Universitat Pompeu Fabra Barcelona

Modelling Tonal Context Dynamics by Temporal Multi-Scale Analysis

Agustín Martorell Domínguez

TESI DOCTORAL UPF / 2013

Directors de la tesi:

Dra. Emilia Gómez Gutiérrez
Dept. of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Copyright © Agustín Martorell Domínguez, 2013.

Dissertation submitted to the Department of Information and Communication Technologies of Universitat Pompeu Fabra in partial fulfillment of the requirements for the degree of

DOCTOR PER LA UNIVERSITAT POMPEU FABRA,

with the mention of European Doctor.

Music Technology Group (http://mtg.upf.edu), Dept. of Information and Communication Technologies (http://www.upf.edu/dtic), Universitat Pompeu Fabra (http://www.upf.edu), Barcelona, Spain.

And so time wags on: but father Cronion has dealt lightly here
(Ulysses - J. Joyce*)*

Al Cor de Cambra Anton Bruckner.

Acknowledgements

This work, as any long-term endeavour, is at the cross of quite a complex interaction of people and circumstances, being impossible any attempt of acknowledge their mutual influences. Xavier Serra, for making possible this opportunity through the Music Technology Group. Emilia Gómez, for her support and supervision, endless discussions and patience, technical and methodological assistance, manuscript revisions and friendship. Hendrik Purwins, for his inspiring mathematical discussions on music. Perfecto Herrera, for opening the Pandora's box of music perception research. Sergi Jordà and Enric Guaus for their support. MTG-ers in general. Graham, Óscar, Moha and Mathieu. Office mates, stable and eventual, José, Dmitri, Sâsô, Juanjo, Álvaro. Justin, for his help with the manuscript. Cristina, Alba and Sònia. Lydia and admin staff. Laura Dempere and the "Ones" team in full. Special thanks to Petri Toiviainen, for making possible a winter at the Centre of Excellence in Interdisciplinary Music Research, at Jyväskylä, all the staff and graduate students there. Martin and Ibi: true warm at -25^{o}C. Fred Lerdahl, Carol Krumhansl and Craig Sapp, for their true inspiration and kind data sharing. All my family and friends in the distance. My special thanks to a group of persons through which I have understood what harmony is all about.

This thesis has been carried out at the Music Technology Group of Universitat Pompeu Fabra (UPF) in Barcelona, Spain from Oct. 2009 to Jun. 2013, and at the Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, in Finland, from Nov. 2011 to Mar. 2012. This work has been supported by an R+D+I scholarship from UPF, and by the projects Classical Planet: TSI-070100-2009-407 (MITYC), DRIMS: TIN2009-14247-C02-01 (MICINN) and MIReS: EC-FP7 ICT-2011.1.5 Networked Media and Search Systems, grant agreement n. 287711. It also received funding from the European Union Seventh Framework Programme FP7 / 2007-2013 through PHENICX project under grant agreement n. 601166.

Abstract

This work explores the multidimensional, ambiguous and temporal characteristics of tonality from a holistic perspective. The approach relies on interfacing pitch-spaces with time vs. time-scale descriptions. In this combined representation, the spatial and temporal hierarchies of tonality are evidenced simultaneously and in relation to each other. A visual exploration method is proposed for the analysis of tonal context in music works, using a simple model of tonal induction. A geometrical colouring solution, based on the topology of the pitch-space, approaches the perceptual correlation between the tonal properties and the visual representation. A relational taxonomy is proposed for describing tonal ambiguity, which leads to extending the method for the analysis of music based on tonal systems beyond the major-minor paradigm. Two perceptual studies are approached from this descriptive framework. The first study evidences the impact of time-scale in a simple model of tonal induction, and analyses the mathematical artefacts introduced by evaluations in scaled spaces. In the second study, a model of contextual instability is proposed and discussed in relation to the modelling of tonal tension. The analysis and representation methods are then generalised, through a set-class theoretical domain, in order to be applied with any pitch-based music.

Resumen

Esta tesis analiza la naturaleza ambigua, multidimensional y temporal de la tonalidad, desde una perspectiva unificada. El método propuesto parte de la conexión explícita entre espacios tonales y descripciones en tiempo y escala temporal. Esta representación conjunta pone de manifiesto la relación entre las jerarquías espaciales y temporales de la tonalidad. Utilizando un modelo simple de inducción tonal, se propone un método de exploración visual del contexto tonal en obras musicales. La correlación perceptual entre la representación visual y las propiedades tonales se aproxima a través de un modelo geométrico de coloreado basado en la topología del espacio tonal. Tras analizar la ambigüedad descriptiva mediante una taxonomía relacional propia, el método se adapta para el análisis de obras musicales basadas en diferentes sistemas tonales. Dos estudios perceptuales son abordados desde el entorno descriptivo propuesto. En el primer estudio, se pone en evidencia el impacto de la escala temporal como parámetro de un modelo simple de inducción tonal, y se analizan los artificios matemáticos introducidos por evaluaciones en espacios escalados dimensionalmente. En el segundo estudio se propone un modelo de inestabilidad contextual, y se analiza en relación al modelado de la tensión tonal. El método de análisis se generaliza, a través de una categorización contextual en set-classes, para su aplicación con cualquier tipo de música basada en pitch.

Preface

This work contributes to clarify some misconceptions, taken for granted by a substantial amount of research on computational models of tonality. It brings back some unanswered questions from a critical and renewed perspective, which sheds light into some of the most elusive aspects of tonality. The outcomes of this research have been published in a number of international conferences. Our approaches, conceived as methodological research tools, were also devised with an educational purpose in mind. The methods have been featured in a variety of undergraduate and graduate courses and seminars on musicology, music psychology and music technology.

Contents

List of figures

CHAPTER 1

Introduction

Tonality has remained a quite challenging topic of research since ancient times from today. The fascination for such a complex phenomenon permeates the whole history of music composition. Currently, the different facets of tonality are approached from a variety of disciplines, ranging music theory, aesthetics, acoustics, neuroscience, cognitive psychology, or linguistics, to cite a few. Yet it manages to evade any satisfactory explanation.

A primary goal of this work is to approach three elusive aspects of tonality *as a context*, namely multidimensionality, ambiguity and timing. While the first two issues have been at the core of every modern theory of harmony, the problem of timing is far less understood. The general question can be posed in simple terms: how the sense of tonal context arises and evolves over time, and how can we describe it? The first question to face is what we mean by "sense of tonal context". This has been mostly referred, in the so-called Western common-practice tradition, to as the concept of musical *key*.

The very concept of key is elusive, even at theoretical level. It involves referential relations between the music stimulus and a set of assumed categories. However, these categories are part of complex hierarchical relations: pitches, chords and keys are fully understandable only in relation to the others.

An obvious characteristic of any contextual information which has to be apprehended along time, is that it requires time. But, how much time? The only reasonable answer is, it depends. However, by no means it is clear on what exactly depends. The scale of observation is a critical factor in any descriptive endeavour, being tonality a particularly sensible case, since the mentioned hierarchical relations between pitches, chords and keys, are correspondingly embedded in a temporal hierarchy.

At the perceptual side of the tonal experience, the problem even worsen. Apparently identical stimuli, such as the same melody played twice, can induce different psychological responses, not only in different listeners, but in the same

listener as well. On the other hand, that the sense of key is something necessarily *induced* by an external music stimulus is far from realistic. There seems to be a considerable *intentional* factor in tonal comprehension. In addition, there is a wealth of evidence supporting the dependence of tonal perception on the listeners' musical training and cultural background.

Aside these challenges, one can still consider how to treat the *rest* of the music, that which is not understandable under the major-minor paradigm. How the concept of tonal context is to be grasped in, say, minimalistic music which uses just a pair of notes? What about contextual description of atonal or stochastic music?

1.1 The challenge

The main goal of this work is to develop a framework for the study of these challenges. To be more specific, we want a method able to capture the tonal content of music pieces, describe it at contextual level, and represent the information in human-readable ways. More importantly, the method has to facilitate a means for assessing the pertinence of the description. That is, we do not pursue an *automatic* analysis tool, but a method for reasoning about the tonal phenomena.

In particular, the basic framework should assist the inspection of: a) the multidimensional nature of the contextual categories; b) the mutual relations between categories; c) the description of any segment of music with respect to the space of categories, including ambiguity; d) the analysis of music pieces over time, including information at structural level; e) many time-scales simultaneously. Aside this, the method should be adaptable to any kind of pitch-based music, beyond the limitations of specific tonal systems. Ideally, it should provide a means for approaching analytical insights of a certain sophistication.

Arguably the most challenging aspect of this thesis is the evaluation of the proposed method. As we are concerned about the risks at the crossroads of different disciplines, a main goal of this work is to find a methodological balance with respect to computational modelling, empirical sciences and humanities.

1.2 Objectives and thesis outline

1.2.1 Objectives

1. To develop an analysis method for music pieces, able to manage the full dimensionality of the tonal phenomena at contextual level. This includes the multidimensional nature of the tonal concepts, time and time-scale.

2. To develop a proper description and representation for such information. This involves: categorical spaces, multidimensional description of each category, relevant metrics between categories, categorical ambiguity, evolution over time and temporal multi-scale description.

3. To develop interfacing solutions in two dimensions for such representations in human-readable ways, optimising the informativeness, and guaranteeing perceptual consistency between the tonal relations and the visualisations.

4. To exploit existing estimation and representation methods for assisting musicological analysis of certain sophistication. This involves the inspection of theoretical, compositional and aesthetic aspects of music works, in relation with the nature of the descriptors and representations.

5. To elaborate on methodological issues of empirical studies on tonal perception, with respect to time-scale and multidimensionality of description.

6. To develop a parsimonious model of tonal instability, able to capture features only accessible by sophisticated theoretical models of tonal tension.

7. To develop analysis methods beyond the usual major-minor paradigm, usable with any set of contextual categories.

8. To generalise the analysis and representation methods for any kind of pitch-based music. To achieve full systematisation, exhausting the complete set of sonorities of the twelve-tone system.

9. To provide off-the-shelf interfaces for tonal exploration, intended for research and educational purposes alike.

1.2.2 Instrumental objectives

1. To keep it simple. A main goal is to provide a comprehensive understanding of the involved processes and a proper interpretation of the results.

2. To guarantee the reproducibility of the method and the results, using open source solutions.

3. To approach the multidisciplinary challenge of combining musicological argumentation, perceptual evidence and information technology methods in complementary, balanced and insightful ways. Ideally, to derive conclusions pertaining to the three areas of knowledge.

1.2.3 Outline

Chapter 2 The motivation leads to a concise literature review, framing the main analysis and representational issues related to tonal description at contextual level. The structure of the review is adapted to our main approach, in terms of what we call *spatial* and *temporal* aspects of tonality. Space is used to represent the properties of tonality as a system, while the temporal dimensions (time and time-scale) index the tonal properties of actual instantiations in real music. Four problem domains are identified, namely: segmentation, description, representation and evaluation.

Chapter 3 The rationale for our method is presented. The general method is then developed and discussed. The approach relies upon a systematic multi-scale segmentation of the music piece and the analysis of every segment. Representation is solved by an explicit connection between pitch-spaces and time vs. time-scale plots (keyscapes). This is achieved by introducing a novel colouring method. A taxonomy for tonal ambiguity is proposed and discussed. The general method is adapted for scalar systems away of the major-minor paradigm, by means of new spaces of contextual categories. The method is compared in symbolic and audio domains, and it is adapted for providing human-readable quantitative measures of the tonal distances in the keyscapes.

Chapter 4 Two perceptual issues are approached from the general analysis framework. In the first study, we reanalyse empirical data obtained by continuous rating experiments to show the impact of time-scale in the evaluation of a simple model of tonal induction. We also discuss the mathematical artefacts introduced by evaluations in scaled spaces. In the second study, we propose a parsimonious model of tonal context instability, and we discuss it with respect to existing models of tonal tension.

Chapter 5 We adapt our general method to a set-class theoretical domain, extending the systematisation to the description stage through the concept of class equivalence. This description domain complements the previous estimation methods, providing an objective description in unambiguous terms. The new dimensional requirements lead to adapting the visualisation and interfacing solutions. A variety of compact descriptors are derived from the primary information, facilitating the inspection of any possible contextual sonority and the mining of patterns of certain sophistication, not revealed by standard computational methods.

Chapter 6 The interfacing possibilities of the method are described in three different tonal exploration tools.

Chapter 7 General conclusions of the thesis.

CHAPTER 2

Framing tonal context

... perhaps, as you said, Bach never even had accompaniment in mind at all ...
(Sonata for Unaccompanied Achilles - D. R. Hofstadter*)*

2.1 Introduction

2.1.1 On context

The term *context* is defined in varied ways, such as: "the interrelated conditions in which something exists or occurs" (Merriam-Webster dictionary), "the situation within which something exists or happens, and that can help explain it" (Cambridge dictionary) or "the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood and assessed" (Oxford dictionary). The last two definitions point to a relevant aspect of any context: it provides a means for explaining or understanding something. In linguistics, the study of the dependence of context in conveying meaning pertains to the field of *pragmatics* (Mey, 2001). While semantics looks at the conventional meaning in a given language, pragmatics studies how the assessment of meaning depends not only on structural and linguistic knowledge (e.g., grammar or lexicon) at both sides of the communication channel, but also on the context of the utterance and the inferred intents of both the speaker and the listener, among other factors. Pragmatics thus studies the relation between communication and ambiguity.

Contextual information can have notable effects in perception. Fig. 2.1 depicts the Adelson's classic "checker shadow" illusion[1]. The squares A and B in the left pane are perceptually associated to the *black* and *white* categories respectively, although they have exactly the same shade of grey, as it is proved in the right pane. The perception of auditory parameters of musical interest is also subjected to substantial degrees of ambiguity depending on the context. Tonality constitutes a particularly challenging phenomenon in this respect. As

[1] http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html.

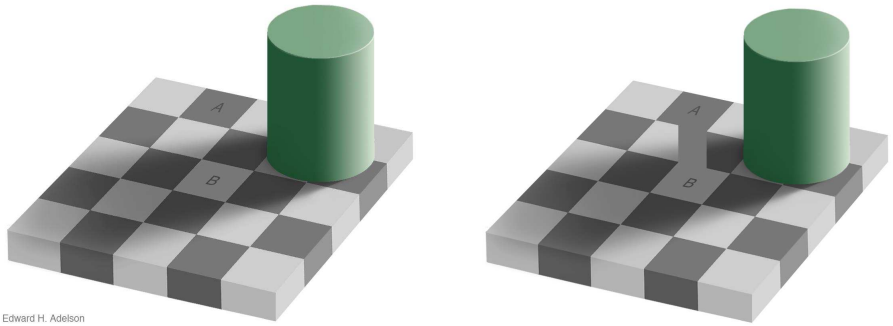

Figure 2.1: Checker shadow illusion. ©1995 by Edward H. Adelson. Reproduced by permission of the author.

an example, let's consider a simple G-C chord progression, connected by standard voicing and played at a slow tempo. In the absence of additional stimuli, the C chord would convey a sensation of resolution from its dominant. However, by preceding the C with the sequence G-D^7-G, the C would sound as a subdominant departure from a G major context, provided that *enough* time has been devoted to establish the contextual stability. It is worth mentioning that the distinct perceptual qualia applies to the C chord, but also to the G-C sequence. That is, the context influences the relational connections between the elements in the sequence, to the extent of *inverting* the sensation (from arrival/relaxation to departure/tension). That the same chord sequence would be perceived as a C:V-I or as a G:I-IV, is both contextual and timing dependent, and of course would also depend on what follows. Simple cases such as this example can be perceived substantially different when played at different tempi (Farbood et al., 2012), or by modifying the relative durations of the elements in the sequence.

2.1.2 Tonal description

Many different definitions have been done for such a complex concept as tonality. If we were to choose an appropriate definition, according to the scope of this work, we would take arguably the broadest one, as stated by Hyer: "[...] it refers to systematic arrangements of pitch phenomena and relations between them" (Hyer, 2013). The study of tonality is approached from many different disciplines, and we will not attempt to survey the enormous amount of research on tonality in general. Instead, in this chapter we will just point to those aspects of tonality and related scientific methods which concern to our specific problem and proposal. Since we will approach the topic from three different perspectives, and with the aim of improving focused readability, most of the details are postponed to the background and discussion sections at the different chapters of this document. This applies particularly to Chapter 5, in

which we reconsider tonal context from a different theoretical domain.

The topic of this work is constrained to tonality. No other features, such as rhythm, timbre or dynamics, will be involved. This endeavour, despite enormously simplified, is far from trivial. The tonal system is hierarchical among several description levels (e.g. pitch, chord, key), each of them sensible to the scale of observation, some levels are multidimensional, and the interpretation of events and contexts are mutually dependent. In order to provide a detailed discussion of our results, taking care of the interpretative artefacts along the way, we will reduce further the scope of our inquiry, by focusing (almost) only in the characteristics of tonality *as context*. The problem of describing tonal context, however, goes far beyond labelling the key of music segments. Three all-embracing aspects have to be taken into account:

1. Tonality is ambiguous in general, as several tone centres or scales can be induced, to different degrees, simultaneously. It is not always appropriate (sometimes, it would be simply absurd) to associate a temporal point or a segment of music with a single key. Tonal implication is most of the times inseparable from a certain degree of uncertainty. At the methodological basis of the tonal cognition research is a concern for ambiguity (see below). From theoretical and analytical perspectives, ambiguity is also natural to tonality, as it is reflected by the terminology. Temperley, for instance, refers the term "tonal implication" to the key implied by a given stimulus, but he also defines "tonal ambiguity" or "tonal clarity" as the degree to which the stimulus implies one key as opposed to several, while "tonalness" grades the stimulus in terms of the ambiguities allowed by a given tonal language (Temperley, 2007). Moreover, the diachronic nature of music listening allows legitimate retrospective revisions of previous events, even in cases of little ambiguity (Temperley, 2001).

2. Tonality is not *static* in general, as it constitutes a central device for inducing a sense of *movement*, *change* or *direction* in the dramatic development of a piece. The study of the modulation is at the core of any theory of harmony. While pivotal chords and cadences are central in the modulation procedures for establishing a target tone centre, equal importance is often given to the means for denying or not conveying it so straightforwardly. This way, Schoenberg highlights the usage of "neutralisations" of certain scale degrees, and the role of "vagrant" harmonies (Schoenberg, 1969) in providing unstable or ambiguous references. So, the study of modulation is not just about *moving* from one tone centre to another in the most efficient manner, but also about managing the uncertainties involved in the process. In many respects, tonality is best described in terms of *processes* than as sequences of *states*. The shift

from functional meaning to functional progression is at the heart of the neo-Riemannian harmonic tradition (Lewin, 1987).

3. Both previous aspects are interdependent through hierarchical relations, and sensitive to the scale of observation.

The previous references to movement are more than shallow analogies. Much of the human conceptualisation about the world rely on spatial and temporal references, which are ubiquitous to sophisticated metaphorical levels in all languages. These "orientational metaphors" are not arbitrary, but they seem to have a basis in our physical and cultural experience (Lakoff & Johnson, 1980, p. 15)[2]. The relation between the spatial representations of tonality and the experiencing of music along time, leads to the usage of *movement* as a natural term and a topic of inquiry. Explicit spatiotemporal relations in tonality have been studied from music theory (Lerdahl, 2001) and psychology (Firmino & Bueno, 2008). An overview of these spatial and temporal perspectives of tonality follows.

2.2 Tonality and spatial models

The spatial representation of music has a long tradition in music theory, music analysis and music psychology, by relating spatial distances with their conceptual (Lerdahl, 2001), aesthetic (Bonds, 2010; Treitler, 1997) or perceptual (Krumhansl, 1990) counterparts. As tonality is concerned, some keys are said to be *closer* than others in relation to a given tone centre, and this also applies to chord or pitch relations within a given key. While some of these distances have a direct connection with the acoustic domain (e.g. pitch proximity) (Bregman, 1990; Shepard, 1982), others are said to be *cognitive* distances (Krumhansl, 1990) or respond to sophisticated theoretical concepts (Tymoczko, 2012).

Since the adoption of the tempered scales, "regional circles" (Heinichen, 1728)[3] became conceptual instruments for reasoning about modulation. In Heinichen's circle, which alternates the major and the relative circles of fifths, close modulations are represented by movements to adjacent positions. The practical limitations of single circles, which did not reflect many aspects of the tonal practice, were approached by many theorists[4]. The proximity between relative and parallel relationships was solved, among others, by Weber's lattice (Weber,

[2]The usage of Lakoff and Johnson's embodied metaphorical schemas of "location", "path" and "attraction", are recurrent in tonal terminology.

[3]The term "region" is recent in theoretical literature, the reference here is from Lerdahl (Lerdahl, 2001, p. 42-43). Heinichen named it "Musicalischer Circul" (Heinichen, 1728, facsimil, p. 837).

[4]See (Lerdahl, 2001) for a concise review.

1821). Among the best well-known chordal spaces, stands Riemann's *Tonnetz* (Riemann, 1893) and its modern revisions (Tymoczko, 2012, for instance), while pitch distance models are mostly related to some variants of Shepard's "melodic map" (Shepard, 1982). The interest in unifying regions, chords and pitches in single representations has been recently approached from both psychology (Krumhansl, 1990, for a review of pairwise solutions), mathematics (Chew, 2000) and music theory (Lerdahl, 2001).

Given its relevance for our own proposal, in this section we will review some spatial representations of tonality, with particular focus on Krumhansl and Kessler's space of inter-key distances (Krumhansl, 1990, for a comprehensive discussion). The rationale is that this space articulates many of our challenges in appropriate ways:

1. It defines a usable set of tonal categories at the contextual level (key).

2. It captures hierarchical relations among three categorical levels (key, chord and pitch).

3. Distance in space quantifies perceptual differences.

4. Orientation in space accounts for tonal relationships (functional harmony).

5. Points in the space represent higher dimensional information (tonal hierarchies).

6. Projections in the space can manage tonal ambiguity.

7. Movement in space accounts for the passing of time.

2.2.1 Cognitive psychology

A proper interpretation of the characteristics and limitations of the tonal spaces, requires to delineate some aspects of the scientific approaches from which they are derived. At the methodological basis of cognitive psychology is a concern for quantification of the perceptual attributes, which are approached by a variety of indirect measurements. In order to argue that the observations are related to the underlying system, and are not a consequence of the methodology, different approaches should converge on similar results. The observable variables are considerably complex, making necessary to delimit the modalities of response in order to be coded with confidence. In addition, choosing the musical stimuli and the participants for listening experiments is particularly challenging in the case of tonality. Given the individual and temporal variability of the subject's responses under apparently identical stimuli, any general

conclusion of a quantitative nature has to be taken cautiously. All these considerations have led to results generally limited to basic stimuli, tested under non naturalistic listening conditions.

From the empirical point of view, Krumhansl and collaborators' systematic approach to tonality is founded upon a principle of perceptual tonal stability. The terms in which this stability is described is supported by two general assumptions in psychology research about reference points. First, that elements can be rated in terms of "goodness" (Garner, 1970) or "typicality" (Rosch & Mervis, 1975) with respect to a given category, which provides a quantitative hierarchical ordering of the elements. Second, that this ordering influences measures of perceptual or cognitive processing (Rosch, 1978). The methodology described next is based upon these two principles, which are closely related with notions of tonal stability in musicological literature[5]. The term *tonal induction*, as used in cognitive psychology, is referred to as the development of a sense of key in listeners, and how it evolves in time (Krumhansl, 2004). This particular definition of tonal induction is quite aligned with our primary question in this thesis (see Chapter 1).

2.2.2 The probe-tone methodology

The probe-tone methodology, introduced in (Krumhansl & Shepard, 1979), provides quantification of the perceived tonal hierarchies in a systematic way. The tonal hierarchies are established by measuring the perception of single musical tones (named "probe tones", and realised in sound so as to be perceived irrespective of the octave) when they are listened under the influence of a given tonal context. Contexts are referred to as the musical stimuli sounded immediately before the probe tone, and they are assumed to induce an unambiguous sense of key (e.g. incomplete diatonic scales or perfect cadences). The probe tones are taken systematically to cover the chromatic circle, and a 12-dimensional vector of "probe-tone ratings" is obtained for each participant. Fig. 2.2 depicts the average results for unambiguous major and minor keys (rooted at C), known as Krumhansl and Kessler's key profiles (hereafter, *KK-profiles*). These profiles resemble some theoretical formalisations of the tonal hierarchy, such as the Lerdahl's "basic space" (Lerdahl, 2001, p. 47), and several modifications of the KK-profiles have been proposed for improving them in specific analytical (Temperley, 2001) and perceptual (Aarden, 2003) scenarios.

[5]Krumhansl cites (Meyer, 1956, p. 214-215). The specific definition of "tonality" (quotations as published) given by Meyer, responds almost exclusively to the concept of tonal hierarchy which supports the psychological reference points assumptions. He refers to the concepts of hierarchy and stability ("activity and rest") of tonality *as a system*, not in reference to particular instantiations in music.

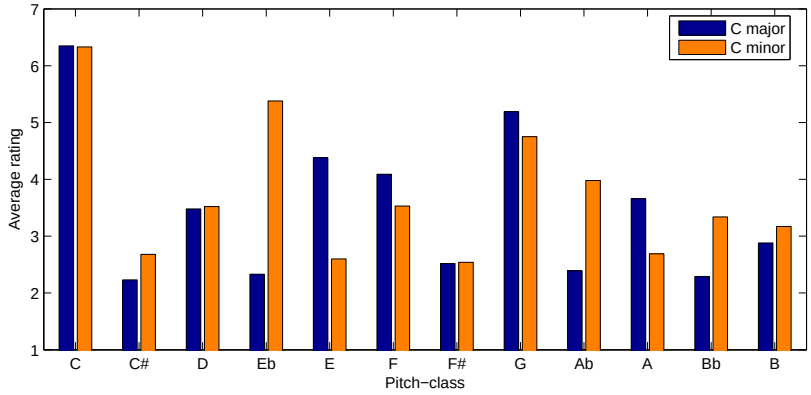

Figure 2.2: Krumhansl & Kessler's key profiles.

2.2.3 Dimensional scaling

A further step by Krumhansl and collaborators was to find suitable low-dimensional representations of the results. In a first stage, the key profiles for all keys were computed by ring-shifting the major and minor prototypes so as to cover the chromatic circle of pitch-classes. Then, the inter-key correlations were computed, and the matrix of dissimilarities fed a nonmetric multidimensional scaling algorithm (Shepard, 1962). An optimal Euclidean solution, minimizing the stress between the dissimilarity values and the spatial distances, was found in 4 dimensions (hereafter, 4-D *KK-space*). In Fig. 2.3 (top-left), the points of such solution are depicted[6]. Two circular structures stand out: 2 dimensions account for the double circle of fifths, while 2 additional dimensions account for the relative and parallel relations.

Given the homeomorphism between a ring torus and the Cartesian product of two circles, the solution points at the 4-D KK-space can be thought of as belonging to a 3-D toroidal surface (hereafter, 3-D KK-space), as depicted in Fig. 2.3 (top-right). The toroidal surface can be unfolded in two angular dimensions, as in Fig. 2.3 (bottom), where the opposite edges in the figure are identified (hereafter, 2-D KK-space). Although each dimensional reduction introduces metric distortion, the 2-D representation results convenient for visualisation purposes.

Similar spatial arrangements of keys are found from other disciplines, although not always explicitly formalised as toroidal structures. From music theory, Weber's lattice (Weber, 1821) and Schoenberg's "charts of the regions" (Schoen-

[6] 4-D coordinates taken from (Krumhansl, 1990, p. 42). Points have been left unlabelled with key names, because of the exact projection of triplets (corresponding to keys rooted at the notes of augmented triads) in the last two dimensions. For a labelled but distorted plot, see (Krumhansl, 1990, p.43).

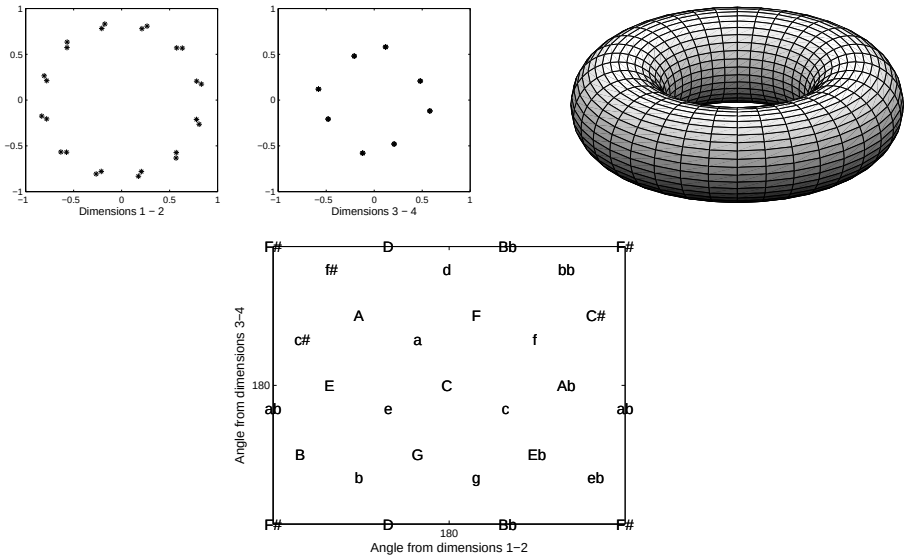

Figure 2.3: Top-left: 4-D KK-space. Top-right: 3-D ring torus. Bottom: 2-D KK-space.

berg, 1969, pp. 20, 30) distribute the keys in quite a similar way, and they can be thought of as toroids if extended in both dimensions so as to close the circles of fifths. Explicit toroidal shapes are proposed in (Werts, 1983) and (Lerdahl, 2001). Virtually identical key distributions arise from other multidimensional scaling solutions (Kohonen, 1997) and from Fourier analysis of the key profiles (Krumhansl, 1990). Toroidal structures emerge from machine learning techniques applied to the analysis of music corpora, considering minimal theoretical assumptions (Purwins, 2005)[7]. The topological and dimensional convenience of using toroidal spaces seems clear for representing the elementary relations of the tonal system.

Some of these spatial models account for several hierarchical categories in the same space. Lerdahl's "chordal-regional space" allows the inspection of the chordal relations within each key, while maintaining the regional picture at sight (Lerdahl, 2001). A geometrical solution for simultaneous management of pitches, chords and keys is given by Chew's "spiral array" (Chew, 2000). The representation of the contextual level, named "center of effect", however, cannot be properly visualised in two dimensions[8]. Other multilayer methods, although not yielding visual representations of distances are proposed from connectionist approaches to tonality using both wired (Tillmann et al., 2000)

[7] Namely, octave equivalence and equal-tempered chromatic division of the octave.

[8] In the spiral array model, pitches are located in a spiral, chords are the geometric summary of the pitches, and keys are the geometric summary of the tonic-dominant-subdominant triads. Thus, the center of effect lies *inside* the spiral structure.

and unsupervised (Tillmann et al., 2003) models.

The limited scope of these solutions to the basic major-minor paradigm in the Western tradition, have raised a number of variants. Stylistic and historical issues have been called for the construction of parallel-mixture[9] and chromatic spaces, as well as non-triadic spaces for accommodating the properties of other common scalar systems[10] (Lerdahl, 2001). Non-linear multidimensional scaling solutions have been proposed for modelling the local plasticity of the spaces, which allows the representation of stylistic differences (Burgoyne & Saul, 2005).

Since the KK-spaces are assumed to represent the internalised knowledge about tonal hierarchies, they have been used for testing models of tonal induction. The rationale is that any segment of music can be *projected* in the space, so as to find the closeness of the segment to the represented tonal categories. Assuming the 4-D KK-space as a fixed scaffold, a multidimensional unfolding technique (Coombs, 1964) can be applied to locate the point in the low-dimensional space best fitting the relative distances to all the tone centres (Krumhansl, 1990, for a number of use cases). The space can thus be used for visual tracking of the tonal estimations as a single point, often referred to as tonal *centroid*. Certain modalities of tonal ambiguity can also be captured by this representation, as the tonal centroid can occupy locations *between* the keys.

An interesting approach to ambiguity representation is provided by the use of toroidal Self-Organising Maps (hereafter, SOM) (Kohonen, 1997). The output of these neural networks is a collection of *codebook vectors* connected in a topological arrangement. The codebook vectors represent *prototypes* and the topological structure imposes an ordering during the training. When trained with the major and minor KK-profiles, a toroidal SOM produces an arrangement of the keys virtually identical than the KK-space. In addition, the SOM fills the space between tone centres with interpolated versions of the key vectors. The SOM thus constitutes a joint representation of the tonal categories (KK-profiles), their mutual relationships (their distances), and an explicit interpretation of the space in between. A music segment can be projected in the trained SOM by comparing its pitch-class profile with every codebook vector. The resulting SOM *activation* constitutes an intuitive visualisation of the tonal content of the segment, in terms of the categories scaffolding the space (Toiviainen, 2008). This method can represent any kind of ambiguity with respect to such categories. By associating each point in the space with a vector, the SOM can be understood as a vector field, while its activation can be seen as a scalar field.

Conveniently coded segments of music or perceptual ratings can be projected

[9] Closer to the usage of the major-minor ambiguity in the Romantic period.
[10] Including the hexatonic, octatonic and Scriabin's *mystic* modes.

in a tonal space, and being observed in terms of the properties embedded by the space. This way, the space serves as a frame of observation under a particular *listening* modality. The activation of the space would be an indicator about how humans would perceive the music segment *if* they were categorising the tonal stimuli in similar terms. While this scenario can be far from a realistic cognitive model in general, the potential of observing music under particular *listening* perspectives is suggestive for analytical purposes. We will discuss these capabilities and limitations in Chapters 3 and 4.

2.3 Tonality and time

Any contextual level of description is closely linked to the scale of observation, a problem which worsen when the phenomenon under study has perceptual implications and the scaling dimension is time. The obvious limitations of the spatial representations above concern to the time dimension, and they are two-folded. First, the uncertain duration of the segment required for capturing such an elusive concept as key. Second, the evolution of the context as music unfolds in time. With respect to analytical inspection or algorithmic implementation, both issues can be observed as a *segmentation* problem.

2.3.1 Duration of the segment

The first issue is the most challenging, as contextual information is not constrained to any temporal duration in general. In Krumhansl and Kessler's key-finding algorithm (Krumhansl, 1990), any segment of music can be used as input. While this makes sense from the point of view of the tonal induction modelling, it presents many problems when it comes to evaluate the algorithm's output in terms of keys. It is clear from the perceptual studies that the sense of key depends (but presumably not suffices) on the pitch-wise evidence included in the segment, as it is clear that the sense of key is ambiguous in general. It is also evident that harmonic analysis, as practised by music theorists or musicologists, diverges in many respects from the concept of tonal induction as seen from a perceptual standpoint[11]. The challenge is thus to asses the pertinence of a given segment of music as being representative of the considered concept of tonal context[12]. Most of the research related with tonal context estimation, thus, approach the evaluation as a classification problem: it is assumed the

[11] Many examples, for instance, in (Krumhansl, 1990; Temperley, 2001).

[12] The confusion between the concepts of *key* (of a given passage) and *key of a musical work* abounds in the literature about key-finding methods. Despite critical voices have been raised (Wiggins, 2009), and it has been warned from certain research communities (Bernardini et al., 2009; Vinet, 2007), basic misconceptions of this kind are replicated throughout. This extends to some evaluation standards, such as the *audio key detection* task at Music Information Retrieval Evaluation Exchange (MIREX). http://www.music-ir.org/mirex/wiki/2013:Audio_Key_Detection.

existence of a *truth* (the *correct* key) taken from a set of categories (the key names), and the algorithm succeeds when its output agrees with the *truth*. It is in these terms, constrained by specific evaluation perspectives applied to specific music stimuli[13], that some general assumptions are often taken for granted with respect to the required duration of the music segments for conveying the sense of key. The segmentation of music according to rhythmic or metric considerations (e.g. bar level) are among the common choices, particularly from music theoretical approaches (e.g. Lerdahl, 2001; Temperley, 2001).

These durations are often assumed from perceptual and cognitive constraints. The short-term memory limitation, agreed about a range of 2-8 seconds, is perhaps the most cited argument, as for instance in (Toiviainen & Krumhansl, 2003). Most of the research related to short-term memory, however, has not addressed the specific problem of the sense of key. Very few studies have attempted this particular issue in a systematic way with respect to time. In (Leman, 2000), the empirical data from (Krumhansl & Kessler, 1982) was reused to test a model of tonal induction based on a leaky memory model. By systematic manipulation of the decay constant of the memory, the best fitting with the empirical data agreed within the short-term memory range conventions. However, the music stimuli was too limited for deriving general conclusions, since it consisted of chord sequences with a maximum length of 9 chords, so the longest stimuli was already about the range assumed by convention. In addition, the empirical data reused by Leman was obtained from a stop-and-rate methodology, which calls for prudent interpretations with respect to time processing in realistic listening conditions.

2.3.2 Context over time

The second temporal issue concerns to the evolution of the tonal context along time, which applies to any music stimuli, but it is particularly suited for music which involves modulation. One can distinguish two main segmentation approaches, borrowed from signal processing and auditory modelling respectively. The first method consists of applying a *sliding window*: a temporal segment of a fixed duration, which is shifted in time so as to cover the whole stimulus systematically. Two parameters are involved: the temporal duration, often referred to as the *window size*, and the temporal shift between two consecutive windows, namely the *hop size*. The terms *window*[14], *segment* and *frame* are often used as synonyms. The second approach is based on a leaky memory

[13] For instance, the evaluations using Bach's *Das wohltemperierte Klavier* are ubiquitous in the key-finding literature. The pertinence of these pieces as *prototypes* of the tonal practice in general is hardly sustainable.

[14] The term *window* or *windowing* applies to the segment's duration, but also refers to the specific *shape* which casts the information within the segment. At the segmentation level discussed here, almost every sliding-window-based approach uses plain chunks of the stimulus, which is referred to as *rectangular* windowing.

model: a temporal *buffer* fed with the music stimulus, which progressively *forgets* the past content as new evidence arrives. The usual implementation consists of a sliding window of a fixed duration shaped by some decay function, which weights the information according to its temporal recency. The usual parameter of this approach is the decay factor of some exponential function, which accounts for the rate of forgetfulness.

2.3.3 Temporal multi-scale approaches

The problem of the segment's duration has been approached from a temporal multi-scale perspective. In (Krumhansl & Kessler, 1982), growing contexts were considered for capturing the sense of modulation, but the results were not discussed with respect to the duration of each context. In (Vos & Leman, 2000; Vos & Van Geenen, 1996), although not stated explicitly in temporal terms, a "parallel processing" method examined both scalar and chordal information simultaneously for estimating the key. In (Leman, 2003) two echoic memories were used simultaneously, accounting for global and local integration of the stimulus. In (Lerdahl, 2001) several simultaneous "paths in pitch-space" are analysed, considering the different "time-span reduction" levels. In (Purwins, 2005), a three-tier method segments the music at note, beat and bar levels, and the key is derived from the joint estimations. For mapping estimations in Chew's "spiral array", several predefined windows account for the pitches, chords and keys, a method that has been used in (Chew, 2006) for estimating the key boundaries in music pieces. Two time-scales are required for the related problem of the joint estimation of chords and keys, as applied by some machine learning approaches (Papadopoulos, 2010). In (Janata et al., 2002), two time-scales are used for simultaneous tracking of key estimations in a SOM. All of these methods assume a small (2 or 3) set of predefined time-scales. A main limitation of these approaches is evident at the representational level: the simultaneous tracking of the estimations computed at several temporal resolutions calls for a dimensional compromise when it comes to represent the results in human-readable ways, much particularly for describing the inherent ambiguity of the estimation.

Sapp proposes a systematic approach to the representation problem with respect to both temporal issues. He introduces the concept of "keyscape" for estimating and visualising the key of every possible segment in a music piece (Sapp, 2005). The method considers many sliding windows, of sizes ranging from fractions of a second to the whole duration of the piece, and compares each segment with a set of key profiles. Each key is coded by a colour, and a two-dimensional triangular plot organises the information: the x coordinate represents time, and the y dimension represents time-scale. Each coloured point in the keyscape represents the best estimation of a segment, indexed by its centre location in time and its duration. This representation provides

visual access to a systematic "parallel multiple-analysis"[15]. Implementations of keyscapes have been proposed for analysing music from MIDI (Sapp, 2005) and audio (Gómez, 2006) encodings.

2.4 Tonality in space and time

Beyond the metaphorical association with movement, the spatial and temporal conceptualisations of tonality are connected by hierarchical relations. When a music segment is projected as a centroid in KK-space, the centroid location in space represent a summary of the segment's tonal content, but it is *associated* to both the segment's position in time and to its duration. A larger segment around the same time point could naturally result in a different projection, as there exists different levels of key[16]. This is actually the information embedded by the keyscapes, although Sapp only associates each segment with a key label without spatial implications[17].

The explicit connection between space (*what*) and time/time-scale (*when/how long*), observed at different temporal resolutions, is at the core of some hierarchical theories of tonality. Lerdahl distinguishes between two different but complementary hierarchical perspectives. On the one hand, "event hierarchies" are referred to as the pitch references[18] *operating* in a particular piece, and to their relations with the rest of the pitch material. Event hierarchies are related to the "prolongational" aspects of the Generative Theory of Tonal Music (hereafter, GTTM) (Lerdahl & Jackendoff, 1983), and they serve to represent the tonal structure of the piece as a tree, whereby any event is related to the stable references in the piece. On the other hand, "tonal hierarchies" are required to define the stability conditions for building such trees. That is, the tonal hierarchies pertain to the tonal system itself, irrespective of the particular instantiations in music. As an oversimplified analogy, one could say that the event hierarchies are to time as the tonal hierarchies are to space, being both of them interdependent. The relationships between event and tonal hierarchies are at the foundations of Lerdahl's model of "tonal tension" (Lerdahl, 1996), elaborated in full in his Tonal Pitch Space theory (hereafter, TPS) (Lerdahl, 2001). An overview on tonal tension and the details of Lerdahl's model will be postponed, for readability reasons, to the background and discussion sections of Chapter 4.

[15] The term has been proposed, among others, in (Lerdahl & Jackendoff, 1983; Temperley, 2007), in reference to a space of multiple analytical hypotheses.

[16] From the main key of the piece, to intermediate keys, to brief tonicisations (Temperley, 2001, p. 187).

[17] Sapp actually embeds some specific *spatial* relations between keys in his colouring method. We will elaborate on this issue in Chapter 3.

[18] The usage of the term "pitch" here refers to the whole pitch system in general. In the most general usage, a "pitch reference" can pertain to the domain of pitches, chords or keys.

2.5 Conclusions of the chapter

The main characteristics and limitations of the previous approaches with respect to our topic of study (tonal context dynamics) can be organised as four problems, namely: segmentation, segment's content description, representation and evaluation[19]. A digested summary follows.

2.5.1 Segmentation

1. The choice of the segment to be analysed is critical for describing contextual information, as the optimal time-scale of observation is unknown in general.

2. The description over time is approached by sliding windows or leaky memory models.

3. Multi-scale methods can be used for parallel processing at several resolutions.

4. The time-scales are mostly chosen so as to optimise specific evaluation methods for specific music stimuli. The use of broad conventions or trial-and-error tuning methods are generalised.

5. Very few systematic studies on tonality have been focused on time-scale properly.

6. The concept of keyscape provides a fully systematic segmentation method.

2.5.2 Description

1. Tonal context plays a fundamental role in the complex hierarchical relations in most tonal systems.

2. Many of the key-finding methods operate under the so-called *bag-of-frames* approach, whereby a single estimation is associated to a music segment as a whole.

3. The description of any arbitrary segment of music in terms of a single key is a misleading approach in general. A simple counterexample is a passage which modulates from one key to another.

4. The description of any arbitrary time point in music in terms of a single key is a misleading approach in general. A simple counterexample is a

[19] Although description and segmentation cannot be regarded as separated problems in music analysis (Cook, 1987, p. 146), it can help to clarify the limitations at this surveying stage. We will elaborate this point along this work.

short tonicisation, which can be understood and perceived at the short-scale or embedded in a larger context. Contexts are embeddable by definition.

5. The ambiguity of the tonal phenomena is best described by multidimensional data.

6. The probe-tone methodology provides multidimensional account of the perceived tonal hierarchies.

7. The sense of key is usually described by some similarity measure with respect to every category (key), which constitutes the lexicon for communicating the results.

8. Most of the methods are limited to key estimation for the major-minor paradigm, generally under loose interpretations of the term *key*.

9. The major-minor paradigm cannot describe pitch-based music in general. Even for the so-called Western tradition, a description in terms of major-minor keys is clearly insufficient: modality in rock, jazz and folk; ancient (pre-tonal) repertoire; symmetric modes; minimalistic or electronic styles using limited pitch material; polytonality; atonality. Not to mention non-Western musics.

2.5.3 Representation

1. The multidimensional descriptions (e.g. from profiling methods) provide access to the ambiguous nature of tonality.

2. The resulting data structures, such as 12-dimensional profiles or 24-dimensional key strengths, are not practical for human-readable applications.

3. The analysis of descriptions over time requires dimensional reductions, which imply the loss of information of potential relevance.

4. An insightful analysis of tonal context requires the characterisation of the segments *and* their mutual relations (e.g. functional analysis). Context is a relational concept.

5. Pitch-spaces solve some human-readability issues, by reducing the dimensionality of the output space.

6. By projecting the estimates in pitch-spaces, some ambiguity modalities become understandable. Multidimensional unfolding provides an intuitive representation of the estimates in relation to several keys simultaneously, using low-dimensional spaces.

7. The projection method provides partial access to the tonal dynamics, by means of trajectories in space. However, this representation lacks of an explicit temporal information.

8. The unfolding method allows the representation of estimates at several time-scales simultaneously.

9. In cases of extreme ambiguity, the unfolding method results in irrelevant information.

10. The SOM approach provides a human-readable information about the estimation of a single frame. Ambiguities of any kind are properly described and are understandable by human observers. It provides access to a *continuous* pitch-space.

11. The SOM approach extends the concept of *space* to the concept of *vector field*, which jointly represents the *meaning* of each position (its codebook vector) and the relationships between all the positions (distance).

12. The projection of a music segment in the SOM provides a *scalar field* (the SOM activation), which represents the segment with respect to the categories scaffolding the space. The SOM thus operates as a human-readable observation framework of the music stimulus, in terms of the tonal categories *and* their relationships.

13. The SOM approach, however, exhausts the screen's dimensional constraints, so it is limited to a single analysis frame. The time dimension is absent. The representation of music over time can only be approached by animation, showing consecutive frames as a temporal sequence, that is, realising the time dimension *in* the observer's time. This gives intuitive access to the tonal dynamics when visualised along the corresponding sound.

14. Animations of the SOM activation, on the other hand, quickly overload the short-term memory of the observer. They are thus insufficient for conveying relational information in the long run. However, it is often of interest to analyse far-reaching relations over time.

15. Any feasible animation of the SOM would make use of a fixed time-scale.

16. The keyscape solves the representation with respect to time and time-scale. However, a single perceptual dimension (colour) is left to describe each segment.

17. The colour coding proposed by Sapp can represent relative distances between estimates in pitch-space, but only in some directions (hue along the circle of fifths and brightness for the relatives). The double circularity of pitch-space is misrepresented.

18. The Sapp's colouring method does not allow for capturing ambiguity between keys. In addition, the observer has no cue about the confidence of the estimation, which in many cases can be misleading or irrelevant.

19. The triangular shape of the keyscapes, a result of the non-overlapping segmentation policy, lacks a proper temporal alignment of segments across scales, which makes indexing the data challenging in the time domain.

20. The keyscapes provide a bird's-eye view of the complete piece, but the visualisations are conceived as the mere output. They lack any interfacing possibility for exploring the missing information (resulted from the dimensional reduction) or the actual music behind each segment. They provide visual access to areas of potential interest, but no means for testing them.

2.5.4 Evaluation

1. The misuse of the concept of *ground truth* abounds in the key-finding literature, and this extends to some evaluation standards.

2. Annotation of a corpora in terms of tonal context can lead to legitimate disagreement among different annotators. Even in cases of agreement, its a very costly task.

3. Empirical annotations of the sense of key, exemplified by the probe-tone methodology, are extremely expensive to collect, and their availability is rather scarce.

4. Empirical approaches to tonal cognition are challenging and the results are noisy. The experimental settings, thus, sacrifice any sophistication in the description, so as to guarantee data consistency. Only elementary tonal concepts have been modelled properly.

5. Data obtained from *naturalistic* listening conditions are very scarce for methodological reasons.

CHAPTER 3

Tonal representation

Du siehst, mein Sohn, zum Raum wird hier die Zeit
(Parsifal, Act I - R. Wagner*)*

3.1 Introduction

In this chapter, a framework for tonal context exploration will be elaborated. First, we will present the rationale and motivation for an integrated solution to some of the limitations discussed in Chapter 2. Second, the proposed method will be elaborated. Two novel contributions will be introduced: a) a geometrical colourspace which facilitates the access to the multidimensional and ambiguous nature of the description; b) an interfacing approach for tonal context exploration, which enhances the informativeness of the representation. Third, the method will be used to illustrate examples of analysis, focusing the ambiguity and multidimensionality of description from a temporal multi-scale perspective. Along the way, we will discuss some spatial and temporal aspects of tonal summarisation, the consideration of contextual information as a matter of stability in time and time-scale, the usage of different tonal spaces as frames of observation, and the suitability of the method for symbolic and audio domains. The applied context of the chapter will be mainly analytical, with emphasis on the interpretative issues (benefits and pitfalls) of the method.

3.2 Rationale for an integrated solution

Considering the individual advantages and drawbacks of the different tonal models presented in Chapter 2, it is clear that the limitations of a given method are partially solved by the benefits of others. Dealing with the representation problem, for instance, a music segment projected in a SOM provides a human-readable high dimensional description of its tonal content, and a proper account of the tonal ambiguity (Toiviainen, 2008). However, this representation is limited to a single analysis frame, and it has no temporal information at all. At

the other side of the problem, a keyscape provides a hierarchical overview of the piece, making explicit both temporal dimensions (time and time-scale) (Sapp, 2005). However, this comes at the price of reducing the description of each segment to a single perceptual dimension (colour), missing the multidimensional and ambiguous information. In this chapter, the different limitations will not be approached in isolation, but simultaneously as a mutual complement. A brief overview of the proposed solution follows.

1. Without prior information about the music to be analysed, the temporal scale of analysis is considered as a parameter. To achieve so, a multiple-time-scale segmentation method is adapted from Sapp's keyscapes. This provides both a means for systematic analysis and the main indexing for the data. This last aspect, indexing, is conceived as a visualisation problem, and is implemented by time vs. time-scale plots.

2. With respect to the description of each segment, a simple key estimation technique is firstly elaborated. The goal at this stage is not to improve existing key-finding methods, but to get the most representational benefit from a simple method, and to provide a comprehensive interpretation of the results. Two types of tonal ambiguity are approached in relation with the confidence and meaningfulness of the description.

3. With the aim of exploiting the full informativeness of the different representation methods, an interfacing mechanism between keyscapes and pitch-spaces is designed. This consists of introducing a novel perceptually-informed colouring method, which in addition addresses the description of some ambiguity modalities.

4. The maximal frame-based informativeness is achieved by projecting the music segments in a SOM. This provides a human-readable account of the multidimensional nature of the estimation, managing even the extreme cases of ambiguity. The SOM representation also serves as an indicator of confidence.

5. The information provided by the different representations is linked in a user interface, which forms the core of the integrated solution[1]. The goal is to provide an intuitive means for exploring a high-dimensional space[2], in terms of analytical pertinence with respect to tonality.

6. This multiple-viewpoint method provides an alternative approach to the evaluation problem. In this sense, the method is a proof-of-concept which

[1] The interfacing possibilities are discussed in Chapter 6.

[2] Within the simplest major-minor scenario, the coverage of the maximal informativeness requires 26 dimensions: 24 key categories, time and time-scale. The dimensionality of the category space is extended later in this chapter, and furthermore in Chapter 5.

> aims to promote an important aspect of music analysis: the very experience of being immersed in an analytical process, which sometimes is more insightful than the analytical conclusions. The discussions will thus be directed to both the (music) analytical results and the analytical process itself. This last point constitutes a major goal of this work, that of a comprehensive interpretation of the method and its components (features, spaces, processes), with the aim of understanding what they actually represent with respect to music.

The use of visualisation and interfacing techniques makes possible to exploit the human visual system in complex tasks of pattern recognition or grouping of similar structures (an imprecise task for computers), assisted by a systematic computational analysis (a costly and prone to errors task for humans). In his classic on visualisation, Tufte demonstrates massive "data maps" (Tufte, 2001, p. 16-20), in similar terms as our proposed method, as powerful "instruments for reasoning about quantitative information". We claim, further, that facilitating the interaction with comprehensive analysis and visualisation methods could be even more beneficial for assisting the analysis loop, the recursive process whereby the analyst test partial observations in order to refine his or her analytical intuitions. As Straus has pointed in analogy to contemplating a painting: "[...] what you see depends upon where you stand", to which he adds: "To appreciate the painting fully, you have to be willing to move from place to place. One of the specially nice things about music is that you can hear a single object like an interval in many different ways at once." (Straus, 2000, p. 10). Is the willing to move within simultaneous points of view in informed ways what motivates the framework described next.

3.3 Framework for tonal analysis

The method is proposed firstly for exploring the analytical potential of combining key estimations projected into continuous toroidal pitch-spaces with time vs. time-scale representations. Such combination provides a bird's-eye view of the tonal structure, and a precise account of the actual content of each analysis segment. The challenge is to preserve or to complement information across the different dimensional reductions, maximizing the informativeness at each of them, while taking into consideration both analytical and perceptual consistency. The main block diagram of the method is depicted in Fig. 3.1.

3.3.1 Preprocessing: chroma extraction

The music stimuli is processed to extract its *chroma* information. The method consists of quantifying the pitch-class content of the signal, and its implementation depends on the encoding of the music stimuli:

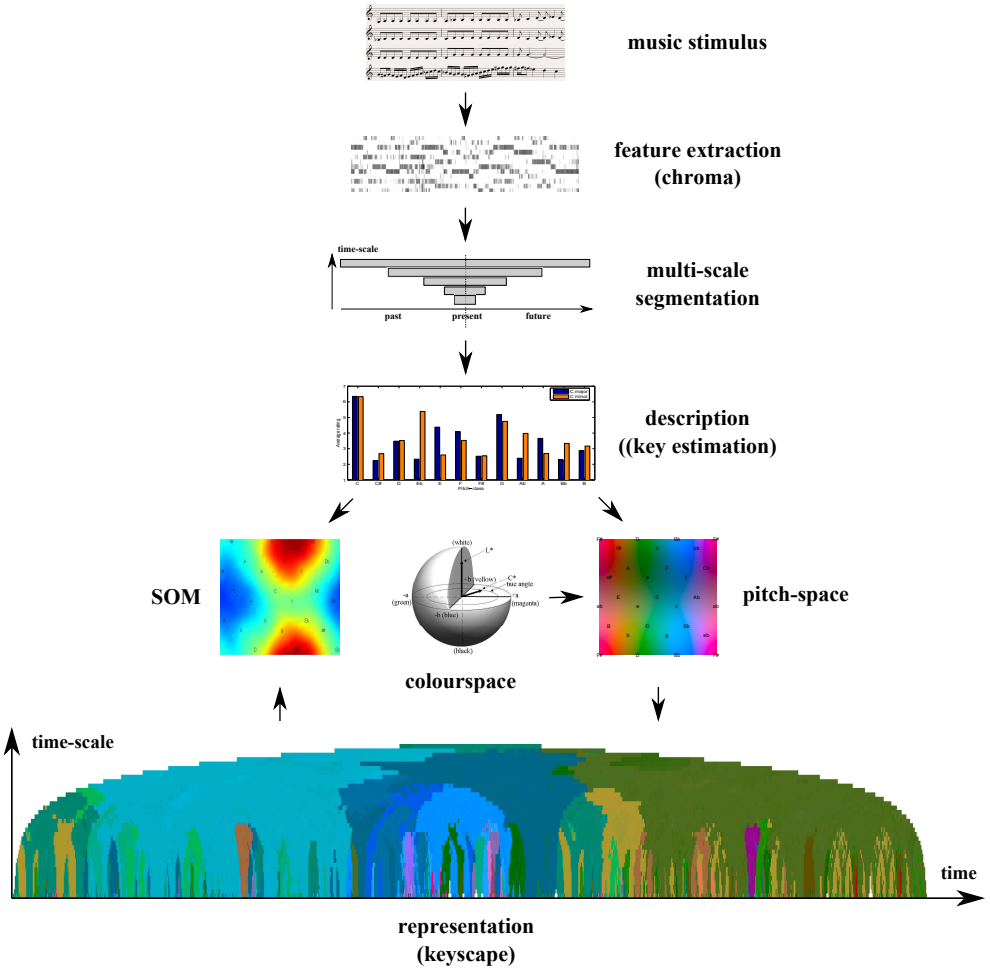

Figure 3.1: General method. Block diagram.

1. In the MIDI domain, each note (its MIDI note number) is encoded as an integer between 0 (C) to 11 (B), following the pitch-class set convention (after Forte, 1964). After that, doubled notes are discarded. This results in the same music mapped to a single octave. Our implementation builds upon the MIDI Toolbox for Matlab (Eerola & Toiviainen, 2004).

2. In the audio domain, many implementations under the common denomination of *chroma features* have been proposed. Any chroma feature can be used, provided that: a) it is properly tuned to fit the 12 pitch-classes of the twelve-tone equal-tempered system (hereafter 12-TET); b) its frame duration is equal or shorter than the minimum time-scale used in the segmentation module (see next). In our implementation, we tested two chroma features, namely the HPCP (Gómez, 2006), and the output of the *mirchromagram* function provided by the MIR Toolbox for Matlab

(Lartillot & Toiviainen, 2007). With the aim of favouring open-source solutions, and since the results were virtually identical for our purposes, we decided for the MIR Toolbox implementation.

3.3.2 Multi-scale segmentation policy

This module defines the temporal boundaries of the analysis frames. An adaptation of the multi-scale segmentation methods in (Sapp, 2005) and (Gómez, 2006) is proposed to overcome their main drawbacks, derived from applying non-overlapping sliding windows for each resolution. Visual aspects aside (see conclusions in Chapter 2), the main concern of these methods is the incremental loss of temporal resolution as time-scale grows. The negative impact of such an implementation is clear for our specific purpose. The method aims to model tonal context dynamics, and this requires both operating at relatively large time-scales, so as to capture the contextual information, and small hop-sizes, so as to guarantee a proper temporal resolution of the contextual changes.

The proposed multi-scale segmentation policy applies many rectangular sliding windows, with time-scales ranging from fractions of a second to the whole duration of the piece. The shortest window size is taken as the hop-size for all the resolutions. The number of time-scales and the common hop-size are parameters to the algorithm, as a trade-off between resolution and computational cost. A logarithmic ratio is applied to the window size for consecutive time-scales. This segmentation policy introduces three benefits with respect to Sapp's and Gómez's approaches. First, the aspect ratio of the keyscapes is more intuitive and pleasant to visualise. Second, it provides a consistent cross-scale temporal indexing for the data. Third, it solves the temporal resolution issue at medium and large time-scales, which is the most likely range for describing modulations.

3.3.3 Chroma segmentation and summarisation

Once the multi-scale segmentation module has provided the temporal boundaries of the analysis frames, each segment is analysed in order to extract a *pitch-class profile* representing its tonal content. The implementation depends on the encoding of the chroma information (see *Preprocessing* above):

1. In the MIDI domain, the pitch-class profile is computed by integrating the duration of each pitch-class within the analysis frame. The duration of the notes partially contained in the frame are accounted in proportion. The resulting pitch-class profile, a 12-dimensional vector, is normalised by the duration of the frame.

2. In the audio domain, the pitch-class profile is computed by integrating the values of each dimension in the chroma feature within the analysis frame. The result is normalised by the duration of the frame.

From now on, the method is agnostic with respect to the symbolic or audio encoding of the music stimuli. The output of the previous modules is a 3-dimensional matrix containing the multi-scale pitch-class-distribution time series. The analysis frames are indexed by the first two dimensions, which account for time-scale and time position respectively, while the third dimension indexes the pitch-class.

3.3.4 Description: key estimation

As mentioned above, we are interested in exploiting the analytical potential of non-sophisticated open-source methods, and this comes for two reasons. First, we want to foster a comprehensive understanding of both the method and its outcome, minimizing the interpretation artefacts. Second, we want to foster the reproducibility of the results. The description method, thus, implements a method virtually identical to the Krumhansl and Kessler's key-finding algorithm (Krumhansl, 1990). Each pitch-class profile, corresponding to each analysis frame, is correlated with ring-shifted versions of KK-profiles (see Chapter 2). The output is a 24-dimensional vector, quantifying the *key strength* for each of the 24 tone centres.

3.3.5 Projection of estimates in the pitch-space

The resulting 24-dimensional estimates, for all frames and time-scales, are projected in a space of inter-key distances. Two complementary methods are proposed. The first method is a multidimensional unfolding technique, virtually identical to that in (Krumhansl, 1990), which finds a tonal centroid in the (angle-based) 2-D KK-space (see Chapter 2). An *ambiguity unfolding* parameter is introduced to promote or relax the strongest candidate with respect to the rest. This method provides the highest dimensional reduction, required for building the keyscapes (see below). An alternative projection method consists of comparing the pitch-class profile with the codebook of a SOM trained with the 24 KK-profiles. The result is the complete activation of the SOM, according to the closeness of the input to each point in the space, providing the maximal (frame-based) informativeness. This is implemented by the *keysom* function of the MIDI Toolbox, adapted to be fed by a pitch-class profile.

3.3.6 Colouring method

The multi-scale information computed so far[3] consists of a time vs. time-scale matrix, each element containing the (2-D) position of the corresponding key estimate in the continuous toroidal pitch-space. In order to be able to represent the keyscape as a 2-D image, a colouring method has to be established. As a main novelty with respect to Sapp's approach (Sapp, 2005), which maps categorical keys to colours, the mapping to our continuous pitch-space implies three main considerations. First, that the torus is conceived to represent a space of perceptual inter-key distances, and the centroid unfolding aims to model the closeness to the different tone centres in similar perceptual terms. Second, the ambiguity of the estimates allows, in principle, the localisation of centroids anywhere in the continuous torus' surface. Third, the toroidal pitch-space features double spatial circularity. The colouring method, thus, should provide: a) a unique colour for each spatial position; b) a perceptual difference between colours according to their distance in the pitch-space; and c) a smooth continuity or colour blending across the space. In addition, these properties should be maintained across the double circularity of the space.

Among the perceptually-informed colourspaces, CIE 1976 L*a*b* (known as CIELAB) colourspace, defined by the Commission Internationale de l'Éclairage as conversion standard (CIE 015, 2004), can approximate the colouring requirements of the model. Three features are relevant for chosing it. First, it uses absolute scales[4]. Second, it is device independent[5], so it can be used in a variety of media. Third, it is compressed to approximate *perceptual uniformity*. That means, the Euclidean distance between any two points in the space is approximately correlated to the perceptual difference of the colours at those locations. CIELAB is a 3-dimensional geometric space in which most of the human visible gamut is covered by a spherical sub-space. Three parameters define colour: $L*$ for luminance, $a*$ for green-magenta colour-opponent axis, and $b*$ for blue-yellow axis. Another convenient parameterisation, referred to as LCh, uses cylindrical coordinates: L for luminance, C for chroma saturation, and h for hue angle.

Assuming perceptual uniformity, the geometrical inscription of any 3-D object within CIELAB results in the colouring of such object in the same terms: the perceptual difference between the colours of any two points would be correlated to their mutual spatial distance. Considering that the torus is also intended to reflect perceptual closeness between keys, the colouring solution becomes

[3] The colouring method only applies to the maximal summarisation of the key estimates, projected in the 2-D KK-space.

[4] The scales are relative to CIE's *standard illuminant* $D50$ white point.

[5] The final result depends on the device's gamut and mapping equations. The prototyped implementation uses Matlab's LAB to sRGB conversion, based on CIE's *perceptual intent* recommendation.

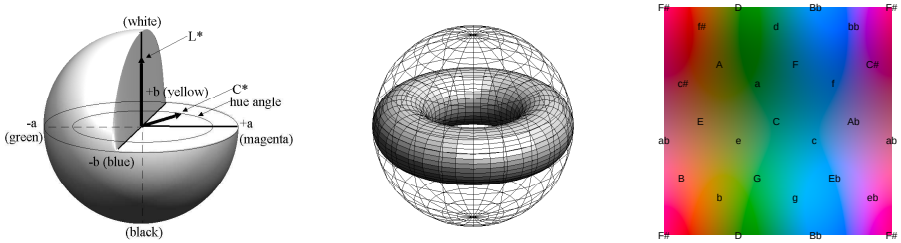

Figure 3.2: Colouring process. Left: spherical sub-space of CIELAB colourspace. Centre: inscription of the toroidal pitch-space in the spherical colourspace. Right: unfolded torus' surface, after colouring.

evident: a geometrical inscription of a 3-D projection of the toroidal pitch-space in the CIELAB's *sphere*. This gives a unique colour for each point in pitch-space (toroidal surface) and gradual colour transitions along any direction, and it also approximates perceptual correlation with distance. In addition, these properties are guaranteed across the torus' double circularity.

The chosen orientation matches the hue angle in CIELAB with the circle resulting from the first two dimensions of 4-D KK-space (see Chapter 2). This provides the maximum hue differences along both circles of fifths. The last two dimensions of the solution are parameterised as a circle in L and C axes. Geometrical rotation of the torus can be used to select a preferred colour for some tone centre, and for compensating the local distortion of the projection[6] with respect to the 4-D space. The colouring method is sketched in Fig. 3.2.

3.3.7 Representation: keyscape

After applying the colouring method to all the tonal centroids, which represent the tonal estimates for all frames at all time-scales, the final step consists of organising the information as a 2-dimensional time vs. time-scale image: the keyscape. In the following sections, the features of the method will be discussed in detail.

3.4 Discussion

The Finale of Haydn's String Quartet *Op.74 n.3* "Rider", in *G* minor, will serve to show the combined representational capabilities of the proposed model. The keyscape in Fig. 3.3 represents all the tonal estimations in time (x-axis) vs. time-scale (y-axis). As it is illustrated by three sample pixels and their corresponding segments, the higher the pixel in the keyscape, the larger the duration of the analysis segment. Each pixel in the image represents a unique segment of music, indexed by it's central position in time (x) and the logarithm

[6] Comprehensive discussion with respect to inter-key distances in (Krumhansl, 1990).

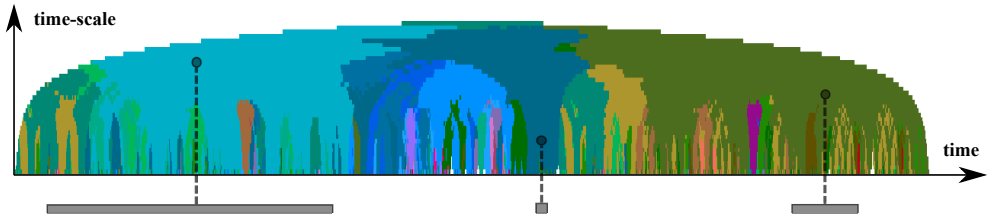

Figure 3.3: Finale of Haydn's *Op.74 n.3*. Keyscape and three sample segments.

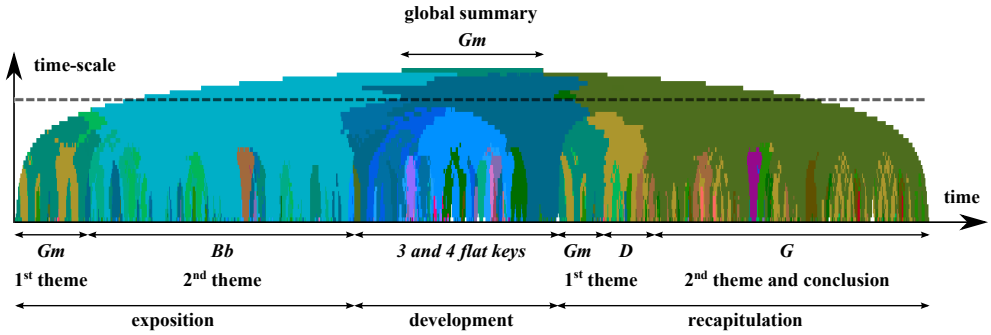

Figure 3.4: Finale of Haydn's *Op.74 n.3*. Tonal structure.

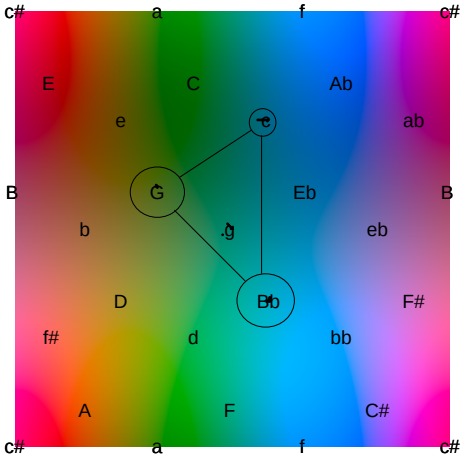

Figure 3.5: Finale of Haydn's *Op.74 n.3*. Tonal structure in space.

of its duration (y). Its colour represents the location of the tonal centroid at the surface of the toroidal pitch-space, which in turn represents the best projection of the segment's tonal content in this dimensionality. The parameters of the centroid unfolding for this example have been chosen to promote the first candidate in the estimation, therefore the centroids are located close to single keys. In the image, homogeneous colours covering large areas reveal stable tonal sections, stability understood in both time and time-scale.

3.4.1 Formal Analysis

An analysis of the main tonal sections[7] has been manually labelled in Fig. 3.4. The piece presents a clear classical sonata structure. It begins with a first theme in Gm, followed by a longer second theme group in $B\flat$. Then, a development section wanders around 3 and 4-flat key signatures (Cm, $E\flat$, $A\flat$ and Fm, each one as a different blue tone in the keyscape). Recapitulation takes the first theme in Gm, and a dominant pedal leads to the second theme group and conclusion in G major.

The complete piece is correctly[8] estimated as Gm (very top of the keyscape), a somewhat surprising result given the relative short duration in this key. The connection between the keyscape and the pitch-space will clarify this analytical aspect at structural level. It will also illustrate a relevant metrical property of the pitch-space when representing information embedded at different time-scales. The temporal resolution of interest is depicted as a dashed line in Fig. 3.4. For this time-scale, the three broadest tonal regions are captured. The pitch-space depicted in Fig. 3.5 serves as a colour legend to the keyscape, and to show all the centroids above the selected time-scale, projected as black dots. Since the centroids fall very close to the categorical keys, their accumulation has been approximated by the circle's sizes for visual convenience. The exposition and recapitulation sections have been summarised as $B\flat$ and G respectively, given the longer duration of their second theme groups. They are almost balanced in duration, and they are located almost symmetrically at both sides of Gm. The shorter development section, summarized at Cm, gently pulls the whole-piece centroid towards the global key of the piece. This spatial interpretation of a *classical symmetry* is a joint consequence of the topology of the pitch-space, and three aspects connecting the key estimation and representation methods with some of the piece's compositional and aesthetic concerns. First, the KK-profiles are generally well suited, in statistical terms, for the pitch-class distributions of classical works. Second, the duration-based computation of the pitch-class profiles matches the classical style's ideal of temporal balance, whereby a simple principle of symmetry rules the duration of phrases and sections. Third, Haydn's choice of tone centres in relation to the tonic: the relative $B\flat$, sharing the main scalar pitch-class set with Gm; the parallel G, sharing the main scale degrees with Gm; and the bias at the development section, deciding in favour of a flattened key signature. The topology of the pitch-space makes the rest, and summarises the global situation right in the correct key. It is worth mentioning that the global tonal summary has not been computed as a systemic reduction from the three discussed tonal regions. All the estimates in the keyscape are computed taking into account the

[7] The author's analysis of the score.

[8] Correctness here responds to the tonal conventions of the classical sonata forms in minor mode.

complete pitch-class content of their segments, and so has been done for the global summary as well.

3.4.2 Space suffices: ambiguity of *Type I*

Haydn's example has pointed to the potential of geometrical pitch-spaces for summarising the tonal content of segments, when they are embedded by larger time-scales. In the next case study, a first type of ambiguity (hereafter, of *type I*) will be discussed. Type I ambiguity considers one of the primary facets of tonality in general, and of modulation processes in particular: that of the estimations legitimately belonging to several keys in different degrees. The estimation method considers all the music segments as ambiguous, as it provides the relative strength of all the key candidates. The ambiguity of type I, however, is not related to the estimation method itself, but with the representation of the estimates in the pitch-space. To be more specific, it has to do with the suitability of the pitch-space, which is to say of the underlying tonal hierarchies, for representing the estimate in a proper way. From the algorithmic point of view, the projection of the segments as tonal centroids in pitch-space is not forced to fall right in the strongest key, but several other candidates are allowed to contribute in the location, according to their relative strengths. Type I ambiguity is thus referred to as centroids whose projections in the torus' surface are not subjected to a large stress in terms of dimensional scaling. While multidimensional unfolding always finds a solution to project the centroids in the target space, some solutions are better than others. An ambiguous estimate of type I can be projected in the space's continuum (between two keys), but its relative distances from the key categories will be similar to their 12-D counterparts before scaling. The centroids with this kind of ambiguity are considered as *tonal* in the same terms as the key profiles used for building the low-dimensional pitch-space: they resemble *both* the pitch-class hierarchies embedded in the tonal categories *and* a proper mixture between neighbouring keys.

Chopin's *Op.28 n.9*, prelude for piano in E major, will serve to illustrate the difference between two unfolding parameterisations, and to show the potential of continuous pitch-spaces to inform about close and far modulations. The keyscape in the top pane of Fig. 3.6 maps centroids in pitch-space, as done in Haydn's example, by promoting the strongest key, showing homogeneous colours and neat boundaries between regions. Centroid unfolding for the second keyscape, in the bottom pane of Fig. 3.6, considers the two strongest candidates and their weights. The image looks similar but it's significantly fuzzier, being informative in a different way. Most areas represent spatial locations between two keys, by a proportional blending of the colours corresponding to both candidates. The continuous colourspace aims to facilitate the distinction between movement towards neighbour keys (soft mixture of both

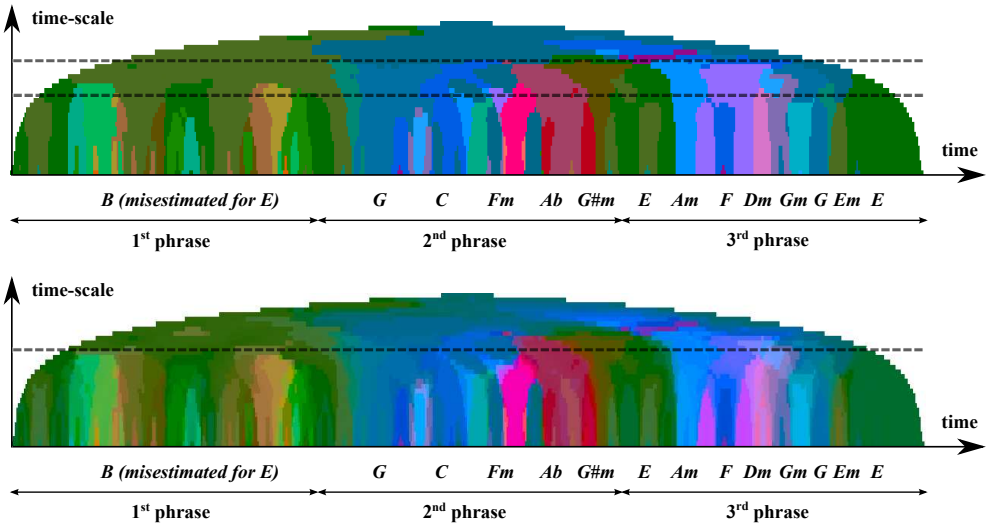

Figure 3.6: Chopin's *Op.28 n.9.* Keyscapes. Top: categorical unfolding. Bottom: ambiguous unfolding.

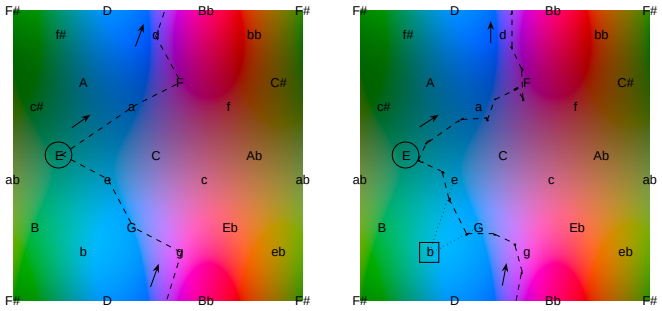

Figure 3.7: Chopin's *Op.28 n.9.* Path of the third phrase. Left: categorical unfolding. Right: ambiguous unfolding.

colours) and far jumps (sudden strong contrast). As a difference with other colouring schemes, like that in (Sapp, 2005), the proposed geometrical mapping approximates gradual colour transitions in any direction, so all neighbour tonal relationships would get benefited from the perceptual closeness. That is, the combination of the colouring scheme and the ambiguous unfolding facilitates the visual recognition of neighbouring modulations.

The left pane of Fig. 3.7 shows the path followed by the centroids for the third phrase of the prelude, under categorical unfolding. The right pane of Fig. 3.7 depicts the same phrase for the ambiguous unfolding case. The time-scale used in both analyses is depicted in the ambiguous keyscape in Fig. 3.6 (bottom) as a horizontal dashed line. Instead of jumping to categorical keys, some intermediate positions soften the way and the corresponding colour transitions

in the ambiguous path. By observing the centroid deviations from their closest keys it is possible to grasp which keys were taken as the second candidates, as it is illustrated in the ambiguous path during the transition between G and Em. It is worth noting that the most critical aspect for achieving gradual colour shifting is the capability of the key estimation and centroid unfolding methods to provide such gradual movement in pitch-space, and this necessarily depends on musical discourse as well.

The former definition of ambiguity of type I, as projections in pitch-space not subjected to a large stress, is also applicable to a related but different scenario: that of passing tone centres during modulations to non-neighbouring keys. A look at the tonal structure of Chopin's prelude will illustrate this aspect. Standard analyses of modulation and tonicisations for this piece are found for time-scales between the horizontal dashed lines in Fig. 3.6 (top). The estimations at these time-scales are labelled as reference below both keyscapes. This piece is structured in three phrases, as it is also labelled in the figures, each of them beginning and ending in E. The last two phrases take distant journeys in pitch-space, and considerable analytical freedom with respect to tone centre induction is allowed. Compared with the tonal stability of the first phrase, a notable key shifting activity is revealed for the last two sections. The second phrase visits G-C-Fm-$A\flat$-$G\sharp m$, and the third phrase crosses Am-F-Dm-Gm-G-Em. A comparable tonal path through pitch-space for this piece can be found in (Lerdahl, 2001, p. 96-98). As in Haydn's example for different time-scales, the topology of the pitch-space also produces summarised descriptions over time for fixed temporal resolutions. This provides a means for estimating passing tone centres during modulations to non neighbouring keys. In this case, the model estimates Am in the third phrase, a key suggested by Lerdahl as a pivotal place, required to fit the *shortest path* rule from E to F (Lerdahl, 2001, p. 97-98)[9]. The music discourse does not rest here, but it would be easy to interpret the passage in such terms by intentional listening. Since an applied dominant precedes the Am chord, an arrival sensation can be induced depending on the performance timings. This kind of analytical freedom is what characterises the ambiguity of type I. Similarly, the model estimates Dm as mediating between F and Gm. Incidentally, the model fails to estimate the first phrase in E, in favour of B, as well as the global key of the piece. This is not surprising, since Romantic music is far less suited for the KK-profiles (see Krumhansl, 1990, p. 88-89) and for the aesthetic premises of temporal balance discussed for Haydn's example.

[9]Lerdahl actually considers the Am chord as the fourth degree of Em, taking a further interpretative step from E to Em, so as to avoid diagonal movements in his regional space.

3.4.3 Space does not suffice: ambiguity of *Type II*

As seen above, the ambiguity of type I can be captured by the reduced dimensionality of the torus, whether as a projection of centroids between neighbouring tone centres, or right in pivoting keys during non neighbouring modulations. For some musical segments, however, the stress introduced by the unfolding method can be too large, up to the point of producing meaningless tonal centroids. For these cases, the estimation method could provide a reasonable description of the segments in 12 dimensions, but the torus surface does not suffice for representing estimations as single points. From now on, this will be referred to as ambiguity of *type II*.

In order to identify the problematic cases, it is useful to inspect the raw estimations before unfolding. Since the estimation method normalises the resulting 24-D vectors with respect to total energy, an ambiguity of type II is expected to be manifested by small values of the maximal correlations of the segment's pitch-class profile with the KK-profiles. High values of maximal correlations indicate strong tone centre implications, as they have been described as an indicator of key clarity (Lartillot & Toiviainen, 2007). On the contrary, very low values of maximal correlation indicate that no key centre gets close to the tonal content of the segment. It may also indicate that several non-neighbouring keys are the best candidates, since two or more distant keys cannot score simultaneously high in a space which represents dissimilarity by distance. That is, the ambiguity of type II express segments of music which cannot be considered as tonal with respect to the key categories and their neighbouring distances. Since the topology of the pitch-space embeds this joint information, it cannot represent properly the estimates as centroids. In the top pane of Fig. 3.8, the keyscape computed from an excerpt taken from the Finale of Haydn's *Op.74 n.3*, is shown. The image depicts several clearly embedded contexts. The arising question is which degree of confidence portrays such representation. In the bottom pane of Fig. 3.8, a visualisation is proposed to answer this question, hereafter referred to as *confidence-scape*. This image represents the maximal correlation value for of each analysis frame[10], clarifying the trustable areas in the keyscape. The colour coding goes from dark blue (minimum) to dark red (maximum). Two sample segments have been selected in both the keyscape and the confidence-scape, for a further inspection of their representational differences.

The ambiguity of type II can also be explored in interactive ways, without the need of leaving the keyscape representation. By a gradual shifting between categorical and ambiguous unfolding[11], the most ambiguous areas are evidenced by notable (even extreme) changes in colour, which manifest the characteristic centroid deallocation of the ambiguities of type II. For getting the best

[10] Computed using the *mirkeyclarity* function from the MIR toolbox.
[11] The unfolding ambiguity parameter of the method is discussed in Chapter 6.

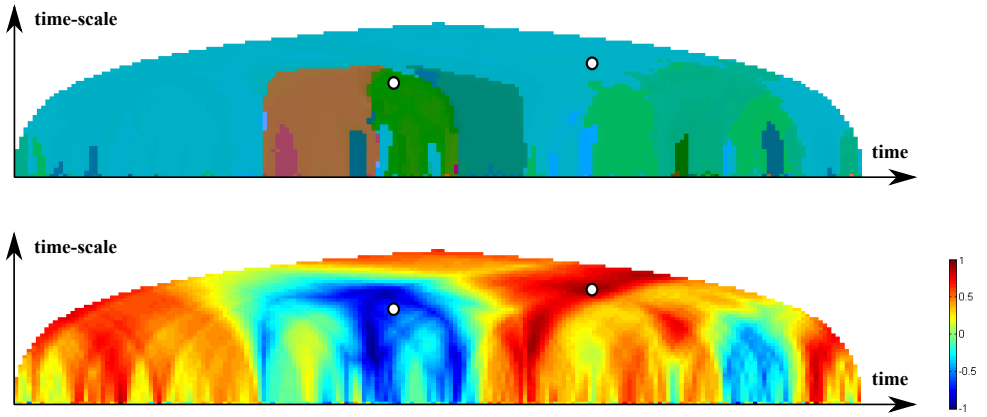

Figure 3.8: Finale of Haydn's *Op.74. n.3* (excerpt). Top: keyscape and two sample segments. Bottom: confidence-scape and the same sample segments.

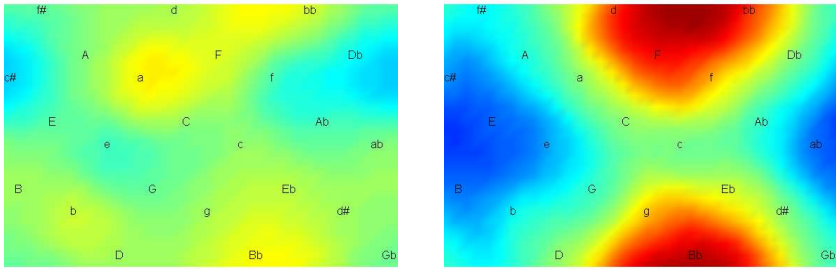

Figure 3.9: Finale of Haydn's *Op.74. n.3* (excerpt). Left: non confident segment. Bottom: confident segment

information, however, one should explore individual segments in higher dimensional spaces. As we have discussed above, the complete activation of a toroidal SOM, trained with the KK-profiles, can represent any input vector in a human-readable way. In Fig. 3.9, the sample frames located as white circles in Fig. 3.8 are represented as the activation of such a SOM[12]. The left pane depicts the activation of the space for the ambiguous segment, while the right pane does the same for the unambiguous one.

3.4.4 More on ambiguity

As other authors have proposed their own terminology and estimation methods with respect to tonal ambiguities, it is relevant to discuss them with respect to the typology proposed here. According to Temperley, the "tonal ambiguity" of a pitch-class set[13] is "the degree to which it clearly implies a single key,

[12] Computed using the *keysom* function from the MIDI toolbox.

[13] Temperley elaborates the concept of ambiguity for pitch-class sets, without considering pitch-class hierarchies.

or is equivocal between two or several keys" (Temperley, 2008, p. 28). This tonal ambiguity, also referred to as "tonal clarity", has to do with the relative strength between the most probable key and the next more probable[14], so in its algorithmic form is connected (although not fully equivalent) to the key clarity descriptor used in Haydn's example. However, this description is agnostic with respect to the vicinity of the involved candidates, which we have claimed as a factor of analytical relevance. The taxonomy proposed in this thesis, in terms of type I and type II ambiguities, captures such refinement and relates it explicitly to the topological properties of the pitch-space. Temperley's concept of "tonalness", according to his definition as "the degree to which a set seems characteristic of the language of common-practice tonality" (Temperley, 2008, p. 29), seems to be appropriate for characterising the ambiguity of type I. Temperley's examples of low clarity but high tonalness, as for the set $\{C,D,E,G,A,B\}$ (p. 30), would be related to an ambiguity of type I, since the two strongest (and equally strong) candidates are C and G, neighbours in the circle of fifths. Similarly, the set $\{C,D,E\flat,F,G,A\flat\}$ (p. 36) is ambiguous between the relatives Cm and $E\flat$, also neighbours in the toroidal pitch-space. In fact, the vicinity of key candidates for both sets are realised in the first two dimensions of the 4-D KK-space, which account for the double circle of fifths connected by the relative relation[15]. On the other hand, cases of extreme ambiguity with respect to the major-minor paradigm, such as the case of symmetric sets[16], are rated low in both clarity and tonalness, and they would correspond to an ambiguity of type II. For these symmetric sets (or pitch-class distributions), several keys can be activated with similar strengths as well. However, instead of being clustered together (as neighbours), they are evenly distributed across the pitch-space, far to each others. Temperley's probabilistic framework for describing ambiguity can thus be conceptualised and implemented in purely topological terms, without involving the analysis of representative corpora.

For music composed under common-practice tonality, in which modulations generally move to neighbouring keys, it would be expected a notable ambiguity of type I, but moderate ambiguity of type II, even in cases of continuous and fast modulations. In Fig. 3.10, a comparison between two ambiguity measures is shown for Chopin's *Op.28 n.9*. In the top pane, the image depicts the confidence-scape computed as maximal correlation[17]. In the bottom pane, the maximal value of the SOM activation is proposed as an alternative indicator of confidence. In both cases, a high value (in red) indicates a low ambiguity

[14] Temperley framework is of a probabilistic nature, based on the analysis of music corpora assumed to be *prototypical* of the standard tonal language. Temperley relies on the so-called Kostka-Payne corpus, taken from (Kostka & Payne, 1995).

[15] See Chapter 2 of this document, or (Krumhansl, 1990, p. 43) for more details.

[16] Formalised by Messiaen as the *modes of limited transposition* (Messiaen, 1944, p. 58).

[17] Computed using the *mirkeyclarity* function of the MIR Toolbox.

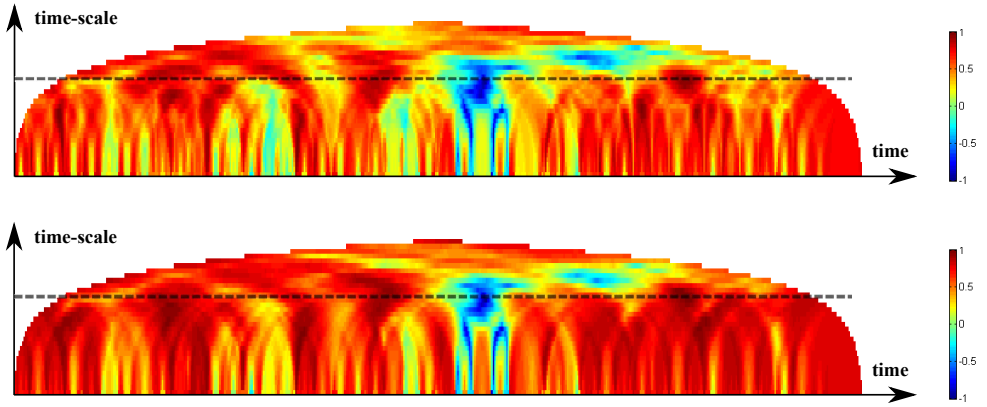

Figure 3.10: Chopin's *Op.28. n.9.* Confidence-scapes. Top: maximal correlation. Bottom: maximal SOM activation.

and vice versa. As seen above for the third phrase of the Chopin's prelude, the long journey through pitch-space evolves smoothly, moving from one tone centre to its neighbours. In both confidence-scapes, the same time-scale used in the previous modulation analysis of the third phrase is depicted as a reference. As expected, the maximal SOM activation yields quite similar values, although visibly higher, than the maximal correlation. The centroids are mostly projected between keys, but their low stress values are revealed by the activation of the SOM in a single, strongly focused, cluster. The bottom confidence-scape reveals that the ambiguity portrayed by the centroid unfolding between keys is compatible with the conventions of the common-practice tonality. We argue, thus, that both the maximal SOM activation and the maximal correlation would qualify as a proper indicator of Temperley's definition of tonalness, which is to say, of ambiguity of type I. This is not to say, however, that the SOM's codebook vectors between tone centres can be thought of as prototypes of the ambiguity in common-practice tonality. These vectors are just an interpolation result of the training process, and a careful testing should be required in order to claim any general conclusion about their appropriateness[18]. It is obvious that training the SOM with other key profiles would result in the corresponding interpolations, and that the highest activation of the space would convey different interpretations.

It is important to notice the differences in both methods. Temperley uses pitch-class sets, and his probabilities have a census origin, from computing a common-practice period dataset. The method presented here operates on real music, and embeds both pitch-class hierarchies and relative durations. That

[18]Temperley's revisions of the KK-profiles (Temperley, 2001, p. 180), derived from theoretical reasoning and trial and error, represent to some extent such kind of *mixture* profiles, intended to characterise certain corpus-specific tonal trends.

both approaches yield similar measures about ambiguity, could be based on the fact that the KK-profiles are generally well suited for tonal content of the kind found in the Kostka-Payne corpus. The probabilistic interpretations of the KK-profiles have remained, however, controversial (see Aarden, 2003, for a critical review).

3.4.5 Contextual stability as information

From the observation of the extremely ambiguous cases, it is evident that the major-minor pitch-space does not suffice for representing ambiguities of type II as centroids. However, a question about the informativeness of the keyscape in these cases still remains open. The discussion so far has been mostly approached as a classification problem: the segments are compared and interpreted with respect to the pitch-space's underlying categories. That means, ambiguities of type II cannot be properly *labelled* in the intended terms, neither as a single key, nor as a reasonable combination of neighbouring keys (ambiguity of type I). However, a considerable accumulation of a similar evidence in time and time-scale, even in cases of ill-defined centroid unfolding, could inform about the contextual stability, which in some cases is of analytical or perceptual relevance. A broad homogeneous area in the keyscape could be an indicator of a more general concept of tonal context, even if one is unable to name it ... yet.

Ligeti's *Polifón etüde* will serve to illustrate this idea. The example will also reconsider the summarisation behaviour of the pitch-space, for cases in which several different contexts do not appear in sequence, but simultaneously. This minimalistic work for 4 hands piano explores the juxtaposition of four melodic ideas, each of them based on a different scalar formation set. Each voice, from the lowest to the highest, enters in sequence right after the previous one has completed a cycle, and each exposition is repeated continuously afterwards, until music suddenly ends in a final chord. From the lowest to the highest, each voice exploits the pitch-class sets {0,2,4,5,7,9}, {1,3,6,8,10}, {1,4,6,8,9,11} and {0,2,5,7,10} respectively. In Forte's notation (Forte, 1964), the left hands of both players make use of the set-class 6-32 (sometimes called Arezzo's major diatonic hexachord), while their right hands uses the 5-35 (pentatonic) set-class, each player at a different transposition and inversion[19].

Fig. 3.11 depicts the SOM activation due to the individual hands (bottom panes), both hands for each player (central panes) and both hands for both players (top pane). The figures are arranged from the lowest voice (1) to the highest (4), according to their order of appearance in the piece. The corresponding pitch-class profiles are depicted below each SOM. A strong key clarity is evidenced for the pentatonic contexts at voices 2 and 4, given its

[19] A complete list of set-classes is available in Appendix A.

similarity to their diatonic counterparts. Some ambiguity of type I appears in voices 1 and 3, in a similar feeling as other hexachordal scalar formations (see Temperley's examples in the previous section). The clarity gets notably reduced for the combination of both hands in any of the players, as shown in the central panes. The involved sets are {0,1,2,3,4,5,6,7,8,9,10} for the voices 1+2 (bass parts, player 1), and {0,1,2,4,5,6,7,8,9,10,11} for the voices 3+4 (treble parts, player 2). As expected, the combination of the four different contexts, in the top pane, results in the most unclear scenario. For computing these SOM activations, each frame begins at the entrance of the last voice (the highest one), and lasts until the end of the piece. This guarantees that all the involved voices are sounding simultaneously, providing a proper tonal context to be analysed.

It can be observed that the combination of the voices 1+2 behaves approximately as expected: the highest activation of the pitch-space lays, more or less, between the highest activations for the individual voices. This is comparable to the summarisation discussed for the Haydn's structural analysis. In the Haydn's case, different contexts presented in sequence resulted in its topological summary in pitch-space, when a larger time-scale fits both of them in the same analysis frame. In Ligeti's example, both contexts share the same temporal range, being topologically summarised at the corresponding time-scale. The combination of the voices 3+4, however, does not behave this way. The resulting summary is similar to the activation for the voice 3 alone, although significantly weaker. This points to the fact that the combination of the different contexts is not required to be represented as a geometrical summary in the pitch-space for all the cases. The tonal categories scaffolding the pitch-space are not only comprised of pitch-class sets, but they also encode a hierarchy of pitch-classes. So, depending on the particular hierarchies of the input vectors, the result may or may not correspond to topological summaries in the expected terms.

While this representation illustrates how the confidence of estimation gets degraded as more contexts are superimposed, it only captures a single frame. The remaining question is how stable is this information in time and time-scale. In Fig. 3.12, three keyscapes computed from Ligeti's *Polifón etüde* are shown. The top pane depicts the keyscape for the full piece. The keyscape for only the bass parts (player 1) is shown at the central pane. The computation for only the treble parts (player 2), which begins later in time, is shown in the keyscape at the bottom pane. The three images suggest that stability is achieved beyond a certain time-scale, depicted as a dashed horizontal line, but this can only be confirmed by the inspection of these frames as a SOM activation[20]. In this case, the time-scale thresholding the contextual stability condition is that spanning a complete cycle of the combined motivic ideas, af-

[20]The interfacing possibilities of the method will be discussed in Chapter 6.

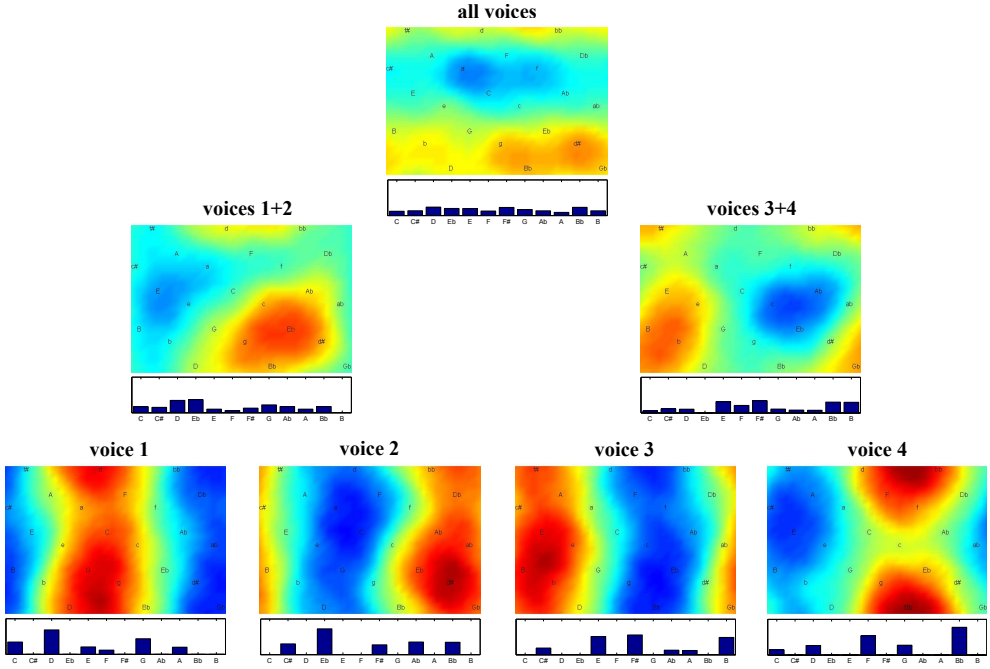

Figure 3.11: Ligeti's *Polifón etüde*. SOM activation.

.

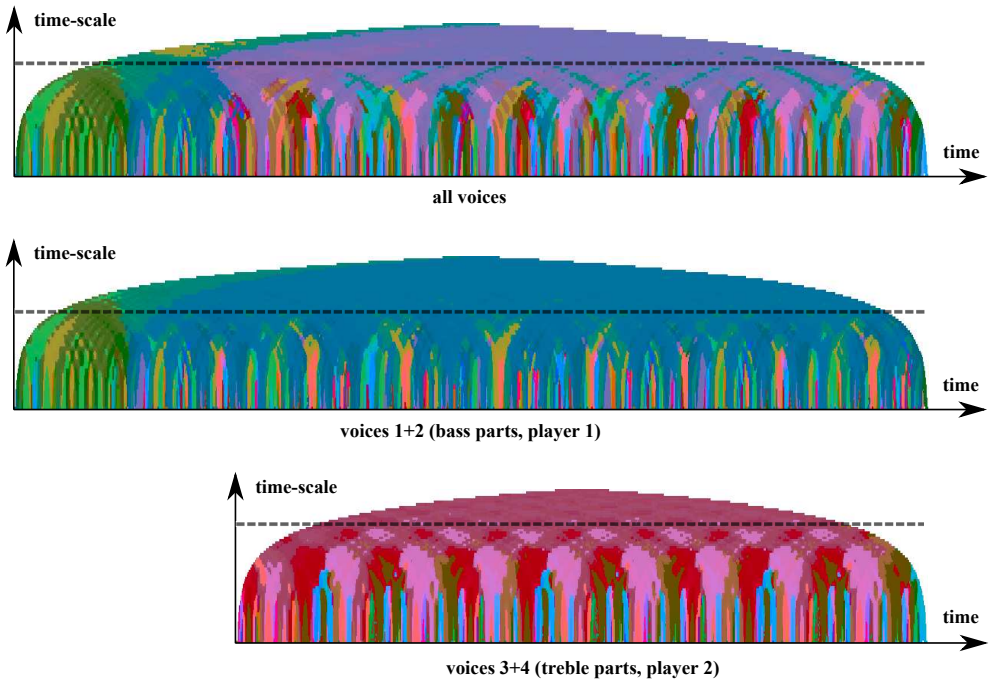

Figure 3.12: Ligeti's *Polifón etüde*. Keyscapes and stability thresholds. Top: all voices. Centre: bass parts. Bottom: treble parts.

ter which the same material is repeated. These periodicities are easily noticed in all three keyscapes, as well as the stable areas above the threshold.

3.4.6 When chords become contexts

As seen in Ligeti's example, an insightful feature of the keyscapes is a visual access to the time-scale thresholds, which separate non-stable from stable contextual information. For tonal music, this threshold often captures the temporal boundary between the description of chords and keys. As mentioned in Chapter 2, the temporal duration required to capture the sense of key is unclear in general. The inspection of keyscapes in terms of stability can constitute a means for reasoning about that. The strict distinction between chords and keys can actually admit a reconsideration, if one interprets the concept of tonal context in terms of stability. What matters for considering a segment as contextual information is what one could call its *internal* stability. That is, during a certain duration and for a certain range of time-scales, the tonal implication of all the sub-segments should be similar to that of the whole. This is what the keyscapes represent by broad connected homogeneous areas. In all the examples so far, even in the extreme case of Ligeti, the contextual information was related to scalar formations conveniently distributed so as to span the whole chromatic circle of pitch-classes. That is, the pitch-class set material was closely related to the concept of scale. However, a more general abstraction of tonal context does not require this condition. Minimalistic and electronic genres make use of more reduced pitch material to construct long contexts, an effective resource to induce a sense of stability from both perceptual and aesthetic standpoints[21]. Long contexts built from reduced sets also pertain to music from the common-practice period. The Prelude of Wagner's *Das Rheingold* builds a unique $\{E\flat,G,B\flat\}$ trichord along 136 bars, with a duration of about four minutes. While it is evident that such a monumental chord induces a clear sense of being in a key of $E\flat$, the contextual information is just a triad. If such a long $E\flat$ triad were followed by a much shorter passage in $A\flat$, the whole passage from the beginning could be interpreted in a key of $A\flat$ major, relegating the $E\flat$ to a dominant role[22]. This, of course, would not be what the keyscape represents, since the estimation algorithm is based on the relative duration of active pitch-classes. The accumulation of similar evidence in time and time-scale in the keyscapes, thus, represents the idea of context as stable information, whether interpretable in terms of keys or not. This idea will be elaborated further in the next case study. A systematic approach to this concept will be elaborated in Chapter 5.

[21] An exponent in the genre is Glass' *Einstein on the Beach* (1976), a five-hour opera in four acts largely based upon this kind of material.

[22] The usage of long dominants preceding the main thematic material in large works is not alien to the Romantic repertoire.

3.4.7 Different categorical spaces

A question still remains open from Ligeti's example. If keyscapes can capture stable tonal contexts, it would be possible, in principle, to characterise such contexts in terms of analytical relevance. A solution can be found away of the major-minor paradigm of tonality, that is, by building pitch-spaces able to represent the particular contexts, whether as categories or as combinations of neighbouring categories. This has been approached as a dimensional scaling problem for characterising tonal spaces suited for different contextual categories, such as North Indian ragas (Castellano et al., 1984, p. 407). However, given the aesthetic implications of the raga-based music, there is a lack of studies about modulation and its representation. It is expected, though, that these spaces would behave similarly to the major-minor space, in terms of summarisation in both spatial and temporal dimensions, provided that a similar duration-based key-finding method were used. The analytical potential of interfacing keyscapes and pitch-spaces, for the case of a different set of contextual categories, is tentatively explored next.

Since the symmetric pitch-class sets are the most ambiguous with respect to the major-minor paradigm, they will be conform the categorical set under study. In the symmetric modes, commonly referred to as the Messiaen's *modes of limited transposition*, the intervals are distributed as evenly as possible along the chromatic circle of pitch-classes, reducing the number of different transpositions. A list of the the modes[23] is shown in Table 3.1. To create the corresponding pitch-space, a method virtually identical to that described in (Krumhansl & Toiviainen, 2001) is followed. The major-minor key profiles are substituted by a flat (binary) version of the symmetric modes, so as to avoid any sense of hierarchy among the pitch-class set members. Although composers often induce a tone centre by varied means, including temporal unbalance, this study attempts to approach ambiguity at its best manifestation. The categorical set includes 7 modes with 2, 3, 4, 6, 6, 6 and 6 transpositions respectively, summing a total of 33 categories. The notation here uses two numbers separated by a hyphen, representing the mode and its transposition (e.g. 3-2 stands for the third mode, second transposition)[24]. A toroidal SOM is trained taking these modal profiles as input. The resulting codebook includes the input vectors, located at places minimising the global stress, and their interpolations covering the full space. Fig. 3.13 shows a comparison between the major-minor space (left pane) and the space of symmetric modes (right pane), after the colouring process.

Scriabin's Op.74 n.5, prelude for piano solo, will serve to discuss some analytical possibilities of this new categorical space. Fig. 3.14 depicts the keyscape and the confidence-scape (based on the maximal SOM activation), computed

[23] The list only considers the fully symmetric modes, without their *truncated* variants.
[24] Not to be confused with Forte's set-class notation.

using the inter-key space in Fig. 3.13 (left). The inadequacy of the major-minor space is evident: the chaotic colour mixing in the keyscape does not reflect any sense of structure, aside a short section at the beginning being repeated at the middle of the piece. The confidence-scape confirms the extreme ambiguity of the estimations, specially at time-scales of contextual level, but it also reveals a repetition pattern, suggesting an underlying structure. In Fig. 3.15, based on the space of symmetric modes in Fig. 3.13 (right), a clear repeated A-B structure appears. The trustability of the information is confirmed by the low ambiguity evidenced in the confidence-scape, at least for time-scales of reasonable contextual relevance (depicted as a horizontal dashed line).

Mode	**First transposition**	**# of transpositions**
1	{0,2,4,6,8,10}	2
2	{0,1,3,4,6,7,9,10}	3
3	{0,2,3,4,6,7,8,10,11}	4
4	{0,1,2,5,6,7,8,11}	6
5	{0,1,5,6,7,11}	6
6	{0,2,4,5,6,8,10,11}	6
7	{0,1,2,3,5,6,7,8,9,11}	6

Table 3.1: Modes of limited transposition

The so-called *acoustic scale*[25], also referred to as the *mystic chord* in its extended scalar form, permeates the whole piece in varied transformations. It has a strong (simultaneous) relation with the Lydian and Dorian diatonic modes, and both the whole-tone and octatonic scales. The aesthetics of the set is distilled by Schloezer, by stating that:

> [...] the concept of the scale is fused with that of the chord, and this chord, embracing the entire scale, appears perfectly stable, reposing upon itself without requiring resolution. It synthesizes and summarizes the scale. From this standpoint any transposition of the chord is equivalent to a freely effected modulation." (de Schloezer, 1987, p. 321-322)

The piece follows a clear A-B-A'-B' structure, labelled in Fig. 3.15. In the A sections, the different realisations of the mystic chord enhance the whole-tone sensation of the set, and prepare the octatonic sonority of the B sections (see Chang, 2006, for a detailed harmonic progression analysis). Both A and A' sections are subdivided in a-a' and a-a'' respectively. The subsection a is related to a' by a four-semitone transposition, and a' is a tritone away from a''.

[25] The scale formed by the 7th to 13th partials from the overtone series. Beginning at C, it forms the set $\{C,D,E,F\sharp,G,A,B\flat\}$.

The tritone is also the transpositional relation between B and B'. By transposing with these whole-tone-based distances, the predominant sonorities of the whole-tone and octatonic scales are quite preserved[26], although the music is in constant modulation. This contributes to perceive the prelude as a clear A-B repeated structure. Both the predominantly whole-tone sonority and the fully realised octatonic passages are neatly captured by the keyscape. The A sections are estimated at or close to 1-2 (second transposition of the whole-tone scale), while the B sections are projected right at 2-1 (first transposition of the octatonic mode). It is worth noting that the space of symmetric contexts used to compute the keyscape does not contain the mystic chord among its categories, but it manages to capture the general whole-tone sensation portrayed by the particular realisations of the chord[27] in the A sections. For the octatonic sections, the chord-scale aesthetic fusion stated by Schloezer is appreciated by a homogeneous estimation in which virtually no chords are distinguishable from the context.

3.5 Additional remarks

3.5.1 Symbolic and audio evaluation challenge

So far, the method has been presented as agnostic with respect the symbolic or audio nature of the input signal. Both representation domains are quite different, though, and a question about the method's performance in both cases has to be raised. In this respect, it is relevant to distinguish between two different aspects. First, the quality of the primary descriptor, in this case the chroma feature. Second, the key estimation and representation methods themselves. The design of more precise features would benefit from an analytical decoupling of both aspects.

To the best of our knowledge, no feature evaluation has been done considering all the possible segmentations of the music, as discussed in Chapter 2. The original Sapp's proposal of keycapes pointed to this direction (Sapp, 2005), as a means for comparing the behaviour of different key profiles. The method can be extended for comparing the performance of different chroma feature implementations, or, as in the following example, for comparing the audio features with an *ideal* approach to chroma provided by the MIDI representation[28].

The Dies Irae from Mozart's *Requiem Kv.626* will serve to explore the evaluation possibilities of the framework. In the top pane of Fig. 3.16, the keyscape computed from a MIDI version is shown, while the bottom pane depicts the keyscape computed from a commercial audio recording. An overall similarity

[26] Both sets are mapped to themselves, a feature of the modes of limited transposition.
[27] In its hexachordal form, the chord is often referred to as the *almost whole-tone scale*.
[28] Which, of course, depends on the particular MIDI encoding.

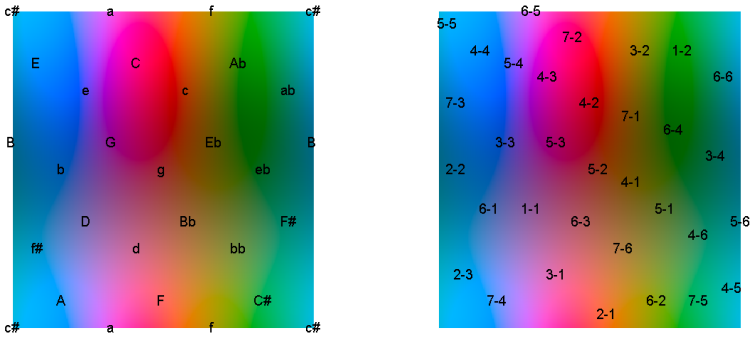

Figure 3.13: Pitch-spaces. Left: major-minor. Right: symmetric modes.

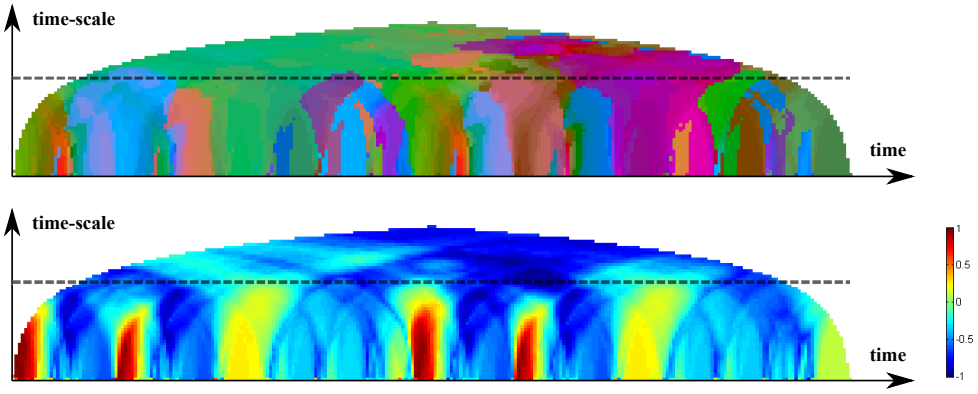

Figure 3.14: Scriabin's *Op.74 n.5.* Keyscape and confidence-scape (major-minor).

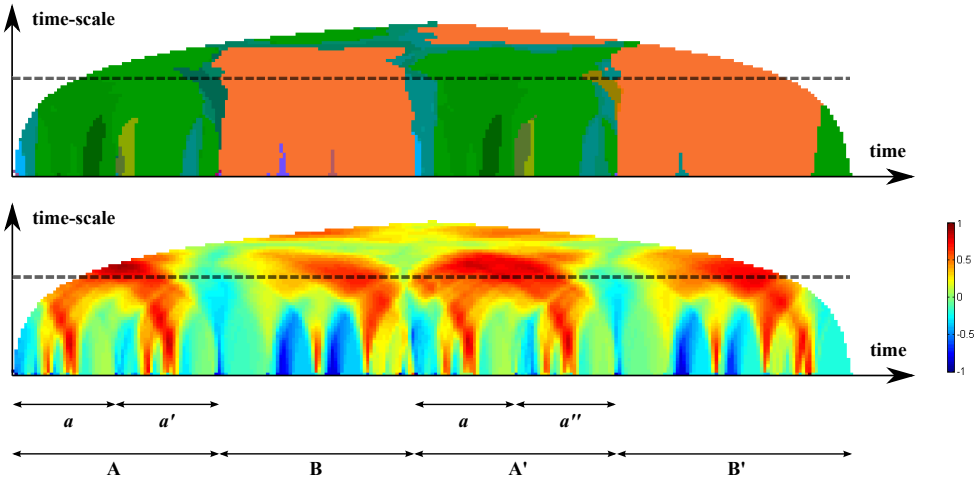

Figure 3.15: Scriabin's *Op.74 n.5.* Keyscape and confidence-scape (symmetric modes).

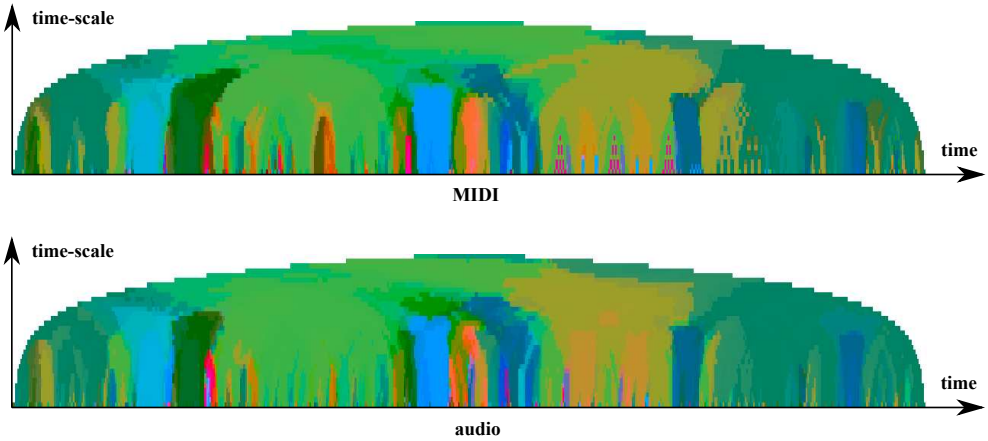

Figure 3.16: Mozart's *Dies irae*. Keyscapes. Top: from MIDI. Bottom: from audio.

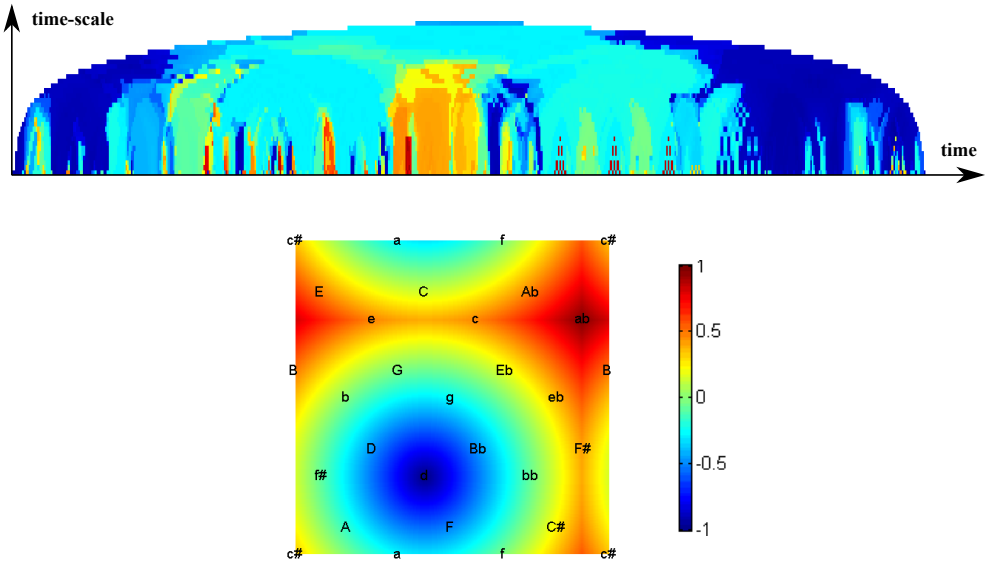

Figure 3.17: Mozart's *Dies irae*. Top: distance-scape. Bottom: pitch-space, coloured relative to *Dm*.

can be appreciated by visual comparison. Quantification is far from straightforward, though. A discussion about that will be done in Chapter 4, in the context of perceptual modelling evaluation.

3.5.2 Scapes of relative distances

The proposed colouring method has two main purposes. First, to represent each contextual category by a distinct colour. Second, to represent the relative distances between the categories in a perceptual manner. While the human vision can *distinguish* many colours easily, it becomes difficult to *iden-*

tify (match them to a given category) more than a few. For the major-minor space, the individual identification of the 24 categories by their colour seems an unlikely task.

Even in the hypothetical case of having a colourspace truly representing the perceptual differences, the colouring method is not practical for *measuring* such distances. There are other colourspaces, though, that can be used for quantify them in a human-readable way. A relative solution, with respect to a given tone centre (point in pitch-space), can be used for that purpose. The top pane of Fig. 3.17 shows a *distance-scape* computed from Mozart's Dies Irae. The bottom pane depicts the pitch-space, coloured according to the distance from the tonic of the movement (Dm), to serve as a colour legend to the distance-scape. It is immediate to observe that most of the piece remains close to the tonic, as well as the section which departs farthest in the piece. This concept will be particularly useful when the number of categories gets larger, as it will be exploited in a systematisation of the method in Chapter 5.

3.6 Conclusions of the chapter

In this chapter, a temporal multi-scale method for tonal context analysis has been presented and discussed in terms of its analytic potential. The main features of the method are:

1. Its conceptual and technological simplicity. The architecture is comprised of standard, open source, non-sophisticated methods.

2. It operates from symbolic or audio signals.

3. A systematic segmentation policy, designed as a practical interactive indexing of the redundant and overlapping data.

4. The exploitation of the human visual capabilities for pattern recognition, a challenging task for computers, for assisting the exploration.

5. A perceptual approximation to the interfacing problem between tonal similarity and human visual perception. The topological aspects of both domains (pitch-space and colourspace) are connected by simple geometrical means.

6. The extension of the idea of keyscape, beyond a mere visualisation, for interactive exploration in more informative spaces. The bird's-eye view of the overall structure is thus linked with he required dimensionality for describing even the most ambiguous music segments.

The method has been illustrated for:

1. The analysis of the tonal context of music pieces or excerpts.
2. A discussion about the summarisation behaviour of the pitch-space at different time-scales, interpreted in terms of the descriptor's nature.
3. A discussion about two different types of tonal ambiguity, framed under the representational capabilities (and limitations) of the toroidal pitch-space of inter-key distances.
4. The introduction of a confidence-scape as a visual tool for assessing the trustability of the multi-scale description.
5. A modification of the confidence-scape for describing ambiguities of type I, by exploiting the information embedded in tonal self-organised maps.
6. A discussion on contextual stability as conveyor of information, even in cases of unreliability of the primary descriptors.
7. A discussion on the summarisation behaviour of pitch-spaces when different (incompatible) contexts are presented simultaneously.
8. The extension of the method to a pitch-space of symmetric modes, able to represent some extreme cases of ambiguity with respect to the major-minor paradigm.
9. A comparative analysis of its performance for symbolic and audio input signals.
10. A modification of the colourspace, for exploring in a quantitative-like way the *actual* distances in pitch-space, relative to a given tone centre.
11. A selection of music examples and application contexts beyond the major-minor paradigm.

The analytical coverage has surrounded a variety of theoretical, compositional and aesthetic aspects, including: classical tonal schemata and symmetry, close and far modulations, ambiguity, polytonality, minimalism and symmetric modes. The method is thus proposed as a useful tool for assisting tonal analysis, extending the usage of a standard descriptor (chroma features) to application contexts of certain degrees of sophistication.

CHAPTER 4

Tonal perception

¿Cómo se puede pensar un cuarto de hora en un minuto y medio?
(El Perseguidor - J. Cortázar*)*

4.1 Introduction

In this chapter, our multi-scale method so far will be adapted, beyond the general descriptive and analytical approaches in Chapter 3, for discussing several aspects of tonal perception. These issues are related to the multidimensional character of the tonal categories, their temporal multi-scale considerations, and the interplay of both domains from a hierarchical perspective.

Cognitive psychology approaches to tonality refer the term *tonal induction* to the development of a sense of key in listeners exposed to music stimuli (Krumhansl, 2004). The modelling of such process is challenged by the elusive description of tonality and by the relatively large and undefined temporal spans required for capturing that sense of context[1]. Even for a short monophonic stimulus, the time-scales involved in the process depend on the complexity of the pitch relations and how (e.g. how fast) the stimulus is delivered over time (Farbood et al., 2012). The effects that timing has in tonal perception remains unclear in the tonal cognition literature, since most of the experimental work has been carried out using retrospective judgements and rather simple stimuli, which do not provide interpretative confidence with respect to the passing of time in realistic music listening situations (Toiviainen & Krumhansl, 2003). The concurrent probe-tone method was conceived for capturing real-time responses from subjects exposed to music of any complexity, providing quantitative ratings of the perceived relative stability of pitch-classes over time. This multidimensional information is related to the concept of key strength in the same terms as provided by the original (stop-and-rate) probe-tone method.

[1] We refer here to contexts with similar perceptual and cognitive implications as the concept of key.

Hence, the concurrent probe-tone method has be used to evaluate computational models of tonal induction. As we pointed in Chapter 2, however, most of the models of tonal induction reported in literature impose rigid temporal assumptions in the analysis of the data, obscuring the interpretation of the tonality phenomena in general, and with respect to time in particular.

Similar concerns apply to the modelling of tonal tension, a central aspect in tonal cognition. As pointed in Chapter 2, the relations between the different categorical levels of description, namely pitches, chords and keys, can be conceptualised in two different but connected domains, which one could name *spatial* and *temporal*. The spatial aspects, whether explicitly by this denomination or not, deals with the pitch relations of tonality *as a system*, irrespective of the particular realisations in music. Both intra-category (e.g. between chords) and cross-category (e.g. between keys and chords) distances have been proposed from theory (Lerdahl, 2001) and modelled by experimentation (Krumhansl, 1990). While a cumbersome amount of empirical research has been done with respect to the hierarchical relations between pitches, chords and keys in spatial terms, very little is known about their temporal counterparts. Some temporal considerations, which one could call *sequential*, have received substantial attention in the cognitive literature, for instance by relating pairs of successive chords including ordering effects (Bharucha & Krumhansl, 1983). However, a temporal multi-scale perspective seems to be required for relating both spatial and temporal domains in a fully hierarchical way. Such connection appears in the theoretical literature, as for instance in the Schenkerian tradition (Forte & Gilbert, 1982) or in the GTTM (Lerdahl & Jackendoff, 1983). However, these approaches oversimplify the time dimension in terms of rhythmic and metric reductions, in which the actual temporal units are absent. That is, the theoretical reasoning does not even consider the tempo in which the music is to be played. Is this perspective, connecting the spatial and temporal hierarchies *while* experiencing music, the one which calls for further attention.

This chapter consists of two case studies. In the first one, we will inspect the impact of time-scale and multidimensionality of description in the evaluation of a simple model of tonal induction, in relation with continuous rating experiments. Particular attention will be driven towards the mathematical artefacts introduced in the comparison between multidimensional time series, depending on the space of representation. In the second study, we will propose a simple model of tonal instability in the context of our general multi-scale method. To do so, we will elaborate on the embeddable nature of any contextual description, and on the concept of keyscape as a joint representation of both spatial and temporal hierarchies in a music piece. As references, we will consider Lerdahl's model of tonal tension, as well as empirical ratings of tonal tension captured by a stop-and-rate methodology. In both case studies, we will reuse existing data from experiments reported in the literature.

4.2 Study I: Tonal induction modelling

4.2.1 Background

Most empirical methodologies aiming to model tonal context cognition, usually under the terms sense of key or key induction, have relied upon stop-and-rate retrospective judgement tasks (Krumhansl & Kessler, 1982). In these settings, listeners are exposed to some music stimulus, after which they are asked to rate a certain subjective perceptual magnitude. These methods provide a notable control of the experimental variables, but they present interpretative concerns about the nature of the captured features, given that the sense of key is derived from indirect measurements[2], and because the ratings are produced after the stimulus (Krumhansl, 1990). Modelling tonal cognition dynamics, conceived as the evolution of the sense of key as music unfolds in time, is even more problematic and time consuming by these approaches, as recognised in (Lerdahl & Krumhansl, 2007).

Real-time response experimental tasks have been proposed to deal with some of these problems, such as the concurrent probe-tone method (Toiviainen & Krumhansl, 2003). Under this approach, a realistic complex music stimulus is used, and the probe tones are played concurrently along music listening. Subjects rate continuously the goodness-of-fit between the probe tones and the music stimulus by dragging a slider in a non-stop setting. Two important aspects make this approach differ substantially from the classical probe-tone tasks. First, the probe tones sound simultaneously with the music instead of being presented after the stimuli. While this may be seen as an advantage with respect to the retrospective judgements, it represents a very different musical reality. A quasi-continuous tone sounding along an independent tonal discourse can have a variety of psychological effects on listeners, and it is not clear whether the listeners would maintain a constant criterion in their attention and judgements. The second problem comes with the calibration of the rating responses, in terms of the consistent use of continuous scales and the unpredictable motoric mediation delays[3]. This has an enormous impact on the statistical significance of the intra-subject and inter-subject analysis.

Several problems arise for evaluating computational models of tonal induction, when continuous ratings are taken as reference for comparison. Temporal scale is a fundamental parameter for describing contextual musical features, as it is

[2]For modelling the sense of key in listeners, the probe-tone method captures the goodness of fit between the stimulus and each probe tone. The ratings are then interpreted in terms of the relative perceptual stability, and the combination of the 12 ratings in a vector is taken as the *key profile*.

[3]See (Koulis et al., 2008) for a detailed analysis of these factors in a far more simpler task (pitch perception). To the best of our knowledge, no studies involving the perception of higher tonal concepts (of the kind of tonal tension) and continuous annotations have reported these critical factors properly.

critical for time series analysis in general. Many window-based key-finding models from symbolic notation (Krumhansl, 1990; Temperley, 2001) analyse the stimuli in beats or bars, in order to produce 12-D profiles comparable to the probe-tone ratings. In the audio domain, in which metric segmentation is not always reliable, this is often implemented by an overlapping sliding window of constant duration (Gómez, 2006). Similar time-scale decisions apply to most models inspired on auditory processing (Leman, 2000), by using a decay constant to simulate leaky memory processing. One recurring argument about the time-scale selection is that it should fall within the short-time memory constraints, agreed around 3-8 seconds. The final choice is then tuned according to rhythmic or metric assumptions, or as a compromise between smoothness and discontinuity of the resulting signals (Toiviainen & Krumhansl, 2003). There have been few systematic attempts of understanding the role of temporal scale in the modelling of tonal induction from probe-tone methods, most of them around the discussion about short-term and long-term memory implications. In (Leman, 2000), the echo constant of the proposed global image model was manipulated in steps of 0.2 seconds, spanning a range of 0.2 to 5 seconds, in order to find the optimal value fitting empirical correlation trends among several context inducing sequences (Krumhansl & Kessler, 1982).

Among the few models of tonal induction which are explicitly concerned with the time-scale issue, it notable the lack of empirical validation with respect to perception. Manipulation of the leaky memory decay constant has been proposed as an interactive parameter in a real-time visualization model of tonal induction (Toiviainen, 2008), but no attempts of empirical validation were done. Several window sizes have been explored using Chew's spiral array for modelling tonal boundaries (Chew, 2006). In her work, evaluation was based on coarse structural (offline) annotations by experts or key indications in the score. As mentioned earlier, the concept of keyscape (Sapp, 2005) provides a systematic approach to the time-scale problem, as it does our multi-scale method so far, but also lacking perceptual validation from experimental evidence.

Aside from time-scale, evaluation issues arise from the multidimensional (12-D) nature of the involved time series. Tonal induction models are often evaluated through an indirect space of key strengths. This mapping is usually achieved by correlating the 12-D vectors with the ring-shifted KK-profiles or similar templates. Vectors from both ratings and model are projected in this space in order to be compared, and model evaluation is discussed in this after mapping scenario (Krumhansl, 1990; Toiviainen & Krumhansl, 2003). This mapping has been proposed in a variety of dimensional reduction solutions, most of them for visual comparison purposes, including multidimensional unfolding into toroidal inter-key spaces (Krumhansl & Kessler, 1982; Krumhansl & Toiviainen, 2003) and self-organized maps (Janata, 2008; Toiviainen, 2008).

Despite their visual informativeness, such frame-based representations do not provide a proper quantitative comparison between models, given that they do not represent the actual listener's ratings, but their relationships with predefined (assumed) categories. Very few direct quantitative careful annotations of key induction dynamics along a complete complex music piece have been approached. In a case study using the Krumhansl and Kessler's key-finding algorithm, a Bach's prelude was analysed offline by two expert music theorists, which provided bar-based estimations of key according to a variety of musicological criteria, and rating up to four possible weighted key candidates (Krumhansl, 1990, pp. 96-106). In most studies, however, subjects rated their perceptual relative stability of pitch-classes instead of direct key strengths, due to obvious methodological reasons. Despite general warnings are posed about this issue (Krumhansl, 1990), no study has covered the quantitative stress or evaluation artefacts introduced by the different dimensional projections. Such concerns include the mapping through correlation, which is non-metric in nature since it does not hold triangular inequality. Thus, the comparison between two input vectors, the actual ratings and the prediction, is not equivalent to their comparison after mapping, and this has a quantitative impact on the evaluation. In experimental settings capturing continuous ratings, the autocorrelation of the involved signals is a very relevant factor as well, introducing notable statistical significance problems. However, there is no consensus about its treatment in literature (Schubert, 2001), which challenges the comparative analysis of models and the reproducibility of the experiments.

4.2.2 A continuous rating experiment

In this study, we show the impact that time-scale and multidimensional mappings have in the evaluation of a simple tonal induction model against ratings collected by the concurrent probe-tone technique. We reanalyse empirical data from a previous experiment (Toiviainen & Krumhansl, 2003), to unveil some limitations of working with pre-existing (often preprocessed) data. From the tonal perception modelling standpoint, we question the timing conventions in tonal cognition. From the evaluation perspective, we discuss the mathematical artefacts introduced by the usage of indirect measurement spaces. To do so, we propose an adaptation of our temporal multi-scale analysis method.

4.2.2.1 Method

Empirical data gathering In (Toiviainen & Krumhansl, 2003), concurrent probe-tone tasks were conducted for capturing real-time responses to Bach's organ duet BWV 805 in A minor. While the full experimental details can be consulted in the original article, the main ones follow. 8 highly trained musicians participated in the rating task. Stimuli consisted on 12 versions of the duet, each of them including a quasi-continuous probe tone from the

chromatic scale. A church organ timbre was used for rendering the resulting files. In order to prevent blending and peripheral sensory dissonance, probe tones were slightly interrupted at the end of each measure, and there were presented to the opposite ear than the duet through headphones. The tempo was fixed to 75 bpm. After a training session with the interface using similar stimuli, subjects were exposed to the 12 versions in random order, adjusting a horizontal on-screen slider with the mouse according to the perceived degree of fit between the music and the probe tones. The position of the slider was recorded each 200 ms. The raw data were smoothed by averaging over each 800 ms. and then averaged across subjects. This resulted in a 12-dimensional time series of 216 samples, representing the relative perceived stability of each pitch-class over time, which is the empirical data used in what follows (hereafter, *ratings*).

Model of tonal induction In order to show the evaluation impact of time-scale and dimensional mapping, a very simple model of tonal induction is implemented, avoiding sophistications that would obscure the interpretation. Most of the discussion, however, would apply similarly for more refined models, provided that sliding windows or decay constants were used for analysing the music signal.

The stimulus[4] (free of probe tones) is first converted into a chroma representation. Both MIDI and audio representations were tested, in order to observe their differences. In the audio domain the Harmonic Pitch-Class Profiles (HPCP) (Gómez, 2006) are computed from the signal every 50 ms. This results in a 12-dimensional time series, representing an estimation of the pitch-class relative energies. In the MIDI domain, the process is identical as described in Chapter 3. Then, our general multi-scale temporal segmentation is applied to the preprocessed signals. The minimum time-scale is fixed to 800 ms., matching the sampling period of the empirical ratings, and the maximum window size fits the whole musical piece. A hop-size of 800 ms. is used for all the time-scales, so as to provide temporal alignment with the ratings time series. The alignment matches the endings of the corresponding analysis and rating frames, that is, tonal context is defined from past to present exclusively. This approximates the experimental setting conditions, in which participants rated continuously along listening, without having access to the *future* of the stimulus. To avoid artefacts in the estimation as time-scale increases, only full-sized segments of music are analysed, so we discard the beginning of the piece for each time-scale accordingly.

A pitch-class profile is then computed for each segment. In the audio domain, the HPCP vectors within the frame are averaged and normalised to estimate

[4] As used in the original experiment, provided in both MIDI and audio encodings by Petri Toiviainen (November, 2011).

the relative pitch-class energies of the segment. In the symbolic domain, we adapt the method implemented in the MIDI Toolbox (Eerola & Toiviainen, 2004) to get the relative duration of each pitch-class within the frame. Parncutt's "durational accent" (Parncutt, 1994), applied by default in the original algorithm, is not used here because of the particular experimental conditions. The psychological effect of sustained pipe organ sounds, as it was the case for the stimulus, is not likely to be well represented by the predominant effect of the onsets assumed in Parncutt's model. The perceptual presence of sustained notes against shorter figurations (which abound in this piece) is substantially different when played by a harpsichord compared with a pipe organ, a distinction not considered by Parncutt. The resulting 12-D time series, one for each time-scale of analysis, are the output of our simple tonal induction model, and the information to be evaluated against the empirical ratings. Since both signals from audio and MIDI resulted in very similar keyscapes, we decided to use the time series from the audio in the rest of the study.

Measures Quantification will be discussed with respect to two different representations of the tonal-related time series, as both alternatives are usual in literature. For both the model and the ratings, we will consider: a) the 12-D time series described above; b) the 24-D time series obtained after correlation of the 12-D vectors with the ring-shifted KK-profiles. As a measure of similarity, the frame-to-frame distance between the model and the ratings is first computed as one minus the correlation of the corresponding 12-D vectors. The final comparison between both multidimensional time series is computed as the root mean-square of the frame-to-frame distances along time. This value is taken as the global distance between the model and the ratings.

Representation Quantification of the overall similarity between both time series, however, is incomplete for representing the quality of a continuous tonal induction model. Aside from the numeric result, it is of interest to evaluate which sections of the music stimulus are being poorly represented by the model and for which ones the algorithm performs the best. A visual inspection method is proposed as a complementary qualitative validation of the model, with respect to both the impact of time-scale and the alignment of similar frames between the ratings and the model time series. This visualization is used as an index for exploring the different dimensional mappings over time and across time-scales[5]. The algorithm is an adaptation of our general multi-scale analysis method (see Chapter 3), in order to be fed with the output of both the tonal induction model and the experimental ratings time series.

[5]The interface used for this particular exploration context is discussed in Chapter 6.

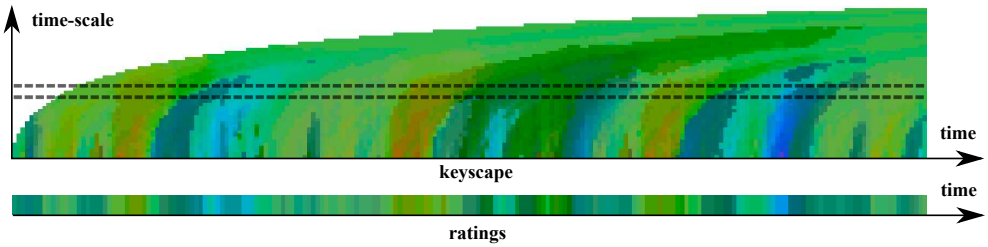

Figure 4.1: Bach's *BWV 805*. Keyscape and ratings.

4.2.2.2 Results

Qualitative and quantitative analysis In Fig. 4.1 the keyscape computed from the stimulus is shown. The resulting coloured centroids for the ratings time series is shown below the keyscape, aligned in time. A qualitative visual analysis of the figure suggests an overall coarse alignment between the estimates and the ratings, as similar colours account for close distances in pitch-space. The influence of time-scale in the model's output is also evident. This visual representation introduces a notable stress through the mapping of the 12-D estimates into the 4-D inter-key space, and additional distortion appears as a consequence of the geometrical colouring through a 3-D projection of the pitch-space. However, it provides a visual intuition about the quality of matching between the involved time series, and a useful index for interactive exploration along time and across time-scales (see Chapter 6).

With respect to quantitative analysis, we considered time-scales ranging from 800 ms. (the sampling period of the perceptual ratings) to 60 secs. Larger time-scales were discarded for two reasons. First, the frequent modulations of the music stimuli make very unlikely that subjects were considering such long segments in their ratings. Additionally, we want to compare a representative amount of music over time and the impact of time-scale over the same stimuli, but larger time-scales require discarding the corresponding segment at the beginning. Consequently, we remove the data for the first 60 seconds from both the model and the ratings, and the time-scales above that value from the model. As side effects of this decision, we evaluate only the most interesting and tonally rich excerpt of the music stimulus, and we avoid potential response artefacts during the subject's ratings at the beginning of the piece. Fig. 4.2 shows the root mean-square deviation (considering error as the frame-to-frame distance) vs. time-scale, computed for both the 12-D and 24-D representations discussed above. Two aspects become evident from the figure. First, the best matching is achieved for time-scales of 11.5 s. (12-D) and 15 s. (24-D), annotated as dashed horizontal lines in Fig. 4.1. Both curves show a clear trend with a minimum around that value, and they evolve quite similarly across time-scales. Second, the 24-D representation introduces a clear metric artefact in

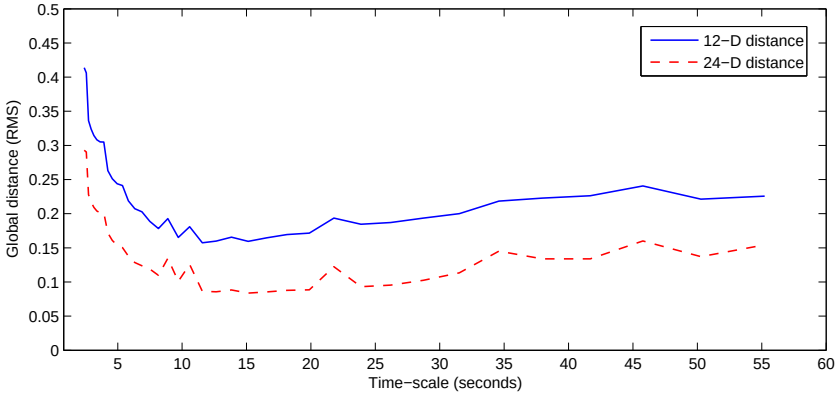

Figure 4.2: Bach's *BWV 805*. Global distances vs. time-scale.

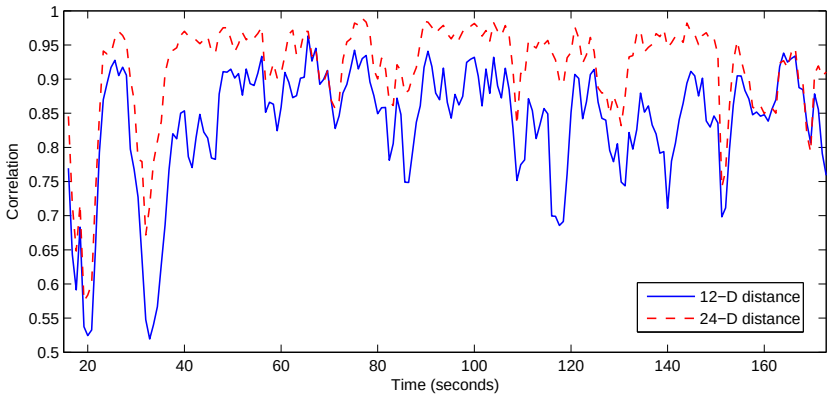

Figure 4.3: Bach's *BWV 805*. Frame-based correlations over time.

the evaluation, seeming to produce a noticeable better matching.

4.2.2.3 Discussion

The best performing time-scales, around 13 seconds, are significantly larger than the agreed conventions for short-term memory, usually in the range of 3-8 secs. Several explanations could account for this result. First, we could argue that such large time-scales can be actually involved in the processing of contextual tonal information, as several hierarchical tonal theories support (Lerdahl, 2001). This specific music stimulus is far more complex than simple chord progressions, but it is quite tonal, so the main tonic references are not difficult to be sustained for highly trained musicians, bridging the short tonicisations and uncertainties even in the presence of foreign probe tones. Some aspects of the composition would probably contribute to this: the main theme lasts 8 bars, which corresponds to 12.8 secs. at 75 bpm. The piece is

structured around such duration, not only for the thematic expositions but for the progressions as well. Moreover, the writing style can impose a strong thematic listening from the very beginning, when the main theme is presented in isolation by a single voice. A second factor which can contribute to such results has a signal processing nature. The available empirical ratings data were the result of, at least, two averaging processes: one intended to minimize the erratic fluctuations of the subject's motor action during the task, and an additional inter-subject averaging. The resulting signals, although downsampled to reduce its autocorrelation, keep a considerable degree of smoothness. In our temporal multi-scale model, smoothing is inherent to the use of large time-scales, and this may contribute to a better matching with the ratings time series.

With respect to metric artefacts, we should notice that the 24-D space used in the evaluation by key strengths is heavily coupled. In fact, the 24-D vectors resulting from the correlation with KK-profiles are just extended versions of the input 12-D vectors, covering a reduced 24-D subspace. However, correlation between vectors after mapping assumes a true space with 24 degrees of freedom, which results in a better fitting for essentially the same information than the original 12-D vectors. Fig. 4.3 shows the evolution over time of the frame-to-frame correlations computed for the best performing time-scale (according to the 12-D representation). Here, the first 15 seconds of music were removed, in order to analyse a more significant portion of the music stimulus. The figure clearly shows that the 24-D mapping outperforms the plain 12-D representation for virtually all frames, but this is actually a mathematical artefact. It is also evident that, even quite similar to each other, both measures are not equivalent, as can be observed from their temporal evolution. This is not surprising, since correlation does not hold triangular inequality, so we cannot expect it to behave as a metric in the strict sense. Both distance time series are, however, well correlated (r =0.89, p <0.001)[6]. This may indicate a certain usefulness of the evaluation in the 24-D space. Even if it cannot be considered in absolute (quantitative) terms, it seems to account for many of the temporal points for which the model performs poorly.

This last point, related to the evaluation in terms of target spaces, deserves particular attention. The KK-profiles are assumed to be the categorical references with respect to the perceived relative pitch-class stability, but these profiles were derived from the responses of participants exposed to clearly unambiguous tonal contexts. However, that music consists of unambiguous tonal material is not the case in general. Much experimental work seems to confirm

[6]Given the notable autocorrelation of the involved time series in our study, the p values of the correlations have been computed by a method identical to the one described in (Pritchard & Theiler, 1994), for the one-dimensional case. In addition, they have been double checked after considering the effective number of degrees of freedom of the signals, as described in (Pyper & Peterman, 1998).

that the KK-profiles are proper indicators of how listeners respond to music stimuli which is similar to clear tonal (major-minor) contexts. In this sense, it is expected that the listener's ratings will be projected closely to the model's prediction *if* the stimulus itself is well suited for the profiles that created the space. However, this is far from claiming that listeners are *using* these categories as the reference to which *any* music stimulus is compared. The use of a target space for a general evaluation against experimental ratings seems conceptually inappropriate. As Krumhansl warns with respect to the key profiles and the resulting inter-key spaces: "[...] the patterns arise from the tonal hierarchies and not direct perceptual judgements of key distances" (Krumhansl, 1990, p. 40). It is evident that by using other key profiles, such as Temperley's (Temperley, 2001) or Aarden's (Aarden, 2003), this kind of evaluation would result in different similarity results.

In any after-mapping scenario, the quantification of similarity between the model and the ratings depends on the ability of the pitch-space for representing the 12-D information from both sources. For instance, it seems unlikely that the probe-tone ratings for highly ambiguous stimulus (with respect to the KK-profiles) would map consistently in the major-minor space. This is observable in any keyscape computed from highly ambiguous music with respect to the pitch-space: even small changes in few dimensions of the segment's profile can result in unpredictable activations of the pitch-space, and consequently in erratic centroid unfoldings. As it was discussed for Scriabin's example in Chapter 3, the mapping results can be radically different depending on the target space. As empirical studies about *polytonality* have suggested (Krumhansl & Schmuckler, 1986)[7], "no evidence was found to support the notion that the two tonalities were perceptually functional as organizational entities abstracted from the perceptually immediate context" (Krumhansl, 1990, p. 233). Rather, the results supported that it was the distribution of tones in the context what determined the ratings. The best fitting of the probe-tone ratings was consistent with Van der Toorn's octatonic interpretation of the passage (Van den Toorn, 1983). This suggests that the sense of context is presumably constructed from the evidence irrespectively from a priori categorical assumptions. It would be of high interest to perform a concurrent probe-tone study for a music piece of certain length and complexity with a high ambiguity (of type II), and test the ratings with respect to the major-minor profiles in comparison with the tone distributions alone. Unfortunately, this kind of experimental data is prohibitively expensive to collect, and we had no access to further existing datasets.

An inverse thought exercise, constrained to a mere representational domain, can be proposed to foresee the result of such a study: we can observe the

[7] This study about the tonal organisation of the so-called Stravinsky's *Petroushka chord* constitutes a landmark with respect to the perceptual relevance of the assumed key categories.

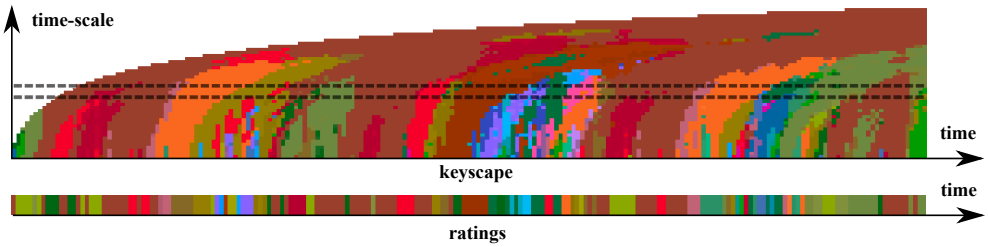

Figure 4.4: Bach's *BWV 805*. Keyscape and ratings (symmetric modes space)

.

effect of mapping the output of tonal induction models and ratings for non-ambiguous stimulus in ambiguous target spaces. In Fig. 4.4, the keyscape and the ratings from Bach's duet are projected in the space of symmetric modes discussed for Scriabin's example. The same reference time-scales as in Fig. 4.1 are depicted as dashed horizontal lines. It can be observed, by comparison with Fig. 4.1, that the matching between the model and the ratings do not follow the same patterns as for the projections in the KK-space. Under the light of the current discussion, we do not include a quantitative comparison between the two mappings, because of the different dimensionality of the target spaces (24-D in the KK-space and 33-D in the space of symmetric modes).

One could then consider up to which point our simple model of tonal induction is successful. The notable correlation between the model and the ratings for most of the frames could be dependent on the fact that the music stimulus is statistically well represented by Krumhansl and Kessler's key profiles. Since the profiles themselves were derived from empirical responses to unambiguous major-minor stimuli, they are good predictors of the listeners' ratings for stimuli of similar characteristics. For these cases, simple algorithms as our duration-based pitch-class distribution have proved to capture the elementary aspects of tonal perception (Krumhansl, 1990). The confidence-scape computed for the Bach's duet (not shown here) is mostly non-ambiguous (of type II) with respect to the KK-space, so it could be in this respect that the model of tonal induction succeeds. General conclusions beyond this point are not supported by the experimental data. We can conclude that, in a general case, the most appropriate information to be used in a quantitative evaluation of tonal induction models, against ratings from probe-tone methods, are the plain 12-D output vectors from the model, and not their projections in target spaces. These spaces, however, can be useful tools for analytical exploration purposes.

About fixed time-scales in modelling The variable performance of the model along time leads to questioning the motivations of using fixed time-scales in the modelling of tonal context cognition. Aside computational convenience, there are no supporting evidence in literature for time-scale to be fixed in

general. Most discussion about time in tonality focuses on the short-term vs. long-term debate, but fixed time-scales are implemented in virtually all cases. However, such temporal constraint seems counterintuitive to music experience in general. Even assuming that tonal context is induced in listeners just from the stimuli, which is quite a strong claim in a general sense, we could argue that the time-scales involved in the listeners' processing of tonal information would depend on the tonal material itself, which can be, and in general is, quite variable as music unfolds in time. This is not in conflict with some parsimony arguments claiming for a need of minimizing the temporal memory resources. Such dynamic mechanism would optimise the memory resources, adjusting the required *window size* to the complexities of the information to be processed.

In our multi-scale evaluation scenario, it is evident that the model can outperform with respect to any single resolution by using a dynamic time-scale over time. This would be actually consistent with some intuitions about musical listening. Tonal cognition is unlikely a passive process with respect to time, as attention can be driven towards short-term or long-term activity dynamically, depending on both the musical content and the listener's background and intentions. Under this hypothetical dynamic listening mode, the time-scale would be adjusted along time to find an optimal path across the keyscape maximizing the fitting between the model and the perceptual ratings. However, the design of an experiment for testing this point would be very challenging.

4.2.2.4 Conclusions of the case study

This study has raised two issues related to the evaluation of models of tonal induction over time, in which the concurrent probe-tone method is used. A systematic exploration of the time-scale involved in the model suggested that the agreed conventions about the short-term memory limitations may not apply to this particular case. A more relevant problem with respect to quantification was identified in the use of target spaces at the evaluation stage, evidencing the distortion introduced by both the statistical artefacts and the choice of the space of categories. This last point highlighted the lack of generality of the evaluation results. Both aspects have been discussed by adapting our temporal multi-scale method to be used for evaluation purposes. The study also called for more adequate quantification methods, beyond global statistics, in which the performance of the model needs to be characterised along time. In passing, we argued that the availability of shared datasets of raw perceptual ratings, methodologically challenging and expensive to collect, would benefit the research on tonal cognition, facilitating replication studies and model comparison, and helping to improve the statistical adequateness of the methods.

4.3 Study II: Tonal stability modelling

4.3.1 Background

Tonal tension, agreed as a fundamental aspect of the tonality phenomena, has received a notable attention from cognition studies (Bigand & Parncutt, 1999; Cuddy & Smith, 2000; Krumhansl, 1990), musicology (Huron, 2006; Meyer, 1956), performance studies (Palmer, 1996) and music theory (Lerdahl, 2001; Narmour, 1992), among others. It also plays a fundamental role in models of musical tension in general (Farbood, 2006). The concept of tension is central to most of the theories of common-practice tonality. However, each model considers different aspects of such an elusive and complex concept, and a variety of terms are used to refer them. Different models aim to express different psychological realities, such as sense of "closure" or "expectation", and the terms "tension and relaxation" are often referred to as "stability and instability" or "consonance and dissonance". One has to consider whether it is possible to generalise some feature shared by all or most of them. We observe two factors shared by many models. The first element is the assumption of a tonal hierarchy, establishing that a certain instance among a given category set (e.g. a pitch-class or a chord) receives the highest consideration in a certain musical context, and so it becomes the stability centre of that context. The rest of instances in the category set are then subordinate to this reference. Several hierarchical layers are often defined around the reference (e.g. pitch, chord and key levels). The second element is an explicit account of the relationship between the reference and the subordinate elements. This is, for instance, the case of the different scale degrees with respect to the tonic pitch-class. Sometimes these relations are given in quantitative forms (e.g. inter-key spaces or intra-key chord distances).

These two elements appear, under different denominations, in a variety of perspectives about the tonality phenomenon. Arguably the most influential hierarchical theory of music, Schenkerian analysis introduces the concept of "prolongation" (Forte & Gilbert, 1982; Schenker, 1935), whereby some notes are "composed-out" (e.g. as a passing note) from others in virtue of their subsidiary status, and it introduces the concepts of "Urlinen", "middleground" and "foreground" in relation to different hierarchical layers of the music structure. One of the pillars of the Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff, 1983) is a reinterpretation of Schenkerian prolongations, referred to as "elaborations", which, together with the concept of "time-span reduction", builds a "prolongational tree" over which a set of "well-formedness rules" operate. In (Werts, 1983), the concepts of "primary", "secondary" and "tertiary scale references" are introduced to account for "essential events" and their harmonic and non-harmonic "projections". In (Rosen, 1972), conceptual "dissonance" degrees are given to tonal regions in relation to the tonic of the

music piece, so modulations can be conceived as large-scale dissonances. From a psychoacoustic perspective, (Parncutt, 1988, 1989) models virtual pitches in terms of the perceptual "salience" from global acoustic stimuli. In (Tillmann et al., 2000), a top-down "back-reverberation" from key to chord units characterise the hierarchical interplay of the pitch categories in their connectionist model of tonality. A substantial research in cognitive psychology considers the stability relations between "events" and their tonal "contexts" at their methodological foundations, constituting the basis of the probe-tone method (Krumhansl, 1990).

4.3.1.1 Temporal multi-scale approaches to tension

Elaborating upon his previous model of tonal induction (Leman, 2000), which is based on a two-tier short-term memory[8], Leman compares the tonal induction time series computed at both resolutions to each other (Leman, 2003). The result is a measure of "tension" between the surface[9] events (short-term integration) and the tonal reference (long-term integration). This is implemented by pitch mappings of echoic memories with different decay times, and proposed as a measure of the stability and instability of the tonal discourse. In (Janata et al., 2002), the interplay between short-term and long-term memories is even suggested to affect the dynamic *topographical* activity in the human cortex. The need of considering events within their contexts for modelling tension is posed in (Lerdahl & Krumhansl, 2007). With respect to how neural networks are able to capture basic properties of the pitch space from minimal theoretical assumptions, such as octave equivalence and chromatic distribution of pitch classes, Lerdahl and Krumhansl suggest the potential usage of temporal multi-resolution analysis for describing simultaneously several hierarchical levels, in ways that would capture the prolongational components of the GTTM.

4.3.1.2 Lerdahl's model of tonal tension

Lerdahl review and extend the GTTM in his Tonal Pitch Space theory (hereafter TPS) (Lerdahl, 2001), which incorporates a substantial outcome from psychology research. The model of tonal tension and attraction introduced in (Lerdahl, 1996, 2001), is based upon the prolongational component of GTTM, accounting for the *event hierarchies*, and the tracking of the musical events through pitch space paths at several hierarchical levels, accounting for the

[8] One level for local events and other level for their context, each of them operating at a different time-scale.

[9] The term *surface* stands for a variety of concepts, depending on the theoretical framework. In the analytical literature, it is usually related to some variant of the Schenkerian or GTTM concepts of *prolongation*. In Leman's usage, it refers to events occurring near the *present time*, therefore being related to short time-scales. Krumhansl's usage of the term is similar to Leman's.

tonal hierarchies. Lerdahl's model of tension has two components: a) a hierarchical pitch-space, modelling the distances between any two consecutive event-context pairs; b) a set of rules for relating any event with the stable references in the piece. The model operates in two steps. Firstly, a *sequential tension* between any consecutive pairs of events is taken as their distances in pitch-space. This measure is composed as a linear combination of their distances across the *basic*, *chordal* and *regional* spaces. Secondly, a *hierarchical tension* reconsiders the sequential tensions from the prolongational perspective. The algorithm is based on rules of inheritance following the departures and arrivals from/to the governing tonic in the prolongational tree.

4.3.1.3 On the evaluation of tension models

As seen in the first case study of this chapter, the evaluation of models of tonal cognition abound in limitations. Collecting empirical data is costly and prone to many error sources, particularly for capturing time-evolving perceptual variables. The methodological problems worsen as the psychological variable gets of higher level, as it is the case of tonal tension. For such elusive variables, a point can be stressed about model comparison when quantitative measures are involved. The risk of deriving misleading explanations from quantitative correlations of curves without the appropriate supporting hypotheses, is highlighted by Meyer:

> [...] quantification must occur in the context of hypotheses that are explicit and specific enough that the predictive inferences derived from them can be tested -provisionally confirmed or definitively disconfirmed. [...] "tension" [...] is only loosely defined and [...] the roles of the various musical parameters (especially those that are not pitch-specific) that presumably create it are not carefully isolated and delineated. (Meyer, 1996, p. 468-9)[10]

The point to be stressed is the very definition of the variable to be modelled. Tension is a very general term, with a wide range of theoretical, psychological, physiological and semantic implications, which can operate simultaneously. If capturing quantitative measures of tension over time from participants is challenging, it seems even more problematic to discriminate which *modalities* of tension they were actually rating, and to discern whether such modalities evolved along time during the experiment.

4.3.1.4 A tension-related failed experiment

A conflictive decision arises for capturing ratings of tonal tension over time. This methodological conflict was identified during a failed experiment aiming

[10] Quotations on "tension" as published. These comments are part of Meyer's critical review of the analyses of a Mozart's piano movement by Narmour, Lerdahl, Gjerdingen, Bharucha, Krumhansl and Palmer (same special issue of the journal), who discuss a variety of tension-related models.

to model the time-scales involved in the perception of tonal context over time. A description of the method follows.

Stimuli 18 sequences of 2, 3 and 4 chords, ending in 6 points of modulatory interest, were selected from a chordal reduction of Chopin's *Op.28, n°9*. Each sequence was randomly transposed to prevent training effects, and played at 30 and 60 bpm, resulting in 36 chord sequences of durations ranging 2-8 seconds. A second stimuli set consisted in the complete prelude's reduction, played at 30 and 60 bpm. The MIDI sequences were triggered from Sibelius, sonified by a Yamaha Disklavier Pro, and recorded in a professional studio.

Procedure 16 musicians, with varied experience and training in music theory, participated in two tasks. In the first task, the participants listened each chord sequence and rate two sensations: a) the degree of tension/relaxation of the last chord with respect to the previous ones; b) the key clarity of the overall excerpt. Participants were able to listen each chord sequence as many times as wished before rating, so as to encourage the confidence and consistency of the responses. The order of presentation of the sequences was randomised for each participant. In the second task, the participants listened the complete chordal reduction at 60 and 30 bpm in a non-stop mode, while rating the degree of tension/relaxation by mouse action over a large cursor on the screen. The cursor position was recorded every 250 ms. Before each presentation, the cursor was centred and a countdown was presented on screen for preparing the actual task. Stimuli presentation and response collection for both tasks were performed through a Max/MSP graphical interface and headphones, with on-screen written instructions.

Lessons learned The experiment failed its intended goal, since no significant results were obtained with respect to the time-scale problem. However, two methodological conclusions were clear. For the stop-and-rate task, the intra-subject consistency was high for most of the participants, but the inter-subject correlation was not significant enough so as to average across subjects. Very consistent responses were given for the clearly cadencing (relaxing) sequences, but not for the rest. The most tensional sequences were rated consistently only by participants with substantial musical background and training in harmony, but not by the rest. Informal verbal reports about the experiment showed that participants conceived tension/relaxation in varied ways, which seemed to be consistent with the given individual responses. The continuous rating task was far more problematic. The statistical analysis of the time series resulted in non significant inter-subject correlations, so averaging was not further considered. Among the identified problems were: a) variability with respect to the tension concept, as reported by the participants; b) varied sources of autocorrelation in the ratings, which reduced notably the effective degrees of freedom of the

signals; c) motor response delays, clearly noticeable by comparing the curves obtained for stimuli at different tempi, with high inter-subject variability; d) inconsistent usage of the scale range, reported as notably difficult to attend by 9 out of 16 participants.

The non-stop rating experiment, closer to a naturalistic situation, resulted impractical for deriving relevant conclusions from the data. The controlled conditions in the stop-and-rate experiment, regardless of the negative results with respect to our intended goals, seemed to be precise enough to be used in more carefully designed experiments, but at the price of sacrificing the temporal evolution. The overwhelming experimental problems related to the passing of time, and the need of a reasonable confidence about the correspondence between the ratings and the actual variable being measured, biased our decision towards using stop-and-rate methods.

4.3.2 A stop-and-rate experiment

In this section we propose a simple model of tonal instability, by adapting our multi-scale analysis framework. The method takes advantage of the joint spatial and temporal hierarchical information conveyed by the keyscapes, and elaborates our previous concept of contextual stability in time and time-scale (see Chapter 3). From the evaluation perspective, the model is compared with both empirical ratings of tension (collected by stop-and-rate methods) and a theoretical model of tonal tension. We discuss our model as a tentative approach to some unsolved questions about modelling of tonal tension by parsimonious methods.

4.3.2.1 Method

Empirical data gathering Bach's chorale *Christus, der ist mein Leben* in F major has been discussed in detail by both theoretical and empirical approaches. In (Lerdahl & Krumhansl, 2007), this piece is analysed by Lerdahl's tonal tension model (Lerdahl, 2001), and it constitutes the stimuli for a tonal tension perception experiment. This piece is also subjected to a detailed prolongational analysis in (Lerdahl, 2001, pp. 7-29). Empirical tension ratings were obtained from a stop-and-rate method, in which the participants rated the perceived tension at each change in the vertical sonority. For each of these changes, the stimulus was played from the beginning and stopped right after the event to be rated. Participants were thus exposed to excerpts of a growing duration. This left an interesting but non-controlled degree of freedom, that of the *amount* of context involved in the tension perception, since all the past was *available* for each event. The data from both the theoretical model and the perceptual ratings were kindly provided by the authors of the study[11].

[11] From both Lerdahl and Krumhansl's personal communications (June, 2009).

Model of tonal instability Since the term *tonal tension* seems prone to much semantic confusion, we will avoid to use it in our model. Instead, we will approach the concept of *stability* in the specific terms discussed in Chapter 3 for Ligeti's example: a contextual stable section as evidenced by a homogeneous (connected) area in the keyscape spanning both time and time-scale dimensions. The *ground-truth*, from both theoretical and empirical sources, provides *instantaneous* discrete values of tension, although not as a regular time series[12]. This calls, in our model, for a frame-based description of stability. The method consists of performing a *vertical* reading of the keyscape for each time frame. In this case, we used a centred alignment for all the sliding windows at the same time-frame, thus considering past and future frames evenly. While this seems contrary to the setting of the experimental ratings, for which an only-past policy seems to be the appropriate choice, it gets closer to Lerdahl's approach, in which the hierarchical branching from a stable point can be leftward (expectation towards the stable point) or rightward (departing from the stable point). This alignment decision is taken upon the impossibility of isolating factors contributing to tension from the empirical data, while Lerdahl's model relies on a theoretical formalisation, which facilitates a comparative evaluation of certain features. For establishing a reference measure in the model, we questioned about the situation that should be referred to as the most stable in a general case. In tonal hierarchical terms, a main stability reference would be a tonic pitch, part of a tonic triad, sounding in the tonic key of the piece. Such is the case, for instance, of a prototypical final cadence. Within the descriptive limitations of our multi-scale method, this situation corresponds to a vertical reading of the keyscape evidencing the same information from bottom to top, all the way through the embedding (growing in duration) contexts.

Quantifying similarity between the segments at a given time point requires a measure able to manage cross-scale comparisons, while keeping our concept of stability in mind. A direct comparison between pitch-class distributions cannot be used in this case, because the profiles from short segments will differ substantially from those at larger time-scales, without implying necessarily an instability. The tonic triad of a given key should be rated as stable within that key, but both profiles would be generally quite different. The cross-scale comparison can be approached by the fact that the KK-profiles embed substantial hierarchical relationships between pitches, chords and keys. The KK-profile for major keys is almost mimicked by Lerdahl's basic space of five hierarchical levels: tonic, tonic/dominant, triadic, diatonic and chromatic (Lerdahl, 2001, p. 47). That is the reason why a single pitch or a major triad is often (mis)estimated as the homonym key in our model. Even though the correlations between the tonic triad and the key profiles are not high in absolute value,

[12]The figurations in Bach's chorale are not regular with respect to the experiment's definition of event.

the homonym key would still get the best score. This well known drawback of the profiling methods for key estimation at short time-scales[13], becomes advantageous for the task at hand. The inter-key torus seems thus a convenient measurement space. For relying on centroids in pitch-space, on the other hand, it is essential to have very low values of ambiguity of type II (see Chapter 3). An inspection of the confidence-scape for this piece (not shown here, but quite unambiguous for this simple tonal content), decided in favour of using centroids. In order to minimize the stress introduced by the unfolding method, the 4-D KK-space was used. The Euclidean nature of this space provides a simple means for quantification in our descriptor. The computation process projects all the segments at each time-frame, for all the available time-scales at that position, in the 4-D space. The framewise stability measure is taken as the variance of the corresponding 4-D points. The lower the variance, the closer the centroids are clustered, indicating a higher stability. So, the resulting time series is to be interpreted as a measure of *tonal instability*.

4.3.2.2 Results

Before quantification, the instability curve was sampled at the time points for which the ground truth was defined[14]. The correlation of the instability curve with the theoretical prediction is $r = 0.58$ ($p < 0.001$), while the comparison with the ratings gives $r = 0.66$ ($p < 0.001$). The given p values consider the correction according to the effective degrees of freedom (35.4 and 36.4 respectively). The number of independent events provided by the ground truth was 41.

4.3.2.3 Discussion

The quantitative overall comparison gives a poor result, which is not surprising. Our method is not aimed to quantify tonal tension in general, but to capture the most stable time points in the musical piece. A visual comparison between the instability curve and the empirical ratings is depicted in the bottom pane of Fig. 4.5. In the top pane, the keyscape is shown as reference. The bottom pane of Fig. 4.6 shows a comparison between the instability curve and the output from Lerdahl's model of tonal tension[15]. The central pane depicts a piano roll representation of the score, aligned for reference. Since both the theoretical and the empirical signals were given as instantaneous values at the onsets of each event, both curves were time-stretched accordingly to span the actual duration of the events, just for visual convenience. In the top pane of Fig. 4.6, the main branching decisions from Lerdahl's prolongational analysis are delineated[16], including all the strong and weak prolongations (empty and

[13] See, for instance, (Temperley, 2001).

[14] More precisely, the samples were taken at the centre of the corresponding events.

[15] The data corresponds to the hierarchical tension model considering attractions.

[16] Adapted from (Lerdahl, 2001, p. 22)

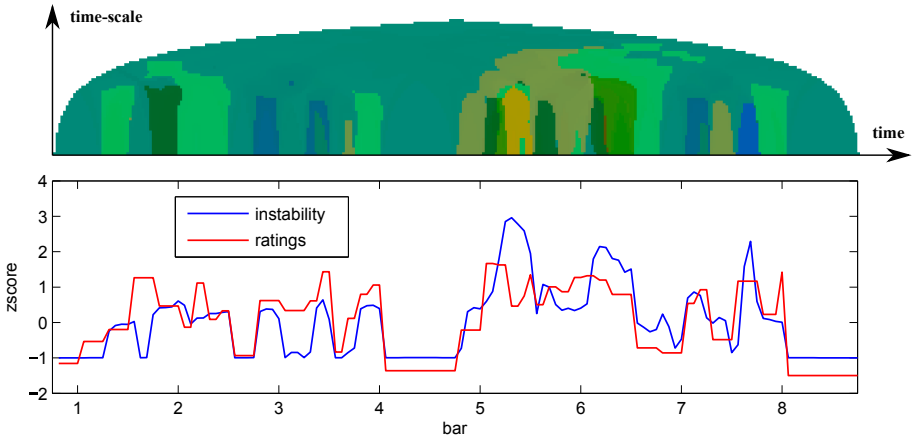

Figure 4.5: Bach's *Christus, der ist mein Leben.* Top: keyscape. Bottom: instability vs. ratings.

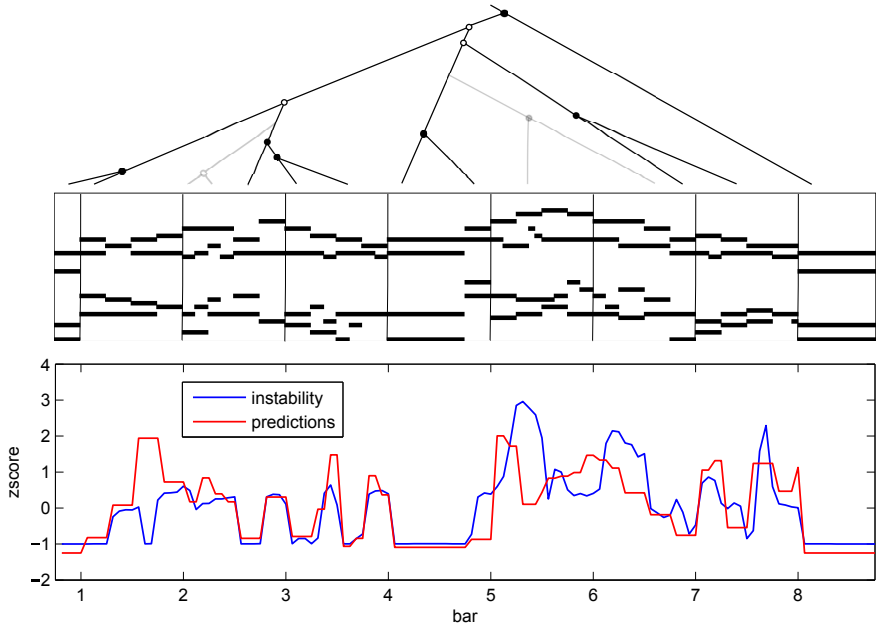

Figure 4.6: Bach's *Christus, der ist mein Leben.* Top: branching. Centre: piano-roll. Bottom: instability vs. predictions.

filled-in circles respectively).

According to Lerdahl and Jackendoff's GTTM, a prolongation is referred to as an event which "composes out" or "elaborates" another event. The former event (the elaborating event) is subordinate to the latter (the elaborated event). In GTTM, the term "prolongation" usually refers to a specific type of elaboration, in which the elaborating event is functionally identical to the elaborated event[17]. In strong prolongations, the event repeats literally, whether in weak prolongations the event is restated in an altered form (e.g. chord inversion). The prolongational analysis follows a top-down approach, from the global to the local, in order to judge the prolongational function of an event from its surrounding context. Despite there are many differences between Lerdahl's concept of prolongation and our instability curve, they share the same essential information for cases of predominantly chordal and cadential music, such as the Bach's chorale. In the prolongational analysis, all the branches in bold are prolongations of the main tonic function, on top of the diagram, which represents the last event as the most stable of the piece (establishing the overall key). In the keyscape, the prolongations of the governing tonic are evidencing as the same estimation (F major) spanning from top to bottom. For these events, the estimations at intermediate time-scales, in its pathway from top to bottom, account for the influence of the governing tonic at that summarisation level. This is related to the prolongational branchings situated at different time-span reductions (the height of the circles in the prolongational diagram).

The lowest values of the instability curve correspond to the branches in bold. Only a strong prolongation (connecting the first and second beats of bar 2) and a weak one (relating the second beat at bar 5 with the third beat at bar 6) were not captured. Both cases are depicted in the top pane of Fig. 4.6 in a grey tone. None of them, however, are prolongations of the tonic. The strong prolongation case, left-branching from the tonic, does not present much interest: it is just a two-beat dominant, leading directly to the most stable tonic, and, accordingly, it is located at the lowest hierarchical level. The weak prolongation, right-branching from the tonic, also plays a dominant role, so their events do not qualify for maximal stability, a fact confirmed by both the theoretical predictions and, to less extent, the empirical ratings. The instability curve, however, peaks at its maximum for the first event of this prolongation. A look at the alignment between the keyscape and the instability curve in Fig. 4.5 explains this result. The peak of the instability curve involves four different tone centre estimations (a vertical reading of the keyscape at those time-frames) close to Em, C, Am and F, resulting in a considerable variance even though they are neighbours in pitch-space. The instability curve presents an additional minimum at the third beat of bar 1, which is in strong contra-

[17] See (Lerdahl, 2001, p. 15) for a comparative terminological discussion.

diction with both the prediction and the ratings. This event corresponds to an F^7 chord created by a chromatic descending voicing, which is misestimated as F. Since our method is agnostic with respect to the distinction between chords and keys, and it is not sensible to voicing (tensional) dissonances, this estimation is interpreted as a returning point to the tonic, thus resulting in a minimal instability score.

On cross-scale alignment The last discussed event, an F^7 chord estimated as an arrival to the tonic F, raises a relevant issue in our model. Given the relative simplicity of the music example, it may seem that the maximal stability is just associated with an F estimation at short-time scales. However, the model does not consider *any* of such F estimations as maximal stability, but only in cases for which F reaches from bottom to top in the keyscape. This obviously depends on the cross-scale alignment choice. The centred decision in this case was taken because Lerdahl's model considers both left- and right-branching around the stable events. Our model was tested covering the whole range of alignment possibilities: from purely *inertial*, considering context from past to present, to purely *expectational*, considering context from present to future, covering a gradual mixture of the two. This alignment policy results in keyscapes with different degrees of skewness, which affects the vertical readings.

The best global quantitative comparison between the instability curve and both the predictions and ratings (see above), was actually obtained from the centred version of the keyscape. However, local aspects of tension could be approached by different alignments, provided that the keyscape estimations were not fallible. For instance, left-branching tensions represent events that *expect* to be resolved in a future stable point, so a future-wise cross-scale alignment would account for that. A high value in the instability curve would represent a large distance between the short-scale event and its larger contexts built from itself to the future. Such a keyscape looks skewed to the left. The opposite behaviour, capturing right-branching tensions would be achieved by a past-wise cross-scale alignment, which results in a keyscape skewed to the right. Branching decisions, relating the present with past or future events, are local, so the exploration of these aspects could or could not improve significantly the global measure. They are better explored interactively, in order to inspect the instability curve locally by gradual skewing of the keyscape[18].

In Fig. 4.7, an example of this feature is shown. The figure sketches three different choices for cross-scale alignment of the sliding windows with respect to the present time (vertical dashed line). The top pane shows a past-wise biased policy, in which context is mostly related to the past, but includes some future as well. The central and bottom panes show a centred and a future-wise policy respectively. In Fig. 4.8, the corresponding keyscapes computed from

[18]The interfacing possibilities of the method are discussed in Chapter 6.

Bach's chorale are depicted. A time point, corresponding to the discussed F^7 chord, appears as a dashed vertical line. As it can be observed in the centred keyscape, the short-scale event spans from bottom to top, thus its minimal value in the instability curve. In the past-wise keyscape (top), however, the previous dominant chord (in light green) is considered as a context of this event at a certain time-scale. This case would introduce a slight source of instability, of the right-branching type (departure from a more stable context). Finally, the future-wise keyscape (bottom) shows clearly that this policy would result in a higher instability, in this case of the left-branching type, as two different contexts are found in the path to the highest time-scales. Lerdahl's model considers this F^7 chord as a left-branching tension that resolves in the dominant (in light green) two chords later, which corresponds to the situation shown in the future-wise keyscape. The interpretation of tension in these terms, however, is sensible to the choice of a proper set of time-scales, an unmanageable feature for our simple model. An *equivalent* aspect in Lerdahl's model is the choice of the time-span reduction level in which the branching decisions are taken.

The perceptual implications of the cross-alignment choice are clear, but difficult to be modelled in experimental ways. In our first case study, we propose that the *listening* time-scales are presumably variable with respect to context. However, there is no reason for such contextual information to be based on the past only, since expectations may also play a role. What is actually being considered as context before/after a modulation point is subjected to much controversy, and it seems to depend on many factors. It is evident that the choice of a given cross-scale alignment also has an impact in the evaluation of models of tonal perception from continuous ratings experiments. In the first case study, we also explored the whole range of alignments in the keyscape. The best matching time-scale was the same as previously discussed, but the best cross-scale alignment was not the purely past-wise. By allowing about a 20% of future in the context, the best matching with the ratings was achieved. Whether this would be indicative of a slight *expectation-wise* listening mode, is of course impossible to elucidate from the data.

On model comparison When different models are compared, it is of interest to observe their conceptual similarities and differences. Lerdahl's model of tonal tension and our proposal of instability curve share the assumption of hierarchical tonic orientation and a model of distances in pitch space. However, their differences are considerable.

1. Lerdahl's model computes tensional states between consecutive events, and builds the tension discourse by inheritance, following the hierarchical patterns down the prolongational structure. It relies on precise event notation, including rhythm and meter information. It is reductionist

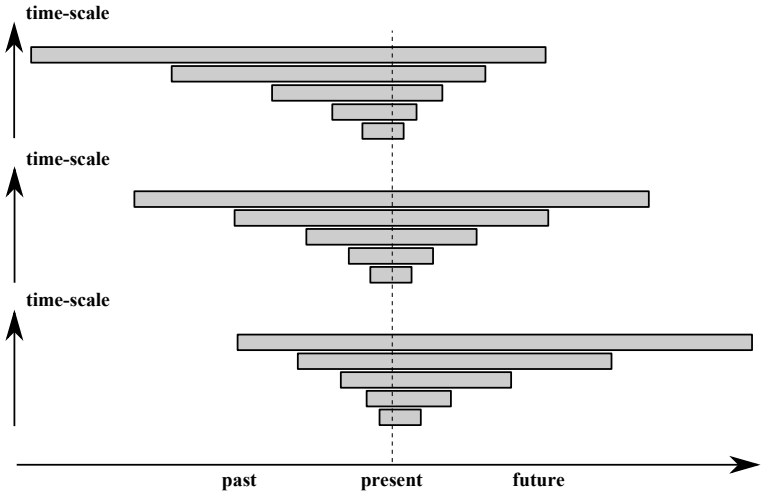

Figure 4.7: Cross-scale alignment policies. Top: past-wise. Centre: centred. Bottom: future-wise.

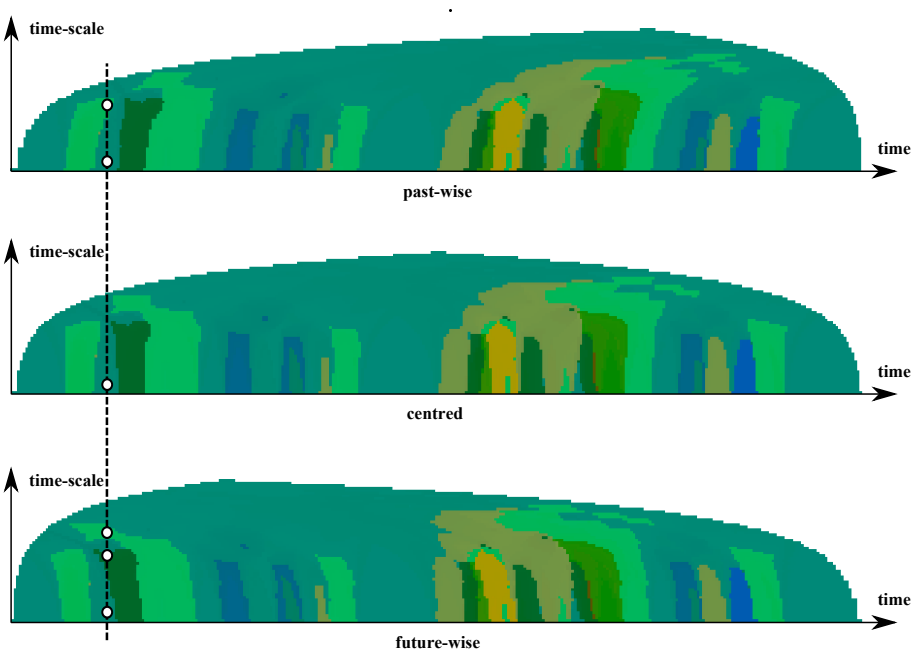

Figure 4.8: Bach's *Christus, der ist mein Leben.* Keyscapes and cross-scale alignment. Top: past-wise. Centre: centred. Bottom: future-wise.

at each time-span level. It requires human computation for deriving a proper prolongational structure. That is, the decisions about which events are relevant for stability and which ones should be removed at the next time-span level, are taken by the analyst. Once branching has been established, the rest of the method (tension inheritance) is mostly algorithmic, although the analyst can still decide about additional rules for specific events.

2. Our algorithm, on the other hand, is fully unassisted[19]. It models the stability at a given time point by a vertical reading of the keyscape, which embeds contextual hierarchical information at many time-scales. The segments are agnostic temporal chunks, not considering any rhythmic or metric information. It operates in audio and symbolic domains alike. No reductionist approach is taken whatsoever, always considering the complete information within each segment. The instability curve considers both past and present events, but their relationships are digested in the progressive embedding of contexts. Lerdahl's concepts of event hierarchy and tonal hierarchy are associated with the temporal and spatial conceptualisations in our model. The inter-key space accounts for the hierarchies of tonality as a system, and the time and time-scale dimensions account for what really happens in time and how it is embedded by its growing contexts. The keyscape summarises and articulates both domains in an appropriate single representation, so as the stability information is readily available from it.

The proposed instability descriptor, despite the discussed drawbacks, is simple and easy to be interpreted. It would thus approach the call for parsimonious models that "could describe the hierarchical relations embodied by the prolongational component of the *TPS*" (Lerdahl & Krumhansl, 2007, p. 357).

4.4 Conclusions of the chapter

In this chapter, the temporal multi-scale method has been extended for discussing the next aspects related with tonal perception:

1. Highlight the implications that time-scale imposes in the evaluation of tonal cognition models.

2. Highlight the implications that multidimensionality has in the evaluation of tonal cognition models.

3. Discuss the problems of a quantitative evaluation of tonal cognition models in terms of target spaces.

[19] Beyond deciding the minimum window size and the number of time-scales, which can reach arbitrary fine-grained resolutions.

4. Discuss the need of temporal-concerned evaluation of the time series, beyond global measures.

5. Extend the framework of our multi-scale method for exploratory evaluations of tonal cognition models in continous rating experiments.

6. Propose a quantification method for comparing two multidimensional time series, for the specific context of tonality.

7. Propose a simple descriptor of tonal instability, which exploits the multi-scale tonal estimations and the properties of the inter-key space.

8. Evaluate the tonal instability curve against empirical ratings of tonal tension.

9. Evaluate the tonal instability curve against a theoretical model of tonal tension.

10. Comparative discussion between the instability descriptor and Lerdahl's model of tonal tension, with emphasis on the prolongational aspects.

In both case studies, the replicability problem has been highlighted. It has been stressed the need of sharing raw data from empirical tests, since they involve careful experimental settings for guaranteeing consistency of results, and they are very expensive to collect. A call for higher transparency in the statistical treatment of the data has been posed in the context of comparative studies between models.

CHAPTER 5

Tonal context generalised

That would guarantee what I call the indifference to harmony and melody.
(Doctor Faustus - T. Mann)[1]

5.1 Introduction

In Chapters 3 and 4, tonal context has been mainly related to a given set of predefined categories, with emphasis on the major-minor key paradigm. In all the cases, the method has involved the comparison of the pitch-class profiles of the music segments with a set of prototype profiles, assumed to represent the categorical references of the tonal system under consideration. Given the multidimensionality of the tonal phenomena, and the implicit ambiguities of the descriptive concepts, both the analysis and the representation have relied on quantitative estimates. Tonal contexts, however, can be conceptualised in different ways.

The description of the musical surface[2] is a fundamental element for analysis, in the sense that it provides the raw material to the analyst. More often than presumably recognised, an appropriate surface analysis is not only a critical stage conditioning analytical decisions, but it also requires considerable effort from the analyst. Systematic approaches to surface analysis are often relegated to a mere instrumental role, as non-analytically relevant *event-counting* algorithms. It is not without irony that many argumentations in analysis are actually supported upon census information of some kind and, perhaps more risky, that such accounts of evidence are often biased by a prior selection of those features that will support the argumentation. This is part of the analytical method, and it does not constitute an issue in itself. In many situations, however, the selective criteria may obscure other sources of relevant evidence,

[1] After Th. W. Adorno.

[2] Here, the term *surface* gets closer to a Schenkerian perspective: surface description as opposed to deep structure, with respect to the referential (analytical or compositional) value of each of the events in the composition.

which can eventually contradict the argument. This is particularly risky in reductionist approaches, for which the connection with the surface may get easily unattended. If analysis is about digest, however, proper information reductions would require the availability of adequate raw materials to be distilled from. Here comes a call for systematicity.

The three main aspects considered by our multi-scale analysis method, namely segmentation, description and representation, are relevant for this endeavour. Our method so far was systematic, but biased at the description and representation stages. The goal of this chapter is to approach a more systematic analysis of the pitch content of music pieces, without assuming any underlying tonal model. That is, the method aims to describe any kind of pitch-based music[3], providing information of analytical or perceptual relevance, while guaranteeing a neutral and objective level of description. Among the possible conceptual frameworks for describing tonal content, we will adopt the most systematic one, namely *pitch-class set analysis*, as it also provides a flexible and widespread analytical lexicon. Although the pitch-class set theory was developed mainly to approach the specific descriptive needs of atonal music, its fully systematic nature makes it adequate for describing any possible combination of pitch-classes. This, of course, includes any scale-based tonal system.

The outline of the chapter follows. First, some relevant characteristics of pitch-class set analysis will be reviewed in terms of dimensionality and descriptive lexicon. Emphasis will be given to the problems of segmentation and representation in set-theoretical analysis. Then, our temporal multi-scale method will be adapted to the specific challenges and opportunities featured by the set-class descriptions. Along the way, the development of the model will be illustrated by examples of use.

5.2 Set-class description

Given that our systematic endeavour deals with large amounts of data, we need first to reframe the main set-theoretical concepts in terms of dimensionality. *Pitch-class* (Babbit, 1955) is defined, for the TET system, as an integer representing the residue class modulo 12 of a pitch, that is, any pitch is mapped to a pitch-class by removing its octave information[4]. *Unordered pitch-class set* (hereafter *pc-set*) is a set of pitch-classes without repetitions in which the order of succession of the elements in the set is not of interest[5]. In the TET system, there exist $2^{12} = 4096$ distinct pc-sets, so a vocabulary of 4096 symbols

[3] We actually assume octave equivalence and a chromatic subdivision of the octave. In this work, in addition, we only consider the twelve equal-tempered (TET) system. On the other hand, many of the set-theoretical principles are applicable to any pitch-based system.

[4] By convention, pitch-classes are represented as ordinals ranging from 0=C to 11=B.

[5] For instance, the set {G5,C3,E4,C4} is represented by the pc-set {0,4,7}.

is required for describing any possible segment of music. Any pc-set can also be represented by its intervalic content (Hanson, 1960). Intervals considered regardless of their direction are referred to as *interval classes*[6]. There exist 6 different interval classes, spanning from 1 to 6 semitones. The total account of interval classes in a pc-set can be arranged as a 6-dimensional data structure coined as an *interval vector*[7] (Forte, 1964).

A relevant relational concept is the *class equivalence*, whereby two arbitrary pc-sets are considered equivalent if and only if they belong to the same class. *Cardinality*, referred to as the number of elements in the set, is the simplest equivalence (Rahn, 1980, p. 74). However, the radical differences among the possible sonorities within a given cardinality does not provide a proper equivalence domain in many practical cases[8]. Interval vector equivalence (hereafter *iv-equivalence*) groups pc-sets sharing the same interval vector. It is thus possible to map any segment of music to one of the 197 different iv-types. Two pc-sets related to each other by transposition belong to the same *transpositional set class*, and they share a similar sonority in many musical contexts. For instance, this applies to all pentatonic segments of music, for which we say that they are *Tn-equivalent*, or that they are of the same Tn-type. There exist 348 distinct Tn-types. The *inversional/transpositional set-class* (*TnI-equivalence*, or Tn/TnI-type) groups all the pc-sets that are related by transposition and/or inversion of their intervals (Forte, 1964). For instance, all major and minor triads are TnI-equivalent (any major triad can be transformed into any minor triad by inversion and transposition), although not Tn-equivalent (no major triad can be transformed into any minor triad by only transposition). The TnI-equivalence maps any pc-set to one of the 220 distinct TnI-types. Both iv- and TnI-equivalence share most of their classes, with some exceptions, named *Z-relations* (Forte, 1964) or *isomeric relations* (Hanson, 1960), for which the same interval vector does not group pitch-class sets under TnI-equivalence (Lewin, 1959). A complete list of set-classes, and an example of description under the three equivalence systems, is given in Appendix A.

5.2.1 On segmentation

A main concern for set-class analysis is the grouping of music events into relevant pitch-class sets. As pointed by Cook, "no set-theoretical analysis can be more objective, or more well-founded musically, than its initial segmentation." (Cook, 1987, p. 146). Forte himself was of course sensible to this problem, as elaborated in *The Structure of Atonal Music*, where he clarifies that "By

[6] Also known as *unordered pitch-class intervals*, *undirected intervals* and *interval mod 6* (Rahn, 1980).

[7] For instance, any diatonic scale is represented by the interval vector ⟨254361⟩: 2 semitones, 5 tones, 4 minor thirds, 3 major thirds, 6 perfect fourths and 1 tritone.

[8] For instance, comparing the sonority of the diatonic {0,2,4,5,7,9,11} and chromatic {0,1,2,3,4,5,6} heptachords.

segmentation is meant the procedure of determining which musical units of a composition are to be regarded as analytical objects" (Forte, 1973, p. 83). The proposal here follows the argument that any a priori selective segmentation, whether justified by theoretical or perceptual criteria, is necessarily biased, and that the pursuit of a neutral level of analysis can only be approached by a full systematisation of the initial segmentation. In this respect, and without claiming that a neutral level suffices for analysis (Deliège, 1989; Forte, 1989; Nattiez, 2003), this work aims to provide an appropriate means for describing segmentation "not as something imposed upon the work, but rather [...] as something to be discovered" (Hasty, 1981, p. 59).

Among the many systematic approaches to set-class description in literature, to the best of our knowledge, no one has achieved a fully systematic and hierarchical multi-scale description. In virtually all cases, the methods do not pursue systematicity in itself, but to reveal specific analytical aspects of the music corpora. That is, the required analytical bias defines how systematicity is to be understood and applied. The most exploratory *event-counting* approaches are not an exception in general, since these *events* are subjected to prior constraints in most cases. Since our intention here is to pursue systematisation first and to explore its analytical potential later, we will not attempt to survey the considerable amount of proposals in this respect. Instead, we will comment on a particular approach which, in our understanding, includes and summarises the most common characteristics (and limitations) featured by most of the methods. It also constitutes one of the very few explicit attempts to pursue a systematic segmentation, description and representation, being thus aligned with our goals.

In (Huovinen & Tenkanen, 2007) a systematic segmentation algorithm is proposed under the concept of *tail-segment array*, whereby every note in a composition is associated with all the possible segments of a given cardinality that contains it. Four limitations, all of them derived from the concept of tail-segment array, can be highlighted from this method. The first is the note-based indexing in terms of *tail segments*, for which vertical sonorities are arbitrarily ordered from the lowest one. This results in many segments containing only some of the notes in a vertical chord. It is unlikely for most of these segments to have analytical or perceptual pertinence. Second, by imposing cardinality as the main parameter, every note is interpreted in the context of segments of the same number of pitch-classes[9]. This results in rather artificial segmentations, eventually up to the point of meaninglessness. For instance, in a passage composed exclusively from an hexachordal set under a segmentation of cardinality 7, the tail-segments are extended endlessly in order to find the missing pitch-class for cardinality completion. The description of such a passage in terms of cardinality 7 would be misleading. Third, a tail-segment array is

[9]Cardinality is the most recurrent constraint in set-class systematic analysis.

generally comprised of varied combinations of set-classes, therefore obscuring the interpretation of the output. The authors actually warn about a *majority* of tail-segments (in a given tail-segment array) probably not qualifying as *good segments*, but nevertheless they are averaged in their descriptor. One can hardly ignore the interpretative consequences of an average taken over a non-qualified majority of segments of an arbitrary mixture of set-classes. The fourth problem is the non uniqueness of the tail-segment array associated with a given time point. There is a systematic conflict among all the tail-segment arrays associated with notes of the same vertical chord. The described *continuity* in the measurements (Huovinen & Tenkanen, 2007, p. 169) is ill-defined. In their analysis of Bach's *Es ist genug*, for instance, the proposed *changes* in the measures are not considered with respect to time, but to a stream of notes *successive in some sense* (p. 163).

The technique proposed next addresses the temporal continuity limitation, and approaches systematisation from a more holistic standpoint. Along the way, the mentioned representational limitations of the method in (Huovinen & Tenkanen, 2007) are solved.

5.3 Temporal multi-scale set-class analysis

The segmentation policy of our general multi-scale analysis method can be combined with an arbitrary analysis function to describe the segment's content. In what follows, our method will be adapted for characterising the set-class content of each segment.

5.3.1 Systematic multi-scale vertical segmentation

The pursuit of a systematic segmentation method requires, in principle, the identification of all possible segments of music. Two different algorithms are proposed here: a) a truly systematic method, which exhausts all possibilities to provide the ideal solution; b) an approximate technique, more practical for interacting with the data. For the sake of argumentation, the latter will be presented first, since it provides a proper descriptive context and it constitutes a good approximation for most cases. The full systematic method is postponed for the applications in which completeness of representation is necessary.

The approximate method is identical to the segmentation policy of our multi-scale method: many overlapping sliding windows, covering the whole range of time-scales, parameterised by the minimum resolution and number of scales. Centred cross-scale alignment is used, so each segment is indexed by its duration (time-scale) and its centre location in time. Segments are then vertical chunks, as it can be performed from an audio source. This constitutes the unique, but important, limitation of the segmentation method in some analyt-

ical scenarios. As for the keyscapes, however, the method can be applied to individual voices or combination of voices, provided the availability of a proper MIDI encoding of the music.

5.3.2 Feature computation

After segmentation, each frame is analysed as follows. First, the pitch-class set content is computed by checking whether each pitch-class is present or not within the segment. This information is mapped to the class systems previously described: a) iv-equivalence; b) TnI-equivalence; and c) Tn-equivalence. For systematisation completeness, the three class spaces are extended to include the so-called *trivial forms*[10]. With this, the total number of interval vectors increases to 200, while the TnI- and Tn-equivalence classes sum to 223 and 351 categories respectively.

5.3.3 Representation: *class-scape*

The resulting multi-scale class information is then arranged in a time vs. time-scale plot. This structure will be referred to as *class-scape*. As it applies for designing any representation, one has to consider the purpose of the description, in our case in terms of set-classes. Perhaps the simplest task, a useful preprocessing stage in analysis, is to localise all the segments in the music belonging to a given class. Since every segment has been characterised by its class, the problem is solved by a trivial filter: the class-scape only shows those segments matching the chosen class. This is illustrated in Fig. 5.1. The top pane shows the class-scape computed for Debussy's *Voiles* filtered by the set-class 6-35, corresponding to the predominant whole-tone scale of the composition. The bottom pane depicts an aligned piano roll representation of the score for visual indexing of the piece.

As in keyscapes, each point in the class-scape represents a unique segment of music. The location of the points in the plot represents the two time-related parameters: its x coordinate corresponds to the temporal position of the segment's centre and its y coordinate represents the time-scale, that is, the duration of the segment. The higher a point is in the class-scape, the larger the duration of the segment it represents. The segments with the requested sonority, 6-35 in this case, appear as an activated point (in black). The rest of the class-scape, whose points represent the segments with a different class, appear in white. The limitations of this black-and-white representation are evident, since just one class can be localised at a time. The proposed visualisation, however, results practical for visual localisation of a given class out of a minimum of 200 categories (iv-equivalence) and a maximum of 351 (Tn-equivalence). It

[10] The null set or single pitch-classes, the undecachords and the universal pitch-class set.

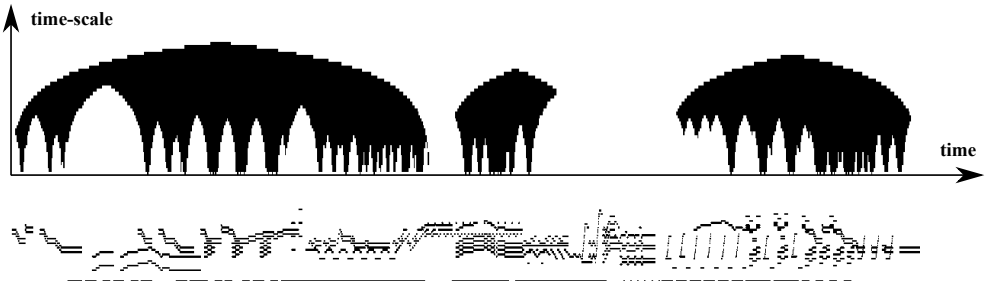

Figure 5.1: Debussy's *Voiles*. Top: class-scape filtered by 6-35. Bottom: piano-roll.

is possible to overcome these limitations, to a reasonable extent, by means of interaction, as it will be discussed in Chapter 6.

5.3.4 Multi-class representation and REL distance

An alternative representation, allowing to inspect all the classes simultaneously, is to assign colours to classes. Given the relatively large number of classes[11], an absolute mapping of classes to colours is unlikely to be informative in general. So, a relative solution is taken. In Chapter 3, the scapes of relative distances were proposed as a human-readable method for *measuring* distances with respect to a chosen key. In that occasion, it sufficed to colour the pitch-space differently, using a linear colourspace for inducing the required monotonic sense of scale. This was possible given the metric properties of the toroidal pitch-space. For using a similar technique with set-classes, thus, it is required to find a pertinent inter-class measure.

A number of geometric set-class spaces have been proposed (for instance, Cohn, 2003; Quinn, 2006, 2007; Tymoczko, 2012). These spaces reflect sophisticated relational properties among sets, but their usage is constrained by cardinality[12], so they do not generalise to distances between arbitrary pairs of sets. Beyond the limitations of geometrical approaches, many set-class similarity measures have been proposed. Among the very few that can manage arbitrary pairs of sets, REL (Lewin, 1979) and IcVSIM (Isaacson, 1990) are popular choices. As it is expected for abstractions of the kind of set-classes, these measures are not without concerns for analytical applications, not to mention with respect to perception (Kuusi, 2001). Our modest purpose here, however, is just to explore the potential of the class-scapes of relative distances to inform about the sonority content of music pieces. Since our method requires a gen-

[11] A minimum of 200 classes, for iv-equivalence.

[12] For instance, Cohn's tetrahedral space measures voice-leading relations between tetrachords. The relational distances in Quinn's and Tymoczko's spaces are also constrained by cardinality.

eralised measure for arbitrary pairwise combinations of classes, and IcVSIM is constrained to iv-equivalence, we decide in favour of the widespread REL.

Unlike IcVSIM, which only measures the standard deviation between the entries of the interval vectors to be compared, REL is often referred to as a *total measure*. That means, REL similarity takes into account the complete subset content of the classes being compared, exhausting the pitch-class set inclusion relations systematically. This measure seems thus convenient to our purpose. The algorithm used in what follows, adapted from (Castrén, 1994, p. 89) to include the trivial forms, is described in Appendix B.

Each point in the class-scape is coloured according to the REL distance between the class it represents and the chosen reference class. Fig. 5.2 shows the class-scape computed for Debussy's *Voiles*, in which the diatonic class 7-35 has been chosen as a reference. This piece does not have a single diatonic segment, so it is clear that the previous all-or-nothing filtering would result in a completely blank image. However, by using REL(7-35,*)[13] and a greyscale[14] to represent the REL similarity (from white = 0, to black = 1), every segment is depicted according to its (REL) closeness to 7-35. It is straightforward to visualise the darkest areas in the class-scape, corresponding to the pentatonic (5-35) passage at bars 42-47, and its superset 7-24B surrounding it. It is also clear that the large whole-tone passages depicted in Fig. 5.1, as well as their building subsets, are poorly (lightly) represented with respect to 7-35. Incidentally, there are some isolated segments (just two dots in the class-scape) belonging to the scalar formation set 6-33B, which are the closest ones to 7-35 in the composition.

This points to an interpretative aspect of multi-scale representations in general, and of class-scapes in particular: the analytical relevance of what is shown is often related with the accumulation of evidence in time and time-scale. A significantly large area of the scape showing the same evidence is most probably representative of a segment of music in which such evidence is of analytical and/or perceptual relevance, in the sense that it accounts for a section in which the inspected properties are stable. Smaller patches, even a single isolated point, capture the class content of their corresponding segments as well. However, they could just be concatenation by-products around the boundaries between more significant segments. Here is where the human visual cognition plays a role, by focusing the viewer's attention on the areas which accumulate similar evidence, an important feature of the multi-scale representations for assisting pattern recognition tasks, in similar terms as we discussed for keyscapes. Of course, localising glimpses of residual evidence could worth a closer inspection. A main difference between keyscapes and class-scapes is the objectivity achieved in the latter, involving no ambiguity or estimation whatsoever.

[13] Where * stands for any class to be compared to the reference class 7-35.

[14] The choice of a greyscale instead a colourful option was motivated by the aesthetics of

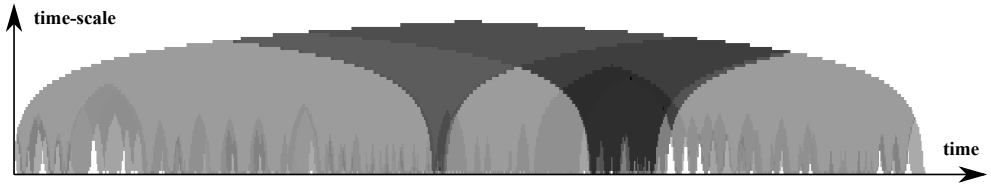

Figure 5.2: Debussy's *Voiles*. Class-scape relative (REL) to 7-35.

5.3.5 Piecewise summarisation: *class-matrix* and *class-vector*

A more compact representation consists of quantifying the presence of each class in the piece. As a data structure, the 2-dimensional class-scape so far can be thought of as a sparse 3-dimensional binary matrix, whose axes represent time, time-scale and class. In the left pane of Fig. 5.3, the raw information computed for Debussy's *Voiles* arranged as such a structure is shown, with the whole-tone class (6-35) used as reference.[15] The reduction process consists of projecting this information into the time vs. class plane. Given the special meaning of the lost dimension (time-scale), this implies growing each point to the actual duration of the segment it represents. The resulting data structure, referred to as *class-matrix*, represents the N classes as N rows, arranged according to Forte's ordinal position. Each row accounts for the activation times of the corresponding class. Activation of a point in the class matrix means that at least one segment of the involved class exists during that time point.

In the right pane of Fig. 5.3, the class-matrix for Debussy's *Voiles* under iv-equivalence is shown, and the prelude's economy of sonorities stands out. Even with the loss of information, it helps to visualise the contribution of each individual class, since colouring is no longer required. The individual duration of each frame from the initial segmentation can be fused, since overlapped segments are projected as their union in the time domain, which has interpretative consequences when looking at individual classes. However, the strict separation of classes allows to capture relational details of certain structural relevance. For instance, two consecutive diatonic passages involving modulation can produce a single long projection in the time axis for the class 7-35. But if both consecutive diatonic passages were in a fifth relation to each other, as in a transition to/from the dominant, the class 8-23 would also be activated at the same time span. Class-matrices are thus suitable for inspecting in detail the inclusion relations down the subclass hierarchy.

An even more compact representation provides a means for quantification. For each row in the class-matrix, its relative active duration is taken and expressed

---

the interactive analysis tool, as an alternative proof-of-concept design.

[15] Each point in the plot represents a segment by the temporal location of its centre (x-axis), the logarithm of its duration (y-axis), and its class content (z-axis).

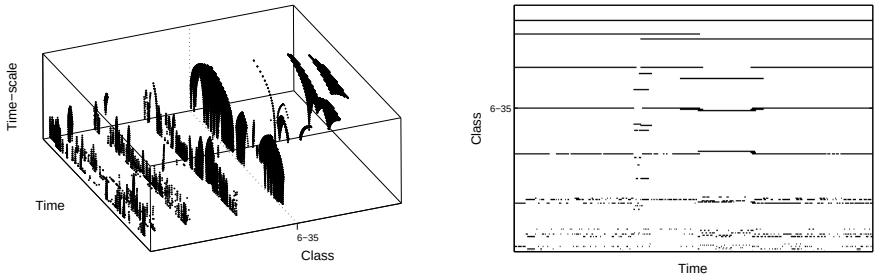

Figure 5.3: Debussy's *Voiles*. Left: raw multi-scale information as a sparse 3-D matrix. Right: class-matrix.

as a percentage with respect to the total duration of the music. This data structure, hereafter *class-vector*, has a dimensionality equal to the number of classes, and quantifies the complete class content in a piece. Its potential application for comparing different pieces of music, however, raises the problem of resolution. The descriptions so far depend on the minimum temporal analysis window. This method limits the number of resolutions for fast interaction, and provides a regular grid for visualisation. Resolution can be tuned as a precision vs. computational cost trade-off. However, a given time-scale may not resolve equally the pitch-class set content of different pieces. Working with MIDI files, however, class-vectors can be computed with absolute precision, by substituting the multi-scale policy by a genuine exhaustive approach. This is done by a vertical segmentation at every change in pitch-class set content, whether product of onsets or offsets, and segmentation is performed for all pairwise combinations of these boundaries. Class-scapes and class-matrices can be computed and exploited by the exhaustive method, but in general they lack a grid regularity, and visualisations do not come without interpretative artefacts. On the other hand, class-vectors are not affected by the grid problem, as they just accumulate durations, so they are free of representational problems. The exhaustive computation of class-vectors has thus been performed in all the applications that follow.

5.4 Subclass analysis of diatonicism in corpora

As class-matrices align the temporal activation of all the classes, they can be mined to describe the subclass content *under* any class, to reveal the building blocks of particular class instantiations. The next case study illustrates the use of the method for comparing two corpora in terms of pure diatonicism, understood as the subset class content under 7-35.

The computation process for the Agnus Dei from Victoria's *Ascendens Christus* mass is depicted in Fig. 5.4. First, the class-matrix is computed. To account for

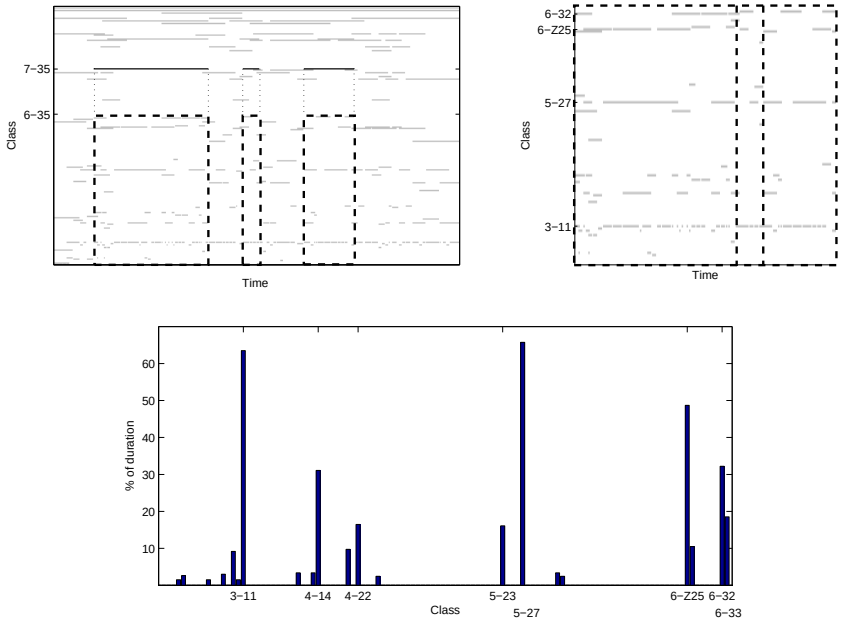

Figure 5.4: Victoria's *Ascendens Christus* mass, Agnus Dei. Computation process of subclass-vector. Top left: filtering under 7-35. Top right: subclass-matrix. Bottom: subclass-vector.

the diatonic-related subclass content, only the time frames below the activated 7-35 frames will be considered, as shown in the top-left pane of Fig. 5.4. The result of this process is a *subclass-matrix*, in the top-right pane of Fig. 5.4, with a number of rows equal to the number of subset classes (up to cardinality 6 here). The number of columns equals the number of activated 7-35 frames in the class-matrix. Partially overlapped classes which are not a subset of 7-35 are then removed from the subclass-matrix. Following the same method as for class-vectors, a *subclass-vector* is then computed from the subclass-matrix. The subclass-vector, in the bottom pane of Fig. 5.4, quantifies the total subclass content contributing to the reference class, that is, it describes what (and how much of it) the particular diatonicism is made of. The most prominent subclasses at 3-11, 5-27, 6-Z25 and 6-32 stand out in the subclass-vector, which also reveals other common scalar formation classes, such as 5-23 and 6-33. The most prominent subset content of 5-27 is featured by 4-14 and 4-22.

In order to characterise a corpus, a dimension-wise average is computed across the subclass-vectors extracted from all the pieces, that is, each class is averaged across all movements, and the standard deviation for each class is also taken. Fig. 5.5 shows the subclass-vectors under 7-35 computed for two contrasting corpora: a) the Victoria's parody masses in Ionian mode[16]; b) the preludes

[16]Including *Alma Redemptoris Mater*, *Ave Regina Caelorum*, *Laetatus Sum*, *Pro Victoria*,

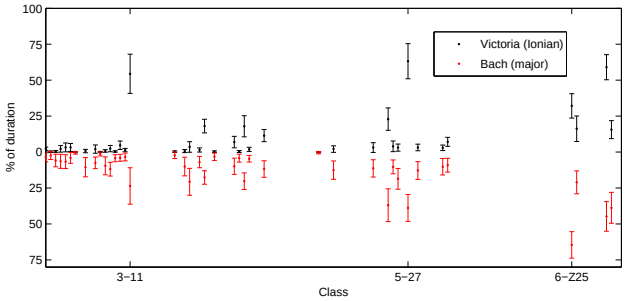

Figure 5.5: Diatonicism in Victoria and Bach: subclass-vectors under 7-35.

and fugues in major mode from Bach's *Das wohltemperierte Klavier*. The selection of the corpora is based on the close relations between the major and the Ionian modes, and the homogeneity criteria with respect to the usage of contrapunctal resources in both composers. For clarity of comparison including standard deviations, Bach's subclass-vector is aligned as if it were negative. This representation reveals the subclass content down the hierarchy at a glance. It is clear a prominent use of major and minor triads (3-11) in Victoria relative to Bach. Similarly predominant is the class 5-27, a far more recurrent cadential resource in Victoria[17]. On the other hand, the Locrian hexachord 6-Z25 is far more present in Bach. Apart from its instantiations as perfect cadences in both corpora[18], 6-Z25 appears consistently in many motivic progressions in Bach[19].

5.5 Multi-scale set-class analysis and serialism

The final use case illustrates the use of the method for characterising structural similarity for a compositional style in which class-equivalence is of analytical relevance. Set-class representation is a standard lexicon for post-tonal analysis, while tone rows and their constituent hexachords are often the starting point in the analysis of serial music.

The top pane of Fig. 5.6 shows the opening of the first movement of Webern's *Variations for piano* op.27, corresponding to the first statement of the series in a virtually palindromic form. In the bottom-left pane of Fig. 5.6, the prime tone row (P0) and its retrogression (R0) are shown. The bottom-right pane of Fig. 5.6 depicts the inverted versions (I0 and RI0) of the row. Throughout the movement, the exposition of the series in one stave (Pn) runs in parallel

Quam Pulchri Sunt, and *Trahe Me Post Te* (Rive, 1969, for a modal classification).

[17] The class 5-27 results from the combination of the dominant and tonic major triads.

[18] The class 6-Z25 results from the combination of a major triad and its dominant 7^{th} chord.

[19] By interfacing class-vectors with class-scapes, the actual content of particular class instantiations can be easily explored, as discussed in Chapter 6.

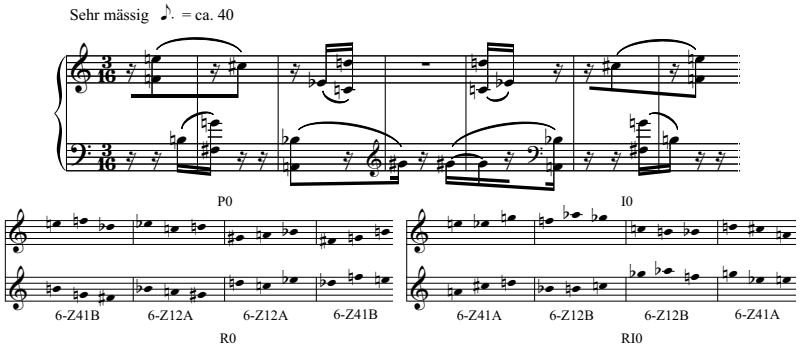

Figure 5.6: Webern's *Op.27/I.* Top: bars 1-7. Bottom-left: hexachordal segmentations for P0/R0. Bottom-right: hexachordal segmentations for I0/RI0.

with its retrogression (Rn) in the other stave at the same transpositional level. This simultaneity is also kept for the inverted statements of the series, as In running in parallel to RIn. What catches the attention here, however, is the trichordal segmentation of the series, and how it harmonises in hexachords with the retrograde. The first 3 notes of the row and the last 3 notes (first 3 notes of the retrograde) can be clearly cut throughout the score. As both series evolve in time, a second hexachord is formed by the six central notes of the row. Then, as the rows progress to their completion, the same two hexachords are palindromically stated.

An unusual presence (92% of the total duration) of the interval vector ⟨332232⟩ is openly manifested at this segmentation level, as it is depicted in Fig.5.7 (top). By construction, the hexachord formed by the extreme trichords of the series belongs to the set-class 6-Z41B, while the central hexachord belongs to the set-class 6-Z12A. Both hexachords are Z-related to each other, sharing the same interval vector although not being TnI-equivalent. The same applies to the inverted series, which are also stated in parallel to their retrogrades. Thus, the extreme trichords of the inverted row form a 6-Z41A hexachord, while the central hexachord belongs to 6-Z12B. Since the whole movement evolves from one TnI to another in quite a neat way, sustaining the direct and retrograde parallel delivery, the pattern is repeated throughout the piece. This holds strictly in the four expositions of the series in section A, where the sequence of hexachords 6-Z41, 6-Z12, 6-Z12 and 6-Z41 is presented 4 times. Section A' virtually reconstructs the same pattern. The well-known structure of the piece is shown in Fig. 5.7 (bottom), segmented according to (Cook, 1987), and sized relatively to the timings computed from the used MIDI version.

A self-similarity matrix is a simple technique for finding recurrences in time series in general, and for pattern discovery in music in particular (Foote, 1999). This data structure represents the distance of all the pairwise points in the time

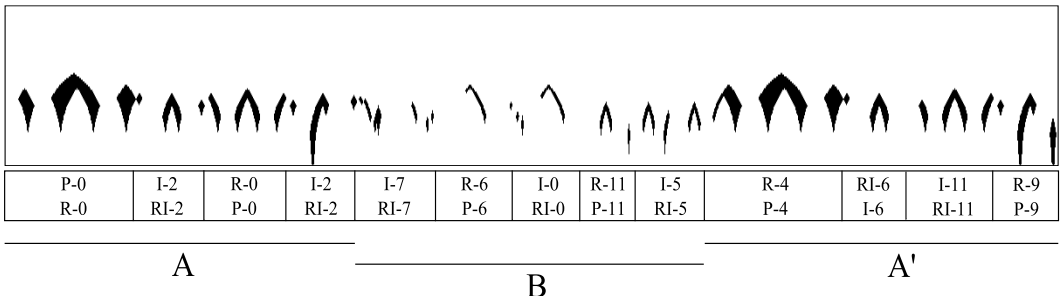

Figure 5.7: Webern's *Op.27/I*. Top: class-scape filtered by ⟨332232⟩. Bottom: structure and row statements.

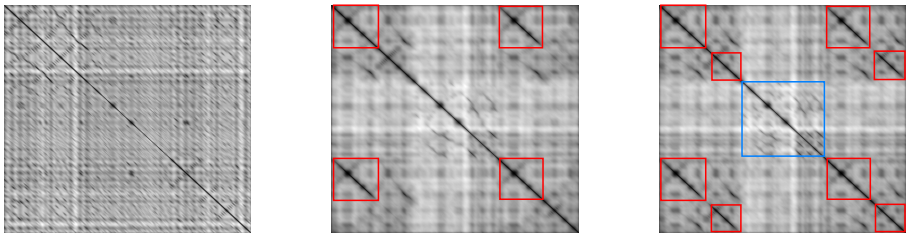

Figure 5.8: Webern's *Op.27/I*. Self-similarity matrix from: Left: pc-set time series. Centre: class-matrix (Tn-equivalence). Right: class-matrix (TnI equivalence).

series according to some similarity measure. The most likely recurrences are revealed as diagonals of high activation in the matrix. Self-similarity matrices were computed for three different descriptors over time: a) the pitch-class set time series[20]; b) the class-matrix under Tn-equivalence; and c) the class-matrix under TnI-equivalence. The computation for the pitch-class set time series in Fig. 5.8 (left) does not show any remarkable recurrence of structural relevance. By allowing transpositional equivalence, the first two row statements at section A appear restated at the beginning of A', as shown in Fig. 5.8 (centre). This is possible because the re-exposed statements are related to their counterparts in the exposition only by transposition: P-0/R-0 and I-2/RI-2 in section A are restated as R-4/P-4 and RI-6/I-6 in A'. However, it fails to match segments as large as the last two statements of the series in the sections A and A', since the direct presentation (R-0/P-0) is re-exposed in inverted form (I-11/RI-11), and vice versa (I-2/RI-2 as R-9/P-9). By using transpositional and inversional equivalence, in Fig. 5.8 (right), all the row statements at A are fully revealed at A'.

The method is thus sensible to capture recurrences of passages under different

[20]This constitutes the symbolic equivalent of the chroma features, extensively used for similar tasks in audio domain (Müller, 2007).

class equivalences. By using the class-scapes and the basic class filtering, as depicted in Fig.5.7 (top), it is possible to grasp potential recurrences as well, but in a more limited way. If the activated class presents a similar shape in the class-scape at different places, this may indicate that the class appears at similar timings, but it does not say anything about the actual content of these instantiations. The guarantee of a true restatement, always under the chosen class equivalence, is possible by considering the class description over time down the full hierarchy. A restatement implies that the class instantiations have to be built from the same subclass content, which has to follow the same temporal sequence. That is why short recurrences appear locally in the self-similarity matrix, but without spanning long distances. The general sonority is ubiquitous, but different subclass associations are distinctly sequenced at those shorter scales.

Given the prominent occurrence of a particular iv-sonority, one could question its pertinence from analytical or perceptual standpoints. As Hasty suggested, "it is the perception of musical articulations which might result from the analyses that offer a test of validity of analytical statements." (Hasty, 1981, p. 55, footnote 2). One could hardly presume that set-class referentiality is justified by arguments of compositional relevance just because a systematic recurrence of a surface phenomenon. However, this particular perceptual qualia is closely related to: a) the structure of the series; b) the verticalisation of non adjacent elements of the row by its simultaneous direct and retrograde statements at the same transpositional level; c) the score punctuation for quite a number of such segments; d) some relational aspects among the hexachordal instantiations, when observed under different class-equivalences. The Z-relations among the sonority statements provide two contrasting pitch-class set material which, while similar to each other through their common interval content, are distinguishable beyond the general transformations by their non TnI-equivalence. The latter aspect provides a means for clarifying the palindromic discourse of each row statement, while the former reinforces the overall coherence. Whether intentional or not, the Z-relations in a set of cardinality 6 always connect with the set's complement, pushing the all-embracing symmetry of the piece to the very stage of the row design.

5.6 Conclusions of the method

The technique described here proposes a reconsideration of the basis for set structure analysis. Pople's three principles in terms of segmentation, reduction and significance judgement (Pople, 1983, p. 151) are reframed by computational systematicity and interfacing methods. In comparison with the approach in (Huovinen & Tenkanen, 2007), this method introduces important changes with respect to systematicity and representational capabilities, with the aim of reducing the impact of interpretation artefacts.

First, with respect to segmentation, it avoids any a priori interpretative selection of the segments of interest, whether explicit or imposed by external constraints. It simply provides a means for revealing them by accumulation of evidence. The method does not impose cardinality, in favour of a comprehensive account for all the different segments of music, which in turns facilitates simultaneous multi-cardinality explorations. Systematic segmentation is proposed in both exact and approximative versions, the former providing appropriate quantification of further reductive descriptions, and the latter facilitating visual inspection while maintaining reasonable comprehensiveness. A relevant feature here is the introduction of an explicit and unambiguous temporal index, as a substitution of the commonly used note-based indexing. This level of description does not introduce artefacts of any kind, beyond the limitation imposed by the vertical segmentation policy, since all the possible segments are captured, characterised and represented uniquely.

Second, it avoids a priori reductive classifications. Systematicity accounts for every class, fed by the actual content of every segment. This provides a neutral level of description and avoids quantification at this stage, which in the author's opinion constitutes a premature information summarisation. The means for focusing material of analytical relevance is relegated to the visual representation domain, whereby the accumulation of similar evidence highlights potential areas of interest. Since locality in this context is understood in terms of time and time-scale, the cognitive system is naturally attracted towards those areas which are stable, without eliminating or obscuring minor (less represented) areas of potential interest, which can be of residual nature or not. Since no average is performed by computational means, it is up to the analyst's criteria to focus on different salient features, facilitating the interaction with the data. An additional difference from other systematic alternatives, is that class content is not mixed up, each segment keeping its own individuality, a feature that can be appreciated by the naked eye.

Third, it provides a different means for assessing the significance of sets, through further information reductions that can account for global properties of single pieces or collections. This includes class-matrices and class-vectors, as well as their subset versions (subclass-matrices and subclass-vectors). In all those reductions, again, the method guarantees a strict separation of classes, avoiding their removal or fusion up to the very moment of the interpretation. Quantification here captures the relative temporal presence of each individual class along the music material, whether globally or with respect to a given reference class, which allows the qualification of details at the level of class inclusion. A particularly useful consequence of the duration-based quantification method is that it facilitates the localisation of potential classes of analytical interest by means of prominent occurrences, as in Webern's example. The method provides values of prominence, but it is compatible with the exploration of any

class, no matter how little represented. Moreover, the interactive version of the method encourages this kind of inspection, since all the information, from the global to the local, is readily available and accessible for testing in real time. The analysis loop is thus assisted by interactive filtering of information which is guaranteed to be systematic, accurately described, significantly quantified, and last but not least, intuitively visualised.

5.7 On content-based metadata

The method so far has been described for the analysis of music given in symbolic (MIDI) representation. This constraint is based upon the premise of guaranteeing objectivity (not involving estimations of any kind) and systematicity (every different segment has to be captured) in the description. Class-vectors and class-matrices constitute a rich and specific content-based metadata about the set-class sonority of the compositions, provided that the computation is performed from good quality encodings of the music scores. This content-based metadata, far more precise and sophisticated than the current standards in music information retrieval[21], can thus be fully representative of the same music pieces in the audio domain, provided a reasonable fidelity of the audio versions with respect to the scores and the existence of an explicit link between them. This motivates the usage of our descriptive framework in applications involving audio datasets, by exploiting content-based metadata which has not been computed from the actual audio, but from the more reliable symbolic domain. Class-vectors can be used for querying large datasets in terms of arbitrary set-class sonorities or their combinations. In this section, both applications are tentatively explored.

5.7.1 On *authoritative* score encodings

That the encoding of a given music work can be considered an *authoritative* source or not, constitutes a big issue in itself. There are a number of initiatives for building curated editions[22] in computer-readable formats, such as the MuseData Project[23]. The impact of the different encodings cannot be elucidated without considering both the repertoire and the application of the analysis. For instance, the description of ancient music, and therefore the analytical results, is sensitive to the editorial choices with respect to the *musica ficta*. For the analysis of music works in terms of set-class sonorities, however,

[21] The ubiquitous content-based tonal descriptors to date, related to our concerns of contextual sonority, are estimations of the global key of the piece, barely informative in general and misleading in many cases.

[22] Not comparable to critical editions in rigour, but considering some editorial issues for musicological research.

[23] http://musedata.stanford.edu

sources with certain degrees of unreliability can be exploited in some applications. Among the MIDI encodings available in Internet, it is common to find piano reductions, others present missing voices, some have playing mistakes, some simplify difficult parts, many live recordings abound in improvised ornaments, and so on. While quantification in class-vectors is for sure affected by these variants, in many cases they constitute quite good approximations to the vectors computed from authoritative sources.

5.7.2 MIDI dataset for testing purposes

The dataset used in what follows is comprised of a selection of MIDI tracks from the common-practice period repertoire, plus some representation of later compositions. The list of composers and the number of movements is: Albéniz (61), Albinoni (49), Alkan (237), J. S. Bach (691), Beethoven (92), Brahms (146), Bruckner (30), Busoni (38), Buxtehude (94), Byrd (109), Chopin (164), Clementi (41), Corelli (55), Couperin (115), Debussy (159), Josquin (35), Dowland (61), Frescobaldi (60), Gesualdo (37), Guerrero (83), Haydn (212), Lasso (72), Liszt (131), Lully (108), Mahler (33), Morales (88), Mozart (280), Pachelbel (59), Palestrina (70), Satie (21), Saint-Saëns (86), Scarlatti (58), Shostakovich (48), Schütz (92), Schumann (97), Scriabin (86), Soler (65), Stravinsky (35), Tchaikovsky (238), Telemann (60), Victoria (333), Vivaldi (40). The dataset also includes anonymous medieval pieces (47), church hymns (362), as well as the Essen folksong collection (8402). Each movement in the dataset was computed for extracting its class-vectors. In what follows, only the iv-equivalence is considered, resulting in a dataset of 13480 vectors of 200 dimensions.

5.7.3 Querying by set-class

Let's reconsider the information conveyed by the class-vectors. For each class, the corresponding value in the vector accounts for the percentage of frames in the piece, which are interpretable in terms of the specific class. That is, for each frame for which the class is active, it exists at least one segment containing the frame which is uniquely defined by some instantiation of the class. We discussed that in terms of the *relative duration* of the class within the piece. They can be interpreted in probabilistic terms as well. Observing a random time frame in the piece, the value of each dimension in the class-vector is the probability of finding a segment with the corresponding set-class around this frame. It is relevant to mention that this probability is not to be interpreted as an *estimation*, since it is guaranteed that the piece actually has this percentage of the sonority. With this in mind, it is possible to perform different queries to the dataset. Localising specific sonorities in a large dataset can be combined with the extraction of the actual segments. Since the systematisation of the description accounts for every possible sonority, it can be exploited in varied

applications, such as in music education. Once the pieces have been localised, a detailed exploration of the sonority in its context can be done by interacting with the class-scape and listening to the segments, as it will be discussed in Chapter 6.

5.7.3.1 Filtering by set-class

A useful and simple task is to sort the feature dataset according to a given set-class sonority. It can be used, for instance, to localise pieces with some *exotic* scalar flavour. Table 5.1 shows the 10 pieces with the largest presence (relative duration) of the sonority 7-22, usually referred to as the Hungarian minor scale[24]. The example also shows how the systematic segmentation and description method is agnostic with respect to the monophonic or polyphonic writing, as it is evidenced by the matching of two folk tunes from the Essen collection. The unique requisite for capturing a given sonority it its existence as a vertical segment, as it would be clipped from an audio source.

retrieved piece	**7-22 (%)**
Busoni - 6 etudes op.16 n.4	61.98
Scriabin - Prelude op.33 n.3	60.29
Essen - 6478	58.02
Liszt - Nuages gris	41.88
Essen - 531	37.18
Satie - Gnossienne n.1	27.11
Scriabin - Prelude op.51 n.2	26.92
Alkan - Esquisses op.63 n.19	26.08
Lully - Persee act-iv-scene-iv-28	25.92
Scriabin - Mazurka op.3 n.9	25.66

Table 5.1: Sorting by 7-22

5.7.3.2 Filtering by combined set-classes

Combined filtering can be used for localising more specific sonorities. The systematicity of the method together with the multi-scale approach, allow finding particular sonorities in specific contexts. As done in the study of diatonicism in Victoria and Bach, this can be achieved by a proper mining of the class-matrices and subclass-matrices. The class-vectors summarise the information in a way in which it is not possible to elucidate the sub-class content under a given class. However, if the queried sonorities have a substantial presence (or a notable absence) in the piece, the class-vectors alone can often account for the combined filter. Table 5.2 shows the 10 pieces with the largest presence of

[24]Sometimes also called Persian, major gypsy, or double harmonic scale, among other denominations. Its prime form is {0,1,2,5,6,8,9}.

the *suspended* trichord (3-9)[25], constrained to cases of mostly diatonic contexts (7-35). This situation, as reflected in the results, is most likely to be found in medieval melodies or early counterpoint.

retrieved piece	**3-9 (%)**	**7-35 (%)**
Anonymous - Angelus ad virginem 1	54.05	83.78
Anonymous - Ductia	44.55	99.00
Anonymous - Instrumental dances 7	41.94	99.75
Lully - Phaeton acte-i-scene-v	40.19	82.15
Anonymous - Instrumental dances 9	38.07	99.54
Lully - Persee prologue-3b	37.97	98.73
Anonymous - Cantigas de Santa Maria 2	36.67	99.65
Lully - Persee prologue-3c	35.29	98.52
Anonymous - Danse royale	32.86	99.30
Frescobaldi - Canzoni da sonare-11	31.59	81.76

Table 5.2: Sorting by 3-9 with high 7-35

5.7.4 On dimensionality of description

In feature design, the ratio between the size of the feature space and the informativeness of description is a relevant factor. The class content of a piece, as described by class-vectors, have 200, 223 or 351 dimensions, depending on the class equivalence chosen (iv, TnI and Tn respectively). Compared with other feature spaces, the dimensionality may seem quite large. However, the benefits of class vectors are the systematicity, specificity and precision of the description. Several relevant differences with respect ot other features are to be noticed. A single class-vector, computed by the fully systematic segmentation approach, accounts for:

1. Every different segment in the piece, regardless of the required time-scales for capturing them. No sampling artefacts of any kind are introduced.

2. Every possible sonority among the set-class space. No pc-set is left out of the description.

3. An objective description of the sonority. No probability or estimation is involved.

4. A description in (high level) music theoretical terms, readable and interpretable by humans.

5. An objective quantification of every possible sonority in terms of relative duration in the piece. No estimation or probability is involved.

[25] A major trichord with the third degree substituted by the fourth. The term *suspended* refers to its common usage in early counterpoint. Its prime form is {0,2,7}.

6. A content-based description of the piece in its own terms only. Neither statistics nor properties *learned* from datasets are involved. The piece itself suffices for its accurate description.

7. In cases of large presence of some sonorities, an approximation to the hierarchical subclass content under these sonorities[26].

With this in mind, it seems to us that a piecewise description in 200 dimensions is a reasonable trade-off between size and informativeness. Considering the somewhat sophisticated tonal information conveyed by class-vectors, they would constitute a useful feature for complementing existing sources of content-based metadata.

5.8 Conclusions of the chapter

In this chapter, our temporal multi-scale method has been extended to a pitch-class set analysis domain. The main features of the method are:

1. Two systematic segmentation approaches: a) a fully systematic method, which considers every possible different segment in the music; b) a multi-scale sliding-window method (the same as for keyscapes), just an approximation to the full systematisation, but more practical for interactive inspection.

2. An unambiguous temporal indexing of the music segments, solving the problems of other systematic approaches to pitch-class set analysis.

3. An unambiguous and objective description of each segment, in terms of set-classes.

4. Three set-class spaces, following three standard class equivalences (iv, TnI and Tn).

5. A piece-wise summarisation method (class-matrix), which captures all the hierarchical information about set-class inclusion relations over time.

6. A piece-wise summarisation method (class-vector), which quantifies the presence of each class in the piece.

7. A method for characterising the sub-class content over time under a chosen class (sub-class matrix).

8. A method for quantifying the sub-class content under a chosen class (subclass-vector).

[26] This is only guaranteed by the class-matrix description, which has much larger size.

9. An indexing method for linking the different spaces in interaction.
10. Its limitation to the symbolic domain.

The method has been illustrated or proposed for:

1. The use of class-scapes for simultaneous visualisation of all the instantiations of a chosen class.
2. The use of class-scapes for simultaneous exploration of every segment of the composition, relative to a chosen class, using an inter-class measure (REL).
3. The use of class-matrices and subclass-matrices for exploration of set-class inclusion relations over time.
4. A hierarchical subclass comparative analysis of diatonicism in different corpora.
5. A similarity and structural analysis of serial music, under different class equivalences.
6. The generation of content-based metadata from symbolic datasets.
7. The querying of datasets for single and combined class sonorities.

The analytical coverage has surrounded a variety of theoretical, compositional and aesthetic aspects, including: symmetric modes, theoretical distance between set-classes, closeness to diatonicism in non-diatonic contexts, atonalism seen from different class equivalences, *exotic* scales and ancient music. The extension of our method is thus proposed as a complementary tool for tonal analysis, extending its usage to application contexts of higher degrees of sophistication.

CHAPTER 6

Interfacing tonality

6.1 Introduction

So far, the features of our multi-scale method have been discussed by examples of analysis. Along the way, we mentioned the potential of interfacing between the different representational domains. The method itself was conceived for approaching some of the representational challenges which arise in the research of the tonal phenomena, but also keeping in mind its potential usage for musicological or educational purposes. Is in this last scenario for which part of the method's *processing* is relegated to the human cognitive abilities. The method benefits from a systematic feature extraction from the music stimuli, a time consuming and prone to errors task for humans, while the analyst takes advantage of the human visual and auditive capabilities, far more sophisticated than computers for dealing with ambiguous and subjective evidence. In turn, the analyst tunes the model's parameters to focus the exploration accordingly. This is what we referred to as the *analysis loop*.

In this Chapter, we will discuss the interfacing potential of the method. All the methods in this work have been prototyped as Graphical User Interfaces (GUI) for Matlab, relying upon part of both the MIDI Toolbox (Eerola & Toiviainen, 2004) and the MIR Toolbox (Lartillot & Toiviainen, 2007). Three of such interfaces will be discussed in what follows, namely:

1. The basic keyscape to pitch-space explorer, as the general framework used in Chapter 3.

2. The explorer of keyscapes and empirical ratings, as the tool used for the first case study in Chapter 4.

3. The set-class explorer, as the main tool used in Chapter 5.

6.2 The basic tonal explorer

In Fig. 6.1, the interface of the basic tonal explorer is shown. A MIDI encoding of Bartok's *Mikrokosmos, n.97* has been loaded and computed for estimating its tonal content. The segmentation parameters, at the left side of the interface, show a minimum resolution of 0.5 seconds (*minres*) and 30 different time-scales (n^o *scales*).

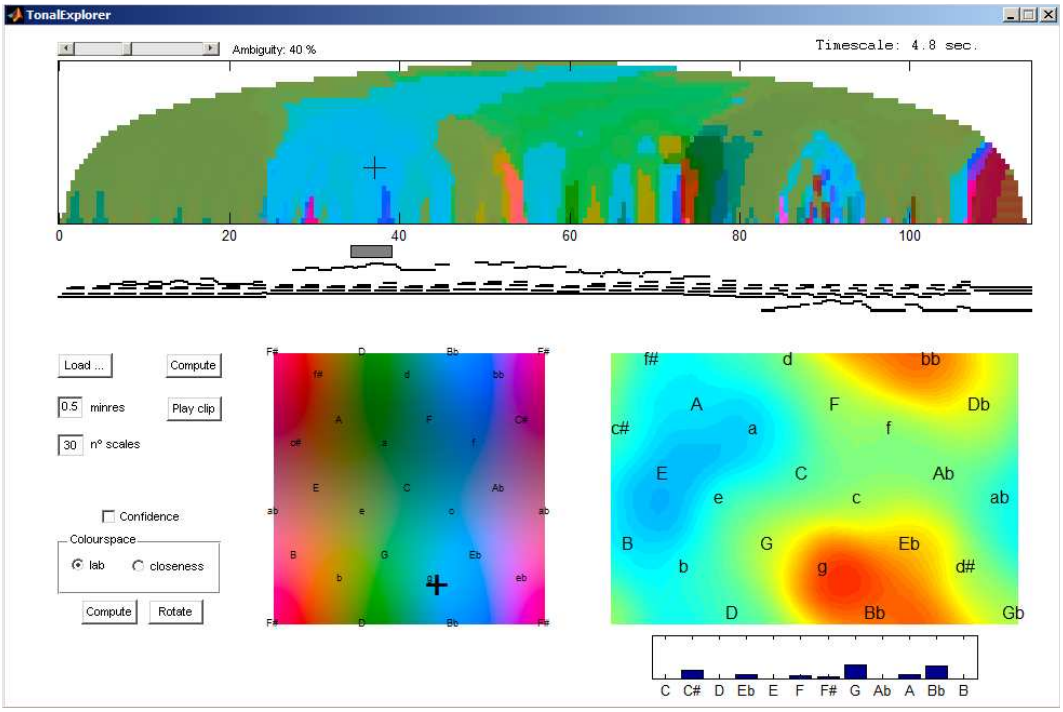

Figure 6.1: Basic tonal explorer.

Two plots account for the bird's-eye view of the piece's pitch content, and three plots complement the information in different representation domains. From top to bottom and from left to right:

1. The keyscape, summarises the tonal estimations for all the segments in the piece as a time vs. time-scale plot. The colour of each pixel represents the centroid's position after its projection in the coloured pitch-space, according to the colouring method described in Chapter 3. Individual segments in the keyscape can be inspected by a movable cursor (black + sign). By dragging it, all the segment-wise information is updated in real time:

   a) The pitch-class profile of the new segment is depicted in its corresponding plot (below the SOM plot).

b) The segment's pitch-class profile is projected as the complete activation of the SOM.

c) The segment's pitch-class profile is summarised as a tonal centroid in the coloured pitch-space.

d) The cursor below the class-scape is relocated and/or resized according to the time and time-scale position of the cursor, aligning the inspected segment with the piano roll. At the top-right corner above the keyscape, the selected time-scale (duration of the segment in seconds) is also shown.

2. Piano roll representation of the piece, which serves as a visual index to the composition. Between the class-scape and the piano roll, a horizontal cursor shows the specific segment selected in the keyscape.

3. The coloured pitch-space, which serves as a colour legend of the keyscape in relation to the key categories. The projection of the selected segment in the pitch-space is depicted with a black '+' sign. The projection of the centroid is computed by multidimensional unfolding, considering the different strengths of the key candidates and the ambiguity unfolding parameter (see additional controls below).

4. The pitch-space as a self-organised map (SOM) representation. The SOM is activated by comparing each vector in its codebook with the pitch-class profile of the selected segment. The the segment's content in terms of the key categories is best represented in this space. It can account for any degree of ambiguity in the estimation in a human-readable way.

5. The pitch-class profile of the selected segment, as the raw estimation of the tonal content of the segment in 12 dimensions.

Six additional controls complement the interfacing possibilities:

1. Ambiguity unfolding parameter, at the top-left corner above the keyscape. It introduces the ambiguity parameter of the centroid unfolding method, mentioned in Chapter 3. It ranges 0 % (no ambiguity, the centroid is projected in the pitch-space right at strongest candidate) to 100 % (maximal ambiguity, all the candidates contribute to the unfolding, proportionally to their strengths). In this example, a 40 % of ambiguity allows a moderate contribution of the next strongest key candidates. All the centroids in the keyscape are affected by this control in real time. The degree of fuzziness in different areas of the keyscape is thus indicative of the confidence of the estimation. By acting on the slider, the less confident areas in the keyscape manifest more noticeable colour changes.

2. Colourspace configuration, at the bottom-left side of the interface. Three different colouring methods are available for representing three different aspects of the keyscapes, related with multidimensionality and ambiguity.

   LAB colourspace. As described in Chapter 3, it approximates perceptual uniformity across the double circularity of the toroidal pitch-space. As geometrically related, the colourspace can be rotated with respect to the pitch-space, in order to match any specific colour to any key. This is done by acting on the *Rotate* button, and selecting a point in the in the coloured pitch-space. The selected position will take the colour associated (by default) to the plot's centre (C). This can be used to maintain the same colour criteria with respect to the tonic when analysing different pieces.

   Closeness. It uses a linear colourspace for representing the distances in pitch-space with respect to the point chosen as a reference (see previous note on the *Rotate* button). Distance is coded from dark blue (minimum) to dark red (maximum). It is useful for fast localisation of the closest or farthest segments of the composition in the keyscape, in relation to any key.

   Confidence. The keyscape is not coloured according to the centroid's location in pitch-space, but according to the maximal correlation of the estimates. That is, the keyscape turns into a confidence-scape, showing the areas for which the key estimations are strong or weak. The colourspace is the same linear space as for colouring by closeness.

3. The play button, which reproduces only the selected segment at the keyscape, by triggering the corresponding MIDI or audio segment. This constitutes a useful complement for the exploration or analysis, by means of sonification.

6.3 Evaluating tonal perception models

In Fig. 6.2, the interface for comparing models of tonal induction with empirical ratings is shown. The audio version of Bach's organ duet *BWV 805* has been loaded and computed for estimating its tonal content. The segmentation parameters, at the left side of the interface, show a minimum resolution of 0.8 seconds and 30 different time-scales.

One plot accounts for the bird's-eye view of the piece's tonal content. The rest of the plots complement the information in different representational domains. From top to bottom and from left to right:

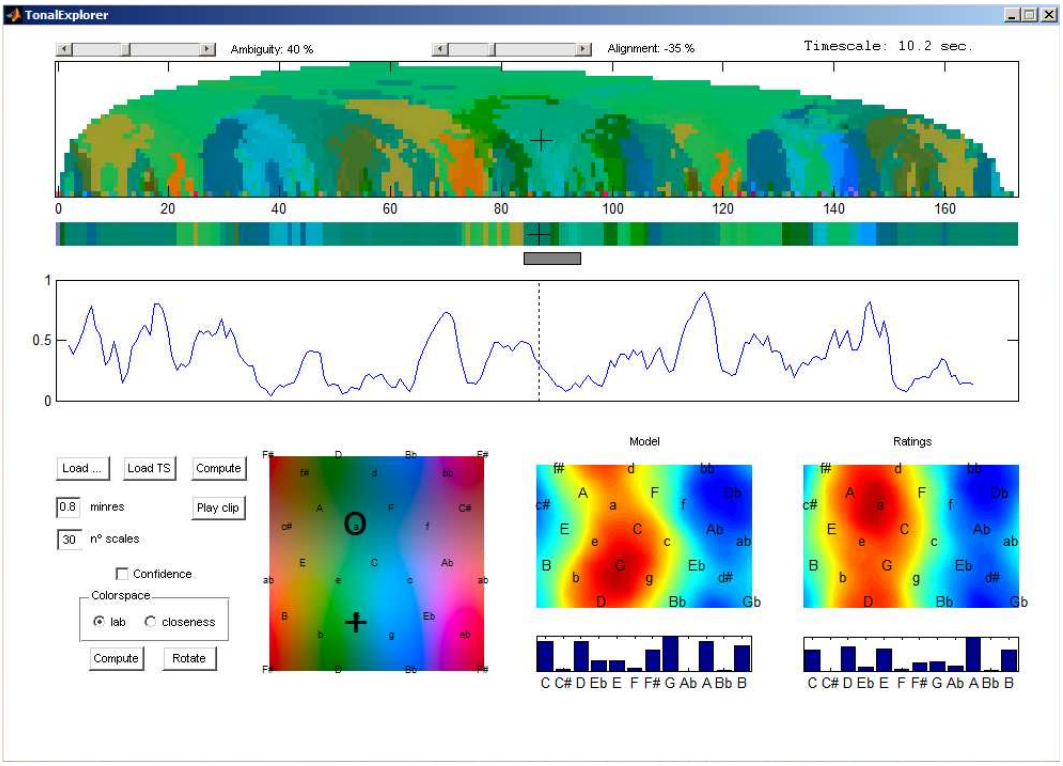

Figure 6.2: Model vs. ratings tonal explorer.

1. The keyscape, summarising the tonal estimates for all the segments in the piece, in the same terms as in the basic explorer. A cursor allows the selection of individual segments in the keyscape, which also selects the corresponding time frame in the ratings time series. Both frames from the keyscape and the ratings are then aligned in time, allowing a frame-based evaluation of the model of tonal induction. By dragging the cursor in the keyscape, all the segment-wise information is updated in real time:

   a) The pitch-class profile of the new segment is depicted in its corresponding plot (below the model's SOM plot).

   b) The model's SOM is activated accordingly to the corresponding pitch-class profile.

   c) The pitch-class profile of the corresponding ratings frame is depicted in its plot (below the rating's SOM plot).

   d) The rating's SOM is activated accordingly to the corresponding pitch-class profile.

   e) The cursor below the keyscape and the ratings plots is relocated and/or resized according to the time and time-scale position of the cursor

and the selected cross-scale alignment (see additional controls below). It represents the selected segment being considered by the model.

2. The empirical ratings. This plot represents the projection of the ratings time-series in the coloured pitch-space, in the same terms as done for the keyscape. Any 12-dimensional time-series can be loaded (*Load TS* button), in order to be compared with the model. If a ratings signal is loaded, the minimum resolution of the keyscape is set to the sampling period of the signal, to guarantee a frame-based alignment between them.

3. The distance curve between the model's estimations at the selected time-scale and the ratings time series. The comparison is done just for the available frames in the keyscape at the chosen time-scale. As discussed in Chapter 4, the distance is computed as one minus the correlation between each pairs of frames (from the keyscape and the ratings). Correlation considers the 12-dimensional pitch-class profiles, to avoid after-mapping mathematical artefacts. A dotted vertical line localises the time frame of the segments being selected for further inspection in the rest of the spaces.

4. The coloured pitch-space, which serves as the colour legend of the keyscape in relation to the pitch-space. A black '+' sign localises the projection of the selected segment in the keyscape. A black 'O' sign localises the projection of the corresponding ratings frame. The distance between them is related to the actual distance between both pitch-class profiles, but considering the distortion introduced by the multidimensional scaling.

5. The SOM activation and pitch-class profile of the selected segment in the kescape.

6. The SOM activation and pitch-class profile of the corresponding frame in the ratings.

Four additional controls complement the interfacing possibilities:

1. Ambiguity unfolding parameter. The same as in the basic explorer, but affecting to all the centroids from both the keyscape and the ratings. Both plots are represented in the same fuzzy terms, helping to inspect simultaneously the ambiguity of the representation for both the model and the ratings.

2. Cross-scale alignment slider. Ranging from -100 % (full future-wise) to 100 % (full past-wise). In this example, the setting at -35 % shows a moderate *expectational* model, whereby the model is biased towards the future, but considering a relevant amount of past as well. By acting on this control, all the segment-wise information is updated in real time:

a) The keyscape is skewed accordingly, in this case towards the left.

b) The cursor below the keyscape and the ratings, indicating the temporal boundaries of the selected segment in the keyscape.

c) The comparison curve between the model and the ratings is recomputed to the new situation, and the result is shown in the comparison plot.

3. Colourspace configuration. The same options (LAB, closeness and confidence) as in the basic explorer. It affects to both the keyscape and the ratings, so as they can be compared in terms of confidence or relative to a given position in pitch-space.

4. The play clip button, which triggers the MIDI or audio segment, corresponding to the selection in the keyscape.

6.4 The set-class explorer

In Fig. 6.3, the set-class explorer interface is shown. A MIDI encoding of Debussy's *Voiles* has been loaded and computed for extracting the set-class content. The segmentation parameters, at the left side of the interface, show a minimum resolution of 0.3 seconds and 40 different time-scales.
Four large images, covering most of the interface, account for the bird's-eye view of the piece's pitch content. From top to bottom:

1. The class-scape, representing the class content of all the segments in the piece. Since the REL distance is activated (see additional controls below), the colour of each pixel is relative (REL) to the selected class, which in this case is the hexatonic sonority 6-35. Three large black areas in the class-scape show the predominant hexatonic sonority of the piece. The reader might want to compare this image with the examples in Chapter 3, which show the all-or-nothing filter (without activating the REL filter) and the scape of relative distances with respect to the diatonic set 7-35. Individual segments in the class-scape can be inspected by a movable cursor (red + sign). By dragging it, all the segment-wise information is updated in real time:

   a) The pitch information of the segment is displayed on top of the interface, showing from left to right: the pitch-class set (specific instantiation of the class), the interval vector, the Forte's name, and the prime form of the class. At the top-right corner, the selected time-scale (duration of the segment in seconds) is also depicted.

   b) The cursor below the class-scape is relocated and/or resized according to the time and time-scale position of the cursor, aligning the segment under inspection with the piano roll.

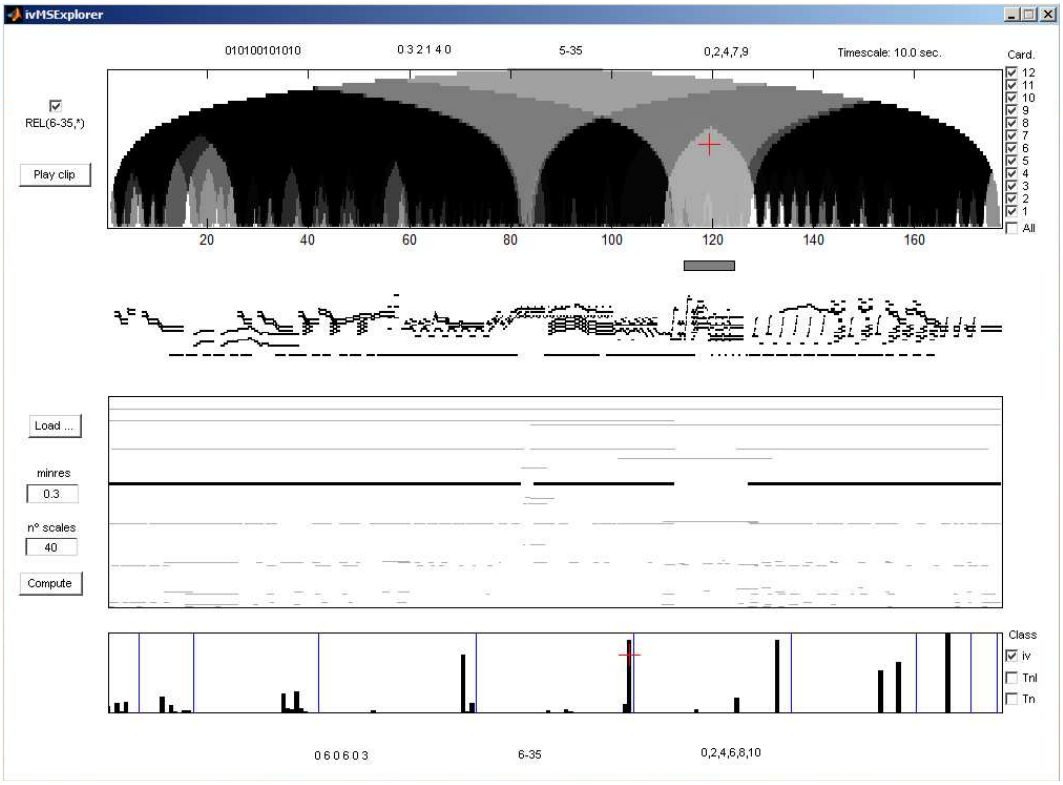

Figure 6.3: Set-class explorer.

2. Piano roll representation of the piece, which serves as a visual index to the composition. Between the class-scape and the piano roll, a horizontal cursor shows the specific segment selected in the class-scape.

3. The class-matrix. It shows the projection of the class-scape in the time vs. class plane. Each row represent the activated frames (if any) of each class over time. An activated point at a given row means that, around that time position, it exists at least one segment which belongs to the corresponding class. Classes are sorted from bottom (1-1) to top (12-1), according to Forte's numbering. The selected class is represented by a bold horizontal line, the rest of the classes remaining in grey. In this example, the three large hexatonic (6-35) sections of the piece are clearly projected in the time axis.

4. The class-vector. It shows all the set-class content of the piece in its most summarised way. It represents the duration percentage of all the possible classes under the chosen equivalence (see additional controls below). The vector is sorted from left (1-1) to right (12-1), according to Forte's numbering. The different cardinalities are separated by vertical

blue lines. The vector is normalised from 0 (non existence of the sonority) to 1 (the sonority spans the whole duration of the piece), helping to localise the prominence of each sonority. A movable cursor (red + sign) in the class-vector plot allow for selecting any class as reference (in the example, 6-35 is chosen). By dragging the class cursor, the class-wise information is updated in real time:

a) The properties of the selected class are updated below the class-vector, from left to right: the interval vector, the Forte's name, and the prime form of the class.

b) The REL filter is configured with respect to the chosen class.

c) The class-scape shows the resulting information, according to both the chosen class and the state of the REL and the cardinality filters (see additional controls below).

The indexing method by navigating the class-vector allows a fast localisation of any segment of music which has the chosen sonority (if any), and the closeness of all the segments with respect to the chosen sonority (even if it does not exist in the piece).

Four additional controls complement the interfacing possibilities:

1. The class-equivalence selector, at the right side of the class-vector. It allows to choose the class space, among the three equivalences discussed in this work (iv, TnI and Tn). The class-equivalence can be modified at any moment. The class-vector, the class-matrix, and (if it applies) the class-scape, are updated in real time, opening the exploration to the description possibilities of the new class space.

2. The REL filter. If checked (as in the example), the keyscape shows all the segments relative to the chosen class. If unchecked, the keyscape only shows the segments belonging to the chosen class (if any).

3. The cardinality filter, at the right side of the class-scape. It allows to select any combination of cardinalities to be displayed in the class-scape. It allows for an easy exploration of ranges of sonorities, and for fast localisation of the involved time-scales.

4. The play button, which reproduces only the selected segment at the class-scape, by triggering the corresponding MIDI sequence. This constitutes a useful complement of the analysis, by realising in sound the particular instantiations of the set-class.

CHAPTER 7

Conclusions

A distilled review of the contributions of this work follows.

Tonal representation

Our general multi-scale method for tonal analysis was firstly conceived as a joint representation of the *spatial* and *temporal* dimensions of tonality. The most convenient domains to achieve so, namely pitch-spaces and time vs. time-scale plots, were interfaced by a novel colouring mechanism, mapping the metric properties of the pitch-space to a unidimensional perceptual variable (colour). This connection between keyscapes and pitch-spaces, intended as an exploration tool beyond a mere visualisation, resulted even more insightful than expected.

Along with the development of the method, some unsolved questions, dating back to the first experiments which led to the considered pitch-spaces, were raised again from a temporal multi-scale perspective. These issues were related with the natural ambiguity of the tonal induction, the stress introduced by the multidimensional scaling solutions, the projection of music segments in pitch-spaces, and the representational limitations of pitch-spaces depending on their underlying tonal categories. The assessment of trustability for the highly summarised and ambiguous information conveyed by the keyscapes, was solved by introducing the concept of confidence-scapes.

We proposed a novel typology of tonal uncertainty, whereby ambiguity is not considered an inherent feature of the music segment alone (as it is usually done), but a *relational* feature which measures the suitability of a given pitch-space for representing a given segment of music. This involves both the similarity of the music stimulus with the tonal categories scaffolding the space *and* the neighbouring relations between the best estimates.

The information summarisation featured by the low-dimensional pitch-spaces, mostly covered in literature as the static properties of tonality *as a system*,

was then explored with respect to both temporal dimensions, namely time and time-scale. That is, we *realised* and analysed the properties of the tonal system with respect to actual pieces of music. We showed that the ambiguity of type I (the natural ambiguity *allowed* by a given tonal system) results in tonal summaries consistent with some compositional and aesthetic principles common in the Classical repertoire. We showed that the ambiguity of type I also arises in cases of fast modulations to non-neighbouring keys (proper of the Romantic repertoire), approaching some theoretical concepts of certain sophistication, such as Lerdahl's "shortest path rule". We also showed the consequences of observing polytonal music through the ambiguity lens. We formalised our typology of ambiguity from the SOM activation perspective, and we discussed a comparison with other formalisations.

The concept of contextual stability was defined in relation to both time and time-scale. We discussed stability as conveyor of tonal information even in cases of extreme ambiguity (of type II). This inspired the extension of our general multi-scale method to different categorical spaces (symmetric modes), as a means to provide alternative *listening* perspectives, which could accommodate tonal systems with different pitch organisations. The analytical potential of the method was demonstrated with music composed under scalar and aesthetic principles alien to the major-minor paradigm.

An alternative colouring method was proposed for improving the quantitative readability of tonal distances in the keyscapes. The feasibility of the method for both audio and symbolic encodings of music was proposed as a means for decoupling the evaluation of the key-estimation methods, in terms of low-level (chroma) and high-level (key estimation) features.

Tonal perception

In the first case study, reusing empirical data obtained by continuous response methods, we evidenced the methodological consequences that time-scale and multidimensional scaling have in the evaluation of a simple model of tonal induction. We adapted our analysis and representation method for a simultaneous inspection of both the music stimuli and the perceptual ratings. We found that the best fitting of the model with the ratings contradicts the general agreement about the short-term memory limitation in tonal perception. Since our model was quite similar to the ones found in the early literature on cognitive psychology, we questioned some taken-for-granted assumptions.

We evidenced the quantitative impact of comparing multidimensional time series in scaled spaces, a practice reported in cognitive psychology literature, with respect to the dimensionality of the empirical ratings. This raised methodological consequences with respect to both model comparison and replicability of

experiments. We discussed how our taxonomy for tonal ambiguity is related with the evaluation problem. We showed evidence supporting that the claimed *success* of some tonal induction methods could be a matter of suitability of the specific stimuli for the model, therefore questioning its generalisation power. We discussed the insufficient information provided by global measures in multidimensional time series comparison.

In the second study, we highlighted some methodological limitations in the modelling of high level tonal concepts, such as tonal tension. A failed experiment was discussed with respect to the challenge of guaranteeing consistency between the participants' responses and the intended perceptual variables being captured. Tonal tension was found a too abstract variable for reliable measurements in the most *naturalistic* experimental setting. The stop-and-rate methodology, although not proved accurate enough for the original purpose of the experiment, supported its feasibility for capturing confident ratings of tonal tension from participants with considerable training in harmony.

We proposed a simple model of tonal instability. We associated the spatial and temporal information in the keyscapes with the *tonal hierarchies* and the *event hierarchies* respectively. This way, the keyscapes captured the essential information of the GTTM's hierarchical tress, through our concept of stability in time and time-scale. We reconsidered a well-known drawback of the KK-profiles at short time-scales, that of (mis)estimating chords for keys, as a convenient means for defining the cross-scale stability conditions in our model. The model was compared with both the theoretical predictions and the empirical ratings. We discussed the relation between the prolongational decisions of the GTTM and the cross-scale alignment in keyscapes.

Tonal context generalised

The limitations of the profiling technique in our general method were approached by extending systematisation, which only applied at the segmentation stage, to the description itself. A set-class level of description was chosen as a compromise between dimensionality and generalisability, and to provide a standard analytical lexicon. The achieved goals were a true objectivity in the description (not involving estimations whatsoever), and the guarantee of characterising every possible sonority (in terms of set-classes) for every possible segment[1]. Three categorical spaces were proposed, corresponding to the most common class-equivalences, namely interval vector (iv), transpositional-inversional (TnI), and transpositional (Tn) equivalences.

The concept of keyscape was adapted to the challenges introduced by the new

[1] Within the limitation of a strict vertical segmentation.

class-spaces: a much larger number of categories, and the lack of generalised low-dimensional spaces able to represent their relations properly. Two solutions were proposed for that: a plain all-or-nothing query-by-class, limited to visualising one single class at a time, and a relative method allowing the simultaneous exploration of all the possible segments in terms of their similarity to any arbitrary sonority (even absent ones). We proposed class-matrices and class-vectors as compact but quite descriptive features. The former provides a complete overview of the set-class inclusions over time, while the latter quantifies the presence of each possible sonority in the piece.

Additional data structures, namely subclass-matrices and subclass-vectors, were proposed for characterising the subclass content underlying any arbitrary sonority. This provides a means for distinguishing different instantiations of a given sonority in real music. The method was demonstrated for the characterisation of different usages of diatonicism in corpora. The method was also the basis of a structural analysis of a serial piece under different class-equivalences. We showed the usage of plain self-similarity matrices fed by class-matrices, for capturing sophisticated recurrences, out of reach from the usual chroma-based approaches.

The method was proposed for the generation of sophisticated, accurate and compact content-based metadata from symbolic music datasets, which would push tonal description to an unprecedented level with respect to the current standards in music information retrieval. The method has been demonstrated in a test dataset for simple and combined queries-by-sonority.

Interfacing tonality

The different multi-scale analysis methods discussed in this thesis, have been implemented as graphical user interfaces for Matlab. Beyond being a mere outcome of the research, these interfaces have provided a means for reasoning along the work. It has been through the exploration of hundreds of musical works in joint representation domains, that several of the relevant questions addressed in this work have actually arisen. It is in this respect that we have proposed the method as a tool for assisting the analysis loop.

These tools have been demonstrated in international conferences. In addition, both the interfaces and the subsidiary modules have been featured for assisting research and educational activities in a number of undergraduate and graduate seminars and courses on musicology, music psychology and music technology (see *Dissemination* in Appendix C).

Agustín Martorell, Barcelona, July 16, 2013.

Bibliography

Aarden, B. J. (2003). *Dynamic Melodic Expectancy.* Ph.D. thesis, The Ohio State University.

Babbit, M. (1955). Some Aspects of Twelve-Tone Composition. *The Score and I.M.A. Magazine*, pp. 53–61.

Bernardini, N., Serra, X., Leman, M., Widmer, G., & De Poli, G. (2009). Sound and Music Computing Network Roadmap.

Bharucha, J. J. & Krumhansl, C. L. (1983). The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context. *Cognition*, *13*(1), 63–102.

Bigand, E. & Parncutt, R. (1999). Perceiving Musical Tension in Long Chord Sequences. *Psychological Research*, *62*(4), 237–254.

Bonds, M. E. (2010). The Spatial Representation of Musical Form. *Journal of Musicology*, *27*(3), 265–303.

Bregman, A. S. (1990). *Auditory Scene Analysis.* Cambridge, MA: MIT Press.

Burgoyne, J. A. & Saul, L. K. (2005). Visualization of Low Dimensional Structure in Tonal Space. In *International Computer Music Conference.*

Castellano, M. a., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *Journal of experimental psychology. General*, *113*(3), 394–412.

Castrén, M. (1994). *RECREL. A Similarity Measure for Set-Classes.* Helsinki: Sibelius Academy.

Chang, C.-L. (2006). *Five Preludes Opus 74 by Alexander Scriabin : The Mystic Chord as Basis for New Means of Harmonic Progression.* Ph.D. thesis, University of Texas at Austin.

Chew, E. (2000). *Towards a Mathematical Model of Tonality.* Ph.D. thesis, M.I.T.

Chew, E. (2006). Slicing It All Ways: Mathematical Models for Tonal Induction, Approximation, and Segmentation Using the Spiral Array. *INFORMS Journal on Computing*, *18*(3), 305–320.

CIE 015 (2004). *Colorimetry*. Commission Internationale de L'Éclairage, 3rd. edn.

Cohn, R. (2003). A Tetrahedral Model of Tetrachordal Voice-Leading Space. *Music Theory Online*, *9*(4).

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

Coombs, C. H. (1964). *A Theory of Data*. New York: Wiley & Sons.

Cuddy, L. L. & Smith, N. A. (2000). Perception of Tonal Pitch Space and Tonal Tension. In *Musicology and Sister Disciplines: Past, Present, Future*, pp. 47–59. Oxford: Oxford University Press.

de Schloezer, B. (1987). *Scriabin: Artist and Mystic*. Berkeley, Los Angeles.

Deliège, C. (1989). La Set-Theory ou les Enjeux du Pléonasme. *Analyse Musicale*, *17*, 64–79.

Eerola, T. & Toiviainen, P. (2004). MIDI Toolbox: MATLAB Tools for Music Research.

Farbood, M. M. (2006). *A Quantitative , Parametric Model of Musical Tension*. Ph.D. thesis, M.I.T.

Farbood, M. M., Marcus, G., & Poeppel, D. (2012). Temporal Dynamics and the Identification of Musical Key. *Journal of Experimental Psychology*, pp. 1–24.

Firmino, E. A. & Bueno, J. L. O. (2008). Tonal Modulation and Subjective Time. *Journal of New Music Research*, *37*(4), 275–297.

Foote, J. (1999). Visualizing Music and Audio Using Self-Similarity. In *ACM Multimedia*, pp. 77–80. Orlando, FL: ACM Press.

Forte, A. (1964). A Theory of Set-Complexes for Music. *Journal of Music Theory*, *8*(2), 136–183.

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

Forte, A. (1989). La Set-Complex Theory: Élevons les Enjeux! *Analyse Musicale*, *17*, 80–86.

Forte, A. & Gilbert, S. E. (1982). *Introduction to Schenkerian Analysis*. New York: Norton.

Garner, W. R. (1970). Good Patterns Have Few Alternatives. *American Scientist*, *58*, 34–42.

Gómez, E. (2006). *Tonal Description of Audio Music Signals.* Ph.D. thesis, Universitat Pompeu Fabra.

Hanson, H. (1960). *The Harmonic Materials of Modern Music: Resources of the Tempered Scale.* New York: Appleton-Century-Crofts.

Hasty, C. (1981). Segmentation and Process in Post-Tonal Music. *Music Theory Spectrum*, *3*, 54–73.

Heinichen, J. D. (1728). *General-Bass in der Composition.* Dresden: Facs. Hildesheim: Olms 1969.

Huovinen, E. & Tenkanen, A. (2007). Bird's-Eye Views of the Musical Surface: Methods for Systematic Pitch-Class Set Analysis. *Music Analysis*, *26*(1-2), 159–214.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Hyer, B. (2013). Tonality. In *Grove Music Online. Oxford Music Online.* Oxford University Press, accessed June 21, 2013, http://www.oxfordmusiconline.com/subscriber/article/grove/music/28102.

Isaacson, E. J. (1990). Similarity of Interval-Class Content between Pitch-Class Sets: the IcVSIM Relation. *Journal of Music Theory*, *34*(1), 1–28.

Janata, P. (2008). Navigating Tonal Space. In W. B. Hewlett, E. Selfridge-Field, & E. Correia Jr. (Eds.) *Tonal Theory for the Digital Age*, pp. 39–50. Stanford, CA: CCARH.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The Cortical Topography of Tonal Structures Underlying Western Music. *Science*, *298*(5601), 2167–70.

Kohonen, T. (1997). *Self-Organizing Maps.* Berlin: Springer.

Kostka, S. & Payne, D. (1995). *Workbook for Tonal Harmony.* New York: McGraw Hill.

Koulis, T., Ramsay, J. O., & Levitin, D. J. (2008). From Zero to Sixty. Calibrating Real-Time Responses. *Psychometrika*, *73*(2), 321–339.

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

Krumhansl, C. L. (2004). The Cognition of Tonality – as We Know it Today. *Journal of New Music Research*, *33*(3), 253–268.

Krumhansl, C. L. & Kessler, E. J. (1982). Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. *Psychological Review*, *89*, 334–368.

Krumhansl, C. L. & Schmuckler, M. A. (1986). The Petroushka Chord. *Music Perception*, *4*, 153–184.

Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the Hierarchy of Tonal Functions Within a Diatonic Context. *Journal of Experimental Psychology*, *5*, 579–594.

Krumhansl, C. L. & Toiviainen, P. (2001). Tonal Cognition. *Annals of the New York Academy of Sciences*, *930*, 77–91.

Krumhansl, C. L. & Toiviainen, P. (2003). Tonal Cognition. In I. Peretz & R. Zatorre (Eds.) *The Cognitive Neuroscience of Music*, pp. 95–108. New York: Oxford University Press.

Kuusi, T. (2001). Set-Class and Chord : Examining Connection between Theoretical Resemblance and Perceived Closeness. Tech. rep.

Lakoff, G. & Johnson, M. (1980). *Metaphors we Live By*. Chicago: University of Chicago Press.

Lartillot, O. & Toiviainen, P. (2007). MIR in Matlab (II): A Toolbox for Musical Feature Extraction From Audio. In *International Conference on Music Information Retrieval*.

Leman, M. (2000). An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings. *Music Perception*, *17*(4), 481–510.

Leman, M. (2003). Foundations of Musicology as Content Processing Science. *Journal of Music and Meaning*, *1*.

Lerdahl, F. (1996). Calculating Tonal Tension. *Music Perception*, *13*(3), 319–363.

Lerdahl, F. (2001). *Tonal Pitch Space*. New York: Oxford University Press.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lerdahl, F. & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception*, *24*(4), 329–366.

Lewin, D. (1959). Re : Intervallic Relations between Two Collections of Notes. *Journal of Music Theory*, *3*(2), 298–301.

Lewin, D. (1979). Some New Constructs Involving Abstract Pcsets and Probabilistic Applications. *Perspectives of New Music*, *18*(1), 433–444.

Lewin, D. (1987). *Generalized Musical Intervals and Transformations*. New Haven, CT: Yale University Press.

Messiaen, O. (1944). *Technique de mon language musical*. Paris: Alphonse Leduc.

Mey, J. L. (2001). *Pragmatics: An Introduction*. Oxford: Blackwell, 2nd. edn.

Meyer, L. B. (1956). Emotion and Meaning in Music.

Meyer, L. B. (1996). Commentary. *Music Perception*, *13*(3), 455–483.

Müller, M. (2007). *Information Retrieval for Music and Motion*. Berlin, Heidelberg: Springer.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press.

Nattiez, J.-J. (2003). Allen Forte's Set Theory, Neutral Level Analysis and Poietics. In M. Andreatta, J.-M. Bardez, & J. Rahn (Eds.) *Around Set Theory*, pp. 1–18. Paris: Delatour-France / Ircam.

Palmer, C. (1996). Anatomy of a Performance: Sources of Musical Expression. *Music Perception*, *13*, 433–453.

Papadopoulos, H. (2010). *Joint Estimation of Musical Content Information From an Audio Signal*. Ph.D. thesis, IRCAM.

Parncutt, R. (1988). Revision of Terhardt's Psychoacoustical Model of the Root(s) of a Musical Chord. *Music Perception*, *6*, 65–94.

Parncutt, R. (1989). *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.

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

Pople, A. (1983). Skryabin's Prelude, Op. 67, No. 1: Sets and Structure. *Music Analysis*, *2*(2), 151–173.

Pritchard, D. & Theiler, J. (1994). Generating Surrogate Data for Time Series with Several Simultaneously Measured Variables. *Physical Review Letters*, *73*(7), 951–954.

Purwins, H. (2005). *Profiles of Pitch Classes Circularity of Relative Pitch and Key – Experiments, Models, Computational Music Analysis, and Perspectives*. Ph.D. thesis.

Pyper, B. J. & Peterman, R. M. (1998). Comparison of Methods to Account for Autocorrelation in Correlation Analyses of Fish Data. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(9), 2127–2140.

Quinn, I. (2006). General Equal-Tempered Harmony. *Perspectives of New Music*, *44*(2), 5–60.

Quinn, I. (2007). General Equal-Tempered Harmony: Part II. *Perspectives of New Music*, *45*(1), 6–65.

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

Riemann, H. (1893). *Harmony Simplified, or the Theory of the Tonal Functions of Chords*. London: Tr. H. Weberunge, London: Augener, 1896.

Rive, T. N. (1969). An Examination of Victoria's Technique of Adaptation and Reworking in his Parody Masses - With Particular Attention to Harmonic and Cadential Procedure. *Anuario Musical*, *24*, 133–152.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Rosch, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, *7*, 573–605.

Rosen, C. (1972). *The Classical Style*. New York: Norton.

Sapp, C. S. (2005). Visual hierarchical key analysis. *Computers in Entertainment*, *3*(4), 3.

Schenker, H. (1935). *Free Composition*. New York: Longman.

Schoenberg, A. (1969). *Structural Functions of Harmony*. New York: Norton, rev. ed. edn.

Schubert, E. (2001). Continuous Measurement of Self-Report Emotional Responses to Music. In *Music and Emotion*, pp. 393–414.

Shepard, R. N. (1962). The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function (I & II). *Psychometrika*, *27*, 125–140, 219–246.

Shepard, R. N. (1982). Geometrical Approximations to the Structure of Musical Pitch. *Psychological Review*, *89*, 305–333.

Straus, J. N. (2000). *Introduction to Post-Tonal Theory*. New Jersey: Prentice-Hall.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: The M.I.T. Press.

Temperley, D. (2007). *Music and Probability.* Cambridge, MA: MIT Press.

Temperley, D. (2008). The Tonal Properties of Pitch-Class Sets : Tonal Implication , Tonal Ambiguity , and Tonalness. In W. B. Hewlett, E. Selfridge-Field, & E. Correia Jr. (Eds.) *Tonal Theory for the Digital Age*, pp. 24–38. Stanford, CA: CCARH.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit Learning of Tonality: A Self-Organizing Approach. *Psychological Review*, *107*(4), 885–913.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2003). Learning and Perceiving Musical Structures: Further Insights from Artificial Neural Networks. In *The Cognitive Neuroscience of Music*, pp. 109–123. New York and Oxford: Oxford University Press.

Toiviainen, P. (2008). Visualization of Tonal Content in the Symbolic and Audio Domains. In *Tonal Theory for the Digital Age*, vol. 35, pp. 187–199.

Toiviainen, P. & Krumhansl, C. L. (2003). Measuring and Modeling Real-Time Responses to Music: The Dynamics of Tonality Induction. *Perception*, *32*(6), 741–766.

Treitler, L. (1997). Language and the Interpretation of Music. In J. Robinson (Ed.) *Music and Meaning*, pp. 23–56. London: Cornell University Press.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Tymoczko, D. (2012). The Generalized Tonnetz. *Journal of Music Theory*, *56*(1), 1–5.

Van den Toorn, P. (1983). *The Music of Igor Stravinsky.* New Haven, CT: Yale University Press.

Vinet, H. (2007). Science and Technology of Music and Sound: The IRCAM Roadmap. *Journal of New Music Research*, *36*(3), 207–226.

Vos, P. G. & Leman, M. (2000). Tonality Induction. *Music Perception*, *17*(4), 401–548.

Vos, P. G. & Van Geenen, E. W. (1996). A Parallel-Processing Key-Finding Model. *Music Perception*, *14*, 185–223.

Weber, G. (1821). *Versuch einer geordeneten Theorie der Tonsetzkunst.* Mainz: Schotts Söhne.

Werts, D. (1983). *A Theory of Scale References.* Ph.D. thesis, Princeton University.

Wiggins, G. a. (2009). Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music. *IEEE International Symposium on Multimedia*, pp. 477–482.

Appendix A: Set-classes

class : Forte's name of the set-class (cardinality - Forte's ordinal).
iv : interval vector.
prime (A) : prime form.
prime (B) : inverted prime form.

Example with set-class 3-11 (cardinality 3, Forte's ordinal 11):
iv-equivalence: ⟨**001110**⟩ : 1 minor 3^{rd}, 1 major 3^{rd}, 1 4^{th} (all minor and major triads).
TnI-equivalence: **3-11** : {0,3,7} (all minor and major triads).
Tn-equivalence: **3-11A** : {0,3,7} (all minor triads) or **3-11B** : {0,4,7} (all major triads).

class	iv	prime (A)	prime (B)	class	iv	prime (A)	prime (B)
1-1	⟨000000⟩	{0}		4-24	⟨020301⟩	{0248}	
2-1	⟨100000⟩	{01}		4-25	⟨020202⟩	{0268}	
2-2	⟨010000⟩	{02}		4-26	⟨012120⟩	{0358}	
2-3	⟨001000⟩	{03}		4-27	⟨012111⟩	{0258}	{0368}
2-4	⟨000000⟩	{04}		4-28	⟨004002⟩	{0369}	
2-5	⟨000010⟩	{05}		5-1	⟨432100⟩	{01234}	
2-6	⟨000001⟩	{06}		5-2	⟨332110⟩	{01235}	{02345}
3-1	⟨210000⟩	{012}		5-3	⟨322210⟩	{01245}	{01345}
3-2	⟨111000⟩	{013}	{023}	5-4	⟨322111⟩	{01236}	{03456}
3-3	⟨101100⟩	{014}	{034}	5-5	⟨321121⟩	{01237}	{04567}
3-4	⟨100110⟩	{015}	{045}	5-6	⟨311221⟩	{01256}	{01456}
3-5	⟨100011⟩	{016}	{056}	5-7	⟨310132⟩	{01267}	{01567}
3-6	⟨020100⟩	{024}		5-8	⟨232201⟩	{02346}	
3-7	⟨011010⟩	{025}	{035}	5-9	⟨231211⟩	{01246}	{02456}
3-8	⟨010101⟩	{026}	{046}	5-10	⟨223111⟩	{01346}	{02356}
3-9	⟨010020⟩	{027}		5-11	⟨222220⟩	{02347}	{03457}
3-10	⟨002001⟩	{036}		5-Z12	⟨222121⟩	{01356}	
3-11	⟨001110⟩	{037}	{047}	5-Z36		{01247}	{03567}
3-12	⟨000300⟩	{048}		5-13	⟨221311⟩	{01248}	{02348}
4-1	⟨321000⟩	{0123}		5-14	⟨221131⟩	{01257}	{02567}
4-2	⟨221100⟩	{0124}	{0234}	5-15	⟨220222⟩	{01268}	
4-3	⟨212100⟩	{0134}		5-16	⟨213211⟩	{01347}	{03467}
4-4	⟨211110⟩	{0125}	{0345}	5-Z17	⟨212320⟩	{01348}	
4-5	⟨210111⟩	{0126}	{0456}	5-Z37		{03458}	
4-6	⟨210021⟩	{0127}		5-Z18	⟨212221⟩	{01457}	{02367}
4-7	⟨201210⟩	{0145}		5-Z38		{01258}	{03678}
4-8	⟨200121⟩	{0156}		5-19	⟨212122⟩	{01367}	{01467}
4-9	⟨200022⟩	{0167}		5-20	⟨211231⟩	{01568}	{02378}
4-10	⟨122010⟩	{0235}		5-21	⟨202420⟩	{01458}	{03478}
4-11	⟨121110⟩	{0135}	{0245}	5-22	⟨202321⟩	{01478}	
4-12	⟨112101⟩	{0236}	{0346}	5-23	⟨132130⟩	{02357}	{02457}
4-13	⟨112011⟩	{0136}	{0356}	5-24	⟨131221⟩	{01357}	{02467}
4-14	⟨111120⟩	{0237}	{0457}	5-25	⟨123121⟩	{02358}	{03568}
4-Z15	⟨111111⟩	{0146}	{0256}	5-26	⟨122311⟩	{02458}	{03468}
4-Z29		{0137}	{0467}	5-27	⟨122230⟩	{01358}	{03578}
4-16	⟨110121⟩	{0157}	{0267}	5-28	⟨122212⟩	{02368}	{02568}
4-17	⟨102210⟩	{0347}		5-29	⟨122131⟩	{01368}	{02578}
4-18	⟨102111⟩	{0147}	{0367}	5-30	⟨121321⟩	{01468}	{02478}
4-19	⟨101310⟩	{0148}	{0348}	5-31	⟨114112⟩	{01369}	{02369}
4-20	⟨101220⟩	{0158}		5-32	⟨113221⟩	{01469}	{02569}
4-21	⟨030201⟩	{0246}		5-33	⟨040402⟩	{02468}	
4-22	⟨021120⟩	{0247}	{0357}	5-34	⟨032221⟩	{02469}	
4-23	⟨021030⟩	{0257}		5-35	⟨032140⟩	{02479}	

Table 1: Set-classes, cardinalities 1-5.

class	iv	prime (A)	prime (B)	class	iv	prime (A)	prime (B)
6-1	⟨543210⟩	{012345}		7-10	⟨445332⟩	{0123469}	{0234569}
6-2	⟨443211⟩	{012346}	{023456}	7-11	⟨444441⟩	{0134568}	{0234578}
6-Z3	⟨433221⟩	{012356}	{013456}	7-Z12	⟨444342⟩	{0123479}	
6-Z36		{012347}	{034567}	7-Z36		{0123568}	{0235678}
6-Z4	⟨432321⟩	{012456}		7-13	⟨443532⟩	{0124568}	{0234678}
6-Z37		{012348}		7-14	⟨443352⟩	{0123578}	{0135678}
6-5	⟨422232⟩	{012367}	{014567}	7-15	⟨442443⟩	{0124678}	
6-Z6	⟨421242⟩	{012567}		7-16	⟨435432⟩	{0123569}	{0134569}
6-Z38		{012378}		7-Z17	⟨434541⟩	{0124569}	
6-7	⟨420243⟩	{012678}		7-Z37		{0134578}	
6-8	⟨343230⟩	{023457}		7-Z18	⟨434442⟩	{0145679}	{0234589}
6-9	⟨342231⟩	{012357}	{024567}	7-Z38		{0124578}	{0134678}
6-Z10	⟨333321⟩	{013457}	{023467}	7-19	⟨434343⟩	{0123679}	{0123689}
6-Z39		{023458}	{034568}	7-20	⟨433452⟩	{0125679}	{0234789}
6-Z11	⟨333231⟩	{012457}	{023567}	7-21	⟨424641⟩	{0124589}	{0134589}
6-Z40		{012358}	{035678}	7-22	⟨424542⟩	{0125689}	
6-Z12	⟨332232⟩	{012467}	{013567}	7-23	⟨354351⟩	{0234579}	{0245679}
6-Z41		{012368}	{025678}	7-24	⟨353442⟩	{0123579}	{0246789}
6-Z13	⟨324222⟩	{013467}		7-25	⟨345342⟩	{0234679}	{0235679}
6-Z42		{012369}		7-26	⟨344532⟩	{0134579}	{0245689}
6-14	⟨323430⟩	{013458}	{034578}	7-27	⟨344451⟩	{0124579}	{0245789}
6-15	⟨323421⟩	{012458}	{034678}	7-28	⟨344433⟩	{0135679}	{0234689}
6-16	⟨322431⟩	{014568}	{023478}	7-29	⟨344352⟩	{0124679}	{0235789}
6-Z17	⟨322332⟩	{012478}	{014678}	7-30	⟨343542⟩	{0124689}	{0135789}
6-Z43		{012568}	{023678}	7-31	⟨336333⟩	{0134679}	{0235689}
6-18	⟨322242⟩	{012578}	{013678}	7-32	⟨335442⟩	{0134689}	{0135689}
6-Z19	⟨313431⟩	{013478}	{014578}	7-33	⟨262623⟩	{012468A}	
6-Z44		{012569}	{014569}	7-34	⟨254442⟩	{013468A}	
6-20	⟨303630⟩	{014589}		7-35	⟨254361⟩	{013568A}	
6-21	⟨242412⟩	{023468}	{024568}	8-1	⟨765442⟩	{01234567}	
6-22	⟨241422⟩	{012468}	{024678}	8-2	⟨665542⟩	{01234568}	{02345678}
6-Z23	⟨234222⟩	{023568}		8-3	⟨656542⟩	{01234569}	
6-Z45		{023469}		8-4	⟨655552⟩	{01234578}	{01345678}
6-Z24	⟨233331⟩	{013468}	{024578}	8-5	⟨654553⟩	{01234678}	{01245678}
6-Z46		{012469}	{024569}	8-6	⟨654463⟩	{01235678}	
6-Z25	⟨233241⟩	{013568}	{023578}	8-7	⟨645652⟩	{01234589}	
6-Z47		{012479}	{023479}	8-8	⟨644563⟩	{01234789}	
6-Z26	⟨232341⟩	{013578}		8-9	⟨644464⟩	{01236789}	
6-Z48		{012579}		8-10	⟨566452⟩	{02345679}	
6-27	⟨225222⟩	{013469}	{023569}	8-11	⟨565552⟩	{01234579}	{02456789}
6-Z28	⟨224322⟩	{013569}		8-12	⟨556543⟩	{01345679}	{02345689}
6-Z49		{013479}		8-13	⟨556453⟩	{01234679}	{02356789}
6-Z29	⟨224232⟩	{023679}		8-14	⟨555562⟩	{01245679}	{02345789}
6-Z50		{014679}		8-Z15	⟨555553⟩	{01234689}	{01356789}
6-30	⟨224223⟩	{013679}	{023689}	8-Z29		{01235679}	{02346789}
6-31	⟨223431⟩	{014579}	{024589}	8-16	⟨554563⟩	{01235789}	{01246789}
6-32	⟨143250⟩	{024579}		8-17	⟨546652⟩	{01345689}	
6-33	⟨143241⟩	{023579}	{024679}	8-18	⟨546553⟩	{01235689}	{01346789}
6-34	⟨142422⟩	{013579}	{024689}	8-19	⟨545752⟩	{01245689}	{01345789}
6-35	⟨060603⟩	{02468A}		8-20	⟨545662⟩	{01245789}	
7-1	⟨654321⟩	{0123456}		8-21	⟨474643⟩	{0123468A}	
7-2	⟨554331⟩	{0123457}	{0234567}	8-22	⟨465562⟩	{0123568A}	{0134568A}
7-3	⟨544431⟩	{0123458}	{0345678}	8-23	⟨465472⟩	{0123578A}	
7-4	⟨544332⟩	{0123467}	{0134567}	8-24	⟨464743⟩	{0124568A}	
7-5	⟨543342⟩	{0123567}	{0124567}	8-25	⟨464644⟩	{0124678A}	
7-6	⟨533442⟩	{0123478}	{0145678}	8-26	⟨456562⟩	{0134578A}	
7-7	⟨532353⟩	{0123678}	{0125678}	8-27	⟨456553⟩	{0124578A}	{0134678A}
7-8	⟨454422⟩	{0234568}		8-28	⟨448444⟩	{0134679A}	
7-9	⟨453432⟩	{0123468}	{0245678}				

Table 2: Set-classes, cardinalities 6-8.

class	iv	prime (A)	prime (B)	class	iv	prime (A)	prime (B)
9-1	⟨876663⟩	{012345678}		9-11	⟨667773⟩	{01235679A}	{01245679A}
9-2	⟨777663⟩	{012345679}	{023456789}	9-12	⟨666963⟩	{01245689A}	
9-3	⟨767763⟩	{012345689}	{013456789}	10-1	⟨988884⟩	{0123456789}	
9-4	⟨766773⟩	{012345789}	{012456789}	10-2	⟨898884⟩	{012345678A}	
9-5	⟨766674⟩	{012346789}	{012356789}	10-3	⟨889884⟩	{012345679A}	
9-6	⟨686763⟩	{01234568A}		10-4	⟨888984⟩	{012345689A}	
9-7	⟨677673⟩	{01234578A}	{01345678A}	10-5	⟨888894⟩	{012345789A}	
9-8	⟨676764⟩	{01234678A}	{01245678A}	10-6	⟨888885⟩	{012346789A}	
9-9	⟨676683⟩	{01235678A}		11-1	⟨AAAAA5⟩	{0123456789A}	
9-10	⟨668664⟩	{01234679A}		12-1	⟨CCCCC6⟩	{0123456789AB}	

Table 3: Set-classes, cardinalities 9-12.

Appendix B: REL distance

The algorithm for computing the REL distance between two classes A and B follows:

1. For each of the classes A and B, a *subset vector* is computed. This subset vector is a 357-dimensional vector, which consists on the interval vector of the class (6-dimensional), followed by a 351-dimensional vector[2], which accounts for the number of occurrences of each *Tn-type chord* in the class. As an example, the class 4-20 has the subset vector:

   [1 0 1 2 2 0 4 1 0 1 2 2 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 ... zeros ...]

   The prime form of 4-20 is {0,1,5,8}, realised as pcset as [110001001000].

   The interval class of 4-20 is [1 0 1 2 2 0], as it appears at the beginning of the subset vector.

   Following Forte's cardinality-ordinal arrangement of classes, the first class is the single pitch-class (1-1), whose prime form is {0} (no intervals). The class 4-20 contains 4 of such sets, and that is the value of the corresponding position in the subset vector. The next class to check is 2-1, whose prime form is {0,1}. It corresponds to dyads a semitone away. The class 4-20 contains 1 of such dyads, so that is the next value in the subset vector. The method follows up to exhaust the 351 Tn-types.

2. We denote sub(X,i) as the i^{th} element of the subset vector corresponding to the class X.

3. The REL distance between the classes A and B is then computed as:

$$REL(A,B) = \frac{\sum_{i=1}^{357} \sqrt{sub(A,i) \cdot sub(B,i)}}{\sqrt{\sum_{i=1}^{357} sub(A,i) \cdot \sum_{i=1}^{357} sub(B,i)}}$$

[2]REL distance distinguishes classes at the level of Tn-equivalence, which sums to 351 different classes, including the trivial forms.

Appendix C: Publications

Submitted to ISI-indexed journals

Martorell, A., Gómez, E. (2013). Hierarchical Multi-Scale Set-Class Analysis. *Journal of Mathematics and Music.*

Full-article contributions to international peer-reviewed conferences

Martorell, A., Toiviainen, P., Gómez, E. (2012). Temporal Multi-Scale Considerations in the Modeling of Tonal Cognition from Continuous Rating Experiments. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, K. Pastiadis (Eds.), *Proc. of the 12th International Conference on Music Perception and Cognition.* Thessaloniki: University of Thessaloniki, 660-665.

Martorell, A., Gómez, E. (2011). Two-Dimensional Visual Inspection of Pitch-Space, Many Time-Scales and Tonal Uncertainty Over Time. *Int. Conf. on Mathematics and Computation in Music* (*MCM*), IRCAM, Paris. LNAI 6726. C. Agon, E. Amiot, M. Andreatta, G. Assayag, J. Bresson and J. Mandereau, eds. Berlin:Springer-Verlag, pp. 140-150.

Dissemination (research)

Martorell. A. (2012). Tonal Spaces: Theoretical Foundations, Music Perception and Mediation Technology. Graduate Seminars on Musicology. Universitat de Barcelona.

Martorell, A. (2011). Modelling of Non-Stationary Tonal Processes by Temporal Multi-Scale Analysis. Graduate Seminars on Music Perception. Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland.

Dissemination (educational)

Martorell. A. (2013). Tonality: MIR Meets Interaction. Postgraduate Course in the Design of Interactive Musical Systems. IDEC-UPF, Barcelona.

Martorell, A. (2011, 2012). Pitch-Spaces and Tonality: Concepts, Representations and Interactive Issues. Postgraduate Course in the Design of Interactive Musical Systems. IDEC-UPF, Barcelona.

Martorell, A. (2011). Temporal Multi-Scale Description of Musical Audio. Undergraduate Seminars on Sonology, Escola Superior de Musica de Catalunya (ESMuC), Barcelona.

Tonal Analysis Toolbox for Matlab

Built upon the MIDI Toolbox and the MIR Toolbox, it is comprised of the set of functions and related data developed along this thesis. All the musical examples discussed in this work have been analysed using these tools. The Toolbox includes two of the interfaces discussed in Chapter 6, namely the basic tonal explorer and the set-class explorer, readily usable as end-user prototypes.