Chapter 6

Microstructural Module
(Local Boundaries, Accents & Metre)

Introduction

In this chapter a general model will be introduced that allows the description of a melodic surface in terms of local grouping, accentuation and metrical structures. Firstly, a formal model will be proposed that detects points of maximum local change that allow a listener to identify local perceptual boundaries in a melodic surface. The Local Boundary Detection Model (LBDM) is based on rules that relate to the Gestalt principles of proximity and similarity. Then it will be shown that the local accentuation structure of a melody may automatically be inferred from the local boundary grouping structure. This is based on the assumption that the phenomenal accents of two contiguous musical events are closely related to the degree by which a local boundary is likely to be perceived between them. Finally, the metrical structure is revealed by matching a hierarchical metrical template onto the accentuation structure. It is suggested that the Local Boundary Detection Model presents a more effective method for low-level segmentation in relation to other existing models and it may be incorporated as a supplementary module to more general grouping structure theories. The rhythmic analyses obtained by the methods described herein are tentative, and complementary to higher-level organisational models (see chapters 7, 8 & 9).
6.1 Musical Rhythm

Many contemporary theories of rhythm (Cooper and Meyer, 1960; Epstein, 1995; Lerdahl and Jackendoff, 1983; Kramer, 1988; Yeston, 1976) consider rhythm to be the organisation/structuring of musical sounds into groups (grouping structure) of more or less salient elements (accentuation structure) that are in constant interplay/interaction with a hierarchy of beats (metrical structure). Metre receives somewhat different treatment in each of these theories and is to a varying extent integrated into the ways rhythm is defined (Moelants, 1997).

For instance, Lerdahl & Jackendoff's (1983) definition of rhythm is based on two kinds of structures: namely grouping structure that 'expresses a hierarchical segmentation of a piece into motives, phrases and sections' (p. 8) and metrical structure that 'expresses the intuition that the events of a piece are related to a regular alternation of strong and weak beats at a number of hierarchical levels' (p. 8). They define three kinds of musical accents: phenomenal accents which are due to local intensification such as dynamic stress, high or low register, long notes, harmonic changes and so on, structural accents which result from higher-level structural relations such as cadences, and metrical accents that correspond to relatively strong beats in a metrical context. Defining a metrical structure is finding a well-formed grid of metrical accents that fits best onto the structure of phenomenal accents: "... the listener's cognitive task is to match the given pattern of phenomenal accentuation as closely as possible to a permissible pattern of metrical accentuation. ... Metrical accent, then, is a mental construct, inferred from but not identical to the patterns of accentuation at the musical surface." (p.18). In their theory, grouping structure is considered to be independent of metrical structure and hence different preference rules are formulated for each: one set of preference rules for the description of groupings and a different independent set for the description of phenomenal accentuation structure from which metrical structure is inferred (see figure 6.1a).

The concept that rhythm relates to cognitive grouping of musical events is a Gestalt-based one. The Gestalt principles of perceptual organisation are a set of rules-of-thumb that suggest preferential ways of grouping mainly visual events into larger scale schemata. Two of the Gestalt principles state that objects closer together (Proximity principle) or more similar to each other (Similarity principle) tend to be perceived as groups. These principles have been used as a basis for some contemporary theories of musical rhythm. Tenney (1964) discusses the use of the principles of proximity and similarity as a means of providing cohesion and segregation in 20th century music and, later, Tenney & Polansky (1980) develop a
computational system that discovers grouping boundaries in a melodic surface. Musical psychologists (Bregman, 1990; Deutsch, 1982a,b; McAdams, 1984) have experimented and suggested how the Gestalt rules may be applied to auditory/musical perception and Deutsch & Feroe (1981) further incorporate such rules in a formal model for representing tonal pitch sequences. The grouping component of Lerdahl & Jackendoff’s Generative Theory of Tonal Music (1983) is based on the Gestalt theory and an explicit set of rules is thereby described - especially for the low-level grouping boundaries (the formulation of these rules has been supported by the experimental work of Deliège (1987)).

![Diagram](image)

Figure 6.1  a. Lerdahl & Jackendoff's theory of musical rhythm  
b. Proposed model of musical rhythm

In the first part of this chapter a systematic theory will be described that attempts to define local boundaries in a given melodic surface. The proposed segmentation model (Local Boundary Detection Model - LBDM) will be based on two rules: the Identity-Change rule (which is more elementary than the Gestalt principles of proximity and similarity) and the Proximity rule (which relates to the Gestalt proximity and similarity principles). The aim has been to develop a formal theory that may suggest *all* the possible points for local grouping boundaries on a musical surface with various degrees of prominence attached to them rather than a theory that suggests *some* prominent boundaries based on a restricted set of heuristic rules. The discovered boundaries are only seen as potential boundaries as one has to bear in mind that musically interesting groups can be defined only in conjunction with higher-level grouping analysis (parallelism, symmetry, etc.). Low-level grouping boundaries may be coupled with higher-level theories so as to produce 'optimal' segmentations (see figure 6.2).
It will be shown that the formulation of the boundary discovery procedures defined by Lerdahl & Jackendoff (1983) and Tenney & Polansky (1980) have limitations and can be subsumed by the proposed theory. Some examples and counter-examples will be given in relation to the influential formulation of the local detail grouping preference rules - mainly GPR 2 & 3 - by Lerdahl & Jackendoff.

In section 6.4 it will be maintained that low-level grouping structure and phenomenal accentuation structure are strongly associated in such a way that if one is defined then the other may automatically be inferred. In other words, if local boundaries for a given melodic surface have been defined then strengths for phenomenal accents may be inferred (the reverse is also possible although not examined in this thesis). It is assumed that the phenomenal accents of two contiguous musical events are closely related to the degree by which a local boundary is likely to be perceived between them. A method then is described that mechanically derives accent strengths from the local boundary strengths detected by LBDM.

The strong link between grouping and accentuation structures is important in that it allows one to develop a model that does not need two separate independent methods for the detection of the local boundaries and the phenomenal accents respectively. In contrast with Lerdahl & Jackendoff's model (figure 6.1a) the proposed model directly links phenomenal accentuation structure with grouping structure (figure 6.1b). This enables a more economic and efficient formulation of a theory for rhythm.

Once the phenomenal accentuation structure has been defined an attempt can be made to match a well-formed metrical structure to it; this may be possible for a number of hierarchic metric levels of beats or only for one level or possibly for no level at all depending on the kind of music. Metrical structure may be inferred from the accentuation structure but, at the same time, it influences the perception of the accentuation/grouping structure. The interplay between these two kinds of structures is addressed further in section 6.5.
In the following sections, formal methods will be described, firstly, for the discovery of local boundaries (low-level grouping structure) in a melodic surface, secondly, for the derivation of the phenomenal accentuation structure from the grouping structure and, lastly, for the selection of a metrical structure that fits best onto the accentuation structure.

6.2 The Gestalt principles of proximity and similarity in theories of rhythm

Some problems in the way the low-level Gestalt principles of perceptual organisation have been applied in the organisation of temporal musical sequences are briefly discussed below.

The Gestalt principles of proximity and similarity have been applied in both Tenney & Polansky's and Lerdahl & Jackendoff's models in such a way as to allow one to interpret them as being different descriptions of the same phenomenon, namely a local maximum in the distance between consecutive musical events for any musical parameter, e.g. pitch, start-times, dynamics and so on. Tenney and Polansky (1980) state explicitly that the similarity principle - as they define it - actually includes the proximity principle as a special case: "In both, it is the occurrence of a local maximum in interval magnitudes which determines clang-initiation" (p. 211). Lerdahl & Jackendoff's (1983) grouping rules are defined in such a way that it seems rather plausible that the proximity rules can be subsumed by the change (similarity) rules and the reverse. For example, GPR3a (register rule) states that a greater pitch interval in between smaller neighbouring intervals initiates a grouping boundary. This can been seen in two ways: a) that the pitches of the first and last intervals are more similar to each other than the pitches of the middle interval or b) that there is a greater proximity between the first two pitches - and the last two - rather than between the middle pitches (see Handel, 1989:198).

It is herein maintained that although this formalisation of the Gestalt principles provides the most important factor for discovering local boundaries a more general approach should account for any change in interval magnitudes. For example, in the following sequence of durations: ♩♩♩♩♩♩♩ a listener easily hears a possible point of segmentation for which neither the Tenney & Polansky nor the Lerdahl & Jackendoff formalisms suggest any boundary. For this reason a different, more elementary rule will be introduced based on the principle of Identity-Change. This issue will be discussed further in the next section and it will be shown that the above example can naturally be accommodated within the proposed model.
The low-level Gestalt principles of proximity and similarity are usually applied on symmetrical non-directional spaces. On applying them to musical temporal spaces, one has to make certain concessions by removing all possible asymmetrical directional properties (e.g. direction of pitch-intervals). There is though one aspect of musical asymmetry that cannot be avoided. This relates to the fact that musical objects are asymmetric objects themselves - even the most simplified homogeneous description of a note distinguishes between its attack and the rest of its body. This asymmetry is reflected in that, for instance, the temporal grouping rules can never give an identical grouping structure to the original and the retrograde form of a melody. It relates to the way that rules of perceptual organisation give different grouping boundaries for musical duration sequences and for start-time interval sequences. It will be shown below how the interaction between these duration and start-time interval groupings results in the asymmetric perceptual organisation of a sequence of musical events.

We will now attempt to define the Identity-Change rule and the Proximity rule which will form the basis of the LBDM. These rules will be discussed initially for any sequence of two or three objects and then will be applied to longer sequences of musical objects.

6.3 The Local Boundary Detection Model (LBDM)

A formal model that attempts to determine local boundaries in a given melodic surface will now be presented.

6.3.1 The Identity-Change and Proximity Rules

As we have seen above, the Gestalt principles of proximity and similarity can be interpreted as being different sides of the same coin. In the Local Boundary Detection Model (LBDM) an elementary rule will be introduced based on the principle of identity. The Identity-Change rule is more elementary as it can be applied to a minimum of two entities (i.e. two entities can be judged to be identical or not) whereas the Proximity/Similarity rule requires at least three entities (i.e. two entities are closer or more similar that two other entities). This Identity-Change rule, in conjunction with the Proximity rule, forms the basis of the proposed low-level segmentation model.

**General Identity-Change Rule:** Grouping boundaries may be introduced only between two different entities. Identical entities do not suggest any boundaries between them.
This rule is supported by an experiment realised by Garner (1974) wherein an eight-element pattern composed of two different pitch elements, for example XXXOXOOO, is looped indefinitely and listeners are asked to describe the pattern they perceive. Various preferential ways of organisation were recorded (there are eight possibilities starting on each element of the sequence) but hardly ever did any listener break a run of same elements.

If the entities compared are *intervals* (intervals for pitch, start-times, dynamics, etc.) then this rule can be formulated more specifically:

**Identity-Change Rule (ICR):** Amongst three successive objects, boundaries may be introduced on either of the consecutive intervals formed by the objects if these intervals are different. If both intervals are identical no boundary is suggested.

When the application of ICR on two consecutive intervals detects a change and suggests a local boundary, this boundary is ambiguous (i.e. the boundary can be placed on either side of the middle object) and each interval receives the same boundary strength value. The second rule (PR) resolves the ambiguity by giving preference to the larger of the two intervals.

**Proximity Rule (PR):** Amongst three successive objects that form different intervals between them, a boundary may be introduced on the larger interval, i.e. those two objects will tend to form a group that are closer together (or more similar to each other).

### 6.3.2 Applying the ICR and PR rules on three note sequences.

We will assume that for each parametric feature of a musical surface we can construct a sequence of intervals on which the ICR and PR rules may be applied. We will start by presenting the application of the rules to the following parameters: pitch, dynamics, rests and articulation (slurs, staccatti, breath-marks etc. are considered to be expressional rests and are inserted between the notes they mark as normal rests that have a value that is a fraction of the preceding note). The grouping boundaries resulting from the sequence of start-time intervals and durations will be presented at the end of this section.

The relation between two intervals can be of two types: *identity* or *change*. For reasons of asymmetry that will be introduced later on we will depict the change relation in two directional forms: '+' and '-' (figure 6.3 b,c). In the following figures, dots represent parametric values of musical events and the distances between the dots the interval sizes between these
values (Dx, Dy are interval values and are placed at the left-hand side of the interval). In figure 6.3a Dx=Dy and the identity relation is represented by a zero. In figure 6.3b Dx>Dy and in figure 6.3c Dx<Dy, and the change relations are represented by the '+' and '-' signs respectively.

At this stage we will introduce numeric values for the strength of the ICR and PR rules (more research is necessary for the selection of the most appropriate values). A numeric value is given to each interval as indicated below:

**ICR:**
- 0 for the identity relation (0 for each interval)
- 2 for the change relation (1 for each interval)

**PR:**
- 0 for the identity relation (0 for each interval)
- 1 for the change relation (1 for the larger interval)

We get thus the total interval boundary strengths as depicted in figure 6.3 (bottom line).

We can now examine the duration and start-time interval sequences. The duration of a musical note is an internal attribute of that note whereas start-time intervals are temporal distances between two different successive events. We have thus the application of the ICR and PR rules for the start-time intervals exactly as described above and, additionally, the application of the General ICR for the sequence of durations (numeric strength 2). We now have the following kinds of relations for two start-time intervals delimited by 3 start-time points (dots) and the two corresponding durations (rectangles) (figure 6.4).

**Figure 6.3** Boundary strengths (last row) calculated by the use of the ICR and PR rules for three parametric values (e.g. pitch, dynamics etc.) separated by two intervals.

**Figure 6.4** Boundary strengths (last row) calculated by the use of the ICR and PR rules for three start-time values separated by two start-time intervals and durations.
It is now clear that the '+' and '-' change relations are not symmetric. It is not possible to apply
the principles of perceptual organisation in the musical temporal domain without introducing
local asymmetry.

6.3.3 Applying the ICR and PR rules on longer melodic surfaces

For a given parametric interval profile of a musical surface one finds all the kinds of interval
relations (0, +, -) that exist between every two successive intervals. If there are 3 or more
consecutive '+' or '-' relations (e.g. +++, - - - - -), then only the ones at the ends are considered -
the others do not contribute to the numeric strengths. Then, the numeric strengths for each
kind of relation are calculated and added for each interval. For a single numeric strength
sequence the local maxima suggest the most preferable local boundaries (when a local
maximum consists of more than one same or almost the same values then an ambiguous
boundary is suggested).

In figure 6.5 we give a first example of how one can use the ICR & PR rules to calculate the
strengths of grouping boundaries for -+ sequences. As it happens, almost all of the grouping
preference rules 1 of Lerdahl & Jackendoff (1983), and all the grouping rules suggested by
Tenney & Polansky (1980) fall under the -+ category of sequences - see figure 6.7 for the
application of the LBDM rules to the local detail examples of Lerdahl & Jackendoff's grouping
theory. The formulation of the boundary discovery procedures defined by Tenney & Polansky
and Lerdahl & Jackendoff are specific instances of the proposed theory.

Figure 6.5 Examples of boundary strengths (last row) determined by the LBDM.

1 Exception: GPR3d (equal note length) and the articulation changes from legato to staccato and the
opposite, fall under the 0 + 0 and 0 - 0 combinations

90
Figure 6.6  Examples of boundary strengths (last row) determined by the \textit{LBDM}. These are ambiguous boundaries which may be resolved if higher-level organisational principles are taken into account.

The boundaries in the examples of figure 6.5 are detected by Tenney \& Polansky's and Lerdahl \& Jackendoff's methods whereas their models do not suggest any boundaries for the examples in figure 6.6. By contrast, the \textit{LBDM} suggests ambiguous boundaries for all the examples of figure 6.6 (such ambiguous boundaries may be resolved if higher-level grouping organisational principles are taken into account).
Figure 6.7 Application of the Local Boundary Detection Model to the Lerdahl & Jackendoff (1983:44-46) local detail grouping examples 3.14-3.17. For the examples not accounted for by the GPR2 and GPR3 rules, the proposed theory suggests ambiguous boundaries (depicted as $\wedge$).

The above procedure is realised for every parametric interval profile of interest. Then the total sum of all the numeric strength sequences is calculated (weighted or not). The local peaks are the points in a melodic sequence in which boundaries may preferably appear. In figure 6.8 the preferred grouping structure is presented for Mozart's opening of the *Symphony in G min*. The boundary strengths for each parametric interval profile are calculated and then added to produce the total boundary strength sequence $A$. Sequence $B$ is given by a refined version of $LBDM$ which takes in account the degree of difference between two intervals and other factors discussed in section 6.3.5.
Figure 6.8  Low-level grouping structure for the theme of Mozart's *Symphony in G min.* The boundary strengths sequence A is determined by the *LBDM* whereas sequence B is determined by the refined version of *LBDM* described in section 6.3.5 (slurs are not taken into account).

*LBDM* has been successfully applied to many kinds of melodic surfaces - from traditional tonal melodies to contemporary atonal surfaces - such as the song Frère Jacques (figure 6.9), the beginning of J.S.Bach's *Concerto for Harpsichord in D min.* (figure 6.10), an excerpt from Xenakis' *Keren* (figure 6.11) and an excerpt from Stravinsky's *Three pieces for solo clarinet,* no. III (figure 6.12). This method can be further enriched if, for example, harmonic chord distance or scale-degree tonal distance profiles of the melodic surfaces are incorporated.
Figure 6.9: Accentuation and metric structure for the song *Frère Jacques*.

Figure 6.10: Accentuation and metric structure for the beginning of J.S. Bach's *Concerto for Harpsichord in D min*. 
Figure 6.11 Suggestion of local boundaries for an excerpt from Xenakis' Keren for solo trombone. Boundary strength sequence A has been calculated without taking into account slurs; sequence B with slurs. Note that the grouping structure proposed by the composer happens to be very close to the boundaries suggested in sequence A; on the contrary, the Lerdahl & Jackendoff local boundaries are very weak without GPR2a (slur/rest).

Figure 6.12 Suggestion of local boundaries for an excerpt from Stravinsky's Three pieces for solo clarinet, no. III. Sequence A (without articulation marks) and sequence B (with articulation marks) indicate the most probable local boundaries calculated by the refined version of LBDM (see section 6.3.5 for further details).
6.3.4 Further comments on the application of the LBDM rules

- Most formal grouping theories define exclusively clear boundaries that appear unambiguously between two musical events. However, there are cases where a boundary is ambiguously suggested. This phenomenon is conveniently accommodated within the present theory wherein numeric peaks with two identical or similar values suggest a blurred boundary (higher level grouping mechanisms may support one interpretation over other possibilities). Deliège (1987) suggests that in the following sequences (figure 6.13) the grouping boundary perceived by listeners tends to appear after the first half-note and staccato note respectively. The current theory suggests an ambiguous boundary on those notes.

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
& & & & & & & \\
L&J rules: & 3d & 3d & 3d & 3d & 3d & 3d & 3c \\
& & & & & & & \\
& 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
& 0 & 0 & 3 & 2 & 0 & 0 & 1 & 0 & 2 & 0 \\
\end{array}
\]

Figure 6.13

- It may be preferable in some cases to use subjective scales for interval sizes instead of acoustic ones. For example, in the following series of equally timed elements (figure 6.14) the ones that are more intense tend to be perceived as beginnings of groups (Handel, 1989:386-389). In other words, it may be said that the interval \( p \to f \) is larger than the reverse \( f \to p \). The sequence below will have the following grouping boundaries:

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{physical intervals} & f & p & f & p & f & p & f & p \\
& 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
& 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

(no boundaries)

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{subjective intervals} & f & p & f & p & f & p & f & p \\
& 2 & 3 & 2 & 3 & 2 & 3 & 2 & 3 \\
\end{array}
\]

Figure 6.14

- Deliège (1987) suggests that a change in melodic contour contributes weakly towards the establishment of a local boundary. This may be incorporated in the current theory by detecting changes of contour of the form 0 * 0 (e.g. U U D D) and at the point of change applying the ICR rule - 1 numeric value for each interval (figure 6.15a).
Deliège (1987:353) reports that the analysis of the responses of listeners to the change of the melodic contour 'revealed a preference for cutting before the pivot sound.' Taking this observation into account it would seem plausible to give an extra numeric weight at the first interval (figure 6.15b).

### 6.3.5 The Refined Local Boundary Detection Model

The LBDM can be enhanced in various ways so as to accommodate further nuances of musical perception that contribute towards a more accurate description of the low-level grouping structure of a musical surface. Some of these are described below:

1. The various parametric profiles may be given different weights depending on the degree of prominence they may have for a given melodic surface. If, for instance, start-time intervals are considered more important, then the start-time profile may be given a higher weight factor before it is added to the other strength profiles.

2. The numeric value of the PR rule may be augmented (e.g. have a value of 2). This will produce sharper local maxima.

3. The $0, +, -$ identity/change relations may be refined by taking into account the ratio/difference between two interval sizes (factor $\alpha$ – this may be calculated using a function such as $\alpha = \frac{\left| x-y \right|}{x+y}$ where $x, y$ are positive integer interval sizes, and $0<\alpha<1$). As Deliège (1987:328) points out, the sensation of a boundary is strengthened in correspondence to the

---

2 If the absolute value of an interval is 0 (e.g. repeated pitches) it is replaced by an arbitrary non-zero value smaller than the interval unit of measurement (e.g. for pitch: half semitone i.e. 0.5); this way a zero denominator for the factor $\alpha$ formula is avoided. Alternatively, the algorithm could check for the case where both intervals are 0 and force $\alpha=0$. 

---

97
increase in difference between two intervals. For example, the second of the following two sequences suggests a stronger boundary:

\[
\begin{array}{c}
\downarrow&\downarrow&\downarrow&\downarrow&\uparrow&\uparrow \\
\end{array}
\]

4. A further factor that contributes to the perceived strength of a boundary relates to the total sum of the two intervals; the larger the sum is, the greater the prominence of the perceived boundary (factor $\beta$ - this may be calculated using a function such as $\beta=1-1/(x+y)$ where $x$, $y$ are positive integer interval sizes and $0<\beta<1$). For example, the second of the following two sequences suggests a stronger boundary:

\[
\begin{array}{c}
\uparrow&\uparrow&\downarrow&\downarrow&\downarrow&\downarrow&\downarrow&\downarrow \\
\end{array}
\]

A refined version of the LBDM has been devised that takes in account suggestions 1, 3 and 4:

For each interval of a specific parametric profile, factor $\alpha$ is calculated for this and the next interval, and this value is multiplied with the absolute size of the current interval (and the next interval); then the second value that had been calculated for the preceding two intervals is also added to the value of the current interval. This process is applied to each interval of the parametric profile; when the process is complete the calculated values are normalised (from 0-100). Finally, the strength values for each parametric profile are averaged (weighted or not) and the overall local boundary strength profile is obtained. The refined LBDM has been applied on a number of melodic surfaces - see examples illustrated in figures 6.8, 6.12, 7.8, 9.1, 9.8, 9.12.

For the theme of Mozart's G minor Symphony (figure 6.8) it is clear that the middle and last boundaries are more prominent and could be considered as best candidates for higher level groupings (actually, these boundaries would emerge if the second-order local maxima were selected i.e. the maxima of the first-order maxima). This is a rather interesting result, especially if one bears in mind that no higher level organisational principles have been employed (e.g. symmetry, parallelism).

---

3 Factor $\alpha$ encapsulates the degree of change/difference between two successive intervals (refined version of ICR rule). By multiplying factor $\alpha$ with the absolute size of each interval the change strength value of factor $\alpha$ is distributed according to the size of each interval, i.e. the largest interval receives a stronger boundary value (refined version of PR rule); at the same time, suggestion 4 (see above) is also satisfied without the use of a factor $\beta$ function.
A second example is given for an excerpt from the 3rd piece from the *Three pieces for solo clarinet* by I.Stravinsky (figure 6.12). Lerdahl and Jackendoff apply their grouping preference rules on the beginning of the 1st of these pieces to show that the grouping component of their theory is general and style-independent. However, if a different excerpt from this set of monophonic pieces (figure 6.12) is examined the local boundaries proposed by Lerdahl and Jackendoff show limitations in two respects: firstly, not all the perceptually significant points of segmentations are accounted for (see, for example, the third grouping boundary - after the 10th note); secondly, many points are given excessive grouping boundary importance (see, for example, the second half of the excerpt in which strong GPR 2a & 2b boundaries are placed on every rest). On the contrary, the refined version of LBDM gives a more integrated account of the possible local boundaries (the peaks of the boundary strength sequence A suggest boundaries which correspond closely to the composer's articulation marks).

The refined LBDM encompasses facets of similarity more effectively as it accounts for the degree of difference between two intervals. The refined LBDM may be incorporated in real-time systems that attempt to segment input musical data. If, for instance, two input durations are almost the same - but not identical - factor $\alpha$ will tend to become zero so this slight performance difference will not contribute towards the establishment of a boundary (there is no need for quantisation of musical parameters before segmentation). It can also cope with the longer strings of only + or - change relations (e.g. ++++) in a more refined manner because these changes will receive different strengths according to their relative factor importance.

### 6.4 Phenomenal Accentuation Structure

It is herein maintained that local grouping and phenomenal accentuation structures are not independent components of a theory of musical rhythm but that they are in a 'one-to-one' relation, i.e. accentuation structure can be derived from the grouping structure and the reverse. If, for instance, one develops an elaborate model of local grouping structure (such as LBDM) then, from this, the accentuation structure can automatically be inferred. This hypothesis is fundamentally different from much common practice whereby one set of rules is given for the detection of grouping boundaries and a different set for the determination of accents of musical notes.

The above hypothesis is based on the observation that group boundaries are closely related to the accented/salient events between which they occur. A perceived boundary in a given continuum indicates that the elements that delimit it are more prominent than other events
further away. Epstein (1995) states: "Demarcation in effect means emphasis - the emphasis required at that moment when a border of some time segment is to be delineated" (p.24).

In figure 6.16 the local boundary strengths are given according to the Local Boundary Detection Model. It is hypothesised that if the boundary strength values are added for every two successive intervals the local accentuation structure of the surface is revealed. The local maxima in this sequence of accent strengths indicate the elements in the surface that are perceived as being more prominent. In particular, the events delimited by two approximately equal local boundary values (e.g. figure 6.16d) are considered to be most salient, i.e. an element that is preceded and followed by a significant boundary indication (ambiguous boundary) tends to be unambiguously highlighted into perception.

![Figure 6.16 Examples of phenomenal accent strengths derived from the LBDM boundary strengths by merely adding every two adjacent boundary strength values.](image)

Figure 6.16 Examples of phenomenal accent strengths derived from the LBDM boundary strengths by merely adding every two adjacent boundary strength values.

For the cases where the two events delimiting a boundary receive equal (or almost equal) accent strength values (figure 6.16c) there is a general tendency to consider the element that initiates a group as more intense although there are cases where this isn't true (see Handel, 1989, chapter 11). As the proposed formal model is considered merely to be complementary to other higher-level organisational factors (e.g. metre, parallelism, symmetry, learned structural schemata etc.) these ambiguities are left unresolved at this low level. For example, a given metrical context for the melodic excerpt of figure 6.16c may assist in resolving the ambiguity by adding metrical accent to one or the other of the two accented notes.

The accentuation structure has been calculated for a variety of melodic surfaces and has produced rather reliable results. In figures 6.9 & 6.10 the accentuation structure is presented for two melodic examples. The local maxima - and the relatively large numeric strengths - indicate the most accented events. Note that most of the strong accents correspond to events that a listener may perceive as most prominent and that the ones that may be considered counter-intuitive (e.g. accent on the 4th and 8th quarter-note of Frère Jacques) are due to the fact that metrical accents and higher-level principles of organisation have not been taken into
account (especially for Frère Jacques, parallelism/repetition plays a paramount role in the determination of grouping structure - see section 7.7).

In the next section it will be shown that the rudimentary phenomenal accentuation structure revealed with the help of the simple mechanism described above may be sufficient for the derivation of the corresponding metrical structure - whenever such a metrical structure does exist. This further supports the validity of the proposed method for determining accentuation structures.

6.5 Metrical Structure

Musical time is structured around a cognitive framework of well-formed hierarchically ordered time-points (at least for metric music). Metrical structure is an abstract system of reference that facilitates the structuring of sequentially emitted/received musical events (Clarke, 1987).

A metrical structure consists of a number of levels of steady patterns of beats (the beat level at which listeners might tap their foot or clap their hands will be referred to as the tactus). The simplest and most 'natural' tactus is when beats are separated by equal time-span units and are delivered at a rate in the neighbourhood of 1.7 beats/sec (not much slower than 1 beats/sec, not much faster than 4 beats/sec) (Handel, 1989). It is possible though to have a tactus where beats are separated by non-regular time-span units as in much of the traditional music of the Balkans (e.g. dance songs in 7/8 metre are usually danced/clapped at $1^{1/2}:1:1$ beat time-span ratios). Time-spans between beats may be further divided into smaller units down to the elementary unit or 'fastest pulse' (Seifert et al., 1995). Above the tactus, beats may be organised into larger measures (usually in regular binary/ternary patterns) and, often, into even larger hypermeasures. In figure 6.17 some well-formed metrical structures are presented. It should be noted, though, that some music doesn't have metric structure at all (e.g. much contemporary music) or only a tactus without higher-level metrical hierarchies (e.g. much of African music - see Arom, 1991).

A metrical hierarchic grid may be matched onto the accentuation structure of a musical piece - more on template-matching models in (Parnutt, 1994). It is asserted that if the grouping/accentuation structure of a piece has been defined then the most appropriate metrical structure may be induced. But, conversely, the metrical structure - once a listener has made a selection - strongly influences and resolves ambiguity in the grouping/accentuation structure.
Metrical accents are added onto the accentuation strengths and thus regulate the grouping structure of a piece. Metre is not simply a mental artefact induced from the music but actually has an autonomous psychological existence that is developed within a cultural context and influences actively the way music is performed/perceived - see Clarke (1985) for an experiment that highlights the influence of different metrical frameworks on the performance of the same melody.

Figure 6.17 Examples of well-formed metrical grids.

Let us examine now how a metric grid may be matched onto a given accentuation structure. The total accent strength that corresponds to a given metric grid can be calculated by adding the accents of all the events whose inception coincides with the points of the grid. If between different positions/displacements of a metric grid one finds a 'significantly' greater total value, then this is considered to be the best fit. If the various placements of a grid receive similar values, then metrical ambiguity is suggested as to that grid. Computational models of the perception of metre - mainly for plain sequences of inter-onset intervals - are described in (Lee, 1991; Longuet-Higgins and Lee, 1982, 1984; Povel and Essens, 1985; Rosenthal, 1992; Steedman, 1977).

The two examples presented above (figures 6.9 & 6.10) are taken from the Western metric tonal musical tradition, so we would expect that a regular metre of binary/ternary beat patterns would be appropriate (figure 6.16a,b). For both of these examples we consider that the tactus appears at the quarter-note durational value (depending on the tempo). A discussion on the metrical structures of these two melodies is presented below.
In figure 6.9 we see that at the half-note metric level the total accent strength (indicated at the end of each metric grid) of the binary grid that starts on the first note is much stronger than that of the one that starts on the second quarter-note. This agrees with the metrical perception listeners have and the way metre is indicated on the score. Ternary metrical grids do not suggest any strong preferences (and obviously parallelism considerations would immediately rule them out). Once a binary grid is established, we can examine the next metric level of a whole-note grid. There is no strong preference (there is ambiguity) between the two possible arrangements although the one that starts on the third note is slightly preferred, i.e. if articulation and the song word prosody are not taken into account the structure of the piece suggests a gavotte-like metre (bar-lines shifted to the right by two quarter-note beats). Interestingly enough, the prosodic structure of the Greek version of the song adheres to this alternative metrical structure.

The first six bars of Bach's *Concert for Harpsichord in D min.* (figure 6.10) are already ambiguous at the tactus; the metrical structure becomes clear only after the seventh bar. The quarter-note beat grid that starts on the first note and the one that starts after an eighth durational value have almost the same total accent strengths (the ambiguity is maintained at the half-note level as well). The first two notes are heard as an upbeat and the listener makes a first selection of a metrical structure that considers the 3rd, 5th and 7th notes as metrically stronger. This assumption is overturned in bar 2 - where the metrical grid is in-phase with the indicated metre on the score - and the beginning of bar 3 is perceived as a suspension. But as more information arrives there is a tendency to shift the metre again and place strong metrical beats on the 'syncopated' notes. The section that comprises sixteenth notes is metrically ambiguous. The second half of bar 5 and the first half of bar 6 suggest a metrical structure that conforms with the metric grid that is displaced/shifted by an eighth-durational value. From the second half of bar 6 onwards the metrical structure becomes clear matching the metre indicated in the score. In figure 6.10 (top) the melody has been segmented in such a way that the accentuation strength difference in each segment is maximised for the two alternative positions. This metrical analysis seems to correspond to the metrical ambiguity that the composer has intentionally implanted in the melodic surface and that is perceived by the listener.

---

4 A more integrated analysis should also take into account the implied harmony and polyphony.
Conclusion

In this chapter a formal theory for the low-level rhythmic description of a melodic surface was presented. The *Local Boundary Detection Model* is based on the Identity-Change and Proximity rules and detects points of maximum change that allow a listener to identify local boundaries in a melody. This model is more general than either Tenney & Polansky's (1980) or Lerdahl & Jackendoff's (1983) grouping models, it can easily be implemented as a computer program and may readily be incorporated as a supplementary module to higher-level theories of rhythmic organisation.

It has also been maintained that grouping and accentuation structures are very closely related. Once a grouping structure is defined, the accentuation structure emerges naturally and, from this, the metrical structure may be inferred. It is suggested that the proposed theory is more economic and coherent than most theories of rhythm that treat grouping and accentuation structures as independent components. The evidence presented in this study accounts only for low-level structural features of grouping and accentuation organisation. It may be the case that at higher-levels of organisation these structures may be partially independent and conflicting. It still is very interesting to see how much is embodied in and can be inferred from a well defined local grouping structure (viz. accentuation and metrical structures).
Chapter 7

Macrostructural Module I
(Musical Parallelism & Segmentation)

Introduction

Music becomes intelligible to a great extent through self-reference, i.e. the relations of new musical passages to previously heard material. Structural repetition and similarity are crucial devices in establishing such relations. Similar musical entities are organised into musical categories such as rhythmic and melodic motives, themes and variations, harmonic progression groups etc. (see chapter 8). Musical parallelism not only establishes relationships between different musical entities but enables - in the first place - the definition of such entities by directly contributing to the segmentation of a musical surface into meaningful units (section 7.6).

Despite the importance of musical parallelism, even the most elaborate contemporary musical theories avoid tackling the problem of parallelism in a systematic way (e.g. it is simply stated in the GTTM - rule GPR6, Lerdahl & Jackendoff, 1983:57). Theories that attempt to formalise musical similarity either restrict themselves to a very well circumscribed and rather limited area of musical knowledge - e.g. Ruwet's machine (Ruwet, 1987), similarity relations in pitch-class set theory (Forte, 1973) - or allow a fair amount of musical intuition to the analyst - e.g. traditional thematic analysis, Reti's thematic processes (Reti, 1951), paradigmatic analysis (Nattiez, 1975; 1990). Empirical studies of musical similarity often restrict themselves to very simple (usually artificially constructed) musical examples although there exists a rather small number of studies that investigate similarity for more complex real melodic excerpts (see Pollard-Gott, 1983; Deliège, 1996; Lamont and Dibben, 1997).

Pattern-matching techniques have been employed in attempts to describe musical parallelism and to build computational systems that recognise or induce musical patterns. An overview of
pattern-matching algorithms used for musical purposes is given in (McGettrick, 1997) and a survey of general string pattern-matching techniques that may be useful for musical analysis and musical information retrieval is presented in (Crawford et al., 1997).

In this chapter the concept of musical parallelism/similarity will only partially be examined in relation to the notion of identity (two musical passages are parallel if they share at least one identical pattern for at least one parametric profile of the melodic surface or a reduction of it); a computational model that discovers significant melodic patterns and contributes towards melodic segmentation will be proposed. Musical similarity will be fully described in chapter 8 wherein the notion of categorisation is introduced and the two are brought into a close relation.

7.1 Similarity and Pattern-matching

Full pattern-matching is aimed at finding instances of given patterns or inducing identical patterns. However, pattern-matching may be used for revealing or establishing similarity between different patterns as well. What kind of pattern-matching methodology, though, is most adequate when attempting to establish similarities between complex entities such as melodic passages?

There are two main approaches:

a) Partial pattern-matching applied on the unstructured musical surface, and,

b) Full pattern-matching applied on the musical surface and on a number of reduced versions of it that consist of structurally more prominent components.

The first approach is based on the assumption that musical segments construed as being parallel (similar) will have some of their component elements identical (for example, two instances of a melodic motive will have a 'significant' amount of common notes or intervals but not necessarily all) - some partial pattern-matching algorithms based on this approach are described in (Bloch and Dannenberg, 1985; Cope, 1990, 1991; Rowe and Li, 1995; Stammen and Pennycook, 1993). The second approach is based on the assumption that parallel musical segments are necessarily fully identical in at least one parametric profile of the surface or reduction of it (for example, two instances of a melodic motive will share an identical parametric profile at the surface level or some higher level of abstraction, e.g. pattern of metrically strong or tonally important notes/intervals and so on) - a computational technique based on this approach is described in (Hiraga, 1997).
What are the pros and cons of each of the above pattern-matching methodologies? Perhaps an example will help clarify the relative merits of each approach. Consider the tonal melodic segments of figure 7.1. How similar are segments $b$, $c$, $d$ to segment $a$? Let us suppose, for convenience, that each melodic segment is represented as a sequence of pitch and inception-time note tuples (figure 7.1, bottom).

Partial pattern matching would show that each of the segments $b, c, d$ is 71% identical to segment $a$ as 5 out of 7 note tuples match (mismatches are indicated by asterisks in figure 7.1). Depending on the threshold that has been set the three melodic segments are equally similar - or dissimilar - to segment $a$. It is quite clear though to a musician that segment $b$ is much more similar to segment $a$ than any of the other segments because segments $a & b$ match in exactly the 'right' way, i.e. more prominent notes match and less important ornamentations are ignored.

In order for the second pattern-matching methodology to be applied a significant amount of pre-processing is required - for instance, the melodic segments are not simply examined at the surface level but various more abstract levels of representation that reflect structural properties of the melodic segments have to be constructed (e.g. longer notes, metrically stronger notes, tonally important notes etc.).

Both methodologies can handle musical similarity and parallelism, but the second can give rise to more sophisticated similarity judgements as it takes into account structural properties of the musical materials - the trade-off being that it is a more complicated procedure. A further advantage of the second pattern-matching methodology is that the reasons for which two musical segments are judged to be parallel/similar are explicitly stated, i.e. the properties
common to both are discovered and explicitly encoded. Such explicit knowledge may be used constructively for further analytic - or compositional - tasks.

In the current study the second methodology has been selected. Full pattern-matching is applied on a number of independent parametric profiles of a melodic surface. Separate analyses are performed for the different parameters of a melody (pitch, rhythm, dynamics etc.) for different levels of abstraction for each of these (e.g. for pitch intervals: exact intervals, scale-steps, contour etc.); additionally, the analyses may be performed on reduced versions of the surface. Then, the results obtained for each parametric profile are combined in order to discover significant melodic patterns and to segment the melodic surface. The interleaving of these different and often conflicting profiles into a single overall analysis has already been addressed in chapter 6 (combination of local boundaries for a number of parametric profiles) and will be examined further in the following sections.

7.2 Overlapping of Patterns

Many contemporary theories - especially theories that have been influenced by linguistic theory - make hypotheses about the way a musical surface should be segmented that are too restricting and limiting. For example, the Generative Theory of Tonal Music (Lerdahl and Jackendoff, 1983) assumes two kinds of rules the first of which are referred to as well-formedness rules. These rules allow grouping interpretations of a piece that comply with a strict tree-like hierarchic non-overlapping structure (limited one-note overlaps and elisions are occasionally allowed as exceptions to these rules).

It is herein suggested that such well-formedness rules should be considered simply as preference rules in a theory where the overlapping of patterns is the norm. Even in the classical tonal system it seems that the cases where such rules apply precisely are rather limited. Most music has a fair number of ambiguous passages where not only the different parametric profiles conflict with each other making it impossible to find a well-formed description, but even within a single profile a non well-formed description may be the most appropriate. For instance, in figure 7.2 a possible description of a melodic surface in terms of a heavily overlapping pattern is depicted. This heavy overlapping may be interpreted as producing a sense of ongoingness or ambiguity. Alternatively, the significant 7-note motive may be broken down into two sub-motives which describe bars 3-4 in a non-overlapping fashion.
Our cognitive skills attempt to impose a well-formed interpretation on a musical surface which is the preferred interpretation mainly for reasons of cognitive economy. This process though often fails leaving an unresolved ambiguity and uncertainty which is central to musical meaning. Music seems to have much weaker 'parsing' rules to which an analysis should comply than natural language has. There are better or worse descriptions, more or less economic, closer or more remote to cognitive models, preferred or avoided within a certain context. In this sense, we consider closer to musical understanding theories that are non-exclusive, i.e. 'theories which do not view new pieces as being true or false, but rather regard all representable musical surfaces as possible' (Conklin and Witten, 1991:2) and all musical analyses as well.

7.3 Pattern-matching and Pitch-Interval Representation

The importance of pitch-interval representation in the designing of a pattern-matching process that detects repetition of pitch-interval patterns will be examined in this section. Our discussion will revolve around a matching process proposed by West, Howell & Cross (1992:7) which they illustrate concisely in the example of figure 7.3.

Figure 7.3 'A simple figure (a), requires at least three different methods of encoding pitch intervals for repetition to be detected by a matching process. Repetition with in-scale transposition (b) requires scale step encoding; repetition with simple transposition (c) requires chroma (pitch class) encoding; and repetition with contour preservation (d) requires contour encoding.' (West, Howell & Cross, 1992:7).
Although this process is very general and economic and gives successful results for the detection of repetitions in the majority of musical surfaces presented to the system, there are some inherent deficiencies relating to the way pitch-intervals are encoded. This procedure will be examined in two respects:

1. If the levels of representation of the pitch-intervals are considered to be strictly hierarchical - i.e. matchings that are detected first, starting from the lowest level (chroma) upwards, are the ones to be selected (it is understood that this is suggested by the authors) - then the system exhibits the following problems:

   a. It disregards important differences by matching (considering identical) enharmonic intervals in tonal surfaces. This shortcoming appears because the chroma level does not effectively represent a tonal surface. The process is not strictly hierarchical as it is possible to find situations, as in figure 7.4, where a higher (more abstract) level contradicts (does not match) a repetition detected at a lower level.

   ![Figure 7.4](image)

   Figure 7.4

   b. The scale-step diatonic matching level is arbitrary in a distributional atonal environment (based on the 12-tone system). A quantification of the chroma level into equal numbers of semitones may be less arbitrary (e.g. 2-semitone intervals, and so on).

   c. Hierarchical tonal systems other than the 7-tone diatonic system are not efficiently represented neither in the chroma level nor in the scale-step level. The pitch and pitch-interval properties of such systems are not appropriately accounted for and thus the analyses obtained from this matching procedure are apt to diverge from the expected results.

---

5 For example, the minor 3rd and the 'rare' augmented 2nd intervals are classified together as 3 semitone intervals. This way the important distinction between them is disregarded altogether. The opposite situation occurs when 12-tone music is analysed by a 7-tone scale-interval representation, i.e. non-significant information is encoded as significant.
2. If the levels of representation are considered to be complementary to each other (e.g. chroma and scale-step levels) then the problems discussed in 1a and 1b may be eliminated as it is possible to infer implicitly the dissimilarity of enharmonic intervals in a 7-tone environment or to deactivate the scale-step level in a distributional 12-tone environment. This means that the system needs additional mechanisms that can control these inter-level relations; but this way it loses on its simplicity and economic outlook. Even with the aid of an extra mechanism, problem 1c cannot be accounted for if the initial representations are not altered.

It is suggested that the general pitch-interval representation proposed in chapter 5 may explicitly represent a wider range of pitch structures in a purely hierarchic fashion. In figure 7.5, the first pitch pattern is matched to each of the following patterns within: a) a 7-tone diatonic representation and b) a 12-tone representation.

For 7-tone diatonic representation:

\[
\begin{align*}
\{\text{dir,nci}, \text{mdf}\} : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} \\
\{\text{dir,nci}\} : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} \\
\{\text{dir,nci}\}' : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\}
\end{align*}
\]

intermediate levels: 

\[
\begin{align*}
\{\text{dir}\} : & \{\ast, +, +, -, \ast, +, +, \ast, +, +, -, \ast, +, +, -, \ast, +, +, -, \ast, +, +, -, \ast, +, +, -
\end{align*}
\]

For 12-tone representation (\{\text{dir,nci}, \text{mdf}\} level is redundant as mdf is always A):

\[
\begin{align*}
\{\text{dir,nci}, \text{mdf}\} : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} \\
\{\text{dir,nci}\} : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} \\
\{\text{dir,nci}'\} : & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\} & \{\ast, +1, +1, -7\}
\end{align*}
\]

intermediate levels: 

\[
\begin{align*}
\{\text{dir}\} : & \{\ast, +, +, -, \ast, +, +, \ast, +, +, -, \ast, +, +, -, \ast, +, +, -, \ast, +, +, -
\end{align*}
\]

Figure 7.5 The first pitch pattern is matched against each of the subsequent patterns within: a) a 7-tone diatonic representation and b) a 12-tone representation.

This pattern-matching procedure gives rise to different analyses of a musical surface for different scaling systems. It is also possible to make use of more than one analysis in a multiple-viewpoint approach implementation.

6 If hybrid musical systems are taken into consideration, e.g. 12-tone music with 7-tone micro-structural properties, then additional evaluation-selection mechanisms should be employed to combine different matching procedures.
7.4 The String Pattern-Induction Algorithm (SPIA)

A brute-force pattern-matching algorithm that can be applied to any sequence of entities will be described below - a formal description of an almost identical algorithm can be found in (Crow and Smith, 1992). The aim of the algorithm is pattern induction, i.e. the discovery of patterns that recur in a string of symbols. The String Pattern-Induction Algorithm (SPIA) is employed in a bottom-up fashion, i.e. starting from the smallest patterns and extending them to maximum length. The well-formedness demands posed by a hierarchical structure of discrete levels with approximately equal length non-overlapping groups are by-passed; overlapping of patterns is allowed.

For a given sequence of entities (e.g. a parametric profile of scale-step pitch intervals), the matching process starts with the smallest pattern length (2 elements) and ends when the largest pattern match is found. For a given pattern length, every possible pattern of the string (starting with the first) is matched against the remainder of the string by a shifting stepwise motion. The patterns for which at least one match is found are separated and labelled (melodic patterns may be matched in their original form or in their retrograde, inversion and retrograde inversion forms). Patterns for which no match is found are disregarded after the introduction of a break marker in their place. Pattern-matching cannot override such markers and the initial sequence is in essence fragmented into shorter sequences. As the matched patterns grow in size, the search space is reduced. When the last matching is found for the largest possible pattern, the matching process ends.

The String Pattern-Induction Algorithm is exhaustive, i.e. it discovers all possible matches, and although it is computationally expensive (polynomial time), it becomes more efficient through the reduction of the initial search space. This procedure can become significantly faster if break markers are inserted in the initial sequence for positions that are thought to be important boundaries in the sequence (e.g. for a melody, points suggested by the LBDM or positions marked in a score by breath marks, large rests, slurs, fermatas, and so on). It is also possible to pre-define a limited range of pattern lengths for which the SPIA will be employed.

For hierarchically ordered melodic profiles (e.g. exact interval - scale step interval - contour profiles) the pattern matching process can be applied first to a more general profile and, then,
the search may proceed within the patterns previously discovered. There is no reason to employ an exhaustive search for every individual parametric profile. This again reduces significantly the search space and the computational time involved (this procedure is not as yet implemented).

The SPIA is applied to as many parametric profiles as are considered necessary (e.g. pitch, duration, start-time, dynamic intervals and so on) for the melodic surface and/or reductions of it.

![Figure 7.6 A great number of pitch-interval pattern matches is found by the SPIA in this short trivial melodic sequence.](image)

It is apparent that such a procedure for the discovery of parallel melodic segments will produce a very large number of possible patterns (figure 7.6) most of which would be considered by a human musician-analyst counter-intuitive and non-pertinent. How can the most prominent patterns be selected and the unimportant ones be filtered out? The next section addresses this issue and proposes a possible solution.
7.5 The Selection Function

Rowe attaches a strength value on each pattern depending on its frequency of occurrence: 'Each known pattern has an associated strength: the strength is an indication of the frequency with which the pattern has been encountered in recent invocations of the program.' (Rowe, 1993:248).

In an attempt to devise a procedure that can attach a prominence value to each of the previously discovered patterns a hypothesis is made whereby the importance of a given pattern relies on the following three factors:

- Prefer longer patterns
- Prefer most frequently occurring patterns
- Avoid overlapping

Below is a function\(^8\) that calculates a numerical value for a single pattern according to the above principles:

\[ f(PL,F,DOL) = Fa^{a}PL^{b}/10^{c\cdot DOL} \]

where
- \( PL \): pattern length, i.e. number of elements in pattern
- \( F \): frequency of occurrence for one pattern
- \( DOL \): degree of overlapping\(^9\)
- \( a, b, c \): constants that give different prominence to the above principles

Any of the three principles can be neutralised by setting the relevant constant to zero. For instance, if \( c=0 \) then \( f(PL,F,DOL) = Fa^{a}PL^{b} \) and the Selection Function is independent of the degree of overlapping. The importance of each principle can be adjusted by assigning different values to the constants. Additionally, the shape of the function may be changed by altering the constants, e.g. for same relative importance of each principle such as

\[ f(PL,F,DOL) = Fa^{a}PL^{b}(1-c\cdot DOL). \]

\(^8\) In this function, the avoidance of patterns that exhibit a degree of overlapping increases exponentially in relation to DOL - for a linear relation a possible function is:

\[ f(PL,F,DOL) = Fa^{a}PL^{b}(1-c\cdot DOL). \]

\(^9\) DOL is defined as the number of elements shared by some patterns divided by the number of all the elements in those patterns or more precisely: DOL = \((T-U)/U\) where: \( T \) is the total number of elements in all the matchings discovered for a pattern (\( T=F\cdot PL \)); \( U \) is the number of elements in the union set of all the matchings discovered for a pattern (this definition allows DOL to be in some cases greater than 100%).
(a,b,c)=(3,3,3) the function produces a curve with sharper peaks than for (2,2,2) which means more prominence for greater length, greater frequency and less overlapping.

For every pattern discovered by the matching process a value is calculated by the use of this function (the same constants should be used for all the patterns). The patterns that score the highest should be the most significant ones.

Returning to figure 7.6, for a=2, b=2, c=2 the system gives the highest value for pattern p4-0; for a=2, b=3, c=2 the system selects p2-0; for a=2, b=2, c=2 and for original matchings only (without retrograde patterns) p3-0 is selected. All of these patterns (along with p8-0) receive the highest values for the above function and are separated from the rest which score much lower.

The pattern analysis and the resulting segmentation is significantly improved when many analyses are performed for multiple profiles and then combined to give an overall multi-faceted description (see next section). Further examples of the application of the SPIA & Selection Function on a variety of melodies are presented in figures 7.7, 9.2, 9.9 and 9.14.

7.6 Segmentation based on musical parallelism

It has been suggested in section 6.1 that the segmentation of a musical surface is not only affected by local discontinuities (detected by the LBDM) but by higher-level processes as well. Perhaps the most important of these higher-level mechanisms is musical parallelism, i.e. similar musical patterns tend to be highlighted and perceived as units/wholes whose beginning and ending points influence the segmentation of a musical surface.

The computational model that consists of the String Pattern-Induction Algorithm and the Selection Function provides a means of discovering such 'significant' patterns. Figure 7.7 illustrates the most prominent pitch patterns for the song Frère Jacques selected by the SPIA & Selection Function. There is though a need for further processing that will lead to a 'good' description of the surface (in terms of exhaustiveness, economy, simplicity etc.). It is likely that some instances of the selected pitch patterns should be dropped out or that a combination of patterns that rate slightly lower than the top rating patterns may give a better description of the musical surface.
In order to overcome this problem a very simple but crude methodology has been devised. According to this, pattern-matching is applied to as many parametric profiles of the melodic surface and reductions of it as required (see section 9.2 for selection of parametric profiles in the current study). No pattern is disregarded but each pattern contributes to each possible boundary of the melodic sequence by a value that is proportional to its Selection Function value. That is, for each point in the melodic surface all the patterns are found that have one of their edges falling on that point and all their Selection Function values are added together. This way a Pattern Boundary strength profile is created (normalised from 1-100). It is hypothesised that points in the surface that have local maxima are more likely to be perceived as boundaries because of musical parallelism (see, for instance, the local maxima that appears at the end of bars 1, 2 and 6 in the Pattern Boundary strength profile of fig. 7.8 - more examples in sec. 9.2).

In order to overcome this problem a very simple but crude methodology has been devised. According to this, pattern-matching is applied to as many parametric profiles of the melodic surface and reductions of it as required (see section 9.2 for selection of parametric profiles in the current study). No pattern is disregarded but each pattern contributes to each possible boundary of the melodic sequence by a value that is proportional to its Selection Function value. That is, for each point in the melodic surface all the patterns are found that have one of their edges falling on that point and all their Selection Function values are added together. This way a Pattern Boundary strength profile is created (normalised from 1-100). It is hypothesised that points in the surface that have local maxima are more likely to be perceived as boundaries because of musical parallelism (see, for instance, the local maxima that appears at the end of bars 1, 2 and 6 in the Pattern Boundary strength profile of fig. 7.8 - more examples in sec. 9.2).
7.7 Interaction with microstructural module

The boundaries revealed by the *LBDM* may assist or complement the pattern boundary detection mechanism described in the previous section.

Firstly, significant boundaries discovered by the *LBDM* can be used as a guide for inserting *break markers* in the musical surface (as suggested in section 7.4). This practice may improve significantly the efficiency of the *String Pattern-Induction Algorithm* by breaking down the musical surface into shorter sequences and thus reducing the available search space. The assumption underlying this procedure is that a listener may use strong local boundary cues as tentative points of segmentation which are unlikely to be overridden by a pattern.

Two types of break markers have been implemented: a) *hard* breaks which cannot be overrun by any pattern, and b) *soft* breaks that can be slightly overrun (e.g. by one element) by either side of a pattern. The exact thresholds for defining *hard* or *soft* break markers need further investigation. In the current study two factors have been selected for designating points where break markers may be inserted: strength of local boundary in relation to its two adjacent neighbouring values, and strength of local boundary in relation to the average of all the boundary strengths (see figure 7.8 - hard breaks indicated by double cross - soft breaks by single cross).

Secondly, the boundaries discovered by the pattern-matching process may complement the local boundaries detected by the *LBDM* in defining the Total Boundary strength profile. In the melodic example of figure 7.8 the Pattern Boundary strength profile has been calculated by applying the *SPIA* to the scale-step, contour and duration profiles (patterns are allowed to reach maximum lengths and the *Selection Function* constants are set to (a,b,c)=(3,3,4)) - if a limited range of pattern lengths is allowed (e.g. 3-4 notes), as suggested in section 9.1.2 and implemented in section 9.2.1, then the peaks of the Pattern Boundary profile become much sharper. The Total Boundary strength profile is calculated as a weighted average of the Local Boundary and Pattern Boundary strength profiles - in this implementation they contribute by 40% and 60% respectively. The local maxima in the Total Boundary strength profile can be taken as a guide for the segmentation of the musical surface (see examples in section 9.2).
Conclusion

An analysis of a given melodic passage involves establishing a way of discovering significant musical patterns. In this chapter a computational model has been introduced that discovers such patterns for a given parametric profile of a melody. The matching process allows overlapping of patterns and then a selection method singles out the most prominent ones taking into account their length, frequency of occurrence and degree of overlapping. This method can be applied to a number of parametric profiles of a melody and the results of each of these can be combined to produce a Pattern Boundary strength profile indicating the most prominent boundary positions due to musical parallelism. This, in conjunction with the local boundaries highlighted by LBDM (chapter 6), leads to an integrated segmentation of a melodic surface.
Chapter 8

Macrostructural Module II
(Musical Categories)

Introduction

Musical parallelism has been discussed to a certain extent in chapter 7. It has been assumed (section 4.5) that similar musical passages are organised into musical categories such as rhythmic and melodic motives, themes and variations, harmonic progression groups etc. But when are two different musical passages similar? And when are two passages different enough to be considered dissimilar? Which musical passages belong to the same paradigm/category? What happens with ambiguous passages?

Following the discussion on similarity and categorisation in chapter 4, a detailed description of a working formal definition of these notions will be given according to which similarity a) is contextually defined, b) may be applied to any property ascribed to an entity (not only to perceptual properties such as visual appearance) and (c) has an associated notion of corresponding categories. This definition inextricably binds together similarity and categorisation in such a way that changes in similarity ratings between entities result in category changes, and vice versa.

In line with these definitions, the *Unscramble* algorithm will be presented which, given a set of objects and an initial set of properties, generates a range of plausible classifications for a given context. During this dynamically evolving process the initial set of properties is adjusted so that a satisfactory description is generated (taking into account the general cognitive principles outlined in section 4.1). There is no need to determine in advance an initial number of classes nor is there a need to reach a strictly well-formed (e.g. non-overlapping) description. At every stage of the process both the extension and the intension of the emerging
categories are explicitly defined. One general example and one musical example will be presented that illustrate the capabilities and effectiveness of the model.

**8.1 A Working Formal Definition of Similarity and Categorisation**

Let \( T \) be a set of entities and \( P \) the union of all the sets of properties that are pertinent for the description of each entity. If \( d(x,y) \) is the distance between two entities \( x \) and \( y \), \( h \) is a distance threshold, and \( s_h(x,y) \) is a function inversely related to the distance, e.g. \( s_h(x,y) = h - d(x,y) \), then:

\[
s_h(x,y) = \begin{cases} 
    \geq 0 & \text{iff } d(x,y) \leq h \quad \text{(similar entities)} \\
    < 0 & \text{iff } d(x,y) > h \quad \text{(dissimilar entities)} 
\end{cases}
\]

In other words, two entities are *similar* if the distance between them is smaller than a given threshold and *dissimilar* if the distance is larger than this threshold.\(^{10}\)

The above definition of similarity is brought into a close relation with a notion of category. That is, within a given set of entities \( T \), for a set of properties \( P \) and a distance threshold \( h \), a category \( C_k \) is a maximal set with the following property:

\[ C_k = \{x_1,x_2,...,x_n\} \text{ such that: } \forall i,j \in \{1,2,...,n\}, \ s_h(x_i,x_j) \geq 0 \quad \text{(II)} \]

In other words, a category \( C_k \) consists of a maximal set of entities that are pairwise similar to each other for a given threshold \( h \). A category, thus, is inextricably bound to the notion of similarity; all the members of a category are necessarily similar and a maximal set of similar entities defines a category.

The distance threshold may take values in the range of \( 0 \leq h \leq d_{\text{max}} \) where the distance \( d_{\text{max}} \) is defined as the maximum distance observed between all the pairs of entities in \( T \), i.e. \( d_{\text{max}} = \max(d(x,y)) \).

If \( h = 0 \) and \( s(x,y) = 0 \), then \( x = y \) (identity) and every individual in \( T \) is a monadic category.

\(^{10}\) Alternatively, the function \( s_h(x,y) \) may be defined in a binary manner - for instance: \( s_h(x,y) = 1 \) iff \( d(x,y) \leq h \) (similar entities) and \( s_h(x,y) = 0 \) iff \( d(x,y) > h \) (dissimilar entities).
If $0<h<d_{\text{max}}$ then the set of entities $T$ is not a category but may be exhaustively described by $m$ categories (possibly overlapping) such that $C_k \subseteq T$, $k \in \{1,2,...,m\}$ and $C_1 \cup C_2 \cup ... \cup C_m = T$ and $m$ sets of properties such that $P_k \subseteq P$, $k \in \{1,2,...,m\}$ and $P_1 \cup P_2 \cup ... \cup P_m \subseteq P$.

If $h=d_{\text{max}}$ then all the entities in $T$ define a single category $C$ with the property set $P$.

### 8.2 The Unscramble algorithm

The above definitions of category and similarity readily lend themselves to form the basis of a dynamic process for discovering pertinent categories and similarities. Given a set of entities and properties the Unscramble algorithm (see figure 8.1) generates a categorisation (i.e. organisation of the space of entities into a number of categories); as categorisation descriptions are refined so are similarities between entities and the prominence of different properties. The term 'categorisation description' or simply 'categorisation' corresponds, in this text, to the term 'clustering' used in the standard machine learning terminology.

The threshold $h$ can take values in the range of $0 \leq h \leq d_{\text{max}}$, but a finite subset of values that is equal to the number of possible distances between the $n$ objects of set $T$ (total number of distances $= n \cdot (n-1)/2$ - it often is smaller as some entities are equidistant) is sufficient for the calculations of all the possible categorisations according to definition (II). Each of these thresholds defines a number of sets of objects in each of which all the members are pairwise similar, i.e. they are categories.

From the above possible categorisations for all the possible thresholds a selection mechanism can select the 'best' categorisation. The selection criteria for determining good categorisations are: a) an exhaustive description of the object set, b) minimum overlapping between the categories, and c) avoiding categorisations that are too specialised (each object a category of itself) or too general (all objects form one category).
The **UNSCRAMBLE** algorithm

1. Select a general set of properties that are pertinent for the description of the set of objects to be organised in categories; select a distance metric.
2. Initialise weights for each property to w=1 (variable weights in the range 0≤w≤1 may also be defined if the prominence of property is known in advance).
3. Calculate all possible distances between every pair of objects.
4. Set the threshold values equal to the distances calculated in (3).
5. For each threshold, compute all the similarities for every pair of objects according to definition (I).
6. Find maximal sets that satisfy definition (II), i.e. maximal sets for which all their members are pairwise similar.
7. Select *preferred* classifications according to the following preference rules:
   a. prefer categorisations with minimal overlapping between the various categories;
   b. prefer number of categories m to be in the range: 1<m≤N/2, where N is total number of objects;
   c. prefer categories with more than one member.
8. The preferred categorisation(s) is considered *satisfactory* if it satisfies predefined constraints for the preference rules of stage (7), i.e. maximum degree of overlapping (e.g. zero or less than 10% etc.), limited range of permitted number of categories and maximum percentage of monadic categories.
9. For the selected satisfactory categorisation(s) - or the preferred one(s) if no satisfactory categorisation has emerged:
   a. if categorisations for more than one threshold have been selected delete, if any, all duplicate categories.
   b. calculate weights for each category according to definition (III).
   c. find average weights for each property from all the weights that have been computed from (8b) for each category.
   d. normalise weights so that maximum weights equal 1.
10. If a satisfactory categorisation has emerged, define the prototype of each category, i.e. find the weighted set of properties that is characteristic for each category, and STOP the algorithm.
11. If a satisfactory categorisation has not emerged, proceed with preferred classification and repeat process from stage (3) for the new weights.

Figure 8.1
When a threshold is chosen, then the initial weights of properties can be altered so as to optimise the distinctiveness of the category's intension. Weights for each property may be adjusted in relation to the diagnosticity of that property for a given category, i.e. properties that are unique to members of one category are given higher weights whereas properties that are shared by members of one category and its complement are attenuated (in other words, the dimensions in a multi-dimensional space are adjusted in such a way that distances between members of different categories are maximised). For example such a function that calculates the weight of a single property $p$ could be:

$$w = \left| \frac{m}{n} - \frac{m'}{N-n} \right|$$  \hspace{1cm} (III)

where:

$m$ = number of objects in category $C_k$ that possess property $p$

$m'$ = number of objects not in category $C_k$ that possess property $p$ (i.e. objects in $T-C_k$)

$n$ = number of objects in $C_k$

$N$ = number of objects in $T$

The weights of each property calculated for each category can then be averaged and normalised for a given categorisation. If an acceptable classification has not been arrived at, the whole process may be repeated for the new set of weighted properties until a satisfactory categorisation is achieved.

One general example will be presented in the next section to illustrate the utility of the above definitions and processes. Then, in section 8.4, the $Unscramble$ algorithm will be applied on a set of melodic segments.

### 8.3 An Illustrative Example

#### 8.3.1 Category Formation

Let us assume that the set of objects $T$ (figure 8.2) is described by a set of properties which, in this example, are taken to be the following attributes with nominal values:
Figure 8.2 Set of objects T for categorisation

Each object X is represented by an array of n=5 attribute values: (x₁,x₂,x₃,x₄,x₅), e.g. for object E: (circle, small, white, heart, bold).

Let us also assume that the distance (0 ≤ d(x,y) ≤ 1) between two objects is given by the following function (based on the Hamming distance):

\[ d(x,y) = \sum_{i=1}^{n} w_{xi} w_{yi} |x_i - y_i| \]  

where:

\[ |x_i - y_i| = 0 \text{ if } x_i = y_i \]
\[ |x_i - y_i| = 1 \text{ if } x_i \neq y_i \]

If stages 2-6 of the *Unscramble* algorithm are applied to the above set of objects and set of attributes we get:

Threshold: h=4
Similarities: s_{AB}=1 s_{AC}=1 s_{AD}=0 s_{AE}=0 s_{AF}=0 s_{BD}=1 s_{BE}=0 s_{BF}=0 s_{CD}=1 s_{CE}=0 s_{CF}=0 s_{DE}=0 s_{DF}=0 s_{EF}=2
Categories: {A,B,C,D,E,F}
Threshold: \( h=3 \)
Similarities: \( s_{AB}=0 \quad s_{AC}=0 \quad s_{BD}=0 \quad s_{CD}=0 \quad s_{EF}=1 \)
Categories: \{A,B\}, \{A,C\}, \{B,D\}, \{C,D\}, \{E,F\}

Threshold: \( h=2 \)
Similarities: \( s_{EF}=0 \)
Categories: \{A\}, \{B\}, \{C\}, \{D\}, \{E,F\}

None of the above categorisations is satisfactory according to the selection preference rules of stage 8 (where constraints have been set as follows: overlapping is less than 10\%, \(1<m\leq3\), fewer than two monadic categories). So, the algorithm proceeds to stage 9 for a preferred categorisation, e.g. for \( h=2 \) (containing the most stable category \{E,F\}) for which new weights are calculated (weights other than 1 in parentheses):

- \( A'_1 \): shape \{square(0.8), triangle(0.8), circle\}
- \( A'_2 \): size \{small(0.6), big(0.6)\}
- \( A'_3 \): shade \{white(0.6), grey(0.6)\}
- \( A'_4 \): content \{dot(0.8), cross(0.8), heart\}
- \( A'_5 \): outline \{plain(0.8), double(0.8), bold\}

Since stage 10 fails, the *Unscramble* algorithm is now repeated from stage 3 for the new weighted attribute set \( A' \). As there are now five possible distances between the objects we have five values of \( h \), and we get:

Threshold: \( h=2.76 \)
Similarities: \( s_{AB}=1.39 \quad s_{AC}=0.83 \quad s_{AD}=0.75 \quad s_{AE}=0 \quad s_{AF}=0 \quad s_{BC}=0.75 \quad s_{BD}=0.83 \quad s_{BE}=0 \)
\( s_{BE}=0 \quad s_{CD}=1.39 \quad s_{CE}=0 \quad s_{CF}=0 \quad s_{DE}=0 \quad s_{DF}=0 \quad s_{EF}=2.04 \)
Categories: \{A,B,C,D,E,F\}

Threshold: \( h=2.0 \)
Similarities: \( s_{AB}=0.63 \quad s_{AC}=0.07 \quad s_{AD}=0 \quad s_{BC}=0 \quad s_{BD}=0.07 \quad s_{CD}=0.63 \quad s_{EF}=1.29 \)
Categories: \{A,B,C,D\}, \{E,F\}

Threshold: \( h=1.92 \)
Similarities: \( s_{AB}=0.55 \quad s_{AC}=0 \quad s_{BD}=0 \quad s_{CD}=0.55 \quad s_{EF}=1.21 \)
Categories: \{A,B\}, \{A,C\}, \{B,D\}, \{C,D\}, \{E,F\}
Threshold: $h=1.36$
Similarities: $s_{AB}=0$ $s_{CD}=0$ $s_{EF}=0.65$
Categories: \{A,B\}, \{C,D\}, \{E,F\}

Threshold: $h=0.72$
Similarities: $s_{EF}=0$
Categories: \{A\}, \{B\}, \{C\}, \{D\}, \{E,F\}

From these categorisation descriptions, only the ones for $h=1.36$ and $h=2$ are preferred (stage 7) and also fulfil the selection criteria of stage 8. For the categories that have emerged for $h=1.36$, - i.e. \{A,B\}, \{C,D\}, \{E,F\} - the final set of weighted attributes $A''$ is given below (note that the attributes of 'shade' and 'size' are not included as they have received zero values, i.e. they are non-diagnostic):

$A''_1$: shape \{square(0.25), triangle(0.25), circle\}
$A''_2$: content \{dot, cross, heart\}
$A''_3$: outline \{plain, double, bold\}

For these new weights, each of the categories \{A,B\}, \{C,D\}, \{E,F\} is defined for the following range of thresholds and set of weighted attributes (prototypes):

Category: \{A,B\}
Threshold Range: $0.06 \leq h < 2.06$
Attributes: $A''_1$: shape \{square(0.25), triangle(0.25)\}
$A''_2$: content \{dot\}
$A''_3$: outline \{plain\}

Category: \{C,D\}
Threshold Range: $0.06 \leq h < 2.06$
Attributes: $A''_1$: shape \{square(0.25), triangle(0.25)\}
$A''_2$: content \{cross\}
$A''_3$: outline \{double\}

Category: \{E,F\}
Threshold Range: $0 \leq h < 2.25$
Attributes: $A''_1$: shape \{circle\}
$A''_2$: content \{heart\}
$A''_3$: outline \{bold\}
The final set of weighted attributes along with the lowest of these threshold values describe the core of the category whereas the highest threshold values the outermost possible category boundaries.

For the threshold h=2 two categories are defined: \{A,B,C,D\} and \{E,F\}. The prototype for category \{A,B,C,D\} is:

A1: shape \{square(0.5), triangle(0.5)\}  
A2: content \{dot(0.5), cross(0.5)\}  
A3: outline \{plain(0.5), double(0.5)\}  

Category \{A,B,C,D\} cannot be defined in monothetic terms (i.e. by singly necessary and jointly sufficient conditions) as there is no single property shared by all its members (but it can be defined by disjunctive conditions, e.g. (square OR triangle) AND (dot OR cross) AND (plain OR double)).

If the two descriptions for h=1.36 and h=2 are combined then a hierarchical categorisation description emerges (figure 8.3). Overlapping of categories is discussed in sections 8.3.3 and 8.4.

Figure 8.3

If the process started with different initial attribute weights then obviously different similarities/categorisation could emerge. If, for instance, the attribute 'shape' was given a higher weight (e.g. double weight) in the above example then objects would be categorised mainly by shape: \{A,D\}, \{B,C\}, \{E,F\}. If weights are given to some properties that are individually higher than the sum of all the other weaker properties, then monothetic categories would result.

If an object (or attribute) is found more frequently in the initial set then this affects the weights of the attributes (see section 4.1). For instance, if object A appeared five times in the

---

11 All the known category members belong to the core (these members are used in the membership prediction tests in section 8.3.2); however, the core of a category may contain more members that do not appear in the initial set of entities T for different combinations of the attributes in the prototype.
initial set then we would get eventually for category \{A,B\} the following 'shape' attribute weights: shape\{square(0.57), triangle(0.27)\}, i.e. 'square' would be more predictive of the category members than 'triangle' because it is encountered more frequently.

In the next section it will be shown how these category descriptions can be used to make predictions of category membership for new objects.

8.3.2 Category Membership Prediction

When a new object is presented and category membership is sought for it, there are two alternative options:

1. If the initial set of objects T is considered to be representative of objects and correlations among those objects' attributes in the context of a rather stable world, then an attempt may be made to categorise the new object into one of the existing categories. In this case, the above descriptions of categories can be used to predict membership of the new object by calculating all the distances of the new object to all the objects in each category's core (h minimum) and checking if all these pairs are similar (s ≥ 0). If this succeeds, then the object is a member of the core of one or more categories. If it fails, the similarity of the new object to all the members of each category's core may be calculated for the category's outermost boundaries (h maximum); this may succeed for one or more categories in which case the new object lies within the broader limits of one or more categories (it is a member but not a core member of each category). If an object is found to be a member of more than one of the existing categories then ambiguous membership results. This ambiguity may be resolved if the whole categorisation process is applied on the reduced set of the objects in the overlapping categories.

2. If a more permanent categorisation of a new object is desired then the new object(s) may be incorporated into the initial set of objects T, any new properties embodied in the initial attribute set A (or even in an adjusted attribute set) and the whole similarity/categorisation process activated from the beginning. This will most probably result in new categories and new weighted attribute sets.

Below are some examples of membership of new graphic objects (figure 8.4) according to option 1 in relation to the previously defined categories \{A,B\}, \{C,D\}, \{E,F\}:

- object G is a core member of \{A,B\} for h minimum.
• object H (similarly, object I) is a member of both \( \{A,B\} \) and \( \{E,F\} \) for \( h \) maximum and if the categorisation process is applied to the set \( \{A,B,E,F,H\} \) then H is shown to be more likely a member of \( \{A,B\} \).

• object J is a member of both \( \{A,B\} \) and \( \{E,F\} \) for \( h \) maximum and if the categorisation process is applied to the set \( \{A,B,E,F,J\} \) then J is shown to be more likely a member of \( \{E,F\} \).

• object K is a member of both \( \{A,B\} \) and \( \{C,D\} \) for \( h \) maximum and there is no preference in being a member of either of the two (object K is also a core member of \( \{A,B,C,D\} \) for \( h \) minimum).

• object L is a member of \( \{E,F\} \) for \( h \) maximum (notice the existence of new attribute value 'hexagon').

![Figure 8.4 Membership predictions for new previously unseen objects.](image)

It is suggested that human aspects of making membership judgements are reflected in the above options. Firstly, a subject checks if a new object is clearly a member of a known category. If it is not, then a small number of possible categories to which it may belong is selected. The membership process may stop there by simply stating that there is some ambiguity and the new object is a sort of hybrid in between different categories or it may continue by a closer examination of membership to the shortlisted categories. If the new object(s) is considered very important so that an elaborate study of its properties and a re-evaluation of the importance of the properties of the other known objects is rendered necessary then the whole similarity/categorisation process may be started right from the beginning after having incorporated the new object(s) and its (their) properties in the initial set of objects and properties.

### 8.4 A Musical example

Paradigmatic analysis (Nattiez, 1975, 1990; see section 2.1) is concerned with the organisation of a musical piece into columns (categories) of similar musical segments. Some musical segments that appear in Nattiez's paradigmatic analysis of Debussy's *Syrinx* are depicted in figure 8.5.
Figure 8.5

Segment D is placed by Nattiez in the column with motives E, F and G although one might initially think it would be more obvious to place segment D with A, B and C. How would this limited set of musical entities be categorised according to the Unscramble algorithm?

Let's assume we have a rudimentary set of pitch-interval and duration parametric profiles for each of these musical segments, i.e. exact pitch intervals (in semitones), contour and durations:

\[
\begin{align*}
A_{\text{rh}}: \{r_{\text{h}1}, r_{\text{h}2}\} & & A_{\text{pex}}: \{p_{\text{ex}1}, p_{\text{ex}2}, p_{\text{ex}3}, p_{\text{ex}4}\} & & A_{\text{pcont}}: \{p_{\text{cont}1}, p_{\text{cont}2}\}
\end{align*}
\]

If the initial weights for all the properties are \(w_{ij}=1\), we have the following categories (similarity values are not depicted) according to the similarity/categorisation algorithm (there are 4 possible distances therefore 4 useful thresholds):

- Threshold: \(h=3\) → Categories: \{A,B,C,D,E,F,G\}
- Threshold: \(h=2\) → Categories: \{A,B,C,D\}, \{D,E,F,G\}
- Threshold: \(h=1\) → Categories: \{A,B,C\}, \{D,E\}, \{E,F,G\}
- Threshold: \(h=0\) → Categories: \{A,B,C\}

If some overlapping is allowed then the two descriptions for \(h=2\) and \(h=1\) are acceptable according to the selection criteria. The description for \(h=2\) is somewhat simpler so preferable. It is obvious that segment D is ambiguous as it can be placed with \{A,B,C\} and/or \{E,F,G\}.

If no overlapping is allowed then one might select the most stable category \{A,B,C\} for \(h=0\), calculate new weights for the attribute set (\(w_{\text{rh}1}=0.75\), \(w_{\text{rh}2}=0.75\), \(w_{\text{pex}1}=1\), \(w_{\text{pex}2}=0.5\), \(w_{\text{pex}3}=0.25\), \(w_{\text{pex}4}=0.25\), \(w_{\text{pcont}1}=1\), \(w_{\text{pcont}2}=1\)) and then apply the similarity/categorisation algorithm to the segments for the new weights. This yields among other classifications:

- Threshold: \(h=0.68\) → Categories: \{A,B,C\}, \{D,E,F,G\}
This conforms with Nattiez's preference in placing musical segment D with the segments of the column/category that includes segments E, F and G. From the above weights it is clear that, for this classification, contour and pitch pattern pex1 are more diagnostic.

The process could have started with different initial attribute weights, e.g. the attribute 'rhythm' could have double weight (this would be quite reasonable in the sense that rhythm and pitch profiles would be overall equally important). In this case among other classifications we have:

Threshold: \( h=5 \)  \( \rightarrow \)  Categories:  \{A,B,C,D\}, \{D,E,F,G\}
Threshold: \( h=2 \)  \( \rightarrow \)  Categories:  \{A,B,C,D\}, \{E,F,G\}

In this case, where the initial weight of the attribute 'rhythm' is higher, the musical segment D is categorised with segments A, B and C (for \( h=2 \), if no overlapping is allowed, as one might have initially guessed (the attribute weights in this case are: \( w_{\text{rh}1}=1 \), \( w_{\text{rh}2}=1 \), \( w_{\text{pé}x1}=0.75 \), \( w_{\text{pé}x2}=0.08 \), \( w_{\text{pé}x3}=0.33 \), \( w_{\text{pé}x4}=0.33 \), \( w_{\text{cont}1}=0.75 \), \( w_{\text{cont}2}=0.75 \)).

The set of weighted attributes for each category along with the range of thresholds for which this category occurs can be used to make membership predictions of new unseen musical segments.

This musical example illustrates the flexibility and adaptiveness of the Unscramble algorithm. Segment D can either be grouped with segments \{A,B,C\} or with segments \{E,F,G\} depending on the initial weighting of the musical parameters or may simply be considered as an ambiguous hybrid of the two classes (although most analytic theories that are based on strict hierarchic non-overlapping descriptions would reject ambiguous overlapping descriptions). When human analysts make a paradigmatic analysis of the same musical piece it is almost certain that they will arrive at different descriptions. This is due to the fact that each analyst gives different prominence to the various musical parameters or might even use somewhat different parameters altogether and, of course, may choose different thresholds for what is considered to be similar/dissimilar. All of these possibilities are accommodated in the proposed system of categorisation.

### 8.5 Relative merits of Unscramble algorithm

The Unscramble algorithm has been applied successfully to a number of musical categorisation tasks whereby a number of melodic segments are organised into pertinent categories (motifs, themes etc.) - see also examples of organising melodic segments into categories in sections 9.2.1, 9.2.2 & 9.2.3. However, the real test of Unscramble will be to see
if and how it differs from and what relative merits it may have in comparison to other relevant concept formation algorithms (see Gennari et al., 1989; Van Mechelen et al., 1993, part II; Michalski, 1987; Langley, 1996).

Some possible useful characteristics of the Unscramble algorithm are:
- learning is unsupervised
- there is no need to define in advance a number of categories
- the prominence of properties is discovered by algorithm
- categories may overlap
- the categorisation descriptions for various thresholds are necessarily hierarchic
- knowledge about emergent categories is explicit and can be used for new membership predictions.

Many of these characteristics are accommodated in various algorithms. For instance, Cluster/2 (Michalski, 1983) is an unsupervised learning algorithm that enables explicit intensional definitions of categories to emerge (conceptual clustering); Cobweb (Fisher, 1987) encompasses most of these characteristics except overlapping (it is though different from Unscramble as it is based on a probabilistic approach and also performs categorisation in an incremental manner). Adclus (Arabie, 1977) is an indirect clustering model and its main common characteristic with Unscramble is that it allows overlapping of categories - see (Arabie et al., 1981) for potential utility of overlapping approaches to categorisation.

A much wider comparison with these and other relevant unsupervised learning algorithms is necessary for establishing and assessing the relative usefulness of Unscramble; the algorithm itself may benefit from other approaches (e.g. Cobweb's category utility criterion for evaluating the quality of categorisation descriptions).

**Conclusion**

In this chapter, a working formal definition was given according to which similarity is contextually-defined and is inextricably bound to a notion of corresponding categories. This definition was used as the basis for a dynamic process whereby, given a set of objects and properties, a range of plausible classifications of similar entities for a given context is generated and the most diagnostic properties highlighted. Unscramble has been successfully applied on a number of melodic categorisation tasks; however, further research is necessary to highlight the potential uses of the algorithm in domains other than music.