Chapter 4
Contextual Set-Class Analysis

Agustín Martorell and Emilia Gómez

Abstract In this chapter, we review and elaborate a methodology for contextual multi-scale set-class analysis of pieces of music. The proposed method provides a systematic approach to segmentation, description and representation in the analysis of the musical surface. The introduction of a set-class description domain provides a systematic, mid-level, and standard analytical lexicon, which allows for the description of any notated music based on a fixed temperament. The method benefits from representation completeness, a balance between generalization and discrimination of the set-class spaces, and access to hierarchical inclusion relations over time. Three new data structures are derived from the method: *class-scapes*, *class-matrices* and *class-vectors*. A *class-scape* represents, in a visual way, the set-class content of each possible segment in a piece of music. The *class-matrix* represents the presence of each possible set class over time, and is invariant to time scale and to several transformations of analytical interest. The *class-vector* summarizes a piece by quantifying the temporal presence of each possible set class. The balance between dimensionality and informativeness provided by these descriptors is discussed in relation to standard content-based tonal descriptors and music information retrieval applications. The interfacing possibilities of the method are also discussed.

4.1 Introduction

The representation of the musical surface is a fundamental element for analysis, as it constitutes the raw material to be explained by the analyst. As for any information representation problem, favouring the observation of some specific musical parameters comes at the price of misrepresenting others. The choice of an appropriate surface characterization is not only a critical step that conditions the whole analysis,

Agustín Martorell · Emilia Gómez
Universitat Pompeu Fabra, Barcelona, Spain
e-mail: {agustin.martorell, emilia.gomez}@upf.edu

but it also often requires considerable effort from the analyst. Proper music information representations and interfacing techniques can free analysts from the most systematic, time-consuming and error-prone tasks, and assist them in finding and testing material of analytical relevance. Human analysts can then exploit the preprocessed data using their perception, intuition and inference abilities, in order to make appropriate analytical decisions. The best processing capabilities of computers (good systematization, poor inference) and humans (good inference, poor systematization) can then complement each other in a constructive way.

In this chapter, we propose an approach to the musical surface, in terms of generalized *contextual* pitch-wise information. This way of conceiving of the surface extends the usual application domain of analysis in terms of *events* (e.g., chords), giving access to richer, multi-scale, hierarchical information about the music. We propose that three aspects of the analytical process are particularly relevant for such an endeavour: first, the segmentation of the music into analysable units; second, the description of these units in adequate terms; and, third, the representation of the results in ways that are manageable and provide insight. We consider each of these three aspects from a systematic point of view. First, we present a comprehensive segmentation policy that extracts every possible (different) segment of music. Second, we propose that these segments be described in terms of pitch-class set theory. Third, we propose usable data structures for representing the resulting data.

The remainder of this chapter is organized as follows. In Sect. 4.2, we present the problems of segmentation, description and representation in the context of surface analysis, and introduce basic set-theoretical concepts. In Sect. 4.3, we describe the proposed computational approach. In Sect. 4.4, we introduce the basic set of data structures, while in Sects. 4.5 and 4.6, we discuss these data structures in the context of several music information retrieval (MIR) applications. Finally, in Sect. 4.7, we consider the interfacing possibilities of the descriptors.

4.2 Background and Motivation

4.2.1 On Segmentation

A main concern for analysis is the grouping of musical elements into units of analytical pertinence, a complex process that we refer to here by the term *segmentation*. According to Forte (1973, p. 83),

> by segmentation is meant the procedure of determining which musical units of a composition are to be regarded as analytical objects.

In the context of pitch-class set analysis, Cook (1987, p. 146) observes that

> no set-theoretical analysis can be more objective, or more well-founded musically, than its initial segmentation.

Considering the joint problem of finding appropriate segments and assessing their pertinence, a pursuit of systematization and *neutrality* in the surface characterization arises as a convenient preprocessing stage.[1] On the other hand, there is no such thing as a truly neutral representation, as pointed out by Huron (1992, p. 38):

> a representation system as a whole may be viewed as a signifier for a particular assumed or explicit explanation. . .

In our work, we aim for neutral representations in order to allow for a wide range of musical idioms to be analysed using only a limited set of assumptions. Our method assumes the concept of a *musical note*, temporally limited between its onset and offset. We also assume octave equivalence in a twelve-tone equal-tempered pitch system. The general framework, however, can be adapted for any discrete pitch organization of the octave. We do not, of course, claim that systematization and neutrality suffice for analysis. Indeed, one could distinguish between a *systematic* segmentation (as a preprocessing step), and an *analytic* (i.e., meaningful) segmentation

> not as something imposed upon the work, but rather. . . as something to be discovered.
>
> (Hasty, 1981, p. 59)

In our approach, the term *segment* is used formally to mean a temporal interval (i.e., chunk) of music, irrespective of its actual musical content. This content may be analysed by subsequent processing stages.

Systematization can, of course, be conceived of in various ways. Conventional segmentation strategies, such as those applied by human analysts working on scores, rely on some prior rhythmic or metric knowledge of the music. For instance, the analysis of chords benefits from a temporal segmentation into beats or bars. This information, however, may not be explicitly available in computer-based music encodings, such as a MIDI file generated from a live performances or automatically transcribed from an audio file. Moreover, this kind of segmentation may not be meaningful or appropriate for certain types of music (e.g., free rhythm, melismatic or *ad-libitum* passages).

A useful, general-purpose, segmentation approach is to apply a *sliding window* with an appropriate *resolution*. In this context, a *window* is a temporal *frame*, which isolates a segment of music to be analysed. A window which is gradually *displaced* over time is referred to as a *sliding window*, and allows sequential discrete observations. The *resolution* of a sliding window is given by its duration (time scale or window size) and the amount of time displaced between consecutive frames (the *hop size*). This resolution is obviously a critical parameter. It depends on the musical parameter under inspection, as well as on the music itself, so adaptive methods may be required in general. In this sense, a systematic segmentation may be thought of as the problem of inspecting *every* possible (different) segment in the music, no matter its position in time or its duration. By adopting the concept of *note*, it is possible to reduce these infinitely many segments to a finite set of segments capturing every distinct pitch aggregate in the piece.

---

[1] For a discussion about neutrality and the risks of circular reasoning in the context of set-theoretical analysis, see Deliège (1989), Forte (1989) and Nattiez (2003).

Forte claimed that this kind of systematic segmentation (which he called 'imbrication') would be impractical for analysis, as it would result in an unmanageable number of overlapping segments, most of which would not be musically significant (see Cook (1987, p. 147) for a discussion). Forte thus claims a need for 'editing' these segments according to criteria which would depend on the specific music, a process unlikely to be systematized in any useful way (Forte, 1973, pp. 90–91). These observations were made with the conventional methods and goals of (academic) musical analysis in mind. For more general (and modest) analysis-related applications, however, one can take advantage of computational methods. In our work, we adopt some interface-design strategies and filtering techniques that can handle such a cumbersome amount of information, and we then apply them to specific music information retrieval tasks.

Among the most systematic approaches to general pitch-class set analysis, Huovinen and Tenkanen (2007) propose a segmentation algorithm based on a 'tail-segment array', in which each note in a composition is associated with all the possible segments of a given cardinality that contain it.[2] This segmentation is combined with certain 'detector functions' to obtain summarized information from the piece. This segmentation method has a number of shortcomings (see Martorell and Gómez, 2015). First, the note-based indexing in terms of 'tail segments' can result in many segments that only contain some of the notes in a vertical chord. Second, considering only segments of a specified cardinality undermines systematization by defining a content-dependent segmentation. Moreover, it undermines neutrality, by defining cardinality to be an analytical parameter, resulting in most of the possible segments being ignored and, in some cases, absurd segmentations. Third, the non-uniqueness of the tail-segment arrays associated with a given time point can result in an ill-defined temporal continuity. For example, in their analysis of Bach's *Es ist genug* (BWV 60), Huovinen and Tenkanen (2007, p. 163) order the notes lexicographically, sorting first by time and then by pitch height. This means that two 'successive' notes may either have different onset times or they may be from the same chord and thus have the same onset time. In such cases, the temporal ordering of the notes is therefore ill-defined.

4.2.2 On Description

In our work, the description stage consists of associating each segment with some content-based property of analytical interest. The descriptive value of such information depends on both the property of the segment that we are interested in (e.g., pitch, rhythm or timbre), and on the way in which the retrieved information is represented (e.g., different ways of encoding or summarizing pitch information). The design of a proper feature space requires some considerations and compromises, among which we identify the following.

[2] See Sect. 4.2.2.1 for a formal definition of *cardinality*.

1. *The trade-off between discrimination and generalization* On the one hand, the descriptive *lexicon* should be able to identify distinct musical realities. On the other hand, such a lexicon should be able to connect segments related by some analytical and/or perceptual criteria. A description *for analysis* should facilitate the observation of relationships.
2. *The trade-off between dimensionality and informativeness* Related to the previous point, this involves maximizing the overall informativeness, using a reasonable (i.e., manageable) lexicon of musical objects.
3. *Completeness* We want a system that is capable of describing *any* possible segment in meaningful ways, covering a wide range of musical idioms.
4. *Communicability* The descriptive lexicon should not only be meaningful, but, ideally, readily understandable by humans, so that analytical conclusions can be derived and explained. This implies that some standard musical terminology is desirable.

4.2.2.1 Pitch-Class Set Concepts

In this section, we will review some basic concepts from pitch-class set theory, as they constitute the descriptive basis of our method. The *pitch class* (Babbitt, 1955) is defined, for the twelve-tone equal-tempered (TET) system, as an integer representing the residue class modulo 12 of some continuous integer representation of pitch, that is, any pitch (represented as a single integer) is mapped to a pitch-class by removing its octave information. A *pitch-class set* (henceforth *pc-set*) is an unordered set of pitch classes (i.e., without repetitions). In the twelve-tone equal-tempered system, there exist $2^{12} = 4096$ distinct pc-sets, so a vocabulary of 4096 symbols is required for describing any possible segment of music. Any pc-set can also be represented by its intervallic content (Hanson, 1960). Intervals considered regardless of their direction are referred to as *interval classes*. There exist 6 different interval classes, corresponding to 1, 2, 3, 4, 5 and 6 semitones respectively. The remaining intervals in the octave are mapped to these 6 classes by inversion. For instance, the perfect fifth (7 semitones up) is mapped to the perfect fourth (5 semitones down), and represented by the interval class 5. The total count of interval classes in a pc-set can be arranged as a 6-dimensional data structure called an *interval vector* (Forte, 1964). For instance, the diatonic set $\{0,2,4,5,7,9,11\}$ is represented by the interval vector $\langle 254361 \rangle$, as the set contains 2 semitones, 5 whole-tones, 4 minor thirds, 3 major thirds, 6 perfect fourths and 1 tritone.

Relevant relational concepts for analysis are the *set-class equivalences*, whereby two pc-sets are considered equivalent if and only if they belong to the same *set class* (defined below). Put differently, set-class spaces result from applying certain equivalence relations among all the possible pc-sets. As pointed out by Straus (1990), equivalence is not the same thing as identity, rather it is a link between musical entities that have something in common.[3] This commonality underlying the surface

[3] In this sense, a class equivalence can be conceived of as an all-or-nothing *similarity measure* between two pc-sets. Later on in this chapter, we will discuss similarity measures between different

can potentially lend unity and/or coherence to musical works (Straus, 1990, pp. 1–2). In the context of pc-sets, the number of pitch classes in a set is referred to as its *cardinality*. This provides the basis for perhaps the coarsest measure of similarity (Rahn, 1980). Despite its theoretical relevance, comparing pitch class sets on the basis of their cardinalities provides too general a notion of similarity to be of use in most analytical situations. Among the many kinds of similarity in the set-theoretical literature, three are particularly useful:[4]

1. *Interval vector equivalence* (henceforth *iv-equivalence*), which groups all the pc-sets sharing the same interval vector. There exist 197 different iv-types.
2. *Transpositional equivalence* (henceforth T_n*-equivalence*), which groups all the pc-sets related to each other by transposition. There exist 348 distinct T_n-types.
3. *Inversional and transpositional equivalence* (henceforth T_nI*-equivalence*), which groups all the pc-sets related by transposition and/or inversion. There exist 220 different T_nI-types (also referred to as T_n/T_nI-types).

The compromise between discrimination and generalization of these equivalence relations fits a wide range of descriptive needs, hence their extensive use in general-purpose music analysis. Of these relations, iv-equivalence is the most general (197 classes). It shares most of its classes with T_nI-equivalence (220 classes), with some exceptions, known as *Z-relations* (Forte, 1964), for which the same interval vector groups pitch-class sets which are not T_nI-equivalent (Lewin, 1959). The most specific of these three relations is T_n-equivalence (348 classes). T_n-equivalence satisfies two of the most basic types of perceptual similarity between pitch configurations: octave and transposition invariance. Two pc-sets related to each other by transposition only belong to the same transpositional set class, and they share a similar sonority in many musical contexts. For instance, it is clear that a pentatonic melody would sound very similar to any of its transpositions. The set-class counterpart to this, under T_n-equivalence, states that all pentatonic music, based on the same set or any transposition of the set, would sound pentatonic.

More general class equivalences have also been proposed in the set-theoretical and analysis literature, such as the *K complexes* and *Kh sub-complexes* (Forte, 1964), or *set genera* (Forte, 1988). Despite their theoretical importance, their resulting classes group together too distinct musical realities to be of practical use for our general purposes. More specific class equivalences, such as pitch (invariant to timbral transformations) or pitch-class (invariant to octave transformations), constitute the usual descriptive lexicon in score-based descriptions and MIR applications alike. However, despite their descriptive power, they cannot encode even basic analytical relations or perceptual implications, without further processing.

We thus propose iv-, T_nI- and T_n-equivalences as a convenient mid-level descriptive domain, bridging the gap between (too specific) pitches or pitch-classes and (too

---

classes, which are used to quantify the distances between pc-sets beyond mere class-belonging relations. In order to avoid misconceptions with the usage of the term ‘similarity’ in this work, the reader is encouraged to attend to the proper contextualization.

[4] See Forte (1964) for a comprehensive formalization or Straus (1990) for a pedagogical approach. A worked example is given in Sect. 4.3.2.

generic) higher-level concepts. Set-classes do not refer *explicitly* to other mid-level musical objects, such as chords or keys. However, they encode an analytically important property of such pitch collections, with the additional benefit of covering the complete set of possible pitch aggregates.

4.2.3 On Representation

Here, we use the term *representation* to mean the data structures serving as a substrate of the descriptions—that is, the final descriptors as they will be processed, whether by human analysts or by automatic algorithms. As discussed above in relation to Huovinen and Tenkanen's (2007) approach, the enormous number of overlapping segments, together with the somewhat large lexicon of classes, gives rise to problems of practical indexing and summarization of the data.

Summarization, in general, implies a loss of information. The time dimension is usually the first parameter to be sacrificed when summarizing pitch information. Some methods based on statistical averaging rarely retain usable time-related information. For instance, Huovinen and Tenkanen (2007) characterize the interval content of a given tail-segment array as a *mean interval-class vector (MICV)*, by computing the dimension-wise mean from the interval vectors of each segment belonging to the array. A more compact summary computes the mean from all the MICVs in a piece (Huovinen and Tenkanen, 2007, p. 172). This results in a single six-dimensional vector, which is an indicator of the relative *prevalence* of the different intervals in the music, but it is not straightforwardly interpretable—to a large extent, because most of the intervals are accounted for several times in the tail-segment arrays, and this number varies according to the polyphonic writing.

Another common method of summarization consists of quantifying the number of occurrences of each class from some given lexicon of classes (e.g., the number of instances of each chord type). While histograms from these *event counting* methods may provide useful indicators of the relative importance of each class, they present an inherent conceptual problem: they assume that the pitch content is to be understood in terms of *events*, as instances that can be unambiguously segmented and counted (for instance, beat-wise chords). Aside from requiring prior knowledge about how to segment the music, this approach fails to take into consideration the description of tonal contexts in a wider sense. For instance, a diatonic passage usually embeds several overlapped diatonic subsegments, so *counting* the number of diatonic instances in a piece may not be very informative. An alternative to this may be to quantify the diatonicism by *measuring* its overall temporal existence in the piece.

Here, we propose alternative summarization and quantification strategies. We avoid thinking in terms of *events*, by conceptualizing every segment as a *context*. Each segment (context) can thus be embedded in larger contexts, and can embed shorter ones. Quantification is not conceived of as the number of instances of the different classes, but as the amount of time these classes are *active* in the music.[5] Deviating from other systematic approaches, this allows for the preservation of a great deal of class-wise and temporal information through each reduction stage. An additional goal is that the various scaled representations should complement each other, in order to maximize overall informativeness when it comes to interface design (see Sect. 4.7).

4.2.3.1 Visualization

In representation design, the final *user* of the information constitutes a major factor. As humans have tremendous, innate capabilities for inferring useful information from images, visualizing information effectively is often the best way of communicating it to humans. Musical information is not an exception. When talking about music, the use of terms related to space is ubiquitous (e.g., references to the *height* of a pitch, a harmonic *attraction*, a modulation to a *distant* key, a *vertical* chord or a *horizontal* interval between notes in a melody). Even time is often conceived of as a spatial dimension, as reflected in many musical notations. Spatial representations and associations seem to be an inherent part of human thinking, presumably related to the fact that we live in a physical (three-dimensional) world with which we interact through our senses (Leman, 2008).

Many visual representations have been proposed for capturing and revealing pitch-related concepts, usually relying on mid-level musical concepts, such as chords or keys.[6] A variety of spaces, based on set-theoretical concepts, are at the core of many mathematical music theories (see Chap. 3, this volume). Tymoczko (2011), for instance, represents voice-leading relations between a given set of chord types in geometric spaces. In those spaces, it is possible to represent somewhat abstract relations as *pictorial* patterns, reinforcing the auditory phenomena with intuitive visual cues.

The analysis of complete pieces of music in these spaces is challenging, however. The time dimension in these representations can be conveniently introduced by movement-in-space metaphors.[7] When it comes to representing long passages, these methods present visual problems relating to leaving traces of the past, compromising any practical temporal indexing of the music. For instance, it may be difficult to distinguish among different instances of the same recurring pattern, as they can appear superimposed on each other in the space. However, it is often of interest to analyse long segments of music, in order to observe far-reaching relationships over time. The temporal dimension, thus, arises as a convenient (explicit) dimension to be considered in the design of visual representations of music (Martorell, 2013).

[5] A large number of event instances does not imply prevalence, as the events themselves can be very short in duration, e.g., in *tremolo* passages.

[6] See Lerdahl (2001, pp. 42–47) for a review.

[7] In Tymoczko's chordal spaces, this is more than a metaphor: while a *sequence of states* can represent chord progressions, the representation of their voice-leading relations needs the actual (continuous) *paths* followed by each individual voice (Tymoczko, 2011, pp. 43–44).

4.2.3.2 Temporal Multi-scale Representation Methods

Sapp (2005) proposes a systematic, user-oriented approach to tonal analysis and representation. After a comprehensive multi-scale (temporal multiresolution) segmentation of the music, the tonal centre is estimated for each segment. The different estimations are coded through a colourmap, and all the segments are plotted as a two-dimensional image, called a *keyscape*, whose dimensions represent time on the horizontal axis and, on the vertical axis, time-scale (i.e., time-scale of observation, or window size, corresponding to the durations of the analysis segments). The method, which gives access to hierarchical descriptions of tonality, has been adapted by Martorell (2013) to form the basis of a general framework for *tonal context* analysis. This method can operate in both the symbolic and audio domains, and extends the concept of keyscape to tonal systems other than the major–minor system.

4.3 Hierarchical Multi-scale Set-Class Analysis

The method proposed here takes as its starting point the concept of a keyscape and extends its systematization (with respect to segmentation and representation) to the description stage. For that, the key estimation is substituted with a set-class description. Figure 4.1 sketches the general framework proposed by Martorell (2013, p. 26), here adapted for set-class analysis. A detailed description of each stage (segmentation, description and representation) follows.[8]

4.3.1 Segmentation

The input to the system is a sequence of MIDI events, which can be of any rhythmic or polyphonic complexity. Segmentation is implemented by two different algorithms: a) an approximate technique, non-comprehensive but practical for interacting with the data; and b) a fully systematic method, which exhausts all the segmentation possibilities. Figure 4.2 illustrates both cases.

The approximate method consists of applying many overlapping sliding windows, each of them scanning the music at a different time-scale (duration of observation). The minimum window size and the number of time-scales are user parameters, and can be fine tuned as a trade-off between resolution and computational cost. The actual time-scales are defined by applying a logarithmic law covering the range between the minimum window size and the whole duration of the piece.[9] In order to provide a regular grid for indexing, visualization and interfacing purposes, the same hop size

[8] The method is explained following the same blocks of Sect. 4.2: segmentation, description and representation. For computational convenience, the actual implementation differs slightly, as depicted in Fig. 4.1.

[9] Motivated by the fact that larger time-scales usually yield coarser changing information.

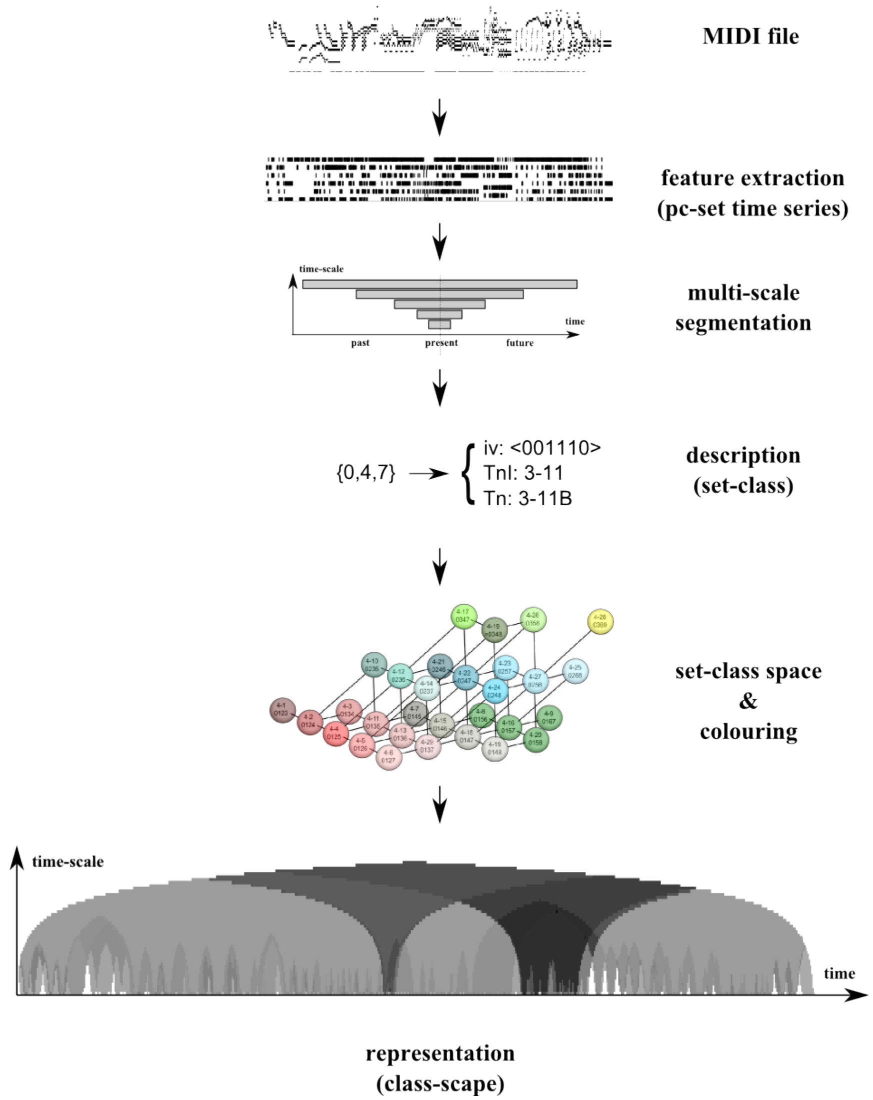

Fig. 4.1 General framework: block diagram

is applied for all the time-scales. Each segment is thus indexed by its centre location (time) and its duration (time-scale).

The fully systematic method is required for the quantitative descriptors in which completeness of representation is necessary (see Sect. 4.4.2.2). The algorithm first finds every change in the pc-set content over time, regardless of whether the change is produced by the onset or the offset of a note. The piece is then segmented according to every pairwise combination among these boundaries.

4.3.2 Description

Denoting pitch-classes by the ordinal convention (C=0, C♯=1, ..., B=11), each segment is analysed as follows. Let $b_i = 1$ if the pitch-class i is contained (totally or partially) in the segment, or 0 otherwise. The pc-set of a segment is encoded as an integer $p = \sum_{i=0}^{11} b_i \cdot 2^{11-i} \in [0,4095]$. This integer serves as an index for a precomputed table of set classes,[10] including the iv-, T_nI- and T_n-equivalences (defined in Sect. 4.2.2.1). For systematization completeness, the three class spaces are extended to include the so-called *trivial forms*.[11] With this, the total number of interval vectors rises to 200, while the T_nI- and T_n-equivalence classes sum to 223 and 351 categories respectively. In this work, we use Forte's cardinality-ordinal convention to name the classes, as well as the usual A/B suffix for referring to the prime/inverted forms under T_n-equivalence. We also follow the conventional notation to name the Z-related classes, by inserting a 'Z' between the hyphen and the ordinal.

As an example, a segment containing the pitches {G5,C3,E4,C4} is mapped to the pc-set {0,4,7} and coded as $p = 2192$ (100010010000 in binary). The precomputed table is indexed by p, resulting in the interval vector $\langle 001110 \rangle$ (iv-equivalence, grouping all the sets containing exactly 1 minor third, 1 major third, and 1 fourth), the class 3-11 (T_nI-equivalence, grouping all the major and minor trichords), and the class 3-11B (T_n-equivalence, grouping all the major trichords). The discrimination between major and minor trichords is thus possible under T_n-equivalence (no major trichord can be transformed into any minor trichord by transposition only), but not under iv- or T_nI-equivalences (any major trichord can be transformed into any minor trichord by transposition and inversion).

The beginning of Bach's chorale *Christus, der ist mein Leben* (BWV 281), shown in Fig. 4.2(a), illustrates the segmentation and description stages. Figure 4.2(b) (bottom) depicts the pitch-class set content of the excerpt over time. Figure 4.2(b) (top) sketches the approximate segmentation method at three different time scales, using window sizes of 1, 3 and 7 quarter notes, respectively. The same hop size (1 quarter note) is used in all three cases, providing a regular indexing grid. Three sample points (in light grey) show the relation between the indexing and the actual segments.

The set-class description of the corresponding segments under T_n-equivalence is also provided. For example, the lowest sample segment contains the pitch classes {C,F,A}={0,5,9} which belongs to the class 3-11B under T_n-equivalence: transposing (mod_{12}) the set {0,5,9} by 7 semitones up gives {0+7,5+7,9+7}={0,4,7},[12] which is

[10] As formalized by Forte (1964). See http://agustin-martorell.weebly.com/set-class-analysis.html for a comprehensive table of set classes.

[11] The *null set* and *single pitches* (cardinalities 0 and 1, containing no intervals), the *undecachords* (cardinality 11) and the *universal set* (cardinality 12, also referred to as the *aggregate*). These classes were not included in Forte's formalization. However, any truly systematic formalization of the set-class space should include them, as segments expressing these sets do occur in real music.

[12] Pitch-class sets are, by definition, *unordered* sets, meaning that the ordering of the elements is not of importance.

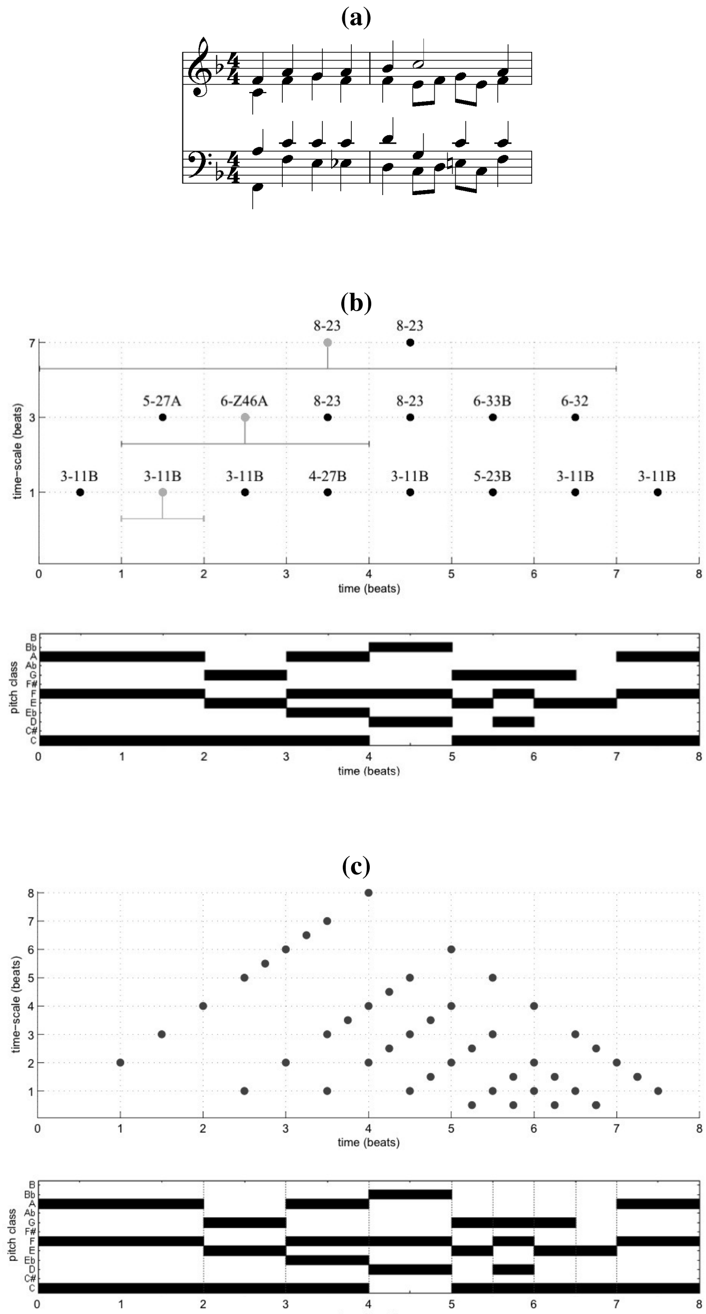

Fig. 4.2 J. S. Bach's *Christus, der ist mein Leben* (BWV 281) (excerpt). (a) Score. (b) Approximate segmentation and description. (c) Fully systematic segmentation

the standard form for all the major trichords under T_n-equivalence, and named 3-11B in Forte's convention.

This particular segmentation is only approximate, as it does not capture the pitch-class-set content of every different segment. For that, the fully systematic segmentation approach is required. The music is first segmented by finding every change in the pitch-class set (vertical) content over time, as in Fig. 4.2(c) (bottom). Then, a segment is defined for every pairwise combination of boundaries. As depicted in Fig. 4.2(c) (top), this exhausts every possible (different) pitch-class aggregate in the piece. In general, as for this example, this results in non-regular indexing grids.

4.3.3 Representation

The representation stage relies upon the general framework depicted in Fig. 4.1. The substitution of the inter-key space used by Martorell (2013, p. 26) by a set-class space is followed by a convenient colouring of the set-class space. The set-class content (description) of each segment is thus mapped to position in the class space, and associated with a colour. All the segments are then represented as coloured pixels in a two-dimensional plot, called a *class-scape*, and indexed by time (x-axis) and time-scale (y-axis). The designs of the class-space and the colouring strategy are application specific, depending on the pitch-related properties of interest. Sections 4.4.1 and 4.4.1.1 describe the method for two application scenarios.

4.4 Basic Set-Class Multi-scale Descriptors

This section presents the basic set of data structures (henceforth, *descriptors*) obtained by the proposed method. Each descriptor is explained and interpreted in the context of analytical examples.

4.4.1 The Class-Scape

A simple but useful task is to localize all the segments in the music belonging to a given class. This is illustrated in Fig. 4.3, where the main features of the multi-scale representation are also introduced. The top figure shows the class-scape computed for Debussy's *Voiles*,[13] filtered by the class 6-35 (pixels in black), which corresponds to the predominant hexatonic (whole-tone) scale in the composition. As a visual

[13] The example has been chosen for illustrative purposes. It has clearly distinctive sonorities and is 'easy' to listen to, however, it is barely describable using conventional tonal description methods. As with any music not based on the major–minor paradigm, its description in terms of standard (non-systematic) dictionaries of chords or keys is nonsensical.

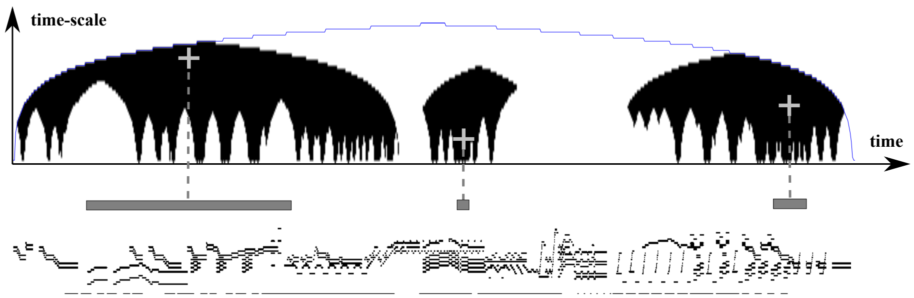

Fig. 4.3 Debussy's *Voiles*. Class-scape filtered by 6-35, piano roll, and 3 sample segments

reference, a thin blue line delineates the boundary in time and time-scale of the complete, non-filtered, class-scape information. At the bottom of Fig. 4.3 is shown an aligned piano-roll representation of the score for visual indexing of the composition, as used in our interactive analysis tool (see Sect. 4.7).

Each pixel in the class-scape, visible or not after the class filtering, represents a unique segment of music. Its x-coordinate corresponds to the temporal position of the segment's centre, and its y-coordinate represents its duration in a logarithmic scale. The higher a pixel is in the class-scape, the longer the duration of the represented segment. Three sample points (+ signs) and their corresponding segments are sketched as an example. In the figure, three large-scale structural hexatonic contexts are clearly revealed.

4.4.1.1 Multi-class Representation and Inter-class Distances

An alternative representation, allowing for the inspection of all segments and classes simultaneously, consists of assigning colours to classes. Given the relatively large number of classes, an absolute mapping of classes to colours is unlikely to be informative in general, so a relative solution is adopted. The method is a set-class adaptation of the concept of 'distance-scape', introduced by Martorell (2013, pp. 48–49). A distance-scape is a keyscape, in which each pixel (segment) is coloured according to its distance in pitch-space to a chosen tonal centre (e.g., the tonic of the home key of the piece).

The adaptation of the method to our set-class domain requires the definition of a systematic inter-class measure, able to relate every possible segment with any chosen reference. Among the many set-class distances proposed in the literature, a number of them can handle any pair of classes.[14] Lewin's *REL* (Lewin, 1979), Rahn's *ATMEMB* (Rahn, 1979), and Castrén's *RECREL* (Castrén, 1994) are among

[14] Most measures require pairwise cardinality equality. Some very informative measures, such as Tymoczko's voice-leading spaces (Tymoczko, 2011), could operate with different cardinalities. However, the class equivalences for capturing such sophistications are more specific than our set classes (iv, T_n and T_nI), and they even require the use of *multisets* (pc-sets *with* repetitions).

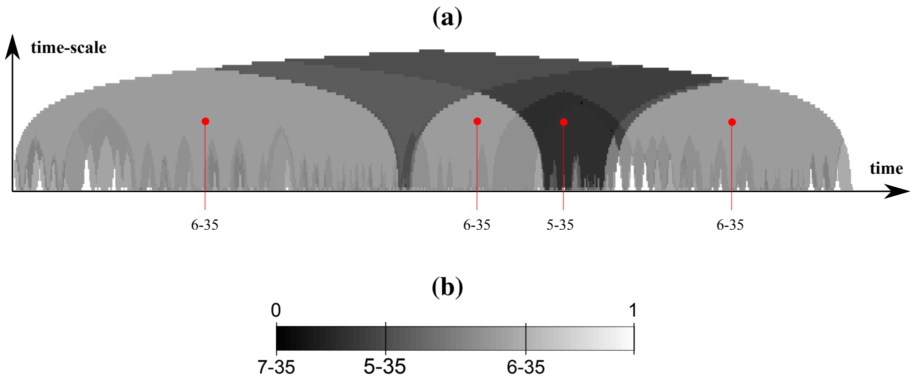

Fig. 4.4 Debussy's *Voiles*. (a) Class-scape relative (*REL*) to 7-35. (b) *REL* distances from 7-35

the so-called *total measures*. They are based on vector analysis, considering the complete subset content of the classes under comparison, and exhausting all the possible pc-set inclusion relations. For systematization completeness, these measures are adapted in our method, in order to accommodate the trivial forms.[15]

Each pixel in the class-scape is coloured according to the distance between the class it represents and any chosen reference class. Figure 4.4(a) shows the class-scape computed for Debussy's *Voiles*, in which the diatonic class 7-35 has been chosen as the reference class,[16] and *REL* as the inter-class measure. This piece does not have a single diatonic segment, so it is clear that the basic all-or-nothing filtering would result in a completely white image. We denote by *REL*(7-35,*) the *REL* distance from any class (asterisk) to the reference class 7-35, and encode its value with a greyscale (from black = 0 to white = 1), as depicted in Fig. 4.4(b). This way, every segment of the music (every pixel in the class-scape) is represented according to its *REL*-closeness to 7-35.

In Fig. 4.4, it is straightforward to visualize the darkest areas in the class-scape, corresponding to a brief pentatonic (5-35) passage, and its closest vicinity. It is also clear that the large whole-tone (6-35) passages depicted in Fig. 4.3, as well as their component subsets, are represented as being far from 7-35. Figure 4.4 clearly reveals the overall structure, while evidencing the radical analytical and perceptual differences between symmetric (whole-tone) and asymmetric (diatonic-based, here the pentatonic subset) scalar contexts.[17] Incidentally, there exist some

---

[15] See Harley (2014) for a comprehensive survey of set-class measures, formalized according to their suitability for tonal analysis. Harley also discusses the accommodation of the trivial forms.

[16] The diatonic class has been chosen for illustration purposes, as a common and well-known set.

[17] By *symmetric*, in this work we refer to maximally even sets (Clough and Douthett, 1991) featuring a reduced number of distinct transpositions, often known as Messiaen's *modes of limited transposition*. These sets present highly symmetric distributions in the chromatic circle. For instance, the whole-tone-scale set (6-35) has 2 different transpositions and 6 symmetry axes in the chromatic circle, in contrast with the anhemitonic pentatonic set (5-35), which features 12 distinct transpositions and 1 symmetry axis.

isolated segments (just two pixels in the class-scape) belonging to the scalar formation set 6-33B,[18] which are the closest ones to 7-35 in the composition.

This points to an interpretative aspect of the class-scapes: the analytical relevance of what is shown is often related to the accumulation of evidence in time and time-scale. A significantly large area of the class-scape showing the same material is probably representative of a section of music in which such evidence is of analytical and/or perceptual relevance, in the sense that it accounts for a passage in which the inspected properties are stable (i.e., a stable context). Smaller patches, even single isolated pixels, capture the class content of their corresponding segments as well. However, they could just be pitch-aggregate by-products lacking analytical interest, such as the spurious 6-33B in *Voiles*. Human cognitive (visual) abilities play an important role here in focusing the viewer's attention on the areas where homogeneous evidence accumulates, an important feature of multi-scale representations for assisting pattern recognition tasks.[19] On the other hand, localized spots of residual evidence may be worth closer inspection, and so interaction with the class-scape is facilitated.

A relevant benefit of pc-set-based spaces, as opposed to continuous ones,[20] is that music can be analysed in terms of different class systems at no extra computational cost. Being finite and discrete spaces (4096 classes at most for the TET system), the whole equivalence systems, including their inner metrics, can be precomputed. The mapping from pc-sets to set-classes, as well as the distances between any pair of music segments, can thus be implemented by table indexing. Once the pc-set of each possible segment has been computed (which constitutes the actual bottleneck of the method), the rest of the process is inexpensive, and multiple *set-class lenses* can be changed in real time, allowing fast interactive explorations of the massive data. This alleviates part of Forte's concerns about the unfeasibility of systematic segmentation and filtering decisions (Sect. 4.2.1). This feature, along with a variety of filtering options for visual exploration, can be tested with our proof-of-concept set-class analysis tool (see Sect. 4.7).

4.4.2 Piecewise Summarization: Class-Matrix and Class-Vector

The class-scape provides a comprehensive bird's-eye view of the piece, well adapted for visual exploration. However, other applications would require more compact and manageable, yet informative, data structures. As discussed in Sect. 4.2.3, it is of interest to preserve the class integrity and as much temporal information as possible. In this subsection, we present two such dimensional reductions, derived from the class-scape: the *class-matrix* and the *class-vector*.

---

[18] The class 6-33B results from removing the third degree from a major diatonic set.

[19] In his classic work on visualization, Tufte (2001, pp. 16–20) discusses similar cognitive implications to demonstrate the usefulness of his massive 'data maps' as instruments for reasoning about quantitative information.

[20] Such as those accommodating the so-called *chroma features*, which are finite but continuous.

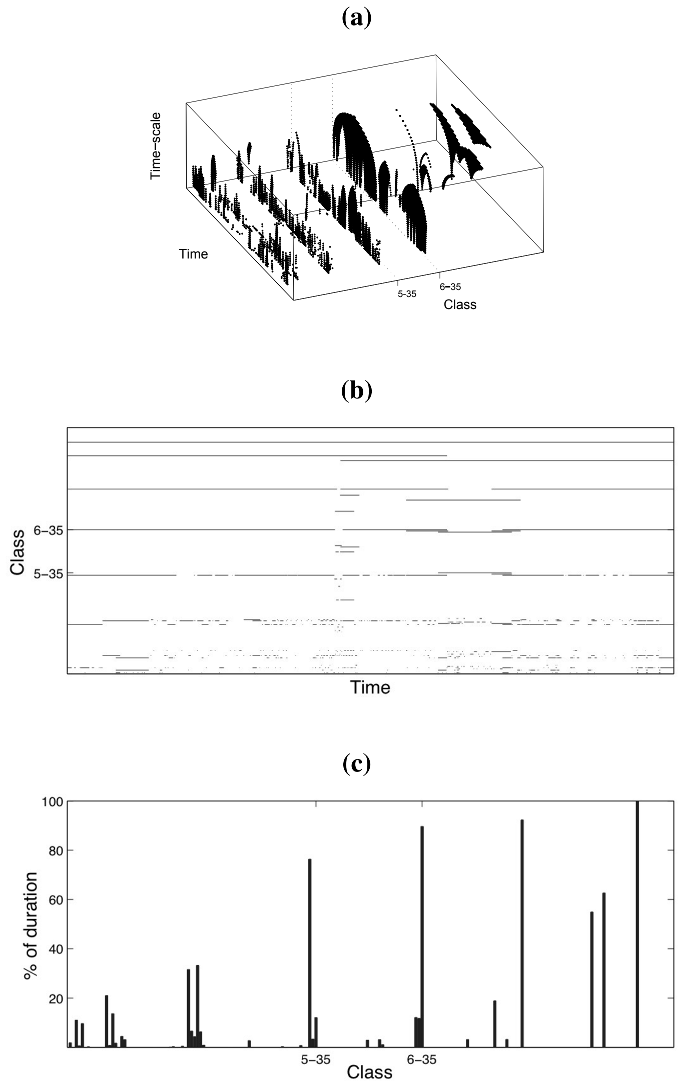

Fig. 4.5 Debussy's *Voiles*. (a) Class-scape as a sparse 3-D matrix. (b) Class-matrix. (c) Class-vector

4.4.2.1 The Class-Matrix

As a data structure, the 2-dimensional class-scape presented above can be thought of as a sparse 3-dimensional binary matrix, in which each segment is indexed by the temporal location of its centre (x-axis), its duration (y-axis), and its class content (z-axis). In Fig. 4.5(a), such an arrangement is shown for Debussy's *Voiles*, with the classes 5-35 (pentatonic) and 6-35 (whole-tone) annotated as reference. The first reduction process consists of projecting this information into the time–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 a *class-matrix*, is a multi-dimensional time series of a special kind, which requires a proper interpretation. It represents the N classes as N rows, arranged from bottom to top according to Forte's cardinality-ordinal convention (Forte, 1964). An *activated* position (t,c) in the class matrix (black pixels) means that at least one segment belonging to the class c includes the time position t. A position on the time axis of the class-matrix is not associated with a unique segment, but with all the segments containing that time point. The class-matrix, thus, summarizes information from all the time-scales simultaneously. Unlike the approach of Huovinen and Tenkanen (2007), which also accounts for several time-scales in a single tail-segment array, the class-matrix preserves a strict separation of classes. This feature can be exploited in more subtle descriptive tasks, as discussed in Sect. 4.5.2.

Figure 4.5(b) shows the class-matrix for Debussy's *Voiles* under iv-equivalence, revealing the prelude's economy of sonorities, characteristic of works mostly based upon symmetric scalar formations. Even with the loss of information, it helps to visualize the contribution of each individual class, since colouring is no longer required. The individual duration of each frame from the initial segmentation may not be appreciated in the class-matrix, since overlapped segments belonging to the same class 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 one to capture relational details of certain structural relevance, by analysing the inclusion relations down the hierarchy of classes, as will be discussed below in Sect. 4.5.2.

4.4.2.2 The Class-Vector

An even more compact representation provides a means for quantification. For each row in the class-matrix, the activated positions are accumulated and expressed as a percentage of the total duration of the piece. This data structure, henceforth *class-vector*, has a dimension equal to the number of classes, and quantifies the temporal presence of each possible class in a piece. Figure 4.5(c) shows the class-vector computed for Debussy's *Voiles*. The potential of class-vectors for comparing different pieces of music, however, raises the problem of resolution. The segmentation used so far was convenient for visualization and fast interaction with the data, but it is clear that the same segmentation parameters may not resolve equally the class content of different pieces, compromising any quantitative comparison. However, class-vectors can be computed with absolute precision, by substituting the multi-scale policy by a truly systematic approach, which captures all the possible different segments, as described in Sect. 4.3.1 (see also the worked example in Sect. 4.3.2). Such exhaustive computation is thus performed for any application involving quantification.[21]

[21] Class-matrices are also suitable for precise quantitative applications. Our method relies upon lists of temporal segments for each class.

4.5 Mining Class-Matrices

This section presents two applications of the information conveyed by the class-matrices.

4.5.1 Case Study: Structural Analysis

Self-similarity matrices (SSMs) are simple tools commonly used for finding recurrences in time series (Foote, 1999). In music-related applications, the typical inputs to an SSM are spectral or chroma feature time series. Some of the SSM-based methods can handle different time-scales, and some of the chroma methods allow for transpositional invariance (Müller, 2007). These functionalities are usually implemented at the SSM computation stage,[22] or as a post-processing step. In the class-matrices, both the equivalence mappings (including their inherent hierarchies) and the multi-scale nature of the information are *already* embedded in the feature time-series, so a plain SSM can be used for finding sophisticated recurrences. For instance, a passage comprised of a chord sequence can be recognized as similar to a restatement of the passage with different arpeggiations and/or inversions of the chord intervals (e.g., from major to minor triads). A *vertical* chord and its arpeggiated version may not be recognized as very similar at the lowest cardinalities, but their common T_nI-sonority will certainly emerge at their corresponding time-scales. Moreover, any sonority containing the chords (i.e., *supersets*) will also be captured at their proper time-scales, climbing up the hierarchy until we reach the whole-piece segment, everything indexed by a common temporal axis. A quantification of similarity between variations may thus be possible at the level of embedded sonorities.

We now consider the application of the above method to the analysis of large-scale recurrence in Webern's *Variations for piano* op.27/I, as an example of a challenging musical work, barely describable by standard tonal features (such as chroma, chords or keys). This dodecaphonic (early serialism) piece presents an A–B–A' structure, built upon several instantiations of the main twelve-tone row under different transformations, involving transpositions, inversions and distinct harmonizations. Figure 4.6 (top) depicts the class-scape of the piece, filtered by the prominent hexachordal iv-sonority $\langle 332232 \rangle$, and Fig. 4.6 (bottom) shows the well-known structure of the row instantiations (extensively analysed in the literature), annotated according to Cook (1987).[23] The large-scale A–B–A' structure can be observed with the naked eye in Fig. 4.6 (top).

[22] For instance, transpositional invariance is often achieved by computing twelve SSMs, accounting for each of the twelve ring-shifted versions of the chroma features.

[23] The standard terminology in row analysis is followed: *P* stands for the prime form of a row, *R* for a row's retrograde, *I* for its inversion, and the ordinal number indicates the transposition. For instance, *RI*-6 refers to a row instantiation, which appears transposed 6 semitones above the original row, with all its intervals inverted, and retrograded. The double labelling stands for the simultaneous occurrence of each row instantiation with its own retrograde version throughout the piece. For a

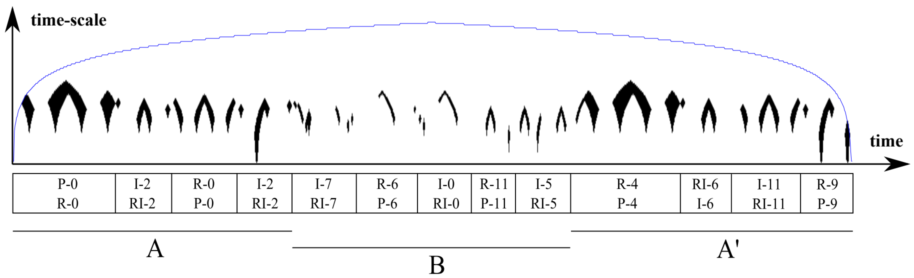

Fig. 4.6 Webern's op.27/I. Top: class-scape filtered by $\langle 332232 \rangle$; Bottom: structure

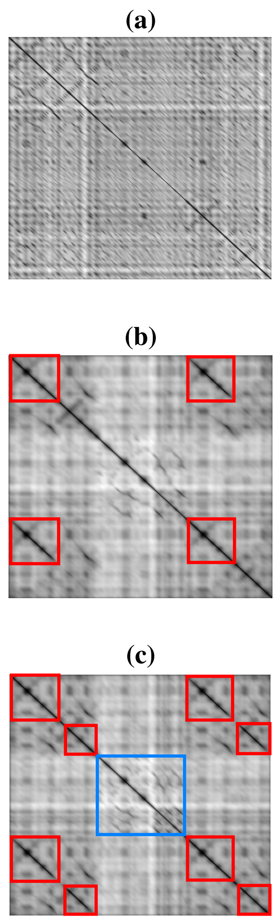

Fig. 4.7 SSMs for Webern's op.27/I. (a) pc-equivalence. (b) T_n-equivalence. (c) T_nI-equivalence

Finding this large-scale A–B–A' form thus requires us to look for transformed repetitions. This can be done with an SSM, fed with an input time-series that is invariant to those transformations. Figure 4.7 depicts the output of a plain SSM, computed from three different inputs: a) the pc-set time series;[24] b) the class-matrix under T_n; and c) the class-matrix under T_nI. The pc-equivalence does not capture any large-scale recurrence. The restatement of the first two phrases in A is captured by the T_n-equivalence, as these phrases are mainly related by transposition in A'. Finally, the T_nI-equivalence reveals the complete recapitulation, including the last two phrases of A, which are restated in A' in both transposed and inverted transformations.

It is worth noting that the method is agnostic with respect to the general sonority, the ubiquitous $\langle 332232\rangle$ that reveals the large-scale structure (as in Fig. 4.6). Moreover, it is not necessary for the recurrences to be *exact* transformations (e.g., exact transpositions). This is because, by using class-matrices as input to the SSM, similarity is conceived of as a construction down the subclass (hierarchical) inclusion relations.[25] This allows for the discrimination of the B section, built upon the same iv-hexachordal sonority and the same kind of row instantiations as A and A', but presented in distinct harmonizations.

4.5.2 Case Study: Hierarchical Subclass Analysis

Class-matrices can also be exploited to describe the subclass content *under* any class, revealing the building blocks of particular class instantiations. The next case study analyses two corpora in terms of pure diatonicism, by characterizing the subset content of only the diatonic segments.

The computation process for the Agnus Dei from Victoria's *Ascendens Christus* mass is depicted in Fig. 4.8. The diatonic-related subclass content is isolated by considering only the information *below* the activated 7-35 positions in the class-matrix, as shown in Fig. 4.8(a). The result of this process is a *subclass-matrix*, in Fig. 4.8(b), with a number of rows equal to the number of subset classes (up to cardinality 6 here). 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 Fig. 4.8(c), quantifies the total subclass content contributing to the reference class, describing what (and how much of it) the particular diatonicism consists of. The most prominent subclasses at 3-11, 5-27 and 6-Z25 stand out in the subclass-vector, which also reveals other common scalar formation classes of cardinalities 4 to 6.

The method can be extended for characterizing a corpus, according to some specific sonorities, by taking a class-wise average across the subclass-vectors extracted

detailed analytical discussion of the piece, in terms of this hexachordal sonority under different equivalences, see Martorell and Gómez (2015).

[24] In a sense, the discrete *equivalent* of the chroma features.

[25] The SSM compares multi-dimensional (*vertical*) views of the class-matrix, each of them accounting for all the embedded contexts around its corresponding temporal location.

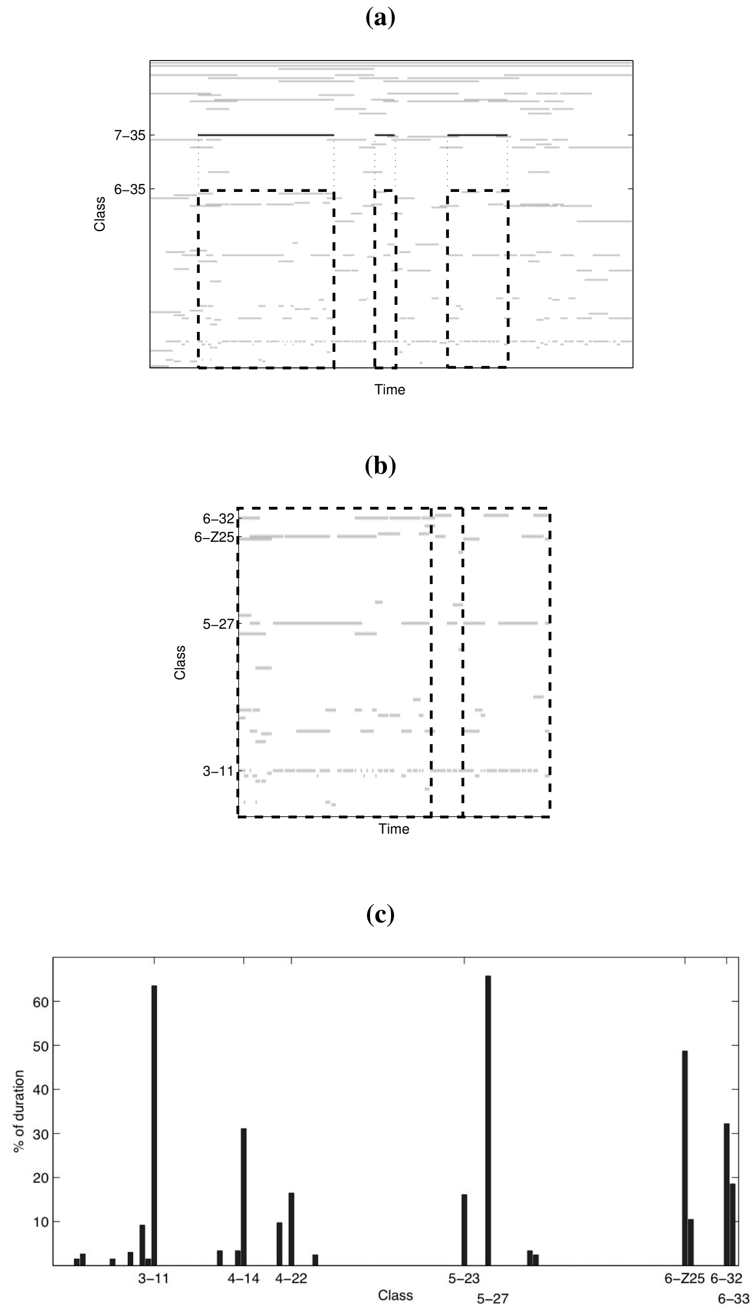

Fig. 4.8 Victoria's *Ascendens Christus*, Agnus Dei. (a) Filtering under 7-35. (b) Subclass-matrix. (c) Subclass-vector

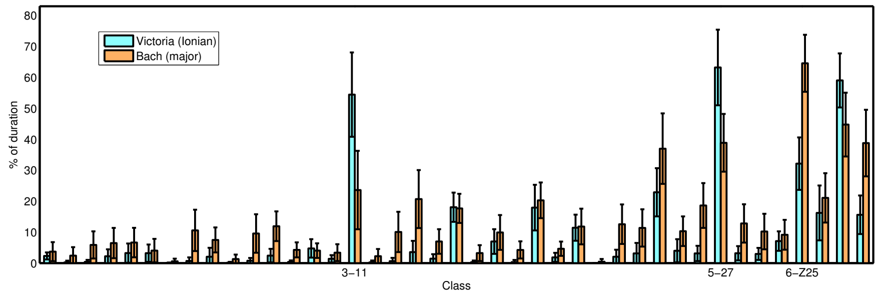

Fig. 4.9 Diatonicism in Victoria and Bach. Mean subclass-vectors under 7-35

from all the pieces. Figure 4.9 depicts the mean subclass-vector under 7-35 computed for two contrasting corpora: a) Victoria's parody masses in Ionian mode;[26] and b) the preludes and fugues in major mode from Bach's Well-Tempered Clavier. The selection of the corpora is based on the close relationship between the major and the Ionian modes, the comparable number of voices and movements in both corpora, and the (loose) assumption of homogeneity in the usage of contrapuntal resources by each composer. For clarity of comparison including standard deviations, only the actual subclasses of 7-35 are represented. The prominent use of major and minor triads (3-11) in Victoria's diatonicism relative to Bach's stands out. Similarly predominant is the class 5-27, a far more recurrent cadential resource in Victoria.[27] On the other hand, the Locrian hexachord 6-Z25 is far more prevalent in Bach. Apart from its instantiations as perfect cadences[28] in both corpora, 6-Z25 appears consistently in many motivic progressions in Bach, which are not idiomatic in Victoria's contrapuntal writing.[29]

4.6 Mining Class-Vectors

This section elaborates upon the information conveyed by the class-vectors. As mentioned, a class-vector quantifies the temporal presence of every class sonority, relative to the duration of the piece. Finding specific sonorities in large datasets can be combined with the extraction of the actual segments from the MIDI files. This can be exploited in various applications, ranging from the analysis of corpora to music

[26] Including *Alma Redemptoris Mater*, *Ave Regina Caelorum*, *Laetatus Sum*, *Pro Victoria*, *Quam Pulchri Sunt*, and *Trahe Me Post Te*. See Rive (1969) for a modal classification.

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

[28] The class 6-Z25 results from the combination of a major triad and its dominant seventh chord.

[29] By interfacing class-vectors with class-scapes, the content of particular class instantiations can be easily explored and listened to with our interactive analysis tool (see Sect. 4.7).

education. For this work, a dataset of class-vectors was computed from more than 16000 MIDI tracks.[30]

4.6.1 Case Study: Query by Set-Class

A simple but useful application is querying a database of pieces for a given set-class. It can be used for finding pieces with a relevant presence of any chordal or scalar sonority, such as *exotic* scales. This kind of task is problematic using standard tonal features, which (usually) support neither contextual descriptions nor tonal systems other than the common-practice harmonic idioms. Table 4.1 shows 10 retrieved pieces with a notable presence (relative duration) of the sonority 7-22, usually referred to as the Hungarian minor set-class.[31] Both monophonic and polyphonic pieces are retrieved, ranging over different styles and historic periods, as the unique requisite for capturing a given sonority was its existence as a temporal segment. The instantiations of 7-22 in the retrieved pieces are also varied, ranging from passages using an actual Hungarian minor modality, to short but frequent (colouring) deviations from the minor mode, including some non-relevant pitch concatenation by-products.

Table 4.1 Retrieved pieces: 7-22

Retrieved piece	7-22 (%)
Scriabin - Prelude op.33 n.3	68.61
Busoni - 6 etudes op.16 n.4	63.22
Essen - 6478	62.50
Liszt - Nuages gris	42.41
Essen - 531	36.67
Scriabin - Prelude op.51 n.2	31.74
Lully - Persee act-iv-scene-iv-28	29.73
Alkan - Esquisses op.63 n.19	28.87
Satie - Gnossienne n.1	28.15
Scriabin - Mazurka op.3 n.9	24.61

[30] Including works by: Albeniz, Albinoni, Alkan, Bach, Bartók, Beethoven, Berlioz, Bizet, Brahms, Bruckner, Busoni, Buxtehude, Byrd, Chopin, Clementi, Corelli, Couperin, Debussy, Dowland, Dufay, Dvořák, Fauré, Franck, Frescobaldi, Gesualdo, Guerrero, Handel, Haydn, Josquin, Lasso, Liszt, Lully, Mahler, Mendelssohn, Messiaen, Monteverdi, Morales, Mozart, Mussorgsky, Pachelbel, Paganini, Palestrina, Prokofiev, Rachmaninoff, Rameau, Ravel, Reger, Saint-Saëns, Satie, Scarlatti, Schoenberg, Schubert, Schütz, Schumann, Scriabin, Shostakovich, Soler, Stravinsky, Tchaikovsky, Telemann, Victoria and Vivaldi. It also includes anonymous medieval pieces, church hymns, and the Essen folksong collection.

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

4.6.2 Case Study: Query by Combined Set-Classes

The strict separation of classes in the class-vectors allows for the exploration of any class combination, whether common or unusual. For instance, the first movement of Stravinsky's *Symphony of Psalms* is retrieved by querying for music containing substantial diatonic (7-35) and octatonic (8-28) material, certainly an uncommon musical combination. The class-vector also reveals the balance between both sonorities, as 30.18 % and 29.25 % of the piece duration, respectively.

As discussed in Sect. 4.5.2, the class-matrices allow for the hierarchical analysis of specific sonorities. The class-vectors, on the other hand, summarize the information in a way that does not, in general, elucidate the subclass content under a given class. However, if the queried sonorities have a substantial presence (or absence) in the piece, the class-vectors alone can often account for some hierarchical evidence. One can, for instance, query the dataset for pieces with a notable presence of the so-called *suspended* trichord (3-9),[32] constrained to cases of mostly diatonic contexts (7-35). This situation is likely to be found in folk songs, medieval melodies, early counterpoint, or works composed as reminiscent of them. It is worth noting that the 3-9 instantiations may appear as a result of a variety of pitch aggregates, such as in monophonic melodies, as a combination of melody and tonic-dominant drones, as a combination of tenor and parallel motion in fifths and octaves, and as actual (voiced) suspensions.

Similarly, as non-existing sonorities may also reveal important characteristics of music, the dataset can be queried for combinations of present and absent classes. For instance, the sonority of purely diatonic (7-35) pieces depends on whether they contain major or minor trichords (3-11) or not. If we constrain the query to polyphonic music, the retrieved pieces in the latter case (diatonic, not triadic) belong mostly to early polyphony, prior to the establishment of the triad as a common sonority.

These results point to interesting applications related to music similarity, such as music education or recommendation systems. Music similarity is, to a great extent, a human construct, as it depends upon cultural factors and musical background. Of particular interest is the possibility of retrieving pieces sharing relevant tonal-related properties, but pertaining to different styles, composers, or historical periods. Music discovery or recommendation systems could thus serve as music appreciation or ear-training tools, by suggesting pieces across standard taxonomies (e.g., composer or style) subjected to musically relevant similarity criteria. This would help to reinforce listening skills for specific sonorities, while enriching the contextual situations in which they appear in real music. Recommender systems able to *explain* the basis of their recommendations, using some standard musicological lexicon (e.g., as provided by the set-classes), would be a particularly useful application for such purposes.

[32] A major or minor trichord with the third degree substituted by the fourth or the second.

4.6.3 On Dimensionality and Informativeness

As pointed out in Sect. 4.2.2, the compromise between the dimensionality of the feature space and the informativeness of a description is a relevant factor in feature design. The class content of a piece, as described by its class-vector, has 200, 223 or 351 dimensions, depending on the chosen equivalence (iv, T_nI or T_n). Compared with other tonal feature spaces, these dimensions may seem quite large. However, the benefits of class vectors are the systematicity, specificity and precision of the description. Several relevant differences with respect to other tonal-related features are to be noticed. A single class-vector, computed after a fully systematic segmentation,

1. accounts for every distinct segment in a piece, regardless of its time position or duration;
2. accounts for every possible sonority in the set-class space, which is *complete*;
3. provides an objective and precise description of the *set-class sonority*;
4. provides a description in music-theoretical terms, readable and interpretable by humans;
5. provides an objective quantification of every possible sonority in terms of time;
6. provides a content-based, model-free, description of the piece; and,
7. in some cases, provides an approximation to the hierarchical inclusion relations.

In contrast, the most common tonal piecewise and labelwise feature (global key estimation) provides

1. a single label for the whole piece;
2. 24 different labels, but only two different sonorities (major and minor);
3. an *estimation* of the key, which is often misleading;
4. a description in music-theoretical terms, but of little compositional relevance;
5. quantification (sometimes) by key strength, but with no temporal information at all;
6. a description based on specific and biased models (e.g., profiling methods); and
7. no access to the hierarchical relations of the piece's tonality.

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

4.7 Interfacing Representations

As pointed out in Sect. 4.2.3, the design of the proposed descriptors aims to maximize the *overall* informativeness of the method. This is proposed through interaction with the user, in our proof-of-concept set-class analysis tool. In this section, we briefly describe the basic functionalities of the tool.[33]

Figure 4.10 depicts a screenshot of the interface, in which the basic descriptors (class-scape, class-matrix and class-vector) are readily accessible. Debussy's *Voiles* serves again as an analytical example. Both the class-scape and the class-vector can be navigated with a cursor. Any class can be chosen from the class-vector, establishing a reference sonority which affects the visualization of the class-scape. The class-scape has two visualization modalities (*single-class* and *multi-class*), corresponding to those described in Sects. 4.4.1 and 4.4.1.1, both controlled by the reference class chosen from the class-vector. The equivalence system as a whole can be changed (iv, T_n and T_nI are supported), as well as the inter-class measure (*REL*, *RECREL* and *ATMEMB* are supported). The class-scape can also be filtered by any set of cardinalities. Any segment can be selected from the class-scape, in order to inspect its properties, and it can be played for auditory testing. All the information is interactively updated according to the user actions, allowing a fast exploration of different filtering possibilities. This facilitates the *analysis loop*, understood as the process whereby an analyst refines the observation parameters according to his or her findings and intuitions.

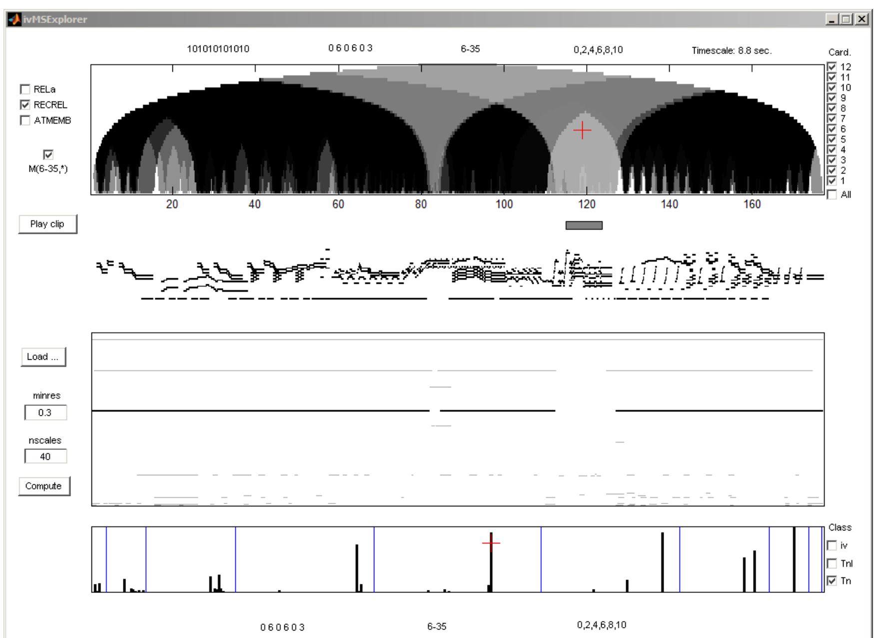

Fig. 4.10 Interfacing set-class analysis

[33] See http://agustin-martorell.weebly.com/set-class-analysis.html for details.

4.8 Conclusions

This work presents a systematic approach to segmentation, description and representation of pieces of music. The methodology is designed for analysing the musical surface in terms of embedded contexts, as an alternative to the usual event-based approaches. A set-class description domain provides a mid-level lexicon, which encodes useful tonal properties of pitch aggregates of any kind, allowing the description of any chordal or scalar sonority. The class-scapes provide a bird's-eye view of a complete piece of music, furnishing information about the class content of every possible segment, whether in absolute terms or relative to any reference class. The class-matrices are multi-dimensional time series, invariant to time-scale and to several transformations, representing the existence of every possible class over time. They can be combined with the simplest pattern finding methods for capturing complex tonal recurrences, and they can provide information about the hierarchical construction of any class sonority. The class-vectors are piecewise tonal summaries, quantifying the temporal presence of every possible class in a piece. They can be exploited in querying tasks of certain sophistication, beyond the possibilities of standard tonal features, while providing a means for describing (and explaining) similarity in alternative and insightful ways. The compromise between dimensionality and informativeness of the class-vectors may constitute a useful complement to existing content-based features. The insights provided by the class-scape visualizations can be interfaced with class-vectors, different class-spaces, inter-class measures, and standard pitchwise filters, in order to provide fast interactive assistance to the analyst. The examples in this chapter show that these descriptors can provide information about very different musical idioms, ranging from monophonic folk tunes and early polyphony to pieces that use exotic scales as well as atonal music.

Acknowledgements This work was supported by the EU 7th Framework Programme FP7/2007-2013 through PHENICX project [grant no. 601166].

Supplementary Material The interactive potential of the methods discussed in this work can be tested with our multi-scale set-class analysis prototype for Matlab, freely available from http://agustin-martorell.weebly.com/set-class-analysis.html. A detailed manual of the tool, a comprehensive table of set-classes, and a (growing) dataset of class-vectors, are also available at this site.

References

Babbitt, M. (1955). Some aspects of twelve-tone composition. *The Score and I.M.A. Magazine*, 12:53–61.

Castrén, M. (1994). *RECREL. A similarity measure for set-classes*. PhD thesis, Sibelius Academy, Helsinki.

Clough, J. and Douthett, J. (1991). Maximally even sets. *Journal of Music Theory*, 35(1/2):93–173.

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

Deliège, C. (1989). La set-theory ou les enjeux du pléonasme. *Analyse Musicale*, 17:64–79.

Foote, J. (1999). Visualizing music and audio using self-similarity. In *Proceedings of the Seventh ACM International Conference on Multimedia (MM99)*, pages 77–80, Orlando, FL.

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*. Yale University Press.

Forte, A. (1988). Pitch-class set genera and the origin of modern harmonic species. *Journal of Music Theory*, 32(2):187–270.

Forte, A. (1989). La set-complex theory: Elevons les enjeux! *Analyse Musicale*, 17:80–86.

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

Harley, N. (2014). Evaluation of set class similarity measures for tonal analysis. Master's thesis, Universitat Pompeu Fabra.

Hasty, C. (1981). Segmentation and process in post-tonal music. *Music Theory Spectrum*, 3:54–73.

Huovinen, E. and 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. (1992). Design principles in computer-based music representations. In Mardsen, A. and Pople, A., editors, *Computer Representations and Models in Music*, pages 5–39. Academic Press.

Leman, M. (2008). *Embodied Music: Cognition and Mediation Technology*. MIT Press.

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

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(1–2):433–444.

Martorell, A. (2013). *Modelling tonal context dynamics by temporal multi-scale analysis*. PhD thesis, Universitat Pompeu Fabra.

Martorell, A. and Gómez, E. (2015). Hierarchical multi-scale set-class analysis. *Journal of Mathematics and Music*, 9(1):95–108.

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

Nattiez, J. J. (2003). Allen Forte's set theory, neutral level analysis and poietics. In Andreatta, M., Bardez, J. M., and Rahn, J., editors, *Around Set Theory*. Delatour / Ircam.

Rahn, J. (1979). Relating sets. *Perspectives of New Music*, 18(1):483–498.

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

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.

Sapp, C. S. (2005). Visual hierarchical key analysis. *ACM Computers in Entertainment*, 4(4):1–19.

Straus, J. N. (1990). *Introduction to Post-Tonal Theory*. Prentice-Hall.

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

Tymoczko, D. (2011). *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press.