Hierarchical multi-scale set-class analysis

A. Martorell[*] and E. Gómez

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

This work presents a systematic methodology for set-class surface analysis using temporal multi-scale techniques. The method extracts the set-class content of all the possible temporal segments, addressing the representational problems derived from the massive overlapping of segments. A time versus time-scale representation, named class-scape, provides a global hierarchical overview of the class content in the piece, and it serves as a visual index for interactive inspection. Additional data structures summarize the set-class inclusion relations over time and quantify the class and subclass content in pieces or collections, helping to decide about sets of analytical interest. Case studies include the comparative subclass characterization of diatonicism in Victoria's masses (in Ionian mode) and Bach's preludes and fugues (in major mode), as well as the structural analysis of Webern's *Variations for piano* op. 27, under different class-equivalences.

1. Introduction

The description of the musical surface is a fundamental element for analysis, in the sense that it provides the raw material to the analyst. An appropriate surface characterization is not only a critical stage conditioning subsequent decisions, but it also often requires considerable effort from the analyst. In the context of pitch-class set analysis, the usage of adequate computational tools can free the analysts from the most systematic, time consuming and prone to errors tasks, and assist them in finding and testing material of analytical relevance. Proper representations and interfacing techniques can also facilitate a closer contact with the surface in reductive analysis methodologies, in which some selective criteria may obscure other surface features of interest. Aiming to provide such assistance tools, we propose three aspects as relevant for an appropriate surface description. First, a segmentation of the music in units of analytical pertinence. Second, a description of these units in adequate terms. Third, a representation of the results in insightful and manageable formats. In this work, all three aspects are considered from a systematic point of view. First, a comprehensive segmentation method scans every possible segment of the music. Second, a set-class level of description provides a systematic, general-purpose and standard analytical lexicon. Third, the massively redundant, overlapped, and hierarchical information is inspected by multi-scale and multi-dimensional representation techniques, adapted for analysis-related tasks. The methods discussed in this work can be tested by an interactive analysis tool, freely available at the first author's website (see the Supplemental material section at the end of the article).

---

[*]Corresponding author. Email: agustin.martorell@upf.edu

2. Background

2.1. *Set-class description*

Given that our systematic endeavour involves large amounts of data, it is convenient to reframe the main descriptive concepts in terms of dimensionality. *Pitch class* (Babbit 1955) is defined, for the twelve-tone equal-tempered (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. A *pitch-class set* (henceforth *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. In the TET 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*. The total count of interval classes in a pc-set can be arranged as a six-dimensional data structure called an *interval vector* (Forte 1964).

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 *class* (defined below). As pointed out by Straus, equivalence is not the same thing as identity, rather it is a link between musical entities that have something in common.[1] This commonality underlying the surface may eventually lend unity and/or coherence to musical works (Straus 1990, 1–2). In the context of pc-sets, the number of pitch classes in a set is referred to as its *cardinality*. This is perhaps the coarsest measure of similarity (Rahn 1980). Despite its theoretical relevance, cardinality is too general a notion of similarity to be of use in many analytical situations.[2] Among the many kinds of similarity in the set-theoretical literature, three of them are particularly useful[3]:

(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/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 class-equivalence systems fits a wide range of descriptive needs, thus their extensive usage in general-purpose music analysis. From them, iv-equivalence is the most general (197 classes). It shares most of its classes with T_nI-equivalence (220 classes), with some exceptions, named *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 from the three systems is T_n-equivalence (348 classes).

2.2. *On segmentation*

A main concern for analysis is the grouping of musical events into units of analytical pertinence, a complex problem which is sometimes encompassed under the term *segmentation*. According

---

[1] In this sense, a class equivalence can be conceived as an all-or-nothing *similarity measure* between two pc-sets. Later on in this work, we will discuss a similarity measure between different classes, which is used to quantify the distance between pc-sets beyond mere class-containment 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.

[2] For example, the cardinality-type 7 does not resolve differences among all the possible heptachords.

[3] See Forte (1964) for a comprehensive formalization, or Straus (1990) for a pedagogical approach.

to Forte, "by segmentation is meant the procedure of determining which musical units of a composition are to be regarded as analytical objects" (Forte 1973, 83). As pointed out by Cook, "no set-theoretical analysis can be more objective, or more well-founded musically, than its initial segmentation" (Cook 1987, 146). Considering the joint problem of finding appropriate segments and assessing their pertinence, a pursuit of systematicity and neutrality in the surface description arises as a convenient preprocessing stage. Without pretending that a neutral level of description suffices for analysis, one can conceive segmentation "not as something imposed upon the work, but rather [...] as something to be discovered" (Hasty 1981, 59).

Among the most systematic approaches to set-class surface analysis, Huovinen and Tenkanen propose a segmentation algorithm 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 (Huovinen and Tenkanen 2007). This segmentation is combined with certain "detector functions" to obtain summarized information from music pieces and collections. Despite the usefulness of the method, mostly discussed in the context of style characterization, four descriptive limitations can be observed, all of them derived from the segmentation policy. First, the note-based indexing in terms of "tail segments", which results in many segments that only contain some notes of a vertical chord. Second, the constraint by cardinality, which leaves most of the possible segments out of analytical consideration. Third, the combination of set classes in each tail-segment array, which does not allow the inspection of individual sonorities. And fourth, the non-uniqueness of the tail-segment arrays associated with a given time point, which results in an ill-defined temporal continuity.[4] In the pursuit of a practical indexing to the data, the last two issues lead to the application of averaging processes, operating simultaneously in both the set-class and the time domains, which do not provide clear-cut qualitative and quantitative interpretations.[5]

2.3. *Temporal multi-scale representation*

Multi-scale analysis, a widespread framework for data processing problems, is particularly suited for information characterized by unknown or unstable relevant time-scales. A multiple-time-scale computation and visualization method for tonal centre estimations of music pieces has been proposed in Sapp (2005). This systematic technique approximates an exhaustive segmentation of the music, by means of many non-overlapping sliding windows of sizes ranging from a fraction of a second to the whole duration of the piece. The key is then estimated for every segment, and the output of the method is a time versus time-scale plot of the estimations, named "keyscape", in which each tonal centre is coded by a different colour. Although the description is based on simple key estimation algorithms, it allows the inspection of relevant hierarchical tonal relations over time.

3. Temporal multi-scale set-class analysis

In this study we address the four descriptive issues in Huovinen and Tenkanen (2007), mentioned in Section 2.2, by conceiving systematization from a more holistic perspective. Our method extends the systematization of the multi-scale segmentation and representation methods in Sapp (2005) to the description itself, by providing a set-class level of description. The main features of

---

[4] In their analysis of Bach's *Es ist genug*, the authors do not consider the "changes" in the measures with respect to time, but to a stream of notes "successive in some sense" (Huovinen and Tenkanen 2007, 163).

[5] All the descriptions derived from the tail-segment arrays involve averaging the properties of many segments belonging to different classes, "the majority of which would probably not qualify as 'good' segments – according to neither a compositional nor a listening grammar" (Huovinen and Tenkanen 2007, 168).

the method are: (a) a systematic temporal multi-scale segmentation policy; (b) a systematic and neutral level of description, in terms of set classes; (c) an unambiguous temporal indexing of the music segments, in time and time-scale; (d) a set of data structures and visualizations, appropriate for different analysis domains. The usefulness of the method is discussed for several applications: (a) simultaneous visualization of every segment of a given class; (b) simultaneous exploration of every segment relative to a given class, using an inter-class measure; (c) exploration of set-class inclusion relations over time; (d) characterization of diatonicism in corpora; (e) structural analysis of serial music under different notions of equivalence.

It is important to clarify that our technique is intended to provide assistance to human analytical decisions. It is not conceived as a *model* (e.g. automatic structural segmentation or style classification), but as an instrumental *method*, which pursues the maximization of informativeness and the minimization of the interpretation artefacts. As a terminological clarification, in what follows, the term "segmentation" will be exclusively referred to as the systematic scanning of the musical surface, and not to the musical sections that may arise as potential elements of analytical interest.

3.1. Systematic multi-scale segmentation

Our point of departure 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; (b) a fully systematic method, which exhausts all the segmentation possibilities. The fully systematic method is postponed for the quantitative applications in which completeness of representation is necessary (see Section 3.3.3).

The approximate method consists of applying many overlapping sliding windows, each of them scanning the music at a different time-scale. The minimum window size[6] 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.[7] The same hop size[8] is applied for all the time-scales, in order to provide a regular grid for visualization and interfacing purposes. Each segment is thus indexed by its centre location (time) and its duration (time-scale).

3.2. Segment 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 in the 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,[9] including the iv-, T_nI-, and T_n-equivalences (discussed in Section 2.1). For systematization completeness, the three class spaces are extended to include the so-called *trivial forms*.[10] With this, the total number of interval vectors rises to 200, while the T_nI- and T_n-equivalence classes sum to 223

---

[6] Window size refers to the duration of each analysis frame, that is, the time-scale.

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

[8] Hop size refers to the amount of time that the window is displaced between two consecutive analysis frames. The overlapping factor of a sliding window is defined as the ratio between the hop size and the window size.

[9] As formalized in Forte (1964), and available at the first author's website (see the Supplemental material section at the end of the article).

[10] The null set and single pitch classes (cardinalities 0 and 1, containing no intervals), the undecachords (cardinality 11) and the universal pc-set (cardinality 12, also referred to as the *aggregate*).

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 ($T_n I$-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 (3-11A for minor, 3-11B for major), but not under iv- or $T_n I$-equivalences.

3.3. *Representation*

The joint outcome of the segmentation and description stages is an arrangement of the class content of the piece, indexed by time and time-scale. This data structure will be referred to as a *class-scape*. In order to design representations for assisting human analysts, it is convenient to consider different application scenarios.

3.3.1. *Basic class selection*

A simple but useful task is to localize all the segments in the music belonging to a given class. This is illustrated in Figure 1, where the main features of the multi-scale representation are also introduced. The top figure shows the class-scape computed for Debussy's *Voiles*, filtered by the class 6-35 (pixels in black), which corresponds to the predominant whole-tone scale in the composition. As a visual reference, a thin blue line delineates the boundaries in time and time-scale of the complete, non-filtered, class-scape information. The bottom figure depicts an aligned piano roll representation of the score for visual indexing of the composition.

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.

3.3.2. *Multi-class representation and REL distance*

An alternative representation, allowing the inspection of all the 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 taken by means of an inter-class measure. Among the many distances proposed in the set-theoretical literature, Lewin's (1979) REL can handle any pair of classes,[11] and it takes into account the complete subset content of the classes being compared, exhausting the pc-set inclusion relations systematically. The REL similarity between two classes *A* and *B*, adapted from the algorithm in Castrén (1994) in order to include the trivial forms, is computed as follows:

(1) For each of the classes *A* and *B*, a 357-dimensional *subset vector* is computed. It is comprised of the interval vector of the class (6-dimensional), followed by a 351-dimensional vector,[12] which accounts for the number of occurrences of each T_n-*type chord* in the class, arranged

---

[11] Unlike most measures, REL does not require pairwise cardinality equality.
[12] REL discriminates distances at the level of T_n-equivalence (351 classes).

Figure 1. Debussy's *Voiles*. Class-scape filtered by 6-35, piano roll, and 3 sample segments.

Figure 2. Debussy's *Voiles*. Class-scape relative (REL) to 7-35.

according to Forte's cardinality-ordinal notation. For example, the class 4-20 has the subset vector:

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

It begins with [1 0 1 2 2 0], the interval vector of 4-20, and this is followed by the 4 occurrences of 1-1 in 4-20, 1 occurrence of 2-1, 0 occurrences of 2-2, and so on, up to exhaust the 351 T_n-types.

(2) We denote by $\mathrm{sub}(X, i)$ the ith element of the subset vector corresponding to the class X.

(3) The REL similarity between the classes A and B is then computed as:

$$\mathrm{REL}(A, B) = \frac{\sum_{i=1}^{357} \sqrt{\mathrm{sub}(A, i) \cdot \mathrm{sub}(B, i)}}{\sqrt{\sum_{i=1}^{357} \mathrm{sub}(A, i) \cdot \sum_{i=1}^{357} \mathrm{sub}(B, i)}} \in [0, 1].$$

Each pixel in the class-scape is coloured according to the REL distance between the class it represents and any chosen reference class. Figure 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 segment belonging to 7-35, so it is clear that the previous 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 white = 0 to black = 1). This way, every segment of the music (every pixel in the class-scape) is represented according to its REL-closeness to 7-35. It is straightforward to visualize 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 Figure 1, as well as their building subsets, are poorly (lightly) represented with respect to 7-35. Incidentally, there exist some isolated segments (just two pixels 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 the class-scapes: the analytical relevance of what is shown is often related with the accumulation of evidence in time and time-scale. A significantly
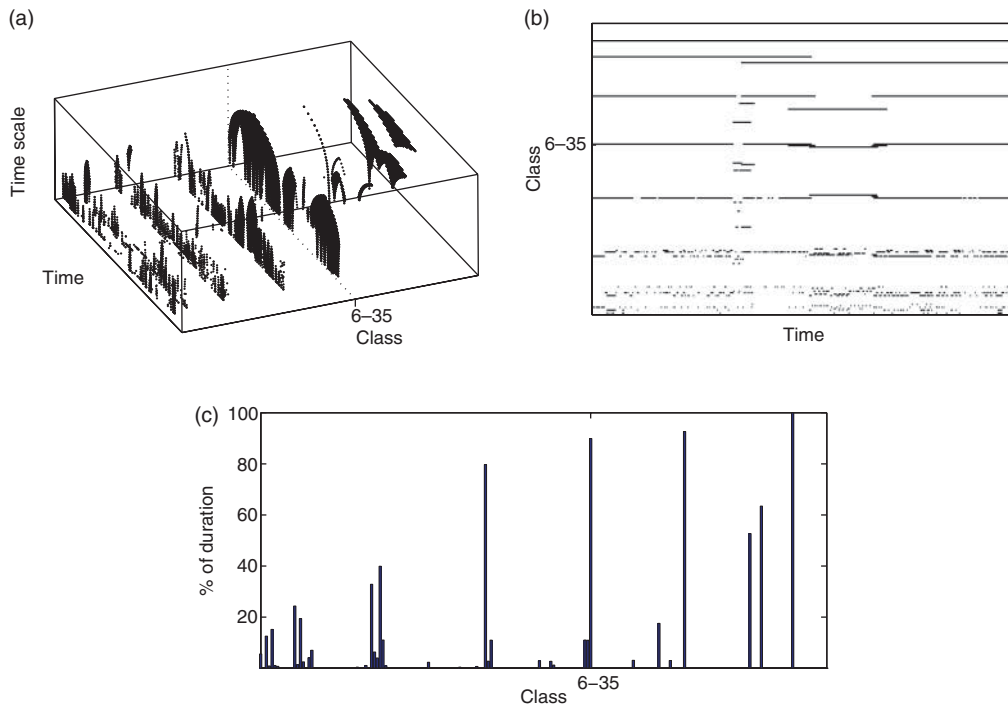
Figure 3. Debussy's *Voiles*: (a) class-scape as a sparse 3D matrix, (b) class-matrix, and (c) class-vector.

large area of the scape showing the same evidence 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. Smaller patches, even single isolated pixels, capture the class content of their corresponding segments as well. However, they could just be concatenation by-products around the boundaries between more significant sections.[13] The human visual cognition plays a role here, by allowing the viewer's attention to focus on the areas which accumulate homogeneous evidence, an important feature of multi-scale representations for assisting pattern recognition tasks.[14] On the other hand, localizing glimpses of residual evidence may be worth a closer inspection, and so facilitates the interaction with the class-scape.

3.3.3. *Piecewise summarization:* 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 two-dimensional class-scape so far can be thought of as a sparse three-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 Figure 3(a), such an arrangement is shown for Debussy's *Voiles*, with 6-35 annotated as reference. The first reduction process consists of projecting this information into the time versus class plane. Given the special meaning of the lost dimension (time-scale), this implies growing each point to the actual duration of

---

[13] As in the case of the spurious 6-33B in *Voiles*, arising at the boundary between the pentatonic passage in bars 42–47 and the arpeggiated section beginning in bar 48.

[14] In his classic on visualization, Tufte discusses similar cognitive implications to demonstrate his massive "data maps" (Tufte 2001, 16–20) as useful instruments for reasoning about quantitative information.

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 according to Forte's cardinality-ordinal convention. An *activated* position (t, c) in the class matrix (pixels in black) means that at least one segment belonging to the class c includes the time position t. A position in the time axis of the class-matrix is not associated with a unique segment, but to all the segments containing that time point. The class-matrix, thus, summarizes information from all the time-scales simultaneously. Unlike the approach in 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 next.

Figure 3(b) shows the class-matrix for Debussy's *Voiles* under iv-equivalence, revealing the prelude's economy of sonorities. 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 overlapping 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. For instance, two consecutive diatonic passages involving modulation may 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 during the same time span. Class-matrices are thus suitable for inspecting in detail the inclusion relations down the hierarchy of classes.

An even more compact representation provides the means for a quantitative description. 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 3(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 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. This is done by locating every change in the pc-set content, whether the product of onsets or offsets, and segmenting the piece by considering all the pairwise combinations among these boundaries. The exhaustive computation of class-vectors has thus been performed in all the applications that follow.

4. Subclass analysis of diatonicism in corpora

The class-matrices can 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 Figure 4. 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 Figure 4(a). The result of this process is a *subclass-matrix*, in Figure 4(b), with a number of rows equal to the number of subset classes (up to cardinality 6 here). Partially overlapping classes which are not a subset of 7-35 are then removed from the subclass-matrix. Following the same method as for
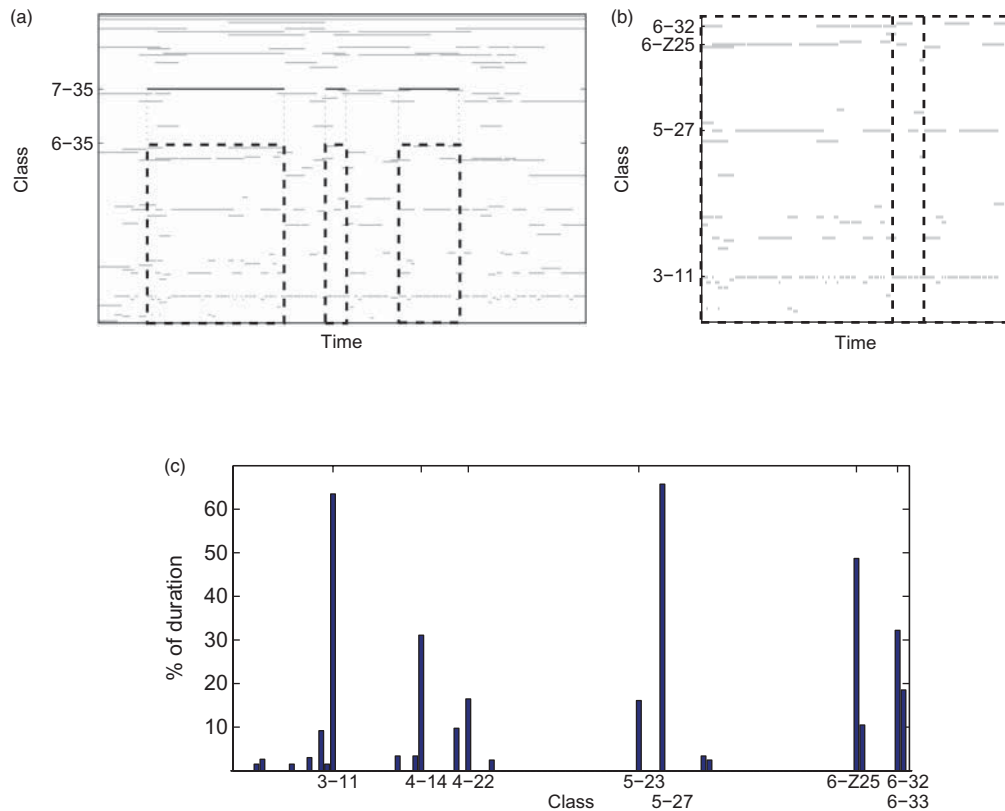
Figure 4. Victoria's *Ascendens Christus*, Agnus Dei: (a) filtering under 7-35, (b) subclass-matrix, and (c) subclass-vector.

class-vectors, a *subclass-vector* is then computed from the subclass-matrix. The subclass-vector, in Figure 4(c), quantifies the total subclass content contributing to the reference class, describing 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 scalar formation classes, such as 5-23 and 6-33. Relevant tetrachordal instances are featured by 4-14 and 4-22.

The method can be extended for characterizing a corpus, by taking a classwise average across the subclass-vectors extracted from all the pieces. Figure 5 depicts the mean subclass-vector under 7-35 computed for two contrasting corpora: (a) the Victoria's parody masses in Ionian mode[15]; (b) the preludes and fugues in major mode from Bach's Well-Tempered Clavier. The selection of the corpora is based on the close relations 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. A prominent usage 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.[16] On the other hand, the Locrian hexachord 6-Z25 is far more present in Bach. Apart from its

---

[15] 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.

[16] The class 5-27 results from the combination of the dominant and tonic major triads.
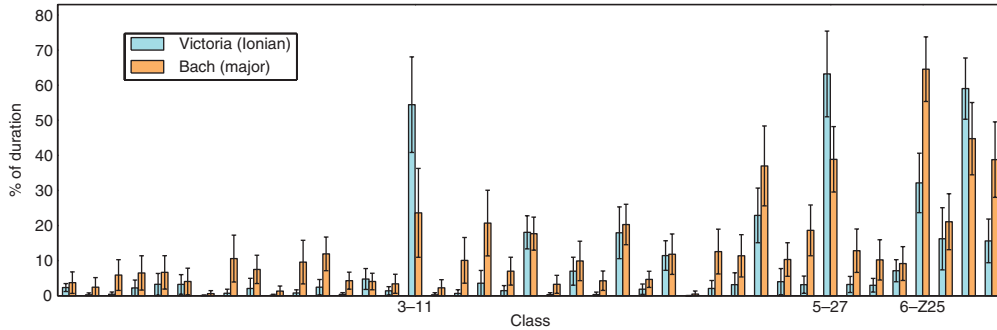
Figure 5.   Diatonicism in Victoria and Bach. Mean subclass-vectors under 7-35.

instantiations as perfect cadences[17] in both corpora, 6-Z25 appears consistently in many motivic progressions in Bach.[18]

The method is thus sensible to capture differences in the usage of classes beyond a mere *event counting* approach, by revealing and quantifying the building blocks of specific sonorities in a hierarchical way. It is important to clarify that the method aims to facilitate the inspection of the class and subclass content over time in objective, neutral, and analysis-friendly terms. That this information may be used to discriminate stylistic features (e.g. between historical periods or between works of the same composer), obviously depends on the specific corpora used and the analyst's interpretation.

5.   Class-equivalences and serialism

The final example illustrates the use of the method for analysing structural similarity in serialism, a style in which the concept of class-equivalence has compositional relevance.

Figure 6(a) shows the opening of the first movement of Webern's *Variations for piano* op. 27, including the first statement of the series in its well-known palindromic form. Figure 6(b) and 6(c) depict the prime tone row P-0, its inversion I-0, and their retrogrades R-0 and RI-0. Throughout the movement, the exposition of the series in one stave (P-n) runs in parallel with its retrograde at the same transpositional level (R-n) in the other stave. This parallel exposition is also kept for the inverted statements of the series (I-n and RI-n). What catches our attention here, however, is the trichordal segmentation of the series, and the resulting hexachordal harmonization with the retrograde. The first three notes of the row and the last three notes (first three notes of the retrograde) can be clearly isolated 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 $\langle 332232 \rangle$ is openly manifested at this segmentation level, as it is illustrated in Figure 7 (top). The hexachord formed by the extreme trichords of the series belongs to the set-class 6-Z41B, while the central hexachord belongs to 6-Z12A. Both hexachords constitute a *Z-related pair*: they share the interval vector but they are not T_nI-equivalent. The same applies to the inverted series: the extreme trichords of the inverted row form a 6-Z41A hexachord, while the central hexachord belongs to 6-Z12B. The

---

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

[18] By interfacing class-vectors with class-scapes, the content of particular class instantiations can be easily explored (and heard) using the interactive analysis tool available at the first author's website (see the Supplemental material section at the end of the article).
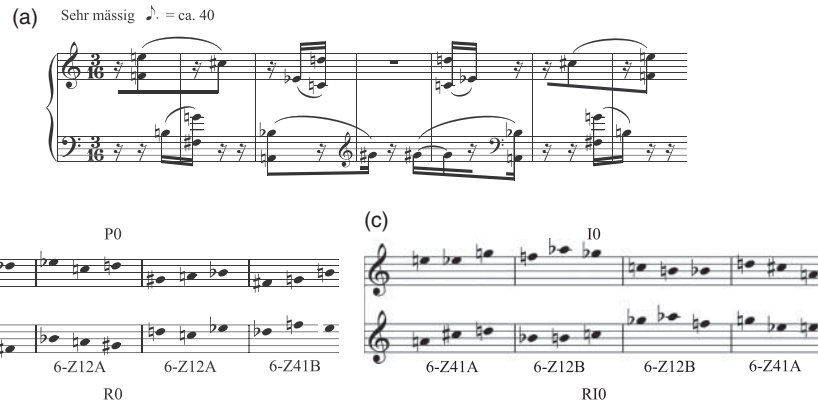
Figure 6. Webern's *Variations for piano* op. 27/I: (a) bars 1–7, (b) hexachordal segmentations for P-0/R-0, and (c) hexachordal segmentations for I-0/RI-0.

whole pattern is restated throughout the movement, evolving at different transpositional and/or inversional levels. 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 four times. Section A' virtually reconstructs the same pattern. The well-known structure of the piece is shown in Figure 7 (bottom), annotated according to Cook (1987), and time aligned with the class-scape (top).

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 elements[19] in the time series according to some similarity measure. The most likely recurrences are revealed by diagonals in the matrix accumulating high similarity scores. Self-similarity matrices were computed for three different descriptors over time: (a) the pc-set time series, *realized* in time[20]; (b) the class-matrix under T_n-equivalence; and (c) the class-matrix under T_nI-equivalence. The self-similarity matrix computed from the pc-set time series, depicted in Figure 8(a), does not reveal 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 Figure 8(b). 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 Section 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). Finally, by using transpositional and inversional equivalence, in Figure 8(c), all the row statements at A are fully revealed at A'.

The prominent recurrence of a surface phenomenon, here an hexachordal iv-sonority, leads one to 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, 85, footnote 2). As we have shown, the set of sonorities revealed by our method is justified by: (a) the structure of the series; (b) the harmonization of the direct and retrograde rows in hexachords; (c) the score punctuation delineating those hexachords; (d) the relations among the hexachordal instantiations under different

---

[19] Usually, time *frames* of a multidimensional time series.

[20] The information conveyed by a piano roll representation, which constitutes a symbolic *substitute* of the *chroma features*, which are extensively used for similar tasks in the audio domain (Müller 2007). Both signals do not represent identical information, although many chroma implementations actually aim to capture the *ideal* pc-sets.
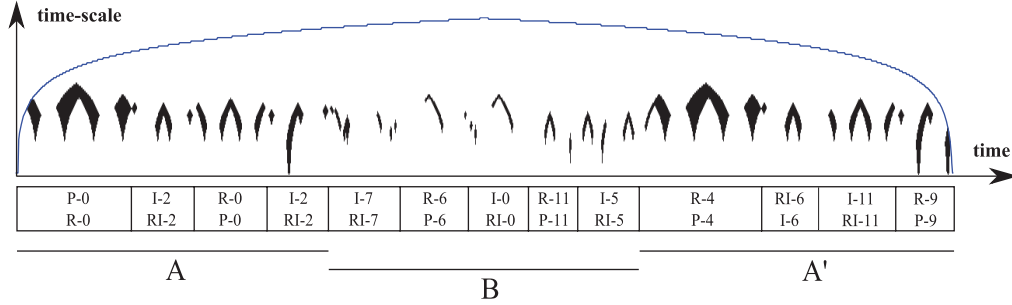
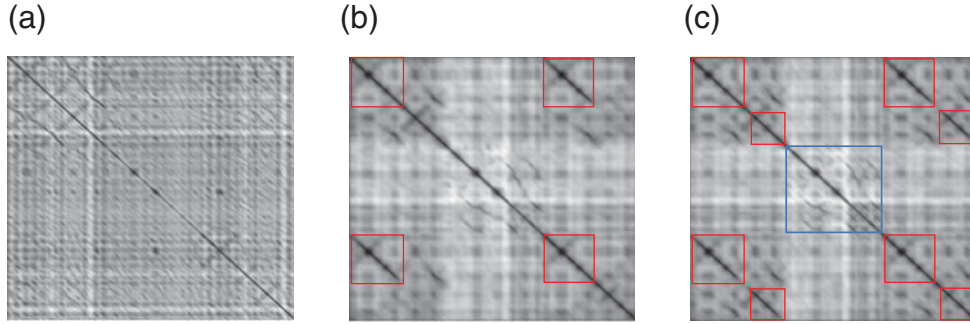Figure 7. Webern's *Variations for piano* op. 27/I. Top: class-scape filtered by ⟨332232⟩. Bottom: structure.

Figure 8. Webern's *Variations for piano* op. 27/I. Self-similarity matrices: (a) Pc-set time series, (b) class-matrix (T_n-equivalence), and (c) class-matrix (T_nI-equivalence).

class-equivalences. We find this last observation of particular analytical interest. The Z-relations among the hexachordal statements provide two contrasting pc-set materials of a special kind: they are similar to each other (iv-equivalent), but distinguishable beyond the basic transformations (non-T_nI-equivalent). The latter aspect provides a means for clarifying the palindromic discourse of each row statement, while the former reinforces the overall coherence. Moreover, as the Z-relations of a set of cardinality 6 connect with the set's complement, the all-embracing symmetry of the piece is further supported at the row level.

6.   Conclusions

This work proposes a reconsideration of the basis for set structure analysis. Pople's three principles in terms of segmentation, reduction, and significance judgement (Pople 2004) are reframed by computational systematicity and interfacing methods.

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 the means for revealing them by accumulation of evidence. The method does not impose cardinality, in favour of a comprehensive account of all the different segments of music, which in turns facilitates simultaneous multi-cardinality explorations. Systematic segmentation is proposed in both exact and approximate versions, the former providing appropriate quantification of further reductive descriptions, and the latter facilitating visual inspection while maintaining reasonable comprehensiveness. A main novelty is the introduction of an explicit and unambiguous temporal index, as a substitute of the commonly used note-based indexing, which allows a proper

description with respect to time. This segmentation level does not contribute to interpretation artefacts of any kind, as all the possible different segments are individually considered.

Second, it avoids *a priori* reductive classifications. Systematicity accounts for every class, computed from the actual content of every segment. This provides a neutral level of description and avoids quantification at this stage, which in our opinion constitutes a premature information summarization. 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 human visual attention can easily grasp the most stable sections, 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. Unlike the approach in Huovinen and Tenkanen (2007), our temporal indexing method makes possible the precise inspection of each individual class.

Third, it provides a 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, a strict separation of classes is guaranteed, avoiding their removal or fusion up to the very moment of the interpretation. Quantification accounts for the relative temporal presence of each class in the piece, whether globally or with respect to a given reference class, which allows the qualification of details at the level of class inclusion. A useful feature of the duration-based quantification is that potential classes of analytical interest may be revealed by 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. By interfacing the different representations, the interactive version of the method aims to assist dynamic explorations of the music. The massive multidimensional information, from the global to the local, is readily available and accessible for testing. The analysis loop is thus assisted by interactive filtering of information which is guaranteed to be comprehensive, precisely described, objectively quantified and intuitively visualized.

Two main shortcomings of the method, however, are derived from combining a temporal-based segmentation with descriptions in terms of sets. First, the descriptive *saturation* of any segment containing the universal pc-set, inherently featured by many segments in highly chromatic and atonal music, for which the informativeness is often limited to short time-scales. Second, the fact that many pc-sets of analytical or perceptual relevance can only be found by inspecting individual voices and/or fragmentary polyphonic groupings,[21] a kind of surface description that cannot be derived from plain temporal-based segmentations. The method can be applied to individual or combined staves, provided the availability of proper score encodings, but this comes at the price of increasing the data dimensionality. Future research may involve the design of systematic representations and practical mining techniques for such information.

---

[21] See Perle (1991, 84–110) for a detailed discussion about the concept of *simultaneity* in atonal music.

Supplemental material

The interactive potential of the methods discussed in this work can be tested by our proof-of-concept multi-scale set-class analysis tool, built upon part of the MIDI toolbox (Eerola and Toiviainen 2004), and implemented as a user interface (GUI) for Matlab. The prototype can be downloaded from http://agustin-martorell.weebly.com/set-class-analysis.html. A concise user manual of the tool, including a comprehensive table of set-classes, is also available at this site.

References

Babbit, M. 1955. "Some Aspects of Twelve-Tone Composition." *The Score and I.M.A. Magazine* 12 (June): 53–61.

Castrén, M. 1994. "RECREL. A Similarity Measure for Set-Classes." PhD thesis, Sibelius Academy, Helsinki.

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

Eerola, T., and P. Toiviainen. 2004. *MIDI Toolbox: MATLAB Tools for Music Research*. Jyväskylä: University of Jyväskylä. www.jyu.fi/musica/miditoolbox/

Foote, J. 1999. "Visualizing Music and Audio Using Self-Similarity." In *ACM Multimedia,* edited by J. F. Buford, S. M. Stevens, D. C. A. Bulterman, K. Jeffay and H. J. Zhang, 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*. New Haven: Yale University Press.

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 (1): 54–73.

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

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.

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

Perle, G. 1991. *Serial Composition and Atonality*. Berkeley: University of California Press.

Pople, G. 2004. "Modeling Musical Structure." In *Empirical Musicology: Aims, Methods, Prospects,* edited by E. Clarke and N. Cook, 127–156. New York: Oxford University Press.

Rahn, J. 1980. *Basic Atonal Theory*. New York: 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." *Computers in Entertainment* 4 (4): 1–19.

Straus, J. N. 1990. *Introduction to Post-Tonal Theory*. Upper Saddle River, NJ: Prentice-Hall.

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