A regularity-constrained Viterbi algorithm and its application to the structural segmentation of songs

Gabriel Sargent, Frédéric Bimbot, Emmanuel Vincent

A REGULARITY-CONSTRAINED VITERBI ALGORITHM AND ITS APPLICATION TO THE STRUCTURAL SEGMENTATION OF SONGS

Gabriel Sargent
Université de Rennes 1,
IRISA (UMR 6074)
gabriel.sargent@irisa.fr

Frédéric Bimbot
CNRS,
IRISA (UMR 6074)
frederic.bimbot@irisa.fr

Emmanuel Vincent
INRIA Rennes
Bretagne Atlantique
emmanuel.vincent@inria.fr

ABSTRACT

This paper presents a general approach for the structural segmentation of songs. It is formalized as a cost optimization problem that combines properties of the musical content and prior regularity assumption on the segment length. A versatile implementation of this approach is proposed by means of a Viterbi algorithm, and the design of the costs are discussed. We then present two systems derived from this approach, based on acoustic and symbolic features respectively. The advantages of the regularity constraint are evaluated on a database of 100 popular songs by showing a significant improvement of the segmentation performance in terms of F-measure.

1. INTRODUCTION

Music structure is one of the properties which contributes to the characterization of a music piece. It describes its temporal organization at a high level, by means of segments labeled according to their musical content and their relationships with one another. The automatic structural segmentation of songs is generally addressed by analyzing the homogeneity and the repetitiveness of the musical content over time (timbre, harmony, rhythm, melody).

Recent work [2] proposes a single-level definition of the structure of a music piece based on a regularity assumption. It implies the prevalence of one (or a few) typical segment duration(s) within each song, *i.e.* structural pulsation period(s). Indeed, a large part of western popular music is built on musical patterns (rhythmic cells, chord progressions, melodies...) which show cyclic behaviors and which are fully or partly repeated over time. This induces some sort of regularity in the structure of songs.

The present work is based on this regularity assumption in music. We introduce a general segmentation framework, which consists of an optimization method to find the best segmentation combining the similarities and/or the contrasts in musical content and the regularity of the segments. An implementation of this method is proposed by means of a Viterbi algorithm.

A similar segmental Viterbi algorithm was briefly sketched in [10] in the context of a probabilistic model (segmental HMM). In this paper, we make it more explicit and we extend it to any type of cost function. This makes it possible to exploit combinations of clustering-based and similarity-matrix-based approaches and to a wider variety of situations outside the probabilistic framework. We also discuss the importance of the regularity cost in the estimation of the segment boundaries, and provide experimental results with several choices for the two terms of the segmentation cost.

The structure of the paper is as follows. In section 2 we present the general music segmentation method, without considering a particular musical feature or temporal scale. Section 3 describes its implementation by means of a Viterbi algorithm, and discusses the expression of segmentation costs. In section 4, after briefly reviewing former work on music structure, we apply the proposed segmentation method to this particular problem. We then present two structural segmentation systems based on the algorithm developed above. Section 5 evaluates the effect of the incorporation of regularity constraints thanks to the evaluation of these systems on the RWC popular music database [6].

2. GENERAL APPROACH

This section presents a general method for the temporal segmentation of music pieces, when regularity assumptions can be hypothesized on the segment length. It consists of an optimization process where the optimal segmentation is searched simultaneously considering the properties of the data and the regularity of the segmentation.

A music piece X can be described as a sequence of N features $\{x_t\}_{1 \leq t \leq N}$ along a particular temporal scale (*e.g.*

frames, or beats...). We denote $X_{t_i}^{t_j} = \{x_t\}_{t_i \leq t < t_j}$ the sequence of features associated to the temporal interval $[t_i, t_j[$.

Let us define a segmentation $S = \{s_k\}_{1 \leq k \leq n}$ of X as a sequence of n intervals $s_k = [t_k, t_{k+1}[$, with the following conventions :

- $t_1 = 1 < ... < t_k < ... < t_n < t_{n+1} = N + 1$,
- $s_0 = [t_0, t_1[= [0, 1[$, for the algorithm initialization,
- $m_k = t_{k+1} - t_k$ is the length of s_k.

We aim at finding the optimal segmentation, by minimizing a certain cost function.

We assume that the cost function C can be written as

$$C(S) = \sum_{k=1}^{n} \Gamma(s_k) \tag{1}$$

with

$$\Gamma(s_k) = \Phi(s_k) + \lambda(\tau)\Psi(s_k) + \epsilon \tag{2}$$

where

- $\Phi(s_k)$ is a content-based segmentation cost, which takes low values when the sequence of features in s_k is likely to correspond to a structural segment. This cost can be described according to different families of functions, like change detection functions or similarity functions. It can also, for instance, be derived from a probabilistic function $P(s_k)$, as $-\log P(s_k)$.
- $\Psi(s_k)$ is a regularity cost. We consider that the regularity of a segmentation depends on the deviation of the length of its segments to a prior reference length τ called the structural pulsation period (as a consequence, $\Psi(s_k)$ decreases as m_k approaches τ). Note that, if the values of m_k are expected to follow a particular distribution $\pi(m_k)$ around τ, $\Psi(s_k)$ can be set as $\Psi(s_k) = -\log(\pi(m_k))$.
- $\lambda(\tau)$ is a balance parameter between these two costs.
- In practice, we add a small constant $\epsilon > 0$ to give a slight advantage to longer segments in the case where Φ and Ψ would be equivalent for several segmentations.

3. IMPLEMENTATION

This section presents an implementation of the approach presented above, and describes possible choices of cost functions Φ, Ψ and parameter λ.

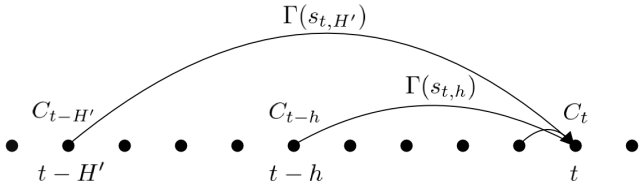

Figure 1. Admissible predecessors for t and their costs

3.1 Viterbi algorithm

Let $s_{t,h}$ be the interval corresponding to $X_{t-h}^{t} = [x_{t-h}, x_t[$, the set of features which precede the temporal index t within a window of length h. We denote H as the maximal window length considered [1] .

- *Initialization ($t = 1$)*

  We set $S_1 = \{[0, 1[\}$ and $C_1 = 0$.

- *For $t = 1 : N - 1$*

  We consider $\{s_{t,h}\}_{1 \leq h \leq H'}$, with $H' = \min(t-1, H)$ as the set of admissible predecessors for temporal index t.

  The optimal segmentations $\{S_{t-h}\}_{1 \leq h \leq H'}$ ending at indexes $\{t - h\}_{1 \leq h \leq H'}$ are assumed to be known, as well as their associated cumulative costs $\{C_{t-h}\}_{1 \leq h \leq H'}$.

  Then, the best partial segmentation S_t is built by choosing the extension of the former partial segmentation S_{t-h} with the lowest cost. We evaluate respectively :

  1. $\Gamma(s_{t,h})$ for $1 \leq h \leq H'$,
  2. $b(t) = \text{argmin}_{1 \leq h \leq H'}\{C_{t-h} + \Gamma(s_{t,h})\}$,
  3. $C_t = C_{t-b(t)} + \Gamma(s_{t,b(t)})$

  We can note that $S_t = S_{t-b(t)} \cup \{S_{t,b(t)}\}$.

The optimal segmentation for X, noted S_{opt} with cost C_{N+1}, is obtained by backtracking the optimal predecessors stored in $b(t)$. The associated temporal indexes $\{t_k\}_{1 \leq k \leq n_{\text{opt}}}$ are then found thanks to the following recursion :

1. $t_{n_{\text{opt}}+1} = b(N + 1)$,
2. $t_k = b(t_{k+1})$, for $1 \leq k \leq n_{\text{opt}}$.

n_{opt} is the number of boundaries of S_{opt}, obtained after this backtracking process.

[1] Typically, $H = N$, but smaller values can be used (*e.g.* multiples of τ).

3.2 Design of the cost functions

3.2.1 *Content-based segmentation cost* Φ

The objective of the content-based segmentation cost is to evaluate a set of segments according to the redundancy of their content. Segmentations with lower costs are expected to consist of segments built on the same musical patterns. Different families of functions can be considered, like abrupt change detection criteria or similarity functions.

Abrupt change detection criteria assign a low cost to segments associated to probable boundaries. In automatic structure inference, [3] uses for example a "novelty function" based on the analysis of the local homogeneity of the song over time.

Similarity functions aim to assign a low cost to segments made of sequences of features repeated elsewhere in the song. We can define such a function as

$$\Phi(s_k) = \min_{\theta \in Z_k}\{\phi(X_{t_k}^{t_k+m_k}, X_\theta^{\theta+m_k})\}. \quad (3)$$

The lowest dissimilarity ϕ is taken between the sequence of features $X_{t_k}^{t_k+m_k}$ from s_k (of length m_k) and any other sequence of the same length contained in a portion Z_k of X. In particular, $\Phi(s_k) = 0$ when the sequence of features of s_k is exactly repeated elsewhere in Z_k, $\Phi(s_k) > 0$ otherwise.

$Z_k = [1, t_{k-m_k}] \cup [t_{k+m_k}, N]$ can be chosen to avoid intra-segment comparisons. In the case of a binary dissimilarity, where a song is described as a sequence of symbolic features, the following function can be chosen :

$$\phi(X_{t_k}^{t_k+m_k}, X_\theta^{\theta+m_k}) = \sum_{p=0}^{m_k-1} 1 - \delta(x_{t_k+p}, x_{\theta+p}), \quad (4)$$

where δ is Kronecker's delta (equals 1 when arguments have the same value, 0 otherwise). More generally any non-binary function can be used in equation (3).

3.2.2 *Regularity cost* Ψ

The regularity cost Ψ of a segmentation is based on the measure of the deviation between the length of its segments from a reference length τ. It can show the following properties :

1. $\Psi(\tau) = 0$,
2. $\Psi(m_k) > 0$, taking higher values as the segment length m_k moves away from τ.

A lot of functions can satisfy these properties. We consider two categories of functions : convex and non-convex functions. As non-convex functions verify the property :

$$\Psi(\tau) + \Psi(\tau + \Delta) < \Psi(\tau + \Delta_1) + \Psi(\tau + \Delta_2) \quad (5)$$

with

$$\Delta = \Delta_1 + \Delta_2, \quad (6)$$

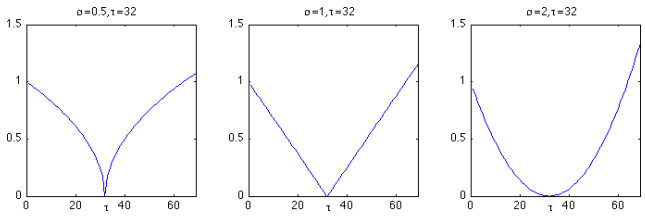

Figure 2. Examples of regularity costs Ψ_α for $\alpha = \{0.5, 1, 2\}$ and $\tau = 32$

$$\Delta_1 > 0 \text{ and } \Delta_2 > 0 \quad (7)$$

they favor segmentations made of fewer irregular segments. By contrast, convex functions tend to favor segmentations with irregularities spread across several segments.

As an illustration, we consider the following family of symmetric functions derived from the l_α norm :

$$\Psi_\alpha(m_k) = |\frac{m_k}{\tau} - 1|^\alpha \quad (8)$$

m_k is the length of interval s_k, and α controls the convexity of the function (we have a non-convex function if $0 < \alpha < 1$, and a convex one if $\alpha \geq 1$). Figure 2 shows Ψ_α for $\alpha = \{0.5, 1, 2\}$.

3.3 Balance parameter λ

We consider that λ depends on τ as the probability of having irregular segments grows with the number of segments, and therefore with the inverse of τ. We choose the linear relation $\lambda(\tau) = \lambda\tau$, where λ is a constant parameter to be tuned.

4. APPLICATION TO THE STRUCTURAL SEGMENTATION OF SONGS

The work presented in section 3 is primarily intended to the structural segmentation of songs. Automatic music structure inference is a difficult task, because the problem to be solved is usually ill-posed. Moreover, it requires the analysis and the complex combination of features and criteria through the development of sophisticated metrics and algorithms. In this section, we review briefly the main state-of-the-art methods for automatic structural segmentation of songs, before describing two structural segmentation systems implemented from the proposed method.

4.1 State-of-the-art

Different approaches have been proposed to the problem of automatic structure inference. They generally use acoustic features, such as Mel-Frequency Cepstral Coefficients (MFCCs) and Chroma vectors, which characterize the instrumental timbre and the harmonic content respectively. Other features are described in [17], [1], and [8]. Structural segments are assumed to show stable instrumentation

(often associated to homogeneous timbre) and therefore to appear as blocks with specific textures in similarity matrices [3, 12], or sequences of similar states in Hidden Markov Models (HMMs) [11].

Repeated harmonic progressions can be detected by localizing the sequences of high similarity coefficients in subdiagonals of the chroma-based similarity matrix [4]. Other approaches, like HMMs [10, 14], or more recently Non negative Matrix Factorization [20] are also used for the recognition of repeated harmonic patterns. Some methods use dynamic programming : Shiu *et al.* interpret the chroma-based similarity matrix as a time-state representation and use the Viterbi algorithm to find the path with the highest score in terms of similarity through it [15]. A constraint is set to give priority to the diagonal direction for the path, and implicitly influence the length of the estimated structural segments.

Some other approaches combine these content-based methods by means of optimization problems, as in [8, 12]. A more detailed state of the art is available in [13].

In the following section, we present two systems that infer the structural segmentation of a song, incorporating the idea of "structural pulsation period" [2] .

4.2 Presentation of the systems

These systems perform a structural segmentation of songs combining content-based segmentation under a regularity constraint by means of the Viterbi algorithm presented in section 3.1. System 1 uses acoustic features to compute change detection criteria and estimates the main structural pulsation period τ from the audio. System 2 analyzes symbolic features, uses a similarity function and prior knowledge of τ (fixed at 32 beats). As features are considered at the beat scale, a beat detection system is needed. We evaluate for these 2 systems the impact of incorporating a regularity constraint on the relative performance of the segmentation.

4.2.1 System 1 : combination of change detection criteria on acoustic features

The system we consider is the one described in [16]. In this paper, we consider variants of this system both with and without the regularity constraint in order to analyze its impact on structural segmentation inference. The content-based segmentation cost is based on 3 statistical criteria which measure for each temporal index the likelihood ratio of a structural segment boundary. This criterion combines instrumental changes, short events and contrastive patterns over time.

The criteria are combined in a weighted sum to form what we name here the content-based segmentation cost. A linear regularity cost function is used to perform the Viterbi approach described in section 3.1, to find the segmentation with lowest cost. The main structural pulsation period of the song is estimated by a Fourier transform on the instrumental change criterion.

4.2.2 System 2 : similarity function on symbolic features

It is interesting to consider symbolic features for structure inference as other means of music description. The joint use of various features in a global and versatile retrieval system may increase the accuracy of the estimated segmentation [19]. The symbols can be obtained for instance from a score of the piece. System 2 uses chords estimations to compute the similarity function described with the equations (3) and (4) of section 3.2.1. Each chord class is associated to a different symbol, to obtain a quite neutral symbolic description of the song. The size of the alphabet of symbols we use is the number of chord classes used by the chord estimator (*e.g.* 24 classes for major and minor chords). Each symbol corresponds to a duration of 2 beats, in order to be consistent with the temporal scale used in [2].

The structural pulsation period value τ is considered as prior knowledge and used in the regularity cost Ψ_α of equation (8), section 3.2.2. The content-based cost and the regularity cost are then combined using equations (1) and (2) from section 2, and the segmentation with lower cost is found using Viterbi algorithm from section 3.1.

5. EVALUATION

5.1 Evaluation database

The algorithms have been evaluated using the RWC popular music database [6], and the set of reference annotations provided by [2], used in MIREX 2010. This database consists of 100 songs written and produced for research purposes.

5.2 Evaluation metrics

The evaluation of the segmentation is done by Precision (P), Recall (R) and F-measure (F) metrics. Let s_R be the set of reference boundaries (annotations) and s_E the set of estimated ones, they are respectively defined as :

$$P = \frac{|s_E \cap s_R|}{|s_E|}; R = \frac{|s_E \cap s_R|}{|s_R|}; F = \frac{2PR}{(P+R)}. \quad (9)$$

The matching of reference and estimated boundaries is performed within particular tolerance windows. We consider 0.5 s and 3 s as in MIREX 2010. Note that each boundary is used only once during the matching process.

5.3 Feature extraction and algorithm parametrization

System 1 (which uses change detection criteria) uses 20 MFCCs (including the 0th coefficient), extracted from

[2] This can be seen as a way to constrain the ill-posed problem of structural segmentation towards a well-defined solution.

Tolerance window = 0.5 sec.					
system	α	λ	$P(\%)$	$R(\%)$	$F(\%)$
1	-	0	10.1	53.6	17.0
	1	0.5	23.9	24.3	**23.8**
Tolerance window = 3 sec.					
system	α	λ	$P(\%)$	$R(\%)$	$F(\%)$
1	-	0	16.8	89.3	28.2
	1	0.5	61.2	63.0	**61.4**

Table 1. Average Precision (P), Recall (R), and F-measure (F) for two versions of System 1 on the RWC pop database.

frames of length 23.2 ms, and a hop size of 11.6 ms (using scripts from MA toolbox by Beth Logan and Malcolm Slaney [3]). Chroma vectors (12 coefficients) are extracted from frames of length 92.9 ms, and a hop size of 23.2 ms. Chroma vectors and beats estimation are computed thanks to LabRosa scripts [4].

System 2 (based on a similarity function) inputs the chords transcriptions obtained by the algorithm from Ueda *et al.*, described in [18], and uses the downbeat annotations available with the RWC database [5]. The reference annotations show that more than 80% of the songs have a main structural pulsation of 32 beats. We will then use $\tau = 32$ beats as prior knowledge for our evaluation, and $H = 3\tau$ as the maximal number of admissible predecessors for each temporal unit.

A preliminary study on a subset of RWC popular was carried out to identify reasonable values of λ which fall within the interval [0, 1]. Three values of α are chosen to consider regularity costs functions with different convexities : a non-convex regularity cost function ($\alpha = 0.5$), a convex regularity cost function ($\alpha = 2$), and the intermediate case $\alpha = 1$.

5.4 Results

The values gathered in Tables 1 and 2 for System 1 and 2 show that the overall mean F-measures increase significantly when the regularity cost is introduced.

Figure 3 shows the average F-measure obtained with System 2 for the 3 regularity costs mentioned in 5.3. The values of λ corresponding to optimal performance appear in Table 2 for each case. The value of α has an impact on the accuracy of the estimated boundaries : it can be seen that, for a small tolerance, a non-convex regularity cost function gives better boundary accuracy than a convex one. This can be explained by the fact previously mentioned, that the convex case ($\alpha = 2$) tends to spread structural irregularities (deviations from the ideal segmentation with segments of length τ) across several structural segments. On the contrary, the non-convex case ($\alpha = 0.5$) tends to concentrate them on a few segments. These results therefore show not only the advantage of the regularity constraint but also the importance of the fine properties of the corresponding cost function.

[3] http://www.ofai.at/elias.pampalk/ma/documentation.html

[4] http://labrosa.ee.columbia.edu/projects/coversongs/

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

Tolerance window = 0.5 sec.					
system	α	λ	$P(\%)$	$R(\%)$	$F(\%)$
2	-	0	17.9	31.9	22.0
	0.5	0.30	37.7	34.8	**35.6**
	1	0.30	34.7	32.3	33.0
	2	0.95	29.3	26.8	27.5
Tolerance window = 3 sec.					
system	α	λ	$P(\%)$	$R(\%)$	$F(\%)$
2	-	0	36.1	64.7	44.5
	0.5	0.15	63.1	63.1	**62.0**
	1	0.15	63.4	64.1	**62.7**
	2	0.60	64.5	60.0	61.2

Table 2. Average Precision (P), Recall (R), and F-measure (F) for System 2 (optimally tuned, considering $\lambda \in [0 : 1]$), on the database described in 5.1.

As a point of comparison, the best system in structural segmentation at MIREX 2010 [6] (MND1) obtained F-measures of 35.9% and 60.5% (for tolerance windows of 0.5 s and 3 s respectively) on the same database. Note however that System 2 relies on a manual annotation of the downbeats.

6. CONCLUSION

The work presented in this paper has highlighted the relevance of incorporating a regularity constraint in the task of structural segmentation. Even with very basic cost functions as the ones considered in the present work, the very existence of the regularity constraint favors the retrieval of a well-defined solution. The Viterbi implementation, which we have detailed, allows a fast calculation of the optimal solution, and it can be applied in a generic way to any type of cost function.

The corresponding Matlab code will be made available to the MIR community [7] for enabling further experimental investigation within diverse structural segmentation systems and possibly for other tasks in MIR where the regularity constraint can be meaningful.

[6] http://nema.lis.illinois.edu/nema_out/mirex2010/results/struct/mirex10/summary.html

[7] http://www.irisa.fr/metiss/logiciel/

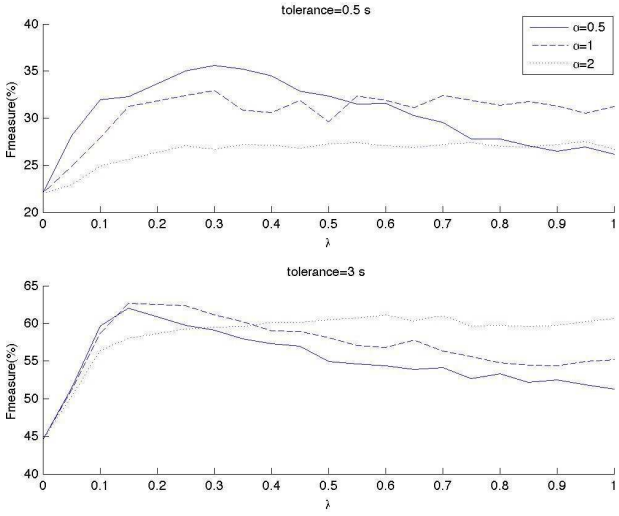

Figure 3. Evolution of the average F-measures of System 2 on the database described in 5.1, as a function of balance parameter λ, for 3 types of regularity cost function ($\Psi_{\alpha=\{0.5,1,2\}}$, $\tau = 32$).

7. ACKNOWLEDGEMENTS

The authors would like to thank Yushi Ueda and Nobutaka Ono for their help in the collection of chord transcriptions used in this article. This work was partly supported by the Quaero project [8] funded by Oseo and by the associate team VERSAMUS [9] funded by INRIA.

8. REFERENCES

[1] L. Barrington, A. B. Chan, G. Lanckriet, "Modeling music as a dynamic texture", *IEEE Transactions on Audio, Speech, and Language Processing*, Volume 18 Issue 3, March 2010.

[2] F. Bimbot, O. Le Blouch, G. Sargent and E. Vincent, " Decomposition into autonomous and comparable blocks : a structural description of music pieces", *Proceedings of the International Symposium on Music Information Retrieval*, pp. 189–194, 2010.

[3] M. Cooper and J. Foote, "Media segmentation using self-similarity decomposition" *Proceedings of the SPIE Storage and Retrieval for Multimedia Databases*, San Jose, California, USA, pp. 167–175, January 2003.

[4] M. Goto, "A chorus-section detecting method for musical audio signals" *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, China, pp. 437–440, April 2003.

[5] M. Goto, AIST Annotation for the RWC Music Database, *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pp. 359–360, October 2006.

[6] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC Music Data- base : RWC music database : Popular, Classical, and Jazz Music Databases" *Proceedings of the International Symposium on Music Information Retrieval*, USA, pp. 287–288, October 2002.

[7] T. Jehan, "Hierarchical multi-class self similarities" *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New York, USA, October 2005.

[8] K. Jensen, "Multiple scale music segmentation using rhythm, timbre and harmony", *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[9] A. Klapuri, M. Davy (Editors) *Signal processing methods for music transcription*, Springer, New York, 2006.

[10] M. Levy and M. Sandler, "New methods in structural segmentation of musical audio", *Proceedings of European Signal Processing Conference*, pp. – September 2006

[11] B. Logan and S. Chu, "Music summarization using key phrases" *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, pp. 749–752, June 2000.

[12] J. Paulus and A. Klapuri, "Music structure analysis by finding repeated parts", *Proceedings of AMCMM*, Santa Barbara, California, USA, pp. 59–68, October 2006.

[13] J. Paulus, M. Muller, A. Klapuri, "Audio-based music structure analysis", *Proceedings of the International Symposium on Music Information Retrieval*, pp. 625–636, 2010.

[14] G. Peeters, A. La Burthe, and X. Rodet, "Toward automatic music audio summary generation from signal analysis" *Proceedings of the International Conference on Music Information Retrieval*, Paris, France, pp. 94–100, October 2002.

[15] Y. Shiu, H. Jeong, and C. C. Jay-Kuo, "Similarity matrix processing for music structure analysis" *Proceedings of AMCMM*, Santa Barbara, California, USA, pp. 69–76, October 2006.

[16] G. Sargent, F. Bimbot and E. Vincent, "A structural segmentation of songs using generalized likelihood ratio under regularity assumptions," *MIREX evaluation campaign*, 2010. http://hal.inria.fr/inria-00551411/en

[17] D. Turnbull, "A supervised approach for detecting boundaries in music using difference features and boosting", *Proceedings of the International Symposium on Music Information Retrieval*, pp. 057–060, 2007.

[18] Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono and S. Sagayama, "HMM-based Approach for Automatic Chord Detection Using Refined Acoustic Features", *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 5506–5509, March 2010.

[19] E. Vincent, S. A. Raczynski, N. Ono and S. Sagayama " A roadmap towards versatile MIR", *Proceedings of the International Symposium on Music Information Retrieval*, pp. 662–664, 2010.

[20] R. Weiss, J. Bello, "Identifying Repeated Patterns in Music Using Sparse Convolutive Non-Negative Matrix Factorization", *Proceedings of the International Symposium on Music Information Retrieval*, pp. 123–128, 2010.

[8] http://www.quaero.org/
[9] http://versamus.inria.fr/