysis and interpretation. It can assist a music analyst in the analytic exploration of a broad universe of possibilities. Additionally, it may also contribute to the further investigation of aesthesic or poietic factors, such as highlighting potential compositional relations that a composer may have employed or to suggest similarity relations that may be investigated further via cognitive modelling and empirical psychological experiments.

The second question which arises is, can paradigmatic or segmentational analysis be fully automated? Computational models may automatically produce musical results (nowhere near the analyses a competent analyst may produce, at this stage). However, since the aim of such an analysis is to involve creative investigation and exploration of various structural relations of the musical material, and since no single 'correct' paradigmatic or segmentational analysis for a given piece exists, the human analyst remains central in the whole process. Computational music

analysis becomes meaningful when a human analyst supervises the whole process, guides the analytical choices, and interprets the results. Analyst and computer can work together to produce interesting and novel results, some of which might have been impossible to detect manually.

One can argue that computational music analysis might come slightly closer to what Nattiez had originally envisaged regarding analysis at the neutral level of music, in that one could achieve 'objectivity', and formalize the analytical process, in relation to specific criteria that are made explicit. However, while working on the neutral level, an analyst has an intuitive perception and interpretation of the piece and of the analytical method to be followed, and makes choices that cannot truly be considered 'neutral'. Analysts thus work on their own poietic level. Although the intention might be 'scientific objectivity', one also makes subconscious or intentional analytical choices; this is the analytical freedom discussed above, which is in our opinion necessary, and which supports the diversity inherent in music analyses.