This paper was presented at the 53rd Conference of the Audio Engineering Society (January 2014) as paper number P1-1. The official published version can be found at http://www.aes.org/e-lib/browse.cfm?elib=17115

SEMIOTIC DESCRIPTION OF MUSIC STRUCTURE :

AN INTRODUCTION TO THE QUAERO/METISS STRUCTURAL ANNOTATIONS

FREDERIC BIMBOT, GABRIEL SARGENT, EMMANUEL DERUTY
CORENTIN GUICHAOUA, EMMANUEL VINCENT

METISS (PANAMA) Research Group, INRIA, IRISA CNRS-UMR 6074 & Université de Rennes 1,
Campus Universitaire de Beaulieu, 35042 Rennes cedex, France

frederic.bimbot@irisa.fr gabriel.sargent@irisa.fr emmanuel.deruty@inria.fr
corentin.guichaoua@irisa.fr emmanuel.vincent@inria.fr

Interest has been steadily growing in semantic audio and music information retrieval for the description of music structure, i.e. the global organization of music pieces in terms of large-scale structural units. This article presents a detailed methodology for the *semiotic* description of music structure, based on concepts and criteria which are formulated as generically as possible. We sum up the essential principles and practices developed during an annotation effort deployed by our research group (Metiss) on audio data, in the context of the Quaero project, which has led to the public release of over 380 annotations of pop songs from three different data sets. The paper also includes a few case studies and a concise statistical overview of the annotated data.

1 INTRODUCTION

Among the wide variety of audio signals, music is a very specific type of “content”, which can be defined as “the art, process and result of deliberately arranging sound items with the purpose of reflecting and affecting senses, emotions and intellect” [1].

In fact, music plays a central role in many human activities and it has been tremendously impacted, over the past century, by the development of analog and digital audio engineering technologies.

Because music is a very sophisticated signal, content-based management of music data currently remains a challenge. In fact, there are many possible types of symbolic metadata which can be used to describe a musical audio content : notes, chords, instruments, singer IDs, tempo, genre, moods, etc… As such, music holds today an important position in semantic audio research activities.

This profusion of simultaneous information sources in music creates specific difficulties. In this context, music structure is frequently considered as a central element to semantic or symbolic music description and modeling, because it constitutes the backbone over which these various sources of information develop.

However, given the wide variety of music signals, describing music structure turns out to be a scientific challenge, not only from the algorithmic perspective but first and foremost from a conceptual viewpoint [2].

Significant effort has been dedicated in the MIR community, towards the production of annotated resources [3][4][5] and the organization of evaluation campaigns [4][6] for automatic extraction of music structure. Indeed, the availability of exploitable experimental material appears as a key factor in producing reliable and reproducible research results.

Given the need for formal and operational concepts [7][8], our research group has been developing, over the past few years, methodological landmarks for describing the structure of musical pieces [9][10], with the concern of resorting to concepts and criteria which are formulated as independently as possible from the music genre and which accommodates multi-dimensionality. The method has reached a level of maturity that has enabled the recent release of a set of 383 structural annotations of pop music pieces, downloadable at :

musicdata.gforge.inria.fr/structureAnnotation.html

This article sums up and illustrates the essential aspects of the annotation methodology, by presenting them as introductive guidelines to semiotic annotation, including general principles, practical considerations and occasional advice in specific situations.

The paper is divided in three parts. The first part elaborates on the various viewpoints on music structure and defines, in general terms, the semiotic approach adopted in this work. The second part presents practical aspects of the annotation methodology, including case studies on specific musical passages. The third part provides a few examples of semiotic descriptions and a concise statistical overview of the annotated data.

This article is intended to be useful to understand the nature of the released annotations, to anyone who may either use the data or contribute to the annotation effort. But we believe that the proposed methodology also provides interesting insights on music structure, paving the way towards a more robust definition and extraction of structural metadata in large sets of music pieces from various genres.

The concepts and the methodology proposed in this article are primarily applied to what we will call *conventional music*, which covers a large proportion of current western popular music and also a large subset of classical music (see the annex for an example). However, we keep in mind that some other types of music (in particular, contemporary music) are much less suited to the proposed approach.

The method has been primarily designed for the description of music material in *audio* form, but most concepts can be straightforwardly adapted to scores or transcribed music. Reading this document will fruitfully be coupled with an in-depth examination of the annotated data, which can be readily consulted via a web interface at :

metissannotation.irisa.fr

2 A SEMIOTIC APPROACH TO THE DESCRIPTION OF MUSIC STRUCTURE

2.1 Time scales

It is commonly agreed that the composition and the perception of music pieces rely on simultaneous processes which vary at different timescales. Similarly to [11], we consider the three following levels corresponding to three different ranges of timescales :

- the **low-scale** elements which correspond to fine-grain events such as notes, beats, silences, etc… We call this level the *acoustic level* and its time scale is typically below or around 1 second.
- the **mid-scale** organization of the musical content, based on compositional units such as bars or hyperbars or on perceptual units such as musical cells, motifs and phrases, ranging typically between 1 and 16 seconds. We will refer to this level as the *morpho-syntagmatic* level.
- the **high-scale** structure of the musical piece, which describes the long term regularities and relationships between its successive parts, and which we will consider typically at a time scale around or above 16 seconds. This typically corresponds to the *sectional form* of the piece.

2.2 Semiotic structure

At the scale of an entire piece, music structure is a concept which can be approached in several ways, in particular :

a. **The acoustic structure,** which describes the active instruments and/or timbral textures over time : singer(s), lead entries, instrumentation, etc…

b. **The functional structure,** which is based on usual designations of the different parts in terms of their role in the music piece, for instance : intro – verse – chorus – bridge – etc… (cf. [12], for instance),

c. **The semiotic structure,** which aims at representing, by a limited set of arbitrary symbols (called *labels*), the similarities and interrelations of structural segments within the piece [10].

These various views of music structure have influenced the design of methods and algorithms for the automatic analysis of audio data, for instance [13-16]. They are also explicitly considered as independent "layers of labels" in the SALAMI project annotation scheme [3].

In the present work, we focus on *semiotic structure*, i.e. the description and annotation of similarities between segments. Note that we use the term *semiotic* in a quite restricted scope, as denoting the high-level *symbolic* and *metaphoric* representation of musical content.

Of course, semiotic structure annotation requires the determination of proper segment *boundaries*. This question is explicitly treated in this work, while it is seldom addressed in concurrent approaches.

Concretely, the *semiotic structure* of a music piece is something that may look like :

A B C D E F B C D E G D E D E H

thus reflecting :

✓ some sort of high-level decomposition/segmentation of the whole piece into a limited number of blocks (here 16 blocks) of comparable size, and

✓ some degree of similarity or equivalence relationship between blocks bearing identical labels (here, 8 distinct symbols).

Providing a semiotic description for a music piece requires primarily the identification of a proper *granularity* (block size and number of blocks) which then conditions the inventory of labels.

Indeed, choosing a finer granularity in the previous example could lead to a sequence of labels such as:

AA'BB'CC'DD'EE'FF'BB'CC'DD'EE'GG'DD'EE'DD'EE'HH'

where any symbol X is systematically followed by symbol X', thus yielding a rather redundant description.

Conversely, a coarser granularity would require either the uneven grouping of the units into *irregular* segments (i.e. of more diverse sizes) :

A BC DE F BC DE G DE DE H

or a very misleading representation such as :

AB CD EF BC DE GD ED EH

which would completely hide the similarities between portions of the piece which had identical labels at a lower scale.

This example thus illustrates a simple case where there clearly exists a preferable granularity at which the semiotic level of the music piece can be described with some optimal *compromise* in terms of :

- coverage of the set of labels,
- accuracy of the sequence of labels,
- regularity of the segment decomposition,
- economy / parsimony of the description.

2.3 Description criteria

Producing consistent annotations across pieces, genres and annotators requires :

1) The definition of a *target time scale* providing an adequate granularity for the semiotic structure of the piece,

2) A segment model to locate as unequivocally as possible the *segment boundaries*,

3) Clear criteria to qualify and denote the *similarities* between segments.

A major difficulty resides in the necessity to formulate segment models and similarity criteria in a generic way, i.e. as independently as possible from the genre of the piece while accounting for the versatility and the multiplicity of musical dimensions which contribute to the structure.

It is also essential for the method to be based on a multidimensional analysis of the musical content, rather than considering a particular dimension (for instance, harmony) as being necessarily well adapted for the structural description of all music pieces. These specific issues are addressed with particular care in this work and constitute driving principles of the proposed approach.

It may be argued that the resulting description of the musical piece may not correspond to that intended by the composer [17] and that it may not necessarily reflect perceptual characteristics that listeners would identify as primarily salient.

In that sense, the proposed description should not be understood as a ground "truth", but as a standardized and codified representation of structural information, in terms of *similarity relationships* between segments, around a target time scale.

We believe that it is ultimately easier to come to agree on such a description, which keeps its distance with potentially subjective considerations.

3 ANNOTATION METHODOLOGY

Following the principles stated in the previous section, the process of describing the semiotic structure of a piece is based on the following five steps :

i. Define a proper working time scale for the piece,
ii. Locate structural blocks by matching them to a prototypical segment model, called System & Contrast (S&C),
iii. Cluster segments based on the most salient S&C similarities and their relative position in the piece,
iv. Analyze in details similar blocks to adjust their class membership and to determine their type of variant,
v. Finalize, revise, adjudicate the annotation.

3.1 Step 1 : time scale definition

The annotation process starts by defining an arbitrary fine-grained unit of length u, around 1 second, synchronized with the musical time scale, which we call the *snap* (generally corresponding to the downbeat and/or the tactus). By convention, structural boundaries are synchronized on snaps.

The coarse target time scale U for semiotic description (called *structural pulsation period*) is set as a multiple of u ranging preferably between 12 and 24 (usually 16) chosen so as to match the main period of repetition of musical content in the song.

As mentioned in section 2.1, this period (around 16 s) corresponds to the typical time scale at which long term regularities tend to develop in music. In many cases the structural pulsation period corresponds to 8 bars.

It is useful to identify at this stage a candidate *central section* of the piece, i.e. a characteristic structural block around time scale U, which can serve as a calibration for the structural pulsation period over the entire piece. Usually, the chorus (when it exists) is a good candidate.

It must be understood that the actual size of semiotic blocks may vary within the piece around the target size U and that the choice of U may be reconsidered if a better value emerges from the next steps of the annotation process.

3.2 Step 2 : segmentation

A key principle that governs the segmentation of the piece into structural blocks is to match the musical content to a segment model, called System & Contrast (S&C) [18]. Block boundaries are then identified as the beginning and the end of successive instances of the model around the target time scale.

Subsection 3.2.1 presents the S&C model in its canonical form (the square system), 3.2.2 extends it to a wider range of configurations and 3.2.3 explains how it can be used to segment musical content.

3.2.1 *The System & Contrast model (square form)*

Assuming a structural block S, we consider its subdivision into 4 *morphological units* (MU) of comparable size, $\boldsymbol{x_{00}}\ \boldsymbol{x_{01}}\ \boldsymbol{x_{10}}\ \boldsymbol{\bar{x}_{11}}$, and thus form a 2 × 2 matrix *system* :

$$S = \left[\!\left[\begin{matrix} \boldsymbol{x_{00}} & \boldsymbol{x_{01}} \\ \boldsymbol{x_{10}} & \boldsymbol{\bar{x}_{11}} \end{matrix} \right]\!\right]$$

We assume that, when these 4 MUs belong to a same structural unit, some salient *syntagmatic relationships* (SR) exist between them, in particular between the *primer* $\boldsymbol{x_{00}}$ and the subsequent elements :

- horizontal relationship : $\boldsymbol{x_{01}} = f(\boldsymbol{x_{00}})$
- vertical relationship : $\boldsymbol{x_{10}} = g(\boldsymbol{x_{00}})$

Relations f and g are merely *similarity* relationships, i.e. transformations which preserve globally (or at least locally) some conformational properties or features of the MUs (see list further below).

We call *contrast*, the divergence between $\boldsymbol{\bar{x}_{11}}$ (the actual element in S) and $\boldsymbol{x}_{11}$, the virtual element which would form a complete logical system with $\boldsymbol{x_{00}}$, $\boldsymbol{x_{01}}$ and $\boldsymbol{x_{10}}$, i.e. which would be obtained by combining f and g and applying this combination to the primer. Therefore, $\boldsymbol{x}_{11}$ can be noted $(gof)(\boldsymbol{x_{00}})$.

Figure 1 depicts (metaphorically) the main components of an S&C. The contrast function γ appears as the discordance of $\boldsymbol{\bar{x}_{11}}$ w.r.t. a subset of properties forming the system. The carrier sequence in the S&C builds up a situation of expectedness which is resolved by the presentation of the contrast, acting as the closure of the segment.

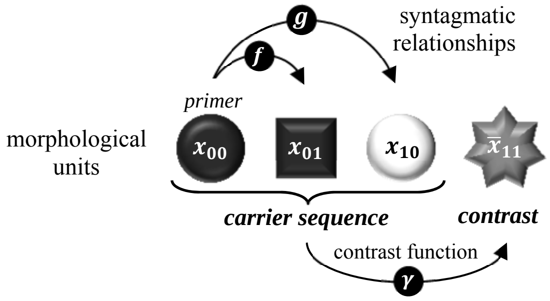


Figure 1 : a schematic illustration of the S&C model

MU properties are represented metaphorically as visual properties (shape, gray level, size, etc…)

This discordance created by the contrast is detected in reference to the carrier sequence $\boldsymbol{x_{00}\ x_{01}\ x_{10}}$ by first deducing the syntagmatic relationships and then finding out in what respect the last element $\boldsymbol{\bar{x}_{11}}$ is deviating from a purely "logical" sequence (see Figure 2).

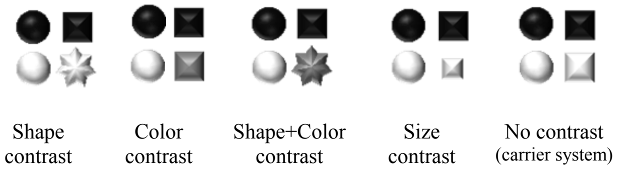


Figure 2 : different contrasts based on a same carrier system (metaphorical representation)

Properties involved in the S&C have different status :

- *descriptive properties* (all properties required to characterize completely the elements of an S&C),

- *structuring properties* (the subset of descriptive properties which participate to the syntagmatic relationships within the S&C),

- *contrastive properties* (the subset of structuring properties on which the contrast applies), and

- *distinctive properties* (the subset of structuring properties which distinguish the carrier systems of two classes of S&Cs within the piece).

Identifying a musical S&C requires that the syntagmatic relationships (and, to a certain extent, the contrast γ) are rather simple, in order to enable their direct detection from the musical content. This viewpoint relies on and extends the concept of cognitive rules as introduced by Narmour [19].

In practice, syntagmatic relationships often operate on the notes or on the underlying harmony :

- exact or almost exact repetition,
- starts-like, ends-like, …
- chromatic or diatonic transposition,
- shift of note placements or durations,
- inversion, complementation (in various possible ways),
- etc…

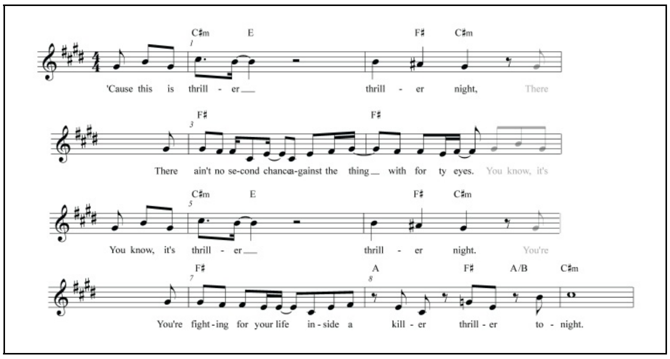
but they may also operate on the amplitude, time, timbral or other musical dimensions such as :

- amplitude increase / decrease / zeroing
- fragmentation / augmentation / expansion
- extension / simplification (i.e. insertion / deletion of auxiliary musical material, such as ornaments)
- adjunction / suppression / change of instrument(s)
- etc…

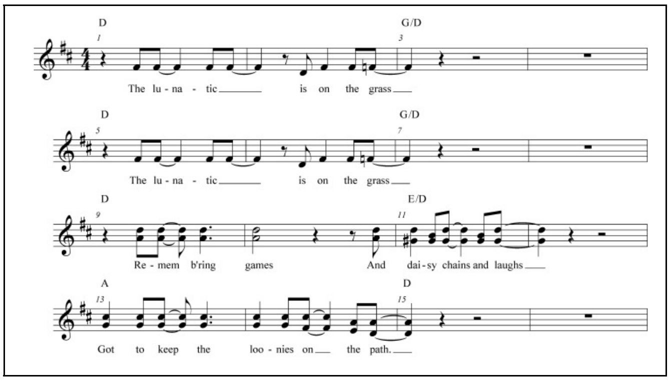
In general, syntagmatic relationships can be considered as varied forms of (exact, approximated, piecewise or local) mathematical transformations, such as identity, translation, dilation, rotation, symmetry, inversion, etc… applied to some particular musical dimensions (time, intensity, rhythm, melodic contour, chords progressions, tonality, etc…).

In some cases, one of the SR can be an *in extenso* substitution of one MU by another one, which can be viewed as a function "*new*". In that case, the other SR is very straightforward (typically, "*identity*", "*near-identity*" or "*starts-like*"), which leads to well-known structural patterns such as *abac* (period-like pattern, see Ex. 1) and *aabc* (sentence-like pattern, see Ex. 2). The S&C model is indeed able to encompass these familiar constructions together with a wide variety of other syntagmatic patterns, including broken progressions [10, 18] for which $g = fof$ (see Ex. 3).

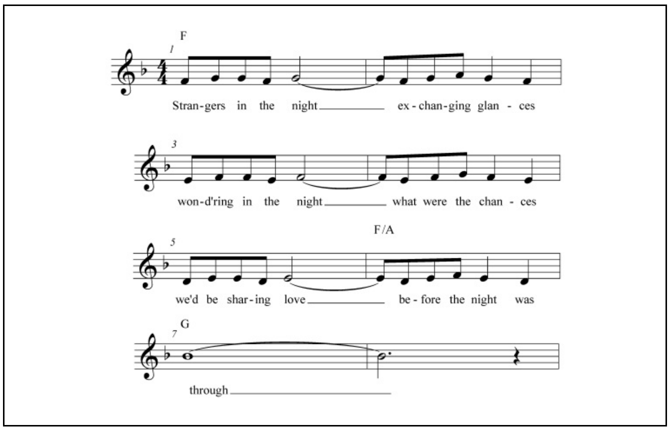
The consistency of a musical S&C ultimately results from the realization of syntagmatic patterns over several musical dimensions, even if, at the same time, some other musical dimensions may not follow any such patterns : for instance, the harmony may go $abac$, the drums $aaab$ and the rhymes $abab$, while the melody goes $abcd$. We underline once more the multi-dimensionality of the model as a very relevant property.


C♯m E F♯ C♯m
'Cause this is thrill - er thrill - er night, There
F♯ F♯
There ain't no se-cond chance a-gainst the thing with for ty eyes. You know, it's
C♯m E F♯ C♯m
You know, it's thrill - er thrill - er night. You're
F♯ A F♯ A/B C♯m
You're fight-ing for your life in-side a kill - er thrill - er to - night.

Ex.1 : an "abac" square S&C (Michael Jackson – Thriller)


D G/D
The lu - na - tic is on the grass
D G/D
The lu - na - tic is on the grass
D E/D
Re - mem b'ring games And dai-sy chains and laughs
A D
Got to keep the loo - nies on the path.

Ex. 2 : an "aabc" square S&C (Pink Floyd – Brain Damage)


F
1
Stran-gers in the night ex-chan-ging glan - ces
3
won-d'ring in the night what were the chan - ces
F/A
5
we'd be shar-ing love be-fore the night was
G
7
through

Ex. 3 : a broken progression (F. Sinatra – Strangers in the Night)

3.2.2 *Extension of the S&C model to non-square forms*

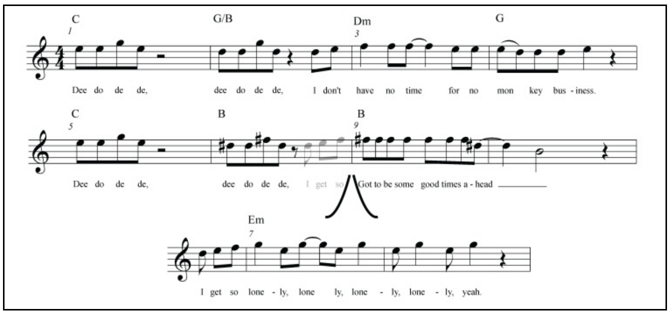
The S&C model can be generalized to systems with fewer or more MUs, by assuming missing, extraneous or redundant MUs, and/or iterated syntagmatic functions.

While dyadic systems (i.e. systems with 2 MUs) can be approached as square systems at half scale, triadic S&Cs are based on a sequence of three elements, the last of which deviates partly from the progression installed by the first two (here, only one syntagmatic function f is involved and the contrast function γ applies to $f o f$).

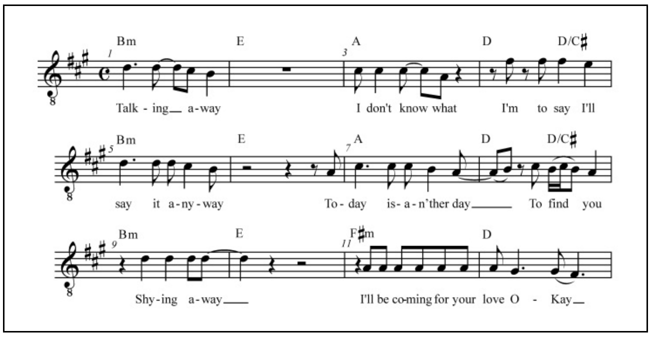
Pentadic systems, i.e. systems with 5 MUs, are treated as a *stem* square system enriched by the insertion of an additional MU, which we call an *affix*. This additional element can be either redundant with the rest of the system and/or create some sort of temporary diversion before the actual contrast (see Ex. 4).

Hexadic systems (6 MUs) combine the principles of a square S&C system with the iteration of one of the two syntagmatic relationships (f or g), creating patterns such as $a_1b_1a_2b_2a_3c$ (tall hexadic system), $a_1a_2ba_1a_2c$ (wide hexadic system) or even $a_1b_1a_2b_2cb_3$ or $a_1b_1a_2b_2c_1c_2$. Ex.5 illustrates a "tall" hexadic system.

Table 1 provides a concise inventory of system configurations with different sizes and their description in reference to the square system, together with their corresponding semiotic annotation convention, as used in the released data (right column). If needed, more details on non-square systems are provided in [18].


C
G/B
Dm
G
1
3
Dee do de de, dee do de de, I don't have no time for no mon key bus-iness.
C
B
B
5
9
Dee do de de, dee do de de, I get so Got to be some good times a-head
Em
7
I get so lone-ly, lone ly, lone - ly, lone - ly, yeah.

Ex. 4 : a pentadic S&C (4&1) (F. Mercury – Living on my Own)


Bm
E
A
D
D/C♯
1
3
Talk - ing a-way I don't know what I'm to say I'll
Bm
5
E
A
7
D
D/C♯
say it a-ny-way To-day is-a-n'ther day To find you
Bm
9
E
F♯m
11
D
Shy-ing a-way I'll be co-ming for your love O - Kay

Ex. 5 : an hexadic S&C (6T.01) (A-Ha – Take on Me)

Square system - prototypic form		
4.0	$a_{00}\ a_{01}\ a_{10}\ a_{11}$	A_0
4.1	$a_{00}\ a_{01}\ a_{10}\ \bar{\mathbf{a}}_{\mathbf{11}}$	A
Triadic sequences - based on a single SR : f. Sequence $a_0\ a_1\ a_2$ forms a progression based on two iterations of f		
3.0	$a_0\ a_1\ a_2$	$(3/4)\ A_0$
3.1	$a_0\ a_1\ \bar{\mathbf{a}}_{\mathbf{2}}$	$(3/4)\ A$
Dyadic sequences are essentially square systems at the immediately lower scale		
2.0	$a_0\ a_0$	$(1/2)\ A_0$
2.1	$a_0\ \bar{\mathbf{a}}_{\mathbf{0}}$	$(1/2)\ A$
Pentadic sequences result from the insertion of a redundant or an extraneous MU (*affix*) within a *stem* square system		
4&1 4+1 4.0:1	$a_{00}\ a_{01}\ a_{10}\ b\ \bar{\mathbf{a}}_{\mathbf{11}}$ $a_{00}\ a_{01}\ a_{10}\ \bar{\mathbf{a}}_{\mathbf{11}}\ \bar{\mathbf{a}}_{\mathbf{11}}$ $a_{00}\ a_{01}\ a_{10}\ a_{11}\ \bar{\mathbf{a}}_{\mathbf{11}}$	$(5/4)\ A$
Hexadic sequences rely on a rectangular system of properties, where either f or g are iterated, leading respectively to either a *wide* (W) or a *tall* (T) configuration. The contrast can apply to the last MU (.01), to the last but one (.10), or to both (.11).		
6W.00 6T.00	$a_{00}\ a_{01}\ a_{02}\ a_{10}\ a_{11}\ a_{12}$ $a_{00}\ a_{01}\ a_{10}\ a_{11}\ a_{20}\ a_{21}$	$(3/2)\ A_0$
6W.01 6T.01 6W.10 6T.10	$a_{00}\ a_{01}\ a_{02}\ a_{10}\ a_{11}\ \bar{\mathbf{a}}_{\mathbf{12}}$ $a_{00}\ a_{01}\ a_{10}\ a_{11}\ a_{20}\ \bar{\mathbf{a}}_{\mathbf{21}}$ $a_{00}\ a_{01}\ a_{02}\ a_{10}\ \bar{\mathbf{a}}_{\mathbf{11}}\ a_{12}$ $a_{00}\ a_{01}\ a_{10}\ a_{11}\ \bar{\mathbf{a}}_{\mathbf{20}}\ a_{21}$	$(3/2)\ A$
6W.11	$a_{00}\ a_{01}\ a_{02}\ a_{10}\ \bar{\mathbf{a}}_{\mathbf{11}}\ \bar{\mathbf{a}}_{\mathbf{12}}$	$(3/2)\ A$
6T.11 [4.0][2.1]	$a_{00}\ a_{01}\ a_{10}\ a_{11}\ \bar{\mathbf{a}}_{\mathbf{20}}\ \bar{\mathbf{a}}_{\mathbf{21}}$ $a_{00}\ a_{01}\ a_{10}\ a_{11}\ b_p\ b_q$	$(3/2)\ A$ $[A_0]\ [B/2]$
4&2	$a_{00}\ a_{01}\ a_{10}\ b_p\ b_q\ \bar{\mathbf{a}}_{\mathbf{11}}$	$A\ \&\ (B/2)$
Hexadic systems of type 6T.11 can be ambiguous with a sequence formed by a *plain* square system (4.0) followed by a dyadic sequence (2.1). The first option is preferred if there exist obvious common properties relating the six MUs, otherwise (or in case of doubt), the second option is favored.		
Some sequences of 6 MUs are in fact better described as square systems with a double affix (here denoted as 4&2).		

Table 1 : concise inventory of the main S&C configurations used to characterize the dominant inner organization of structural segments

3.2.3 *S&C model matching for segmentation*

The description of a musical passage or piece in terms of successive S&Cs can be viewed as a generalization of the grouping operation, as defined and applied by Lerdahl & Jackendoff to contiguous musical elements [20].

Indeed, elements forming a S&C share privileged relationships, the existence of which creates a sense of musical consistency within structural segments (even if not all musical dimensions participate in the system).

However, in our case, the matrix scheme assumed in the S&C model is able to account for tight relationships between elements which may not be contiguous. This is typically the case for period-like constructions $aba'c$.

As the contrast can take a very unexpected form, two elements play an essential role in identifying the boundaries of successive S&Cs : the primer of the current segment and the primer of the forthcoming one. As an illustration, let's consider the sequence of Figure 3.

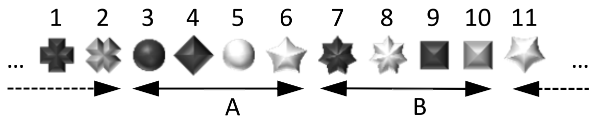


Figure 3 : A sequence of 11 morphological elements and its most likely segmentation into S&Cs

The preference for the proposed segment boundaries arises from the identification of valid primers at positions 3 and 7 while MUs at positions 6 and 10 turn out to be plausible contrasts (but are not decisive as such). In fact, MUs at position 3 and 7 are *essential* to explain economically the forthcoming MUs by simple syntagmatic relationships within the S&C framework.

As a practical consequence, two S&Cs based on the same carrier system will usually be similar over their first three quarters, but may terminate very differently (see for instance Ex. 6).

It is therefore essential to understand that the segmentation into structural blocks does not rely on the occurrence of a pre-determined musical event, but on the detection of sequences of units globally matching the generic S&C model.

3.2.4 *Law of parsimony*

The determination, within a piece, of the underlying S&Cs around a given time scale is based on the *joint estimation* of morphological units, structuring properties and syntagmatic relationships. Between several concurrent hypotheses, the one which provides the decomposition with lowest complexity is retained.

Anticipating on the labeling step can also help arbitrating between multiple hypotheses in favor of the one which is bound to lead to a more compact set of semiotic labels.

3.3 **Step 3 : clustering and labeling**

3.3.1 *S&C similarities*

Once structural blocks have been determined, the clustering and labeling steps consist in forming classes of segments whose S&C share common features, and more particularly, a similar primer and a homologous set of syntagmatic relationships.

Note that similarity is not judged on the basis of a particular property across the whole piece and/or common to all pieces (such as harmonic progression or melodic line). It is determined by considering the homology of systems formed by the structural blocks across the piece.

Two structural segments are considered as homologous in any of the following cases :

a. there exists a simple and smooth transformation between their carrier systems (for instance transposition, intensity variation, instrumental support, etc…)

b. there exists a set of common meta-properties on which their carrier systems are equivalent (for instance, the two melodic lines may not be absolutely identical but the shape of the melodic contour exhibits similar systemic variations).

c. the differences between the two systems are sporadic and/or erratic (i.e. they are viewed as small variations which do not impact the carrier system).

While systems A_0, A_0' and A_0'' in figure 4 illustrate 3 homologous (non-contrastive) systems, A_1 and A_2 are based on the same carrier system (A_0) and only differ by their contrasts. All these S&Cs are considered as belonging to a same semiotic class and are labeled with a same root symbol A, but with different sub- or superscripts.

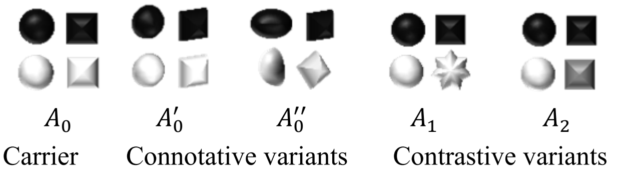


Figure 4 : examples of system variants and their annotations

3.3.2 *Neighbourhood analysis*

As music pieces often follow some regularity in their construction, two blocks surrounded with similar neighbors should be considered as potentially belonging to a same semiotic class, even if their content seems to differ at the surface level.

Indeed two structural blocks will be considered to be *a priori* more likely to belong to the same equivalence class if they appear in similar contexts in the piece, i.e. if they are located beside similar left and/or right segments within the piece. For instance, in a sequence **ABxDAByDECDCDD**, **x** and **y** are more likely to belong to the same semiotic class than in **ABxDyBCDECDCDD.** Guidance can be obtained by the reference to a *prototypical structural pattern*, as developed in [10].

3.3.3 *Proto-functional labels*

The alphabet of basic semiotic labels is composed of 25 symbols (all alphabetic letters except O). Even though semiotic labels are theoretically arbitrary, we have chosen to use *as much as possible* the symbols in correlation with the place and role of the segment in the piece.

In particular, C denotes the central element of the piece (typically, the chorus for songs), possibly followed by a second central element D, perceived as a development of C. Symbols A (resp. B) are used to denote the first (second) element before C (and E and F, after). I, J, K, L are used to denote instrumental sections acting as intros, bridges or outros. X, Y, Z are used for singular sections such as solos or middle-8s. M, N can be used for codas.

Table 3 recapitulates the functional value assigned to each basic semiotic label. In practice, only a small number of symbols is used in each piece, and a dozen of them cover almost all needs (see section 4).

3.3.4 *Intermediate labeling*

Ultimately, step 3 results in the temporary assignment of one or several basic semiotic symbols to each structural block, accounting for the potentially multiple hypotheses that the annotator is willing to consider and partially disambiguate in the next step.

3.4 **Step 4 : semiotic analysis**

This last but one step aims at consolidating the inventory of semiotic classes and at characterizing finely the variants of each segment within its assigned semiotic class. This involves the determination of distinctive properties (3.4.1), the characterization of system variants (connotative and/or contrastive) (3.4.2), and when needed, the introduction of specific notations for truncated systems (3.4.3) or multi-class segments (3.4.4).

3.4.1 *Distinctive properties*

The determination of distinctive properties aims at deciding whether variations across similar segments should be considered as the sign of their affiliation to distinct classes or as simple connotative variations within members of a same semiotic class.

Here, the convention adopted is as follows : if there exists in the carrier sequence, a well-definable property which varies in correlation with the relative position of the block w.r.t. other segments within the piece, then this property should be considered as distinctive and the segments should be assigned to two distinct semiotic classes. Otherwise, the 2 segments are considered as variants of a same class and treated as described in subsection 3.3.6.

In particular, differences between two blocks should not be appreciated in the same way if they are immediately next to each other or at some distance in the piece : a slight difference between two successive similar blocks may be distinctive (especially if this opposition is recurrent) whereas a stronger difference at a long distance may just be a connotative variation, especially if the two blocks occur in similar contexts, or if the difference is occasional.

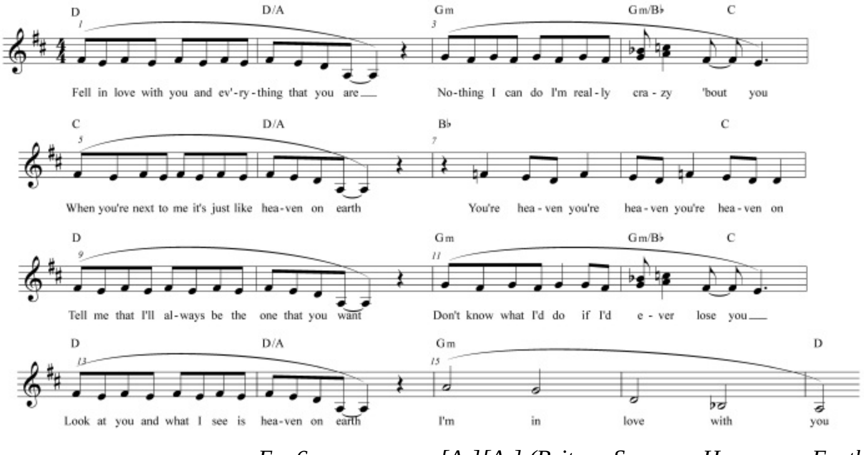
3.4.2 *Annotation of variants*

Table 2 below inventories the most frequent types of variants and their annotation.

Reference square S&C	
$a_{00}\, a_{01}\, a_{10}\, \bar{\boldsymbol{a}}_{\mathbf{11}}$	A
Contrastive variants	
$a_{00}\, a_{01}\, a_{10}\, \bar{\boldsymbol{a}}_{\mathbf{11}}^{(\boldsymbol{i})}$	A_i
$a_{00}\, a_{01}\, a_{10}'\, \bar{\boldsymbol{a}}_{\mathbf{11}}^{(\boldsymbol{j})}$ $a_{00}\, a_{01}'\, a_{10}\, \bar{\boldsymbol{a}}_{\mathbf{11}}^{(\boldsymbol{k})}$	A_j A_k
Connotative variants	
$v(a_{00})\, v(a_{01})\, v(a_{10})\, \boldsymbol{v}(\bar{\boldsymbol{a}}_{\mathbf{11}}^{(\boldsymbol{i})})$	A_i^v or A_i'

Table 2 : basic notations for contrastive (subscripted) variants and connotative (superscripted) variants of a given S&C

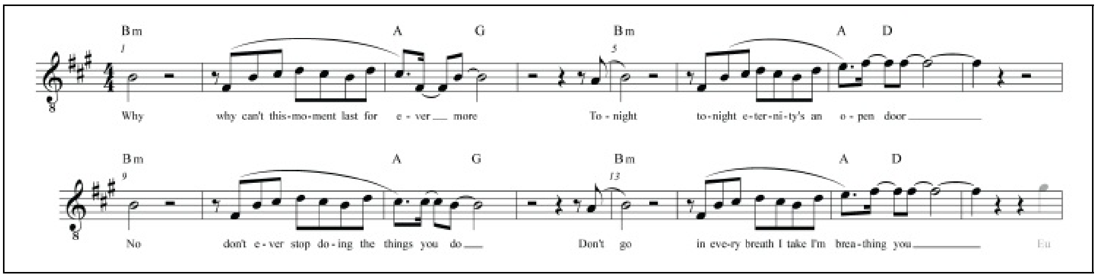
Contrastive variants are annotated by concatenating a variant index to the semiotic label (here a subscript) to reflect variations of the contrast, i.e. modifications occurring on the last quarter of the segment (snaps 13-16 of a regular square block). It is however tolerated that the contrastive variations also affect snaps 11 and 12 (second half of 3rd MU) or 7 and 8 (second half of 2nd MU), by some sort of contaminative effect. The symbol * is used to annotate exceptional forms of contrasts. Ex. 6 illustrates a contrastive variant.


D
D/A
Gm
Gm/B♭
C
1
3
Fell in love with you and ev'-ry-thing that you are
No-thing I can do I'm real-ly cra - zy 'bout you
C
D/A
B♭
C
5
7
When you're next to me it's just like hea-ven on earth
You're hea - ven you're hea - ven you're hea - ven on
D
Gm
Gm/B♭
C
9
11
Tell me that I'll al-ways be the one that you want
Don't know what I'd do if I'd e - ver lose you
D
D/A
Gm
D
13
15
Look at you and what I see is hea-ven on earth
I'm in love with you

Ex. 6 : a sequence $[A_1][A_2]$ (Britney Spears – Heaven on Earth)

Connotative variants of (say) class A are annotated as A', A'', $A^{(i)}$, or any non-numerical symbol (here superscripted) designating the type of variant : A^+ (stronger variant), A^- (weaker variant), $A^\#$ (upwards transposition), A^b (downwards transposition), $A^\sim$ (exceptional variant), … Ex. 7 illustrates a connotative variant.

Table 3 : proto-functional labels	Intro	Pre-central	Central	Post-central	Relay	Other (recurrent)	Other (sporadic)	Outro
Primary set	**I, J**	**A,B**	**C,D**	**E,F**	**J,K**	**M,N**	**X,Y,Z**	**K, L**
Secondary set	**G,H**	**P,Q**	**R,S**	**T,U**	**G,H**	**U,V,W**		**G,H**


Bm
1
A
G
Bm
5
A
D
Why
why can't this-mo-ment last for
e - ver __ more
To - night
to-night e-ter-ni-ty's an
o - pen door
Bm
9
A
G
Bm
13
A
D
No
don't e - ver stop do - ing the things you do
Don't go
in eve-ry breath I take I'm brea-thing you
Eu

Ex. 7 : a sequence [A][A'], case (c) (Loreen - Euphoria)

3.4.3 *Incomplete systems*

Incomplete systems are annotated in reference to the complete form of A, as schematized in Table 4 :

<table>
<tr><th colspan="6">Incomplete systems</th></tr>
<tr><td colspan="3">$a_{00}\ a_{01}\ a_{10}\ \cancel{\boldsymbol{\bar{a}_{11}}}$ A ...</td><td colspan="3">$\cancel{a_{00}}\ a_{01}\ a_{10}\ \boldsymbol{\bar{a}_{11}}$ ... A</td></tr>
<tr><td colspan="2">$a_{00}\ a_{01}\ \cancel{a_{10}\ \boldsymbol{\bar{a}_{11}}}$
$A\,|$</td><td colspan="2">$a_{00}\ \cancel{a_{01}\ a_{10}}\ \boldsymbol{\bar{a}_{11}}$
A ... A</td><td colspan="2">$\cancel{a_{00}\ a_{01}}\ a_{10}\ \boldsymbol{\bar{a}_{11}}$
$|\,A$</td></tr>
<tr><td colspan="3">$a_{00}\ \cancel{a_{01}\ a_{10}\ \boldsymbol{\bar{a}_{11}}}$
$(1/4)\,A$ ...</td><td colspan="3">$\cancel{a_{00}}\ \cancel{a_{01}\ a_{10}}\ \boldsymbol{\bar{a}_{11}}$
... $(1/4)\,A$</td></tr>
<tr><td colspan="6">$\cancel{a_{00}}\ a_{01}\ \cancel{a_{10}\ \boldsymbol{\bar{a}_{11}}}$ $\cancel{a_{00}\ a_{01}}\ a_{10}\ \cancel{\boldsymbol{\bar{a}_{11}}}$
... A ...</td></tr>
</table>

Table 4 : configurations and notations for incomplete systems

3.4.4 *Overlaps, collisions and ambiguities*

Overlaps, collisions and ambiguities correspond to cases where a given segment results from the combination of musical content stemming from several segments and/or distinct semiotic classes. These are quite frequent and, rather than trying to arbitrate between fundamentally ambiguous hypotheses, we handle these situations with specific notations accounting for constructions involving elements from several systems, as summarized in Table 5 (see also Ex. 8 and 9).

A particular (yet frequent) case is the mutation, illustrated on figure 5 :

Mutations correspond to singular blocks (say X) which relate to another class of blocks (say A) by sharing a subset of structuring properties with that class, while other structuring properties of A show distinct behaviors in X (including disappearance). This is typical of solos based on a verse or a chorus, in which the melodic lead behaves in a weakly structured way (see Ex. 9), or intros based on the accompaniment of another section, but with no melodic lead at all.

<table>
<tr><th colspan="2">Reference square S&Cs</th></tr>
<tr><td>$a_{00}\ a_{01}\ a_{10}\ \boldsymbol{\bar{a}_{11}}$
$b_{00}\ b_{01}\ b_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$A$
B</td></tr>
<tr><td colspan="2"><b>Mixed systems :</b> <i>i.e.</i> the segment is composed of MUs stemming from 2 classes of segments (A <b>or</b> B) – cf. Ex. 8</td></tr>
<tr><td>$b_{00}\ b_{01}\ a_{10}\ \boldsymbol{\bar{a}_{11}}$</td><td>$B\,|\,A$</td></tr>
<tr><td>$a_{00}\ b_{01}\ a_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$A \setminus B$ or $B < A$</td></tr>
<tr><td>$a_{00}\ a_{01}\ b_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$A\,|\,B$ or $B < A$</td></tr>
<tr><td colspan="2">Notation B < A is used when the mixed system is observed just after A.</td></tr>
<tr><td colspan="2"><b>Hybridation, intrication :</b> <i>i.e.</i> the segment is composed of the superposition or aggregation of 2 classes (A <b>and</b> B)</td></tr>
<tr><td>$\begin{Bmatrix} a_{00} & a_{01} & a_{10} & \boldsymbol{\bar{a}_{11}} \\ b_{00} & b_{01} & b_{10} & \boldsymbol{\bar{b}_{11}} \end{Bmatrix}$ played simultaneously</td><td>AB</td></tr>
<tr><td>$a_{00}\ b_{00}\ a_{01}\ b_{01}\ a_{10}\ b_{10}\ \boldsymbol{\bar{a}_{11}}\ \boldsymbol{\bar{b}_{11}}$
$a_{00}\ a_{01}\ b_{00}\ a_{10}\ b_{01}\ \boldsymbol{\bar{a}_{11}}\ b_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$A \,\&\, B$</td></tr>
<tr><td colspan="2"><b>Overlap (tiling), connection (hinge)</b> : two segments partly overlap or are connected by a short standalone MU</td></tr>
<tr><td>$a_{00}\ a_{01}\ a_{10} \begin{Bmatrix} \boldsymbol{\bar{a}_{11}} \\ b_{00} \end{Bmatrix} b_{01}\ b_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$[A\_][A\_B][\_B]$</td></tr>
<tr><td>$a_{00}\ a_{01}\ a_{10}\ \boldsymbol{\bar{a}_{11}}\ x\ b_{00}\ b_{01}\ b_{10}\ \boldsymbol{\bar{b}_{11}}$</td><td>$[A][\_AB\_][B]$</td></tr>
<tr><td><b>Mutation</b> : partial difference of systemic properties of B w.r.t. A – cf. Ex. 9</td><td>B/A</td></tr>
<tr><td><b>Unresolvable ambiguities</b> : i.e. undecidable hypotheses</td><td>$A\tilde{}\,?\,B$</td></tr>
</table>

Table 5 : codification of various cases of overlaps, collisions and other semiotic ambiguities

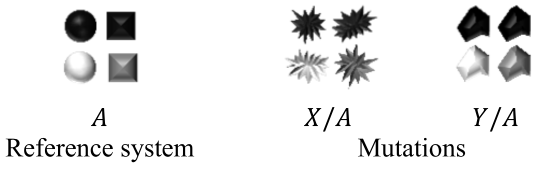
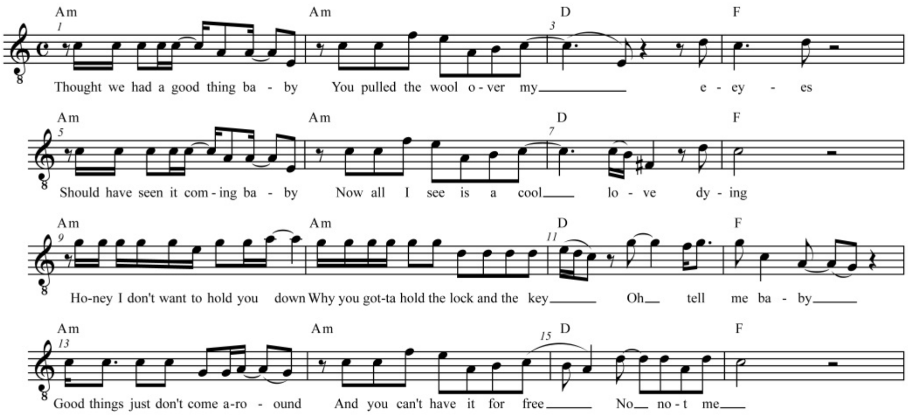
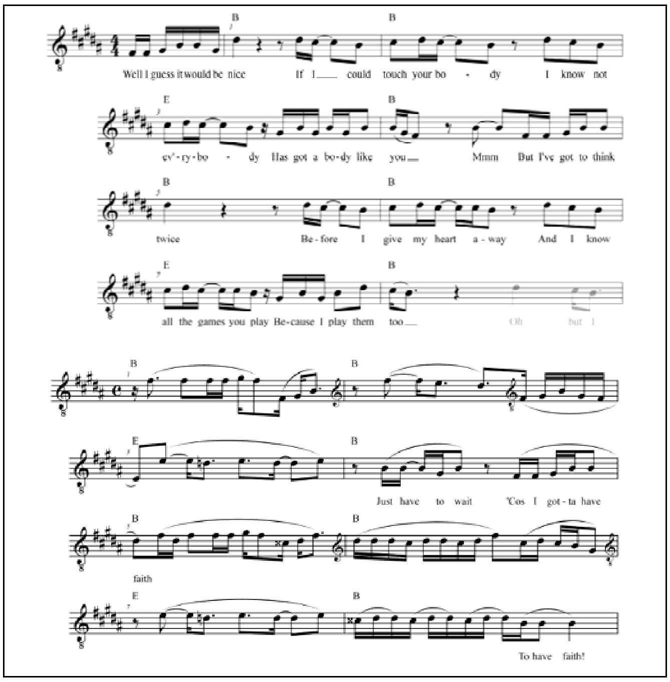


Figure 5 : two (metaphoric) examples of system mutation


Am
Am
D
F
Thought we had a good thing ba - by You pulled the wool o - ver my e - ey - es
Should have seen it com - ing ba - by Now all I see is a cool lo - ve dy - ing
Ho-ney I don't want to hold you down Why you got-ta hold the lock and the key Oh tell me ba - by
Good things just don't come a-ro - ound And you can't have it for free No no - t me

Ex. 8 : a sequence [A] [B|A] (Faith Hill – Breathe)


B
E
Well I guess it would be nice If I could touch your bo - dy I know not
ev'-ry-bo - dy Has got a bo-dy like you Mmm But I've got to think
twice Be-fore I give my heart a-way And I know
all the games you play Be-cause I play them too Oh but I
Just have to wait 'Cos I got-ta have
faith
To have faith!

Ex. 9 : an example of [A] (top) vs [X/A] (bottom).
Both segments occupy the same position as first verses in two different parts of the song. (George Michael – Faith).

3.5 Step 5 : finalization, revision, adjudication

Ultimately, the verification and harmonization of the annotations are essential steps in the finalization of the proposed process.

For the data reported on here, two annotators (also coauthors of this article) were involved : initial annotations were produced by one (or the other) annotator, and cross-checked by the second annotator. Disagreements were treated during an adjudication phase, where the two annotators had to reach a consensus on a final annotation (which, of course, sometimes led to the annotation of some segments as ambiguities and undecidable configurations).

Reviewing the annotation has also been the occasion to detect inconsistencies or unnecessary details, which were corrected and smoothed to improve the overall coherence of the annotations across the entire dataset.

4 DATA OVERVIEW

The structural annotations produced and released in the context of this work are composed of three data sets : RWC Pop (100 titles) [21], Eurovision 08-10 (124 titles) [10] and Quaero 09-11 (159 titles selected by IRCAM) [4]. A detailed list of titles is provided with the downloadable data.

4.1 Data sample

Readers who wish to explore the annotations may want to examine in priority the entries in Table 6 as representative of particular aspects of the proposed methodology.

For these songs, we first present the reduced form of the annotation, i.e. the sequence of semiotic symbols without the contrastive nor connotative indications. When several symbols are used for a same segment, this means that the extensive annotation contains a composite label with two symbols.

Below the reduced form, we provide the extensive form, which exhibits more details on the actual type of variant observed for each symbol occurrence. It is worth noting that, with a little training, these representations give a meaningful view of the musical narration process developing throughout the various songs.

Eurovision 2010 Cyp	*I A AB P C C I B P C C C C*
	$[I_1^-]\ A\ [B < A]P\ C\ C\ I'_2\ B\ [P \vert P^+]C^+C^+[C^\sim \vert C']C_*^\sim$ *No particular difficulty - clear example of B<A (B reused alone later)*
RWC Pop 55	*I A A B C D_I A A B C D_X Y C D_ED_I*
	$\left[\frac{5}{4}I^- \middle\vert I\right] A_1A_2\ \frac{5}{4}B\ C\ D\_[\frac{5}{4}I]\ A_1A_2\ \frac{5}{4}B\ C\ D\_X\ Y\ C\ D\_E \vert D\_[\frac{5}{4}I]$ *A song with a pentadic system and several instances of block tiling*
Quaero 0233	*I I J A AB P C D A AB P C D K KL*
	$[I^{--} \vert I_0^-]I\ [J^- \vert J_0]\ A_1[B \vert A_2]\ P\ C\ D\ [A^- \vert A_1][B' \vert A'_2]\ P\ C\ [D \vert D^-]\ K_0[\frac{1}{4}L/K]$ *A nice illustration of connotative symbols with no particular difficulty*
Quaero 0301	*I A B C A B C I A XC C XC*
	$[I/2]\ A_0[B/2]\ C_1A_0^+[B/2]\ C_2[I'/2]\ [A^- \vert A_1^\sim][(X/C) \vert X]\ C_1\ [XC_2^\sim]$ *A song easy to segment but with sophisticated collisions towards the end*
RWC Pop 70	*C_I A A B C A B XC Y C_C_I*
	$C_{1\_}[I/2]\ A\ A\ \frac{5}{4}B\ C_2\ A\ \frac{5}{4}B\ [(X/C)\,?\,C^\sim]\ Y\ C_{3\_}C_{1\_}I^*$ *This song starts by the presentation of the chorus*
Eurovision 2010 Isr	*I A B C B C C X YC C C*
	$[I/4]\ A\ B_1C_1\ B_2\ C_1\ C_2\ X\ [Y/C^\#]\ C_3^{\#+}[C^{\#\sim} \vert C_*^{\#+}]$ *A song showing various type of connotations with increasing intensity*
Quaero 0012	*I J A B C I A B C C J*
	$[\dots I^-]\ J\ A\ B\ C_1\ [\frac{3}{2}I^+]\ A\ B\ C'_2\ C'_*[\vert J_*]$ *A simple song to label but somehow tricky to segment at the beginning*
Eurovision 2008 Fra	*IC IE C CD_E C CD CX C_E*
	$[I_1/C]\ [I_2/E]\ C_1\ [C_2\&(D/2)]\_E\ C_1\ [C_2\&(D/2)][C_*^\sim?\,(X/C)][\frac{5}{4}C'_1]\_[\frac{5}{4}E'_*]$ *An electro-pop piece with affixes, tilings and a "polysemic" collision*

Table 6 : Samples from the semiotic structure annotation dataset

4.2 Statistical study

This section provides a statistical 'digest' of the released data. In this section, we consider annotations in their reduced form, i.e. connotative and contrastive marks are ignored. We distinguish *symbols* (single letters used to annotate) and *labels* (actual symbolic codes of segments) : in particular, a label can be composed of several symbols.

Out of 383 songs in the entire data set, 2 were judged as not fitted to the proposed conventions. Both are minimalist techno pieces by the artist *Plastikman*, from the same album (year 1994), for which it was considered as impossible to determine reliable structural segment boundaries. Both songs were therefore annotated as a single segment, labeled with an exclamation mark.

4.2.1 *General statistics*

Table 7 summarizes global statistics over the entire dataset.

Number of annotated songs	383 songs
Total duration	82418 seconds
Number of segments	5552 segments
Alphabet of symbols	25 symbols (+ !)

	mean	median
Song duration (in seconds)	215	201
Duration of segments (in sec.)	14.8	14.5
Number of segments / song	14.5	14
Nb of distinct symbols / song	5.7	6

Table 7 : general statistics of the QUAERO/METISS annotations

The distribution of song durations (not represented here) shows a marked peak around 180 s, corresponding to the maximul duration of Eurovision songs, set to 3 minutes by the contest rules.

4.2.2 *Segment duration*

The distribution of the absolute segment duration across the 383 songs behaves as depicted on Fig. 6, with a mean value around 15 s and a standard-deviation of approx. 6 s.

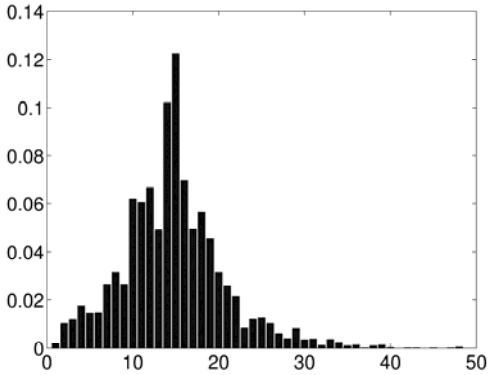


Figure 6 : histogram of absolute segment durations (in seconds) in the QUAERO/METISS structural annotations

This distribution illustrates clearly the fact that the value of 16 s for typical segment duration is only an *a priori* target but that actual durations can significantly deviate from the target value.

When normalized by the song's median segment duration on a song-by-song basis, the *relative* segment duration shows a very concentrated distribution, as is visible on Fig. 7.

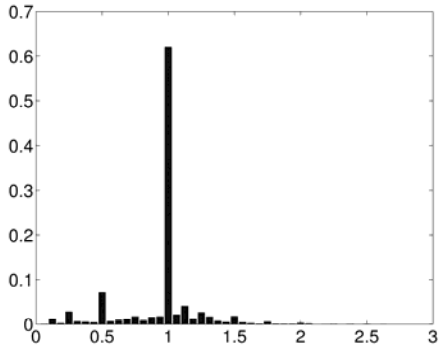


Figure 7 : histogram of the relative segment duration

In fact, 71 % segments deviate from no more than 1/8th of the song's median segment duration, which is a direct consequence of the use of a target time scale for each song.

Local maxima are visible at relative durations $1 \pm {}^{k}/_{4}$, corresponding to dyadic (0.5), triadic (0.75), pentadic (1.25) and hexadic (1.5) systems, when the median duration segment is a square system (as is often the case).

4.2.3 *Symbol distribution*

Table 8 reports the proportion of songs as a function of the number of distinct symbols in their annotation (mean 5.7). Most songs (i.e. 85 %) are annotated with 4 to 8 symbols.

# symbols	1	2	3	**4**	**5**	**6**	**7**	**8**	9	10+
% songs	1	2	5	**15**	**27**	**22**	**13**	**8**	5	2

Table 8 : relative distribution of the number of distinct symbols across songs

Table 9 indicates the proportion of songs in which a given symbol is occurring at least once. In that respect, the top seven symbols are : A, B, C, D, I, J and X.

Symbol	**C**	**I**	**A**	**X**	**B**	**D**	**J**	K	Y	H	E	P	others
% songs	**100**	**89**	**88**	**57**	**57**	**38**	**36**	24	17	12	12	8	≤ 5

Table 9 : rate of usage of the various symbols across songs

Figure 8 displays the number of occurrences of labels in the data set as a function of their frequency rank. In this case, labels resulting from the combination of primary symbols are considered as distinct from one another. In log-log scale, the plot exhibits an almost linear behavior, characteristic of a Zipf-like Law.

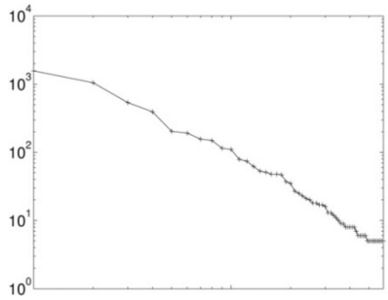


Figure 8 : number of label occurrences in the annotations as a function of their frequency rank (log-log plot)

Table 10 indicates the proportion of labels composed of 1, 2, 3 or more symbols. About 4/5 of the labels are mono-symbolic, and most of the rest are di-symbolic (i.e. described as combinations of 2 symbols).

Symbols per label	**1**	**2**	3	4+
Segments (% total)	**80.1**	**18.6**	1.1	0.2

Table 10 : proportion of n-symbolic labels in the annotations

Note also that undecidable labels (i.e labels including a question mark) represent 1.33 % of the entire population of segments, and are found in 12.5 % of the 383 songs.

Figure 9 depicts the number of symbol occurrences (weighted by their relative occurrence in each label, i.e. 1 for mono-symbolic labels, $^1/_2$ for di-symbolic ones, $^1/_3$ for tri-symbolic, etc…).

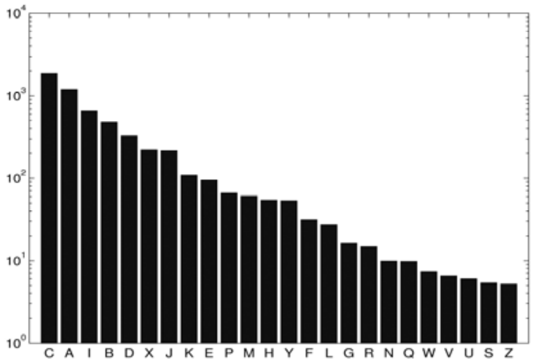


Figure 9 : number of symbol occurrences in the annotations as a function of their frequency rank (log-lin plot)

The histogram exhibits a rather constant slope in log-linear scale, indicating that the distribution behaves as an exponential function of the rank. In average, each symbol is about 25% less frequent than its previous one in the ranked list. Here too, the top 7 symbols are C, A, I, B, D, J and X, which covers almost 90 % of the entire set of segments. Adding K, E, P, Y, M and H brings the coverage to 97.5 %.

4.2.4 *Syntactic dependencies*

Finally, we estimated unigram and bigram models on the sequence of symbols occurring in the annotations, and used these zero- and first-order models to evaluate the entropy of each subset of annotations, reported in Table 11 in terms of perplexity figures (i.e. $2^{Entropy}$).

Data set	**RWC Pop**	**Eurovision**	**Quaero**	**All pooled**
Unigram	8.0	7.5	9.8	8.8
Bigram	4.3	4.6	5.9	5.4

Table 11 : zero- and first-order perplexities of the annotated data

These results indicate that the annotations exhibit definite first-order syntactic regularities and that the structural patterns of the Quaero songs are slightly less predictable than that of RWC Pop and Eurovision, probably because of a broader coverage and diversity of the Quaero subset.

5 CONCLUSIONS

Describing the organization of music pieces in terms of similarity relationships and structural patterns is unquestionably a desirable yet challenging objective in the domain of semantic audio and music information retrieval. The Quaero/Metiss annotation effort has been the occasion to deeply investigate concepts underlying music structure and to develop an expertise on the various dimensions involved in its description.

This initiative has led to the production and release of *semiotic* annotations for more than 380 songs stemming from 3 distinct datasets, following a well-defined methodological process and a clear set of conventions, as described in this article.

The statistical overview of the released annotations shows that they comply well with the methodology and provide, in that sense, a consistent set of resources. Given their public availability, we expect them to contribute in a near future to reproducible (and hopefully fruitful) research in various areas of music science and technology. Indeed, they have been used for evaluation at the end of the Quaero project, and a subset of a former version of these data was introduced in the MIREX evaluations, in 2010.

Readers of this article will probably wonder about the time taken to annotate the 383 songs with the proposed conventions. It is indeed hard to answer this question, because in the present case, the process of resource production has been spread over time and closely intertwined with that of scientific investigation, yielding simultaneously the annotated data and the reported methodology. Nevertheless, we believe that a reasonable order of magnitude for trained annotators would be to allow for 45 minutes annotation time per song broken down as 20 minutes for the primary annotation, 10 minutes for the cross-check phase, 2×5 minutes for adjudication and 2×2.5 minutes for data handling. Based on a typical song duration of 3-4 minutes, this means between 10 and 15 times the duration of the audio material.

Beyond the Quaero/Metiss resource production effort, we believe that the proposed concepts for characterizing systemic relationships inside musical segments and codifying segment similarities within music pieces opens new perspectives towards a more robust definition, extraction and exploitation of structural metadata in semantic audio applied to music signals.

REFERENCES

[1] Mixture of several definitions from various sources.

[2] J. Paulus, M. Müller, A. Klapuri, Audio-based music structure analysis. *Proc. ISMIR 2010.*

[3] SALAMI Project : http://salami.music.mcgill.ca

[4] QUAERO Project : http://www.quaero.org

[5] G. Peeters, K. Fort : Towards a (better) definition of annotated MIR corpora. *Proc. ISMIR 2012*.

[6] MIREX : http://www.music-ir.org/mirex

[7] G. Peeters, E. Deruty : Is Music Structure Annotation Multi Dimensional ? *LSAS, Graz (Austria) 2009.*

[8] J.B.L. Smith, J.A. Burgoyne, I. Fujinaga, D. De Roure, J.S. Downie : Design and Creation of a Large-Scale Database of Structural Annotations. *Proc. ISMIR 2011.*

[9] F. Bimbot, O. Le Blouch, G. Sargent, E. Vincent : Decomposition into Autonomous and Comparable blocks : a Structural Description of Music Pieces. *Proc. ISMIR 2010.*

[10] F. Bimbot, E. Deruty, G. Sargent, E. Vincent : Semiotic structure labeling of music pieces : concepts, methods and annotation conventions. *Proc. ISMIR 2012.*

[11] Snyder B. : *Music and Memory*, M.I.T. Press, 2000.

[12] Ten Minute Master n°18 : Song Structure - *Music Tech Magazine,* October 2003, pp. 62-63.

[13] J. Foote. Automatic audio segmentation using a measure of audio novelty. *IEEE ICME*, pp. 452–455, Aug. 2000.

[14] J. Paulus. Signal Processing Methods for Drum Transcription and Music Structure Analysis. *PhD Thesis,* 2009.

[15] M. Müller and F. Kurth. Towards structural analysis of audio recordings in the presence of musical variations. *EURASIP Journal on Advances in Signal Processing*, 2007.

[16] G. Sargent, F. Bimbot, E. Vincent : A Regularity-Constrained Viterbi Algorithm and its Application to the Structural Segmentation of Songs. *Proc. ISMIR 2011.*

[17] J.-J. Nattiez : *Musicologie générale et sémiologie*, Christian Bourgois Ed., 1987.

[18] F. Bimbot, E. Deruty, G. Sargent, E. Vincent : System & Contrast : a Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces (Extensive Version). *IRISA Internal Publication PI 1999*, December 2012.

[19] E. Narmour : Music expectation by cognitive rule-mapping. *Music Perception*, XVII/3 (2000), p. 329-398.

[20] F. Lerdahl, R. Jackendoff : A Generative Theory of Tonal Music, MIT Press, 1983 (reprinted 1996).

[21] RWC : http://staff.aist.go.jp/m.goto/RWC-MDB

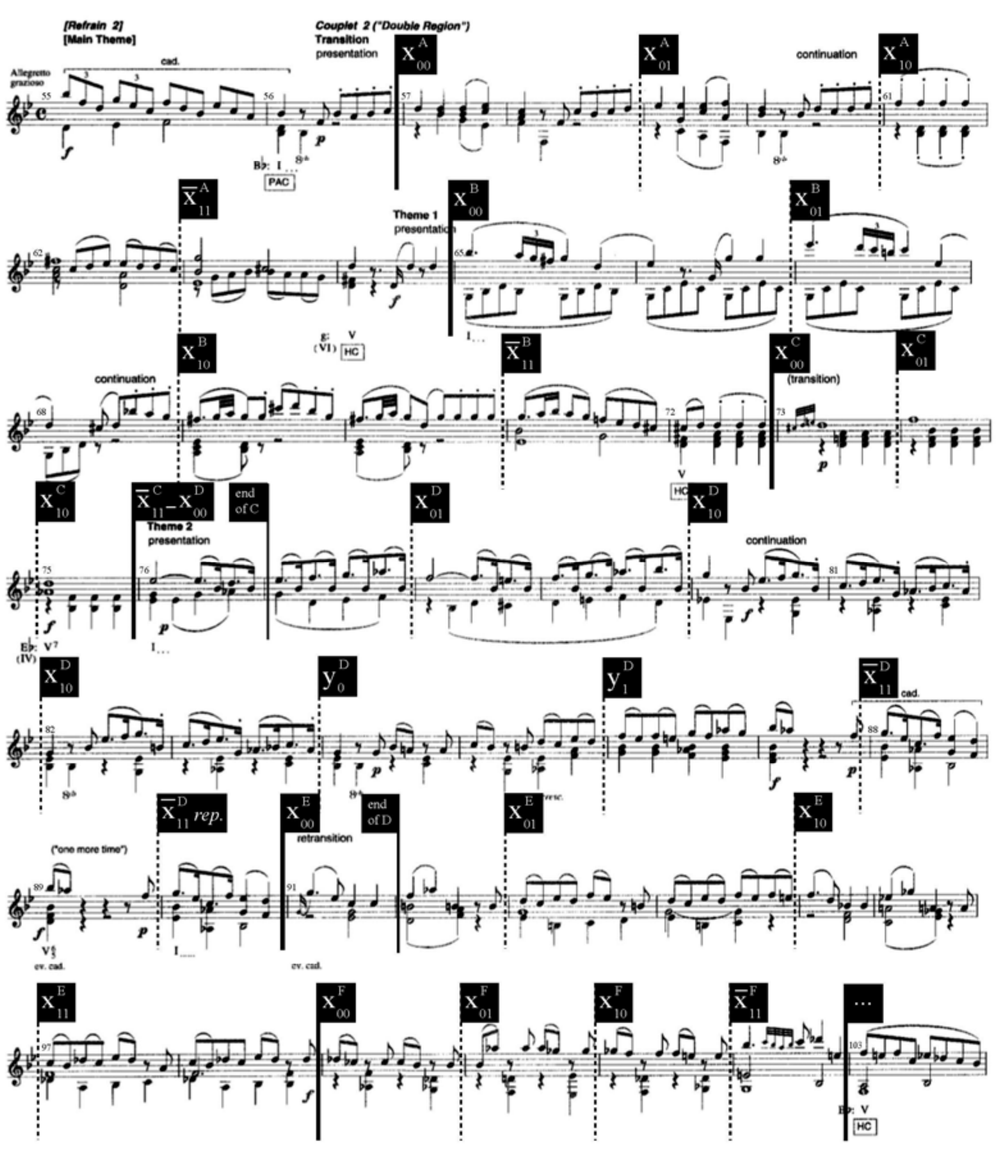
[22] W.E. Caplin : Classical Form : A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven. *Oxford University Press, New York*, 1998.

Acknowledgments : this work was partly funded by OSEO in the framework of the Quaero project.

ANNEX

In this example, we annotate an entire passage from Mozart's piano sonata in B-Flat, K333/315c, third movement into successive S&Cs. The example is borrowed from [22], p. 240, who uses it as an illustration for a particular part of the *sonata-rondo* form.

EXAMPLE 16.8 Mozart, Piano Sonata in B-flat, K. 333/315c, iii, 55–102



In this passage, 6 successive S&Cs can be identified. They are denoted A-F, and their main features are as follows :

Segments A and B are standard square S&Cs dominated by a sentence-like structure

Segment C is a shorter S&C at a half time scale acting as a fast transition and tiled with the next segment over 1 bar.

Segment D is a sophisticated 8-element segment whose inner organization can be decomposed as $[\boldsymbol{a_1 a_2 b} bxyc_0\boldsymbol{c}]$, i.e. a period-like square stem (elements in bold font) and 4 affixes : 1 redundant MU (the second occurrence of b), 2 nested MUs (xy, forming a sub-system) and a pre-contrast (c_0).

Segment E, which is tiled with D over 1 bar, is a quasi non-contrastive S&C exhibiting a period-like pattern while F is a half-time scale S&C, and it can be argued that EF forms an hexadic system of type 6T.11.