Methodology and conventions for the latent semiotic annotation of music structure

Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, Emmanuel Vincent

Publications Internes de l'IRISA
ISSN : 2102-6327
PI 1993 – février 2012

Methodology and conventions for the latent semiotic annotation of music structure

Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, Emmanuel Vincent

Equipe METISS, Département DSIR

Abstract: This document proposes and describes guidelines for the annotation of music structure into elementary units called semiotic blocks, reflecting high-level and high-scale similarities across the music piece.

The annotation process requires both the localization of block boundaries (segmentation) and the determination of block classes (labeling).

The proposed methodology is based on concepts and principles inspired and adapted from semiology. It combines elements resulting from the analysis of internal properties forming systems within blocks (morphological analysis), from the characterization of distinctive properties across blocks (paradigmatic analysis) and from the location and context of blocks within the music piece (syntagmatic analysis).

The implementation of these principles has resulted in a set of notations which are also developed in the document.

Méthodologie et conventions pour l'annotation sémiotique latente de structure musicale

Résumé : Ce document propose et décrit un cadre méthodologique pour l'annotation de structure musicale sous forme d'unités élémentaires dénommées blocs sémiotiques, reflétant les ressemblances de haut-niveau et à grande échelle dans un morceau de musique.

Le processus d'annotation consiste à la fois à localiser les frontières de blocs (segmentation) et à déterminer leurs classes d'équivalence (étiquetage).

La méthodologie proposée repose sur des concepts et des principes inspirés et adaptés de la sémiologie. Elle combine des éléments d'analyse résultant de l'étude des systèmes de propriétés internes au sein des blocs (analyse morphologique), de la caractérisation des propriétés distinctives entre blocs (analyse paradigmatique) et de la position et du contexte des blocs dans le morceau (analyse syntagmatique)

La mise en oeuvre de ces principes a débouché sur un système de notation, lui aussi développé dans ce document.

Methodology and conventions for the latent semiotic annotation of music structure

Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, Emmanuel Vincent

Rapport Interne IRISA - Février 2012

Summary

This document proposes and describes guidelines for the annotation of music structure into elementary units called semiotic blocks, reflecting high-level and high-scale similarities across the music piece. The annotation process requires both the localization of block boundaries (segmentation) and the determination of block classes (labeling).

The proposed methodology is based on the following key ideas, inspired and adapted from semiology :

1. The localisation of block boundaries results from the identification of complete and locally autonomous morpho-syntagmatic systems (carrier system + contrast + affixes) defined on one or several musical information layers (morphological analysis).
2. The determination of block labels is achieved by identifying distinctive semiotic properties across morpho-syntagmatic systems (paradigmatic analysis) and by taking into account the position and context in which the blocks are observed, in reference to a structural pattern (syntagmatic analysis).
    a. Recurring and systematic differences across music blocks are prone to be semiotically distinctive while occasional and surface differences are preferentially considered as expressive variations of a same semiotic element.
    b. When the semiotic structure appears to derive from a clear structural pattern, blocks occurring in similar contexts are more prone to belong to a same semiotic class.
3. While the semiotic structure of music tends to be based on recurrent components of the musical discourse, occasional irregularities and singularities are commonly observed in the realization of music structure. They tend to be more frequent and more spectacular towards the end of the piece.
4. The existence of class ambiguities is considered as inherent to the semiotic structure of (some) music pieces and specific symbols are designed to render them accurately.

The implementation of these principles has resulted in a set of guidelines and notations which are developed in the document.

---

1. Presentation

This document presents principles[1], notations and methodological guidelines designed at IRISA for the description and annotation of *the latent semiotic structure* of musical pieces.

Music structure, as understood in this work, does not correspond to the acoustic structure (who/what plays/sings when) nor to the functional structure (verse / chorus / bridge).

The *latent semiotic structure* reflects high-level and high-scale systematic relationships and distinctive properties across segments of comparable size (and of same-scale morpho-syntagmatic organization) within the music piece[2].

2. General concepts

The semiotic description and annotation process consists in representing the music piece as a sequence of labels (transcription) from a limited set of symbols called (semiotic) alphabet. Each occurrence of the labels corresponds to a segment of the musical piece and each symbol represents metaphorically a class of musical segments.

Structural blocks constitute the elementary units to describe the semiotic structure. They are primarily identified by considering morpho-syntagmatic patterns resulting from local relationships on one or several musical information layers

Block boundaries are located at instants when morpho-syntagmatic patterns are re-actualized, under regularity assumptions.

Each structural block is assigned a semiotic label which denotes :

(i) Properties which are common to segments of a same class within the piece (one label per class)
(ii) Properties which are distinctive across classes (two different labels for two distinct classes)
(iii) Variations of the ending of the morpho-syntagmatic pattern of the block (subscripts)
(iv) Variations of surface properties which are considered as expressive and which develop over the semiotic structure (superscripts) but which do not affect the morpho-syntagmatic pattern of the block

Some specific semiotic labels are used to render situations when structural blocks show properties which are clearly ambiguous or undecided between two classes.

The annotation process consists in :

- Identifying which musical layers are relevant to the structural description of the piece
- Determining which musical layers are relevant to locate structural block boundaries (structuring information layers)

[1] The fundamentals behind these principles will be subject to a separate document.
[2] A future version of this document will contain a brief survey of previous attempts to define music structure

- Determining which properties of these layers are relevant to identify the alphabet of semiotic classes (distinctive semiotic properties)
- Transcribing the piece with this alphabet (labels, subscripts and optionally, superscripts)
- Identifying a structural pattern to which the structure of the piece is assumed to correspond.

Trying to achieve all these tasks on the only basis of the musical piece is an ill-posed (under-determined) problem.

Assumptions are introduced to reduce the indetermination of the solution, which boils down to considering that :

- Components of the high-level musical process which are periodic, cyclic, regular, recurrent, repeated, synchronized, correlated, quantizable, etc… are prone to be linked to the structural level of the process.

- Components of the high-level musical process which are aperiodic, spurious, occasional, unique, asynchronous, erratic, continuous, etc… are less likely to pertain to the structure.

This general dichotomy must not be used to decide blindly on the nature of an observation, but rather as a guideline to orientate investigations and arbitrate situations in combination with a careful analysis of the actual musical content.

Note that, considering that a particular element or layer of musical information is not relevant to the structural level does not mean that it is not important to the musical piece. It simply means that it should be considered as relevant to another level of the musical discourse, which we call the expressive level.

As a result of the aforementioned constraints, the annotation process is based on some optimal compromise between the following objectives :

1. The largest possible number of musical layers should be encompassed in the structural description
2. The choice of labels should render as accurately as possible the distinctive properties of semiotic information layers
3. The semiotic alphabet should be reasonably compact (in terms of number and distribution of symbols)
4. The transcription should be as regular as possible and relate as much as possible to a simple structural pattern

Semiotic blocks are defined at a high time-scale (typically 16 seconds). We therefore consider the downbeat (or snap) as the unit of time at which the musical content is sampled at the “low” level (typically 1 s).

3. Musical information layers

The musical discourse is based on a variety of processes and properties which operate at different levels of description.

The musical content generally results from the activity of a number of *musical sources*, such as :

- the main lead (voice, main instrument, …)
- the accompanying instruments
- the lyrics (sung and/or spoken)
- percussions
- arrangements
- sound effects
- musical gimmicks
- etc…

Musical sources, can be viewed as the concrete vehicles of the actual musical content of the piece as a continuous signal.

These various musical sources contribute to what we call *musical information layers*, such as :

- melody (horizontal succession of notes)
- harmony (vertical combination of notes)
- rhythm (relative duration of notes, chant, …)
- phonetic stream (rhymes, alliterations,…)
- accentuation (energy focus)
- effect patterns
- etc…

Musical information layers refer to the organized evolution followed by abstract musical properties, which can generally be coded as discrete observations from a quantized set of configurations.

Note that a source signal does not constitute *per se* an information layer, but that the behavior of the source (i.e. the organized pattern created by its activity) does. Conversely, several sources can contribute (simultaneously, successively or alternatively) to a same information layer.

To complete the overall picture, we may also consider *macroscopic properties* of the piece (or a section of the piece) when dealing with :

- tonality (key, reference tonic)
- modality (scale)
- tempo (execution pace)
- language (reference phonetic system of the lyrics)
- volume (execution strength, number of active instruments)
- timbre (nature and properties of the voice and/or the instruments)

Macroscopic properties are usually evolving slowly over the whole piece and can usually be inferred from the long-term behavior of musical sources and/or the time-integration of musical information layers.

4. Definition of structural blocks and semiotic labels

Semiotic blocks are defined by time boundaries and a label denoting an equivalence class. Their localization and their characterization relies on the internal (morpho-syntagmatic) organization of one or several of their musical information layers.

Definition of blocks boundaries relies on what we call *structuring* properties. Structuring properties participate to the cohesion of the musical content within the blocks.

In a second step, *semiotic* properties are considered for characterizing and comparing blocks in the song[3].

Not all structuring properties act as semiotic properties (and vice-versa) : some properties may strongly contribute to identify block boundaries but not be relevant to define classes of block, and conversely.

4.a – Morpho-syntagmatic properties of blocks

We assume that a block is composed of a stem (usually 16 snaps) and a number of additional (suppressible) elements called affixes [1,2]. In most cases, a stem decomposes into 4 morphological units (called sub-blocks) of 4 snaps each[4]. We further define the first three sub-blocks of the stem as the *carrier system* and the last one as the *contrast* (sometimes, a "flat" contrast)[5].

The (usually simple) network of relationships between the sub-blocks form a graph which represents syntagmatic relations within the block and are the basis of its internal musical consistency. These relationships vary from pure identity to translation, transposition, inversion, conjugation, …

A structural block is therefore characterized by the properties of its morphological units and by a graph which internally relates these units. These systems exist on several musical layers in different forms and at different scales simultaneously, which indeed further contributes to the musical consistency of the segment. The layers taking part to the morpho-syntagmatic cohesion are called structuring layers.

We admit here that a musical information layer is considered as a structuring layer if the morpho-syntagmatic relationships observed in the block stem for this particular layer is affiliated to one of the following morphological patterns :

- aa'a"c (c can be equal to a)
- aba'c (c can be equal to b (or to a))
- aa'bc (c can be equal to b (or to a))

In all 3 cases above, elements a' and/or a" are considered as transformations or variants of element a (including identity).

When some well-defined property of a musical information layer follows one of these patterns at a certain time scale, it constitutes a structuring property, i.e. it induces locally the existence of a syntagmatic cohesion at that scale. The set of musical information layers which follow one or another of these morphological

[3] The distinction between *structuring* and *semiotic* properties is essential.
[4] Conceptual generalization to non-square stems is rather straightforward.
[5] Corresponding publication is in preparation

patterns form the set of structuring properties (one layer may follow a particular pattern and simultaneously another layer follow another pattern).

Note that a constant layer (i.e. a layer for which the behavior is identical on all the sub-blocks of the carrier system, i.e. a pattern aaaa) is not necessarily a structuring property. However, it is bound to acquire the status of structuring property if its apparition and/or disappearance is synchronized with the reactualization of other morpho-syntagmatic patterns.

4.b – Localization of structural blocks (segmentation)

Similarity relationships also occur at a longer distance in the piece, but in general it is possible to identify, at the level of the entire piece, a typical (time-)*radius* above which the density of relations between morphological units becomes irregular and decreases rapidly. This induces a structural pulsation and is the basis for the definition of segment boundaries, beyond which similarities can be somehow "ignored" without hurting the local musical consistency.

As a consequence, structural blocks appear as self-sufficient musical units, i.e. they possess the property of iterability (see former work [1] [2]). They also can be removed from the piece with no severe effect on the rest of the piece, hence the property of suppressibility.

A key element to detect block boundaries is the presence of a contrast, i.e. an element "c" in the morpho-syntagmatic system which departs from the logical sequence induced by the first three elements of the carrier system. This acts simultaneously as a logical modulation and as some sort of punctuation mark which announces an imminent border (located after the contrast)[6].

4.c – Characterization of structural blocks (semiotic properties)

Internal systems forming blocks are refreshed regularly in the following manners :

1. by presenting a new occurrence of a previously observed morphological unit
   a. either for introducing what will appear to be a variant of an already known system
   b. or to expose a new system which partly re-uses material from a previous system
2. by introducing significantly new musical material (which means, the beginning of a totally new system)

For each musical layer, a structural block can be described, by :

1) The structuring property that participates to the system
2) The network of relations between the 4 sub-blocks composing the stem

Some of these attributes constitute *semiotic* properties.

A semiotic property is defined as a property which contributes to the description of the equivalence classes on which the semiotic alphabet is based. In general, semiotic properties are a subset of the structuring properties but they also happen to be defined as *meta-properties* of the structuring properties.

[6] This will be detailed in a separate publication.

As an example, here is a non-exhaustive list of possible semiotic properties of the layer “melody” :

- the melody itselfs
- the support notes of the melody
- the rhythm of the melodic line
- the variations of the melody
- the energy fluctuations caused by the melody
- the placement of the melody w.r.t. downbeats
- etc…
-

These various properties of the melody may be used to explain latent similarities despite surface differences between two segments : for instance, the melody can be distinct but the support notes may turn out to be identical.

Strong and accurate formulation of semiotic properties will be preferred (namely because they convey more information on the actual behavior of the musical layer). But sometimes a weaker formulation will be relevant, when seeking for common properties which are bound to explain the affiliation of two blocks to a same class (or a same root).

Determining block labels can be approached as a standard task of semiotic analysis which can be addressed by the commutation test which involves substituting or exchanging blocks with one another and decide whether this substitution changes the perception of the overall “narrative” organisation of the song.

However, the approach based on commutation test is prone to divergences of interpretation between listeners and we prefer to formulate the task of assigning semiotic labels to structural blocks as a matter of comparing morpho-syntagmatic systems by analyzing what makes them different but also what they have in common, beyond their immediate surface properties (paradigmatic analysis), but also by taking into account their context in the song (syntagmatic analysis).

This is detailed in the forthcoming two sections.

4.d – Paradigmatic comparison of structural blocks

The semiotic comparison of blocks is based primarily on the comparison of their internal morpho-syntagmatic carrier system. For the purpose of explaining this point, let’s denote as X and Y two instances of structural blocks.

For a given music information layer, the carrier systems of X and Y are considered as equivalent if and only if, there exist a property of the considered layer for which the corresponding network of internal relationships of X and Y are homologous.

This comparison can then be carried out for all structuring properties and, if the two blocks are somehow similar, a subset of common properties will emerge as being potentially characteristic of a class grouping X and Y.

At the same time, some of these common properties may also be common to other blocks which however are unlikely to belong to the same class because they strongly differ from X and Y with respect to other structuring layers.

The determination of the subset of structuring properties (and the corresponding network of internal relationships) which are strictly relevant to the semiotic classification of structural blocks is called the paradigmatic analysis and it aims at defining the distinctive properties of the semiotic structure.

4.e – Syntagmatic comparison of structural blocks

The position and context of the structural blocks in the transcription are also taken into account in order to guide their membership to a given class. In other words, two blocks are more likely to belong to the same equivalence class if they appear in similar contexts in the piece.

Blocks are considered to occur in a similar context if they are located beside similar left and right segments within a recurrent sequence. Context definition can be relaxed to left-only or right-only situations and similarity can be understood more or less strictly (for instance, be extended to "functional" similarity).

Therefore, structural blocks which have a similar neighborhood are more likely to be considered from a same class (or at least, one of them to be rooted on the other – see section 7), even if they have quite different surface properties.

It is also worth considering that differences between blocks should not be appreciated in the same way if the two blocks are immediately next to each other or at some distance in the piece : a slight difference observed between 2 successive similar blocks may be distinctive (especially if this opposition is recurrent in the piece) whereas a stronger difference at a long distance may be just an expressive variation, especially if the two blocks are surrounded by a similar context.

Therefore, the guidance of a structural pattern (see section 6) is considered as essential to weigh similarities and differences between and across blocks and interpret those with respect to the global organization of the piece.

However, while the semiotic structure of music tends to be based on recurrent patterns, occasional irregularities and singularities are commonly observed in the realization of music structure. Those tend to be more frequent and more spectacular towards the end of the piece.

5. Notations for semiotic labels

5.a – Primary notation of semiotic labels (transcription)

The choice of the semiotic label of a block is primarily guided by the following considerations :

- When two blocks are from two distinct classes (i .e. when they are considered to be built on 2 distinct carrier systems) they are denoted by 2 distinct alphabetic capital letters : A vs B.
- When two blocks show equivalent carrier systems but different contrasts, this difference is noted as a subscript : A_1 vs A_2. Note that suffixes and infixes within the last sub-blocks are treated as specific forms of contrasts. Blocks showing no (or extremely weak) contrasts on any layer will be denoted A_0. Blocks showing exceptional contrast (usually resulting of combinations of several factors) will be denoted as A_* (they usually tend to occur at the end of the piece).
- When two blocks are considered as equivalent, i.e. when their system+contrast are equivalent but they differ in their (surface) realizations, the variants are denoted with a superscript, for instance : A'

vs A''. Optionally, the superscript may be chosen specifically to indicate the nature of the variant (see extended conventions in section 8.a). The notation A^0 and A^* can be used to denote exceptionally poor or exceptionally rich variants of A.

- When a property or a set of properties disappear gradually or degrades progressively, this will be denoted as a fade-in or a fade-out. This may apply to surface properties such as the intensity $^<A$, $A^>$, the instrumentation support, but also to strengthening or vanishing properties of the carrier system, then denoted $_<A$, $A_>$.
- If a block is realized only in a half form, specific notations are used : A/2 for a block which is half the size of the structural pulsation period but which forms a complete, autonomous system, |A and A| for truncated half blocks corresponding to the right (resp. left) part an entire block A, (usually realized in full form somewhere else in the piece). More general truncations will be denoted as … A or A …, depending on the side of the truncation, and if necessary, accompanied by an indication of the fractional size of the segment, for instance : (3/4) A … which indicates a block based on a square system which has been truncated after the third morphological unit.
- Inevitably, some situations may reveal ambiguities[7]. Therefore, a set of additional notations are designed to render these ambiguities with <u>composed labels</u> : AB, B/A, B|A, B>A, A&B, etc… They are described in details in section 7.

Summary of main notation conventions

	Regular	Minimal	Exceptional	Fade-in Fade-out	Fractional size	Truncated blocks
System variants (subscripts)	A_1, A_2, … A_i, A_j	A_0	A_*	$_<A$ or $A_>$	A/2 or (1/2) A A/4, (3/4) A	|A or A| … A or A …
Surface variants (superscripts)	A', A'', … $A^{(i)}$, $A^{(j)}$	A^0	A^*	$^<A$ or $A^>$		

5.b – Hypermetric notations (size)

Semiotic labels can be combined with size (hypermetric) information in a second field, in brackets. For instance [16+4], [-2+16], [16-4], [16&2], etc… following conventions previously introduced to denote block sizes [1].

It is also admissible to describe the hypermetric information as an operation reflecting the internal organization of the morphological elements, for instance : [4x4], [2x4x2], [6x4],[3x4(+4)+4], etc[8]…

By default, the block size will be considered as equal to 16 snaps.

[7] This is indeed a natural outcome of the semiotic strategy : creating an alphabet of elements, getting the listener to learn them and ultimately **combining** these elements to create hybrid objects (some sort of chord, at the level of the semiotic structure).

[8] These conventions will be detailed in a future document

5.c – Internal morpho-syntagmatic system

A third field, between parenthesis, can be used to denote the internal morphology of the block (usually referring to the semiotically dominant musical information layer). For instance (abac), (aabc), (aba'c), (aa'a"c), etc[9]…

5.d – Comprehensive description

In minimalistic situations, the semiotic annotation will therefore consist of a capital letter and a simple subscript and/or a simple superscript, or both, for instance : A_2, A', $A"_1$, $A°_{>,}$ etc…

Semiotic symbols can be put into brackets to facilitate visual parsing, especially when they are complex :

$[A_0]$ $[A_1]$ $[A_2]$ [B] [C] $[A'_1]$ $[A'_2/2]$ [B'] [C'] [|C'] [D] $[D*^{>}]$

In more extensive descriptions, a semiotic label may look like this : A_2 [4x4+4] (abacc) and the description of the piece will preferably be presented vertically.

6. Structural patterns

As part of the approach, the sequence of semiotic labels is assumed to derive from an underlying *structural pattern*, namely a regular sequence of symbols governed by some systematic organization. In practice, the actual sequence of semiotic labels is less regular than its underlying pattern, but it can be understood as an more or less irregular realization of the structural pattern.

Denomination	**Illustration**	**Codification**
Trivial	A A A A A A A A …	(A)
Binary	A B A B A B A B A B …	(AB)
Ternary	A B C A B C A B C A B C …	(ABC)
Quaternary	A B C D A B C D A B C D …	(ABCD)
Double binary	A A B B A A B B A A B B …	(2A,2B)
Double ternary	A A B B C C A A B B C C …	(2A,2B,2C)
Alternate	A A B C C D A A B C C D …	(2A,B,2C,D)
Cyclic	A B B C C D D D A B B C C D D D …	(A,2B,2C,3D)
Acyclic	A B C D E F G H …	"A,B,C,D, … H, …"
Double acyclic	A A B B C C D D … H H	"2A,2B,2C,2D, …, 2H"
Ergodic	A C D C B C D A A A B C C A …	{ABCD}
Extensive	A B B C C D D D D E C C C D F …	"A,2B,2C,4D,E,3C,D,F"

As can be understood from this table, the use of parentheses indicates that a pattern is cyclic ; the use of inverted commas denotes an enumeration and the use of brackets denotes randomness in the order of elements (the term "ergodic" is used in the sense that the probability of a symbol does not depend on its position in the sequence).

Assuming that the structure of the piece derives from "double" (or "triple") structural patterns (i.e. using symbols recurrently repeated twice or three times) should encourage the annotator to consider the possibility of upscaling the structural pulsation period, or of investigating on the existence of semiotically

[9] This notation is currently under consolidation and is therefore subject to future evolutions.

distinctive properties between successive similar elements. However, such adjustments should not be systematic, as there do exist situations for which double / triple structural patterns are perfectly relevant.

As mentioned in the introduction, the actual *realization* of a structural pattern in a music piece generally shows irregularities (which tend to increase when getting towards the end of the piece). For instance the structural description A B C D A B C D B C D E C D C D D can be viewed as a realization of 5 cycles of structural pattern (ABCD) with increasing irregularities towards the end of the piece. It could indeed be written as follows : A B C D A B C D ~~A~~ B C D ~~A B~~ E C D ~~A B~~ C D D, where double-barred characters are suppressions and underlined ones are insertions.

As already mentioned previously, more irregularities tend to occur as the piece progresses. It is conjectured that the distribution of labels in the realization of musical structures could tend to follow some sort of Zipf law.

7. Handling ambiguities

In this section, we address a number of situations where ambiguities may occur and we detail criteria and solutions for their consistent identification and description. It is worth underlining that the concepts defined so far make it easy to handle very generically a variety of ambiguous situations and to characterize them quite clearly.

In fact, the possibility for the annotation to render ambiguities is essential. As already mentioned in a footnote, ambiguities correspond to existing situations which are inherent to the high-level semiotic process that develops along a music piece ; namely the possibility of creating elements which result not only from simple variations within a given semiotic class, but also from combinations, offsprings, crossbreeds or strains of previously (or, sometimes subsequently) introduced material[10].

Therefore, it is essential that the annotator has hybrid symbols at his/her disposal to recognize ambiguous situations and annotate them as such, rather than being forced to assign a segment to one semiotic category or the other, when in fact this is inherently not decidable.

We consider successively the following situations :

1. Resolving label ambiguities
   - Class ambiguity
   - Alphabet ambiguity
2. Resolving segmentation ambiguities
   - Boundary ambiguity
   - Scale ambiguity

We denote as X (and Y) unlabeled blocks and by A, B (and C) class labels from the semiotic alphabet.

[10] An analogy can be made using concepts stemming from genetics : each main semiotic class possesses its own genotype, shared by all members of that class (beyond surface variations). However, some structural elements result from the hybridation, mutation or transposition of the genotype of other classes and should be identified (and annotated) as such.

7.a – Notations for mixed labels

Before giving indications on how to address the various ambiguities, we introduce the following notations :

AB : block showing undecided prevalence of properties of classes A and B (*vertical hybridation*)

A&B : block showing intertwined properties of classes A and B (*alternation*). A&B has usually a size double than that of A (and B) alone.

B|A : new system (B) in the first half of X but recall of morphological units of A in its second half (*horizontal hybridation*)

B>A : block (B) acceptable as autonomous at the current scale but showing strong cohesion with the previous block (A). The sequence [A][B>A] could be described as a single system+contrast at the upper scale (*kinship*).

B/A : system similar to A (*rooted* in A) but with properties strongly departing from those of the other elements of class A on a subset of layers (*mutation*). This includes cases when the deviating property is purely missing.

We also introduce the following ways of denoting short segments that occasionally intervene in between regular ones :

A_B : *tiling* segment, i.e. segment corresponding to a partial superposition between A and B (*overlap*)

_AB_ : *common affix* located between A and B which can not decidedly be related primarily to A or B (*connexion*)

In a future version of this document, a diagram will illustrate these various situations.

In AB blocks, the hybrid nature is vertical : the corresponding segment clearly results from the superposition of a segment of class A played simultaneously with a segment of class B. BA should be used instead of AB when it is felt that B is slightly prevalent over A.

In A&B blocks, properties of A and B are alternating, resulting in a block whose size is the total of that of A and B. Here too, B&A should be used instead of A&B, if B starts first.

In B|A blocks, the hybrid nature is horizontal : the system of the first half of X may be specific (hence the B label) but the system of the second half of X obviously relates to a block of class A (sometimes, with a different contrast, though, then noted $B|A_i$). Block B|A is considered as a new object at the scale of the semiotic structure, but its strong and obvious similarities with class A reflects in the choice of the hybrid label. At the immediately lower scale, the sequence [A] [B|A] would generally appear as [A A] [B A] or, more generally, as [A B] [C B].

Notation B>A covers situations where the current block (B) and its predecessor (A) would exhibit strong internal relations if considered as a single segment at a scale twice that of the block size. In other words, the sequence [A] [B>A] indicates that the structural pulsation period could have *locally* be chosen twice larger. This typically happens (but not exclusively) when, at the immediately lower scale, the sequence [A] [B>A] appears as [A A] [B C] or [A B] [A' C], with a significant number of properties common to these 4 elements.

In B/A segments, the current segment X can be viewed as some sort of “mutation” of segment A : some musical layers continue to reflect class A but some others are showing major differences with those of a regular block from class A, which are judged to go well beyond a simple variation. However, it usually remains possible to describe these differences through relatively simple transformations of the characters of class A, hence justifying the fact to consider X as “rooted on” A.

- A typical such situation happens when X is a solo (usually towards the end of the piece) where an instrument takes over the vocal lead and revisits a verse or a chorus with strong modifications of the melody supported by the very same harmony as that of block A. The strong change in the melody information layer is considered to create a new semiotic elements B, but it derives so clearly from A that it can be denoted as B/A.

- In many cases, a segment X functioning as an intro or an instrumental passage (with no melody layer) turns out to be unambiguously related to another segment of the piece (assumed to be labeled A) where a clear melodic lead develops. In that case, segment X should be considered as a particular case of mutation of A (by the loss of a character) and denoted B/A. Note however that constant layers are not covered by this situation as they are not considered as structuring layers. It is also important to distinguish this configuration from the observation of a “minimal variant”, i.e. when the the musical content in the block just preserves sufficient information to maintain all distinctive semiotic systems alive.

7.b – Label ambiguities

<u>*Class ambiguities*</u> happen when the annotator is facing the following question : *should the current block X be labeled as belonging to class A or to class B ?*

In this case, the annotator should first identify :

a. The set of common properties and systems for all blocks in classes A and B
b. The set of distinctive properties and systems between blocks of classes A versus B
c. The set of irrelevant properties and systems across blocks of classes A or B indistinctively

On this basis, the annotator should then choose the label corresponding to the characters most in adequation with X. If this is impossible to decide, an option is to use a hybrid notation and/or to consider creating a new symbol – see next point.

<u>*Alphabet ambiguities*</u> occur if the annotator is wondering : *should a new class B be created to account for current block X or should it be considered as a variant of class A ?*

In this case, the annotation should be guided by the following analysis steps :

a. Identify the set of common properties and systems for all blocks in class A and see if they are compatible with those of X
b. Determine the nature of distinctive properties between the various classes across the piece and see if one of them is applicable to X
c. Take into account the position of X in the sequence of labels within the piece (index and context)
d. Avoid inflation of the number of semiotic labels used over the whole piece
e. Consider the option of using a mixed label, especially a “mutation”

Indeed, ambiguities usually arise from the fact that some properties of X deviate from those of class A, but it is difficult to evaluate whether this deviation is (only) a strong variation of A or a new semiotic element. Here are some guidelines which should be considered to resolve the situation :

- If the deviation observed is concerning properties which have been used in other parts of the piece as distinctive properties between classes, this reinforces the plausibility of a new label B.
- On the other hand, if the block is occurring in a context (neighbouring labels) where class A is usually observed in other parts of the piece, this reinforces the plausibility of a variant of A.
- The absolute position of block X in the sequence of blocks should also be taken into account as deviations and/or irregularities tend to increase as the piece progresses. Therefore, more deviant variants of A can be tolerated at the end of the piece as opposed to the beginning.
- Note that reusing existing symbols leads to a more economical decomposition (pleading in favour of A) and that semiotic elements tend to be more obviously distinct from others when they are rare.

After having analyzed all these factors, the annotator must either arbitrate between A or B on the basis of the decisive prevalence of one of the above considerations, or opt for a hybrid notation, for instance B/A, thus describing a partly new semiotic element strongly rooted on an existing one. In case of uncertainty, the second solution is preferable.

7.c – Segmentation ambiguities

In this subsection, we briefly review a few situations which may trigger questions during semiotic labeling about the segmentation or the segmentation scale used to define the boundaries of the semiotic blocks.

Boundary ambiguities: this situation corresponds to the question that can be stated as : *should the border of segments labeled A and B be reconsidered ?*

Here, we shall distinguish 4 cases :

- The last sub-block of the stem of a block (labeled A) also acts as the first block of the stem of the next block (labeled B). This is a situation of tiling and this requires the creation of a specific segment where the overlap takes places, with label A_B, yielding a sequence : [A…] [A_B] […B]
- A sub-block located immediately after the end of the stem of block A (called *suffix*) acts as the first subblock of block B but it is also showing properties which relates strongly to the system of block A. Here, the annotator may also decide to extend the block into a tiling with the next block (with label B) and label the overlapping part as A_B, yielding again : [A…] [A_B] […B]
- A small sub-block may be found in between blocks A and B, but it is not part of the stem of A nor of that of B and its properties partially relates and partially departs both from those of class A and B. In that case, the label _AB_ shall be used. Here, this notation denotes some sort of connection unit, yielding the sequence : [A] [_AB_] [B]
- Finally, some sub-blocks may occasionally function both as the suffix (or prefix) of a block and a semiotic unit of their own (for instance a break that is used both as a contrast and as a re-intro). In that case, it is acceptable to denote this unit as a tiling A_B, where the right boundaries of A and B coïncide : [A…] [A_B]

In general, however, blocks shorter than half of the typical block size should be avoided. Blocks of a quarter size should preferably be merged to neighboring blocks as suffixes, prefixes, etc… unless they clearly act a functional units, especially intros and re-intros. Conversely, an individual affix should not exceed the quarter of a block (but several consecutive affixes may extend the size of a block beyond this recommendation).

Scale ambiguities : a number of situations may raise questions concerning the scale chosen for the structural segmentation, as of whether the local/global scale of the segmentation should be reconsidered, in particular by merging successive symbols :

- Large/complex structural patterns ; for example (ABCDEFGH)
- Recurring multiplicity of similar symbols in sequence; for example (nA,nB,nC,nD)

In both cases, the grouping of segments 2 by 2 should be considered, privileging simple n-ary structural patterns with n preferably around 4 (quaternary structural pattern).

However, this trend should be balanced with the objective of describing the structure around a scale which is comparable across pieces, namely segments with stems of 16 snaps, and a typical snap duration around 1 second per snap.

Note that repeated successions of $[A_1]$ $[A_2]$ (repetition of carrier systems with identical or distinct contrasts, which therefore forms a super-system [A B] [A C] at the immediately lower scale) should not be considered as denoting a scale ambiguity if other parts of the piece clearly decompose nicely at the current scale. Similarly, consistent successions of [A] [B|A] (another super-system) are a priori preferable to less regular decompositions such as [A] [B/2] [|A] , at least as long as [B/2] is not observed alone elsewhere.

Ultimately, for some (rare) cases, two scales may be equally admissible and in that case, the lowest one should be preferred. Appropriate symbols should be carefully chosen, in order to preserve as much information as possible on the relationships between successive blocks at the upper scale. This should result in a transcription with many occurrences of : $[A_i]$ $[A_j]$, [A] [B>A] and/or [A] [B|A].

7.d – Ultimate ambiguities / impossibility to annotate

In (hopefully) rare situations when the annotator is unable to decide on the label of a segment, he/she can use the following notations

- X, Y, Z = multiple possible interpretations between several labels
- X ? Y ? Z = hesitation of the annotator between several incompatible possibilities
- ? : complete inability to label the segment

The choice of “multiple interpretations” corresponds to cases when the annotator is convinced that the ambiguity between several options is somehow intentional and unresolvable. Choosing “hesitation” means rather that the annotator is not able to decide which of the various options is the most likely.

8. Extended conventions

The additional conventions proposed in this section offer a possibility to refine the proposed basic notation, with the two following goals :

1) to offer a finer description of the specific properties of observed variants of the semiotic elements
2) to provide an indication of some (proto-)functional status of the blocks within the piece

These conventions should however be used by experienced annotators only, once the basic notation is well understood and has become familiar. However, their knowledge should be useful to help the annotator identifying the type of variation that is observed in a given situation.

8.a – Expressive variants

The annotator may use a variety of superscripts to make more explicit the nature of observed variants of a same semiotic label A :

A^{f}, A^{ff} : increased loudness
A^{p}, A^{pp} : decreased loudness
$A^{»}$, $A^{»»}$: accelerated tempo
$A^{«}$, $A^{««}$: slowed-down tempo
$A^{\#}$, $A^{\#\#}$: upwards transposition
A^{b}, A^{bb} : downwards transposition
A^{M}, A^{m} : variants in major / minor modes
A^{t+}, A^{t-} : upwards / downwards translation (inducing a change of modality)
A^{i+}, A^{i-} : inclusion / deletion of an infix (can be used for prefixes and suffixes, with letters p and s)
A^{+}, A^{++} : increased support (more sources participate to the carrier system)
A^{-}, A^{--} : decreased support (less sources participate to the carrier system)
A^{x}, A^{xx} : reinforced properties (carrier systems applies to stronger or more well-defined properties)
$A^{\sim}$, $A^{\sim\sim}$: weakened properties (carrier systems applies to weaker or less well-defined properties)
A^{w1}, A^{w2} : variations of lyrics (can be neglected, except if considered as significantly expressive)

These various conventions should be understood as special cases which replace the standard superscript notation (A', A", $A^{(3)}$, …) in situations when it is possible to pinpoint the precise nature of the corresponding variant. However, before using these notations as expressive variations, the annotator must ascertain that the observed variations are not of a semiotic nature.

This list is probably not exhaustive and additional notations on this basis may be designed to account for unpredicted situations.

8.b – Functional indications

Even though latent semiotic structure description is distinct from functional structure description, it can be informative to choose the semiotic labels in such a way that they somehow reflect the functional status of the block within the piece.

We therefore propose to use, as much as possible, the alphabetic letters with the following correspondence :

Main « set » : ABCDEF-IJKLMN-XYZ

- · C : Main central element (C for Central / Chorus)
- · D : Secondary central element (may or may not be situated just after C)
- · A,B : Elements that precede the main central element
- · E,F : Elements that follow the main central elements
- · I, J : Introductive / re-introductive elements (locally or globally)
- · K,L : Conclusive / pre-conclusive elements (locally or globally)
- · M,N : Other recurrent non-central elements
- · X,Y,Z : Singular elements (occurring only once or spuriously)

Secondary « set » : GH-PQRSTUVW

- · R,S : Other central elements
- · P,Q : Elements which precede central elements
- · T,U : Elements which follow central elements
- · G,H : Introductive, conclusive, bridging elements
- · V,W : Any use

A central element is defined as an element that is representative of the piece in terms of musical content and of frequency of occurrence.

Summary of proto-functional conventions

	Intro	Pre-central	Central	Post-central	Relay / bridging	Other (recurrent)	Other (spurious)	Concluding
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	V,W		G,H

Future work

On the basis of the elements presented in this document, future work will be directed towards :

- the further formalization of the concepts and methodology established in this report
- the annotation and distribution of 500 songs annotated with the proposed conventions
- the expression of optimization criteria for structural description, derived from information theory
- the modeling of semiotic structure in terms of adequate hidden state automata
- the design of new algorithms aiming at inferring automatically the semiotic structure

Acknowledgements

Part of this work was achieved in the context of the QUAERO project, funded by OSEO.

We would like to thank our colleague Stanislaw Raczinsky who read, commented and made very useful suggestions to improve the pre-final version of this document.

References

This work is a sequel of previous work formerly published in :

[1] F. BIMBOT, O. LE BLOUCH, G. SARGENT, E. VINCENT. Decomposition into Autonomous and Comparable Blocks : A Structural Description of Music Pieces, in "Proc. 2010 Int. Society for Music Information Retrieval Conf. (ISMIR)", Utrecht, Netherlands, August 2010.

[2] F. BIMBOT, E. DERUTY, G. SARGENT, E. VINCENT. Methodology and Resources for The Structural Segmentation of Music Pieces into Autonomous and Comparable Blocks, in "Proc. 2011 Int. Society for Music Information Retrieval Conference (ISMIR)", Miami, United States, October 2011.

A specific document presenting in detail the fundamentals and the principles of the “system/contrast + affixes model” is in preparation. Please contact Frédéric Bimbot for more information.