# Towards a General Computational Theory of Musical Structure

Emilios Cambouropoulos

Ph.D.

The University of Edinburgh
Faculty of Music and Department of Artificial Intelligence

May 1998

# Abstract

The *General Computational Theory of Musical Structure (GCTMS)* is a theory that may be employed to obtain a structural description (or set of descriptions) of a musical surface. This theory is based on general cognitive and logical principles, is independent of any specific musical style or idiom, and can be applied to any musical surface.

The musical work is presented to *GCTMS* as a sequence of discrete symbolically represented events (e.g. notes) without higher-level structural elements (e.g. articulation marks, time-signature etc.) - although such information may be used to guide the analytic process.

The aim of the application of the theory is to reach a structural description of the musical work that may be considered as 'plausible' or 'permissible' by a human music analyst. As style-dependent knowledge is not embodied in the general theory, highly sophisticated analyses (similar to those an expert analyst may provide) are not expected. The theory gives, however, higher rating to descriptions that may be considered more reasonable or acceptable by human analysts and lower to descriptions that are less plausible.

The analytic descriptions given by *GCTMS* may be said to relate to and may be compared with the intuitive 'understanding' a listener has when repeatedly exposed to a specific musical work. Although the theory does not make any claim of simulating cognitive processes as these are realised in the mind, it does give insights into the intrinsic requirements of musical analytic tasks and its results may be examined with respect to cognitive validity.

The proposed theory comprises two distinct but closely related stages of development: a) the development of a number of individual components that focus on specialised musical analytic tasks, and b) the development of an elaborate account of how these components relate to and interact with each other so that plausible structural descriptions of a given musical surface may be arrived at.

A prototype computer system based on the *GCTMS* has been implemented. As a test case, the theory and prototype system have been applied on various melodic surfaces from the 12-tone equal-temperament system.

# Acknowledgements

I would like to express my gratitude to:

My supervisors, Mr Peter Nelson and Dr Alan Smaill, for their guidance and support throughout this research study, and for all their invaluable comments and advice that helped me crystallise the views presented in this thesis.

My colleagues from the Faculty of Music and the Department of Artificial Intelligence at the University of Edinburgh for providing an inspiring intellectual environment in which this work matured; especially, the meetings of the AI-Music Group and the Musical Communication Colloquium have been an indispensable source of ideas.

The University of Edinburgh for making this research study possible by offering me a three-year postgraduate research award.

My friends who, with their constant support, love and good humour, have made my stay in Edinburgh unforgettable; especially, Evie Athanassiou for sharing so much with me in both stressful and happy times.

My parents and brothers for their continuous whole-hearted support throughout these years.

# Table of Contents

# Chapter 1

# Introduction

In recent years, the need for the development of theories of music based on scientific approaches emerging from disciplines such as cognitive psychology, artificial intelligence, semiotics, computer modelling, psychoacoustics and so on has been argued by a number of researchers (see Laske, 1988, 1992, 1993; Camilleri, 1992; Ashley, 1989; Leman et al., 1997; Selfridge-Field, 1990).[1]

More specifically 'the forces that nowadays pull towards an integration of the music sciences are based on a computation oriented methodology.' (Leman et al., 1997:19). Perhaps the most important aspect of introducing computational methods in musicology is that they force music researchers to formulate explicit theories about musical understanding which can subsequently be tested and substantiated by the use of computer systems. 'Electronic musicology may therefore be expected to continue to pursue the traditional goals of scholarship in both historic and systematic musicology, but in addition it is likely to raise expectations for precision, completeness, and consistency, to foster new methods of research, and ultimately to spawn new theories on the resulting sources of information.' (Selfridge-Field, 1990:305). The primary aim of constructing computational models is not to find solutions to musical problems but rather to assist the formulation of theories that describe musical activities and tasks in an explicit and consistent manner.

[1] The progression, however, in this area has been rather slow and this is due, primarily, to the difficulties of bringing together such diverse fields of inquiry. '1970 to 1973 was a period in which musicology underwent a revolution that has barely begun to bare fruit. ...not only are there few professorships for cognitive musicologists working with computers; communication between musicologists, musical engineers and cognitive scientists remains poor.' (Laske, 1993:226).

Musical theories allow the formulation of hypotheses and models which can be implemented as computer programs and then evaluated, and, conversely, results from the application of the computer programs may force the re-examination and adjustment of the initial theories. Especially the importance of theories of music for designing computer systems should be stressed: 'While it is not a prerequisite for building intelligent music systems to have a full-fledged theory of activity one wants to support, it is certainly more effective to design such systems on as much theory as one can harness.' (Laske, 1988:45).

It is herein suggested that a theory of music is more powerful in terms of its descriptive and predictive capacity and is more useful in terms of providing a framework for building computer systems if it addresses the following points:

- *Explication*. By this term Kassler and Howe (1980) refer to 'the restructuring of a process from an idea apprehended only intuitively to an unambiguous method that effects the process step-by-step, using information definitely provided.' (p.606). They suggest that 'what generally has precluded immediate delegation of a musical or musicological process to a computer is ... that explication of the process has not occurred.' (p.606). Even the most elaborate contemporary theories of music are not fully explicit and require a fair amount of intuited knowledge on the part of the musician in order to reach a plausible description of a musical work or task.

- *Generality*. The broader the scope of a musical theory the more powerful it is. Most current theories may be applied to a relatively narrow musical repertoire, i.e. they are style- or idiom-dependent. This 'raises the serious problem of the demarcation between general assumptions, applicable to other repertoires, and style-bound ones; this demarcation is often underestimated in such studies. We are of course a long way from a general theoretical and applicative framework which could be used to analyse several musical repertoires...' (Camilleri, 1992:181).

- *Induction:* As music is a very complex domain with a great variety of styles and idioms, theories that have an inductive outlook, i.e. that are capable of making generalisations by analysing existing musical works, can be more parsimonious, general and powerful. 'Hand-crafting' rules and grammars based on intuited knowledge of particular styles is a tedious task with many limitations: 'there are too many exceptions to any logical system of musical description, and it will be difficult to ensure completeness of an intuited theory.' (Conklin and Witten, 1995:52).

Most contemporary theories of music (some are examined in chapter 2) have weaknesses on one or more of the above points. The current study attempts to address these issues by proposing a musical theory that is explicit, general and inductive; this theory can be readily used to form a basis for designing computer systems. This research is strongly influenced by principles and methodologies drawn from the domains of artificial intelligence and cognitive psychology. A brief overview of the proposed theory is presented in the next section.

### 1.1 Outline of the *General Computational Theory of Musical Structure*

The *General Computational Theory of Musical Structure (GCTMS)* is a theory that may be employed to obtain a structural description (or set of descriptions) of a musical surface. This theory is independent of any specific musical style or idiom, and can be applied to any musical surface.

The musical work is presented to *GCTMS* as a sequence of discrete symbolically represented musical events (e.g. notes) without higher-level structural elements (e.g. articulation marks provided by the composer or by a performer, or time-signature etc.) - although such information may be used constructively to guide the analytic process.

The aim of the application of the theory is to reach a structural description of the musical work that may be considered as 'plausible' or 'permissible' by a human music analyst. As style-dependent knowledge is not embodied in the general theory, highly sophisticated analyses (similar to those an expert analyst may provide) are not expected. The theory should, however, give higher rating to descriptions that may be considered more reasonable and acceptable by human analysts and lower to descriptions that are less plausible.

The analytic descriptions given by *GCTMS* may be said to relate to and may be compared with the intuitive 'understanding' a listener has when repeatedly exposed to a specific musical work (the listener need not be familiar with the particular style or idiom the work belongs to). Although the theory does not make any claim of simulating cognitive processes as these are realised in the mind, it does give insights into the intrinsic requirements of musical analytic tasks and its results may be examined with respect to cognitive validity.

The proposed theory comprises two distinct but closely related stages of development: a) development of a number of individual components that focus on specialised analytic musical

tasks - such as the *General Pitch Interval Representation (GPIR)* and transcription algorithm, the *Local Boundary Detection Model (LBDM)*, the *Accentuation and Metrical Structure Models*, the *String Pattern-Induction Algorithm (SPIA) and Selection Function*, the *Unscramble* category formation algorithm - and, b) development of an elaborate account of how these components relate to and interact with each other so that plausible structural descriptions of a given musical surface may be arrived at - for instance, the inter-relation between *LBDM,* and *SPIA and Selection Function* for the segmentation of a musical surface or the influence of some of these components on the reduction of the musical surface.

A prototype computer system based on the *GCTMS* has been implemented. As a test case, the theory and prototype system have been applied on various distinct melodic surfaces from the 12-tone equal-temperament system.

The overall form of the theory is illustrated in figure 1.1. A *musical surface (0)* composed of discrete events (e.g. notes) is converted to a *musical surface (1)* which comprises a number of musical interval profiles at a number of levels of abstraction (e.g. for pitch: exact pitch intervals, scale-step intervals, step-leap intervals, contour; and also various profiles of time-intervals, dynamic-intervals, chord-intervals etc.). Especially for pitch, this conversion can be achieved, for instance, by the use of the *General Pitch Interval Representation (GPIR)*.

At the next stage, a process for discovering potential local boundaries is employed (for this task the *Local Boundary Detection Model (LBDM)* has been developed and may be used). Local discontinuities and changes can provide cues as to possible points where local boundaries may be detected. Following the assumption that notes that are immediate neighbours of stronger boundaries will tend to be perceived as being more prominent, the accents of individual events/notes may be calculated. It is hypothesised that these accents are the key to determining low-level metrical structure (e.g. (sub)beat level or the level immediately above the beat level) - if one exists.

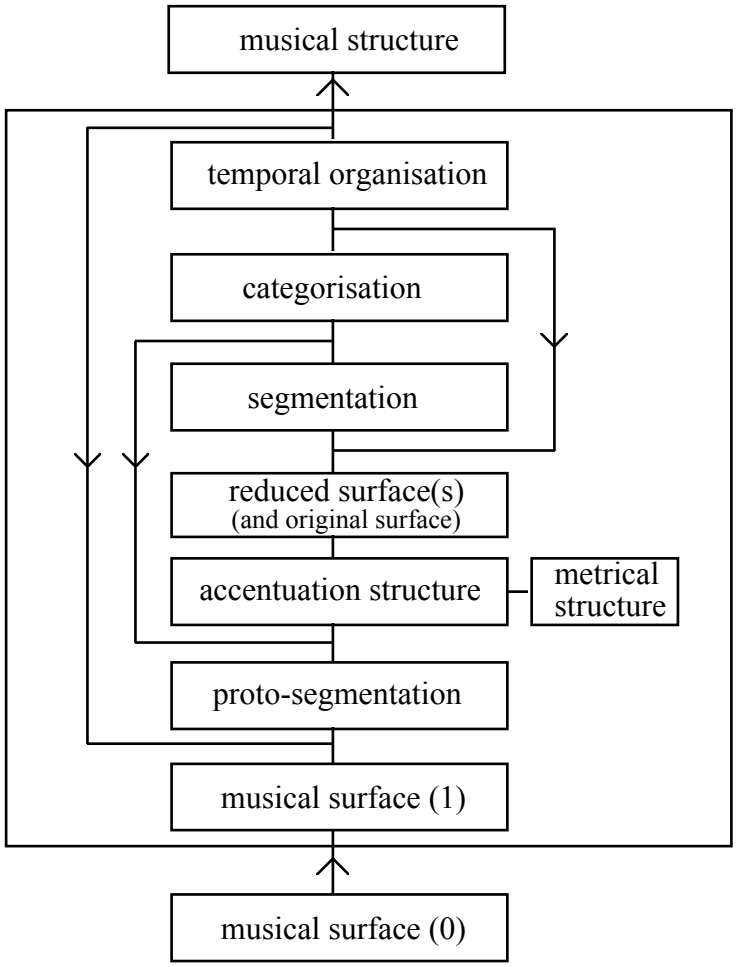



Figure 1.1 Overall form of *GCTMS*

The proto-segmentation provided by the local boundary detection component (e.g. *LBDM)* is tentative and has to be complemented by higher-level processes if a more integrated segmentation[2] is aimed at. Such a higher-level component (for instance, the *String Pattern-Induction Algorithm & Selection Function*) relies heavily on the notion of musical parallelism and similarity - recurring musical patterns are highlighted into perception and suggest

[2] The term *segmentation* refers in this text to the partitioning of a musical surface which may contain ambiguous boundaries - possibly suggesting overlapping of segments - and which is not necessarily regular.

boundaries that may be compatible or contradicting with locally detected boundaries. When the two components are coupled together a more comprehensive segmentation may be achieved.

As low-level structural properties of the musical surface have previously been revealed it is possible to apply the parallelism component on reduced versions of the surface as well (e.g. notes on metrically strong positions, more accented notes etc.). This enables 'deeper' similarities to be established.

Once a segmentation (or set of segmentations) has been obtained, musical segments are organised and labelled into categories based on their similarity (e.g. by the application of the *Unscramble* algorithm). The 'goodness' of the resultant categorisation descriptions may determine which segmentation amongst alternative segmentations should be preferred. The discovered categories can then be organised syntagmatically in terms of their ordered in-time relations (not examined in the present study).

Finally, the *GCTMS* can be applied on the new sequence of labelled musical segments (e.g. motives) so that even higher-level structural descriptions may be derived.

### 1.2 Uses of the *General Computational Theory of Musical Structure*

The proposed theory will be useful in the following areas:

- *Musical Theory:* The *GCTMS* raises interesting issues in the domain of musical theory as it provides a general underlying theory for describing musical structure and it reveals and highlights links between seemingly unrelated specialised theories of various musical idioms.

- *Musical Applications:* For computer systems to respond musically to musical users, they too must 'understand' musical structure. 'Intelligent' computer systems may be developed based on the *GCTMS* to be used in the domains of musical education, musical analysis, composition, interactive human-machine performance, musical information retrieval, artistic enablement for disabled users and so on.

- *Artificial Intelligence:* The proposed models and algorithms are of particular interest to the domains of knowledge representation, machine learning, and pattern matching.

Especially a novel unsupervised machine learning algorithm may prove useful for categorisation tasks in general non-musical domains.

- *Musical Cognition:* This theory also gives insights into the (mainly unconscious) cognitive processes that take place in the human mind when listening to music, especially to musical cognitive problem domains such as Gestalt perception, musical rhythm, musical similarity and category formation (the various predictions made by *GCTMS* might be tested, in the future, against empirical experimental data resulting from psychological experiments).

### 1.3 Outline of the thesis

A brief description of each chapter of the current thesis is given below:

*Chapter 2:* Three contemporary musical theories and three computational models that relate to the proposed theory are presented; various aspects of these theories/models are highlighted that provide useful insights or problem domains that need to be addressed by the current theory.

*Chapter 3:* The principles, methodology and scope of the proposed theory are discussed followed by an overall description of the *General Computational Theory of Musical Structure*.

*Chapter 4:* The cognitive and logical foundations of *GCTMS* are presented with special attention on the notions of identity, similarity and categorisation.

*Chapter 5:* Issues relating to finding an adequate representation for the musical surface are discussed; the focus of this chapter is the *General Pitch Interval Representation*.

*Chapter 6:* Microstructural aspects of the musical surface are presented that provide the means for determining a proto-segmentation of the surface *(Local Boundary Detection Model)* and its metrical structure.

*Chapter 7:* The notion of musical parallelism/similarity is explored and a pattern-matching technique is developed for determining significant parallel musical passages *(String Pattern-Induction Algorithm* and *Selection Function)*. The integration of micro- and macrostructural information for determining an overall segmentation of the surface is also described.

*Chapter 8:* The *Unscramble* machine learning algorithm is described; this algorithm groups similar musical segments into pertinent musical categories/paradigms highlighting at the same time the most characteristic musical properties of each category.

*Chapter 9:* A detailed account of how the various components of the theory interact with each other is given and four analyses obtained by the application of a computer system based on the *GCTMS* on four melodic examples from diverse musical styles is presented.

*Chapter 10:* A discussion of the relative merits and problems of the proposed theory is given and a number of possible further developments are suggested.

Research material from this thesis has been published in a number of conference proceedings and academic publications (Cambouropoulos, 1996a, 1996b, 1997a, 1997b, 1998; Cambouropoulos and Smaill, 1995, 1997).

# Chapter 2

# Background and Related Work

**Introduction**

Three contemporary musical theories have been selected for drawing direct and indirect parallels and comparisons with the proposed computational theory. These theories provide a general background for musical analysis - with a cognitive perspective - and share with the current proposal some of the aims outlined in the previous chapter. The first is *Paradigmatic Analysis* (Nattiez, 1975, 1990) that provides a general methodology for decomposing a piece of music into classes/paradigms of 'significant' units. The second is the *Generative Theory of Tonal Music - GTTM* (Lerdahl and Jackendoff, 1983) that provides a systematic description of tonal music in terms of grouping, metrical and reductional structures. And, the last is *The Implication-Realisation Model* (Narmour, 1990, 1992a) that attempts to describe primarily style-independent bottom-up processes in melodic perception.

In addition, three analytic-compositional musical models implemented on the computer are examined as to their relations with the proposed system. In the first model, the real-time interactive system *Cypher* (Rowe, 1992, 1993), analysis is being pursued dynamically as new events enter the system, whereas in the other two models the analytic system has access to any component part in any order within one or more musical works. These two systems are, *Experiments in Musical Intelligence - EMI* (Cope, 1991, 1992a) which takes primarily a non-linear hierarchical structural approach and the predictive musical model developed by Conklin & Witten (1991, 1995) which takes a linear informational approach.

In Table 1, each of the above theories and models is depicted along with its main musical analytic components and capabilities; only those aspects that relate to the proposed *General*

*Computational Theory of Musical Structure - GCTMS* are shown. All these theories start with a symbolic representation of musical events (viz. notes) and then continue with more or less formal descriptions of how various analytic tasks may be achieved. Apart from the musical surface, these theories often require or presuppose other externally defined ('external' in table 2.1) analytic input (e.g. metre, harmonic description, segmentation etc.). Blank entries in table 2.1 indicate analytic aspects that

| | *Paradigmatic Analysis* | *GTTM* | *Implication Realisation Model* | *Cypher* | *EMI* | *Predictive Model* | *GCTMS* |
|---|---|---|---|---|---|---|---|
| Note Representation | Traditional Notation (mainly) | Traditional Notation | Traditional Notation (mainly) | MIDI (plus elementary categorisations) | MIDI pitch quantised time | MIDI pitch quantised time | GPIR quantised time |
| Metre | External | Metrical Structure | External | Beat-Tracking Algorithm | External | External | Metrical & Accentuation Model |
| Harmony | External | External | External | Chord-Key Finding Algorithms | External | External | |
| Segmentation | External | Grouping Structure | Closure Conditions (mainly) | Based on L&J's Grouping | Based on Grammar (ATN) | External | LBDM |
| Segmentation (repetition similarity) | Semi-formal Intuitive | | | | | | SPIA & Selection Mechanism |
| Categorisation of segments | Semi-formal Intuitive | | | Real-time Partial Pat. Matching | Partial Pattern Matching | Implicit | *Unscramble* Algorithm |
| Temporal Relations/ Functions | (Syntagmatic Analysis) | Prolongation Reduction | Implication Realisation Processes | | Statistical model & Grammar | Information Theory Model | |
| Surface Reduction | External | Time-span Reduction | Transformed Tones | | Based on Grammar (ATN) | External | Accentuation Model |
| Musical-Idiom Dependency | Independent | Mainly Tonal | Independent | Partially Tonal | Partially Tonal | Independent | Independent |

Table 2.1 Brief comparison of various theories and computational models of music (top row) as to their musical components and capabilities (first column). Note the absence of descriptions for musical similarity and categorisation processes.
(Last column: abbreviations explained in figure 3.1 and blank entries discussed in section 10.2)

are not relevant to or are not embodied in the theory or model (the missing components of the *GCTMS* for harmony and temporal organisation are briefly discussed in section 10.2).

A software toolkit for musical analysis is also reviewed that is quite different from all the above in that it is not bound to any specific musical theory for analysis, but simply provides a

general computer format and toolkit with which a user may specify and achieve a great variety of musical analytical tasks.

### 2.1 *Paradigmatic Analysis*

Paradigmatic analysis (Nattiez, 1975, 1990; see also Cook, 1987; Monelle, 1992) is the first stage of semiotic analysis whereby a musical work is segmented and organised into paradigms/categories of 'meaningful' musical units - the temporal relations of these units are disregarded at this stage. The second stage (syntagmatic analysis) involves the description of the temporal distribution and organisation of these 'significant' components. The proposed computational theory mainly addresses issues relating to paradigmatic analysis, as this is in some sense a pre-requisite for syntagmatic analysis and has also resisted full formalisation that may allow the implementation of sophisticated computational musical analytic systems.

Nattiez's attempt to systematise musical analysis introduces three distinct but closely related levels (Nattiez, 1975, 1990) at which analysis may be pursued: a) the neutral level (i.e. immanent configurational properties of a musical work), b) the poietic level (i.e. compositional procedures and intentions) and c) the aesthesic level (i.e. interpretation and perceptual processes). More specifically Nattiez proposes the following primary definition of analysis at the neutral level: "This is a level of analysis at which one does not decide a priori whether the results generated by a specific analytic proceeding are relevant from the aesthesic or poietic point of view. The analytic tools used for the delimitation and the classification of phenomena are systematically exploited, until they are exhausted, and are not replaced by substitutes until a new hypothesis or new difficulties lead to the proposition of new tools. 'Neutral' means both that the poietic and aesthesic dimensions of the object have been 'neutralised', and that one proceeds to the end of a given procedure regardless of the results obtained." (Nattiez, 1990:13). Laske (1977) suggests that the neutral level is a 'methodological artefact' that 'makes it possible for the aesthesic interpreter, to hypothesize a repertory of syntactic relationships from which, in a second step, elements of poietic and/or aesthetic relevance can be selected.' (pp. 221-222).

But what exactly are the 'analytic tools' and 'procedures' that can be used to obtain an analysis at the neutral level? Nattiez[3] adopts the paradigmatic technique proposed by Ruwet (1987)

[3] 'I shall show (...) that the paradigmatic technique suggested by Ruwet, in the tradition of Lévi-Strauss and Jacobson, allows us indeed to analyze a good number of relationships between musical units. Having reached moment *y* in a musical work, we tend to establish a connection with an *x that has already been*

whereby relationships between musical sequences are established mainly because of recurrence and repetition (with or without variants). But can such relationships be established in a true neutral manner (that is, without recourse to aesthesic or poietic processes)? It is suggested herein that if similarity (i.e. not merely exact repetition) is taken into account then analysis at the neutral level becomes unwieldy because any two musical sequences are similar in some respect (see section 4.5). Analysis at the neutral level is useful only if guided by some sort of heuristics - for instance, based on general cognitive principles.

Nattiez seems to acknowledge indirectly the fact that analysis purely at the neutral level is essentially intractable by stressing the interdependency of the three levels of analysis: 'Analysis never stops engineering a dialectical oscillation among the three dimensions of the object. Analysis at the neutral level is dynamic; it displaces itself constantly as the analysis takes place...' (Nattiez, 1990:32). In doing so he seems to introduce human intuition as a necessary component of paradigmatic analysis. In this sense analysis at the neutral level is a methodological device ('methodological artefact' in Laske's words[4]) that enables a human analyst to reach an analysis rather than a systematic theory for analysis that can produce musical analyses in its own right; it is mainly an analytic methodology that forces an analyst to make their own decisions and judgements explicit rather than a general formal analytic theory that provides a set of explicit representations and procedures which may lead to pertinent analyses.

Paradigmatic analysis has been mainly applied to melodic surfaces (e.g. Ruwet, 1987; Nattiez, 1975, 1982; Lidov, 1980; Morin, 1979; Guertin, 1981). It can, however, be extended to other aspects of musical works - for instance, the overall methodology of pitch-class set theory (Forte, 1973), which is mainly concerned with atonal harmony, has significant points of resemblance (Cook, 1987:152,178; Nattiez, 1990:140).

Some practical difficulties in the application of the paradigmatic methodology to the analysis of melodic surfaces are discussed below. These relate to complex issues such as the selection of important musical parameters for the description of musical entities, the hierarchic organisation of musical structure and the segmentation of a musical surface.

---

*heard*. Analysis of the neutral level allows us to categorise possibilities for establishing these relationships. (In this, analysis of the neutral level may constitute a preliminary to aesthesic analysis.)' (Nattiez, 1990:116)

[4] Nattiez (1990:31) endorses Laske's terminology.

The set of features that is important for classifying the musical units of a specific musical work into paradigms is defined in an *ad hoc* manner; each piece of music requires a specially compiled list of features that are relevant for the particular musical context.[5] The paradigmatic methodology does not suggest a general set of features or at least a general strategy as to how such features may be selected.

The more hierarchically structured the elements of a musical surface are, the harder it is usually to perform a paradigmatic analysis of it. This is due to the fact that not only has one to determine a list of features that is relevant for the analysis of the musical surface but also a set of pertinent reductions of the surface at a number of hierarchical levels and a list of features that is relevant for the analysis of each reduction. The additional difficulty lies mainly in the need to determine a set of explicit criteria for distinguishing between more or less structurally salient events at a number of hierarchical levels that may lead to the construction of reduced versions of the surface - no such criteria are provided by the paradigmatic technique.

Perhaps the most difficult aspect of paradigmatic analysis relates to the segmentation of a musical work (this problem is also true of pitch-class set analysis). If this is taken to be a pre-requisite (produced perhaps intuitively by the analyst) then a decisive stage of the analysis lies outside the paradigmatic programme. If, on the other hand, segmentation is taken to be an emerging property of the taxonomic process then this is manageable only in the simplest cases where music exhibits a considerable amount of exact repetitions[6] (melodies on which algorithmic methods such as Ruwet's 'machine' can be successfully applied are extremely simple in the first place and quite rare as well).

The relations between segmentation, similarity and categorisation are quite complex especially when it comes down to developing a computational model.[7] On the issue of

[5] '... wouldn't semiotic analyses be more useful if they all used the same list of features so that one analysis could be directly compared with another in detail? The justification (which I don't consider wholly convincing) is that the purpose of such a list is to identify the features that are important for the relationships between units within the particular context of a given piece or repertoire of pieces; hence the list of features has to be compiled especially for each application.' (Cook, 1987:172).

[6] 'What, then, happens if the relation between segments is not one of simple recurrence at all but of some more complex transformational relation? The answer, of course, is that there are no criteria on which to base the initial segmentation. The result of this in practice is the limitation of semiotic analysis to such styles (Debussy, imitative counterpoint, certain exotic musics) as are characterised by literal repetition. The limitation is not very compatible with the aim of creating a general theory of sign structures in music.' (Cook, 1987:180)

[7] In computational terms, it may be said that the main difficulty with paradigmatic analysis is one of tractability. Although Ruwet and Nattiez propose a method for constructing a 'good' paradigmatic description (i.e. a small number of distinct paradigms that cover most of the musical surface) in the

similarity Nattiez states: 'It is hard to see how a computer could automatically establish an equivalence which depends on a judgement of similarity transcending concrete resemblances and differences.' (Nattiez, 1982:257). Taking this statement as a challenge rather than as a deterrent, a significant amount of the current study is devoted to developing a formal theory that can automatically produce a segmentation concurrently with establishing similarity relations between melodic segments and forming a taxonomic description (see especially chapters 7, 8 & 9).

### 2.2 The *Generative Theory of Tonal Music (GTTM)*

Lerdahl and Jackendoff (1983) propose a generative theory that accounts for the intuitions of experienced listeners in the tonal idiom.[8] The main components of the theory are: grouping structure, metrical structure, time-span reduction and prolongation reduction: "... *grouping structure* expresses a hierarchical segmentation of a piece into motives, phrases, and sections. *Metrical structure* expresses the intuition that the events of a piece are related to regular alternation of strong and weak beats at a number of hierarchical levels. *Time-span reduction* assigns to the pitches of the piece a hierarchy of 'structural importance' with respect to their position in grouping and metrical structure. *Prolongation reduction* assigns to pitches a hierarchy that expresses harmonic and melodic tension and relaxation, continuity and progression." (Lerdahl and Jackendoff, 1983:8-9) - grouping and metrical structure are further discussed in section 6.1. The theory is developed in a rather formal manner and rules are divided into two distinct types: well-formedness rules that define possible structures and preference rules that specify descriptions that correspond more closely to listeners' intuitions. Many aspects of the *GTTM* have been supported by experimental studies (Deliège, 1987; Bigand, 1990). Jackendoff (1992) has shown more recently how the hierarchic structural *GTTM* can form a basis for a processing model.

The *GTTM* attempts to describe musical structure by adopting a stance that is influenced by linguistic theory. In doing so, it may be argued that it sometimes gives rise to formalisms that do not seem to reflect musical structure in the most adequate way. For instance, the well-formedness rules are unnecessarily rigid (see section 7.2). It will be maintained in this study that strict well-formed tree-like structures should not be considered as the norm (with possible

---

course of which appropriate features and segmentations are discovered, they do not offer a tractable algorithm for implementing this (except only in the simplest cases of surfaces consisting mostly of exact repetitions where the search space is sufficiently small).

[8] 'We take the goal of a theory of music to be a formal description of the musical intuitions of a listener who is experienced in a musical idiom.' (Lerdahl and Jackendoff, 1983:1).

divergences such as overlaps and elisions) but rather as a desirable aim for reasons of simplicity and clarity that often need not be reached.

In the *GTTM*, motivic-thematic processes are not explicitly dealt with. Parallelism, i.e. similarity of different musical groups, is stated as a preference rule influencing each of the components of the theory but no attempt is made to describe it further. For example, rule GPR6 (Parallelism) states that 'where two or more segments of the music can be construed as parallel, they preferably form parallel parts of groups.' (Lerdahl and Jackendoff, 1983:51). But when can two or more segments be construed as parallel? *GTTM* does not attempt to answer this question: 'we feel that our failure to flesh out the notion of parallelism is a serious gap in our attempt to formulate a fully explicit theory of musical understanding.' (Lerdahl and Jackendoff, 1983:53).

Grouping and accentuation structure (on which a metric grid is matched) are also unnecessarily considered independent in the *GTTM*. In sections 6.1 & 6.4 it will be argued that the two are closely linked (especially for the lower structural levels) in such a way that if one is given the other may automatically be inferred.

The inter-relations among the four major components of the theory are not clearly described. For example, in (Lerdahl and Jackendoff, 1983: figure 1.1) there are bi-directional arrows linking each component of the theory to every other component whereas in (Lerdahl, 1988: figure 1; Lerdahl, 1992: figure 11.1) there are one-directional arrows leading from grouping and metrical structure to time-span reduction and finally to the prolongation structure component (there is no arrow connection between the grouping and metrical structure components). The *GTTM* suggests some feedback links from higher level structures to lower level ones, e.g. "GPR7 (Time-Span and Prolongation Stability) Prefer a grouping structure that results in more stable time-span and/or prolongation reductions." (Lerdahl and Jackendoff, 1983:52), but no detailed description is given as to how exactly this may be achieved.

Finally, the *GTTM* is a theory of *tonal* music. However, there are aspects of the theory that are style-independent - especially the Gestalt-based grouping rules (these are reviewed in more detail in sections 6.2 and 6.3). More recently, Lerdahl (1989) attempts to adapt the *GTTM* for the description of the intuitions of experienced listeners in atonal music, but Dibben (1994) presents experimental evidence that doesn't seem to support Lerdahl's proposal.

In general, the *GTTM* is a well worked out theory and readily lends itself to further development, comparisons and experimentation, as most of its elements are spelled out in a very clear and precise manner.

### 2.3 The Implication-Realisation Model

Narmour's theory for the analysis and cognition of melody (Narmour, 1990, 1992a) is a theory that attempts to describe primarily 'the specific, note-to-note principles by which listeners perceive, structure and comprehend the vast world of melody' (Narmour, 1990:3). It is based on a small number of style-independent bottom-up general principles that interact with top-down processes relating to intra- and extra-opus knowledge acquired through previous experience.

The main focus of the theory is the bottom-up processes which are presumed to be general, innate and universal.[9] These processes interact with and are influenced by top-down processes (these include, for instance, harmony, meter, intra-opus schemata such as recurring patterns and extra-opus processes such as tonal functions etc.).[10] These top-down learned schemata are not explicitly described by the theory; they are rather considered as independent knowledge which is contributed by the analyst or listener (consciously or unconsciously).

The Implication-Realisation model firstly determines points of implication[11] (implicative intervals) in a melodic surface and then suggests a number of melodic archetypes for possible continuations (realised intervals) that may or may not satisfy implications. The notion of implication has opposite effects to the notion of closure, i.e. implication is weak when closure is strong and implication is strong when closure is weak. Small primitive melodic structures

[9] '... the theory will analyse (and thus partly explain) all melodies ever written or to be written, regardless of stylistic origin. What this surprising assertion means is that the hypotheses of the theory operate independently of any specific style structures, of any learned, replicated complexes of syntactic relations.' (Narmour, 1992a:7). 'Innate, inborn rules govern bottom-up simplex relations (and are thus constant), whereas top-down learning governs complex structural relations (and thus varies from listener to listener).' (Narmour, 1992a:11-12).

[10] "Narmour refers to the pervasive influence of learned, 'top-down' idiom- and style-specific schemata that the listener consciously brings to bear on these [Gestalt-based 'bottom-up' parametric] style shapes: these could include the influence of explicit or implied harmony, which might manifest simply as the listener's awareness of scale-step; the influence of duration and of meter; and the influence of intra-opus and extra-opus style." (Butler, 1992:248).

[11] " 'Implication' is an objective term referring to demonstrable analytical patterning in a piece of music, whereas 'expectation' is a subjective term denoting the listener's psychological response to such a patterning. In other words, from the listener's point of view, one could call the implication-realization model the 'expectation-confirmation model.' " (Narmour, 1992b:69).

can be combined for the description of larger, more complex structures (for instance, notes on which strong closure takes place may be transformed into elements of a higher reduced structural level and may determine grouping boundaries and new implicative intervals on higher levels). Concise descriptions of the Implication-Realisation model can be found in (Krumhansl, 1995, 1997; Butler, 1992; Cross, 1995; Narmour, 1992b). A number of studies seem to support the formulation of some of the bottom-up processes of the model (e.g. Krumhansl, 1995, 1997; Thompson et al., 1995-96).

It is herein suggested that Narmour's theory makes a rather too strong distinction between bottom-up (invariant) and top-bottom (variable) processes (see also Cross, 1995); more importantly, it gives too much emphasis to the formal description of the bottom-up note-to-note pitch processes leaving perhaps unnecessarily too much space for top-down intuitive or semi-intuitive 'except-cases' usually marked as *os* (intra-opus style) and *xs* (extra-opus style).[12] In the computational theory proposed in this study low-level note-to-note processes are complemented by gradually higher-level factors (mainly intra-opus information) in a rather continuous and integrated manner (especially chapters 3 & 9) - it is asserted that this may lead to more coherent and systematic descriptions that depend less on external intuited contribution from the musical analyst.

In the Implication-Realisation model the metrical and rhythmic aspects of melodic processes are not clearly described; their influence is taken into account but no separate theory of metrical and rhythmic structure is given[13] (a similar comment applies to the treatment of harmony). For instance, 'closure' - on which grouping and transformation of notes to higher-levels are based - relies on the interaction of factors such as metrical position, duration (a rest or a short duration followed by a longer one), harmony (dissonance followed by consonance) and pitch (a large pitch interval followed by a smaller one); these factors - especially the way they interact with each other[14] - are not the focus of the Implication-Realisation model and no attempt is made to describe them formally (a formal model that attempts to detect local

[12] 'The if-then, formalizable constants ... govern style shapes (primitive parametric simplexes). The influence of style structures (multiparametric complexes) on such constants is an except-condition (if-then-except).' (Narmour, 1992:168).

[13] 'Given the formality of much of the rest of the theory, his [Narmour's] treatment of rhythm and meter appears too discursive. ... It could have been helpful to articulate theories of meter and rhythm independently before instancing the influence of metrical and durational factors in the overall implication-realization model.' (Cross, 1995:502).

[14] In attempting to describe roughly the interaction of parameters (especially the influence of melody, duration, metric emphasis and dissonance on melodic closure/implication) Narmour states: 'Since formalizing parametric interactions in these terms is beyond the scope of this book, the rules that follow, therefore, are largely pragmatic - informal methodological ones." (Narmour, 1992a:364).

grouping boundaries in a melodic surface and determine metrical structure is described in chapter 6).

As the Implication-Realisation model is primarily concerned with the note-to-note sequential in-time flow of the melodic surface the description of outside-time structural relationships and classifications of the melodic material is essentially absent. According to the theory, similarity of form (especially repetition) plays a significant role in grouping and in low- and high-level implication (Narmour, 1992a:129-132,300-308; see also Krumhansl, 1997) but no attempt is made to describe *when* two melodic patterns may be considered similar (musical similarity is extensively discussed and described in the current thesis - see especially chapters 5, 7, 8 & 9).

The overlapping of successive melodic structures in Narmour's theory reflects musical progression and ongoingness. In contrast to other theories such as Lerdahl and Jackendoff's *GTTM* 'Narmour treats overlap as the norm, the exception being separation' (Cross, 1995:506). The current proposal endorses this view (see especially section 7.2) although, when possible, non-overlapping descriptions of melodic surfaces are preferred to overlapping ones for reasons of clarity and economy.

The generality of the Implication-Realisation model is based on three basic theoretical constants: 'that A+A implies A (i.e., that sameness or similarity causes the subconscious expectation of more sameness or similarity, all other things being equal); that A+B implies C (i.e., that differentiation causes the expectation of further differentiation); and that the definition and evaluation of these two hypotheses in both cognition and analysis depend on syntactic parametric scales (i.e. on gradated, innate cognitive input systems).' (Narmour, 1992a:1).

The first two principles do not correspond to logical implication but are not necessarily valid as general cognitive principles either. Two successive entities do not imply *in general* any further sameness or differentiation; the only thing that can be inferred is that the two entities are either the same or different. Implication (and expectancy) is essentially a generalisation of experience[15] and is also context-dependent[16]. (For instance, if two red Mercedes cars pass

[15] Ian Cross stresses the importance of shared learned experience which is excluded from Narmour's bottom-up processes by saying: 'Narmour does not appear to consider the possibility of trans-genre stylistic constraints that may be oriented around some constant structural core ... through our exposure to the music of the past five centuries.' (Cross, 1995:504). And he continues: "Perhaps Narmour's idea of a large interval as implying 'change of registral direction and a sequence of intervalic differentiation' is

successively in front of a viewer no expectation for a further red Mercedes car is created - on the contrary, one would be surprised if one or more red Mercedes cars did follow! - or - in music, a sequence of two successive ascending sixth melodic intervals doesn't seem to imply a further ascending sixth interval).

The third hypothesised constant is also unnecessarily rigid: "... a syntactic parametric scale is an automatic, 'brute' input system that is domain specific, mandatorily operative, and computationally reflexive... It determines what is similar (A+A) or differentiated (A+B)." (Narmour, 1990:4). It is maintained in this study that similarity and differentiation strongly depend on previous experience and on current context, and that the definition of general concrete thresholds (Narmour, 1992a:15-19) for similarity/differentiation is unwarranted and arbitrary.

Although it is possible to hypothesise general logical or cognitive principles as a basis for a theory of music (e.g. the principle of identity/change), Narmour's hypotheses do not seem to be the best candidates (see chapter 4 for a discussion on the general logical and cognitive principles that form the basis of the current proposed theory).

### 2.4 Computational models

#### 2.4.1 *Cypher*

The real-time interactive music system *Cypher*, developed by Rowe (1992, 1993) consists of two major real-time components: a listener (analytic module) and a player (compositional module). The listener component analyses incoming musical data (MIDI) and the player component responds to this information generating new relevant musical material. The listener classifies input data as to different parametric features (e.g. speed, density, dynamics, beat, harmony on a lower level, and regularity of lower level features of phrases on a second level), and the player reacts to this analysed data moulding it into new musical structures. The listener and player modules are relatively independent and an interface is provided to enable the user to configure the ways that the player should react to the messages sent by the listener.

---

better conceived of as being derived from the examination of Western classical musical style structures rather than from any specific and innate properties of our cognitive systems." (Cross, 1995:507).

[16] Meyer states that 'the implicative effect of repetition depends upon context. For instance, if a reiterated pattern is understood to be part of an ostinato or a ground bass, we do not necessarily expect change. Similarly, repetition in a coda or of a cadential figure repeated as an echo, has quite different effect from repetition which is understood to be part of an on-going process.' (Meyer, 1973:51).

Each component of *Cypher* is a network of interconnected agents operating on various hierarchic levels. *Cypher's* relation to the current theory revolves mainly around the design and implementation of its listener module.

*Cypher's* listener module attempts to make generalisations on the input musical data so that the knowledge acquired may be used by the player component for composing new material. The approach to classification incorporated in *Cypher* often assumes absolute context-independent thresholds set prior to the application of the system. For instance, a partial pattern-matching algorithm[17] is applied in *Cypher* for musical pattern classification in which an absolute threshold is set as a criterion for determining a successful match (two patterns are said to match if at least 4 of their elements match - maximum length of patterns is predefined).[18] One of the main claims made in the current proposal is that similarity and categorisation always depend on context and that fixed absolute thresholds may give rise to dubious results (see especially section 4.5 and chapter 8). Classification of musical patterns in *Cypher* also requires a pre-determined segmentation; the influence of similarity and classification on segmentation is absent from the system (see chapter 7 for an integrated approach to segmentation). Finally, classification in *Cypher* does not take into account patterns of reduced versions of the musical surface that may reflect relations between structurally prominent events; this is perhaps due to the fact that a *partial* pattern matching technique is employed (see section 7.3).

Although the intention of *Cypher* is to be a general interactive compositional system, it has partial orientation towards the Western tonal system. For instance, vertical organisation of pitches is based on tonal harmonic relations of chords in a specific key and specification of grouping boundaries is biased strongly towards tonic and dominant cadential function of chords.[19]

---

[17] This algorithm is based on a performance-to-score matching technique (Bloch & Dannenberg, 1985) which is applied on an absolute pitch representation of a score; Rowe has extended this technique for pattern-induction and pattern-matching on interval representations of a score (Rowe, 1993, 1995; Rowe & Li, 1995).

[18] 'Each element from the larger pattern is successively sent to the matcher, always to be matched against the smaller pattern. ... if at least 4 elements from the smaller pattern were also found in the larger one, the induction is successful, and an attempt is made to add the new entry to the list of known patterns. Now the newly induced pattern must be compared to those already known; accordingly, it is matched against all the patterns already in memory. If the rating after matching the new pattern against a known pattern is 4 or more, the patterns are considered to be the same.' (Rowe, 1993:246).

[19] 'The harmonic sense implemented here models a rather simple version of Western tonality.' (Rowe, 1993:134) 'Following the conventions of Western tonal harmony, tonic and dominant functions are given more weight as potential phrase boundaries than are chords built on other scale degrees.' (Rowe, 1993:155).

In general, as *Cypher* attempts to tackle many aspects and levels of musical analysis in *real-time,* it is led to only generating a simplified analysis of the input musical structures (especially as far as higher-level organisation is concerned). In the trade-off between interactive real-time pragmatic efficiency and elaborate exhaustive analytic expressiveness, this system is biased towards the former.

### 2.4.2 *Experiments in Musical Intelligence (EMI)*

*Experiments in Musical Intelligence*, developed by Cope (1991, 1992a, 1993), is a computer model of musical composition based on style analysis of a composer's body of works. This system focuses on the replication of works in the style of an individual composer, which is grounded on the observation that composers tend to reuse musical patterns throughout their corpus of compositions. The system requires at least two compositions in a similar style from which it induces 'musical signatures'[20] and rules for composition (mainly statistical analysis).[21] In the composition phase, "the program 'fixes' [signatures] to their same locations in an otherwise empty form based on the form of the first of the input works." (Cope, 1993:407). The intervening spaces between signatures are composed based on the rules discovered by the statistical analysis. 'Proper interpolation of this new music relies on an Augmented Transitional Network (ATN). By following protocols similar to those found in linguistics, the program orders and connects appropriately composed materials and fleshes out a new work.' (Cope, 1993:407). The works generated by this model resemble quite successfully music in the style, for instance, of Bach, Mozart, Brahms, Prokofiev, Joplin etc. - see review of CD released with works composed by *EMI* (Vantomme, 1995).

The input works are presented to Cope's model 'as separate lists of phrases of MIDI note numbers.' (Cope, 1992a), i.e. a preliminary segmentation of works is externally defined at the level of phrase structure (in contrast, the current model assumes no initial segmentation). The discovery though of 'signatures' in EMI contributes to musical segmentation at the motivic level by determining important musical patterns. In general, EMI does not attempt to describe an integrated segmentation strategy whereby a musical surface may be broken down to

[20] A *signature* is a set of contiguous intervals (i.e., exempt from key differences) found in more than one work by the same composer.' (Cope, 1991:46).
[21] '[Musical rules analysis] is a series of mathematical subprograms that compute percentages of certain aspects of music such as voice leading directions, use of repeated notes, triad outlining, leaps followed by steps, etc.' (Cope, 1993:406).

'significant' components in terms of both local discontinuities and higher-level musical parallelism (see especially section 7.6 & 7.7).

Perhaps the most interesting aspect of Cope's work, as far as the current proposal is concerned, is the 'signature' discovery methodology. *EMI* employs an exhaustive pattern-matching mechanism on the input musical surfaces - i.e. the matching process shifts in a step-wise manner throughout the sequence of events and all the possible patterns are considered (see Cope, 1990). The match between two patterns may be full or usually partial (see section 7.1 on advantages and disadvantages of partial pattern matching techniques in music); the pattern-matching process is guided by a number of 'tuners', such as 'pattern-size', 'range-tolerance' (the amount by which a given interval may be incorrect during pattern-matching) and 'error-tolerance' (the amount of non-matching between patterns that is accepted) that are set by the user prior to the application of the system. This approach would be practical if it were an intuitively straightforward procedure to define these variables; usually though this is not the case, as the size of patterns can vary significantly even within the same piece and the limits and kinds of variance are context-dependent and difficult to select and define. A built-in procedure that attempts to discover and suggest the most appropriate sizes and kinds of variance (pertinent similarity judgements) that are most relevant to the analysed piece(s) would be of significant help to the user (see especially section 4.5 and chapter 8).

*EMI* relies on a grammar which follows an idiom-specific protocol of musical functions and hierarchic relations (primarily a classical tonal protocol). But 'in *EMI*, one may vary the interpreter protocols. ... This has the effect in tonal music of establishing new arrangements of chords so that tonic need not follow dominant. It can force a new logic into non-tonal musics.' (Cope, 1991:216). These protocols are externally defined - based on previously acquired musical knowledge - and have an overall 'tonal' outlook - even though they may be 'stretched out' to partially cover other musical systems; in this sense, *EMI* is not a genuinely general analytic-compositional model.

#### 2.4.3 A Predictive Musical Model

The analytic-synthetic system developed by Conklin and Witten (1991, 1995) is a computational model in which style analysis is based on an empirical induction approach, whereby the description of a style is developed through the analysis of a corpus of existing compositions, rather than on a knowledge engineering approach whereby musical knowledge about a specific style is 'hand-crafted' into a system in terms of explicit rules and constraints.

A large number of training cases is presented to the system from which a description of the musical style is gradually built. The analytic approach incorporated in this model is grounded on information theory and predictive musical theories (especially Meyer, 1956, 1957).

A musical piece is viewed from different perspectives (multiple-viewpoint approach) which contribute to an overall predictive profile. Prediction of the next-event is reflected in 'the entropy profile of a work which measures the information flow as the piece progresses' (Witten et a1., 1994:70). A system of multiple-viewpoints (i.e. the combination of individual viewpoints) for which the entropy estimate is minimum[22] is considered to be a better description of a style and has better predictive power than other alternative multiple-viewpoint systems. A long-term model represents the general musical style and a short-term model the details of an individual piece. This computational model has been applied in the description of the information content of the Bach chorale melodies; a comparison to human music predictive capabilities is given in Witten et al. (1994).

The overall approach of Conklin and Witten's model relates to the theory proposed herein in terms of its neutrality as to specific musical systems, its inductive outlook and its multiple-viewpoint analytic procedure. Conklin & Witten's model may be of special interest when describing the temporal component of the proposed computational theory (not as yet described; see more on future work in section 10.2).

Perhaps the most significant difference of Conklin and Witten's model to the *GCTMS* is that this model requires pre-defined viewpoints on a number of levels of musical structure whereas the proposed theory gradually builds such viewpoints. For instance, Conklin & Witten - in applying their model to Bach choral melodies - presuppose primitive viewpoints that rely on 'basic types' such as 'timesig' (time signature), 'keysig' (key signature) and 'fermata' - that is, metrical structure, tonality and phrase[23] structure are defined prior to the application of the model. However, the derivation of such higher-level structural information from the musical surface by a listener is anything but trivial; the proposed theory attempts to describe how such information can be automatically inferred and then used for further analysis.

[22] 'The goal of a machine is to reduce its entropy estimate of the concept. The entropy of the chorales is a measure of the amount of nondeterminism present, and is a quantitive measure of the complexity of a musical genre. ... the predictive theory that minimizes the entropy estimate will also generate original, acceptable works.' (Conklin and Witten, 1991:2).

[23] 'Information about phrases is notated in a consistent manner throughout chorales using fermatas.' (Conklin and Witten, 1995:62).

A further point is that Conklin and Witten's main focus seems to be the creation of a learning mechanism that gradually implicitly embodies knowledge of a specific musical work or style rather than the explicit description of important musical structures that characterise a specific work or style (although this is possible if additional mechanisms are devised). A very interesting aspect of their work though is the ability of their model to explicitly determine which (combinations of) viewpoints are most significant in describing a musical work or style; this is a goal shared with the proposed theory (see chapters 4 & 8).

### 2.4.4 *Humdrum*

*Humdrum* is a formal syntax and a set of general-purpose software tools that enable musical analysts and researchers to pose questions and obtain answers about music (Huron, 1994, 1996). The *Humdrum* format is quite abstract and can accommodate an unrestricted number of concrete musical representations. A great variety of musical tasks can be achieved by interconnecting general-purpose tools each performing a simple operation (based on the UNIX 'software tools' design philosophy). Kornstädt (1995-96) states: 'Almost any analytical research task of a quantitative nature can be solved by combining the right tools. ... [Humdrum] is the most versatile and promising tool kit for computer-assisted musicological analysis' (p. 111).

*Humdrum* clearly is not a music analytic system based on a computational model of musical understanding. It does not have a general inference engine based on music theory or music cognition which can automatically generate plausible descriptions of a musical piece. The researcher has to define accurately and in an *ad hoc* manner the kind of question *Humdrum* has to answer (for instance, find all the occurrences in a given piece of a specific pitch-interval pattern under a specific set of constraints, e.g. anchored to specific metric positions). *Humdrum* is a sort of programming environment which enables users to represent musical works and to build specific analytic procedures by combining the *Humdrum* tools - 'In essence, assembling *Humdrum* command lines amounts to a form of computer programming.' (Huron, 1996:35). As *Humdrum* is very abstract, it is possible that parts of the proposed general computational model may be implemented as additional specific tools in the *Humdrum* format.

## 2.5 General Comments and Problems

Some interesting as well as problematic aspects of the above theories and models - at least as far as the current study is concerned - are described below.

a. *Surface representation.* All of the above computational models represent pitch and pitch intervals as integers (e.g. MIDI) although they often attempt to analyse tonal structures. This obscures important qualities of intervals relating to scale structures (see section 5.3) and often leads to oversimplified interpretations of the musical surface.

b. *Musical structure representation.* Some of the above theories and models are biased towards well-formed hierarchical tree-like structures, while others take primarily a linear approach whereby the note-to-note dynamic aspects of musical processes are examined and described. Finding a balance between structural hierarchic and linear dynamic aspects of musical understanding seems to be a particularly difficult task.

c. *Segmentation*. Most systematic theories of music suffer on the issue of surface segmentation (all of the above theories and models, and even formal mathematical theories like Forte's (1973) pitch-class set theory). They all require some sort of pre-processing of the surface into segments which relies on explicit/implicit knowledge on the part of the human musician/analyst.[24] Segmentation is a central part of musical analysis and it can seriously affect subsequent analysis as a selected segmentation automatically excludes a great number of inter-segment musical structures. Segmentation also relies on both low-level discontinuities in the musical surface and higher-level emerging patterns due to musical parallelism; an integrated approach that takes into account these two segmentation factors would be a significant contribution to systematic theories for musical analysis.

d. *Musical Parallelism and Musical Categories*. None of the above theories and models provides an effective sophisticated mechanism for achieving grouping of musical events in terms of musical parallelism, and then for organising musical segments into significant musical categories/paradigms. Musical parallelism and similarity is mentioned in most of these theories as a significant aspect of musical structure but only *EMI* and *Cypher* attempt to formalise it more rigorously (although with the limitations outlined in sections 2.4.1 & 2.4.2).

[24] Even attempts to formalise low-level rules for surface segmentation such as those proposed by Lerdahl and Jackendoff (1983) are anything but rigorous accounts of musical segmentation processes and do not readily lend themselves to the development of computer applications for segmentation of musical surfaces - although there are various attempts such as (Camilleri et al., 1990; Robbie and Smaill, 1995).

e. *External Input*. Most of the approaches described above have a restricted overall coverage - often because the theory or model was not meant to cover the specific area - and require extensive external input to be provided usually by the human analyst/user in the form of prior intuitive non-computational analytic input (see blank entries and 'external' entries in Table 2.1)

f. *Style-dependency*. Most of the above theories and models either provide a general style-independent methodology which requires some form of external style-dependent input or provide a partially-independent mechanism which usually is biased towards the Western tonal system. Generality and style-dependency of musical theories is a thorny issue.

**Conclusions**

Some musical theories and computer models have been outlined in this chapter and some general comments were made that may help to show the potential and capabilities of the proposed *General Computational Theory of Musical Structure* (described from the next chapter onwards). Aspects of these theories and models - especially the influential theory of Lerdahl and Jackendoff - will be examined and evaluated in more detail at the appropriate positions in the main body of the thesis. It will be shown that the proposed computational theory addresses effectively many of the problems outlined in this chapter and provides improved and/or novel models for generating pertinent structural analyses of musical surfaces.

# Chapter 3

# The *General Computational Theory of Musical Structure*

**Introduction**

The *General Computational Theory of Musical Structure (GCTMS)* is a theory that may be employed to obtain a structural description (or set of descriptions) of a musical surface. This theory makes use of a set of general cognitive and logical principles as a basis for modelling the intuitions of a listener and is independent of any specific musical style or idiom. The input to the computational theory will be presented in the form of musical surfaces (only melodic surfaces will be dealt with in this study) and the output will be a set of graded structural analyses which will be evaluated, at this stage, by an expert musical analyst as being 'acceptable' and 'plausible'.

In this chapter the following questions will be addressed briefly: What are the main characteristics of *musical structure?* What is a *computational theory*? To what extent are *cognitive* aspects of musical understanding represented in the computational theory? Is there a set of general cognitive/logical principles that can form the basis of a *general* style-independent theory of musical structure? What is the overall form of the proposed *General Computational Theory of Musical Structure?*

### 3.1 Musical Structure

Making sense of a complex musical[25] phenomenon means being able (consciously or unconsciously) to break it down into simpler components and to make associations between them (Minsky, 1993). *Musical structure* is taken here to be the organisation assigned to a musical surface in terms of its constituent parts and the relations/functions between them at various levels of description. Musical theory[26] is nowadays mainly concerned with the study of musical structure and musical analysis[27] is aimed at eliciting such structural descriptions, often from a perceptual/cognitive perspective.[28]

There are five main aspects of musical structure which the *GCTMS* attempts to formalise:

1. *Musical surface*. This is the lowest level of representation which is chosen as the starting point of analysis. In this study a musical structure is described as merely consisting of primitive atomic elements (e.g. notes or musical intervals). On a psychological level, this roughly corresponds to the level of discrete elements emerging as a result of categorical perception.

2. *Segmentation.* Perceptual discontinuities (e.g. a long note or a large melodic leap) allow a tentative segmentation (proto-segmentation) of a musical surface. Musical similarity also strongly affects the emergence of significant musical entities (e.g. motives) which in turn contribute towards a more integrated segmentation.

3. *Categorisation.* The musical surface may be described in terms of meaningful musical categories. Each musical category consists of a set of musical entities that are associated together by means of a set of criteria. For example, a set of musical segments may be considered as instances of a musical motive in that they share a number of melodic/harmonic/rhythmic characteristics and so on.

4. *Temporal organisation.* Musical categories are ordered and organised in time. It is essential to define the relations and functions between musical materials within the temporal and

[25] "The 'musical' is any sonorous fact constructed, organised, or thought by a culture" (Nattiez, 1990:67)
[26] 'Theory is now understood as principally the study of the structure of music.' (Palisca, 1980:741).
[27] Musical analysis is 'the resolution of a musical structure into relatively simpler constituent elements, and the investigation of those elements within that structure.' (Bent, 1980:340)
[28] 'Underlying all aspects of analysis as an activity is the fundamental point of contact between mind and musical sound, namely musical perception.' (Bent, 1980:341)

logical framework of a musical work. For instance, probabilistic transitional networks may be used to represent the temporal relations between musical categories (e.g. motives) at a certain hierarchic level of description.

5. *Reduction.* Some musical events are perceptually more prominent than others. These may form part of and give rise to more abstract representations of the musical surface. Segmentation, categorisation and temporal organisation can be applied to the musical surface and to a number of reductions of it. This way more sophisticated descriptions of the musical surface may arise that reflect hierarchic qualities of the musical materials.

The *GCTMS* as presented here attempts a systematic description of aspects 1, 2, 3 and 5 of musical structure with special emphasis on musical segmentation. The aspects relating to the temporal organisation of a musical work have not as yet been addressed (see section 10.2).

It will be maintained that the above description of musical structure need not result in a hierarchical non-overlapping tree-like structure as is commonly hypothesised in linguistically oriented musical theories (e.g. Lerdahl and Jackendoff, 1983). Such structures are an idealisation that may assist in highlighting some aspects of musical structure but may disregard or obscure others. In this study such 'tidy' structures will be considered only as special cases of the more flexible - but computationally more expensive - overlapping representations.

The proposed theory of musical structure is taken to be mainly concerned with a musical work in two respects: a) as it exhibits an inherent structural organisation and b) as it becomes intelligible/meaningful to a listener. The former aspect assumes an internal immanent structure that is independent of an external observer (structure at the 'neutral level'); analysis at this level often reveals logical or mathematical relations between various components that are not necessarily perceived by a listener, and usually produces unwieldy analyses. The latter aspect allows the reduction and 'filtering' of such logical possibilities to those that are most likely to be perceived by a listener. But what kind of listener is assumed in the present theory?

Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music (GTTM)* attempts to describe the intuitions of 'a listener experienced in a musical idiom' (p.1) - more specifically in the tonal idiom. "Occasionally we will refer to the intuitions of a less sophisticated listener, who uses the same principles as the experienced listener in organising his hearing of music, but in a more limited way. In dealing with especially complex artistic issues, we will

sometimes elevate the experienced listener to the status of a 'perfect' listener ..." (p.3). Narmour's (1990, 1992a) *Implication-Realisation Model (I-R Model)* attempts to describe primarily the general principles that govern bottom-up style-independent processes of melodic cognition[29] shared by all listeners (naive and experienced)[30] - it is herein argued that the *I-R Model* describes essentially the understanding of *non*-experienced listeners since this is the common denominator of both experienced and non-experienced listeners. The bottom-up processes interact with *independent*[31] learned top-down processes that vary depending on the experience of the listener; the intra- and extra-opus structural knowledge of the experienced listener is essentially an external input to Narmour's model. In these two theories the intuited knowledge of the experienced listener is either directly built into the idiom-dependent theory (*GTTM)* or is an external contribution from the musical theorist/analyst when applying a general theory *(I-R Model)* - see table 3.1.

| | | | | |
|---|---|---|---|---|
| *GTTM:* | non-experienced | ← | experienced | → ideal |
| *I-R Model:* | non-experienced | ← | experienced | |
| *GCTMS:* | non-experienced | → | experienced | |

Table 3.1 Types of *listener* assumed by different theories.

In the *General Computational Theory of Musical Structure (GCTMS)* a listener is assumed that possesses general cognitive abilities (e.g. abilities for abstraction, categorisation, boundary detection, hierarchic organisation, and so on) that are shared with other faculties of the mind (e.g. vision, language). Through exposure to and familiarisation with one or more musical works in a specific idiom the elementary non-experienced listener may acquire a more refined representation of musical structure through the activation of her/his general cognitive capacities. Gradually this listener becomes more experienced and develops more refined cognitive abilities in accordance with the information available in the surrounding musical environment. It is assumed, however, that the general cognitive abilities of an experienced listener remain intact and can always be activated when the listener, for instance, encounters and tries to understand music from a novel musical idiom.

[29] '... the implication-realization model treats melody primarily as a note-to-note phenomenon, as a continuity of melodic relations whose intelligibility fundamentally derives from lower-level, bottom-up structures.' (Narmour, 1992:330).

[30] Narmour's work explores the idea "... that a cognitive 'genetic code' enables both naive and experienced listeners to comprehend the entire world of melody." (Narmour, 1992:ix).

[31] 'Both [bottom-up and top-down] tracks are independent and thus always simultaneously operate in the comprehension and assimilation of incoming stimuli.' (Narmour, 1992:12).

The listener assumed in *GCTMS* is capable of learning. What is presupposed is not knowledge of musical structures themselves but the ability to make generalisations on given musical data and to learn musical structures. It is conjectured that elementary musical concepts such as a discrete pitch space and pitch scales, metrical templates and of course higher-level musical knowledge concerning melody, harmony, tonality and so on can be induced from musical examples. Such acquired knowledge can then be used to facilitate further processing of new musical pieces (for reasons of methodological convenience, the fundamental musical concepts of pitch scale genres and metrical templates are taken as given in the proposed model - see section 3.4.1).

A general theory of musical structure may attempt to formulate a broad underlying theory of musical understanding which, if further elaborated and refined, may lead to more informed descriptions of individual styles and idioms. In this sense, such a general theory should be idiom-independent and compatible with traditional specialised theories (e.g. for classical tonal music, counterpoint, jazz, ethnic musics, atonal music and so on). At this stage, only elementary insights may be suggested as to how more complex musical knowledge may be accommodated in a general theory such as the *GCTMS*.

According to Meyer (1973), musical analysis tends either to describe the individuality of a piece of music (intra-opus description) within a given musical style (*critical* analysis) or to define the common properties of different pieces (extra-opus description) that allows them to be considered as belonging to the same genre (*style* analysis). Both critical and style analysis require a preliminary analysis of a piece (or set of pieces) of music into 'meaningful' constituents parts (segmentation) which is usually done intuitively by the analyst. The *GCTMS* is a theory that serves to define such an analysis - or rather a set of graded analyses - of an individual musical piece without recourse to expert knowledge relating to a musical style; only the immanent structural properties of the piece itself and a set of general cognitive principles are taken into account. Apart from descriptions of individual pieces, the proposed theory can also give rise to style-analytic information if more pieces are examined for their commonalties and differences (not examined in the current study).

### 3.2 Computational Theory

In trying to understand human capabilities *computational models* are often employed. A computer system is built which embodies a theory describing some aspect of human intelligent behaviour and then this system is used to test the theory against empirical data. For example,

given certain musical input a computer system may be built that gives some 'competent' musical response - such as analysing tonality or rhythm, beat tracking, improvising, composing, and so on.

A computational approach to exploring human (musical) capabilities is important in that it allows the development of theories in an explicit, precise and coherent manner so that they can all or in part be implemented as computer programs. Initially a description is made of the nature of a task and assumptions are formulated as to what the possible underlying mechanisms may be, then a computer system is developed that performs this task and makes predictions about new situations, and finally these predictions are evaluated by comparison with empirical data. This way a better understanding of the problem domain is obtained along with a concrete implementation that may be used creatively in its own right.

A classical computational approach to cognitive processes (according to the traditional AI approach) considers the mind to be a symbol-processing system (*physical symbol system hypothesis*, Newell and Simon, 1976). The classical computational architecture (Newell, 1982; Pylyshyn, 1989; Luger and Stubblefield, 1993) assumes that computers and minds exhibit organisation on at least three levels: the *Knowledge level* (the knowledge that is required for achieving a certain goal or performing a specific task), the *Symbol level* (the formalisms that allow this knowledge to be encoded, e.g. predicate logic) and the *Physical level* (the physical/biological continuum on which the system may actually be realised).

There are two main methodological angles from which the understanding of human abilities may be approached and examined (sometimes referred to as the low-road and the high-road towards understanding cognitive processes - Pylyshyn, 1989:62). The first starts off with a limited well-defined problem examined within a closed universe for which a very detailed model of narrow scope is developed (e.g. Minsky's microworlds, Desain and Honing, 1992; Posner's minimodels, Pylyshyn, 1989). In this approach emphasis is given to discovering the exact algorithm/mechanism by which the task is performed and often psychological experiments are set up to validate the model. Only at a later stage do such models get examined as to how they may relate to or be embodied in broader more general contexts. The other methodological road attempts to describe a much broader problem domain. For this to be achieved, attention is focused on the various general characteristics of a problem, its structure, its different constraints, the functions that may map inputs to outputs, its relations to other domains, and so on. The description of the exact processes and mechanisms is postponed for a later stage.

Marr (1982) has suggested three levels at which cognitive theories may be studied: the *Computational level* (what is to be processed and why and what the function that links inputs to outputs is),[32] the *Algorithmic level* (how exactly certain computations are carried out) and the *Mechanism level* (the description of the physical device on which the process is to be realised). Marr takes the high-road (in developing a computational theory of visual processes) by starting the description of visual processes at the computational level; assumptions about possible algorithms are suggested only after the computational level has been described. "...an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than examining the mechanism (and the hardware) in which it is embodied" (Marr, 1982:27).

A theory at the computational level 'constitutes a formal statement of the various outputs resulting from different inputs' (Eysenck and Keane, 1995:18) and focuses on the form and structure of what needs to be computed for a particular task rather than the precise process by which it is actually computed by the brain (Jackendoff, 1987, ch.4). The formulation of such a computational theory, even though it may not make any direct claims of simulating cognitive processes as these are realised in the human mind, does give insights into the intrinsic requirements of a cognitive task and should always have its results examined with respect to cognitive validity (Van Mechelen et al., 1993a:346; Pylyshyn, 1989:89).

Theories at the computational level tend to focus on the formulation of general principles and functions with which knowledge may be acquired and represented rather than on the construction of intricate *ad hoc* descriptions of a task domain. There are two approaches in constructing a musical representation: the first is the *knowledge engineering* approach whereby the entire representation is 'hand-crafted' by the theorist-programmer based on intuited or explicit (e.g. musical text-books) musical knowledge (e.g. systems by Baroni and Jacoboni, 1978; Cope, 1987; Ebcioglu, 1993) and the second is *empirical induction* whereby a representation is developed by making generalisations on a set of (musical) phenomena based on a set of general fundamental principles (e.g. Conklin and Witten, 1991). As the computational theory proposed here attempts to describe musical phenomena starting with the elementary 'understanding' of a non-experienced listener, it is biased towards the empirical induction approach, that is, by means of a general set of logical and cognitive principles,

[32] He suggests, for instance, that Chomsky's theory of linguistic competence 'is a true computational theory' (Marr, 1982:28).

descriptions and generalisations of increasing complexity may be given to a set of musical entities.

In the context of the present thesis, the term *computational theory* will be taken to refer to a theory that focuses mainly on the computational level (as outlined by Marr). Such a theory suggests possible representations and algorithms that may enable the theory to be implemented as a computer program. A *computational model* will be taken to refer to a specific instantiation (all or in part) of such a theory in a specific situation.

### 3.3 General Principles

The cornerstone of the proposed *GCTMS* is a set of general principles which are assumed to be part of the way in which a human makes sense of the world. These are drawn mainly from the domains of cognitive psychology and are examined more extensively in chapter 4.

The most fundamental principle is the logical principle of *Identity:* two entities are identical if they share exactly the same properties in a given domain of discourse (two entities that do not share the same properties are *different).* If metrics can be devised according to which ordered values can be given to a property of an entity or if each entity has many properties and the number of properties it has in common with another entity is taken into account, then a *degree of difference* (or distance) between the two entities can be established.

As the Identity-Difference principle is fundamentally a logical principle, it may give rise to associations which may not be psychologically pertinent (e.g. the opening and closing tonalities in the sonata-form are the same but listeners do not usually make this association and do not notice if a different tonality is employed at the end - see Cook, 1990). For this reason a set of general cognitive principles (Eysenck and Keane, 1995:233-234) will be introduced that constrain the possible associations given by the Identity-Difference principle. These are:

*Economy:* Because of limitations of the processing and memory capabilities of the human mind the world is divided into more manageable constituent parts through abstraction/reduction, categorisation and hierarchic organisation.

*Informativeness:* An abstraction of the world should accommodate sufficient information to enable a human to achieve desired goals. This principle balances the effects of the economy

principle, which, if unconstrained, will give rise to an extremely small number of over-generalised categories in a way that useful information and detail about the world is lost.

*Naturalness:* This principle relies on the fact that perceptual/cognitive systems are conditioned by 'natural' constraints that suggest some abstractions or categorisations as being more plausible than others (this principle provides an ecological link to the development of theories of various aspects of the world).

These principles in conjunction with the Identity-Difference principle give rise to the notions of similarity and categorisation on which the *GCTMS* is based. *Similarity* judgements can be made when at least three entities are compared. Similarity is inversely related to the degree of difference. For a given context of entities a threshold is set for the degree of difference below which entities are judged to be similar and above which dissimilar. It should be stressed that similarity may be applied not only to internal properties of an entity but to relations with other entities as well. Similarity is also inextricably bound to a notion of *categorisation*. Similar entities are grouped together in categories. If categorisation descriptions change so do similarity judgements and the converse. Additionally, both similarity and categorisation are linked to the descriptions of entities in terms of diagnostic properties (i.e. properties become more or less prominent according to emerging categorisations and similarity measurements).

Finally, the *exposure priming effect* relates to the observation that the salience of objects and relations between objects in memory is roughly proportional to the exposure to the stimuli (frequency of occurrence, recency and exposure length). That is, if a stimulus is more recent, occupies larger space/time in a sensory field or is repeated more often than other stimuli, then it is highlighted into perception (this applies basically for implicit memory, i.e. unconscious automatic data-driven memory processes).

The principles of Identity and Economy may be combined to create the simplest forms of structure, namely, *regular* structures. Such structures consist of a single unit or pattern which is simply repeated throughout a given space. In music, for instance, such units are the semitone (or other pitch interval unit) or scale patterns that organise pitch space, and metric beat time-span units or simple patterns of beats that organise time. Such regular structures are very useful in providing systems of reference against which more complex structures may be constructed and perceived.

### 3.4 Overview of the *General Computational Theory of Musical Structure*

In this section the *GCTMS* will be outlined and each constituent part will be briefly described - figure 3.1 presents an overview of the theory.

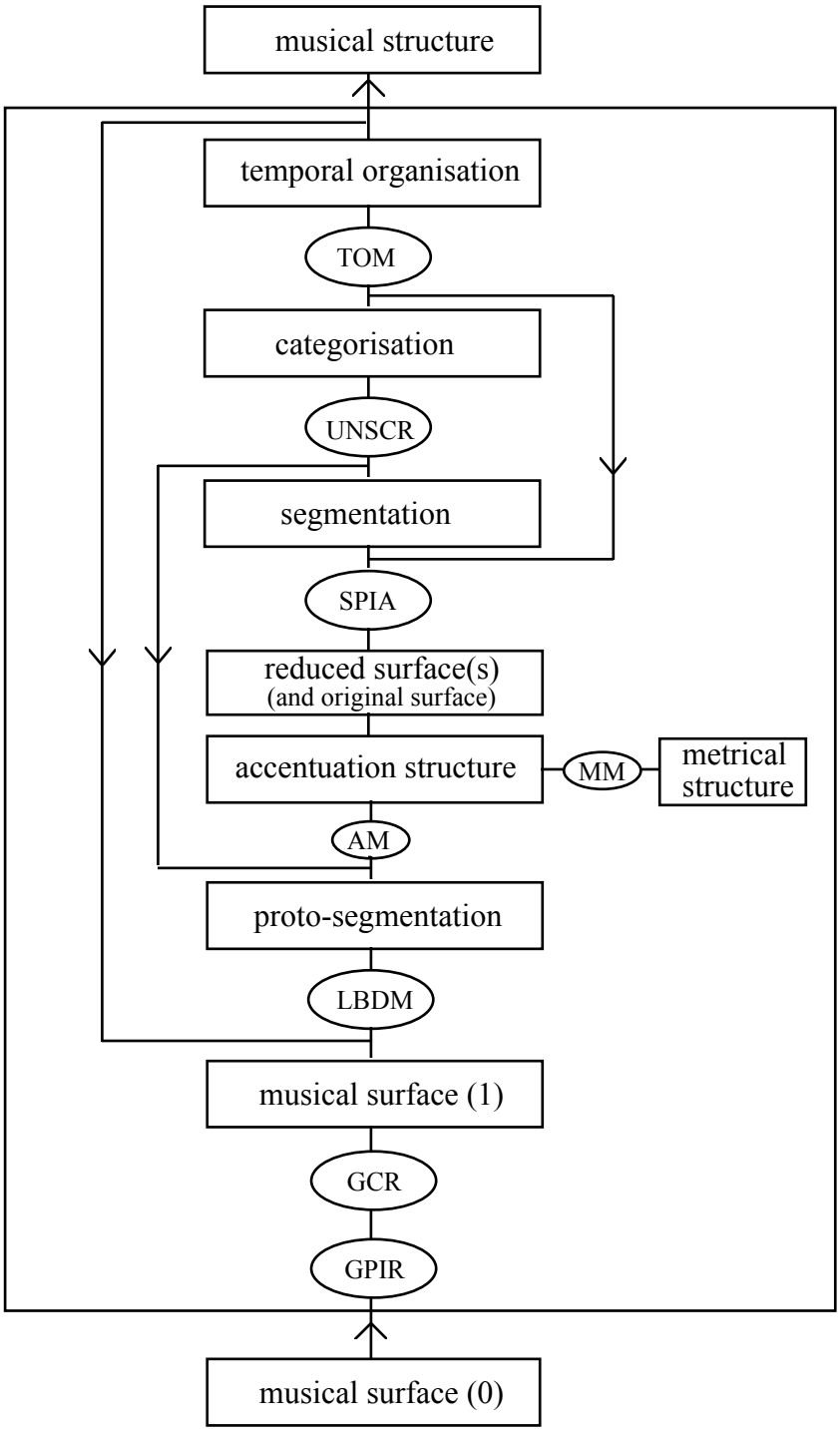
### 3.4.1 *GCTMS:* Musical Input

A musical surface is assumed to be 'the lowest level of representation that has musical significance' (Jackendoff, 1987:219). It has been supported by many studies (see section 5.2) that the acoustic continuum is perceived categorically as discrete quantised musical primitives. Symbols may be used to denote such musical primitives (even though many oral musical traditions do not have explicit symbolic notation systems). For instance, for the Western 12-tone equal-temperament system notes on a staff (not a full score) may be considered to be an adequate representation of this lowest level of representation - musical notes are multi-faceted entities characterised by different independent attributes such as pitch, temporal onsets, durations, loudness, timbre and so on and may thus be represented by an array of symbols. This elementary discrete quantised representation of a musical piece will be referred to as the *musical surface (0)* - a more extended discussion of the musical surface appears in section 5.2.

Features of music that are considered to be primarily expressive (not structural) - such as, in the Western tradition, continuous timing (mainly expressive timing), pitch inflections, expressive timbre variations and so on - are not necessary prerequisites in the present theory even though they play an important role in highlighting underlying musical structural interpretations (see Clarke, 1987). In this respect, articulation features indicated on scores as slurs, breath marks and so on, will be disregarded or simply considered as guides to particular interpretations of a musical surface among many other possibilities. Obviously, the distinction between expressive and structural musical features is specific to a given musical idiom and may vary from idiom to idiom.

For convenience, the following two musical concepts that are idiom-dependent are given as input to the present theory: a) pitch scale genres (e.g. diatonic or pentatonic or whole-tone scale genres) and b) metrical templates (e.g. 3/4, 4/4, 5/8, 7/8 metrical structures). These are composed of *regular* structures at various levels. Such basic templates 'are assumed to be primarily *learned* from environmental sound patterns ... simply by passive exposure' (Parncutt, 1994:149) and it is suggested that computational models can be built that perform such learning tasks by making generalisations on sets of given musical examples. Template-

inducing computational models are not described in the present thesis; taking idiom-specific pitch and metrical templates as given input in the current study is mainly a practical decision taken to allow earlier engagement with the description of higher-level analytic processes.




TOM: Temporal Organisation Model
SPIA: String Pattern-Induction Alg. & Sel. Function
AM: Accentuation Model (event salience)
GCR: General Chord Representation
UNSCR: UNSCRAMBLE Algorithm
MM: Metrical Matching
LBDM: Local Boundary Detection Model
GPIR: General Pitch Interval Representation

Figure 3.1 Overview of the *General Computational Theory of Musical Structure*

### 3.4.2 *GCTMS:* Output analysis

The *GCTMS* is a non-exclusive theory: it produces a set of graded analyses - no analysis is totally disregarded. It is assumed that a musical structure may be interpreted by listeners in many different ways none of which should be excluded as 'false'. However some interpretations may be judged as 'better' by more experienced listeners; one analysis may be selected as the 'best' description for a given musical surface or a number of analyses may co-exist in the final description (denoting ambiguity or transitionality).

A musical surface is segmented into component parts at various levels. These segments may be partially overlapping and may be of different sizes; however, regular non-overlapping partitionings of the surface may emerge in some cases. There are cognitive reasons (relating to the general principles outlined in section 3.3) that make some segmentations more likely than others, whilst there are other cases where such preferred descriptions are not obvious and musical passages may simply be considered ambiguous (e.g. co-existence of many different analyses resulting in a complexity that does not allow the selection or domination of one over the others). The proposed theory attempts to highlight segmentations that are preferred when such segmentations exist.

The theory not only suggests possible segmentations of a surface but also enables the categorisation of segments into classes and the description of relations between them. These categories of musical materials (e.g. motives, themes etc.) may give rise to a hierarchical description of the surface (hierarchies are not necessarily tree-like structures). This way, a surface receives a structured description at various levels that allows it to become more easily accessible to a human listener.

The theory may be evaluated in a number of ways:
a) the resulting analyses may be judged by an *expert musical analyst* as being 'acceptable' and 'plausible' (or may be compared with published analyses).
b) psychological experiments may be set up where the predictions of the theory may be compared to the descriptions suggested by different types of listeners.
c) the analytic data obtained by the theory may be used in a reverse process whereby new pieces may be composed that are of the same 'style' as the original.

At the present stage of this research project only the first (actually weakest) method will be adopted. Psychological experiments and analytic-compositional systems are further possibilities that may produce more concrete evidence as to the effectiveness and validity of the proposed theory. These options are left open for further research.

### 3.4.3 *GCTMS:* Representations and Models

The main body of the proposed theory consists of one pre-module and two main modules. The pre-module allows the derivation of the musical surface (1) - i.e. a representation consisting of musical intervals or compound musical objects such as chords (see section 5.2) - from the musical surface (0). The first module is mainly a *microstructural module* that allows the comparison of contiguous events or intervals of the musical surface. This module results in a proto-segmentation of the musical surface, highlights local salient events, suggests a possible metrical structure and allows a preliminary reduction of the surface. The second module is a more central *macrostructural module* that allows the comparison of an event or pattern of events with all the other events and patterns in the musical surface and/or reduced versions of it. This module complements the microstructural module in producing an integrated segmentation of the musical surface and allows the categorisation of musical events and patterns of events into categories that share a number of properties (the temporal organisation of musical categories has not as yet been described).

The analytic engine of the computational model (which can be seen as a specific instantiation of the *GCTMS*) is based on the individual component models outlined below:

a) the *General Pitch Interval Representation (GPIR).* The initial absolute pitch information of the musical surface (0) is converted to a more sophisticated pitch and pitch-interval representation, that reflects hierarchic qualities of the tones of a given pitch scale over the available background pitch space, using the *GPIR* and a relating transcription algorithm (chapter 5). The resulting *GPIR* pitch interval profiles constitute part of the musical surface (1).

b) the *Local Boundary Detection Model (LBDM).* This is a model that detects points of maximum change/discontinuity in a musical surface which are most likely to be perceived as local boundaries at various hierarchic levels. This produces an initial tentative segmentation of a musical surface (chapter 6).

c) the *String Pattern-Induction Algorithm (SPIA)* and *Selection Function*. This is a pattern-matching algorithm that starts with the smallest patterns of a sequence (e.g. a sequence of musical intervals) and stops when it reaches maximal patterns. From these patterns a selection function selects the most cognitively pertinent ones. This algorithm complements (b) in decomposing a musical surface into 'meaningful' components by revealing significant parallel musical passages, e.g. motives, themes etc. Pattern-matching is done for the musical surface and/or for reduced versions of it (chapter 7).

d) a module that reveals the *Accentuation and Metrical Structure* of the piece. The accentuation structure is automatically inferred from the grouping structure defined by (b) and (c) and then a metrical template is matched onto the accentuation structure. This module can provide cues for the reduction of the musical surface by the elimination of less accented or metrically weaker events (chapter 6 & 9).

e) the *Unscramble* algorithm. This is an unsupervised symbolic machine learning algorithm that organises the musical segments discovered by (b), (c) and (d) into cognitively pertinent categories/paradigms in the fashion of paradigmatic analysis (chapter 8).

The *GCTMS* is not a linear theory whereby analysis is pursued uni-directionally in a bottom-up or top-bottom fashion. Neither is it a theory of totally independent agents freely interacting with each other. It is rather a theory where the different components at the various levels interact with each other (lower level analytic results facilitate the employment of higher-level procedures and, in turn, higher-level results inform and disambiguate lower-level analytic outcomes) but there is a loose overall directionality from lower level descriptions towards higher level ones (for instance, it is not computationally practical nor is it cognitively plausible to start with the categorisation model before some preliminary segmentation has been obtained). In figure 3.1 the analysis proceeds from the bottom of the diagram upwards; arrows indicate the feedback loops of the theory. The exact description of how the above individual components are combined and interact with each other to produce a final analysis is given in chapter 9 - along with four examples of melodic analyses obtained by the application of the overall model.

**Conclusions**

This chapter started with a discussion on the main aspects of musical structure that the *General Computational Theory of Musical Structure* attempts to describe, the computational

level on which this theory is formulated and the general logical and cognitive principles on which it is based. Then, an overview of the *GCTMS* was presented in terms of the mechanisms and models that allow the derivation of a structural description from a musical surface (0). In the next chapter, a more detailed account of the logical and cognitive foundations of *GCTMS* is given and from chapter 5 onwards the full description of each individual component, as well as the interaction between the various components, will be presented.

# Chapter 4

# Logical and Cognitive Foundations

**Introduction**

In this chapter the basic logical and cognitive elements on which the *General Computational Theory of Musical Structure* is based will be discussed. Firstly the principles of economy, informativeness and naturalness - complemented by the exposure effect - will be briefly presented as precursors to the description of similarity and categorisation processes. Then the logical principle of identity will be presented followed by a more detailed examination of the notions of similarity and categorisation. It will be suggested that similarity and categorisation are inextricably bound together and cannot be described independently of each other. This claim will be supported by presenting general and music-specific examples, and will form the basis for developing a formal description of these notions and a novel computational model of categorisation (chapter 8).

**4.1 Basic Principles**

Categorisation is paramount in allowing us to organise the infinitely complex world into concise meaningful constituent parts. Our ability to perceive something as an entity, e.g. object/event/action, is directly linked to our ability to form and use categories. This reduction of information into manageable components is necessary for reasons of storing and processing efficiency in the human mind. It is suggested (e.g. Rosch, 1978; Eysenck and Keane, 1995) that this process of categorisation is guided by the following general cognitive principles.

• *economy principle:* Through abstraction and categorisation it is possible to reduce our experience into manageable constituent parts and further organise these into parsimonious

hierarchical structures. This cognitive principle of economy may be paralleled to the methodological principle of ontological economy referred to as Occam's razor: *Entities are not to be multiplied beyond necessity* - i.e. an explanation should not postulate more kinds of things than are absolutely necessary (Read, 1994).

• *informativeness principle:* An abstraction of the world should accommodate sufficient information to enable a human to achieve desired goals. This principle constrains the economy principle so that over-generalisation may be avoided. If everything is reduced to a handful of categories then we would have an extremely economical description of the world but a lot of useful information would be lost (one wouldn't be able to distinguish, for instance, between a fish and a bird if only the category 'animal' was available).

These two principles are magnificently balanced by the human mind so as to produce useful multi-level taxonomies. Of course, the structure of the world itself - as well as the perceiving agent - plays an important role in the formation of such economical and informative descriptions; this ecological link to the world may be referred to as the *'naturalness' principle*. In line with ecological accounts of perception (see Gibson, 1966) it is herein asserted that this principle may be applied not only in relation to the natural environment but to more abstract cultural systems as well - there is no sharp distinction between nature and culture - see (Clarke, 1997; Clarke and Dibben, 1997) for an ecological account of musical perception.

Finally, the construction of a certain kind of taxonomic organisation is influenced by the *exposure priming effect*. This accounts primarily for the effect that repetition - i.e. the frequency with which entities are presented to a subject - plays in the formation of a concept (recency and exposure duration also contribute to this effect). Frequency of occurrence has been shown to play an important role in the formation of base-levels of categorisation, i.e. intermediate most useful level of a taxonomy, and on category gradedness, i.e. typicality of category members (Hintzman, 1976; Hasher et al., 1979; Barsalou, 1985; Barsalou et al., 1986; for frequency effects in music see Jeffries, 1974).

In the next few sections some - usually implicit - commonly used descriptions of identity, similarity and categorisation will be outlined and discussed. Then, in section 4.5 a definition and description of similarity and categorisation will be given whereby the two are inextricably bound together. Finally, in section 4.6 some psychological experiments will be examined and re-interpreted in a way that is conformant with these definitions.

**4.2 Identity**

Before attempting to describe the notion of similarity it is important to discuss briefly the principle of identity and to try to clarify its usage within this text.

Without getting into a deep ontological discussion, an *entity* is herein taken to refer to a complete and distinct thing - concrete or abstract - such as an object, an event, a structure, a function, a goal, and so on (e.g. a pencil, a robin, a song, an emotion, an action such as running, sleeping etc.). A *property* is any predicate that may be used to describe an entity.

Is it possible for two different entities (that have different spatio-temporal properties) to be identical? For example, is it possible that two drops of water or two middle-C notes played on the same instrument may be identical? Leibniz's response to such a question would be that 'there are no two individuals indiscernible from one another' (Fourth Paper to Clarke, Sec. 4, quoted in Stroll, 1967:122) or 'there are not in nature two individuals indiscernible from one another' (G. VII. 393 (D. 258) in Extracts from Leibniz in Russell, 1949:219). This principle is referred to as the *Principle of the Identity of Indiscernibles*. Stroll (1967) states that "Leibniz's language suggests that he considered this principle to be an empirical law; that if we were to find two items (say two drops of water) apparently possessing exactly the same set of internal features, further investigation (by means of a microscope, for instance) would show that they differed from one another." (p.122). He then continues: "But reflection upon [Leibniz's] use of the expressions 'intrinsic quality' and 'internal difference' suggests that he covertly employed the principle as if it were a logical truth, to which no empirical finding would be a counter-instance." (p.122). Many philosophers have rejected this principle when presented as logically necessary (Black, 1952) but it is accepted when seen as an empirical law.

This principle is connected, according to Russell, to Leibniz's implied assertion 'that every substance has an infinite number of predicates' (Russell, 1949:60). '...individuality involves infinity, and only he who is capable of understanding it [infinity] can have knowledge of the principle of individuation of such or such a thing' (G.V. 268 (N.E. 309) in Extracts from Leibniz in Russell, 1949).

According to Russell's definition of identity two entities x and y are identical if and only if the same properties (predicates) are satisfied by both (Russell and Whitehead, Principia Mathematica, vol. i, def. 13.01). The identity relation is an equivalence relation, i.e. it is reflexive, symmetric and transitive. But is this definition of any use if two entities have

infinitely many properties? How is it that one says that two different drops of water or two middle-C notes are identical?

The key to answering these questions is that two entities are judged identical only when a finite number of properties *that are considered salient* for a given domain of discourse are demarcated. When we say that two objects are identical we mean that all the properties (predicates) that describe the two objects - taken from a set of predefined properties that are considered to be pertinent in a given context - have the same values. Quine (1950) emphasises the value of a domain of discourse: 'In general we might propound this maxim of the *identification of indiscernibles:* Objects indistinguishable from one another within the terms of a given discourse should be construed as identical for that discourse.' He continues that this maxim 'is relative to a discourse, and hence vague in so far as the cleavage between discourses is vague. It applies best when the discourse is neatly closed, like the propositional calculus; but discourse generally departmentalizes itself to some degree, and this degree will tend to determine where and to what degree it may prove convenient to invoke the maxim of identification of indiscernibles.' (p.626).

The most crucial factor in establishing 'meaningful' identities is selecting the set of properties that are pertinent in describing a set of entities in a given situation. This set of properties is not absolute but depends on the task at hand. For instance, two tunes may be most commonly considered identical in the Western tradition if they both are composed of the same sequence of 12-tone equal tempered pitch intervals and quantised integer duration ratios, i.e. same musical surface. If, on the other hand, in a different domain their expressive or spectrographic properties are considered to be most pertinent then they may be judged as being non-identical.

**4.3 Similarity**

Similarity is a difficult and obscure notion. How does it relate to identity? What are the conditions and limits under which two entities may be considered similar?

For a given set of pertinent properties and following from Russell's definition of identity, similarity is very often defined as *partial identity*, i.e. two entities are similar if they have *some* properties (predicates) the same but not necessarily all. Pairs of entities may be compared and one pair may be judged as being more similar than another if its members share more common properties than the members of the other pair.

Similarity between two entities may be calculated by simply counting the number of matches between their properties. Alternatively, similarity may be defined as a function of the differences between all the pairs of properties these objects posses. For example, according to the traditional multidimensional scaling model (Shepard, 1962a,b) similarity between objects x and y is a monotonic decreasing function *f* of interpoint distance:

$$s(x,y)=f(d(x,y))$$

where s(x,y) is a similarity rating between x and y, d(x,y) is the distance between the two points of the objects' attribute vectors in a multidimensional attribute space - for a brief summary of commonly used metrics see (Murtagh, 1993:228-230).

If all properties receive equal weight for the metric d(x,y) then this definition of similarity is equivalent (for binary features and Hamming distances) to the former definition (i.e. partial matching of properties). If, on the other hand, properties are given different weights reflecting the intuition that not all properties are equally important for a given object then there is a significant departure from the former traditional definition of similarity. For instance, the members of a pair of objects that have in common only one important property may be judged as being more similar than the members of another pair that share two or more less salient properties.

The similarity definitions given above imply that the similarity relation is reflexive, symmetric but not transitive. There exist though other models that allow asymmetric definitions of similarity. For example, Tversky (1977) proposed that similarity between two entities may be defined as a function of their common properties minus the properties that are distinctive to either of them:

$$s(x,y) = \theta \cdot f(X \cap Y) - \alpha \cdot f(X-Y) - \beta \cdot f(Y-X)$$

where s(x,y) is the similarity between two objects, X and Y are the feature sets of x and y respectively and $\theta,\alpha,\beta$ are parameters that are used to reflect prominence of common and distinctive features. Tversky's model of similarity has been proved to be very useful in describing (empirically) observed similarities but is rather impractical if used to predict similarities between entities as it requires a very elaborate representation of each individual entity. That is, the model requires that the individual sets of all the features that are important for the description of each object be precisely defined (rather than using only one general set

of features that accounts for all the objects) and/or all three parameters $\theta,\alpha,\beta$ be given in advance for each ordered pair of objects. Tversky's model fails to address the question of how people determine which properties are relevant for a similarity comparison (see Barsalou, 1992:282-284).

Alternatively, Krumhansl (1978) proposes an extension of the multidimensional similarity definition, namely the distance-density model, that accounts for asymmetric judgements and contextual aspects of similarity. The distance-density model is based on the assumption that 'two points in a relatively dense region of a stimulus space would have a smaller similarity measure than two points of equal interpoint distance but located in a less dense region of the space' (Krumhansl, 1978:446). According to this model, the distance d(x,y) in the similarity function of the multidimensional scaling model - $s(x,y)=f(d(x,y))$ - is replaced by a modified distance function d'(x,y): $d'(x,y)=d(x,y)+\alpha\cdot\delta(x)+\beta\cdot\delta(y)$ where d(x,y) is the interpoint distance, $\delta(x)$ and $\delta(y)$ are measures of spatial density in the neighbourhoods of x and y, and $\alpha$ and $\beta$ are constants that reflect the relative weight given to the densities $\delta(x)$ and $\delta(y)$. For instance, 'if $\alpha<\beta$, then $s(x,y)>s(y,x)$ if and only if $\delta(x)<\delta(y)$, that is, in directional similarity tasks, asymmetries would be expected to be associated with differences in the densities in the regions surrounding the two points in the geometric configuration.' (Krumhansl, 1978:453). This definition of similarity augments the traditional definition by incorporating a density factor that relies on local context.

A common characteristic of all the above definitions is that none of them incorporates a notion of categorisation. These definitions of similarity (usually the symmetric ones) are commonly used as prerequisites for other categorisation models that predict possible clusterings of objects but they are not explicitly linked to a notion of categorisation.

**4.4 Categorisation**

In the course of this text the word *category* will be taken to refer to a set of entities which are grouped together on the basis of some criteria. The conditions for classification are commonly referred to as the *intension* of a concept and the set of entities that are members of a category the *extension* of the concept. The term *concept* 'refers to the idea or notion by which an intelligence is able to understand some aspect of the world' (Hampton et al., 1993:13).

According to the classical *monothetic* definition a category is constituted of all the entities that posses a set of properties or satisfy a set of conditions (see Sutcliffe, 1993). Most commonly these conditions are taken to be singly necessary and jointly sufficient.

Classical categories do not rely on similarity measurements but once such a category is formed all its members can be considered similar. It should also be noted that the conditions relate to sets of properties possessed or not possessed by objects rather than weighted combinations of properties.

A different approach to formalising the notion of categories has emerged following Wittgenstein's approach to the notion of 'family' and 'family resemblance' (Wittgenstein, 1953). According to the polythetic view, a category consists of individuals that have a large number of properties from a given set P and each property is possessed by a large number of members but no property is possessed by all the members of the category (Beckner, 1959:21). The problem with this definition is to determine when a 'large number' is large enough, i.e. to define a limit above which entities share enough properties so as to be considered members of a category.

The polythetic definition of categories underlies prototype models (Rosch, 1975; see Hampton, 1993 for an overview) and exemplar models of categorisation (Estes, 1994). According to the prototype view members of a category are determined by their similarity to the category's prototype and 'a prototype concept is constituted by a *set of attributes with associated values* (where a particular attribute-value pair corresponds to a property), each with a particular weight corresponding to its 'definingness' or contribution to the concept's definition." (Hampton, 1993:73). Membership and typicality of an instance is judged in relation to a similarity measurement of the individual to the category's prototype (i.e. the weighted attribute-value set) - or exemplar for exemplar models. There exists a criterion on the similarity scale over which individuals are considered to be members of the category and their typicality is proportional to the similarity rating (i.e. the higher the rating for an instance the higher its typicality).

Prototype models account for many phenomena observed in the way humans make categorisations in everyday situations, e.g. flexibility of category boundaries, gradedness and typicality of members, ambiguity of membership etc. (e.g. is a tomato a fruit or a vegetable?).

The prototype of a concept and the similarity criterion can be determined by direct experimentation and then used for further predictions. If one, though, wants to derive the prototype and the criterion from a set of entities and a general set of properties so that categories may be formed, then the prototype definition of a category reveals its weaknesses. How can one discover a relevant similarity threshold of objects to the prototype to determine their membership if the prototype is not known? How can the prototype (i.e. a weighted set of characteristic attribute-values) be determined? If the extension of a category is given then a prototype can be defined (by finding the most characteristic properties that are possessed by most members), but that means that one knows in advance the category members. But how could the category be defined without reference to the prototype since it is defined in terms of the prototype? Sutcliffe remarks that 'there must first be a family before one can observe any family resemblances, and thus one cannot define a family by reference to family resemblances!' (Sutcliffe, 1993:46).

The prototype view on categorisation relies to some extent on either some form of independent bottom-up, data driven, clustering-like analysis (see Mechelen et al, 1993, part II) or on top-down theory-based approaches (e.g. Murphy, 1993; Murphy and Medin, 1985) or a mixture of the two.

Both of the above descriptions of categories can accommodate conjunctive as well as disjunctive intensional descriptions (especially for monothetic categories, Sutcliffe (1993:59) argues that disjunctive concepts have a sound logical basis). It is asserted herein that disjunctive concepts are hard to work with when dealing with unsupervised category formation tasks. The reason for this assertion is that the space of all possible conjunctive descriptions (for a given set of entities) through which a search has to be pursued is augmented explosively if disjunctive concepts are considered as well. If instances, though, of a category are known in advance - as in supervised learning - then disjunctive descriptions may be convenient (for example, if 'couples' are represented on an instance space by the 'colour' of each partner, then categories such as 'mixed couples' and 'same-colour couples' are not possible unless either disjunctive concepts are accommodated or the initial representation is altered).

The debate between the 'classical' and the 'modern' view is heated. Hampton argues that 'classical monothetic concepts can be treated as special cases of prototype models in which the membership criterion has been set very high on the similarity scale, so that the criterial level of similarity cannot be achieved without the core properties.' (Hampton, 1993:76).

Contrastingly, Sutcliffe argues that "the 'modern view' developed by Rosch on the basis of Wittgenstein's and Beckner's notion of polythetic class, is incoherent and unworkable." (Sutcliffe, 1993:62).

In this study it is suggested that the distinction between the monothetic and polythetic views on categorisation is not as sharp as many would argue (e.g. Lakoff, 1987). For instance, if an exact threshold is set for a polythetic category then a sharp boundary is defined (some form of boundary is necessary in any case: it doesn't make much sense to say, for instance, that a chair is a very atypical member of the category 'bird' - it simply isn't a bird). If overlapping of categories is allowed then ambiguity and gradedness is introduced (for both monothetic categories and polythetic categories with sharp boundaries), i.e. the more categories an entity belongs to, the more ambiguous it is and the less typical a member of a category it is. If the two definitions of category are dissociated from metaphysical claims and are seen simply as formal descriptions of the notion of category then there can be only pragmatic criteria as to their usefulness and efficiency.

It is clear from the above discussion that all the members of a category are necessarily pairwise similar as they necessarily share some common properties (they share at least the property of belonging in the same category!), but the converse is not necessarily true, i.e. similar entities are not necessarily members of the same category. The notions of similarity and category can be brought into a close relation if a threshold is introduced in the definition of similarity (see next section).

**4.5 Similarity and categorisation bound together**

A commonly encountered hypothesis on which many categorisation models are grounded is that categorisation is strongly associated with the notion of similarity, i.e. similar entities tend to be grouped together into categories.

However, there are different views on the relation between similarity and categorisation (Goldstone et al., 1994; Medin et al., 1993). On the one hand, similarity is considered to be too flexible and unwieldy to form a basis for categorisation, i.e. any two entities may be viewed as being similar in some respect (e.g. a car and a canary are similar in that both weigh less than 10 tons, but these objects are not normally considered to be members of the same category!). On the other hand, similarity is regarded to be too narrow and restricting to

account for the variety of human categories (e.g. a whale is more similar to other fish but we still consider it to be a mammal). Goodman (1972) doesn't hesitate to call similarity 'a pretender, an impostor, a quack' (p.437). Rips (1989) claims that "there are factors that affect categorisation but not similarity and other factors that affect similarity but not categorisation. ...there is a 'double dissociation' between categorisation and similarity, proving that one cannot be reduced to the other" (p.23).

The above debate is directly linked to a further issue; that is how entities and their properties are represented. If objects are described in terms of mainly perceptual (e.g. visual or auditory) properties, then, obviously similarity is insufficient for many categorisation tasks, whereas, if any sort of properties - perceptual or abstract or relational - are considered then similarity becomes too flexible.

It seems that the notions of categorisation, similarity and the representation of entities/properties are strongly inter-related. It is not simply the case that one starts with an accurate description of entities and properties, then finds pairwise similarities between them and, finally, groups the most similar ones together into categories (figure.4.1a). It seems more plausible that as humans organise their knowledge of the world, they alter their representations of entities concurrently with emerging categorisations and similarity judgements (figure 4.1b).

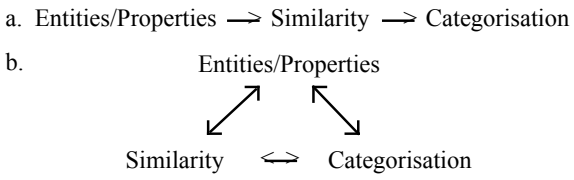



Figure 4.1 Relations between entities/properties, similarity and categorisation

One of the main assumptions made in this study is that similarity *always* depends on context (i.e. it is contextually defined), and when similarity seems to be relatively stable, this is so simply because the context - e.g. the structure of the natural world or a specific cultural system - tends to be quite stable. Of course, there are some general perceptual constraints as to what is perceptible in the first place, but from there on different properties of entities become more prominent in a given context for a specific categorisation task or for a similarity judgement. Tversky (1977) has highlighted the importance of context in similarity judgements and has shown how properties of objects become *diagnostic* within a specific context; he

treats, though, these contextual effects on similarity as specific cases/exceptions rather than the norm (his definition of similarity is independent of categorisation).

As a first general example consider figure 4.2. Which of objects *b, c & d* is most similar to object *a*? One might - cautiously - select one of these objects or refuse to answer the question altogether. If, though, these objects are placed in a context such as a barber shop or an office or a surgical operating room, then it becomes apparent which objects are more similar and are actually categorised together, and which properties of the objects are more prominent and diagnostic in that specific context - for instance, within the context of a barber's shop objects *a & c* are more similar and they tend to be categorised together because they share barber-related properties (such as 'hair-cutting').

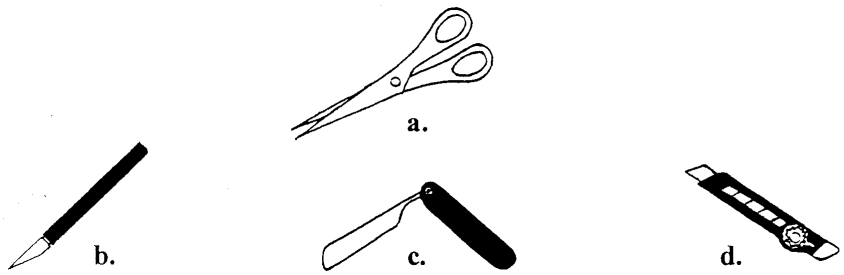



Figure 4.2

A second example from the musical domain will be presented below that highlights the contextual nature of similarity and categorisation. A musical work may be considered as a local context within which things like motives, themes, harmonic progression groups etc. emerge. Trying to discover the similarity of two isolated musical passages will usually produce dubious or relatively uninteresting results. Consider, for instance, the musical passages in figure 4.3. In which of the two pairs are the two passages more similar? Some might select the first pair, others the second pair, and still another group might refuse to make a judgement. It is suggested that perhaps this similarity experiment is simply ill-designed in the first place, and perhaps subjects of the third group are right in refusing to make a judgement. The problem seems to be that these excerpts are taken out of their context. As it happens, the first two passages are very dissimilar - actually contrasting - within the homogeneous minimal context of S. Reich's *Electric Counterpoint,* whereas the second two are very similar within the very diverse context of I.Xenakis' *Keren.* Context seems to be paramount in our establishing similarities and categories between musical passages and it is

asserted that it is not possible to find an absolute criterion for defining what things are similar *in general*.

The psychological theory of musical form proposed by I. Deliège - see overviews in (Deliège, 1997a, 1997b) - examines empirically issues of property prominence *(cue abstraction)*, musical similarity and prototypical description of categories *(imprint formation)* in musical listening. Deliège's work seems to be in line with the description of entities/properties, similarity and categorisation in the current thesis; however, the model presented in this chapter and chapter 8 establishes direct formal links between these notions in a way not encountered explicitly in other cognitive accounts of musical understanding.

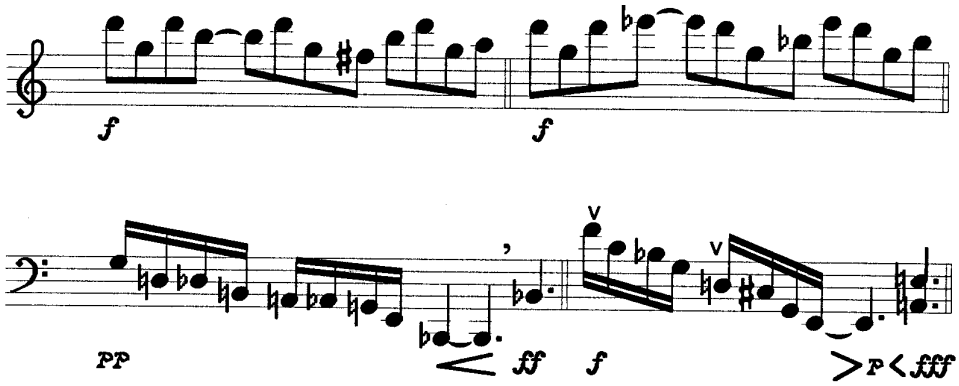



Figure 4.3 In which of the two pairs are the two passages more similar?

In the light of the above discussion, formal definitions of similarity and category will be given wherein the two notions are inter-dependent, i.e. changes in similarity result in category changes, and the converse - a more detailed description will be presented in chapter 8.

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:

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

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.

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:

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

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

In line with the above descriptions, the *Unscramble* algorithm (unsupervised symbolic machine learning algorithm) will be presented in chapter 8 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 an acceptable description is generated.

Finally, some psychological experiments that seem to suggest an incongruity between the notions of similarity and categorisation will be re-visited in the next section and alternative interpretations that are compatible with the proposed formal definitions will be given.

#### 4.6 Re-examining some psychological experiments

We will now examine how the notions of identity, similarity and categories have been applied in three psychological studies and will show that these experiments need not be considered incompatible with the proposed working definitions.

1. Krumhansl (1990:143-152) suggests that two instances of the same musical tones are perceived as being more identical if they are more stable in a given tonal context. 'The first

principle, *contextual identity*, governs the degree to which two instances of a musical tone are perceived as identical.... For two instances of the same tone, a, the psychological distance is denoted d(a,a). The principle says that this distance is less for more stable tones. Contextual identity: d(a,a) decreases as the stability of 'a' increases' (Krumhansl, 1990:143).

One experiment reported by Krumhansl that supports the above principle involves listeners comparing and measuring the degree of sameness of the two instances of the same tone preceding and following the same tonal context. For example, a middle G is played before and after a C major context and a middle F# before and after the same C major context. Listeners gave a higher rating of sameness to the more stable diatonic tone G than to the non-diatonic tone F#. Although both instances of the two tones have the same pitch and both occur in the same tonal context (i.e. they are identical) they are judged to be to a different degree identical.

According to the definition of identity given previously, two entities are identical if they share *all* the same predicates in a given domain of discourse. This means that two entities in a given context are either identical or not - there can be no degree of identity. In the light of this definition the use of the term identity in relation to the above experiment is questioned. It is herein suggested that the two instances of a tone presented to listeners in the above experiment are in the first place *non*-identical as they occur in different temporal positions in relation to the given C major tonal context. Actually the first occurrence of the tones does not have any local context except in retrospect: it may be hypothesised that the listener makes a tentative assumption according to background knowledge that the first standard tone is a tonic or another diatonic tone which may be overturned by the subsequent context - contrastingly, the last tone is clearly placed in relation to the preceding context. Perhaps one way to have the 'same' context for both instances of the tone is to present to listeners the sequence: Cmajor context - tone X - Cmajor context - tone X (possibly looped indefinitely).

Perhaps this issue could be resolved if the word identity appeared in inverted commas, as the principle of contextual 'identity'.

2. Carey (1985) presented to subjects a set of living things plus one mechanical monkey. Then subjects were asked to select an item from this set that was most similar to a human; both children and adults chose the mechanical toy monkey. However, when they were asked about the biological properties of the mechanical monkey all subjects denied that the mechanical monkey had any at all (e.g. it doesn't have a heart, it doesn't sleep etc.). So, although the mechanical monkey was judged to be most similar to humans the two were not considered to

be members of the same category. Murphy (1993) refers to this experiment as an 'impressive demonstration' that it is *not* generally the case that 'the more similar an object is to a conceptual representation, the more likely it is that it will be identified as an exemplar of that concept' (Murphy, 1993:185).

This experiment - if adequately interpreted - seems to be in line with the claims of this chapter on the strong link between similarity and categorisation. One interpretation of this experiment is that perceptual *visual* similarity (appearance) is not generally sufficient (or even relevant) for categorisation. But it doesn't seem to support Murphy's claim. Visual similarity is not the same thing as more general conceptual similarity. Another way to view this experiment is that subjects are not making judgements of perceptual similarity, but are simply using the mechanical toy monkey as a signifier/sign for a real monkey - since all the other objects are living things - and are actually comparing a real monkey to a real human. In this case, subjects may be making use of a more general notion of similarity and actually the experiment may be taken to be in support of the claim that similarity is strongly bound to a notion of categorisation. (Perhaps, if the order of the experimental stages was reversed, i.e. first the discussion on biological properties and then similarity ratings, then the actual mechanical monkey might have not been judged to be similar to the human, as subjects would probably be using a broader notion of similarity.)

3. Barsalou's (1983) 'ad hoc' categories are often used as examples of categories whose members are dissimilar. Examples of such categories are: 'things to take on a camping trip', 'foods to eat on a diet', 'things to take from a burning house' and so on. It is suggested that such categories rely on transient goals rather than similarity between the objects. But such goals may be considered as properties of the objects in a given domain of discourse in which case the objects are similar as to these goal-oriented properties. Murphy states that "Children, jewellery, family photographs, and pets are quite different in most respects, but they are similar in that they are portable, people value them highly and they are irreplaceable. Thus, they are all excellent candidates for 'things to carry out of a burning house'." (Murphy, 1993:186)

**Conclusion**

In this chapter, the logical and cognitive foundations of the proposed computational theory were presented. Special emphasis was given to the description of the notions of similarity and

categorisation, and a working formal definition was given according to which similarity is contextually defined and is inextricably bound to a notion of corresponding categories. Finally, some experiments from psychological research that seem to contradict the proposed definitions were critically re-examined and it was shown that they are not incompatible with the proposed description of similarity and categorisation.

# Chapter 5

# Representation of the Musical Surface

**Introduction**

In this chapter, firstly, a general abstract representation adequate for representing hierarchical musical structures will be presented. Then, the discussion will focus on representational issues of the musical surface with emphasis on the representation of melodies. The core of this chapter is a general representation of pitch intervals and a model for the derivation of the *melodic surface (1)* (a sophisticated representation of melodic intervals) from the *melodic surface (0)* (a mere sequence of discrete musical notes).

## 5.1 The *CHARM* Representation

It is essential that the musical representation, on which our model will be based, will be as flexible, manipulable, expressive and structurally general as possible to support the multitude of tasks described above. A representation which rates high in terms of expressive completeness and structural generality (Wiggins et al., 1993) is the *CHARM* representation developed by Harris, Smaill and Wiggins (1991).

The *Common Hierarchical Abstract Representation for Music (CHARM)* is intended to free the representation of music from application constraints and specific music domain restrictions. This is achieved by separating the 'concrete' representation used in practice by a musician or musical program, from the 'abstract' mathematical properties pertinent to it (based on the computer science notion of abstract data types). For example, the determination of a pitch interval between two notes described by some abstract mathematical property (depending on the two pitch values) can be concretely instantiated in many different ways depending on whether the pitches are represented in Hertz or number of semitones, etc.

At the lowest level of abstraction (below this level properties of events are outside the formal system) the *CHARM* events are discrete entities of any sort (e.g. notes or other primitives - see next section) with explicitly defined properties (e.g. pitch, duration, start-time, dynamics, etc.). This way many kinds of musical systems can be efficiently expressed (e.g. musical systems based on the equal-temperament semitone scale, quarter-tone scales, non-western scales, etc.). For practical reasons, in the current study only music from the 12-tone equal-temperament system will be examined.

Events may be grouped into higher level *constituents* which are collections of *particles* (events or other constituents). These constituents are labelled with a set of first-order logical formulae that describe the grouping properties of the constituent (or its particles) or with a name that defines an *ad hoc* grouping of particles (e.g. a motif or a piece). By the use of such abstractly defined particles and constituents, structural properties and relations of any sort can explicitly be represented and manipulated.

The *CHARM* system is an adequate musical representation for expressing the multiple viewpoint analytic procedures of our proposed system, as for one melodic surface many different constituents may be constructed that describe it from different perspectives. The structural generality of this representation can support almost any demands posed by the multiparametric and multilevel structural needs described in this analytic research study.

**5.2 Musical Surface**

The acoustic continuum is broken down into elementary events by a listener. 'The identification of each event is an endproduct of the ongoing perceiving process. Without rules to segregate elements, events could not be perceived.' (Handel, 1989:217). Xenakis states: "... if events were absolutely smooth, without beginning or end, and even without modification or 'perceptible' internal roughness, time would find itself abolished. It seems that the notions of separation, of bypassing, of difference, of discontinuity, which are strongly interrelated, are prerequisite to the notion of anteriority. In order for anteriority to exist, it is necessary to distinguish entities, which would then make it possible to 'go' from one to the other." (Xenakis, 1989:87).

The elementary events perceived as constituent units of an acoustic continuum are further grouped together into elementary categories. Research in categorical perception has investigated especially various facets of musical pitch and time perception - see overviews

and discussion in (Dowling and Harwood, 1986; Handel, 1989; Lamont, 1997). It is generally admitted that categorical perception depends not only on the physical acoustic source or on the perceptual sensitivities of the human auditory system but on contextual effects and background knowledge as well (Handel, 1989).

Jackendoff (1987) describes the *musical surface* as being the 'lowest level of representation that has musical significance' (p. 219). In relation to tonal music he states: '... the *musical surface*, encodes the music as discrete pitch-events (notes and chords), each with a specific duration and pitch (or combination of pitches, if a chord). Standard musical notation represents the pitch-events of the musical surface by means of symbols for discrete pitch and duration; ...' (p. 218).

But which exactly is the lowest level of representation that has musical significance? Is it the level of discrete musical primitives (e.g. musical notes for the 12-tone equal-temperament system)? Is it the level at which music is perceived as primitive relations between adjacent musical primitives (e.g. musical intervals, chords, clusters, trills etc.)? There is evidence that things such as melodic and harmonic pitch intervals, chords, start-time intervals, dynamic intervals or larger configurations such as tone clusters, tremolos, trills, glissandi and so on are commonly perceived by listeners as wholes rather than combinations of atomic lower-level components. For example, especially for pitch, it has been suggested that the majority of listeners, for whom musical pitch is relative, perceive pitch intervals categorically prior to individual pitches (Dowling and Harwood, 1986; Handel, 1989). Tenney suggests that larger sound complexes such as tone-clusters or other dense chords 'cannot usually be analysed by the ear into constituent tones, and [he suggests] are not intended to be analysed.' (Tenney, 1961:6) - see also (Cook, 1990); even simpler triadic chords may be perceived as elementary chord types - or even tonal chord function types - before being possibly analysed into their constituent tones and intervals. A glissando is also perceived and can be represented as a single entity with start-pitch and end-pitch, duration and intensity (a linear transition between the two pitches may be implied as a default).[33]

In this study, no commitment to any single one level of the above low-level representations is made; instead, all of the above will be considered as possible elements of the musical surface. A working definition of musical surface - loosely associated with levels of categorical

[33] 'The units ... can form groups with other similar ones. ... For example, suppose there is a glide in frequency, bounded by a rise and fall in intensity. Between these boundaries, the change in frequency may be measured by the auditory system and assigned to the unit as one of its properties. This frequency-gliding unit will prefer to group with other ones whose frequency change has the same slope and which are in the same frequency region.' (Bregman, 1990:644).

perception - will be given whereby the notion of musical surface will be broken down into two distinct levels: *musical surface (0)* and *musical surface (1)* - the general term *musical surface* will refer to either or both of these without distinction:

- *Musical Surface (0)* will refer to the lowest-level representation of a musical work which consists merely of discrete quantised musical events (e.g. notes with discrete pitch, duration, dynamic values etc.).
- *Musical Surface (1)* will refer to a slightly higher-level representation of a musical work which consists of discrete quantised musical intervals (e.g. melodic and harmonic pitch intervals, start-time intervals, dynamic intervals etc.), or of larger compound entities such as chords, trills, glissandi, and possibly relations (distances) between them (e.g. chord distances).[34]

The derivation of the *musical surface (0)* from the acoustic signal is a complex process. Ideally a theory like the *General Computational Theory of Musical Structure* should interact with and complement lower-level acoustic and psychoacoustic theories in attempting to quantise the acoustic signal. However, for convenience, it is herein assumed that the description of a musical work as an ordered collection of discrete quantised musical primitives is a prerequisite - a given input - to the *GCTMS*. Once the *musical surface (0)* has been selected as an adequate level of representation for a particular musical idiom the *GCTMS* can be employed in order to obtain higher-level descriptions.

Should higher-level articulatory features of scores or expressive features of performance be considered part of the musical surface? Jackendoff's (1987:217-219) description of the notion of the musical surface refers mainly to properties of individual notes (such as pitch, duration, timbre) and seems to exclude such features at least at the level of phrase structure;[35] in practice, however, Lerdahl and Jackendoff (1983) make extensive use of articulation marks such as slurs and breath marks, e.g. in the strong local detail grouping preference rule GPR2a (p.45). As in the current study the only input to *GCTMS* is the *musical surface (0)* all expressive features[36] relating to a musical score or performance are excluded. The musical surface is amenable to various expressive interpretations proposed, for instance, by the

[34] In practice, *musical surface (1)* - for instance, as a *GCTMS* component in figure 3.1 - could refer to an even higher-level representation of a musical work such as a sequence of melodic motives accompanied by some method for determining distances between them - see further discussion at the end of section 9.1.3.

[35] '... the presence of phrase boundaries is not marked explicitly in the printed music; in fact, phrase boundaries are determined by grouping and time-span reduction, so they are not even encoded in the musical surface.' (Jackendoff, 1987:236).

[36] See (Clarke, 1985) on structural and expressive features of music.

composer or imposed by the performer; such expressive preferences may be taken into account[37] and actually may be used to guide the analytic process but they are not considered as a necessary prerequisite.

Since the proposed theory will be applied - as a test case - on melodies based on the 12-tone equal-temperament system the input *melodic surface (0)* is represented by a sequence of discrete quantised notes and rests (as in a traditional score). The *melodic surface (1)* is derived from the melodic surface (0) and is represented by a number of distinct interval profiles (sequences of intervals) for the various parametric properties of the notes at a number of abstraction levels - for instance, for pitch: exact pitch interval profile (in semitones), scale-step interval profile, step-leap profile, contour profile; for time: duration profile, start-time interval profile, relative duration profile (i.e. shorter-longer-equal duration relations) and so on. The derivation of a sophisticated representation of melodic intervals - especially pitch intervals - from the melodic surface (0) is anything but a trivial process as will be shown in the next section.

Pitch and pitch-intervals are most often represented - in the western tradition - either by the traditional pitch naming system and the related pitch-interval names, or as pitch-classes and pitch-class intervals. In the next sections some properties, relationships and limits of these two representations will be presented and a *General Pitch Interval Representation (GPIR)* will be proposed in which the above two systems constitute specific instances. *GPIR* can be effectively used in systems that attempt to represent pitch structures from a wide variety of musical styles (from traditional tonal to contemporary atonal) and can easily be extended to other microtonal environments. Special emphasis will be given to the categorisation of intervals according to their frequency of occurrence within a scale. Finally a model based on the *GPIR* will be presented that enables the derivation of a pitch profile of the melodic surface (sequence of pitch-intervals that embody properties of relevant pitch-scales) from the primitive melodic surface (sequence of discrete notes).

[37] One way of taking into account slurs, staccatti, breath-marks etc. is by considering them to be expressional rests; such rests may be inserted between the notes they mark as normal rests that have a durational value that is a fraction of the preceding note (see section 6.3.2).

### 5.3 Pitch and Pitch-Interval Representation

Many computer-assisted analytic and compositional systems represent pitch intervals as the number of semitones which they consist of. Some other systems, that deal with the tonal system, use the traditional pitch-interval naming system. In the followin sections we will examine the possibility of devising a general representation that can be adapted to different scaling environments according to the musical task at hand.
A major difference between the traditional pitch interval system and the semitone interval system relates to the degree by which each system allows explicit representation of different categories of intervals. On one hand, the traditional interval system allows multi-dimensional encoding of intervals in terms of scale degree distances (e.g. $2^{nd}$, $6^{th}$ etc.), different sizes within the scale degree distances (e.g. major, minor, perfect, augmented, diminished, etc.) and different modality categories, i.e. {perfect}, {major, minor}, {augmented, diminished}. Thus, the traditional system allows explicit representation of different classes of intervals that relate to established hierarchies and functions. On the other hand, the semitone interval system abolishes any such possibility by representing all intervals unidimensionally and thus is adequate for the representation of 12-tone atonal pitch structures.

Various studies of music cognition (Deutsch, 1982b, 1984; Bharucha, 1984a,b; Sloboda, 1985; Dowling and Harwood, 1986; Krumhansl, 1990; McAdams, 1989) suggest that most musical systems establish different degrees of hierarchic taxonomies amongst the various musical elements that facilitate cognitive processing of a musical structure. In this chapter we will examine one facet of such hierarchies, namely the hierarchic organisation of the pitches and pitch intervals of a scale or set of scales over the full space of discrete pitch elements available in a given musical system.

Two enharmonic intervals in a tonal musical domain are very different although they consist of exactly the same number of semitones. The reason for this distinction lies in the structural properties that are assigned to each interval depending on the structural context in which it appears. For example, an isolated ascending interval of three semitones can be heard in the tonal domain either as a minor $3^{rd}$ or an augmented $2^{nd}$. If this same interval is preceded and followed by an ascending semitone, it is recognised as an augmented $2^{nd}$ interval, as this specific sequence is encountered only on the $5^{th}$ degree of a harmonic minor scale. Our mind tries to match the heard sequence to the learned scale schemata of the major-minor system in

an attempt to place the sequence in a higher level tonal framework. In the case of the above sequence, our mind makes a first selection placing the sequence in the minor scale and considering the last note of the sequence as the tonic. As new intervals are encountered the first assumption is either reinforced or altered (if the new data give evidence that a better selection can be made).

The structural/functional properties of intervals within larger pitch schemata allow a finer classification than the one made if only their physical properties[38] are taken into account. This way, the 3 pc-interval (pc:pitch-class) can be further subdivided into the minor 3rd class and the 'rare' and very characteristic augmented 2nd class allowing, thus, an explicit representation of intervallic properties that relate to more abstract tonal schemata.

Such structural properties may either be explicitly represented in a pitch representation of a specific musical system or may be left to be implicitly inferred by other processes. Depending on the musical task at hand, a more refined representation may be more efficient (despite its seeming redundancy at the lowest pitch level) as it allows higher level musical knowledge to be represented and manipulated in a more precise and parsimonious manner.

Brinkman (1990), in his discussion of encoding pitch and pitch intervals for computer applications, proposes a *binomial* system whereby he brings together the 12 pc-set theory (Forte, 1973; Rahn, 1980) and the diatonic set theory (Regener, 1964; Clough , 1979, 1980; Clough et al., 1985). The latter suggests that the 12-tone pc-set formalism can be applied to the seven diatonic name classes; an integer from 0-6 stands for each letter-name (C$\rightarrow$0, D$\rightarrow$1, ... A$\rightarrow$6) and a modulo 7 mathematical formalism is developed. In the binomial system each pitch is represented by an integer couple the first of which is pitch-class and the second name-class (e.g. following the form [pc, nc] the note G# is [8,4] and Ab is [8,5]). Pitch intervals are encoded in a similar manner (e.g. augmented 2nd is [3,1] and minor 3rd is [3, 2]). This representation enables encoding of enharmonic pitches and pitch intervals. Brinkman proceeds to develop a set of mathematical operations that can be performed between the elements of the binomial system.

Following this direction of investigation, we will attempt to propose a *General Pitch Interval Representation (GPIR)* that can be applied to any M-tone scale set over an N-tone equal-

[38] Enharmonic intervals were originally physically different, until the equal-temperament tuning forced them into identity, and, even today, enharmonic intervals, when performed on non-tempered instruments (e.g. voice, violin etc.), appear in different physical sizes (different intonation) depending on musical context (Schackford, 1961, 1962).

tempered discrete pitch space (M≤N). In the *GPIR* system the modality of a name-class interval is explicitly represented by the introduction of a separate symbol which is calculated from its frequency of occurrence - relating to Browne's theory (Browne, 1981) on the importance of intervallic rarity. It will be shown that both the 12-tone and the traditional diatonic representations are conveniently accommodated within the *GPIR* and that this general-purpose representation expresses efficiently a wide range of other scale environments that may illustrate a varying degree of hierarchical organisation.

### 5.3.1 The *General Pitch Interval Representation (GPIR)*

In this study we will deal with equal-tempered scaling systems and more specifically with the 12-tone equal-temperament. The only equivalence assumed is octave equivalence under which any two pitches separated by a number of octaves are considered structurally equivalent - the octave equivalence assumption is an essential part of most musical systems (Dowling and Harwood, 1986; Trehub et al., 1997). All other kinds of equivalence (e.g. inverse interval equivalence) are not embodied explicitly in the *GPIR* but can easily be inferred by the use of simple operations on the *GPIR* primitives.

#### 5.3.1.1 Pitch Representation

In the proposed system two pitch symbols relate directly to the structure of a scale. The first is taken from a set of integers that is used to represent the scale tones. The number of elements of this set is equal to the number of scale tones (i.e. 7 integers for 7-tone scales, 8 for 8-tone scales and so forth). Integer 0 is mapped onto note C of the diatonic system. This integer representation is a natural extension of the diatonic name-class representation discussed above. The second symbol is selected from a set of modifiers-accidentals. For these we use positive integers to stand for sharps, zero for natural and negative integers for flats (e.g. -2 → 𝄫, -1 → ♭, 0 → ♮, 1 → ♯, 2 →𝄪). In the following table the traditional accidental symbols are used for matters of readability.

traditional representation:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7-tone diatonic scale | C | ♯/♭ | D | ♯/♭ | E | F | ♯/♭ | G | ♯/♭ | A | ♯/♭ | B | (C) |

*GPIR* representation[39]:

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7-tone diatonic scale | 0 | ♯/♭ | 1 | ♯/♭ | 2 | 3 | ♯/♭ | 4 | ♯/♭ | 5 | ♯/♭ | 6 | (0) |
| pentatonic scale | 0 | ♯/♭ | 1 | ♯/♭ | 2 | ♯/♭♭ | 𝄪/♭ | 3 | ♯/♭ | 4 | ♯/♭♭ | 𝄪/♭ | (0) |
| octatonic scale | 0 | ♯/♭ | 1 | 2 | ♯/♭ | 3 | 4 | ♯/♭ | 5 | 6 | ♯/♭ | 7 | (0) |
| 12-tone scale | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | (0) |

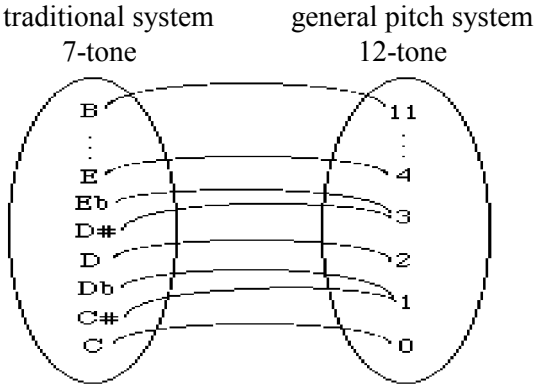
In the *GPIR* every pitch is represented by an array of the sort [nc, mdf, pc, oct] where *nc* (name-class) takes values from {0, 1, 2, ..., M} for an M-tone scale, *mdf* (modifier) takes values from {-u, ..., -1, 0, 1, ..., u} and u is the number of pitch interval units in the largest scale-step interval, *pc* (pitch-class) takes values from {0, 1, 2, ..., N} for an N-tone discrete equal-tempered pitch space and *oct* is octave range (middle C octave is 4). For example, in the diatonic system D4 is [1, 0, 2, 4], D♯4 is [1, 1, 3, 4], E♭5 is [2, -1, 3, 5], G♭3 is [4, -1, 6, 3]. Enharmonic notes are represented with different arrays although enharmonic equivalence can be identified through the *pc* entry. In the 12-tone system D4 is [1, 0, 1, 4], D♯4 is [3, 0, 3, 4], E♭5 is [3, 0, 3, 5], G♭3 is [6, 0, 6, 3] and the first two entries become redundant as *nc* is identical to *pc*, and the modifier symbol is always 0. This representation can easily be applied to any other equal-temperament scaling systems as, for example, the twelfth-tone Aristoxenian system[40] (Aristoxenos; Xenakis, 1992).

[39] Alternatively, integers may correspond to the symbols assigned to the elements of the discrete pitch space (columns in the table below consist of the same letter-symbols) facilitating thus pitch representations especially in cases where within the same piece of music we have changes of scaling systems, as pitch names remain invariant within the overall pitch structure. Of course, in this representation, the modulo M (for M-tone scales) mathematical formalisms do not any longer apply.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7-tone diatonic scale | 0 | ♯/♭ | 2 | ♯/♭ | 4 | 5 | ♯/♭ | 7 | ♯/♭ | 9 | ♯/♭ | 11 | (0) |
| pentatonic scale | 0 | ♯/♭ | 2 | ♯/♭ | 4 | ♯/♭♭ | 𝄪/♭ | 7 | ♯/♭ | 9 | ♯/♭♭ | 𝄪/♭ | (0) |
| octatonic scale | 0 | ♯/♭ | 2 | 3 | ♯/♭ | 5 | 6 | ♯/♭ | 8 | 9 | ♯/♭ | 11 | (0) |
| 12-tone scale | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | (0) |

[40] In the Aristoxenian pitch system (Aristoxenos; Xenakis, 1992) the smallest pitch-interval unit is the twelfth-tone. The *tone* is defined as the difference between the perfect fifth (dia pente) and the perfect fourth (dia tessaron) and can be divided into two parts called *semitones* (6 twelfths), three parts called *chromatic dieseis* (4 twelfths) or four parts called *enharmonic dieseis* (3 twelfths). Three of these are combined to form *tetrachords* (total of 30 twelfths i.e. $2^1/_2$ tones). There are three genres of tetrachords: a. *enharmonic* (3+3+24=30 segments), b. *chromatic* (soft: 4+4+22=30, hemiolon: 4.5+4.5+21=30 and toniaion: 6+6+18=30) and c. *diatonic* (soft: 6+9+15=30 and syntonon: 6+12+12). (If it is required that

Before ending this section on pitch representation, we will briefly address some issues concerning the transcription of a piece of music from a traditional system of pitch notation (Western or otherwise) to the proposed *GPIR*, and the inverse. In general, the relation that allows conversion of a pitch structure from an M-tone to an N-tone representation (where M-tone is a subset of N-tone), is a mathematical *function,* i.e. for every element of the M-tone set there is one and only one element of the N-tone set that corresponds to it. In this case, transcription can be uniquely defined and realised.




When a pitch structure represented by an M-tone notation is converted to an N-tone notation and the M-tone is not a subset of the N-tone notation, the conversion relation is not a function and thus transcription is not a uniquely defined process (e.g. note 1 of the 12-tone scale can be either transcribed as C♯ or D♭ in the 7-tone diatonic scale). In this case, additional rules are necessary to allow selection of one possible transcription over another. This issue will be addressed in section 5.3.3.

---

all intervals, e.g. the ones in the chromatic hemiolon, are expressed in integer numbers then the tone should be divided in 24 segments.) Tetrachords and tones are further combined to form *systems.*

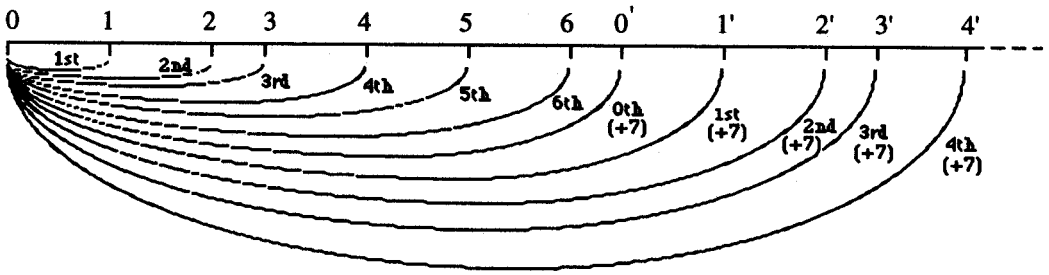
As an example, let us create a system which consists of two syntonon diatonic tetrachords (6+12+12=30) disjunct by a tone. If octave equivalence is further assumed, this *system* is the diatonic genre. This genre can be represented by 7 nc integers {0, 1, ... 6} for the 7-tone scale, 72 pc integers {0, 1, ...71} for the 72-tone discrete pitch space and 25 mdf integers {-12, -11, ...-1, 0, 1, ... 11, 12} since the largest possible scale step interval is the tone (12 units). For instance, between the scale tones [2, 0, 24, 4] and [3, 0, 30, 4] there exist 5 discrete pitches with two possible enharmonic spellings each e.g. for one of these: [2, 2, 26, 4] and [3, -4, 26, 4].

The Aristoxenian scaling system may accommodate a wide gamut of microtonal systems because of its fine resolution of intervals.

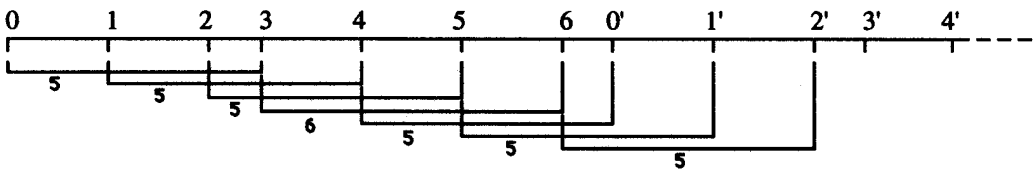
#### 5.3.1.2 Pitch interval representation

The structure of a scaling system affects the functions and properties that may be assigned to other musical quantities, such as pitch intervals, that directly relate to it. In the *GPIR* two interval symbols relate directly to inherent properties of a given scale:

1. *Name-class interval (nci):* this integer indicates the number of scale steps that an interval consists of and is calculated as the modulo M difference between the name-class integers (for an M-tone scale). Taneiev (1902/1962:25-33) first introduced a similar way of naming intervals wherein the symbol 1st was used for the scale step interval - not 2nd as in the traditional interval system (this facilitates direct mathematical operations between intervals, such as addition and subtraction, e.g. 1st+4th=5th). For instance, for a 7-tone scale we would thus have:




2. *Modality*: The second interval symbol is determined by the *frequency of occurrence* of each member of the subset of intervals that relate to the *nci* integer. If we calculate the number of times that all the different modalities of a specific name-class interval occur within a scale (taking as its lower note each degree of the scale), we can classify intervals depending on their frequency of occurrence.[41] For example, the interval of a fourth in the diatonic genre occurs 6 times at the size of 5 semitones (frequency of occurrence F=6/7=0.86) and once at the size of 6 semitones (F=1/7=0.14):

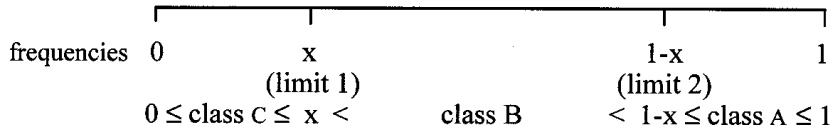



[41] Every genre of scales will have exactly the same set of intervals and frequency of their occurrences, i.e. it doesn't matter which tone is considered to be the tonic in a particular mode.

Table 5.1 illustrates the name-class intervals (as 1st, 2nd etc.), the frequency of their occurrences and the interval size in semitones (top row) for different kinds of genres of scales.

The naming process of the traditional interval system, wherein, for instance, a fourth is called perfect when it contains 5 semitones and augmented when it contains 6 semitones, seems to correspond to the above observation concerning the frequency of occurrence of intervals,[42] e.g. perfect intervals occur most frequently between the degrees of the scale whereas augmented are rare.

The problem in defining the second symbol is the definition of the limits that will classify name-class intervals into different categories. As a default we propose to have 3 classes (borrowed from the traditional system) defined by two symmetric limits:




where *x=0.25* (this is an arbitrary selection of a limit that seems to work well for our purposes; further research may define a better value or range of values for limit x).

The frequency of occurrence of a scale interval of a specific size over the total number of scale degrees on which it can be based is F=n/N, where n=number of occurrences for that interval size and N is total number of scale degrees. For this limit (i.e. lower limit=0.25 and upper limit=0.75), class A contains at maximum one member (as each *nci* may occur only in one modality with a frequency over 75%), class B maximum four elements and class C maximum N elements. So, in general: class A = {A}, class B = $\{B_1, B_2, B_3, B_4\}$ and class C = $\{C_1, C_2, ..., C_N\}$. Intervals that do not appear between scale tones may be encountered between scale tones and non-scale tones or between non-scale tones. For these intervals, the modality symbol is selected from class D.

Table 5.2 depicts the resulting two-symbol names for all the intervals of the genres of scales presented in table 5.1. Some comments on table 5.2 are presented below:

[42] This view seems to relate to Krumhansl's observation (Krumhansl, 1990:273) that there is a link between the consonance of an interval and its frequency of occurrence, although any direct connection of modalities of intervals to degrees of consonance is herein avoided.

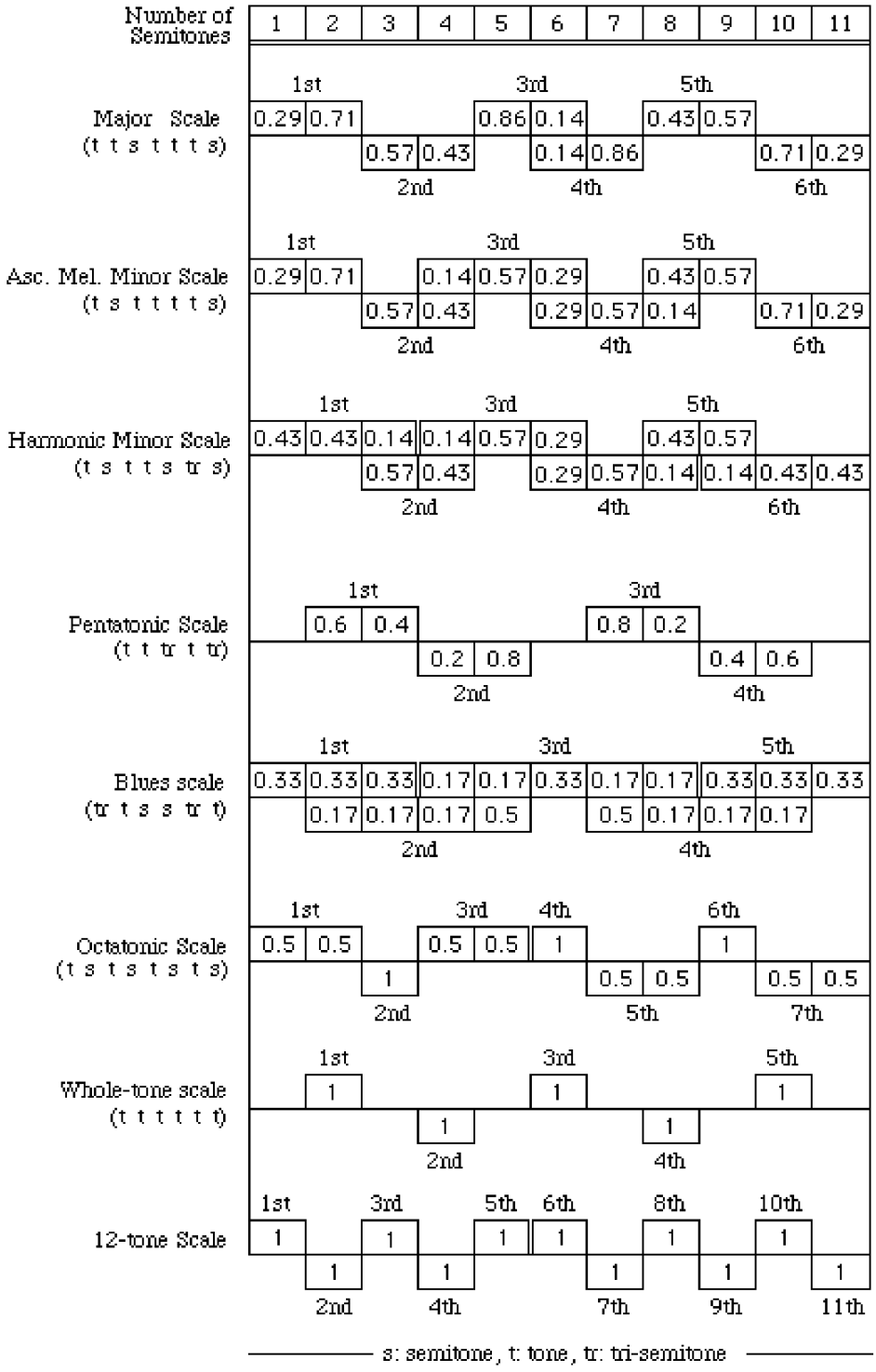

Number of Semitones
1 2 3 4 5 6 7 8 9 10 11
Major Scale (t t s t t t s)
1st 3rd 5th
0.29 0.71 0.86 0.14 0.43 0.57
0.57 0.43 0.14 0.86 0.71 0.29
2nd 4th 6th
Asc. Mel. Minor Scale (t s t t t t s)
1st 3rd 5th
0.29 0.71 0.14 0.57 0.29 0.43 0.57
0.57 0.43 0.29 0.57 0.14 0.71 0.29
2nd 4th 6th
Harmonic Minor Scale (t s t t s tr s)
1st 3rd 5th
0.43 0.43 0.14 0.14 0.57 0.29 0.43 0.57
0.57 0.43 0.29 0.57 0.14 0.14 0.43 0.43
2nd 4th 6th
Pentatonic Scale (t t tr t tr)
1st 3rd
0.6 0.4 0.8 0.2
0.2 0.8 0.4 0.6
2nd 4th
Blues scale (tr t s s tr t)
1st 3rd 5th
0.33 0.33 0.33 0.17 0.17 0.33 0.17 0.17 0.33 0.33 0.33
0.17 0.17 0.17 0.5 0.5 0.17 0.17 0.17
2nd 4th
Octatonic Scale (t s t s t s t s)
1st 3rd 4th 6th
0.5 0.5 0.5 0.5 1 1
1 0.5 0.5 0.5 0.5
2nd 5th 7th
Whole-tone scale (t t t t t t)
1st 3rd 5th
1 1 1
1 1
2nd 4th
12-tone Scale
1st 3rd 5th 6th 8th 10th
1 1 1 1 1 1
1 1 1 1 1 1
2nd 4th 7th 9th 11th
s: semitone, t: tone, tr: tri-semitone

Table 5.1

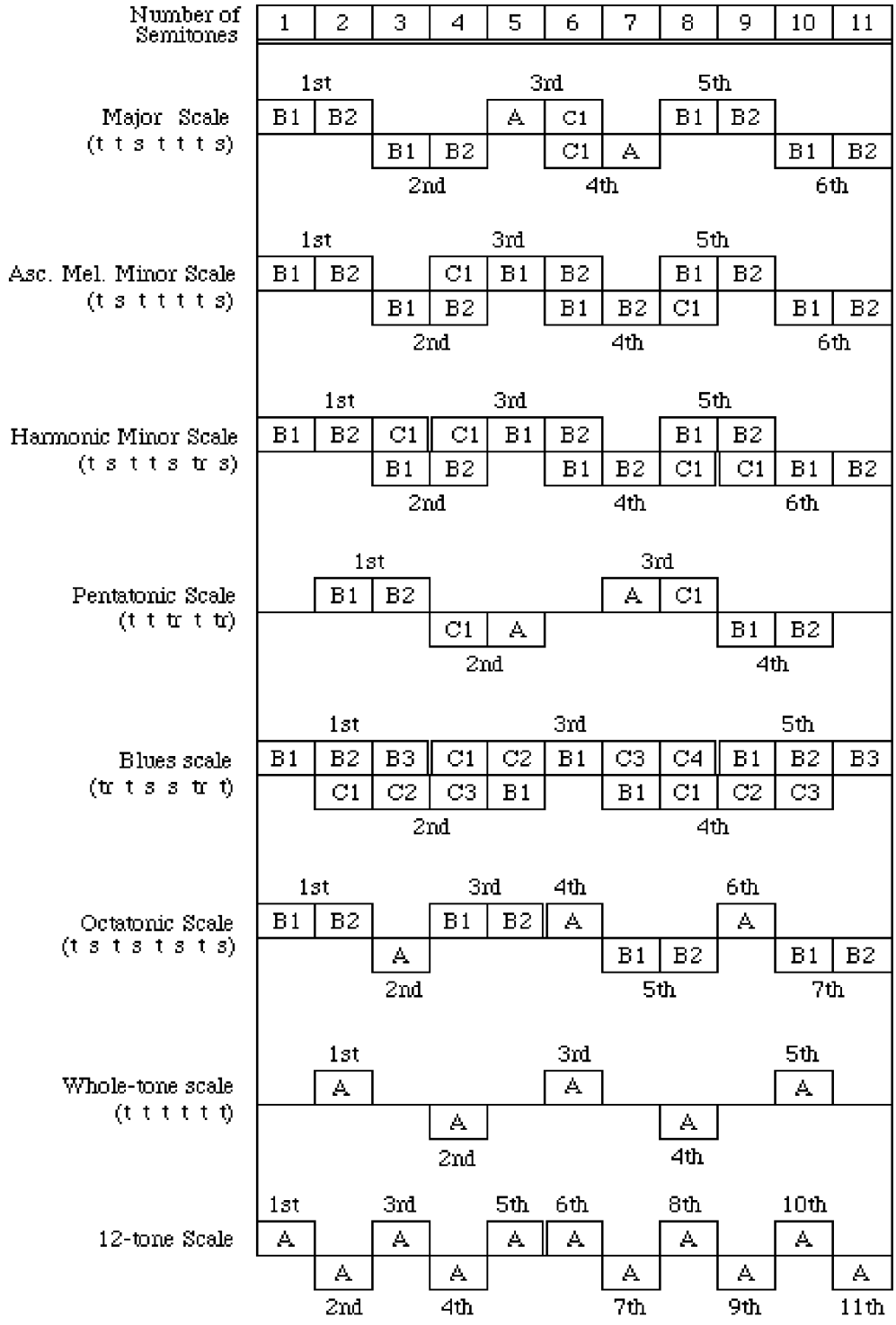

Number of Semitones
1 2 3 4 5 6 7 8 9 10 11
Major Scale (t t s t t t s)
1st 3rd 5th
B1 B2 A C1 B1 B2
B1 B2 C1 A B1 B2
2nd 4th 6th
Asc. Mel. Minor Scale (t s t t t t s)
1st 3rd 5th
B1 B2 C1 B1 B2 B1 B2
B1 B2 B1 B2 C1 B1 B2
2nd 4th 6th
Harmonic Minor Scale (t s t t s tr s)
1st 3rd 5th
B1 B2 C1 C1 B1 B2 B1 B2
B1 B2 B1 B2 C1 C1 B1 B2
2nd 4th 6th
Pentatonic Scale (t t tr t tr)
1st 3rd
B1 B2 A C1
C1 A B1 B2
2nd 4th
Blues scale (tr t s s tr t)
1st 3rd 5th
B1 B2 B3 C1 C2 B1 C3 C4 B1 B2 B3
C1 C2 C3 B1 B1 C1 C2 C3
2nd 4th
Octatonic Scale (t s t s t s t s)
1st 3rd 4th 6th
B1 B2 B1 B2 A A
A B1 B2 B1 B2
2nd 5th 7th
Whole-tone scale (t t t t t t)
1st 3rd 5th
A A A
A A
2nd 4th
12-tone Scale
1st 3rd 5th 6th 8th 10th
A A A A A A A
A A A A A A
2nd 4th 7th 9th 11th

Table 5.2

a. In the octatonic scale there exist three class A intervals one of which is the tritone. There are no class C intervals ('rare' intervals).

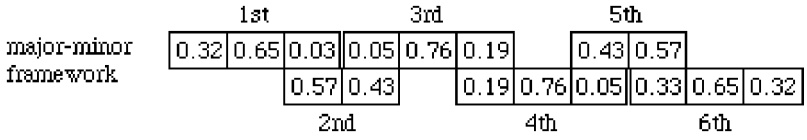
b. The 12-tone scale[43] and the whole-tone scale consist only of class A intervals and, thus, the modality symbol becomes redundant and may be dropped altogether. For the chromatic scale the *nci* integer coincides with the *pci* (pitch-class interval) integer (e.g. the $4^{th}$ interval is identical to the 4 pc-interval and consists of 4 semitones). One can see that the pitch-class interval representation is an instance of the proposed general system.

c. For the diatonic genre (including the major and natural minor scale) the traditional interval names emerge, if the following 'traditional' symbols are used: class A = {perfect}, class B = {minor, major}, class C = {diminished, augmented}.

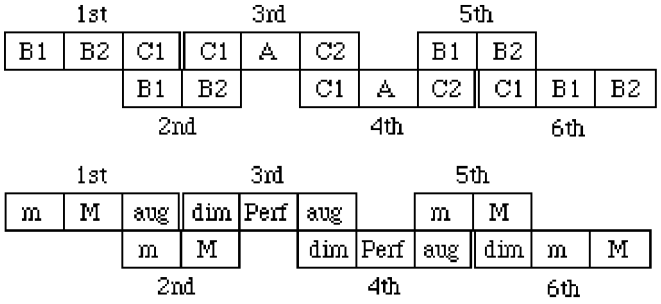
d. For the ascending melodic and the harmonic minor scales naming of intervals is somewhat different from the traditional system (e.g. $3^{rds}$ and $4^{ths}$ have a class B modality instead of class A). One may notice though that these scales hardly ever appear exclusively on their own. They are an integral part of a wider major-minor framework (even a piece of music that is composed solely on the harmonic minor mode cannot eliminate the significance obtained from the absent 'opposite' major mode). If we weigh[44] each kind of scale (e.g. 4 x major scale, 1 x natural minor, 1 x desc. melodic minor, 1 x asc. mel. minor & 2 x harmonic minor, add all occurrences for each interval and divide by 9) we arrive at the following results:

[43] It may be preferable to analyse atonal music with an N-tone (N<12) scale system as an atonal composition may micro-structurally be based on N-tone scale fragments.

[44] This weighting is not a result of any comprehensive analysis (cognitive, statistical or otherwise). Its aim is to represent all the different kinds of the major-minor scales in a balanced manner. It attempts to give half weight to the major scale and half to the minor scales (the natural minor scale actually reinforces both sides as it consists of intervals identical to those of the major scale - they both belong to the same genre of diatonic scales).




From this weighted frequency of occurrence values we derive all the traditional interval names for the major-minor scales:

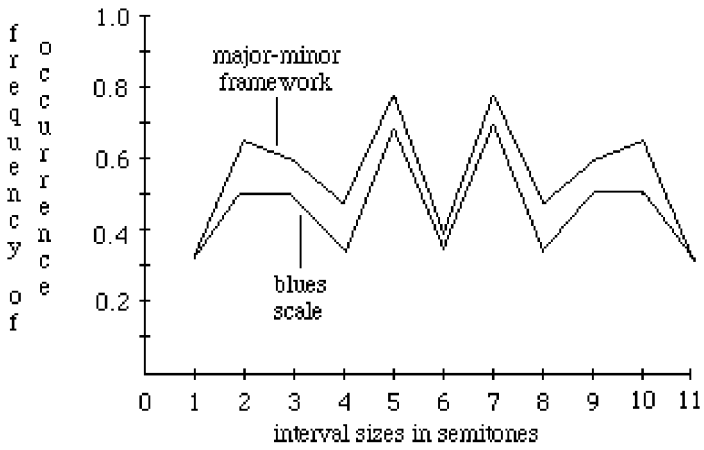






From the above it is obvious that the traditional interval representation is only an instance of the proposed general system.

e. 'Blending' different scales together seems to be a useful method of obtaining a broader interval representation. The use of more than one genre of scales is commonly employed in some musical styles. Such scales usually exhibit a similar interval 'character' i.e. they have a similar frequency of occurrence for all intervals or the most important ones. In the following graph, one can discern the similarity between the major-minor scale framework and the blues scale (the blues scale appears usually in a major-minor context within jazz music). The same interval representation may also be used for the major scale and the pentatonic scale as the tones and intervals of the latter are a subset of the former.



In the *GPIR* every pitch interval may be accurately represented by an array of the sort [dir, nci, mdl, pci, oct] where *dir* (direction) takes values from {-, =, +} depending on the direction of the interval, *nci* (name-class) takes values from {0, 1, 2, ... M} for an M-tone scale, *mdl* (modality)[45] takes values from class A, B, C or D, *pci* (pitch-class interval) takes values from {0, 1, 2, ...N} for an N-tone discrete equal-tempered pitch space and *oct* is the number of octaves within compound intervals. For instance, in the traditional diatonic system an ascending augmented 2nd is [+, 1, C1, 3, 0], a descending minor 3rd is [-, 2, B1, 3, 0], an ascending major 9th is [+, 1, B2, 2, 1] whereas the same intervals in the 12-tone system are [+, 3, A, 3, 0], [-, 3, A, 3, 0] and [+, 2, A, 2, 1]. In the latter case the nci and mdl entries become redundant.

#### 5.3.2 Applications and Uses of the *GPIR*

The *GPIR* has been implemented in a PROLOG programming environment; the user presents to the system the interval array of a selected scale (or weighted set of scale interval arrays) and the system induces and stores the appropriate *GPIR* information (e.g. number of scale tones, number of discrete pitch elements, modality interval names, possible enharmonic spellings of notes and so forth). A set of operations has been developed that can be performed on the *GPIR* primitives in order to compute the interval between two pitches, the inverse of a given interval, the transposition of a pitch by a given interval and so on.

This representation increases the complexity of categorisation of intervals at the lowest level but, as it embodies structural properties that are inherent to the given scaling system, it facilitates reasoning and manipulation of the pitch material at higher levels of analytic and compositional processes. It has the advantage of encoding efficiently pitches and pitch intervals from a hierarchical tonal system down to a distributional 12-tone system.

Probably the most interesting aspects of this representation is the possibility to represent on computers other scaling systems in a way which is most relevant to them - e.g. pentatonic, octatonic, 9-tone scales or even uncommon 7-tone genres (e.g. s-s-t-t-t-t-t). It may be the case that the lack of musical systems residing in the territory in between the traditional highly hierarchical tonal system and the distributional atonal system, is related to inefficiencies

[45] The modality symbol may be broken down into a two element list containing a modality symbol {a, b, c, d} or {1, 2, 3, 4} and an index number that is assigned to different members of the same modality class; the index number may indicate the number of units that an interval is greater or lesser than a reference size in that modality.

inherent in the traditional notation system. How can a composer notate, for instance, a functional 8-tone tonal piece on the traditional 7-tone stave notation? She/he either has to spend endless hours distinguishing the scale tones from the secondary non-scale tones (for instance, see (Gillies, 1993) on pitch notation and tonality in Bartók's music) or invent and learn a new notation system! The *GPIR* may enable computer-assisted compositional systems to compose music in hierarchical/functional systems other than the 7-tone diatonic system.

The *GPIR* could also be used creatively in analytic/compositional programs by forcing an analysis (or composition) based on 'wrong' scaling-interval representations (e.g. analyse 7-tone music with a 9-tone interval representation, etc.). One may impose the structural and functional properties of a given piece to different scale representations. This kind of experimentation could lead to novel and interesting compositions.

This representation may easily be adapted or extended to meet the needs of musical systems (ethnic musics, experimental scaling environments etc.) other than the Western 12-tone equal-tempered system.

It is suggested that a flexible pitch interval representation, such as the *GPIR*, may prove itself indispensable when devising a computer system that attempts to deal with a wide variety of musical styles. Two applications are presented that highlight the representational advantages of the *GPIR* in devising a) a transcription program (next section) and b) a pattern-matching system (section 7.3).

#### 5.3.3 Transcription of melodies based on the *GPIR*

As stated in section 5.3.1.1, the transcription of a piece of music from an M-tone system to an N-tone, where the M-tone system is not a subset of the N-tone, is not a function and, thus, is not a straightforward process. We have implemented a system that converts melodies from a 12-tone notation (MIDI) to the traditional 7-tone notation based on the *GPIR* theory (an important similar system implemented from a cognitive perspective appears in Longuet-Higgins, 1976/1987). The principle of classifying intervals according to their frequency of occurrence is strongly supported by this application.

The transcription system applies two basic principles:

1) *Notational Parsimony* (i.e. 'spell' notes making minimum use of accidentals[46])

2) *Interval Modality Optimisation* (i.e. prefer intervals in the order of their frequency of occurrence - most preferable: class A - least preferable: class D).

A numerical grading of the different parameters that relate to these principles is devised:

Interval Notational Parsimony:

| | |
|---|---|
| non-enharmonic spelling of notes | 0 |
| enharmonic spelling of one note | 2 |
| both notes enharmonic | 6 |

Interval Modality Optimisation:

| | |
|---|---|
| intervals of class A or B[47] | 0 |
| intervals of class C | 1 |
| intervals of class D | 4 |

[46] This actually means to avoid the enharmonic spelling of notes that can be notated without any accidentals e.g. prefer C and avoid B♯ & D𝄫.

[47] It is not possible to have for one name class interval both modalities of class A and B as this would give an overall frequency of occurrence greater than 100% - this could actually be taken as a constraint on the value of threshold x which has to be $x \leq 0.25$.

For any given sequence of MIDI pitch numbers all the alternative spellings of each pitch are found. For example, for the beginning of the theme of Bach's *Musical Offering* we have:

| 60 | 63 | 67 | 68 | 59 | 67 | 66 | 65 | 64 | 63 | 62 | 61 | 60 ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D𝄫 | | A𝄫 | | C♭ | A𝄫 | | G𝄫 | F | | E𝄫 | | D𝄫 |
| \| | E♭ | \| | A♭ | \| | \| | G♭ | \| | \| | E♭ | \| | D♭ | \| |
| C | \| | G | \| | B | G | \| | F | E | \| | D | \| | C |
| \| | D♯ | \| | G♯ | \| | \| | F♯ | \| | \| | D♯ | \| | C♯ | \| |
| B♯ | | F𝄪 | | A𝄪 | F𝄪 | | E♯ | D𝄪 | | C𝄪 | | B♯ |

Then, the program calculates the total sum of the above values for each possible string of traditional pitch names and selects the ones with the minimum sum value.

As the system may find more than one string with the minimum value, we have added one additional rule:

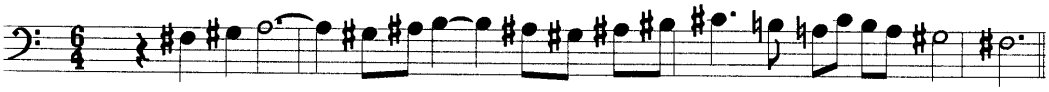
3) *Prefer a sequence in which the higher 'quality' intervals appear last.*

This rule accounts for the asymmetric temporally-ordered aspects of musical perception (Deutsch, 1984; Krumhansl, 1990) according to which listeners, for example, tend to hear the last note of an interval as more prominent. When there are two alternative spellings of two intervals the system prefers the sequence in which the last interval belongs to a 'better' modality class. This rule gives precedence, e.g. to the sequence G - G♯ - A over the equivalent G - A♭ - A (they both have a total value of 4).

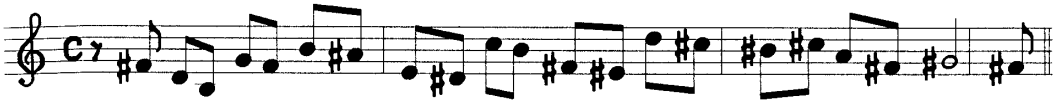
We tested the system over a set of diatonic melodies with unexpectedly good results for such a small and general set of rules (note that there is no higher-level representation of musical knowledge such as keys, tonalities, modulations, tonics etc.).

The transcription programme was applied on the 24 fugue themes from J.S.Bach's *Das Wohltemperierte Klavier I.* All themes were accurately notated with only a few exceptions.

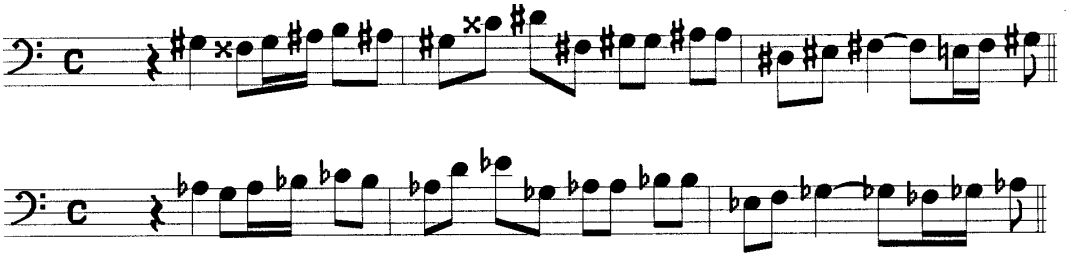
Fugue 14 in F♯ minor (transcription) Identical with original.



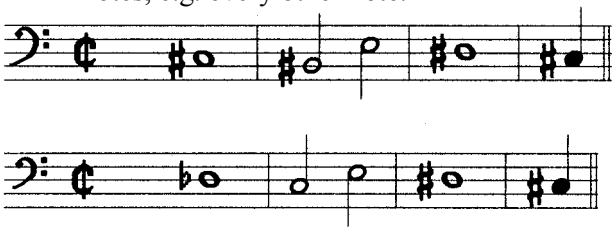
Fugue 24 in B minor (transcription) Identical with original. Note the use of enharmonic spelling of notes in bar 2 (E♯) and bar 3 (B♯).



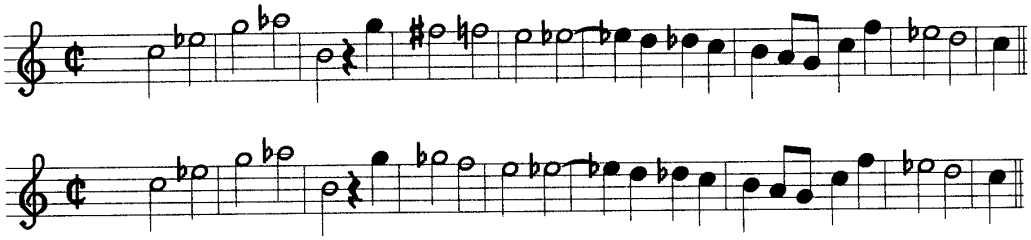
Fugue 18 in G♯ minor (original and transcription) The system prefers the enharmonic key of Ab minor. The same occurs in Fugue 3 (C♯ major) and Fugue 13 (F♯ major).



Fugue 4 (original and transcription) This problem may be bypassed if additional rules are applied such as 'avoid enharmonic spellings of a tone in a single passage', or if the optimisation method is additionally applied to intervals between non-contiguous notes, e.g. every other note.

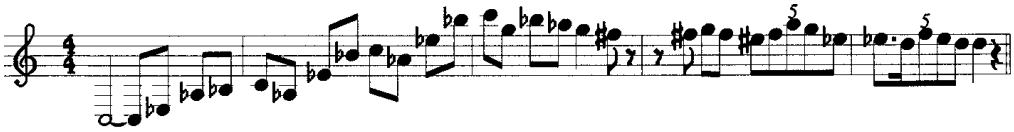


Theme from *Musical Offering* by J.S.Bach (original and transcription) The selection of G♭ in the transcription is due to Rule 3. Both sequences have the same total value. Bach prefers F♯ for harmonic reasons.

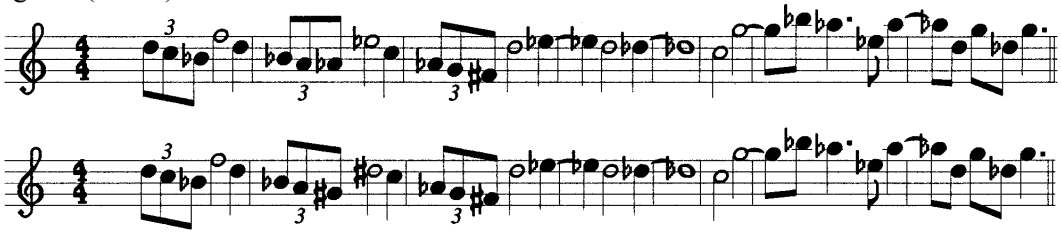


The programme was applied to some melodies from later periods. For example:

Opening from *Ballade* Op. 23 by F. Chopin (transcription). Identical with original.



Excerpt from English Horn solo from third act of *Tristan Und Isolde* by R. Wagner (original and transcription). The incongruence in the second bar is of the same nature as the one in fugue 4 (above).



**5.3.3.1 AI methodology of the transcription programme**

The total number (T) of all possible strings that can be derived from n1 pitches with 2 alternative spellings and n2 pitches with 3 alternative spellings is:

$T = 2^{n1} \cdot 3^{n2}$

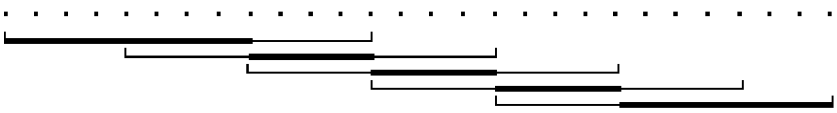
This was significantly reduced by disallowing altogether a) two successive enharmonic notes and b) all class D intervals with the exception of chromatic semitones. T becomes thus approximately[48]:

$T = 2^{n}$ where n = n1 + n2 i.e. total number of notes

[48] For example, two notes with 3 alternative spellings may give $3^2=9$ combinations. Four of these are disallowed by the use of constraint *a* and usually one more by constraint *b* reducing thus the initial number of combinations to approximately $4=2^2$ (e.g. for the interval between MIDI notes 59-67 the spellings C♭-A𝄫 , A𝄪 -F𝄪 , C♭-F𝄪 , Ax-A𝄫 are disallowed by constraint *a* and B-A𝄫 by constraint *b*).

The total number of possible paths given by this function is significantly reduced but still is an *exponential function* of *n* leading, thus, to a combinatorial explosion and making it impossible to calculate the transcription sum values for larger sequences of pitches.

This problem was overcome by implementing an algorithm that transcribes the piece gradually by smaller sections. An overlapping technique was devised in such a way that only the middle part of the transcribed section is selected (marked by the bold segments of the lines in the figure below). This gives stability to the system and avoids misinterpretations of the interval qualities near the edges of the sections.[49]



The above function becomes now:

$T = c \cdot {}^{\nu}/_{\mu} \cdot 2^{\mu} \rightarrow T = ({}^{c}/_{\mu} \cdot 2^{\mu}) \cdot \nu \rightarrow T = k \cdot \nu$ where $\mu$ = number of notes in transcription sections, $\nu$ = total number of notes and c = 3 (each 5-element subsection is transcribed 3 times as beginning, middle and ending of the 13-element transcription sections).

This relation is a *linear function* and melodies of any length can be transcribed within reasonable computational times. The following table shows the values of the three functions for various values of $\nu$:

| | $\nu$ =10 | $\nu$ =20 | $\nu$=50 | $\nu$ =100 | $\nu$ =500 |
|---|---|---|---|---|---|
| $T = 2^{\nu 1 \cdot 3}\ \nu\ 2$ $\nu_1 = \nu_2$ | $8 \cdot 10^3$ | $6 \cdot 10^7$ | $3 \cdot 10^{19}$ | $8 \cdot 10^{38}$ | $3 \cdot 10^{194}$ |
| $T = 2^{\nu}$ | $10^3$ | $10^6$ | $10^{15}$ | $10^{30}$ | $3 \cdot 10^{150}$ |
| $T = k \cdot \nu$ (k=1890, c=3, $\mu$=13) | $8 \cdot 10^3$ ($T = 2^{\nu}$, $\nu$ =13) | $4 \cdot 10^4$ | $9 \cdot 10^4$ | $2 \cdot 10^5$ | $9 \cdot 10^5$ |

How good are the transcription results obtained by this shifting overlapping technique compared to the results obtained by the method that transcribes a whole melody at once? Both methods were tested over a number of melodies generating always identical results. The

[49] An instance of boundary problems caused by a non-overlapping transcription technique can be demonstrated in Bach's fugue in B min. If the transcription section boundary is on 6th note of bar 2 then this note will be spelled E♯ as the last note of the preceding section and F as the first note of the following section!

reason for this is that intervals of class C and D tend to appear isolated in between unambiguous stable sections of class A and B intervals.[50] The sections that may receive alternative spellings with a similar sum value are, in most cases, short - usually just a few notes. This localisation of the transcription process allows a shifting overlapping method to yield good results (although, in general, it is not necessarily true that the results obtained by the two techniques are always identical).

This technique of a step-by-step transcription by overlapping sections is also closer to the processes that take place while a listener is notating little-by-little a heard melody (melodic dictation). The listener hears and notates a few bars at a time making possible alterations to the immediately preceding notes if this is required by the new input.

The results obtained by the simple transcription system described above reinforces the case for having a hierarchical classification of pitch intervals according to their frequency of occurrence within a scale as suggested by the *GPIR*. This system may form a basis for developing more sophisticated software for the transcription of MIDI scores into the traditional notation; it may also be used as a precursor to the construction of a key-finding system - counting the number of sharps or flats proposed by the transcription programme may be the basis of such a system.

**Conclusion**

In this chapter, firstly, the *Common Hierarchical Representation for Music (CHARM)* which is adequate for representing hierarchical musical structures was briefly presented. Then, representational issues relating to the musical surface were addressed; it was argued that the *musical surface* may be represented both as a sequence of discrete primitive events such as notes - termed *musical surface (0)* - and as a slightly higher-level collection of musical interval profiles (or as a succession of multi-event complexes such as chords, trills etc.) - termed *musical surface (1).*

Especially for pitch, it was shown that the proposed *General Pitch Interval Representation* introduces a better way of encoding pitch and pitch intervals depending on the specific scale qualities on which musical works are based. It is maintained that the hierarchy of scale tones

[50] This relates to the fact that 'if X Y Z are three successive notes of a melody which, on paper, are separated by chromatic intervals XY and YZ, then there is always an alternative, simple interpretation of the middle note Y which transforms both intervals into diatonic ones.' (Longuet-Higgins, 1987:113)

over a discrete pitch space makes possible - and even necessary - the more elaborate classification of pitches and pitch intervals according to their higher level structural properties. The flexibility of this representation renders it an ideal candidate for computer systems that attempt to manipulate musical structures from diverse musical domains with a varying degree of hierarchic organisation. A computer application was presented that enables the conversion of a sequence of absolute pitches (MIDI pitch) to the traditional diatonic pitch notation. Some other benefits of adopting the *GPIR* representation are given in section 7.3.

# 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 20$^{th}$ 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)).

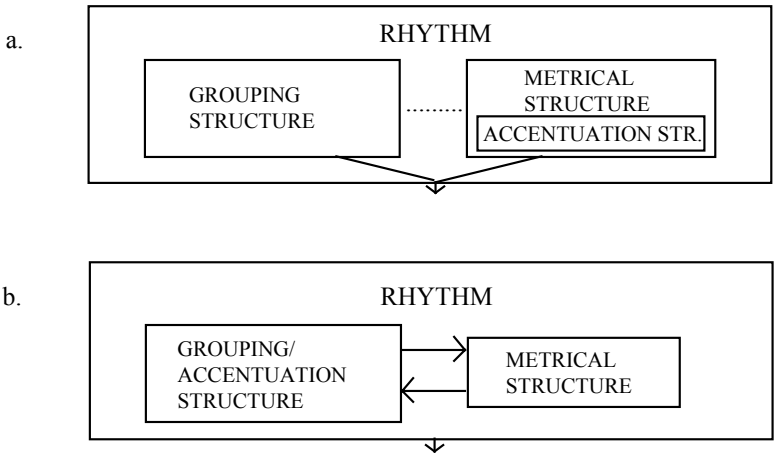



Figure 6.1 a. Lerdahl & Jackendoff's theory of musical rhythm
b. Proposed model of musical rhythm

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).

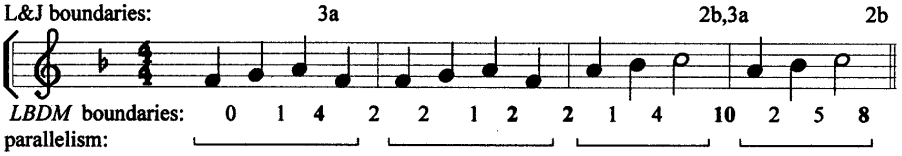



Figure 6.2 Beginning of *Frère Jacques*. Higher-level grouping principles override some of the local detail grouping boundaries (note that *LBDM* gives local values at the boundaries suggested by parallelism - without taking in account articulation - whereas Lerdahl & Jackendoff do so only for the 3rd and 4th boundary).

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).

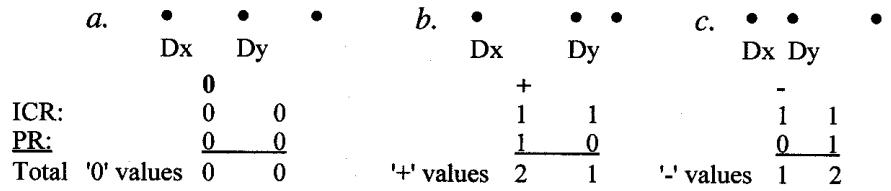



Figure 6.3 Boundary strengths (last row) calculated by the use of the ICR and PR rules for three parametric values (e.g. pitch, dynamics etc.) separated by two intervals.

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).

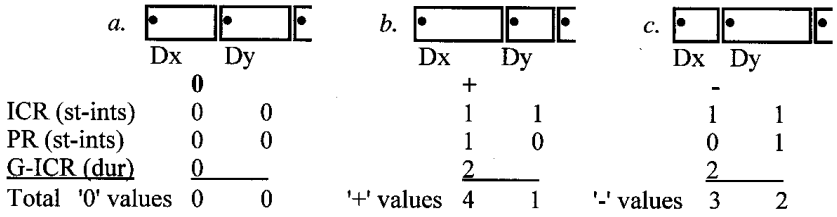



Figure 6.4 Boundary strengths (last row) calculated by the use of the ICR and PR rules for three start-time values separated by two start-time intervals and durations.

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.

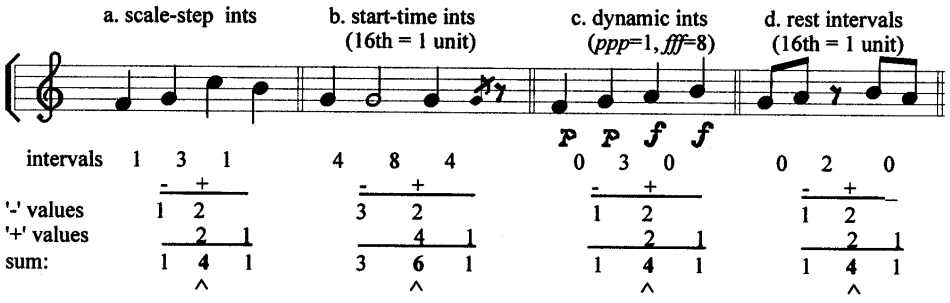



Figure 6.5 Examples of boundary strengths (last row) determined by the *LBDM*.

[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

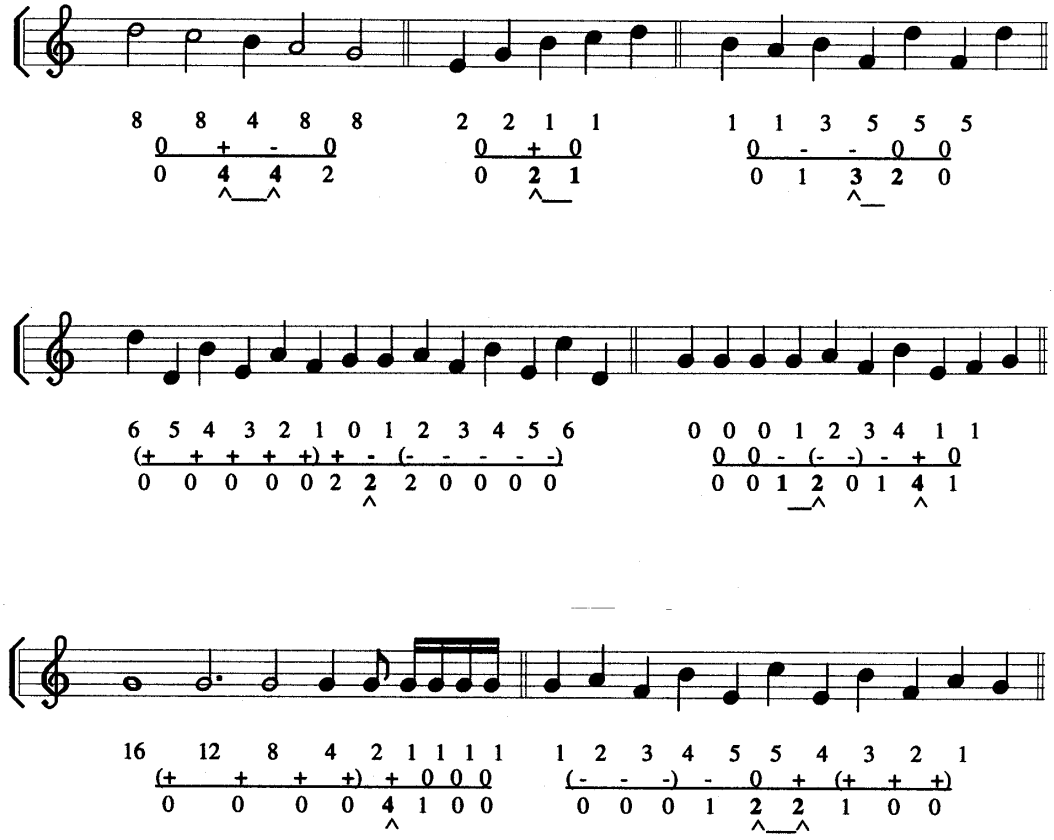



Figure 6.6 Examples of boundary strengths (last row) determined by the *LBDM*. These are ambiguous boundaries which may be resolved if higher-level organisational principles are taken into account.

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 *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).

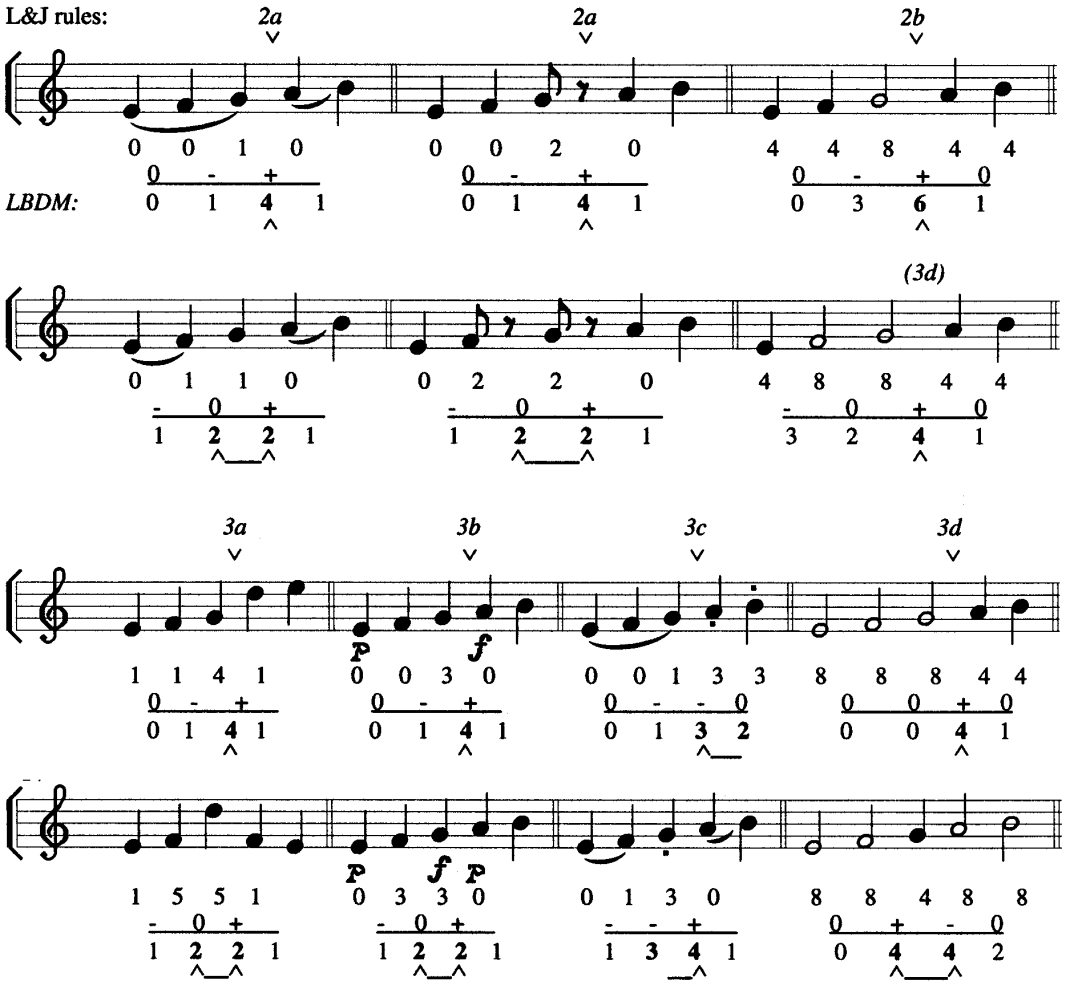



Figure 6.7 Application of the *Local Boundary Detection Model* to the Lerdahl & Jackendoff (1983:44-46) local detail grouping examples 3.14-3.17. For the examples not accounted for by the GPR2 and GPR3 rules, the proposed theory suggests ambiguous boundaries (depicted as ∧__∧ ).

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.

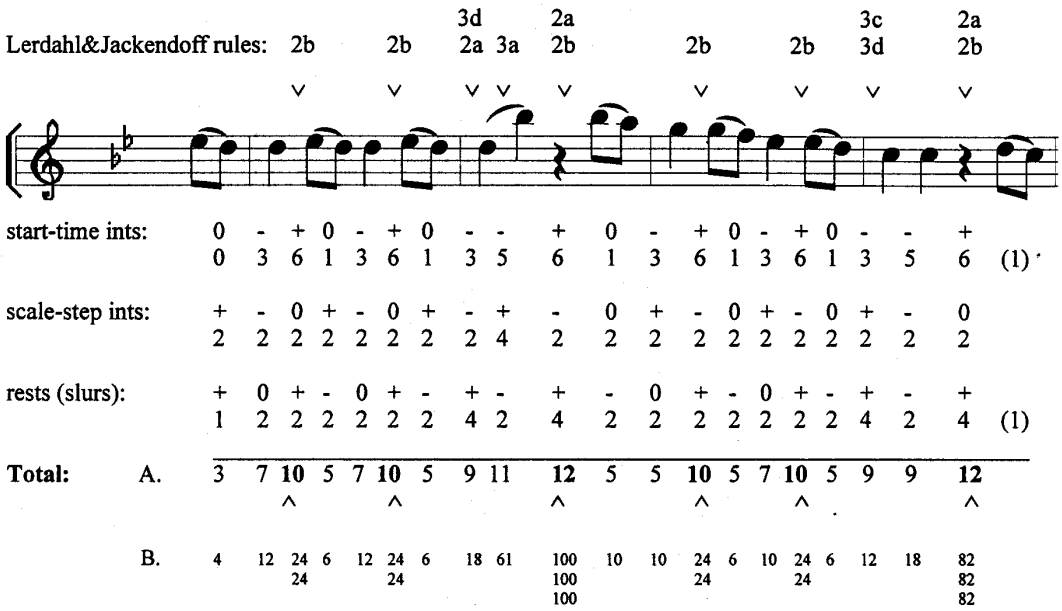



Figure 6.8 Low-level grouping structure for the theme of Mozart's *Symphony in G min.* The boundary strengths sequence A is determined by the *LBDM* whereas sequence B is determined by the refined version of *LBDM* described in section 6.3.5 (slurs are not taken into account)

*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.

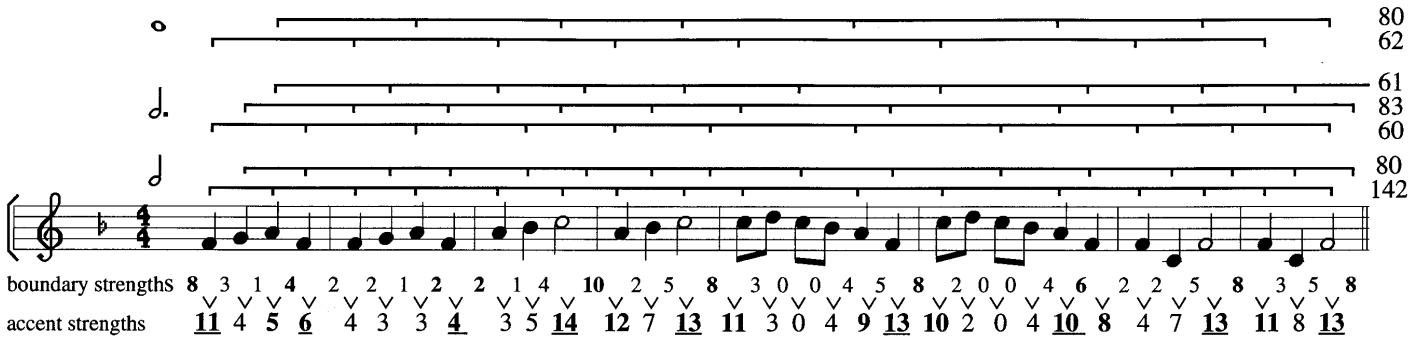



Figure 6.9: Accentuation and metric structure for the song *Frère Jacques*.

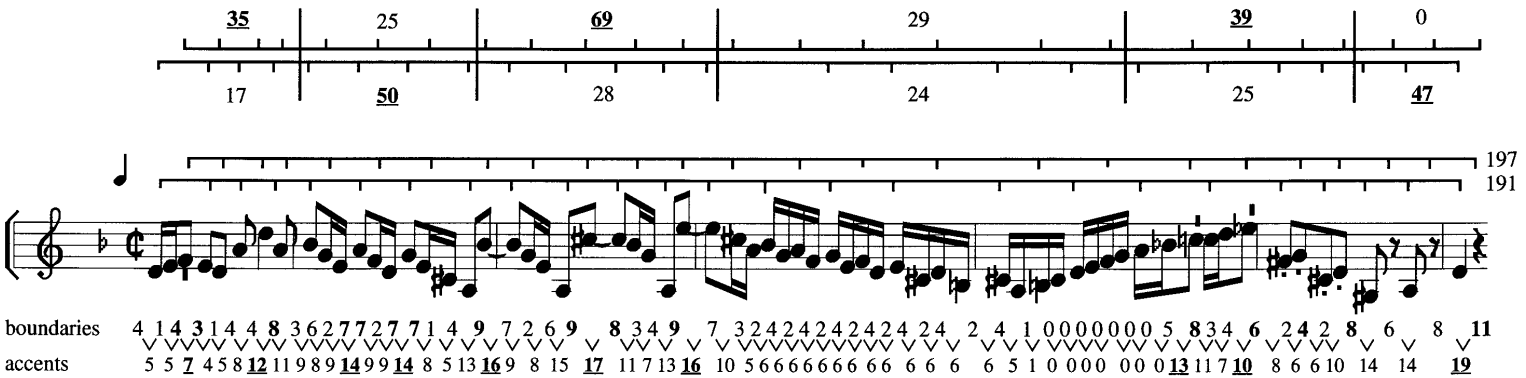



Figure 6.10: Accentuation and metric structure for the beginning of J.S.Bach's *Concerto for Harpsichord in D min.*

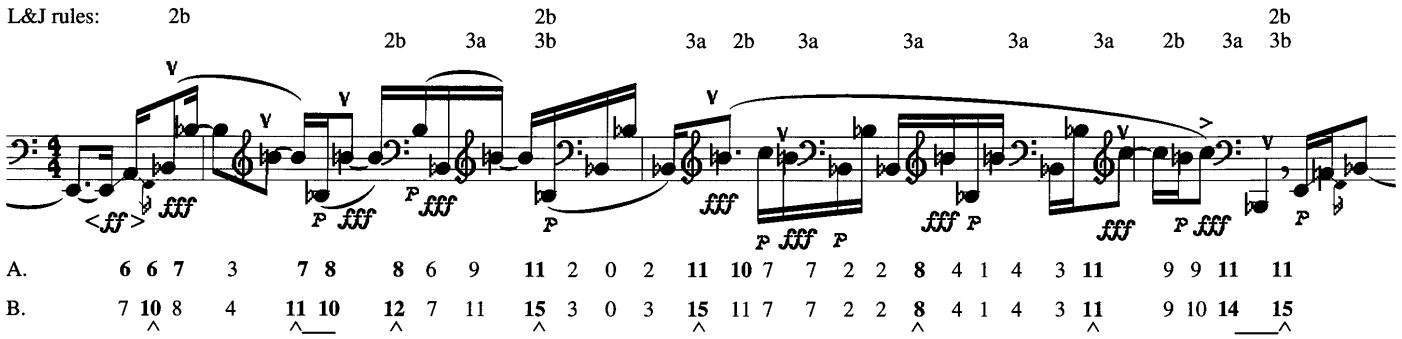



Figure 6.11 Suggestion of local boundaries for an excerpt from Xenakis' *Keren* for solo trombone. Boundary strength sequence A has been calculated without taking into account slurs; sequence B with slurs. Note that the grouping structure proposed by the composer happens to be very close to the boundaries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

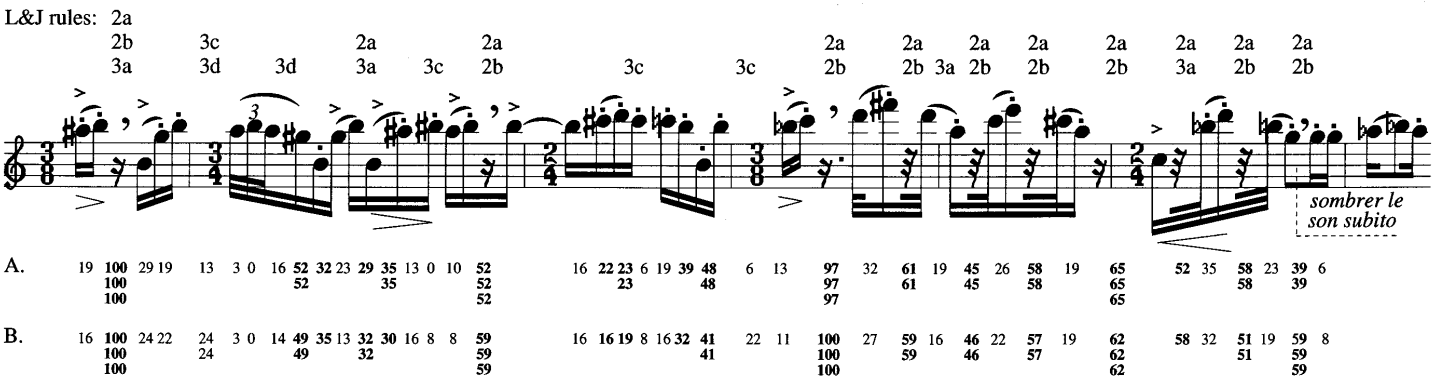



Figure 6.12 Suggestion of local boundaries for an excerpt from Stravinsky's *Three pieces for solo clarinet,* no. III. Sequence A (without articulation marks) and sequence B (with articulation marks) indicate the most probable local boundaries calculated by the refined version of *LBDM* (see section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


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 $p \rightarrow f$ is larger than the reverse $f \rightarrow p$. The sequence below will have the following grouping boundaries:

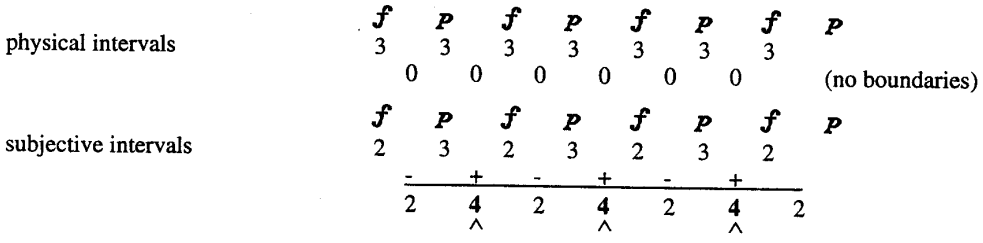



Figure 6.14

• 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).

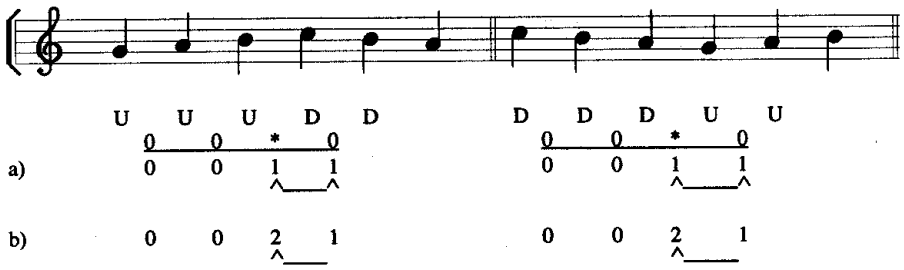



Figure 6.15

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 $\alpha = |(x-y)/(x+y)|$ where x, y are positive integer interval sizes[2] 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 $\alpha=0$.

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

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 β - 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:

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 α 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.*[3] *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 α encapsulates the degree of change/difference between two successive intervals (refined version of ICR rule). By multiplying factor α with the absolute size of each interval the change strength value of factor α 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 β 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.

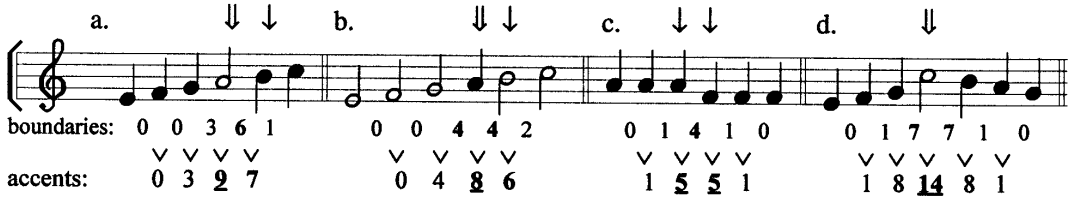



Figure 6.16 Examples of phenomenal accent strengths derived from the *LBDM* boundary strengths by merely adding every two adjacent boundary strength values.

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 (Parncutt, 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.

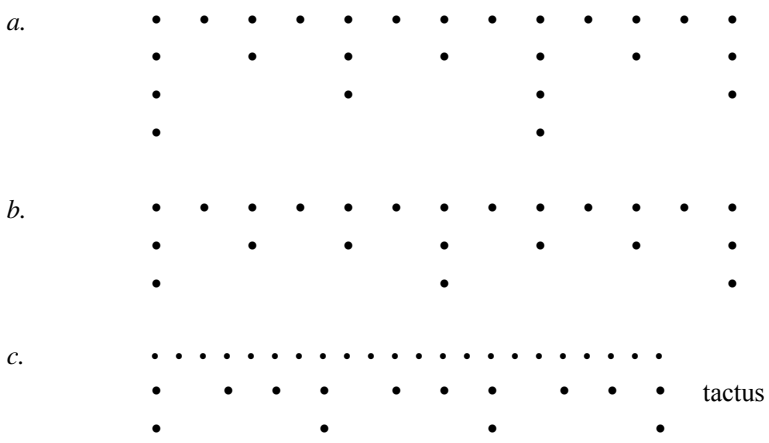



Figure 6.17 Examples of well-formed metrical grids.

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[4] 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

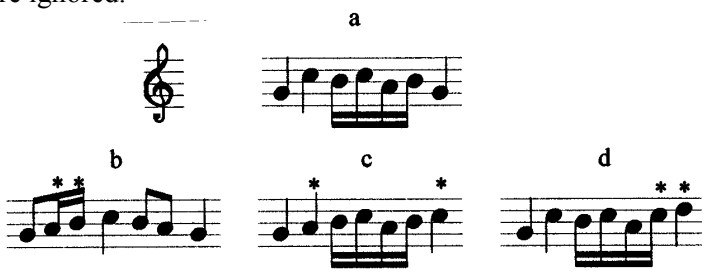
Full pattern-matching is aimed at finding instances of given patterns or inducing identical patterns. However, pattern-matching may be used for revealing or establishing similarity between different patterns as well. What kind of pattern-matching methodology, though, is most adequate when attempting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




segment a: [g,0],[c,4],[b,8],[c,9],[a,10],[b,11],[g,12]
segment b: [g,0],[a,2],[b,3],[c,4],[b,8],[a,10],[g,12]
segment c: [g,0],[a,4],[b,8],[c,9],[a,10],[b,11],[c,12]
segment d: [g,0],[c,4],[b,8],[c,9],[a,10],[c,11],[d,12]

Figure 7.1

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.

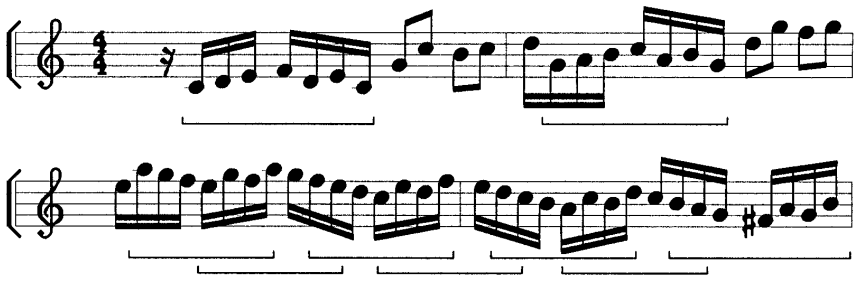


Figure 7.2 An overlapping pattern/motive in the beginning of J.S.Bach's *Two-part Invention No. 1* (highlighted by the *SPIA & Selection Function*).

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.

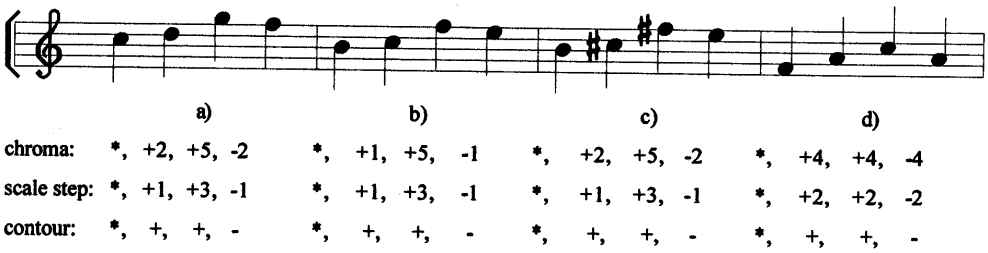



Figure 7.3 'A simple figure (a), requires at least three different methods of encoding pitch intervals for repetition to be detected by a matching process. Repetition with in-scale transposition (b) requires scale step encoding; repetition with simple transposition (c) requires chroma (pitch class) encoding; 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[5] 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.

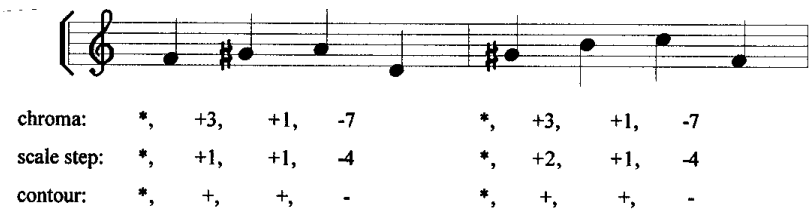



Figure 7.4

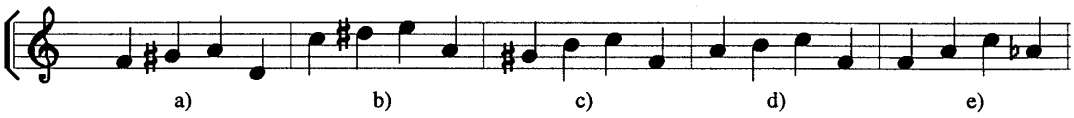
b. The scale-step diatonic matching level is arbitrary in a distributional atonal environment (based on the 12-tone system). A quantification of the chroma level into equal numbers of semitones may be less arbitrary (e.g. 2-semitone intervals, and so on).

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.[6] 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.




| | a) | b) | c) | d) | e) |
|---|---|---|---|---|---|
| **For 7-tone diatonic representation:** | | | | | |
| [dir,nci,mdf]* | *,+1C1,+1B1,-4A | *,+1C1,+1B1,-4A | *, +2B1,+1B1,-4A | *,+1B2,+1B1,-4A | *,+2B2,+2B1,-2B2 |
| [dir,nci]: | *, +1, +1, -4 | *, +1, +1, -4 | *, +2, +1, -4 | *, +1, +1, -4 | *, +2, +2, -2 |
| [dir,*nci'*]:[3] | *, +1, +1, -4 | *, +1, +1, -4 | *, +2, +1, -4 | *, +1, +1, -4 | *, +2, +2, -2 |
| intermediate levels: | . . . . . | | | | |
| [dir]: | *, +, +, - | *, +, +, - | *, +, +, - | *, +, +, - | *, +, +, - |
| **For 12-tone representation ([dir,nc,mdf] level is redundant as mdf is always A):** | | | | | |
| ([dir,nci,mdf]* | *,+3A, +1A,-7A | *, +3A, +1A, -7A | *, +3A, +1A, -7A | *, +2A, +1A, -7A | *, +4A, +3A, -4A) |
| [dir,nci]: | *, +3, +1, -7 | *, +3, +1, -7 | *, +3, +1, -7 | *, +2, +1, -7 | *, +4, +3, -4 |
| [dir,*nci'*]: | *, +3, +1, -7 | *, +3, +1, -7 | *, +3, +1, -7 | *, +2, +1, -7 | *, +2, +2, -4 |
| intermediate levels: | . . . . . | | | | |
| [dir]: | *, +, +, - | *, +, +, - | *, +, +, - | *, +, +, - | *, +, +, - |

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.[7] 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,

[7] An efficient algorithm that computes all the repetitions in a given string is described in (Crochemore, 1981; Iliopoulos et al., 1996) - not as yet been implemented as part of the current prototype system. This algorithm takes $O(n \cdot \log n)$ time where n is the length of the string. It should also be noted that this algorithm does not match retrograde and inverted forms of patterns.

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.

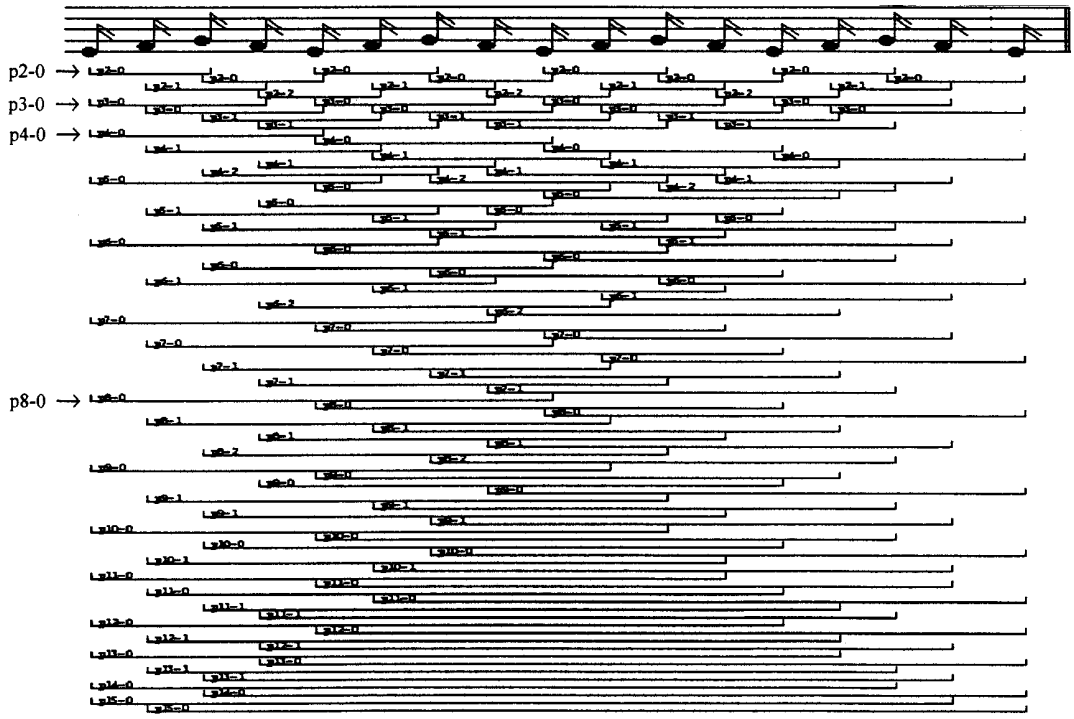



Figure 7.6 A great number of pitch-interval pattern matches is found by the *SPIA* in this short trivial melodic sequence.

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:

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

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 $f(PL,F,DOL)=F^{a}\cdot PL^{b}$ 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

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

[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·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.

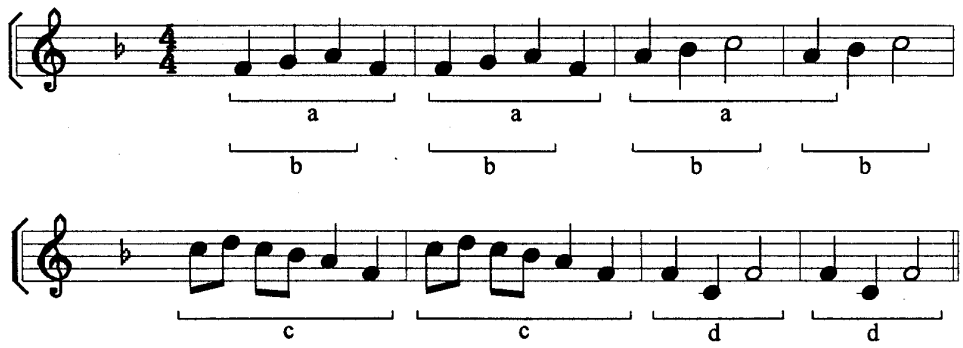



Figure 7.7 *Frère Jacques* - most prominent pitch-patterns highlighted by the *SPIA* and *Selection Function* (*SPIA* applied only on scale-step pitch profile for original patterns, and *Selection Function* constants set to (a,b,c)=(3,3,4))

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).

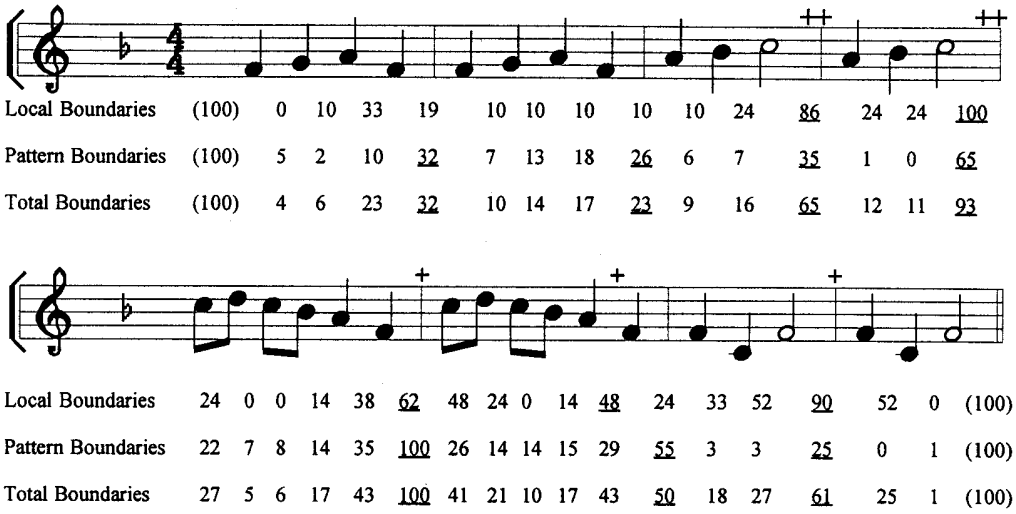



Figure 7.8 Local Boundaries strength profile (refined *LBDM*), Pattern Boundary strength profile and a weighted Total Boundary strength profile for the song *Frère Jacques.*

## 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:

$$s_h(x,y) \begin{cases} \geq 0 \text{ iff } d(x,y) \leq h \quad (\textit{similar} \text{ entities}) \\ < 0 \text{ iff } d(x,y) > h \ (\textit{dissimilar} \text{ entities}) \end{cases} \qquad \text{(I)}$$

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:

$C_k=\{x_1,x_2,...x_n\}$ such that: $\forall i,j \in \{1,2,...n\}$, $s_h(x_i,x_j)\geq 0$ (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 $0\leq h\leq d_{max}$ where the distance $d_{max}$ is defined as the maximum distance observed between all the pairs of entities in T, i.e. $d_{max}=max(d(x,y))$.

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 $0<h<d_{max}$ 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 $C_1 \cup C_2 ... \cup C_m = T$ and m sets of properties such that $P_k \subseteq P$, $k \in \{1,2,...m\}$ and $P_1 \cup P_2 ... \cup P_m \subseteq P$.

If $h=d_{max}$ 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 $0 \leq h \leq d_{max}$, 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 $0 \leq w \leq 1$ 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: $1 < m \leq N/2$, 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:

$w = |m/n - m'/(N-n)|$ where: (III)

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:

$A_1$: shapes {square, triangle, circle}
$A_2$: size {small, big}
$A_3$: shade {white, grey}
$A_4$: content {dot, cross, heart}
$A_5$: outline {plain, double, bold}

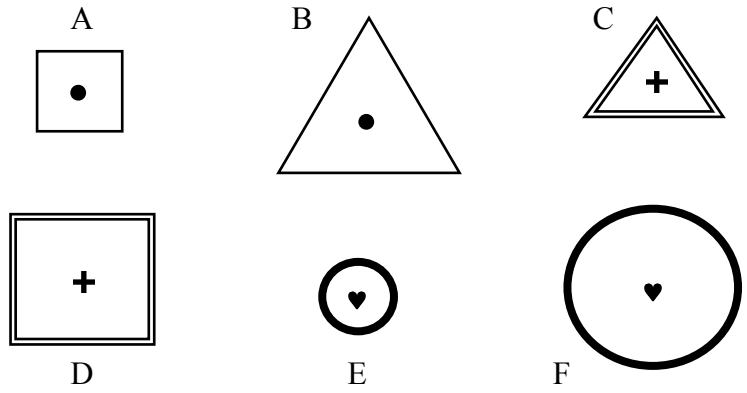



Figure 8.2 Set of objects T for categorisation

Each object X is represented by an array of n=5 attribute values: $(x_1,x_2,x_3,x_4,x_5)$, e.g. for object E: (circle, small, white, heart, bold).

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

$$d(x,y)=\sum_{i=1}^{n} w_{x_i} \cdot w_{y_i} \cdot |x_i - y_i| \qquad \text{(IV)}$$

where: $|x_i - y_i| = 0$ if $x_i = y_i$

$|x_i - y_i| = 1$ 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: $s_{AB}=1$ $s_{AC}=1$ $s_{AD}=0$ $s_{AE}=0$ $s_{AF}=0$ $s_{BC}=0$ $s_{BD}=1$ $s_{BE}=0$ $s_{BF}=0$ $s_{CD}=1$ $s_{CE}=0$ $s_{CF}=0$ $s_{DE}=0$ $s_{DF}=0$ $s_{EF}=2$
Categories: {A,B,C,D,E,F}

Threshold: h=3
Similarities: $s_{AB}=0$ $s_{AC}=0$ $s_{BD}=0$ $s_{CD}=0$ $s_{EF}=1$
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%, $1<m\leq 3$, 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: $s_{AB}=1.39$ $s_{AC}=0.83$ $s_{AD}=0.75$ $s_{AE}=0$ $s_{AF}=0$ $s_{BC}=0.75$ $s_{BD}=0.83$ $s_{BE}=0$ $s_{BF}=0$ $s_{CD}=1.39$ $s_{CE}=0$ $s_{CF}=0$ $s_{DE}=0$ $s_{DF}=0$ $s_{EF}=2.04$
Categories: {A,B,C,D,E,F}

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

Threshold: h=1.92
Similarities: $s_{AB}=0.55$ $s_{AC}=0$ $s_{BD}=0$ $s_{CD}=0.55$ $s_{EF}=1.21$
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[11] 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:

$A_1$: shape {square(0.5), triangle(0.5)}
$A_2$: content {dot(0.5), cross(0.5)}
$A_3$: 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.

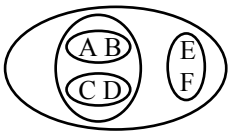



Figure 8.3

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 ($s_h \geq 0$). 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').

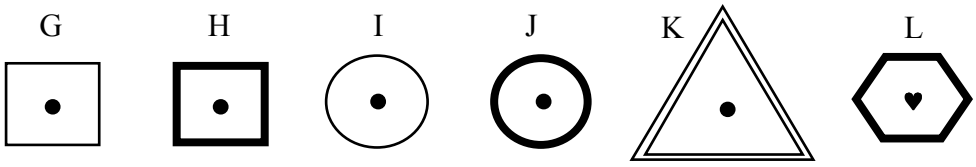



Figure 8.4 Membership predictions for new previously unseen objects.

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.

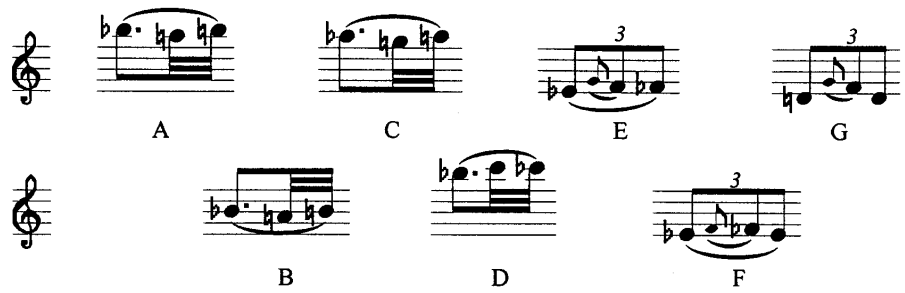



Figure 8.5

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:

$A_{rh}$: {rh1, rh2} $A_{pex}$: {pex1, pex2, pex3, pex4} $A_{pcont}$: {pcont1, pcont2}

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 ($w_{rh1}$=0.75, $w_{rh2}$=0.75, $w_{pex1}$=1, $w_{pex2}$=0.5, $w_{pex3}$=0.25, $w_{pex4}$=0.25, $w_{cont1}$=1, $w_{pcont2}$=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 → Categories: {A,B,C,D},{D,E,F,G}
Threshold: h=2 → 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: $w_{rh1}$=1, $w_{rh2}$=1, $w_{pex1}$=0.75, $w_{pex2}$=0.08, $w_{pex3}$=0.33, $w_{pex4}$=0.33, $w_{cont1}$=0.75, $w_{pcont2}$=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.

# Chapter 9

# Overall Model and Four Analyses

## Introduction

In this chapter the computational components of the *GCTMS* presented in the previous chapters are combined in order to obtain analytic descriptions of four melodies. The main aim of these analytic examples is to highlight the capabilities of the proposed overall model, to give some preliminary evidence of the generality of the theory and to present problem areas that require further study.

Initially the main function of each computational component is summarised and the default settings used for the purposes of the four analytic examples are given. Then, the overall model is applied on four melodies taken from diverse musical idioms in order to obtain analytic structural descriptions of them. Some aspects of the overall computational model which have not as yet been fully implemented as part of the prototype computer system are described in section 9.1.4.

## 9.1 Overall model based on *GCTMS*

The analytic engine of the overall computational model is based on the individual components which have been outlined in section 3.4.3 and have been described in more detailed in the previous chapters. In this section a detailed description will be given as to how these components are combined and interact with each other; additionally the default

values that have been selected in order to obtain the analyses presented in section 9.2 will be given.

The overall design of the computational model based on the *GCTMS* is illustrated in figure 3.1 (attention should be focused on the computational components of the theory depicted in oval shapes). The following computational components are applied on a *melodic surface (0)* in order to obtain an analytic description of it: the transcription program based on the *General Pitch Interval Representation,* the refined *Local Boundary Detection Model*, the *Accentuation Model*, the model for *Metrical Matching*, the *String Pattern-Induction Algorithm & Selection Function*, and finally the *Unscramble* algorithm (the *General Chord Representation* and the *Temporal Organisation Model* have not as yet been described - see section 10.2).

### 9.1.1 Musical Input

Let us assume that a melody is presented to the system as an unstructured sequence of notes where each note is represented by a tuple in the form [MIDI_pitch, Quantised_duration]. Four different melodies from diverse musical styles will be examined in section 9.2.

The only musical knowledge the system has access to is the set of musical scale genres and the set of metrical templates that are relevant to the musical idiom this melody belongs to; no other harmonic, tonal, melodic, metrical and articulatory information is available. Of course, this is a severe restriction but it is an interesting experiment for testing the capabilities of the computational model based on *GCTMS*. How far can the proposed theory take us in terms of providing a pertinent analysis of a melody?

### 9.1.2 From melodic surface (0) to segmentation

The given melodic surface (0) is converted into the appropriate pitch notation for the relevant scale genre with the use of the *General Pitch Interval Representation (GPIR);* then, the melodic surface (1) is constructed (represented as a number of parametric interval profiles).

The next step is to apply the refined *Local Boundary Detection Model (LBDM)* on the melodic surface (1) in order to detect points that are most likely to form low-level grouping boundaries. The refined *LBDM* is applied for scale-step pitch intervals, start-time intervals, rest intervals and contour changes. These strength profiles are averaged and normalised (from 0-100), and the Local Boundary strength profile of the melodic surface is revealed.

Local accents are calculated simply by adding every two consecutive boundary strength values. Then, low-level metrical grids may be matched on the local accent strength profile - accent strength values are added for all the notes whose inception coincides with the points of the metric grid and a total value for each grid is computed. A metrical grid is selected if it has a total value for one of its placements on the melodic surface that is significantly larger (for the following examples larger by at least 40%) than the value of each of its other possible placements with a different offset. If this doesn't succeed for any of the possible metric grids at or above the beat level a melody is said not to have a metrical structure.

It is hypothesised that metrical structure (if it exists) and grouping structure have to be co-extensive (with possible local discrepancies) - however, they may be in-phase or out-of-phase. Cases where metrical and grouping structures seem not to be co-extensive (for instance, a 3/4 metrical structure and a 4/4 grouping structure) could be interpreted either as having ambiguous metrical structure (and perhaps ambiguous metrical structures should not be considered metrical structures at all) or that one of the two has been erroneously assigned to the melodic surface (at least in perceptual terms). This hypothesis is vital for the application of the pattern-matching algorithm (see below).

The local boundaries detected by *LBDM* are tentative and have to be coupled by a higher level model that discovers parallel melodic patterns. This is based on the *String Pattern-Induction Algorithm (SPIA)* which finds for each parametric sequence of musical intervals all the patterns that are encountered at least twice in the sequence - from the smallest to the largest. The *SPIA* may be applied to a number of parametric profiles of the melodic surface and to reduced versions of it (e.g. notes on metrically stronger positions or more accented notes and so on).

Linked with this algorithm is the *Selection Function* which assigns higher values to patterns that are more likely to be more prominent in terms of being a) longer, b) more frequent, and c) allow less overlapping between their members (in the examples below the constants in the Selection Function are (a,b,c)=(3,3,4)). This way from amongst the usually great number of discovered patterns some are selected as being more perceptually significant. However the selection of a final set of patterns that best describes a melodic surface is not as straightforward a process as it may seem. For instance, the highest rating pattern is not necessarily the best as two or three lower rating patterns may give a better description of the overall melody.

In order to overcome this problem a very simple but crude methodology has been devised. According to this, 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.

As the *SPIA* is computationally expensive it is useful to add some heuristics that can reduce the search space. Two such methods are proposed: a) to specify only a limited number of pattern lengths (e.g. 2-3 intervals) so that the algorithm does not need to search for *all* the patterns, or b) to insert break markers in positions where significant local boundaries were detected by the *LBDM* (patterns are not allowed to cross over such marked points) and pattern lengths may be allowed to be much longer.

Both of these pattern-matching techniques have been employed in the present system. The first is extremely useful if one is looking for the grouping structure that is at or immediately above the low-level metrical structure discovered by the microstructural module (in the examples below the *SPIA* is applied for 3-4 note patterns on the scale-step interval and the duration parametric profiles of the melodic surface). Metrical grids that are co-extensive with metrical structure discovered by the microstructural module are matched to the pattern boundary strength profile in the same manner as was done for the local accentuation structure. This way co-extensive in-phase or out-of-phase grouping structures

may be detected. If this succeeds then the resultant segmentation may be fed into the categorisation model. If it fails, the second technique may be employed, i.e. break markers ('soft' or 'hard') are inserted in the melodic surface and the *SPIA* is allowed to find matches for a wider range of pattern lengths, parametric profiles and surface reductions (in the examples below the *SPIA* is applied for 3-7 note patterns on the scale-step, contour, duration and relative duration parametric profiles of the melodic surface, and on the exact and scale-step interval profiles of a reduced version of the surface).

Finally, a weighted average of the Local Boundary strength profile and the Pattern Boundary strength profile is calculated and the peaks in the Total Boundary strength profile are selected as the most likely boundaries for the given melodic surface (in the examples below 40% weight is given to the local profile and 60% to the pattern profile). These boundaries provide the basis on which the surface is segmented.

### 9.1.3 From segmentation to paradigmatic description

Once a segmentation or set of segmentations has been selected, the discovered melodic segments are fed into the *Unscramble* algorithm to be organised into categories. For each segment a number of attributes may be assigned. In the four examples the following attributes have been assigned for each segment: exact pitch interval pattern (semitones), scale-step intervals or near-exact pitch interval pattern for 12-tone scales (see figure 9.13), pitch contour, exact duration pattern, relative duration pattern (i.e. sequence of shorter, longer, equal start-time intervals), all the previous attributes for a reduced version of the surface (see section 9.1.4), and, finally, exact pitch interval between first and last note of each segment, and register of each segment (high or low). Obviously there are a large number of other attributes that may be considered important (e.g. inclusion of same subpatterns, more accented notes, harmonically important notes, etc.) but these should suffice for the purposes of this exercise. As a first rough approximation rhythmic attributes have been given half the weight of the attributes relating to pitch.

Then, the *Unscramble* algorithm is applied to the set of melodic segments and attributes, and a preferred categorisation description is selected. Each class of segments (e.g. motive, theme, etc.) is described by a weighted set of attributes (a sort of prototype) that reflects the diagnosticity of each attribute. If a way of measuring the 'goodness' of emerging

paradigmatic descriptions is established then the selection of a 'better' categorisation may influence the selection of one segmentation out of many alternative options. This way categorisation can affect segmentation.

The sequence of discovered melodic segments, that have been labelled according to the category they belong to, can form a new sequence of entities (e.g. motives) which can be fed back to the initial stages of the model (just before *LBDM* in figure 3.1) and organised into higher-level categories (e.g. sub-themes, themes etc.). The 'interval' or distance between consecutive segments can be measured in relation to their category membership, i.e. melodic segments of a category that exists for lower similarity thresholds may be considered more similar than ones from a category that ceases to exist for higher thresholds; the *LBDM* can thus be applied on the motive 'interval' profile. Pattern-matching can then be applied to the sequence of labelled melodic segments and a higher-order segmentation reached. Categorisation can finally proceed for the new higher-order segments and for higher-level attributes such as inclusion of smaller labelled melodic segments at different positions, higher-order reductions, tonal regions, note densities, etc. This way *GCTMS* penetrates deeper into musical structure and generates higher-level structural descriptions.

### 9.1.4 Manually performed tasks

Some parts of the overall model have not as yet been fully implemented on the computer. These are: the process for selecting the 'best' segmentation or set of segmentations, the link between the segmentation and the categorisation module (i.e. the melodic segments along with their attributes are fed manually into the *Unscramble* algorithm), the selection mechanism of the *Unscramble* algorithm, the construction of reduced versions of a melodic surface and, finally, the application of the overall model on higher-level sequences of discovered motivic categories. Implementing most of these as part of the prototype computer system should be a rather straightforward procedure; the first task, however, requires further design decisions to be made (see next paragraph). The fact that these parts of the model have not been fully implemented on the computer at this stage (this is due to constraints on the time length of this research study) may introduce slight unintentional biases towards a preferred analysis although care has been taken to avoid this as much as is possible.

The Total Boundary strength profile obtained by the processes described in section 9.1.2 indicates the points in a melodic surface that are most likely to be perceived as segmentation boundaries. However, further research is necessary for the construction of a mechanism that will determine exactly which peaks should be considered significant and should be included in the 'best' segmentation. In the examples provided in the next section an arbitrary threshold selection mechanism has been devised: if x,y,z are successive strength values and $0 \leq x,y,z \leq 100$, then y is a significant peak if $y \geq 50$ and $y > 4/3 \cdot x$ & $y > 4/3 \cdot z$ (i.e. greater by 1/3 than both of its neighbouring strength values) - two exceptions in the examples below will be discussed at the appropriate point. Instead of setting a fixed threshold it would probably be more appropriate to select a range of thresholds which would give rise to a number of segmentations and a number of corresponding paradigmatic descriptions; the selection of the 'best' paradigmatic description would help determine the 'best' segmentation.

The reduced version of a melodic surface is also constructed manually. In the analytic examples below, the simple surface reductions consist of metrically stronger notes - usually notes on the beat level (repeated notes are merged when no strong boundary appears between them). The reduced version of each surface is slightly different; for instance, in the example in section 9.2.2 a reduction that consists of notes whose inceptions coincide with the points of the beat level is identical to the surface itself - in this case the reduction consists of notes that coincide with points on the next level up 3/4 metrical grid. In future a mechanism that automatically constructs reductions should be devised; the *SPIA* and *Unscramble* algorithms should also be applied on more than one reduction of the melodic surface.

The description of the way the proposed overall computational model proceeds from the melodic surface (0) towards a structural analysis can probably best be presented through the detailed exposition of a number of specific melodic analyses.

## 9.2 Four melodic analyses

Four melodies have been selected from diverse musical idioms and the same prototype computer system with the same default values has been applied to each of them (any

diversion from the above description will be given at the appropriate point of the analysis of the specific melody). This way the generality of the theory can be tested: if the same analytic mechanism gives rise to 'reasonable' results for very different melodic surfaces then we have an indication that our assumption of style-independence of *GCTMS* may be substantiated. Of course, four analyses are hardly enough; however, it is hoped that the exposition given below will highlight the potential of the system, reinforce the validity of the overall theory and encourage further experimentation and testing.

### 9.2.1 The Finale Theme of Beethoven's *9th Symphony*

At a first glance, the finale theme of Beethoven's *9th Symphony* seems to be almost trivial in terms of providing an analytic description for it - e.g. Lerdahl and Jackendoff (1983:124-126) take the analysis of the melody in terms of phrases and sub-phrases for granted as naturally emerging from musical parallelism and then proceed with other aspects of the analysis. But the description of this melodic surface is actually a rather difficult task when analysis is to be pursued computationally - especially if the surface is presented as a mere unstructured sequence of notes. The reason is that parallelism plays an important role in its description in terms of both segmentation and categorisation. If segmentation is externally provided to a computational system then further analysis is not very complicated; but if the system is also expected to provide a segmentation of the surface then the analysis becomes more complex. An attempt to analyse this melodic surface by applying the computer system based on the *General Computational Theory of Musical Structure* is presented below.

The melodic surface (0) - i.e. [MIDI_pitch, duration]: [54, 1/2], [55, 1/4], [57, 1/4], [57, 1/4], [55, 1/4], ...) - can readily be converted into the traditional pitch notation for the major-minor scale framework with the use of the *General Pitch Interval Representation* transcription algorithm. There are no mistakes in the transcription as far as the use of accidentals is concerned. The melodic surface (1) is represented by pitch and time intervals.

The next step is to apply the refined *Local Boundary Detection Model (LBDM)* on the surface in order to detect points that are most likely to form low-level grouping boundaries. The system is expected to perform poorly as the melodic surface does not provide strong

local grouping cues in terms of pitch or rhythmic changes (see Local Boundary strength profile in figure 9.1).

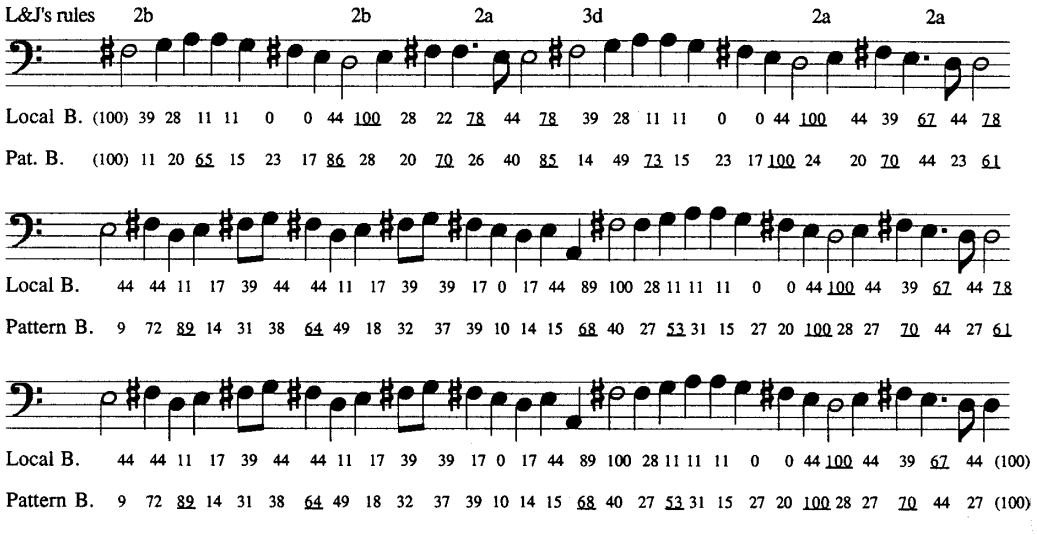



Figure 9.1 Local Boundary strength profile and Pattern Boundary strength profile for the finale theme of Beethoven's *9th Symphony*

The local accents for each note may easily be calculated from the grouping boundary strengths. If metrical grids are now matched to the local accents, the lowest metrical grid on the quarter-note level (beat level) becomes apparent as well as the 2/4 metrical grid every two beats that starts on the first note (Table 9.1). At the next level up - i.e. every four beats - the two competing grids with offsets 0 and 2/4 receive values which are not significantly different. The accent profile of the melody may be enhanced if extra numerical values are added to the notes whose inceptions coincide with the points of the 2/4 metrical grid starting on the first beat.

| Metric Grid | 2/8 | | 3/8 | | | 2/4 | | 3/4 | | | 4/4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/8 | 0 | 1/8 | 2/8 | 0 | 1/4 | 0 | 1/4 | 2/4 | 0 | 1/4 | 2/4 | 3/4 |
| Total Value | 1428 | 189 | 802 | 761 | 862 | 1918 | 938 | 1307 | 1582 | 1394 | 1778 | 646 | 2057 | 1229 |

Table 9.1 Metric grids matched onto the accent strength profile of the melody. Preferred metric grids are underlined. There is strong preference for a metric grid matched on the quarter-note beat level and the 2/4 metrical level.

It is clear that this low-level information is insufficient for a general description of the grouping and metrical structure of the melody although some interesting aspects have indeed been highlighted such as low-level metrical structure - in contrast, Lerdahl and Jackendoff's local detail grouping rules provide hardly any useful information at all without the assistance of the slur rule (see top line of figure 9.1).

At this point musical parallelism should prove itself indispensable for the breaking down of the melody into 'significant' units. Initially, the *String Pattern-Induction Algorithm (SPIA)* and *Selection Function* will be applied to the melodic surface for patterns that consist of 3-4 notes. The algorithm is applied to the scale-step interval and the duration parametric profiles. This should capture a rather low-level grouping structure - if one exists - that is co-extensive with the discovered metrical structure. Since the previously discovered metrical structure was the 2/4 metrical grid a co-extensive structure should have group boundaries every 2 beats (2/4) or multiples thereof (e.g. 4/4). The Pattern Boundary strength profile is depicted in figure 9.1. It is clear that there is a strong preference for a grouping structure that is co-extensive with a 4/4 metrical grid starting on the 1st beat (see Table 9.2).

| Metric Grid | 2/4 | | 3/4 | | | 4/4 | | | |
|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/4 | 0 | 1/4 | 2/4 | *0* | 1/4 | *2/4* | 3/4 |
| Total Value | 1024 | 530 | 738 | 768 | 826 | *1468* | 232 | *581* | 829 |

Table 9.2 Metric grids matched onto the pattern boundary strength profile of the melody. Preferred metric grids are underlined. There is a strong preference for a regular grouping structure starting on the first beat and extending over 4/4 time spans. This strongly suggests a coextensive 4/4 metrical structure (as is indicated in the score).

An alternative methodology would be to apply *SPIA* exhaustively (from smallest to largest patterns) only to the first section of the melody that is quite short. This is not an unreasonable approach as it is plausible that intensive processing takes place at the

beginning of a musical piece in order to establish tentative descriptions which can be used for further analytic processing. For the first section of the melody there is a significant preference for the four-beat metrical grid (4/4) starting on the first beat. In figure 9.2 the most prominent patterns selected by the *SPIA & Selection Function* are depicted. For this first section of the melody the grouping and metrical structures are in-phase. It is a plausible assumption that the metrical structure will pertain throughout the melody and that the grouping structure will continue to be in-phase unless there are strong cues against this initial assumption. The patterns discovered in the first section, which recur in the rest of the melodic surface, also reinforce this assumption.

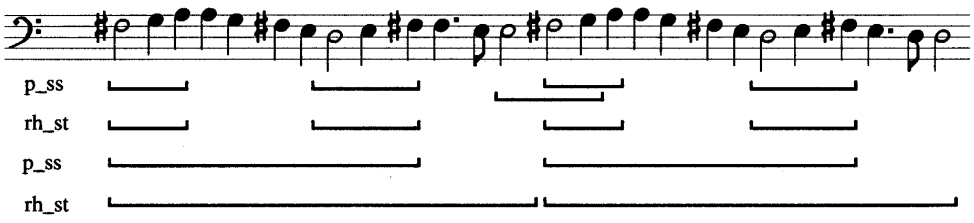



Figure 9.2 The most prominent patterns discovered by the *SPIA* for the scale-step pitch interval (pss) and the duration (rdur) profiles are depicted below. For the smaller patterns the most important factor in the *Selection Function* is the frequency of occurrence whereas for the longer ones the most important factor is pattern length. Overlapping is allowed but strongly discouraged.

The regular four-beat grouping structure highlighted above allows the melodic surface to be segmented as is shown in figure 9.3. An ambiguous position in the segmentation is the 'syncopated' half-note in the middle of the second and third sections of the melody (this note extends over a strong metrical position). This ambiguity is easily resolved if a further rule is added: *in cases where there is no clear boundary between repeating notes these notes may be merged together* (this is a sort of low-level prolongational principle). In the theme under examination this assumption would place a grouping boundary before the ambiguous note (this note is essentially an anticipation) - see also (Lerdahl and Jackendoff, 1983:126) for a different approach to this problem. Alternatively, this ambiguity may be left unresolved at this stage (a boundary may occur before or after the note); between the two alternative segmentations the one may be preferred that provides a 'better' classificatory description of the surface (see below).

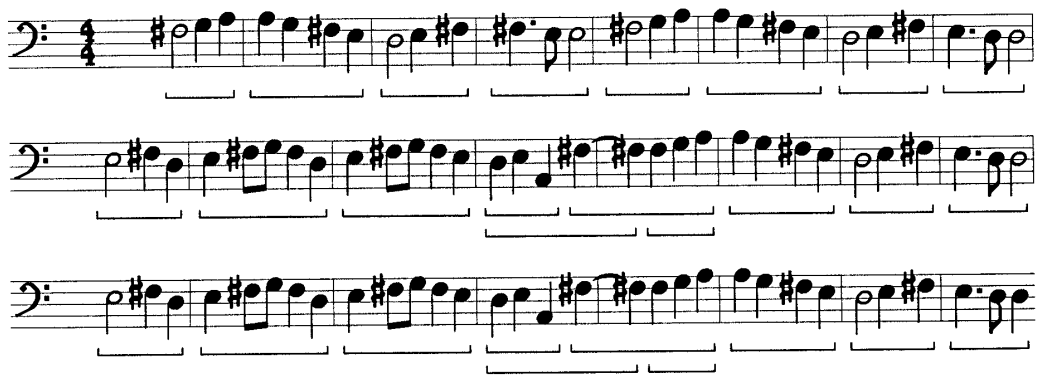


Figure 9.3 Segmentation of the finale theme of Beethoven's *9th Symphony*

If exact repetitions are omitted then the musical segments depicted in figure 9.4 may be presented to the *Unscramble* algorithm for categorisation (the alternative segmentation over the syncopated note is ignored at present).

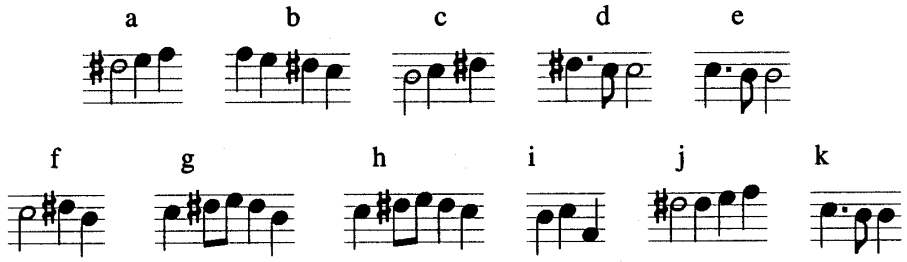



Figure 9.4 The musical segments above are given to the *Unscramble* algorithm for categorisation.

There are a large number of possible attributes that can be assigned to each of these segments for the musical surface or reductions of it. For the sake of brevity only a limited set of twelve attributes has been selected as described in section 9.1.3 - e.g. melodic_segment(*a*, [pex1, pss1, pcon1, rdur1, rrel1, pex_rd1, pss_rd1, pcon_rd1, rdur_rd1, rrel_rd1, pfl, preg]).[1] The reduction of the surface consists of notes whose inception coincides with the beat level positions and repeated notes are merged. As a rough first estimate all the rhythm related attributes are given half the weight of the pitch related ones.

[1] Abbreviations: p:pitch, r:rhythm, ex:exact, ss:scale-step, con:contour, rel:relative(i.e. shorter, longer or equal), rd:reduced (surface consisting of metrically stronger notes), fl:first-last (pitch interval between first and last note of segment) and reg:register (low or high).

The *Unscramble* algorithm gives various categorisations of these melodic segments for the various possible thresholds.

| | | | |
|---|---|---|---|
| Threshold: | 8.0 | Categories: | [[a,b,c,d,e,f,g,h,i,j,k]] |
| Threshold: | 7.75 | Categories: | [[a,c,f,g,h,j],[b],[d,e,k],[f,g,h,i]] |
| Threshold: | 7.5 | Categories: | [[a,c,f,g,h],[a,c,j],[b],[d,e,k],[f,g,h,i]] |
| Threshold: | 7.0 | Categories: | [[a,c,f],[a,c,j],[b],[d,e,k],[f,g,h,i]] |
| Threshold: | 6.5 | Categories: | [[a,c,j],[b],[d,e,k],[f,g,h],[f,i]] |
| Threshold: | 6.0 | Categories: | [[a,c,j],[b],[d,e,k],[f,g],[f,i],[g,h]] |
| Threshold: | 5.0 | Categories: | [[a,c,j],[b],[d,e,k],[f,g],[g,h],[i]] |
| Threshold: | 4.75 | Categories: | [[a,c,j],[b],[d,e,k],[f,g],[h],[i]] |
| Threshold: | 3.75 | Categories: | [[a,c],[a,j],[b],[d,e,k],[f,g],[h],[i]] |
| Threshold: | 3.5 | Categories: | [[a,c],[b],[d,e,k],[f,g],[h],[i],[j]] |
| Threshold: | 2.0 | Categories: | [[a,c],[b],[d,e,k],[f],[g],[h],[i],[j]] |
| Threshold: | 0.75 | Categories: | [[a],[b],[c],[d,e,k],[f],[g],[h],[i],[j]] |
| Threshold: | 0.0 | Categories: | [[a],[b],[c],[d,e],[f],[g],[h],[i],[j],[k]] |

From these, the one that rates best, in terms of least overlapping, least number of monadic categories and is not one single category, is the description for threshold h=6.5 (see figure 9.5). If no overlapping is allowed, then the description for h=4.75 is preferred.

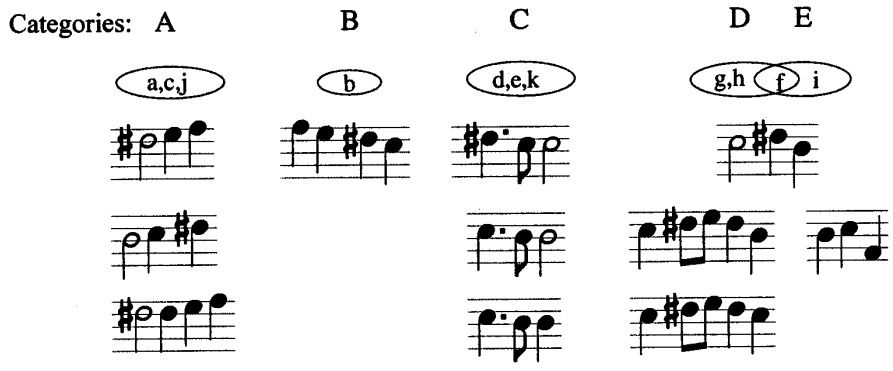



Figure 9.5 The 'best' classificatory description revealed by the *Unscramble* algorithm.

For this categorisation the new weights for each attribute of each melodic segment may now be calculated. For instance, for category A we get:

category: [a,c,j]

weights: [[pex1,0.33],[pex3,0.33],[pex9,0.33],[pss1,0.66],[pss8,0.33],[pcon1,0.66],
[pcon6,0.33],[rdur1,0.54],[rdur6,0.33],[rrel1,0.54],[rrel6,0.33],
[pex_rd_ms1,0.66],[pex_rd3,0.33],[pss_rd1,1.0],[pcon_rd1,1.0],
[rex_rd1,0.29],[rex_rd5,0.33],[rrel_rd1,0.62],[pfl,1.0]]

The most characteristic (in essence, defining) attributes of category A are: the scale-step and contour pitch interval pattern for the reduced surface and the first-last note interval. The pitch register is non-diagnostic and is dropped altogether (all segments have similar average pitch register).

It is also interesting to note how the description of the segments {f,g,h,i} evolves as the similarity thresholds are gradually lowered (figure 9.6). In the beginning all these segments are placed in the same category, then they are placed in two overlapping categories and so on till segments *h* and *i* become monadic categories and {f,g} form one category. If more weight was given initially to the properties relating to rhythmic aspects of the segments then *f* and *i* would be monadic categories and {g,h} in the same category.

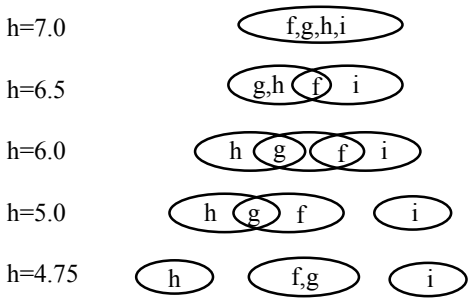



Figure 9.6 The evolution of the categorisation description of musical segments {f,g,h,i} for different thresholds.

If the alternative segments over the syncopated note are included in the initial set of melodic segments to be categorised by the *Unscramble* algorithm, then it is found that the first alternative segment that ends on the long syncopated note becomes a monadic category much faster (i.e. for a higher threshold) and thus leads to lower quality categorical descriptions. This supports the initial segmentation where the syncopated note is the beginning of the next group.

The musical surface can now be described as a sequence of melodic motifs taken from categories A, B, C, D & E (figure 9.7, bottom row). If the *SPIA* and *Selection Function* are

applied to this sequence of musical motifs then further higher level descriptions of the melodic surface emerge (figure 9.7, top row).


Figure 9.7 Organisation of sequence of motives into higher-level (phrase) categories.

### 9.2.2 L'Homme Armé

*L'homme armé* is a 15th century melody on which many polyphonic works, especially masses, have been based - here the version presented in (Lockwood, 1980:712) will be


Figure 9.8 Local Boundaries strength profile, Pattern Boundary strength profile and a weighed Total Boundary strength profile for the melody *L'homme armé.*

analysed. Although this melody has a very clear structure it can be seen as a challenge for an automated analytic system mainly because of its irregular lengths of phrases and subphrases (it does not fit into a symmetric tree-like structure of equal length melodic segments). Below is a description of how the proposed computational model provides a rudimentary structural analysis of the melodic surface (0) of *L'homme armé* (figure 9.8).

The local boundary strengths are calculated with the application of *LBDM* (figure 9.8) and the local accent values are then automatically derived. From the various metric grids the best match occurs for the 3/4 metric grid starting on the first beat (table 9.3).

| Metric Grid | 2/4 | | 3/4 | | | 4/4 | | | | 6/4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/4 | 0 | 1/4 | 2/4 | 0 | 1/4 | 2/4 | 3/4 | 0 | 1/4 | 2/4 | 3/4 | 4/4 | 5/4 |
| Total Value | 988 | 1236 | 1939 | 516 | 880 | 934 | 1427 | 1041 | 1045 | 1731 | 525 | 724 | 2146 | 507 | 1036 |

Table 9.3

At this point the *String Pattern-Induction Algorithm (SPIA)* and *Selection Function* will be applied to the melodic surface for patterns that consist of 3-4 notes. The algorithm is applied to the scale-step interval and the duration parametric profiles. Since the previously discovered metrical structure was the 3/4 metrical grid a coextensive structure should have group boundaries every 3 beats (3/4) or multiples thereof (e.g. 6/4). There is a preference for a grouping structure that is co-extensive to the 3/4 metrical grid starting on the 1st beat; however, there is clearly no single preference amongst the alternative offset positions of the 6/4 metrical grid which indicates that there is no metrical structure and corresponding co-extensive grouping structure above the 3/4 metrical grid (see table 9.4).

| Metric Grid | 2/4 | | 3/4 | | | 6/4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/4 | 0 | 1/4 | 2/4 | *0* | 1/4 | 2/4 | *3/4* | 4/6 | 5/6 |
| Total Value | 582 | 594 | 968 | 217 | 579 | *954* | 216 | 609 | *981* | 218 | 549 |

Table 9.4

The grouping structure that is co-extensive with the 3/4 metrical grid starting on the first beat gives rise to a grouping structure in which most groups consist of only one note or two repeating notes. As such a segmentation may be considered uninteresting the *SPIA* may be applied for a more extended range of pattern lengths (for the melodic surface and a reduction of it) aiming at a higher-level possibly irregular grouping description.

The *SPIA* is now applied on the melodic surface for patterns consisting of 3-7 notes and on a reduction of the surface consisting of notes on strong 3/4 metrical positions and notes bordering strong local boundaries for patterns consisting of 3-5 notes. The resulting Pattern Boundary strength profile is depicted in figure 9.8 (along with the Total Boundary strength profile). In addition to the pattern boundary strength maxima that coincide with the strong local boundaries, there is a significant pattern boundary maximum at the point indicated by

an asterisk in figure 9.8 due to parallelism. This boundary could become even more prominent if there was special provision in the parallelism model for immediate repetitions - see (Monelle, 1992) for a brief discussion of this significant type of repetition. The boundary appearing at the point indicated by a cross in figure 9.8 which is very close to the selection threshold may be diminished for the same reason (it is actually disregarded in this example). If more stress is given to the rhythmic aspect of the melody then segment *h* becomes also highlighted (figure 9.9). Obviously, there are many different possible segmentations.

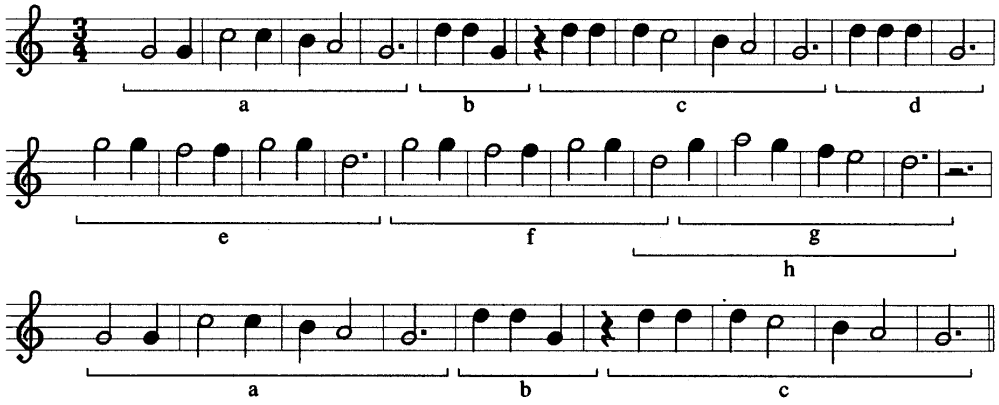


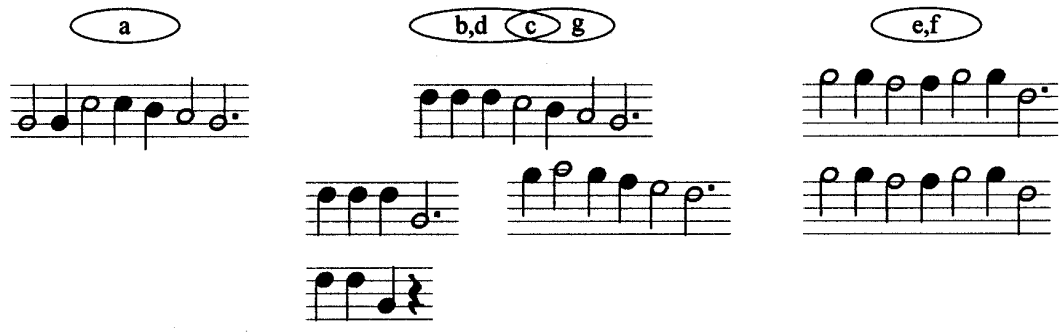

Figure 9.9 Patterns highlighted by the application of the *SPIA* and *Selection Function*.

Let us assume that the grouping description depicted in figure 9.9 provides the musical segments that are to be classified by the *Unscramble* algorithm. Let us also assume that each of these segments is represented by the set of attributes that was described in section 9.1.3. If these segments are presented to the *Unscramble* algorithm the following classificatory description emerges:

| | | | |
|---|---|---|---|
| Threshold: | 10 | Categories: | [[a,b,c,d,e,f,g,h]] |
| Threshold: | 9 | Categories: | [[a,b,c,d],[a,h],[b,c,d,e,f,g],[e,f,g,h]] |
| Threshold: | 7 | Categories: | [[a,h],[b,c,d],[c,g],[e,f]] |
| Threshold: | 6 | Categories: | [[a,h],[b,d],[c],[e,f],[g]] |
| Threshold: | 5 | Categories: | [[a],[b,d],[c],[e,f],[g],[h]] |
| Threshold: | 4 | Categories: | [[a],[b],[c],[d],[e,f],[g],[h]] |

Perhaps the best categorical description is the one for threshold h=7 (only one overlap). For the two alternative segmentations (one with segment *g* and one with *h)* the resulting categorisations are presented in figure 9.10.

Categorisation for melodic segments {a,b,c,d,e,f,**g**}:




Categorisation for melodic segments {a,b,c,d,e,f,**h**}:

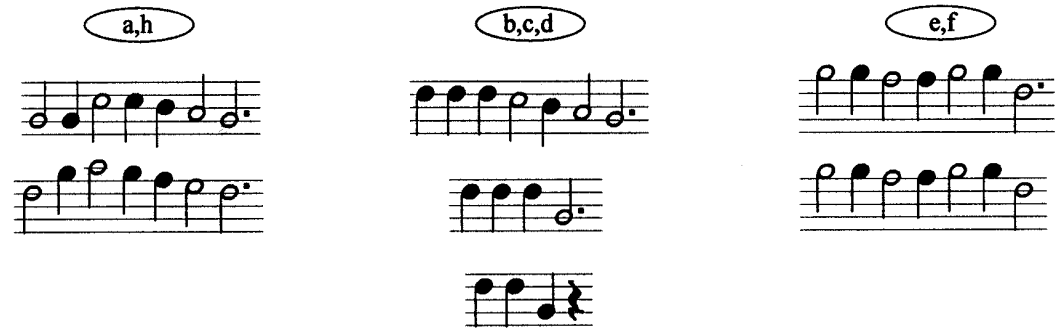



Figure 9.10 The 'best' classificatory descriptions revealed by the *Unscramble* algorithm for the two alternative segmentations.

For each of these categories the most characteristic attributes are the ones with the highest weights; for instance, for category {a,h} the most characteristic attributes are the shared rhythmic pattern, the pitch contour of the reduced surface and the exact pitch interval between the first and last note of the segments:

Category: [a,h]
Weights: [[rdur1,1],[rrel1,1],[pex1,0.5],[pex8,0.5],[pss1,0.5],[pss8,0.5],
[pcon1,0.5],[pcon8,0.5],[pex_rd1,0.5],[pex_rd5,0.5],
[pss_rd1,0.5],[pss_rd5,0.5],[pcon_rd1,1],[pfl1,1]]

If one of these categorisations for a specific threshold is selected and the *Unscramble* algorithm is applied yet again for the newly calculated weights, a whole range of more elaborate classifications arise. Obviously, more or less complex classificatory descriptions than the ones presented above can be constructed if attributes are given different initial weights or a different set of attributes is initially selected.

It is clear that the strongest categories are the ones that hold for the lowest thresholds and these would be for instance: {b,d} & {e,f}. As the patterns *a, b, c* appear twice in the segmented score a categorisation for a lower threshold may be accepted as these single patterns may not be considered monadic categories - for instance, for threshold h=5 the only monadic category is {g} or {h} (only one of the two may be accepted in a single categorisation unless heavy overlapping is allowed). If this categorisation is selected then the musical surface can be described as a sequence of melodic motifs taken from categories A={a}, B={b,d}, C={c}, D={e,f} & E={g/h} (figure 9.11, bottom row). If the *LBDM* and the *SPIA & Selection Function* is applied to this sequence of musical motifs then further higher level descriptions of the melodic surface emerge (figure 9.11, top row).

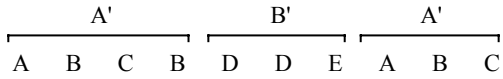



Figure 9.11 Organisation of sequence of motives into higher-level (phrase) categories.

### 9.2.3 A melody from Webern's *Lieder Op. 3*

The atonal melody of the first song from Webern's *Fünf Lieder Op. 3* will be analysed with the use of the computational model based on the *GCTMS* (cf. brief analysis in Wason, 1996:113-117). It will be shown that the proposed general model can provide a sound description of the atonal melody.

The piano part that accompanies the melody is not considered. Although this is a gross simplification in many respects, it is possible to describe the work as consisting of two distinct streams because of different timbral characteristics (i.e. voice and piano sounds). The atonal melody (figure 9.12) is presented to the computational model as a melodic surface (0). Pitch and pitch-intervals are encoded in the 12-tone *GPIR* representation (rather than the *GPIR* diatonic representation).

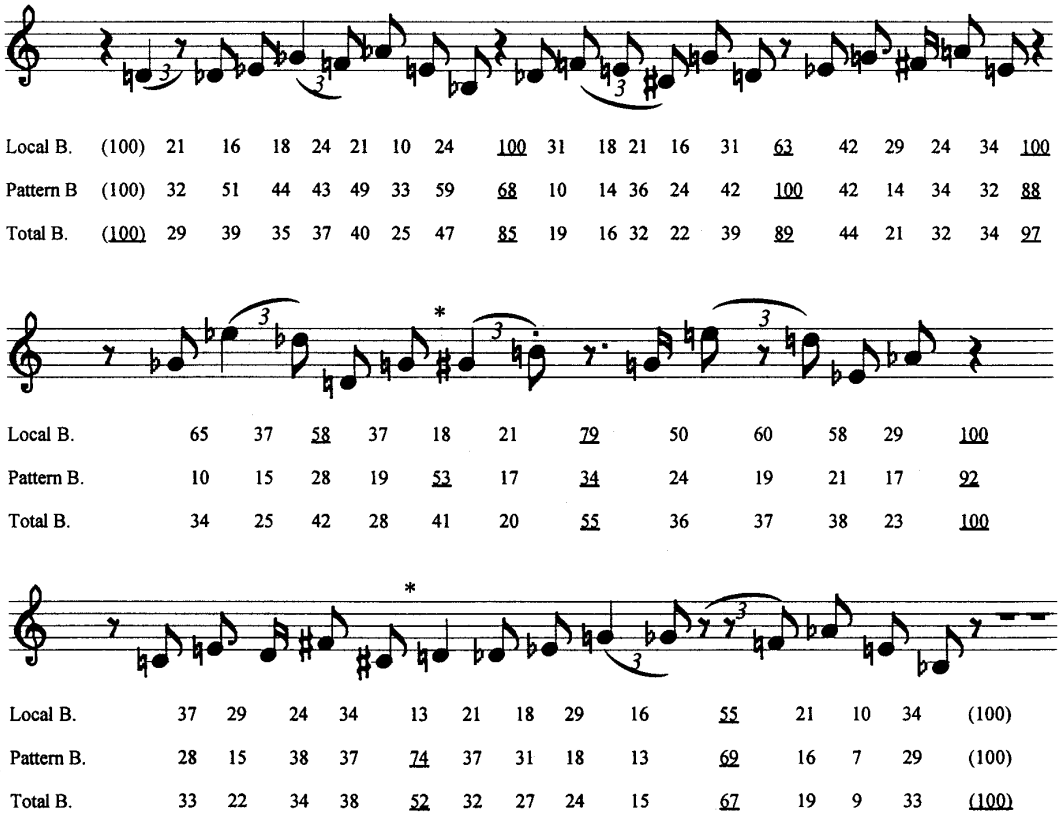



Figure 9.12 Local boundary strength profile, Pattern boundary strength profile and a weighed Total boundary strength profile for Webern's melody.

The Local Boundary strengths are calculated with the application of *LBDM* (figure 9.12) and the local accent values are then automatically derived. From the various low-level metric grids the best match occurs for the eighth-note metric grid starting on the first beat (according to the indicated tempo this is a possible beat level); there is also a weak preference for the quarter-note metric grid starting on the first beat - see Table 9.5. At higher levels there is no strong preference over any specific metric grid; this is consistent with the fact that the composer indicates multiple time signatures through the course of the song.

| Metric Grid | 2/16 | | 3/16 | | | 2/8 | | 3/8 | | | 2/4 | | 3/4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/16 | 0 | 1/16 | 2/16 | 0 | 1/8 | 0 | 1/8 | 2/8 | 0 | 1/4 | 0 | 1/4 | 2/4 |
| Total Value | 344 | 38 | 162 | 186 | 225 | 398 | 289 | 324 | 374 | 333 | 389 | 408 | 334 | 434 | 428 |

Table 9.5

If the *String Pattern-Induction Algorithm (SPIA)* and *Selection Function* are applied to the melodic surface for patterns that consist of 3-4 notes (the algorithm is applied to the scale-step interval and the duration parametric profiles) then a low-level regular grouping structure may be revealed; it is clear from table 9.6 that no preference for any regular grouping structure exists.

| Metric Grid | 2/8 | | 3/8 | | | 2/4 | | 3/4 | | | 4/4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/8 | 0 | 1/8 | 2/8 | 0 | 1/4 | 0 | 1/4 | 2/4 | 0 | 1/4 | 2/4 | 3/4 |
| Total Value | 257 | 152 | 158 | 233 | 222 | 256 | 258 | 175 | 321 | 274 | 266 | 289 | 245 | 227 |

Table 9.6

As there is no clear metrical structure above the quarter-note level it is likely that low-level grouping processes may play an important role in the perception of possible boundaries. As can be seen, especially in the Local Boundary strength profile of the refined version of the *LBDM* (figure 9.12), some boundaries are very strongly marked out. 'Hard' break markers are inserted in these strong boundary positions of the melodic surface. Then the *SPIA* algorithm may be applied for longer sequences of notes more efficiently as the search space is significantly reduced.

The *SPIA* is now applied on the melodic surface for patterns consisting of 3-7 notes (exact and near-exact pitch interval, contour, duration and relative duration matching) and on a reduction of the surface (for 3-5 note patterns) consisting of notes whose inceptions coincide with points of the eighth-note beat level (only exact and near-exact pitch interval matching) - see next paragraph for near-exact matching. The resulting Pattern Boundary strength profile is depicted in figure 9.12. In addition to the pattern boundary strength maxima that coincide with the strong local boundaries, there are significant pattern boundary maxima at the points indicated by asterisks in figure 9.12 due to parallelism.

A special pattern matching process (*near-exact* matching) - similar in some ways to matching scale-step intervals in diatonic surfaces - has been applied to the atonal pitch profile of the melody, i.e. two patterns are matched if each of their corresponding intervals is identical or differs by 1 semitone (see figure 9.13 and patterns in figure 9.14 labelled with infix or suffix 'nex').

Figure 9.13 Near-exact matching: these two patterns match because each of their corresponding intervals is identical or differs by 1 semitone; they do not match either for exact pitch intervals (in semitones) or for diatonic scale-step intervals (the latter actually is arbitrary for a 12-tone pitch structure).

The strongest peaks in the Total Boundary strength profile give a quite plausible segmentation of the musical surface as shown in figure 9.15 - segments *a1, a2, a3* and *d1* have been added manually as they appear at least once as significant independent nearly-identical melodic patterns elsewhere in the melody.

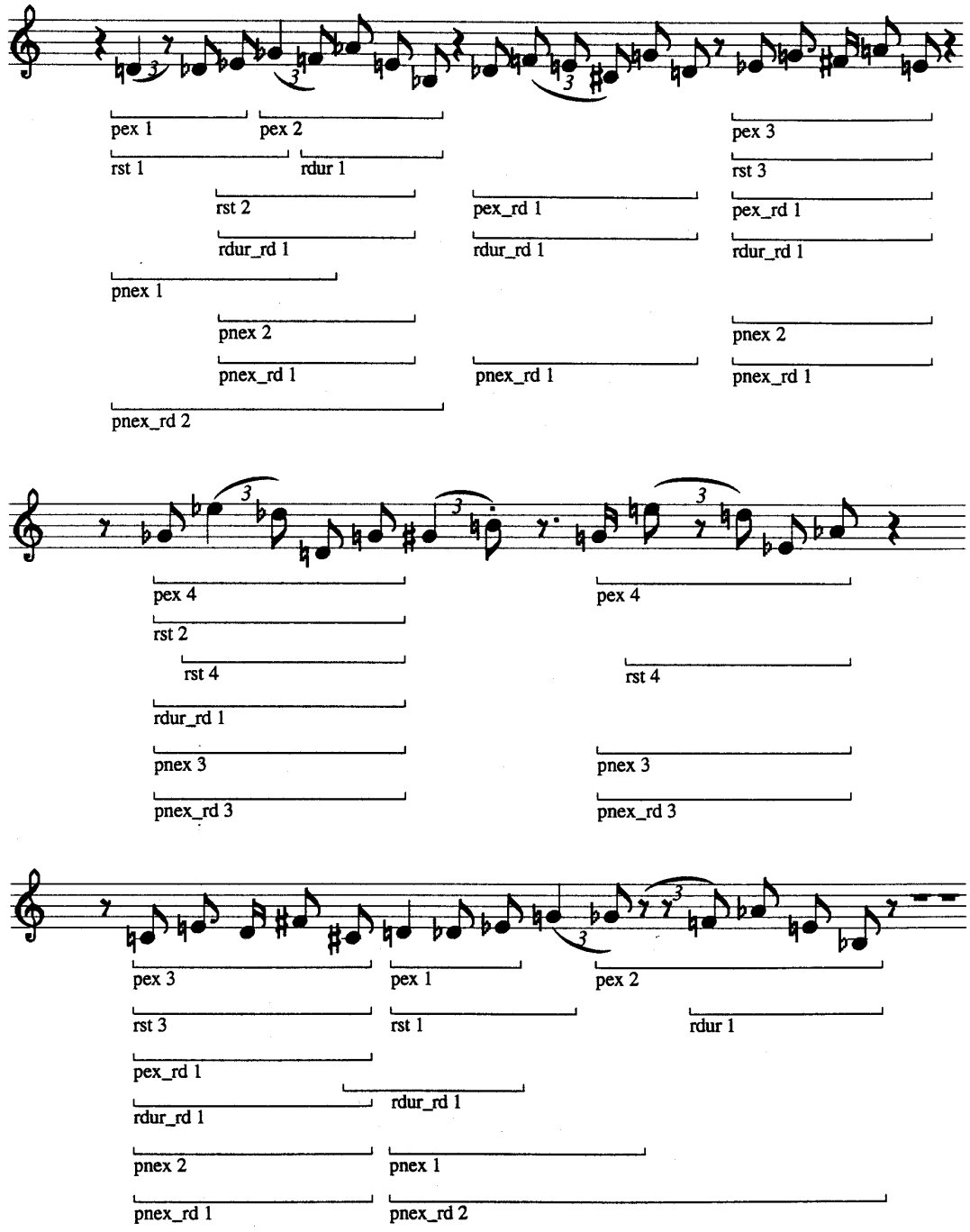



Figure 9.14 Some of the most prominent patterns highlighted by the *SPIA* and *Selection Function* (abbreviations: p:pitch, r:rhythm, ex:exact, st:start-time, dur:duration, nex:near-exact, rd:reduced)

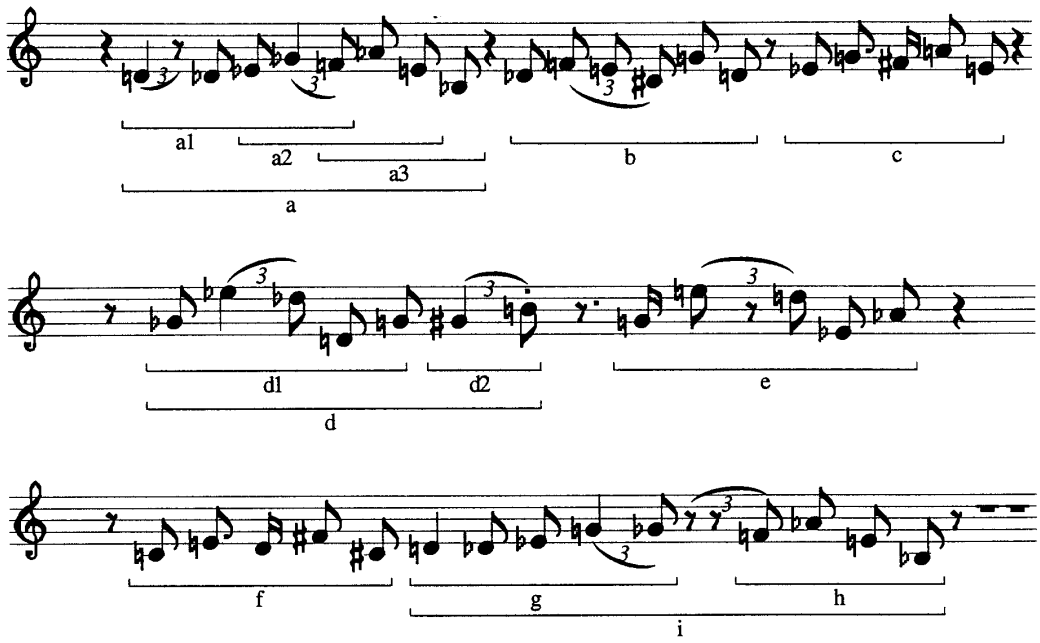



Figure 9.15 Preferred segmentation of atonal melody.

From the above complex segmentation it is possible to select one segmentation that consists of all the smaller patterns {a1, a2, a3, b, c, d1, d2, e, f, g, h}. If these melodic segments (represented by the attributes described in section 9.1.3) are presented to the *Unscramble* algorithm they can be organised into categories/paradigms. The algorithm gives the following categorisation descriptions:

Threshold: 9.0 Categories: [[a1,a2,a3,b,c,d1,d2,e,f,g,h]]
Threshold: 8.0 Categories: [[a1,a2,a3,b,c,f,g,h],[a1,a2,b,d1,d2,e,f,g]]
Threshold: 7.75 Categories: [[a1,a2,b,c,f,g],[a2,b,c,d1,e,f],[a3,h],[d1,d2,e]]
Threshold: 7.5 Categories: [[a1,a2,b,c,f,g],[a2,b,c,d1,f],[a2,c,d1,e,f],[a3,h],[d1,d2,e]]
Threshold: 7.25 Categories: [[a1,a2,b,c,f,g],[a2,c,d1,f],[a2,d1,e],[a3,h],[d1,d2,e]]
Threshold: 7.0 Categories: [[a1,a2,b,c,f,g],[a2,d1],[a3,h],[d1,d2,e]]
Threshold: 4.5 Categories: [[a1,g],[a2,b,c,f],[a3,h],[d1,e],[d2]]
Threshold: 3.5 Categories: [[a1,g],[a2,c,f],[a3,h],[b,c,f],[d1,e],[d2]]
Threshold: 2.25 Categories: [[a1,g],[a2,c,f],[a3,h],[b],[d1,e],[d2]]
Threshold: 2.0 Categories: [[a1,g],[a2],[a3,h],[b],[c,f],[d1,e],[d2]]
Threshold: 0.25 Categories: [[a1],[a2],[a3,h],[b],[c,f],[d1,e],[d2],[g]]
Threshold: 0.0 Categories: [[a1],[a2],[a3,h],[b],[c,f],[d1],[d2],[e],[g]]

From these the 'best' description occurs for Threshold=4.5 (because of least number of monadic categories, no overlapping and not too general) - see figure 9.16. This classificatory description seems to reflect the structure of the melody rather well.

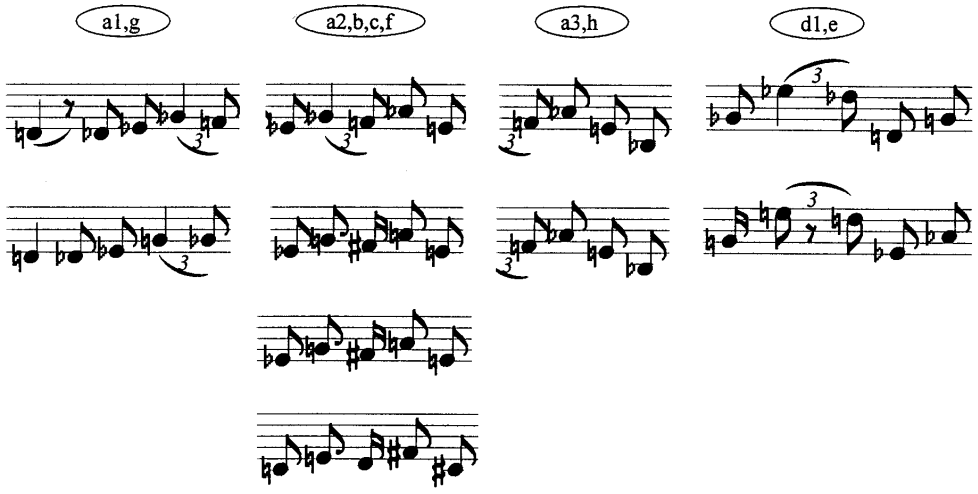



Figure 9.16 Melodic segments organised into categories by the *Unscramble* algorithm (segment *d2* is a monadic category as is not depicted in the figure).

Since segments *a1 & g* and *a3 & h* are nearly identical and they occur in immediate succession in the score (i.e. *a1*→*a3* and *g*→*h)* it is possible to combine *g & h* into a single larger segment *i* - see figure 9.15 (this procedure has been implemented manually). The *Unscramble* algorithm may then be applied on the larger segments {a,b,c,d,e,f,i} for attributes that relate to the previous classification, e.g. the sub-segment categories that each contains. This way the categorisation depicted in figure 9.17 is arrived at (an overlapping description is another acceptable possibility: [[a,i],[a,b,c,f],[d,e]]).

For each of the above categories the properties that are most diagnostic receive higher weights whereas other properties shared between more than one categories are attenuated.

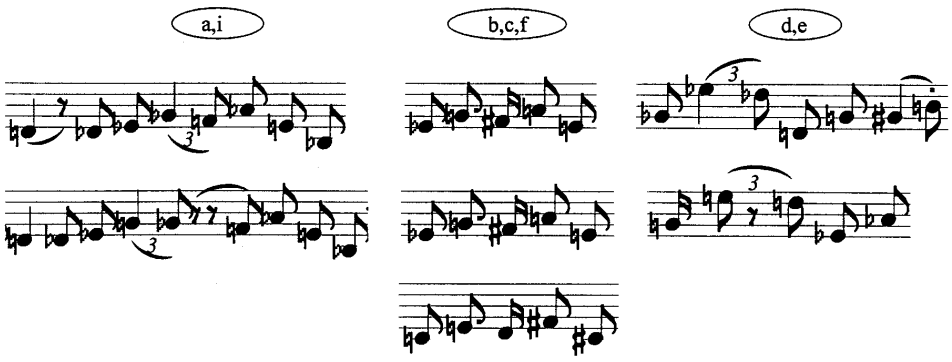



Figure 9.17 Melodic segments organised into categories by the *Unscramble* algorithm

If the *Unscramble* algorithm is applied to the shorter segments where each segment is represented only by its pitch-class set as defined by pitch-class set theory then the following description is generated:

Categories: [[a1],[a2],[a3,h],[b,c,f],[d1,e,g],[d2]]

This description is counter-intuitive as it ignores the many other aspects of a musical work that directly influence the way we perceive it (e.g. rhythm, ordered pitch sequences, reductions etc.). For instance, segments *d1* and *g* are equivalent under pitch-class set theory (same *prime form)* whereas *a1* and *g* are weakly similar (weak similarity relation $R_p$) - see figure 9.18. Pitch-class set theory gives strong context-independent definitions of what is equivalent and what similar, and is thus bound to give unintuitive results in many occasions. Obviously, pitch-class sets can be one amongst many other properties that describe musical segments in which case it may or may not affect the categorisation process.

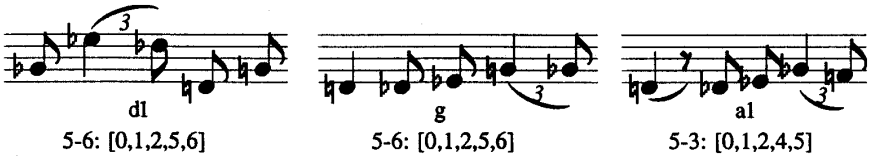



Figure 9.18

### 9.2.4 A melody from Babbit's *Du* song cycle

The voice line of the first song from Babbit's *Du* song cycle is composed of an all-interval twelve-tone series followed by its retrograde inversion form. An analysis of the beginning of this song cycle in terms of the serial technique employed by the composer is presented in (Cogan and Escot, 1976:207-212). However, following a number of experiments, Francès (1988) suggests that a listener hardly recognises the links between a series and its derivative forms and that the serial technique has a very diminished role in the perception of atonal music.[2] The question that arises is how a cognitively-based model may perform when applied to such a serial melody. It will be shown that the model based on the *General Computational Theory of Musical Structure* - applied for the same default values that were used in the previous examples - provides a sound segmentation of Babbit's melody and, even, arranges the segments into an acceptable paradigmatic description (with some weaknesses that will be discussed below).

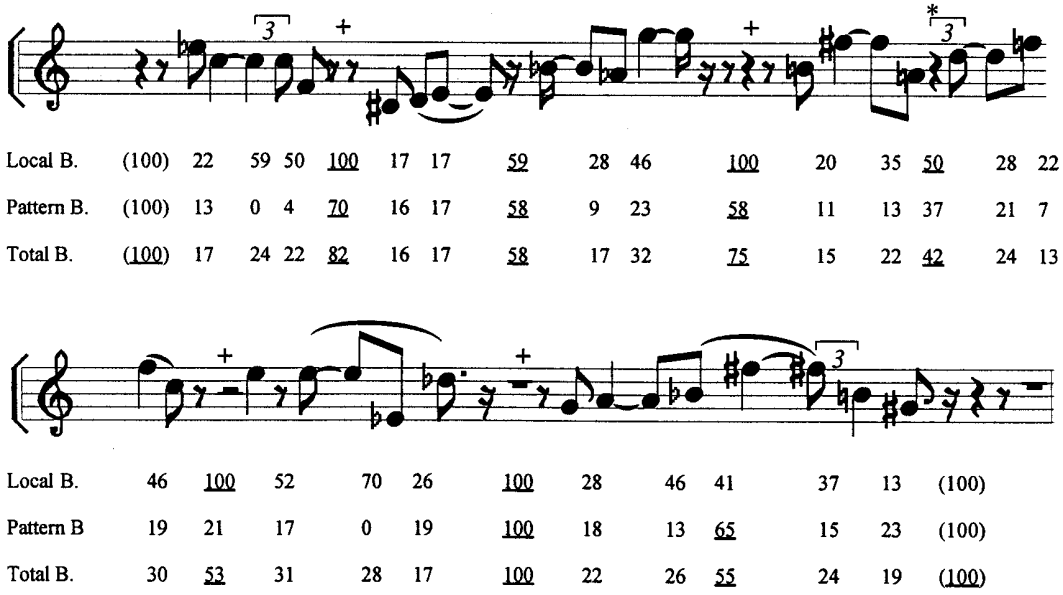



Figure 9.19 Local Boundary strength profile, Pattern Boundary strength profile and a weighted Total Boundary strength profile for Babbit's melody.

[2] Imberty (1993) endorses this view and states: 'Atonal musical structure rests on other polarities, not situated at the level of the series itself which hence cannot serve as a prototype or frame of reference in perception and memory.' (p.327).

The atonal melody (figure 9.19) is presented to the computational model as a primitive melodic surface (0). Pitch (and pitch-intervals) are converted into the 12-tone *GPIR* representation (rather than the *GPIR* diatonic representation).

The Local Boundary strengths are calculated with the application of *LBDM* (figure 9.19) and the local accent values are then automatically derived. From the various low-level metric grids the best match occurs for the eighth-note metric grid starting on the first beat (according to the indicated tempo this is a possible beat level); there is no preference for the quarter-note metric grid or for other higher-level grids - see table 9.7. The melodic structure of this song does not evoke any regular metric structure above the beat level; this is consistent with the fact that the composer indicates multiple time signatures through the course of the song.

| Metric Grid | 2/16 | | 3/16 | | | 2/8 | | 3/8 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Offset | 0 | 1/16 | 0 | 1/16 | 2/16 | 0 | 1/8 | 0 | 1/8 | 2/8 |
| Total Value | 296 | 11 | 204 | 177 | 204 | 316 | 277 | 408 | 159 | 321 |

Table 9.7

If the *String Pattern-Induction Algorithm (SPIA)* and *Selection Function* are applied to the melodic surface for patterns that consist of 3-4 notes (the algorithm is applied to the scale-step interval and the duration parametric profiles) no low-level regular grouping structure is revealed. As there is no clear metrical structure above the eighth-note level it is likely that low-level grouping processes may play an important role for the perception of possible boundaries. As can be seen, especially in the boundary strength profile of the refined version of the *LBDM* (figure 9.19), some boundaries are very strongly marked out - in these positions 'hard' break markers are inserted (indicated by crosses in figure 9.19).

The *SPIA* is now applied on the melodic surface for patterns consisting of 3-7 notes (exact and near-exact pitch interval, contour, duration and relative duration matching) and on a version of the surface in which repeated notes - with no strong boundary between them - are merged (only exact and near-exact pitch interval matching); the reduction of the surface by the elimination of notes that are shorter than an eighth durational value whose inceptions do not coincide with points of the eighth-note beat level is meaningless as there are hardly any such notes (actually only one). The resulting Pattern Boundary strength profile is depicted in figure 9.19. It should be noted that there are hardly any pattern

matches for the exact and near-exact pitch interval and the duration profiles. This could be taken to suggest that either the Pattern Boundary strength profile should make a weaker contribution towards the Total Boundary profile or that the melodic surface should be represented by additional parametric features (e.g. the gap between the exact pitch-interval and the contour representation could be filled by representations such as step-leap sequences etc.) so that more sophisticated matching may be realised (neither of these suggestions have been implemented at this stage).

The strongest peaks in the Total Boundary strength profile give a plausible segmentation of the musical surface as shown in figure 9.20. As the local peak at the point indicated by an asterisk in figure 9.19 is close to the selection threshold it may be selected as an additional segmentation point - the two resulting successive melodic segments are quite similar in terms of step-leap pitch interval and refined relative duration patterns (see below) and are also of the same size with the rest of the segments.

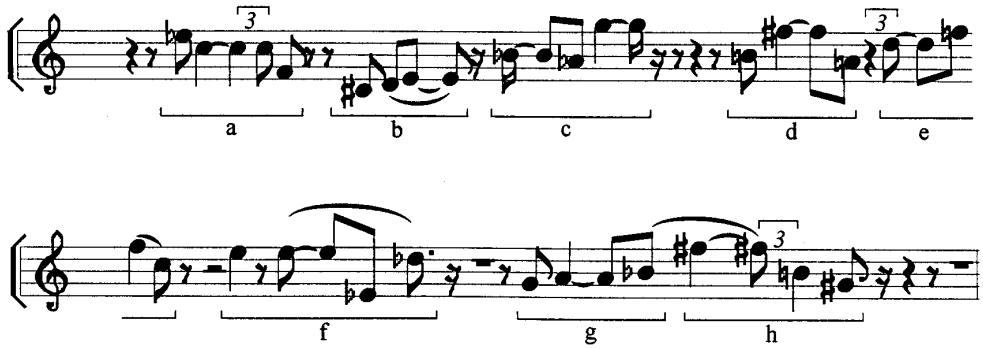



Figure 9.20 Preferred segmentation for given melody.

If these melodic segments are presented to the *Unscramble* algorithm they can be organised into categories. As there are hardly any matches for exact or near-exact pitch and duration patterns these segments will be represented - in addition to the attributes described in section 9.1.3 - by step-leap pitch interval and by refined relative duration patterns (i.e. much_shorter, shorter, equal, longer, much_longer); this more elaborate representation of the segments could - and perhaps should - have been applied to all the previous examples as well. The algorithm gives the following categorisations:

Threshold: 10.0 Categories: [[a,b,c,d,e,f,g,h]]
Threshold: 9.0 Categories: [[a,c,d,e,f,g,h],[b,g]]

Threshold: 8.75 Categories: [[a,d,e,g],[a,h],[b,g],[c,f,h]]
Threshold: 8.5 Categories: [[a,d,e,g],[a,h],[b,g],[c,f],[f,h]]
Threshold: 8.25 Categories: [[a,h],[b,g],[c],[d,e,g],[f]]
Threshold: 8.0 Categories: [[a,h],[b,g],[c],[d,e],[d,g],[f]]
Threshold: 6.25 Categories: [[a,h],[b,g],[c],[d,e],[f]]
Threshold: 6.0 Categories: [[a],[b,g],[c],[d,e],[f],[h]]
Threshold: 4.5 Categories: [[a],[b,g],[c],[d],[e],[f],[h]]

From these the 'best' categorisation description occurs for Threshold=6.25 (few monadic categories, no overlapping and not too general) - see figure 9.21. The properties that are most diagnostic of these categories are those that relate to pitch step-leap patterns, refined relative duration patterns and first-last note pitch interval; for instance, segments *a* and *h* share mainly the same step-leap and contour pattern (repeated notes merged) and first-last note pitch interval. As step-leap, contour and relative duration patterns provide much weaker links between melodic segments than exact or near-exact pitch interval and duration patterns this paradigmatic description is weaker than the descriptions we have found thus far in the previous examples.

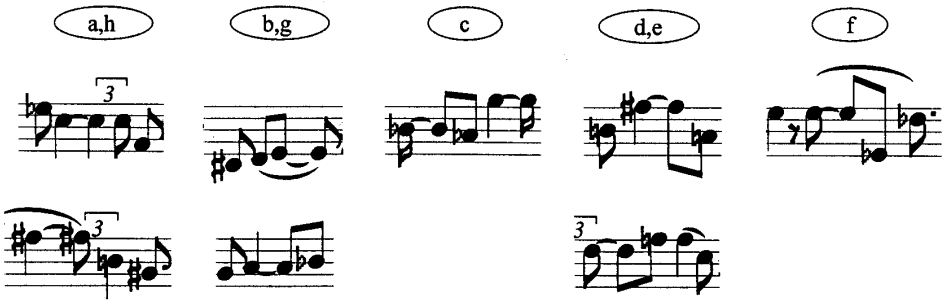



Figure 9.21

The analysis of the melody which is based solely on the properties of the twelve-tone series gives the following classification: [[a,h],[b,g],[c,f],[d,e]] in which each pair of melodic segments shares the same pitch-class interval sequence in a retrograde-inversion relation. As it happens, this description is not very different from the description given by the proposed computational model. However, they *are* different and what is actually radically different is the reasons for which they emerge (in principle, the relation between

analyses given by the proposed system and by serial analytic techniques can range from being identical to being incompatible).

## Conclusion

In the first part of this chapter a description of how the various more specialised computational components of the theory can be combined in order to achieve a analysis of a given melodic surface was presented. Then, the overall model was applied to four melodies from diverse musical repertoires; the results obtained were presented to a number of music analysts for evaluation and were judged as being 'reasonable'. There are still various aspects of the computational model that require further refinements and the tasks described in section 9.1.4 are as yet to be implemented. The four exemplar analyses presented above provide some preliminary evidence on the validity and generality of the proposed theory; however, the system needs further extensive testing on a much larger number of melodies from other styles and idioms, and further evaluation by more sophisticated techniques such as those proposed in section 3.4.2.

# Chapter 10

# Conclusion

## 10.1 Concluding remarks

This research study has gone a long way towards formulating a theory of musical structure that is *fully explicit* in the sense that it can readily be and has been implemented as a computer program, is *general* in the sense that it is style- and idiom-independent, and has an *inductive* outlook in the sense that it can make generalisations from musical examples rather than have musical knowledge embodied beforehand.

The *General Computational Theory of Musical Structure* is grounded on general cognitive and logical principles - especially on the notions of identity/difference, similarity and categorisation. Based on these principles a number of component modules specialised in different analytic tasks have been developed; the co-ordination of these modules enables the elicitation of a cognitively pertinent structural description of a musical surface.

A prototype computer system based on the *GCTMS* has been developed. In the four detailed examples, presented in the previous chapter, this system was applied on four melodies from diverse musical styles. Structural analyses of the melodies were generated by the system for the same set of default values. These analytic descriptions were then presented to a number of musical analysts and were judged as being 'acceptable' and even 'interesting'. Of course these analyses are not as sophisticated as analyses produced by musical style analysis experts; however, they give 'plausible' structural descriptions of a melodic surface that may be said to correspond to the intuitive understanding of a listener

not necessarily experienced in the certain musical style (the assumption that these descriptions should correspond to the intuitions of a listener is yet to be tested experimentally).

The fact that this system is able to produce such interesting analyses seems to support the validity of the assumed hypotheses and models of the *GCTMS*. It is encouraging - if not rather impressive - to see the quality of results from this system especially if one bears in mind that the given melodies have been presented to the system as mere unstructured sequences of notes and that the system does not essentially embody any style-dependent musical knowledge. Further experimentation with the computer system on a wide range of melodies from the 12-tone equal-temperament and other musical systems is essential for revealing the theory's full potential alongside possible shortcomings for which amendments to the theory will have to be made.

In the course of developing the overall theory a number of interesting conclusions have been reached:

- The representation of the musical surface strongly influences the course of the subsequent analytic processes. Especially for pitch, it has been shown that it is possible to create a more sophisticated representation of pitch and pitch intervals *(General Pitch Interval Representation)* that reflects properties of pitch-scale genres; this facilitates further analytic tasks such as pattern-matching and transcription of MIDI-pitch sequences to the traditional notation or other non-diatonic notation.

- The strength of local boundaries is reinforced by *any* discontinuities in the various interval profiles of a musical surface - not only discontinuities that involve effectively a longer musical interval in between shorter ones. The *Local Boundary Detection Model* was developed for finding local boundaries in a melodic surface.

- Local boundaries are closely related to local accents (phenomenal accents) of notes. This enables the automatic inference of the local accentuation structure from the grouping structure on which metrical templates can be matched.

- The segmentation of a musical surface is strongly influenced by both local boundaries and higher-level boundaries highlighted by significant musical patterns; the combination of these gives rise to an integrated overall segmentation or set of segmentations (this process is fully described in the present theory).

- Parallel musical patterns can be highlighted by the use of pattern-matching techniques. It has been suggested that full pattern-matching techniques such as the *String Pattern-Induction Algorithm* applied on multiple parametric profiles of a musical surface may be more useful than partial pattern-matching techniques applied on only the pitch and duration profiles of the surface. Significant patterns are revealed if pattern-matching techniques are coupled by a selection mechanism that rates patterns in relation to a number of characteristics such as frequency of occurrence, pattern length and degree of overlapping.

- Musical similarity is contextually defined and is strongly linked to categorisation processes. The *Unscramble* machine learning algorithm determines similarity between musical segments alongside with placing them in categories/paradigms; it also highlights those musical parametric features that are most characteristic of each category.

- The derivation of musical structure is not a uni-directional procedure that starts with the lowest-level details of the musical surface and ends with the highest-level description; this procedure has a loose overall directionality from lower- to higher-level descriptions but higher-level analytic results inform and disambiguate lower-level ones; this is manifested in the indispensability of collaboration between the pattern-matching mechanism with the lower-level boundary model for the determination of an integrated boundary strength profile, and also in the serviceability of the categorisation model for the selection of a better segmentation.

## 10.2 Future Developments

As the *General Computational Theory of Musical Structure* is not fully developed at the present time there are a number of future tasks yet to be described.

• *General Chord Representation.* Harmonic aspects of the musical surface have not been addressed in the present research study. However a preliminary informal study has been made that examines the possibility of formulating a *General Chord Representation* (similar in many ways to the *General Pitch Interval Representation)* in which various harmonic systems may be accommodated. It seems possible to extract information on the frequency of occurrence of harmonic intervals found in a number of musical works of the same idiom and then use the smallest and the most frequent interval as the intervallic unit which forms the basis for arranging compact versions of different chord types. For instance, in tonal pieces such as J.S.Bach's chorales by far the most frequent small harmonic interval is the 3rd and this may lead to a triadic representation of chords; in some works by B.Bartók the most frequent small interval is found to be the 3-semitone interval or the tritone; in many atonal works the most frequent small interval is the semitone which may lead to a pc-set representation of chords. It is suggested that, by examining musical examples, generalisations on the vertical organisation of pitches can be made and these used to create more elaborate descriptions of various harmonic systems.

As most of the components of the *GCTMS* can be applied to any parametric profile of a musical surface it seems plausible to apply them on a sequence of chord types as well. This way harmonic boundaries may be suggested and harmonic progression patterns may be discovered. These can contribute to the overall description of a musical surface.

• *Temporal Relations and Functions.* The syntagmatic organisation of musical works has not as yet been described. This might involve constructing a grammar (most likely a probabilistic grammar - see Rader, 1993) that would organise the musical structures (motives, themes etc.) discovered by the other components of *GCTMS* temporally/sequentially at various hierarchic levels.

Perhaps the simplest way this may be achieved is by employing first-order stochastic processes, i.e. Markov chains (Ames, 1989; Cambouropoulos, 1994), on many parametric

profiles of a surface at a number of levels of abstraction. Although Markov chains are thought to be very weak in terms of capturing musical in-time processes (they have a very short memory) they may partially overcome this problem if they are applied to many levels of abstraction (such as the note-level, accented-note level, motive-level etc.); this way the temporal hierarchical structure of a musical work may be described more effectively.

Higher-order stochastic processes should also be considered - e.g. context models which are a subclass of such processes described by Conklin and Witten (1995). It is suggested that the *String Pattern-Induction Algorithm* could be useful in constructing such models because a pattern of n-elements may be regarded as a context of n-1 elements followed by the $n^{th}$ element; in this sense, either the resulting patterns of *SPIA* may be re-organised into a stochastic table (consideration should be made of the fact that patterns appearing once are discarded by *SPIA)* or stochastic tables may be constructed concurrently with the application of *SPIA*. A possible advantage of making use of *SPIA* is that one need not define in advance an arbitrary threshold for the highest order of the stochastic process as *SPIA* stops when it discovers maximum length pattern-matches.

• *Interactive Analytic and Compositional System.* Perhaps the ultimate application goal of this whole enterprise would be to implement an interactive system for musical analysis and composition which would allow a musician/user to obtain useful analytic results of a given musical work and to generate interesting novel compositions. Such a system will consist of two main general and sophisticated modules: the analytic and the compositional module. The user will present to the system a musical example, this will be analysed by the analytic engine (based on *GCTMS)* and then the compositional module may be used to re-compose a new musical piece in the 'style' of the one analysed. At an intermediate level the user can alter or redefine the data obtained by the analytic module as demanded. The new musical work may then be fed back as input to the system, its analytic description induced and then possibly used for further compositional purposes (Figure 10.1).

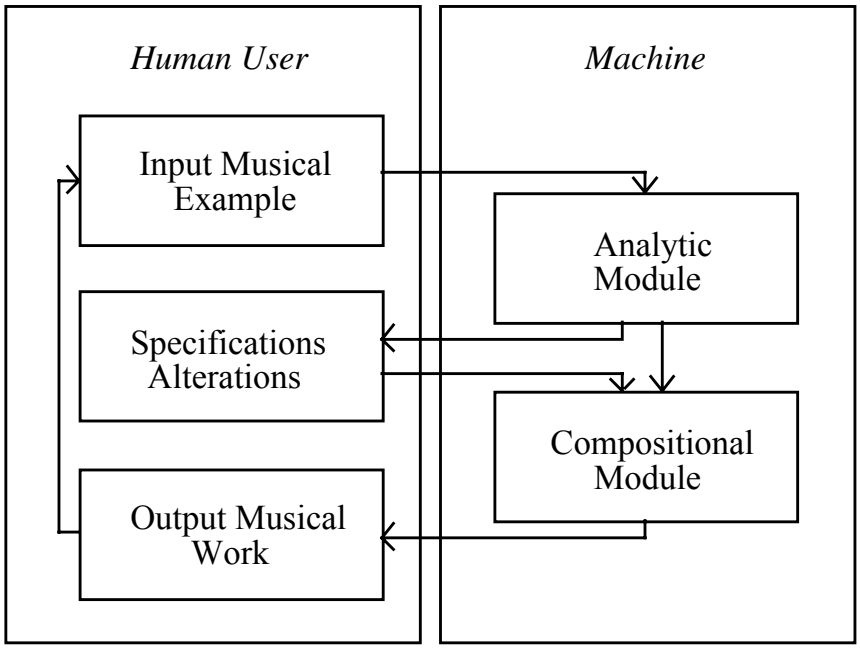



Figure 10.1 The interactive AI system for musical analysis and composition

As most of the analytic information obtained by the *GCTMS*-based analytic module is explicitly represented, it is possible to use this information so that new musical surfaces may be generated. For instance, knowing the exact structure of the various members of a thematic/motivic category, a new altered theme/motive with possible variations may be created; or having constructed a grammar that describes the given musical structure this grammar may be used to generate new surfaces with possible alterations being introduced at various levels.

The compositional module is primarily viewed as a means to test and evaluate the analytic module - and indirectly test the *GCTMS* on which this module is based. If there is no intervention at any intermediate analytic stage, the compositional output should be structurally in the same 'style' as the analysed piece, i.e. all the structural features and relations will be preserved intact, a new surface merely being generated.

The system may also be used for creative compositional purposes. The user/composer may present to the analytic module a musical example of the style of music she/he would envisage to compose and then interact with the system altering features at any level of the analysis so that a novel musical surface may be generated. When a new satisfying musical

work has been created, the user may feed it back to the analytic module so that its structure is learned (i.e. a new analytic description is produced) and further compositional experimentation and generation may take place. This way, the system gradually adapts itself and learns the new 'personal' style of the user/composer. The new descriptions may be used as a means for further compositional explorations.

# Bibliography


Ames, C. (1989) The Markov Process as a Compositional Model: A Survey and Tutorial. *Leonardo*, 22(2):175-187.

Arabie, P. (1977) Clustering Representations of Group Overlap. *Journal of Mathematical Sociology,* 5:113-128.

Arabie, P. et al. (1981) Overlapping Clustering: A New Method for Product Positioning. *Journal of Marketing Research*, 18: 310-317.

Aristoxenos, *Elementa Harmonica.* Translated in A. Barker (1989) *Greek Musical Writings II - Harmonic and Acoustic Theory,* Cambridge University Press, Cambridge.

Arom, S. (1991) *African Polyphony and Polyrhythm.* Cambridge University Press, Cambridge.

Ashley, R.D. (1989) Modelling Music Listening: General Considerations. *Contemporary Music Review*, 4:295-310.

Baker, M. (1989) An Artificial Intelligence Approach to Musical Grouping Analysis. Contemporary Music Review, 3:43-68.

Baroni M. and Jacoboni C. (1978) *Proposal for a Grammar of Melody.* Les Presses de l'Université de Montréal.

Barsalou, L.W. and Hale, C.R. (1993) Components of Conceptual Representation: From Feature Lists to Recursive Frames. In *Categories and Concepts,* I.Van Mechelen et al. (eds), Academic Press, London.

Barsalou, L.W. (1992) *Cognitive Psychology: An Overview for Cognitive Scientists.* Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Barsalou, L.W. (1985) Ideals, Central Tendency, and Frequency of Instantiation and Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4):629-649.

Barsalou, L.W. (1983) Ad Hoc Categories. *Memory and Cognition*, 11:211-227.

Barsalou, L.W. and Ross, B.H. (1986) The Roles of Automatic and Strategic Processing in Sensitivity to Superordinate and Property Frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1):116-134.

Beckner, M. (1959) *The Biological Way of Thought.* Columbia University Press, New York.

Bent, I.D. (1980) Analysis. In *The New Grove Dictionary of Music and Musicians*, *Vol.1*. Macmillan, London.

Bharucha, J.J. (1984a) Anchoring Effects in Music: The Resolution of Dissonance. *Cognitive Psychology,* 16:485-518.

Bharucha, J.J. (1984b) Event Hierarchies, Tonal Hierarchies and Assimilation: A Reply to Deutsch and Dowling. *Journal of Experimental Psychology: General,* 113:421-425.

Bigand, E. (1990) Abstraction of Two Forms of Underlying Structure in a Tonal Melody. *Psychology of Music,* 18:45-59.

Black, M. (1952) The Identity of Indiscernibles. *Mind*, 61(242):153-164.

Bloch, J. J. and Dannenberg, R. B. (1985) Real-Time Computer Accompaniment of Keyboard Performances. In *Proceedings of the International Computer Music Conference*, San Fransisco.

Bratko, I. (1990) *PROLOG: Programming for Artificial Intelligence* (2nd edition). Addison-Wesley, Wokingham, England.

Bregman, A. S. (1990) *Auditory Scene Analysis*. The MIT Press, Cambridge (Ma).

Brinkman, A. R. (1990) *Pascal Programming for Music Research.* The University of Chicago Press, Chicago.

Browne, R. (1981) Tonal Implication of the Diatonic Set. *In Theory Only,* 5(6,7):3-21.

Butler, D. (1992) Review of E. Narmour's: 'The Analysis and Cognition of Basic Melodic Structures'. *Music Perception*, 10(2):243-251.

Cambouropoulos, E. (1998) Musical Parallelism and Melodic Segmentation. In *Proceedings of the XII Colloquio di Informatica Musicale*, Gorizia, Italy.

Cambouropoulos, E. (1997a) The Role of Similarity in Categorisation: Music as a Case Study. In *Proceedings of the Third Triennial ESCOM Conference*, Uppsala, Sweden.

Cambouropoulos, E. (1997b) Musical Rhythm: Inferring Accentuation and Metrical Structure from Grouping Structure. In *Music, Gestalt and Computing - Studies in Systematic and Cognitive Musicology*, M. Leman (ed.), Springer-Verlag, Berlin.

Cambouropoulos, E. (1996a) A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface. In *Proceedings of the III Journées d' Informatique Musicale*, Caen, France.

Cambouropoulos, E. (1996b) A General Pitch Interval Representation: Theory and Applications. *Journal of New Music Research,* 25 (3): 231-251.

Cambouropoulos, E. (1994) Markov Chains as an Aid to Computer-Assisted Composition. *Musical Praxis*, 1(1):41-52.

Cambouropoulos, E. and Smaill, A. (1997) Similarity and Categorisation Inextricably Bound Together: The *Unscramble* Machine Learning Algorithm. In *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, Edinburgh.

Cambouropoulos, E. and Smaill, A. (1995) A Computational Model for the Discovery of Parallel Melodic Passages. In *Proceedings of the XI Colloquio di Informatica Musicale*, Bologna, Italy.

Camilleri, L. (1992) Computational Theories of Music: Theoretical and Applicative Issues. In *Computer Representations and Models in Music*. A. Marsden et al. (eds), Academic Press, London.

Camilleri, L., Carreras, F. and Duranti, C. (1990) An Expert System Prototype for the Study of Musical Segmentation. *Interface,* 19(2-3).

Carey, S. (1985) *Conceptual Change in Childhood,* The MIT Press, Cambridge (Ma).

Carterette, E.C., Hohl, D.V. and Pitt, M.A. (1986) Similarities Among Transformed Melodies: The Abstraction of Invariants. *Music Perception*, 3(4):393-410.

Clarke, E.F. (1997) From Perception to Critique. In *Proceedings of the Third Triennial ESCOM Conference,* Uppsala, Sweeden.

Clarke, E.F. (1987) Levels of Structure in the Organisation of Musical Time. *Contemporary Music Review,* 2:211-238.

Clarke, E.F. (1985) Structure and Expression in Rhythmic Performance. In *Musical Structure and Cognition,* P. Howell et al. (eds), Academic Press, London.

Clarke, E.F. and Dibben, N. (1997) An Ecological Approach to Similarity and Categorisation in Music. In *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, Edinburgh.

Clough, J. (1980) Diatonic Interval Sets and Transformational Structures. *Perspectives of New Music,* 18(2):461-482.

Clough, J. (1979) Aspects of Diatonic Sets. *Journal of Music Theory,* 23(1):45-60.

Clough, J. and Myerson, G. (1985) Variety and Multiplicity in Diatonic Systems. *Journal of Music Theory,* 29 (2) 249-270.

Cogan, R. and Escot, P. (1976) *Sonic Design: The Nature of Sound and Music*. Prentice Hall, Englewood.

Conklin, D. and Witten, I. H. (1995) Multiple Viewpoint Systems for Music Prediction. *Journal of New Music Research,* 24:51-73.

Conklin, D. and Witten, I. H. (1991) Prediction and Entropy of Music. Technical Report 91/457/41, University of Calgary, Dept. of Computer Science, Canada.

Cook, N. (1990) *Music, Imagination and Culture.* Oxford University Press, Oxford.

Cook, N. (1987) *A Guide to Musical Analysis,* J.M.Dent & Sons Ltd, London.

Cooper, G.W. and Meyer, L.B. (1960) *The Rhythmic Structure of Music.* The University of Chicago Press, Chicago.

Cope, D. (1993) A Computer Model of Music Composition. In *Machine Models of Music*, S.M. Schwanauer and D.A. Levitt (eds), The MIT Press, Cambridge (Ma).

Cope, D. (1992a) Computer Modelling of Musical Intelligence in EMI. *Computer Music Journal*, 16(2):69-83.

Cope, D. (1992b) On Algorithmic Representation of Musical Style. In *Understanding Music with AI,* M. Balaban et al. (eds), The MIT Press, Cambridge (Ma).

Cope, D. (1991) *Computers and Musical Style.* Oxford University Press, New York.

Cope, D. (1990) Pattern-Matching as an Engine for the Computer Simulation of Musical Style. In *Proceedings of the International Computer Music Conference*, Glasgow.

Cope, D. (1987) An Expert System for Computer-Aided Composition. *Computer Music Journal,* 11(4):30-46.

Crawford, T., Iliopoulos, C.S. and Raman, R. (1997) String Matching Techniques for Musical Similarity and Melodic Recognition. Research Report, Algorithm Design Group, Dept. of Computer Science, King's College London.

Crochemore, M. (1981) An Optimal Algorithm for Computing the Repetitions in a Word. *Information Processing Letters,* 12(5):244-250.

Cross, I. (1995) Review of E. Narmour's: 'The Analysis and Cognition of Melodic Complexity'. *Music Perception*, 12(4):486-509.

Crow, D. and Smith, B. (1992) DB_Habits: Comparing Minimal Knowledge and Knowledge-Based Approaches to Pattern Recognition in the Domain of User-Computer Interactions. In *Neural Networks and Pattern Recognition in Human-Computer Interaction,* R. Beale et al. (eds), Ellis-Horwood, London.

Deliège, I. (1997) Similarity in Processes of Categorisation: Imprint Formation as a Prototype Effect in Music Listening. In *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, Edinburgh.

Deliège, I. (1996) Cue Abstraction as a Component of Categorisation Processes in Music Listening. *Psychology of Music,* 24:131-156.

Deliège, I. (1987) Grouping Conditions in Listening to Music: An Approach to Lerdahl and Jackendoff's Grouping Preference Rules. *Music Perception*, 4:325-360.

Deliège, I. and Mélen, M. (1997) Cue Abstraction in the Representation of Musical Form. In *Perception and Cognition of Music*. I. Deliège and J. Sloboda (eds), Psychology Press Ltd, Hove, U.K.

Desain, P. and Honing H. (1992) *Music, Mind and Machine*. Thesis Publishers, Amsterdam.

Deutsch, D. (1984) Two Issues Concerning Tonal Hierarchies: Comment on Castellano, Bharucha & Krumhansl. *Journal of Experimental Psychology: General,* 113:413-416.

Deutsch, D. (1982a) Grouping Mechanisms in Music. In D. Deutsch (ed.), *The Psychology of Music,* Academic Press, New York.

Deutsch, D. (1982b) The Processing of Pitch Combinations. In *The Psychology of Music,* D. Deutsch (ed.), Academic Press, New York.

Deutsch, D. and Feroe, J. (1981) The Internal Representation of Pitch Sequences in Tonal Music. *Psychological Review,* 88:503-22.

Dibben, N. (1994) The Cognitive Reality of Hierarchic Structure in Tonal and Atonal Music. *Music Perception,* 12(1):1-25.

Dowling, W.J. and Harwood D.L. (1986) *Music Cognition.* Academic Press, New York.

Ebcioglu, K. (1993) An Expert System for Harmonising Four-part Chorales. In *Machine Models of Music*, S.M. Schwanauer and D.A. Levitt (eds), The MIT Press, Cambridge (Ma).

Epstein, D. (1995) *Shaping Time: Music, the Mind and Performance*. Schirmer books, New York.

Estes, W.K. (1994) *Classification and Cognition.* Oxford University Press, Oxford.

Eysenck, M.W. and Keane, M.T. (1995) *Cognitive Psychology* (3rd edition). Lawrence Erlbaum Associates, Hove, U.K.

Fisher, D.H. (1987) Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning*, 2:139-172.

Forte, A. (1973) *The Structure of Atonal Music.* Yale University Press, New Haven.

Francès, R. (1988) *The Perception of Music.* Lawrence Erlbaum Associates, London.

Garner, W. R. (1974) *The processing of Information and Structure.* Lawrence Erlbaum Associates, London.

Gennari, J.H., Langley, P. and Fisher, D. (1989) Models of Incremental Concept Formation. *Artificial Intelligence*, 40:11-61.

Gibson, J.J. (1966) *The Senses Considered as Perceptual Systems*. George Allen and Unwin Ltd., London.

Gillies, M. (1993) Pitch Notations and Tonality: Bartók. In *Early Twentieth-Century Music,* J.Dunsby (ed.), Basil Blackwell Ltd, Oxford.

Goldstone, R.L. (1994) The Role of Similarity in Categorisation: Providing a Groundwork. *Cognition,* 52:125-157.

Goldstone, R.L., Medin, D.L. and Gentner, D. (1991) Relational Similarity and the Non-independence of Features in Similarity Judjements. *Cognitive Psychology,* 23:222-262.

Goodman, N. (1972) Seven Strictures on Similarity. In *Project and Problems,* by N. Goodman, The Bobbs-Merrill Comany, Inc., New York.

Guertin, M. (1981) Différence and Similitude Dans le Préludes Pour Piano de Debussy. *Revue de Musique des Universités Candiennes (Montreal),* 2:56-83.

Hampton, J. (1993) Prototype Models of Concept Representation. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Hampton, J. and Dubois, D. (1993) Psychological Models of Concepts: Introduction. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Handel, S. (1989) *Listening. An Introduction to the Perception of Auditory Events.* The MIT Press, Cambridge (Ma).

Harris, M., Smaill, A. and Wiggins, G. (1991) Representing Music Symbolically. In *Proceedings of the IX Colloquio di Informatica Musicale*, Associazione di Informatica Musicale, Venice.

Hasher, L. and Zacks R.T. (1979) Automatic and Effortful Processes in Memory. *Journal of Experimental Psychology: General*, 108:356-388.

Hintzman, D.L. (1976) Repetition and Memory. In *The Psychology of Learning and Motivation, Vol. 10,* G.H.Bower (ed.), Academic Press, London.

Hiraga, Y. (1997) Structural Recognition of Music by Pattern-Matching. In *Proceedings of the International Computer Music Conference*, Thessaloniki, Greece.

Huron, D. (1996) Humdrum - Frequently Asked Questions. This file is available by sending a request to: dhuron@watserv1.uwaterloo.ca

Huron, D. (1994) The Humdrum Toolkit; Reference Manual. CCARH, Merlo Park.

Iliopoulos, C.S., Moore, D.W.G. and Park, K. (1996) Covering a String. *Algorithmica,* 16:288-297.

Imberty, M. (1993) How Do We Perceive Atonal music? Suggestions for a Theoretical Approach. *Contemporary Music Review*, 9(1&2):325-337.

Jackendoff, R. (1992) Musical Processing and Musical Affect. In *Cognitive Bases of Music Communication.* M. R. Jones et al. (eds), American Psychological Association, Washington DC.

Jackendoff, R. (1987) *Consciousness and the Computational Mind.* The MIT Press, Cambridge (Ma).

Jeffries, T.B. (1974) Relationship of Interval Frequency Count to Ratings of Melodic Intervals. *Journal of Experimental Psychology*, 102(5):903-905

Kassler, M and Howe, H.S. (1980) Computers and Music. In *The New Grove Dictionary of Music and Musicians*, *Vol.4*. Macmillan, London.

Kornstädt, A. (1995-96) SCORE-to-Humdrum: A Graphical environment for Musicological Analysis. *Computing In Musicology,* 10:105-122.

Kramer, J.D. (1983) *The Time of Music.* Schirmer Books, New York.

Krumhansl, C.L. (1997) Effects of Perceptual Organization and Musical Form on Melodic Expectancies. In *Music, Gestalt and Computing*, Marc Leman (ed.), Springer-Verlag, Berlin.

Krumhansl, C.L. (1995) Music Psychology and Music Theory: Problems and Prospects. *Music Theory Spectrum*, 17:53-90.

Krumhansl, C.L. (1990) *Cognitive Foundations of Musical Pitch.* Oxford University Press, Oxford.

Krumhansl, C.L. (1978) Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship Between Similarity and Spatial Density. *Psychological Review,* 85(5):445-463.

Lakoff, G. (1987) *Women, Fire and Dangerous Things.* The University of Chicago Press, Chicago.

Lamont, A. (1997) Categorical Perception of Musical Pitch: Theory and Evidence. Research Report, University of Cambridge, Cambridge.

Lamont, A. and Dibben, N. (1997) Perceived Similarity of Musical Motifs: An Exploratory Study. In *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, Edinburgh.

Langley, P. (1996) *Elements of Machine Learning*. Morgan Kaufmann Publishers, San Fransisco.

Laske, O. (1993) A search for a Theory of Musicality. *Languages of Design*, 1:209-228.

Laske, O. (1992) Artificial Intelligence and Music: A Cornerstone of Cognitive Musicology. In *Understanding Music with AI,* M. Balaban et al (eds), The MIT Press, Cambridge (Ma).

Laske, O. (1988) Introduction to Cognitive Musicology. *Computer Music Journal*, 12(1):43-57.

Laske, O. (1977) Towards a Musicology for the Twentieth Century. *Perspectives of New Music*, 15(2):220-225.

Lee, C. S. (1991) The Perception of Metrical Structure: Experimental Evidence and a Model. In *Representing Musical Structure,* P. Howell et al. (eds), Academic Press, London.

Leman, M. (1995) *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*. Springer-Verlag, Berlin.

Leman, M. (ed.) (1997) *Music, Gestalt and Computing*. Springer-Verlag, Berlin.

Leman, M. and Schneider, A. (1997) Origin and Nature of Cognitive and Systematic Musicology: An Introduction. In *Music, Gestalt and Computing*, M. Leman (ed.), Springer-Verlag, Berlin.

Lerdahl, F. (1992) Pitch-Space Journeys in Two Chopin Preludes. In *Cognitive Bases of Musical Communication*, M.R. Jones et al (eds), American Psychological Association, Washington DC.

Lerdahl, F. (1989) Atonal Prolongation Structure. *Contemporary Music Review*, 4:65-87.

Lerdahl, F. (1988) Tonal Pitch Space. *Music Perception*, 5(3):315-350.

Lerdahl, F. and Jackendoff, R. (1983) *A generative Theory of Tonal Music,* The MIT Press, Cambridge (Ma).

Lidov, D. (1980) Musical Structure and Musical Significance, Part I. Working Paper. Victoria University, Toronto.

Lockwood, L. (1980) L'homme armé. In *The New Grove Dictionary of Music and Musicians*, *Vol.10*. Macmillan, London.

Longuet-Higgins, H. C. and Lee, C. S. (1984) The Rhythmic Interpretation of Monophonic Music. *Music Perception,* 1:424-441.

Longuet-Higgins, H. C. and Lee, C. S. (1982) The Perception of Musical Rhythms. *Perception,* 11:115-128.

Longuet-Higgins, H. C. (1976/1987) The Perception of Melodies. In *Mental Processes - studies in cognitive science.* The MIT Press, Cambridge (Ma).

Longuet-Higgins, H. C. (1987) *Mental Processes - studies in cognitive science.* The MIT Press, Cambridge (Ma).

Luger, G.F. and Stubblefield, W.A. (1993) *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (2nd edition). The Benjamin/Cummings Publishing Company, New York.

Marr, D. (1982) *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information.* W.H. Freeman and Company, New York.

Marsden, A. (1992) Modelling the Perception of Musical Voices: a Case Study in Rule-Based Systems. In *Computer Representations and Model in Music*. A. Marsden and A. Pople (eds), Academic Press, London.

Marsden, A. (1991) Musical Abstractions and Composers' Practice. In *Proceedings of the IX Colloquio D' Informatica Musicale*, Genova, Italy.

Marsden, A. and Pople, A. (1989) Modelling Musical Cognition as a Community of Experts. *Contemporary Music Review*, 3:29-42.

McAdams, S. (1989) Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review,* 4:181-198.

McAdams, S. (1984) The auditory Image: A metaphor for musical and psychological research on auditory organisation. In *Cognitive Processes in the Perception of Art,* W.R. Crozier and A.J. Chapman (eds), North-Holland Press, Amsterdam.

McGettrick, P. (1997) *MIDIMatch: Musical Pattern Matching in Real Time*. MSc Dissertation, University of York, U.K.

Medin, D.L., Goldstone, R.L. and Gentner, D. (1993) Respects for Similarity. *Psychological Review,* 100(2):254-278.

Meyer, L.B. (1973) *Explaining Music: Essays and Explorations*. University of California Press, Berkeley.

Meyer, L.B. (1957) Meaning in Music and Information Theory. *Journal of Aesthetics and Art Criticism*, 15:412-424.

Meyer, L. B. (1956) *Emotion and Meaning in Music,* University of Chicago Press, Chicago.

Michalski, R.S. (1993) Beyond Prototypes and Frames: The Two Tiered Representation. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Michalski, R.S. (1987) Concept Learning. In *Encyclopedia of AI,* S.C. Shapiro (ed.), John Wiley and Sons, New York.

Michalski, R.S. (1983) Learning from Observation: Conceptual Clustering. In *Machine Learning: An Artificial Intelligence Approach,* R.S. Michalski et al. (eds), Springer-Verlag, Berlin.

Michalski, R.S. and Stepp, R.E. (1987) Clustering. In *Encyclopedia of AI,* S.C. Shapiro (ed.), John Wiley and Sons, New York.

Minsky, M. (1993) Music, Mind and Meaning. In *Machine Models of Music*, S.M. Schwanauer and D.A. Levitt (eds) , The MIT Press, Cambridge (Ma).

Minsky, M. (1988) *The Society of Mind*. Touchstone Editions, New York.

Moelants, D. (1997) A Framework for the Subsymbolic Description of Meter. In *Music, Gestalt and Computing*, Marc Leman (ed.), Springer-Verlag, Berlin.

Monelle, R. (1992) *Linguistics and Semiotics in Music.* Harwood Academic Publishers, Reading.

Morin, E. (1979) *Essai de Stylistique Comparée.* University of Montreal, Montreal.

Murphy, G.L. (1993) Theories and Concept Formation. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Murphy, G.L. and Medin, D.L. (1985) The Role of Theories in Conceptual Coherence. *Psychological Review,* 92(3):289-316.

Murtagh, F.D. (1993) Cluster Analysis Using Proximities. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Narmour, E. (1992a) *The Analysis and Cognition of Melodic Complexity.* The University of Chicago Press, Chicago.

Narmour, E. (1992b) The Influence of Embobied Registral Motion on the Perception of Higher-Level Melodic Processes. In *Cognitive Bases of Music Communication.* M. R. Jones et al. (eds), American Psychological Association, Washington DC.

Narmour, E. (1990) *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realisation Model.* The University of Chicago Press, Chicago.

Nattiez, J-J. (1990) Music and Discourse: Towards a Semiology of Music. Princeton University Press, Princeton.

Nattiez, J-J. (1989) Reflections on the Development of Semiology in Music, *Music Analysis,* 8(1-2):21-75.

Nattiez, J-J. (1982) Varese's 'Density 21.5': a Study in Semiological Analysis. *Musical Analysis,* 1(3):243-340.

Nattiez, J-J. (1975) *Fondements d'une Sémiologie de la Musique.* Union Générale d'Editions, Paris.

Newell, A. (1982) The Knowledge Level. *Artificial Intelligence* 18(1):87-127.

Newell, A. and Simon, H. (1976) Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM,* 19(3):113-126.

Palisca, C.V. (1980) Theory. In *The New Grove Dictionary of Music and Musicians*, *Vol.18*. Macmillan, London.

Parncutt, R. (1994) Template-Matching Models of Musical Pitch and Rhythm Perception. *Journal of New Music Research,* 23:145-167.

Parncutt, R. (1994a) A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Perception*, 11(4):409-464.

Polansky, L. (1996) Morphological Metrics, *Journal of New Musical Research,* 25(4):298-368.

Pollard-Gott, L. (1983) Emergence of Thematic Concepts in Repeated Listening to Music. *Cognitive Psychology*, 15:66-94.

Povel, D. J. and Essens, P. (1985) Perception of Temporal Patterns. *Music Perception,* 2:411-440.

Pylyshyn, Z.W. (1989) Computing in Cognitive Science. In *Foundations of Cognitive Science*, M.I. Posner (ed.), The MIT Press, Cambridge (Ma).

Quine, W.V. (1950) Identity, Ostension, and Hypostasis. *The Journal of Philosophy,* Volume XLVII (22):621-633.

Rader, G.M. (1993) A Method for Composing Simple Traditional Music by Computer. In *Machine Models of Music*, S.M. Schwanauer and D.A. Levitt (eds), The MIT Press, Cambridge (Ma).

Rahn, J. (1980) *Basic Atonal Theory*. Schirmer Books, New York.

Ramscar, M., Hahn, U., Cambouropoulos, E. and Pain, H. (eds) (1997) *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation*, University of Edinburgh, Edinburgh.

Read S. (1994) *Thinking About Logic: An Introduction to the Philosophy of Logic.* Oxford Uviversity Press, Oxford.

Reneger, E. (1964) On Allen Forte's Theory of Chords. *Perspectives of New Music,* 8(1):191-212.

Reti, R. (1951) *The thematic Processes in Music,* The Macmillan Company, New York.

Rips, L.J. (1989) Similarity, Typicality and Categorisation. In *Similarity and Analogical Reasoning.* S. Vosniadou and A. Ortony (eds), Cambridge University Press, Cambridge.

Robbie, C. and Smaill, A. (1995) Implementing Lerdahl and Jackendoff's Grouping Rules in an Interactive System. In *Proceedings of the International Congress in Music and Artificial Intelligence,* University of Edinburgh, Edinburgh.

Rosch, E. (1978) Principles of Categorisation. In *Cognition and Categorisation.* E. Rosch and B.B. Lloyd (eds), Lawrence Erlbaum Associates, London.

Rosch, E. (1975) Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104:192-232.

Rosenthal, D. (1992) Emulation of Human Rhythm Perception. *Computer Music Journal,* 16(10):64-76.

Rowe, R. (1995) Artificial Intelligence and Musical Interaction. In *Proceedings of the International Congress in Music and AI,* University of Edinburgh, Edinburgh.

Rowe, R. (1993) *Interactive Music Systems (Machine Listening and Composing),* The MIT Press, Cambridge (Ma).

Rowe, R. (1992) Machine Listening and Composing with Cypher, *Computer Music Journal,* 16(1):43-63.

Rowe, R. and Li, T.C. (1995) Pattern Processing in Music. In *Proceedings of the Fifth Biennial Symposium for Arts and Technology*. Connecticut College, New London, Connecticut.

Russell, B. (1949) *A Critical Exposition of the Philosophy of Leibniz.* Bradford and Dickens, London.

Russell, B. and Whitehead, A.N. (1964) *Principia Mathematica.* (2nd ed.) Allen & Unwin, London.

Ruwet, N. (1987) Methods of Analysis in Musicology. *Music Analysis* 6(1-2):4-39.

Schackford, C. (1962) Some Aspects of Perception, Part II: Interval Sizes and Tonal Dynamics in Performance. *Journal of Music Theory,* 6:66-90.

Schackford, C. (1961) Some Aspects of Perception, Part I: Sizes of Harmonic Intervals in Performance. *Journal of Music Theory,* 5:162-202.

Searle, J.R. (1992) *The Rediscovery of the Mind*. The MIT Press, Cambridge (Ma).

Selfridge-Field, E. (1990) Reflection on Technology and Musicology. *Acta Musicologica*, 62:302-315.

Seifert, U., Olk, F. and Scheider, A. (1995) On Rhythm Perception: Theoretical Issues, Empirical Findings. *Journal of New Music Research,* 24:164-195.

Shepard, R.N. (1962a) The Analysis of Proximities: Multi-dimensional Scaling with an Unknown Distance Function, Part I. *Psychometrica*, 27:125-140.

Shepard, R.N. (1962b) The Analysis of Proximities: Multi-dimensional Scaling with an Unknown Distance Function, Part II. *Psychometrica*, 27:219-246.

Sloboda, J.A. (1985) *The Musical Mind: The Cognitive Psychology of Music.* Oxford University Press, Oxford.

Stammen, D.R. and Pennycook, B. (1993) Real-time Recognition of Melodic Fragments Using the Dynamic Timewarp Algorithm. In *Proceedings of the International Computer Music Conference*, Tokyo.

Steedman, M. J. (1977) The Perception of Musical Rhythm and Metre. *Perception,* 6:555-569.

Stroll, A. (1967) Identity. *The Encyclopedia of Philosophy*. Macmillan Publishing Co., London.

Sutcliffe, J.P. (1993) Concept, Class, and Category in the Tradition of Aristotle. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Taneiev, S.I. (1902/1962) *Convertible Counterpoint in the Strict Style*. Bruce Humphries Publishers, Boston.

Tenney, J. (1961) *Meta+Hodos.* Frog Peak Music, Oakland (Ca).

Tenney, J. and Polansky L. (1980) Temporal Gestalt Perception In Music. *Journal of Music Theory,* 24: 205-241.

Thompson, W.F. and Stainton, M. (1995-96) Using *Humdrum* to Analyze Melodic Structure: An Assessment of Narmour's Implication-Realization Model. *Computing in Musicology*, 10:24-33.

Todd, N.P. McAngus (1994) The Auditory 'Primal Sketch': A Multiscale Model of Rhythmic Grouping. *Journal of New Music Research*, 23:25-70.

Trehub, S., Schellenberg, E. and Hill, D. (1997) The Origins of Perception and Cognition: A Developmental Perspective. In *Perception and Cognition of Music*. I. Deliège and J. Sloboda (eds), Psychology Press Ltd, Hove, U.K.

Tversky, A. (1977) Features of Similarity. *Psychological Review,* 84(4):327-352.

Tversky, A. and Gati, I. (1978) Studies of Similarity. In *Cognition and Categorisation.* E. Rosch and B.B. Lloyd (eds), Lawrence Erlbaum Associates, London.

Van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (eds) (1993) *Categories and Concepts*. Academic Press, London.

Van Mechelen, I., De Boeck, P., Theuns, P. and Degreef, E. (1993a) Categories and Concepts: Theoretical Views and Inductive Data Analysis. In *Categories and Concepts,* I. Van Mechelen et al. (eds), Academic Press, London.

Vantomme, J.D. (1995) David Cope: Bach by Design - Experiments in Musical Intelligence (review of CD). *Computer Music Journal*, 19(3):66-68.

Wason, R. W. (1996) A pitch-class motive in Webern's George Lieder, Op.3. In *Webern Studies*, K. Bailey (ed.), Cambridge University Press, Cambridge.

West, R., Howell, P. and Cross, I. (1991) Musical Structure and Knowledge Representation. In *Representing Musical Structure,* R.Howell et al. (eds), Academic Press, London.

Wiggins, G., Miranda, E., Smaill, A. and Harris, M. (1993) A Framework for the Evaluation of Music Representation Systems. *Computer Music Journal,* 17(3):31-42.

Witten, I. H., Manzara, L. C. and Conklin, D. (1994) Comparing Human and Computational Models of Music Prediction. *Computer Music Journal,* 18(1):70-80.

Wittgenstein, L. (1953) *Philosophical Investigations.* Macmillan, New York.

Yeston, M. (1976) *The Stratification of Musical Rhythm.* Yale University Press, New Haven.

Xenakis, I. (1992) *Formalised Music.* Pendragon Press, New York.

Xenakis, I. (1989) Concerning Time. *Perspectives of New Music*, 27(1):84-92.