# Decomposition into autonomous and comparable blocks: a structural description of music pieces

Frédéric Bimbot, Olivier Le Blouch, Gabriel Sargent, Emmanuel Vincent

**Publications Internes de l'IRISA**
**ISSN : 2102-6327**
**PI 1948 – Mars 2010**

# DECOMPOSITION INTO AUTONOMOUS AND COMPARABLE BLOCKS : A STRUCTURAL DESCRIPTION OF MUSIC PIECES

Frédéric BIMBOT[*] , Olivier LE BLOUCH[**] , Gabriel SARGENT[***] , Emmanuel VINCENT[****]

**Abstract:** The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is not commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

This article proposes an approach called *decomposition into autonomous and comparable blocks*, based on principles inspired from structuralism and generativism in Linguistics. It specifies a methodology for producing music structure annotation by human listeners based on simple criteria and resorting solely to the listening experience of the annotator.

We show on a development set that the proposed approach can provide a reasonable level of concordance across annotators and we are currently producing a set of annotations on the RWC database, which we intend to release to the scientific community, within the scope of MIREX.

**Key-words:** music structure, annotation, MIR

## *DECOMPOSITION EN BLOCS AUTONOMES COMPARABLES : UNE DESCRIPTION STRUCTURELLE DES MORCEAUX DE MUSIQUE*

**Résumé :** *La structure d'un morceau de musique est un concept auquel il est fréquemment fait référence en musicologie et en informatique musicale, mais pour lequel il n'existe pas de définition communément admise. Ceci soulève un problème méthodologique dans le cadre de la recherche d'informations dans les contenus musicaux lorsque l'on souhaite concevoir et évaluer des algorithmes d'inférence automatique de la structure musicale. Ceci complique aussi la production à grande échelle d'annotations cohérentes réalisées par différents annotateurs.*

*Cet article présente une approche dite de* décomposition en blocs autonomes comparables*, fondée sur des principes inspirés du structuralisme et de la linguistique générative. Celle-ci fournit une méthodologie pour l'annotation manuelle de la structure musicale à partir de critères simples et basés essentiellement sur l'expérience musicale de l'annotateur.*

*On montre que l'approche proposée permet d'atteindre un niveau de concordance raisonnable entre les annotations produites par plusieurs annotateurs sur un même corpus de développement. Cette approche est actuellement utilisée pour produire un ensemble d'annotations de la base RWC, que l'on souhaite mettre à la disposition de la communauté scientifique dans le cadre de la campagne d'évaluation MIREX.*

**Mots clés :** *structure musicale, annotation, recherche d'informations dans les contenus musicaux*

[*] CNRS, IRISA (- UMR6074). *frederic.bimbot@irisa.fr*
[**] INRIA, Centre INRIA Rennes - Bretagne Atlantique. *olivier.le_blouch@inria.fr*
[***] Université de Rennes1, IRISA (- UMR6074). *gabriel.sargent@irisa.fr*
[****] INRIA, Centre INRIA Rennes - Bretagne Atlantique. *emmanuel.vincent@inria.fr*

# 1 INTRODUCTION

## 1.1 Motivations

The automatic inference of musical structure is a research area of growing interest [1, 2, 3, 4, 5, 6], which is illustrated for instance by the creation in 2009 of a task called *structural segmentation* in the MIREX community [7], or the existence of a specific research topic called *music structuring and summarization* in the QUAERO project (started 2008) [8].

Inference of musical structure has multiple applications, such as fast browsing of musical contents, automatic music summarization, chorus detection, unsupervised mashups, music thumbnailing, etc... but also, more fundamentally, it offers great potential for improving the acoustic and musicological modeling of a piece of music with the help of structural information such as the relative position of events within structural elements or the exploitation of recurring similarities across them.

Musical structure deals with the description of the formal organization of music pieces. However, several conceptions of musical structure coexist and there is no widely accepted definition. This raises a methodological issue when the question arises of evaluating and comparing automatic algorithms on a common "ground-truth" (see, in particular [9]).

This article presents an attempt to provide an operational definition and to specify an annotation procedure for producing a structural description of music pieces that can be obtained quasi-univocally and in a reproducible way by several human annotators.

The concepts and the methodology proposed in this article are intended to be 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. However, we keep in mind that some other types of music (in particular, contemporaneous music) are much less suited to the proposed approach.

## 1.2 Objectives

The concepts and methodology presented in this work aim at specifying a process for annotating musical structure, with the following requirements :

- based on the *listening experience* of the annotator (and not on his/her *musicological expertise*)
- *unrelated* to any particular algorithm or application
- *independent* of any *musical role* assigned to structural elements
- *reproducible* across annotators
- *applicable* to a wide variety of music genres

At the current stage of our work, we have focused on the issue of locating structural elements (≈ segmentation) and we postpone to a later step the question of how to label these elements.

We first present, in section 2, the fundamentals of our approach, then we describe, in section 3, the proposed annotation process. We provide in section 4, an evaluation of the consistency of the methodology on a development set and we introduce to the MIR community a set of annotations on the RWC [10] data, based on the proposed approach.

## 1.3 Preliminary definitions

In the rest of this paper, we consider that a piece of music is characterized by *3 reference properties*, which may be constant or vary over time :

- tonality/modality (reference key and scale)
- tempo (speed / pace of the piece)
- timbre (instrumentation / audio texture)

We also consider that a piece of music shows *4 levels of temporal organization* :

- rhythm (relative duration and accentuation of notes)
- harmony (chord progression)
- melody (pitch intervals between successive notes)
- lyrics (linguistic content)

These levels of description form *7 musical layers* which we consider independently.

# 2 FUNDAMENTALS

## 2.1 Framework

The proposed approach relies on concepts inspired from *structuralism*, a school of thoughts initiated by Ferdinand de Saussure in the field of Linguistics [11] and later extended to other domains, in particular to some areas of music semiotics. Our approach also borrows ideas from *generative theory* by Lerdahl and Jackendoff [12].

In this context, we consider a music piece as the layout of a number of constitutive elements governed by a specific assembling process, called *syntagmatic process*. The constitutive elements are related to one another through *paradigmatic relationships* which manifest themselves as a set of *equivalence classes* (i.e. two elements belong to the same subset if and only if the relation holds between them). The entire scheme forms a *system* in the structuralist sense, namely an "entity of internal dependencies", according to Hjelmslev's definition [13].

The piece of music thus appears as a particular observation produced by this system and the problem of musical structure inference consists in determining the constitutive elements of the piece (i.e. *segmentation* task or, more generally speaking, *decomposition*) and to assign equivalence classes to each of them (*labelling* or *tagging* task).

As a consequence, specifying a type of musical structure requires the definition of :

1. the nature and properties of the constitutive elements

2. the assembling process used to combine them

3. the equivalence relation(s) that are referred to, so as to relate them to one another.

## 2.2 Working assumptions

In the present work, the constitutive elements are assumed to be common to the 4 levels of temporal organization. They are limited in time and are assembled mainly by concatenation. They are called *blocks*.

A block is defined as an *autonomous* segment (see 3.2). It is specified by a *starting instant*, a *duration* and a *size* (the distinction between duration and size is also explicitted in 3.2). A block can be decomposed into a *stem* (which is itself autonomous) and one or several *affixes*.

Several equivalence relations can be considered and combined to qualify *comparability* between blocks, in particular : isometry (same size), interchangeability (possibility to swap), similarity in a given layer or combination of layers, etc...

Thus, we approach music structure description as the process of decomposing the music piece into autonomous and comparable blocks. Blocks share similarities with musical phrases but are not strictly identifiable to them. The concept of blocks also has some connections with the notion of *grouping structure* as developed in [12].

# 3 SPECIFICATIONS

In this section, we introduce a number of criteria which are used to specify as univocally as possible the structural decomposition of a music piece. We attempt to formulate these criteria without resorting to absolute acoustic properties of the musical segments nor to their musical role in the piece (chorus, verse, etc...), so as to remain as independent as possible from musical genre.

## 3.1 Musical consistency

For specifying further the decomposition process, we resort to several simple transformations, such as the suppression, the insertion or the substitution of musical segments within the music piece and we consider the *musical consistency* of the result, with respect to the various layers defined in section 1.3.

The concept of musical consistency is somehow tricky to define : for instance, musical consistency of the transformed piece can be considered to be preserved as long as the transformation does not create any morphological singularity nor any syntagmatic disruption with respect to the original piece : in other words, the result of the transformation is valid provided that a similar passage is found somewhere else in the piece, or *could be* found without creating any feeling of heterogeneity with the rest of the piece.

Clearly, this definition is partly subjective but, in the lack of a more accurate formulation (which remains to be found), it provides human listeners with criterion for deciding whether they should consider that the result of a transformation is admissible or not.

## 3.2 Properties of blocks

A block is defined as a musical segment which is *autonomous*, i.e. which fulfils one of the two following properties : either it is *independent* (i.e. it is perceived as self-sufficient when played on its own) or it is *iterable* (it can be looped and the result is musically consistent).

Moreover, blocks within a musical piece have the property of being *suppressible*, i.e. they can be removed from the piece without altering its musical consistency. This test is used to identify the most likely block boundaries. However, suppressibility is a necessary but not a sufficient condition to qualify a block.

It is also worth noting that blocks are not necessarily homogeneous : reference properties may evolve within a block (change of tonality, tempo modifications, appearance/ disappearance of instruments or voice).

The *size* of a musical block is expressed as the number of times a listener would snap his fingers to accompany the music, at a rate which is as close as possible to 1 bps (beat per second). Conventionally, block boundaries are synchronized with the first beat of a bar. Occasionally, unusual situations may arise, such as blocks having a fractional size or for which the listener is unable to decide what the size is.

Blocks can contain *affixes*, i.e. portions which can be suppressed, yielding a reduced block which remains musically consistent with the rest of the piece. The various types of affixes are : *prefixes*, *suffixes* and *infixes* (the latter can be non-connex). A block can therefore be described as a *stem* combined with zero, one or several affixes. In general, affixes are short and not autonomous.

## 3.3 Equivalence relations and comparability

Several paradigmatic relationships between blocks can be considered :

- *isometry* : blocks of the same size (absolute isometry) or blocks reducible to stems of the same size (stem isometry).
- *interchangeability* : blocks that can be swapped within a music piece without altering its musical consistency.
- *similarity* : blocks sharing similarities across most of their musical layers (over the whole blocks or over their stems only).
- *isomorphy* : blocks that can be obtained from each other by a transformation of their reference properties.

As mentioned earlier, these equivalence relations are resorted to in order to determine on what basis blocks are judged comparable with one another.

## 3.4 Structural pulsation and regularity

To specify further the decomposition into blocks, it is hypothesized that the structure of (most) music pieces is rather cyclic and therefore follows some form of regularity characterized by a small set of distinct block sizes.

We thus suppose that the music piece is built upon *structural pulsation periods* which take, in order of preference :

- One value *(type I)*
- Two values *(type II)*
- A limited set of values observed in a quasi-regular sequence, called *structural pattern (type III)*
- A limited set of values, showing no structural pattern *(type IV)*
- A large variety of distinct values *(type V)*

We also designate as *type 0* (or undeterminable) a piece for which it turns out to be impossible (for the listener) to locate clear boundaries of autonomous segments. Blocks whose size complies with one value of the structural pulsation periods are called *regular* blocks.

The regularity property induces decompositions which tend to yield comparable blocks within a given piece.

Figure 1 depicts a block-wise structural decomposition in the case of a *type I* structure (top) and illustrates several configurations of non-regular blocks (bottom) and their corresponding notations :

- Truncated block : block resulting from the suppression of a fragment inside a regular block
- Extended block : block obtained by the insertion of an affix into a regular block
- Bridging block : irregular block, usually of a smaller size, and which is often isolated at the beginning of the piece, at the end or in-between regular parts.

- Tiling : partially overlapping blocks (on all levels of organization simultaneously), as is the case for instance in canon singing or fugue-like compositions.

### 3.5 PIC minimization and target duration

The various properties introduced in the previous paragraphs still do not necessarily elicit a unique structural decomposition. Several possibilities may remain, generally based on structural pulsation periods which are multiples or sub-multiples of each other.

These situations can be decided between by specifying a target duration for regular blocks, the value of which we derive from the minimization of the *predominant informative context.*

Figure 2 illustrates a structural decomposition as a paradigmatic representation showing the correspondence between homologous parts across distinct blocks.

If this decomposition is exploited to predict the musical properties of the music piece on a short time interval, the most relevant portions of the piece for this purpose are, on the one hand, the neighboring portions belonging to the same block (horizontally) and the homologous portions across the other blocks (vertically).

This is what we call predominant informative context (or PIC). It is distinct for each small portion of the music piece and it is solely determined by the structure. It constitutes the predominant source of information within the entity of internal dependencies mentioned in section 2.1.

If the total length of the music piece is equal to $N$ and if the typical block length is equal to $n$, then, the number of blocks in the piece is in the order of N/n and the total coverage of the PIC is given by :

$$C = n + (N/n) - 2 \tag{1}$$

which is minimal when $n = \sqrt{N}$ .

On the basis of music pieces with a typical length of 4 minutes (i.e. $N = 240$ seconds), the value of $n$ which minimizes $C$ is approximately equal 15.5 seconds.

In the present work, we retain the value of 15 s as the *target duration* for blocks, which leads to the following additional criterion : at least one of the structural pulsation periods must have a duration as close as possible to 15 s, on a logarithmic scale. This criterion tends to provide structural decompositions which result from a balanced contribution of the paradigmatic and syntagmatic axes.

More generally speaking a relative weight $\lambda$ can be applied to the two terms in equation (1), leading to a *PIC function* which writes :

$$C(\lambda) = n + \lambda(N/n) - (\lambda + 1) \tag{2}$$

and whose minimization (in $n$) induces decompositions based on an adjustable balance between the syntagmatic and the paradigmatic axes.

The constraint resulting from a target duration criterion enables the disambiguation of situations such as several identical medium-size segments in sequence and it provides blocks which are more adapted to comparisons *across* music pieces.

### 3.6 Subsidiary criteria

In some residual situations, several competing decompositions may locally be compatible with a given structural pulsation period while fulfilling satisfactorily all other criteria. For instance, sequences of an odd number of suppressible segments which have a size equal to half of the structural pulsation period.

In these circumstances, the following *subsidiary criteria* are considered :

- group preferably in a same block short neighboring segments of this passage which are most similar

- favor decompositions that yield the largest possible number of blocks which are interchangeable with other blocks outside this passage.

### 3.7 Procedure (summary)

The box hereafter summarizes the annotation procedure resulting from the criteria presented above.

1) Identify a plausible value $n$ (or set of values $n_i$) for the structural pulsation period(s), from the parts of the piece which are structured with strong regularity. Choose in priority values as close as possible to the target duration.
2) Locate suppressible blocks of size $n$, whether they are regular or can be derived straightforwardly from regular blocks. Also search for tiling at this stage.
3) Continue the decomposition by resorting to less regular (suppressible) blocks considering in priority block sizes that are sub-multiple of $n$.
4) Assess the regularity of the decomposition, and find out to which type (0, I, II, III, IV or V) the decomposition tends to belong. The local structure around the beginning and the end of the piece should be given a lower importance and so should it be for affixes.
5) Consider other options for the value(s) of the structural pulsation period(s) and find out whether they would lead to a simpler decomposition.
6) Refine the decomposition by resolving remaining ambiguities with the help of the subsidiary criteria.

Once the process finalized, the annotator fills up a short report summing up the type of structure, the degree of difficulty, the level of confidence and any relevant information pertaining to the annotation of that piece.

## 4 EVALUATION AND DISSEMINATION

### 4.1 Evaluation protocol

In order to validate the annotation procedure proposed in this paper, we have measured the concordance between several annotators on a same annotation task.

Four annotators are provided with a set of 20 songs in their audio version (development set), the list of which was determined by IRCAM, in the context of task 6.5 of the QUAERO Project (Table 2).

The concordance between annotators is evaluated by taking them pair-wise and computing for each piece the F-measure between their annotations (with a tolerance of ±0.75 s between segment boundaries) and averaging the F-measure over all 20 pieces.

Among the four annotators, none is a musicologist nor a professional musician. However, it is important to mention that they are the 4 co-authors of this paper, which may induce some methodological bias, which needs to be taken into account in interpreting the experimental results.

Scores presented in Table 1 correspond to what we call pre-final concordance between annotators, i.e. results of a round of annotation carried out after the annotation procedure was specified in its main lines, but before the *subsidiary criteria* of section 3.6 were introduced.

Figure 3 details the distribution of concordance scores across pieces. The median score is 95.8% and 2 pieces are responsible for almost 4% absolute error rate.

The subsidiary criteria were added in a last stage to resolve most of the residual ambiguities and a consensual annotation was produced for 19 of the 20 pieces, while the $20^{th}$ piece (#11 in Table 2) was considered as type 0 (i.e. impossibility to define reliable block boundaries).

### 4.2 Dissemination

*At the time of writing this article, we have undertaken the annotation the RWC database [10] using the methodology presented in this paper, and we aim at making it available shortly to the MIR community.*

*A first round of annotation is focused on the Popular Music dataset (100 songs), and a second round will deal with the Music Genre dataset (100 songs).*

*These data are being processed in the following way : given 3 annotators $A_1$, $A_2$ and $A_3$, annotator $A_i$ is in charge of producing an initial annotation, $A_j$ to control it and $A_k$ to arbitrate when $A_i$ and $A_j$ disagree. Roles are swapped so that 6 subsets of the data sets are processed by the 6 distinct combinations of i, j and k.*

*These annotations are planned to be ready in May 2010 (for the Pop set) and in July 2010 (for the Genre set). We will propose them to the MIREX 2010 community, for the structural segmentation task.*

*Once this annotation completed, this section will provide a brief statistical analysis of the annotated databases in terms of block size, duration, distribution of structural types across the pieces, typology of divergences across annotators, etc...*

| Annotateur | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | - | 88.9 | 95.7 | 92.9 |
| 2 | 88.9 | - | 88.7 | 88.7 |
| 3 | 95.7 | 88.7 | - | 92.8 |
| 4 | 92.9 | 88.7 | 92.8 | - |

Table 1: Pre-final concordance between annotators evaluated as the F-measure (%) between annotations averaged over 20 pieces of music. The mean value is 91.3%.

# 5 CONCLUSIONS

In this paper, we have specified and described thoroughly a consistent procedure for the description and the manual annotation of music structure, which is usable by non-expert human listeners, without resorting to external acoustic properties, nor to the analysis of the musical role of the constitutive elements. Our approach does not refer either to a particular family of automatic algorithms.

We are now working on the definition of a procedure for assigning labels to the structural blocks, so as to account for similarities between them. More specifically, we currently investigate an approach based on distinct sets of labels for each layer, yielding a multi-dimensional description of the internal similarities within the music piece.

# References

[1] Peeters, G. "Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach", in *Lecture Notes in Computer Science*, Springer- Verlag, 2004.

[2] Abdallah, S. Noland, K., Sandler, M., Casey, M., and Rhodes, C. "Theory and evaluation of a Bayesian music structure extractor", in *Proc. ISMIR*, London, UK, 2005.

[3] Goto, M. "A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station", *IEEE Transactions on Audio, Speech, and Language Processing*, 2006.

[4] Paulus, J. and Klapuri, A. "Music structure analysis by finding repeated parts", in *Proc. AMCMM,* Santa Barbara, California, USA, 2006.

[5] Peeters, G. "Sequence Representation of Music Structure Using Higher-Order Similarity Matrix and Maximum- Likelihood Approach", in *Proc. ISMIR*, Vienna, Austria, 2007.

[6] Levy, M., Sandler, M. "Structural Segmentation of musical audio by constrained clustering", *IEEE Transactions on Audio, Speech and Language Processing*, 2008.

[7] MIREX 2009 : http://www.music-ir.org/mirex/2009

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

[9] Geoffroy Peeters and Emmanuel Deruty : Is Music Structure Annotation Multi-Dimensional ? A Proposal For Robust Local Music Annotation. *LSAS*, Graz (Austria) 2009.

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

[11] F. de Saussure : Cours de Linguistique Gnrale. 1916.

[12] Fred Lerdahl & Ray Jackendoff : A Generative Theory of Tonal Music, MIT Press, 1983, reprinted 1996.

[13] Louis Hjelmslev : Essays in Linguistics (1959).

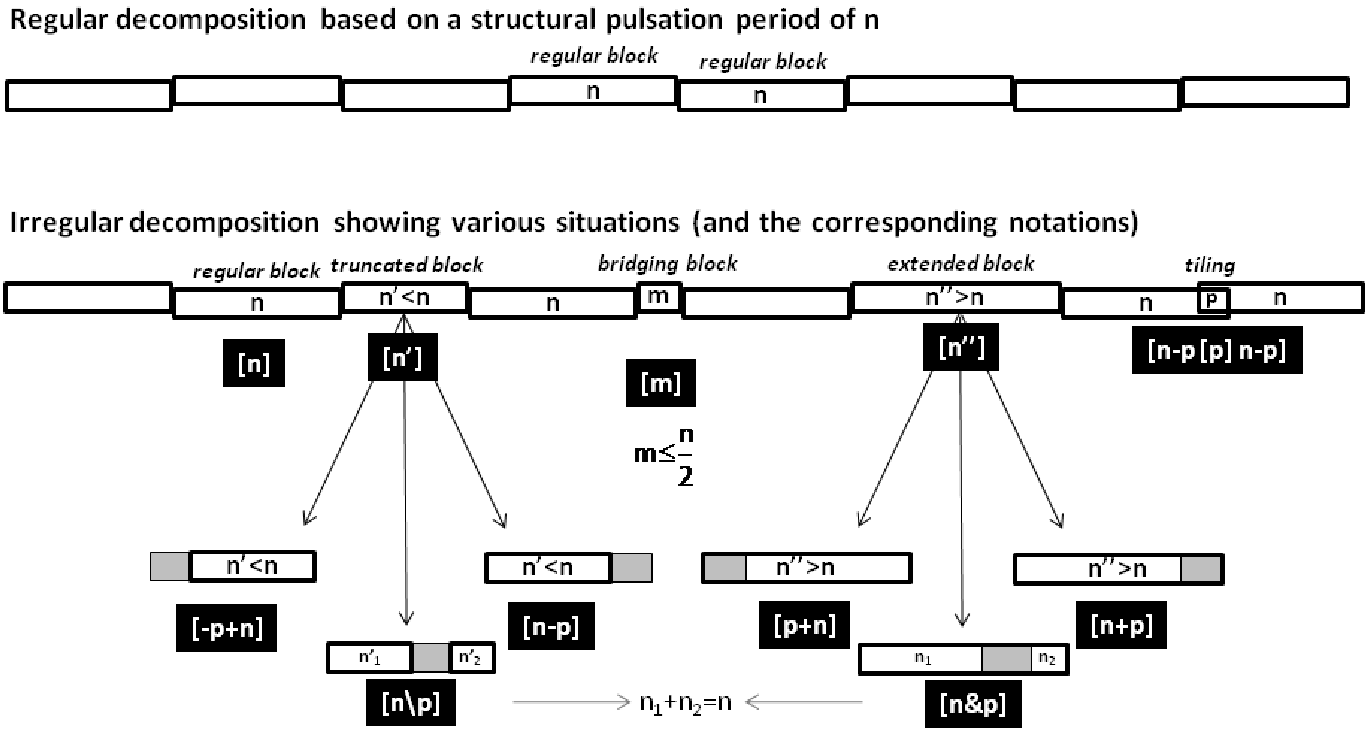
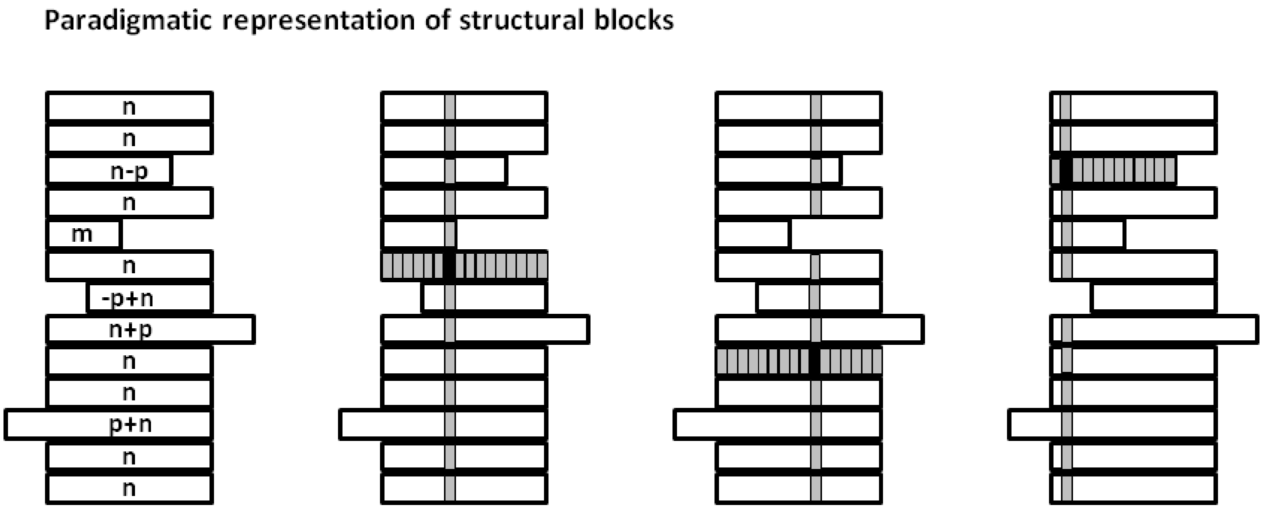


Figure 1: Decomposition into structural blocks (syntagmatic point of view) showing various configurations of irregular blocks.



*Any small portion (in black) shows privileged internal dependencies with the gray parts in the piece : on the one hand, other portions of the same structural block (horizontally) and on the other hand, homologous portions in the other blocks (vertically). These define the PIC (Predominant Informative Context).*

Figure 2: Decomposition into structural blocks (paradigmatic point of view) and visualization of the Predominant Informative Context (PIC).

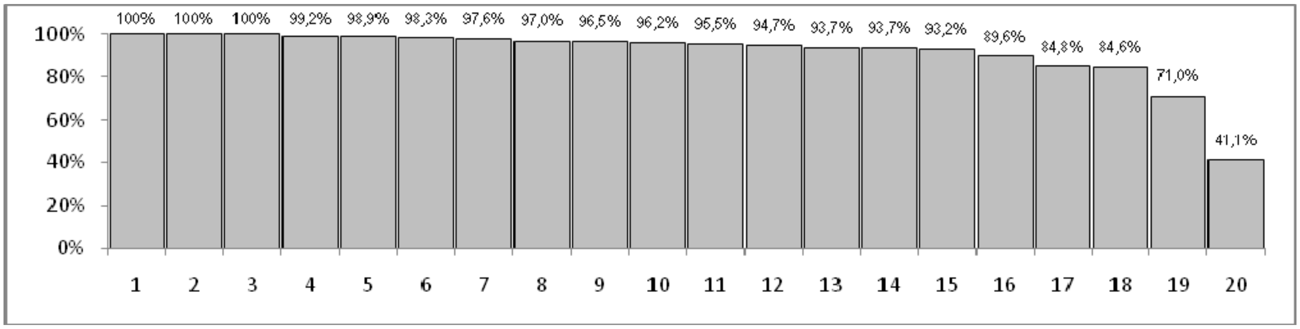


Figure 3: Inter-annotator concordance sorted in descending order for the 20 music pieces in the development set.

| | | |
|---|---|---|
| 01 | Pink Floyd | Brain Damage |
| 02 | Queen | Lazing On A Sunday Afternoon |
| 03 | DJ Cam | Mad Blunted Jazz |
| 04 | Outkast | Return Of The G |
| 05 | ACDC | You Shook Me All Night Long |
| 06 | Eric Clapton | Old Love |
| 07 | Stan Getz & J. Gilberto | O Pato |
| 08 | Enya | Caribbean Blue |
| 09 | Mickael Jackson | Off The Wall |
| 10 | Bass America Collection | Planet |
| 11 | Plastikman | Fuk |
| 12 | Shack | Natalie's Party |
| 13 | Sean Kingston | Take You There |
| 14 | Lil Mama | Shawty Get Loose |
| 15 | Abba | Waterloo |
| 16 | Eiffel 65 | Blue (Da Ba Dee) |
| 17 | Meat Loaf | I'd Do Anything For You |
| 18 | Kaoma | Lambada |
| 19 | Vangelis | Conquest Of Paradise |
| 20 | Nirvana | Smells Like Teen Spirit |

Table 2: List of music pieces used for the experiments reported in this article (development set)