HAL
archives-ouvertes.fr

Inférence de segmentation structurelle par compression via des relations multi-échelles dans les séquences d'accords

Corentin Guichaoua, Frédéric Bimbot

▶ **To cite this version:**

Corentin Guichaoua, Frédéric Bimbot. Inférence de segmentation structurelle par compression via des relations multi-échelles dans les séquences d'accords. JIM 2018 - Journées d'Informatique Musicale, May 2018, Amiens, France. pp.71-79, <http://www.algomus.fr/jim2018>. <hal-01791367>

HAL Id: hal-01791367
https://hal.archives-ouvertes.fr/hal-01791367

Submitted on 14 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INFÉRENCE DE SEGMENTATION STRUCTURELLE PAR COMPRESSION VIA DES RELATIONS MULTI-ÉCHELLES DANS LES SÉQUENCES D'ACCORDS

Corentin Guichaoua
Université de Rennes 1
IRISA – Équipe-projet PANAMA

Frédéric Bimbot
CNRS
IRISA – Équipe-projet PANAMA

RÉSUMÉ

Une des approches possibles pour l'analyse automatique de structure musicale est d'aborder la question comme un problème de compression : une représentation efficace de la structure musicale d'un morceau serait ainsi une représentation qui permet de décrire précisément ce morceau en peu de termes.

La méthode présentée dans cet article exploite les réseaux de relations à courte et moyenne échelle entre les événements musicaux afin d'obtenir des représentations compactes de segments de séquences musicales symboliques. Ces représentations compactes sont ensuite utilisées pour exprimer le morceau analysé comme une suite de segments de faible complexité, d'après le postulat qu'une segmentation découpant un morceau en un nombre limité d'éléments simples est susceptible de s'aligner avec la segmentation annotée manuellement.

Cette méthode est testée sur des séquences d'accords annotées manuellement, correspondant à la base de données RWC-Pop, et est comparée à une autre méthode à base de compression. Les segments obtenus correspondent avec une F-mesure de l'ordre de 70% avec les segmentations de référence.

1. INTRODUCTION

La structure musicale est un concept crucial de l'analyse de la musique, comme en témoigne son utilisation pour des applications variées, telles que la séparation chant-voix [1], la transcription de parole [2] ou la composition automatique [3]. En dépit de son importance, sa définition reste encore très ambiguë [4].

Devant les multiples définitions possible de la structure musicale et la nécessité de trouver un cadre permettant de les regrouper, cet article aborde la question sous l'angle de la théorie de l'information. Plus précisément, nous nous intéressons aux notions de complexité de Kolmogorov et d'entropie de Shannon pour rendre compte de la redondance présente dans les segments musicaux et les compresser. Selon cette optique, la structure musicale réside dans les régularités utilisables pour exécuter cette compression [5].

L'utilisation de données symboliques, comme une partition ou une séquence d'accords au lieu d'un enregistrement, en interposant une couche d'abstraction, permet de faire abstraction des régularités propres au signal audio pour se concentrer sur les régularités musicales. On se limitera ici à l'étude de séquences de triades (échantillonnées à 1 accord par temps), mais les principes décrits peuvent également être appliqués à d'autres types de descriptions symboliques.

Les séquences musicales présentent des régularités plus poussées que ce qui apparaît en examinant simplement leur nature séquentielle. En particulier, on peut observer des dépendances entre les éléments qui sont situés à des positions métriques homologues. Afin de modéliser ces dépendances multi-échelles, nous introduisons un modèle s'appuyant sur des graphes hypercubiques, proche de celui présenté dans [6], développé de façon concomitante dans l'équipe PANAMA, mais axé sur la modélisation de différents type d'organisations au lieu d'une description détaillée. Nous proposons une méthode exploitant ces dépendances afin d'obtenir des représentations compressées de segments musicaux, que nous utilisons ensuite pour trouver une séquence de segments peu complexes.

Dans la section 2, nous introduisons les concepts de *patrons tensoriels* et d'*historique tensoriel*, qui formalisent différents réseaux de relations et les dépendances qui en résultent entre les éléments d'un segment. Chaque patron différent correspond ainsi à une hypothèse sur l'organisation du segment qui, selon son adéquation avec le segment, permet ou non de l'exprimer de façon compacte.

La section 3 décrit comment les réseaux de relations définis par les patrons tensoriels peuvent être utilisés pour tenter d'anticiper successivement chacun des éléments de la séquence.

En section 4, nous appliquons cette méthode de compression des segments pour la segmentation structurelle, en recherchant les séquences de segments résultant en une complexité minimale.

Nous étudions en section 5 l'application de cette méthode de segmentation sur des annotations de séquences d'accords de musique pop, avant de conclure en section 6.

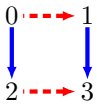
2. PATRONS TENSORIELS

Avant d'utiliser les relations multi-échelle entre les différents éléments d'une séquence musicale, il convient de

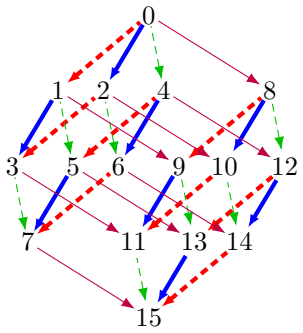
Corentin Guichaoua, Frédéric Bimbot, *Inférence de segmentation structurelle par compression via des relations multi-échelles dans les séquences d'accords*, Journées d'Informatique Musicale (JIM 2018), Amiens, France, éd. L. Bigo, M. Giraud, R. Groult, F. Levé, pp. 71–79, 2018. CC BY-SA 4.0.

définir l'architecture de ces relations. Pour cela, nous introduisons les *patrons tensoriels*, qui décrivent l'organisation des éléments à plusieurs échelles simultanément. Plutôt qu'un modèle universel de l'organisation d'un segment musical, il s'agit d'un catalogue (non-exhaustif) d'organisations simples, parmi lesquelles celle permettant au mieux de décrire la séquence sera sélectionnée (cf. sections suivantes).

L'organisation non-linéaire la plus simple est une organisation de 4 éléments en carré, où chaque élément dépend de ses voisins apparaissant antérieurement dans la séquence, comme illustré en figure 1a. Ces relations sont considérées comme *homologues* lorsque les arêtes associées sont parallèles. Cette organisation peut être généralisée à des séquences de 2^d éléments, organisés selon un hypercube de dimension d (par exemple en dimension 4 sur la figure 1b).



(a) Patron carré (2-cubique)



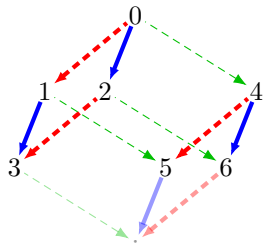
(b) Patron tesseractique (4-cubique)

Figure 1: Exemples de patrons hypercubiques. Les styles de traits indiquent les classes de relations homologues.

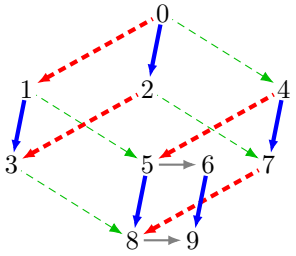
Cette extension, à elle seule, reste toutefois très insuffisante pour modéliser la variété des organisations rencontrées en pratique, même dans des genres considérés comme pauvres à cet égard comme la pop. En effet, on observe souvent des sections musicales dont la longueur en temps musicaux n'est pas une puissance de 2. Afin de pouvoir gérer une plus grande variété de séquences, nous étendons l'inventaire des patrons hypercubiques à l'aide de deux opérations de modification permettant d'augmenter ou de diminuer le nombre d'éléments admis par le patron. La première opération, dite d'insertion, agit sur une hyper-face [1] de l'hypercube définissant le patron, en la dupliquant et en connectant ses éléments avec les éléments correspondants de la nouvelle copie. La seconde opération, dite d'omission, agit également sur une hyper-face, la retirant du patron. Un exemple de chaque cas est présenté en figure 2.

Pour éviter une multiplication excessive du nombre de patrons, on se limite à la manipulation d'hyper-faces dont un des sommets est le dernier élément de l'hypercube, et à une seule insertion et/ou une omission. Cette restriction permet d'identifier les modifications de façon unique par

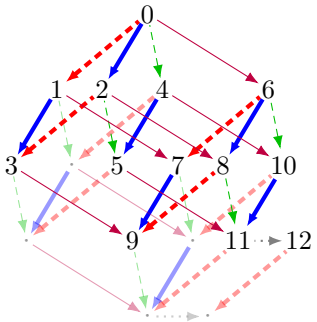
1 . Une hyper-face est un hypercube inclus dans un hypercube de dimension supérieur



(a) Patron avec une 0-omission, admettant une séquence de 7 éléments



(b) Patron avec une 1-insertion, admettant une séquence de 10 éléments



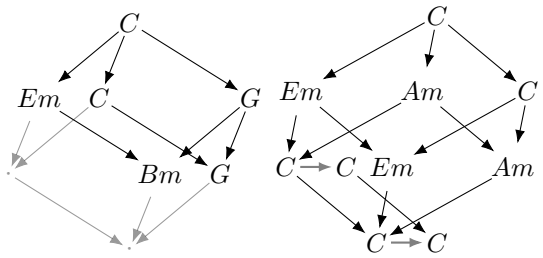
(c) Patron avec une 1-insertion et une 2-omission, admettant une séquence de 13 éléments

Figure 2: Exemples de patrons avec omission, avec insertion et avec une combinaison des deux.

la position du sommet opposé de l'hyper-face. Ainsi, en accord avec le principe d'utiliser des organisations simples, 3 nombres (au plus) suffisent à identifier les patrons : la dimension de l'hypercube, et les positions des modifications éventuelles.

Ces extensions supplémentaires, bien qu'elles ne permettent toujours pas de modéliser toutes les longueurs de séquences, offre un éventail bien plus diversifié d'organisations, qui suffisent à couvrir la majorité de cas rencontrés. La figure 3 montre des exemples de séquences avec des signatures impaires modélisées à l'aide de patrons déformés.

À partir de ces réseaux, on définit pour chaque élément son *historique tensoriel*, le sous-ensemble des relations et des éléments dont dépend cet élément, directement ou transitivement. Deux exemples d'historiques tensoriels sont présentés en figure 4. Ces historiques tensoriels sont pour chaque élément l'information qui est utilisée afin d'établir une anticipation, comme décrit dans la section suivante.



(a) Séquence de structure ternaire modélisée à l'aide d'une omission

(b) Séquence de structure pentadique modélisée à l'aide d'une insertion

Figure 3: Exemples de structures non carrées

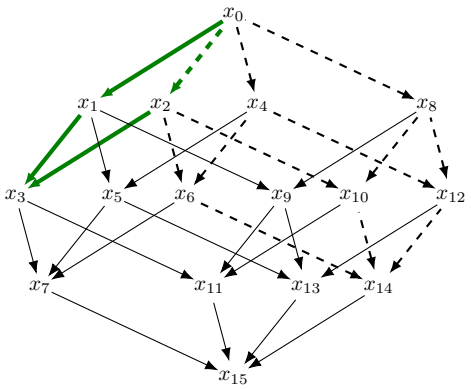


Figure 4: En vert épais l'historique tensoriel de x_3, en pointillé l'historique de x_{14}. Les éléments d'origine des arcs font partie des historiques tensoriels

3. ANTICIPATION SELON UN PATRON TENSORIEL

Une façon de mesurer l'adéquation d'un patron tensoriel à un segment est d'observer dans quelle mesure l'utilisation de ce patron permet d'anticiper chacun des éléments de la séquence correspondante en fonction des éléments passés. D'après les principes énoncés en introduction, un patron approprié permettra ainsi d'anticiper plus d'éléments qu'un patron inadapté. Cette mesure présente l'avantage de pouvoir être effectuée dans un cadre non-supervisé, sans entrée autre que la séquence à modéliser.

Ce procédé peut être exprimé selon un point de vue compressif : pour chaque élément de la séquence, 3 cas peuvent se présenter, qui déterminent la quantité d'information requise pour encoder l'élément :

1. L'élément observé correspond à l'élément anticipé : aucune information supplémentaire n'est nécessaire pour décrire cet élément.
2. L'élément observé ne correspond pas à l'élément anticipé : il est nécessaire de spécifier l'élément observé ou en quoi il diffère de l'élément anticipé (du même ordre dans le cas d'éléments simples).
3. Un élément anticipé n'a pas pu être produit (par exemple le premier élément de la séquence pour lequel aucune information n'est disponible) : il est nécessaire de spécifier l'élément observé.

En effectuant la simplification que le bit utilisé pour distinguer le premier du deuxième cas est négligeable devant la spécification d'un élément, la longueur de la description de la séquence selon le patron tensoriel est proportionnelle au nombre d'éléments n'ayant pas été correctement anticipés.

Avant de décrire le processus par lequel le patron tensoriel est utilisé pour produire des anticipations d'éléments, nous décrivons le modèle que nous utilisons pour la représentation des relations entre les accords [2].

3.1. Relations entre les éléments

Le modèle de représentation des accords et de leurs relations que nous avons choisi d'utiliser consiste en deux simplifications principales.

La première simplification est une restriction du vocabulaire des accords aux 24 triades majeures et mineures [3]. Cette simplification est assez courante dans les approches symboliques et ne sera pas détaillée.

La deuxième simplification porte sur les relations entre les accords deux à deux : ces relations sont modélisées par les rotations envoyant un accord sur un autre sur le cercle des tierces présenté en figure 5. En plus de ses propriétés musicales, par son équivalence avec $\mathbb{Z}/24\mathbb{Z}$, cet espace de représentation des relations a les propriétés clés suivantes (que nous dirons quasi-affines [4]) :

— la composition est définie et respecte la relation de Chasles pour toutes les paires de relations ;
— la composition est commutative et associative ;
— chaque accord a une unique image par une transformation donnée.

3.2. Cas du carré

Le cas le plus simple est celui de l'organisation en carré, qui correspond au modèle Système & Contraste introduit dans [7], illustré avec des formes en figure 6. L'arrangement des 4 éléments x_0, x_1, x_2 et x_3 de la séquence correspond à l'hypothèse d'une relation d'analogie entre ces éléments, où x_3 est à x_2 ce que x_1 est à x_0, soit, en utilisant une notation vectorielle $\overrightarrow{x_0x_1} = \overrightarrow{x_2x_3}$.

En pratique, cette analogie n'est pas toujours vérifiée, on note alors $\overrightarrow{\gamma_3}$ la différence entre l'élément observé x_3 et l'élément *anticipé* $\widehat{x_3}$ qui aurait vérifié l'analogie, que l'on nomme *contraste*.

Cette formulation permet de mettre en avant le rôle joué par chacun des éléments dans l'organisation de la sé-

2 . À l'exception de la section expérimentale, cette description du modèle des accords est la seule partie de cet article qui est spécifique aux accords. À condition de respecter les propriétés clés qui y sont énoncées, tout modèle de représentation des propriétés de la séquence est compatible avec les principes développés.

3 . Cet espace et les relations peuvent être étendus avec un non-accord pour les temps où la notion d'accord ne s'applique pas. Cette solution est un artifice qui permet de traiter par extension les passages où il n'existe pas d'harmonie identifiable.

4 . Ces propriétés s'apparentent à celles d'un espace affine, à ceci près que la multiplication par un scalaire n'est pas toujours inversible.

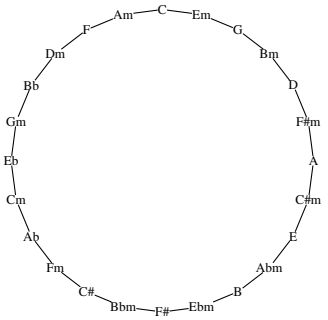


Figure 5: Le cercle des tierces utilisé pour représenter l'espace des accords. Chaque accord partage deux notes avec ses voisins et l'on passe d'un accord à son voisin horaire en décalant d'une tierce (majeure si l'accord de départ est majeur, mineure sinon) vers les aigüs.

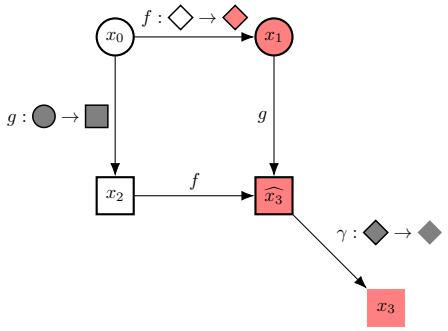


Figure 6: Visualisation des relations entre les éléments du modèle Système & Contraste sur un système d'éléments géométriques. Les relations portent sur le remplissage, la forme et le contour des éléments. L'élément $\widehat{x_3}$ est un élément virtuel qui ne fait pas partie de la séquence observée.

quence (analysée à l'aide du patron carré), et sa contribution à la quantité d'information nécessaire à l'encodage de la séquence. Le premier élément (x_0), appelé *primer*, permet d'ancrer la séquence et sert de point de référence. Il est présenté sans information préalable et doit toujours être encodé. Le deuxième et le troisième élément (x_1 et x_2), appelés sous-primers, établissent les transformations (ou relations) qui définissent le reste de la séquence. Un cas simple se dégage pour chacun de ces éléments, celui où l'élément est identique au primer, pour lequel une quantité d'information faible est requise. Enfin, le dernier élément x_3, en correspondant ou non avec l'élément anticipé (on dit qu'il est *concordant* ou *discordant* respectivement), détermine si la séquence est *contrastive* ou non. Dans le cas où il est concordant, très peu d'information est requise pour l'encoder.

On peut noter que dans le cas d'objets présentant plusieurs propriétés comme dans l'exemple des formes, l'encodage des transformations peut être plus économique en information, même si aucune des transformations n'est triviale. Ce n'est toutefois pas le cas avec notre représentation simplifiée des accords, ce qui limite l'intérêt du cas carré pris en isolation.

3.3. Cas régulier en dimension 3

Une première observation dans le cas des patrons hypercubiques de dimension supérieure est que tous les éléments à l'exception du dernier peuvent être ramenés à des cas de dimension inférieure : par exemple, dans un patron cubique (3-cubique), l'élément x_6 est le dernier élément du carré formé par les éléments x_0, x_2, x_4 et x_6. Ainsi, en procédant en dimension croissante, seule la détermination de l'élément anticipé pour la dernière position doit être adaptée.

Considérons d'abord le cas du cube : comme dans le cas carré, on peut considérer l'analogie définie par $\overrightarrow{x_0x_2} = \overrightarrow{x_1x_3} = \overrightarrow{x_4x_6} = \overrightarrow{x_5x_7}$, illustrée en figure 7. Cependant, contrairement au cas carré, plus d'une de ces relations a été observée dans l'historique tensoriel de x_7, et une contradiction peut être apparue si x_3 ou x_6 est discordant. S'ils sont concordants, on peut définir un élément anticipé pour x_7 d'après le parallélogramme formé par les éléments x_0, x_2, x_5 et x_7, de la même façon que pour une organisation carrée.

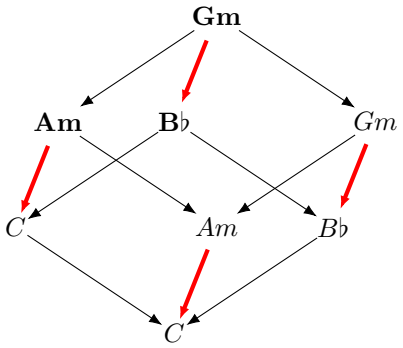


Figure 7: Mise en évidence d'une analogie sur un patron cubique, à partir de la séquence d'accords du pont (1'38 à 1'52) du morceau n°10 de RWC (*Getting Over*). La séquence est ici prise à 1 accord par mesure et peut être entièrement déduite des 3 seuls accords en gras, qui en constituent ainsi une représentation compressée.

Par symétrie, on peut également considérer les analogies utilisant les relations $\overrightarrow{x_0x_1}$ ou $\overrightarrow{x_0x_4}$. Ces analogies supplémentaires sont contredites respectivement si x_3 ou x_5 sont discordants et si x_5 ou x_6, et ne sont donc pas équivalentes, contrairement au cas carré où les deux directions étaient interchangeables.

Un autre type d'analogie que l'on peut effectuer est en considérant les faces parallèles formées d'une part par x_0, x_1, x_2 et x_3, et d'autre part par x_4, x_5, x_6 et x_7. Cette analogie est toutefois équivalente à l'analogie selon la relation $\overrightarrow{x_0x_4}$. Ici encore, on retrouve par symétrie deux autres analogies entre faces, équivalentes aux analogies selon les arêtes.

On peut remarquer que si plus d'un élément parmi x_3, x_5 et x_6 est contrastif, aucune des analogies n'est vérifiée sur les éléments observés : dans ce cas, on ne peut pas construire d'élément anticipé de cette façon, et il est nécessaire d'encoder x_7 non trivialement. Une conséquence

de la nature quasi-affine de l'espace des relations est que les analogies non contredites aboutissent au même élément anticipé. On a alors un unique point de référence auquel comparer l'élément observé x_7 pour l'encoder trivialement s'il est concordant.

3.4. Cas en dimension supérieure

En dimension 4 et au delà, le même processus peut être appliqué en considérant les analogies définies par les hypercubes contenant le primer. Comme dans le cas du cube, ces constructions sont complémentaires 2 à 2 : pour la définition de l'élément projeté pour x_i, l'analogie définie par l'hypercube entre x_0 et x_h est équivalente à l'analogie définie par l'hypercube orthogonal entre x_0 et x_{i-h}. Ces deux analogies, si elles ne comportent pas de contradiction, peuvent être réduites au parallélogramme formé par x_0, x_h, x_{i-h} et x_i. Un exemple de paire d'hypercubes complémentaires et du parallélogramme qui leur est associé est présenté en figure 8.

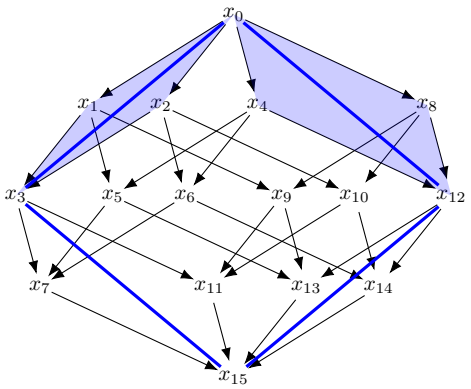


Figure 8: Paire de carrés complémentaires et l'organisation en parallélogramme correspondante

L'inventaire des conditions pour lesquelles l'analogie est contredite par les éléments observés peut sembler laborieux de prime abord, mais s'exprime relativement aisément en fonction de ces hypercubes complémentaires : l'analogie pour x_i basée sur un hypercube entre x_0 et x_h est contredite s'il existe au moins un élément discordant dans l'hypercube entre x_0 et x_i qui n'est ni dans l'hypercube entre x_0 et x_h, ni dans son complémentaire (entre x_0 et x_{i-h}).

Là encore, les propriétés quasi-affines de l'espace de relations assurent que les analogies non-contredites aboutissent au même élément anticipé.

3.5. Influence des déformations

Jusqu'ici, nous avons décrit comment sont construit les éléments anticipés dans le cas des patrons basés sur des hypercubes non déformés. Afin d'utiliser tout l'éventail des patrons définis en section 2, nous étendons ces constructions aux patrons avec omission et/ou insertion.

L'extension aux patrons avec omission est presque triviale : l'omission ne fait que retirer des éléments à anticiper, et grâce à la contrainte sur les positions des hyperfaces supprimées, les éléments supprimés n'apparaissent que dans les historiques tensoriels d'autres éléments supprimés. La re-numérotation des éléments rompt toutefois la compatibilité des formules s'appuyant sur les index, mais cette difficulté est aisément contournée en appliquant ces calculs sur les index non modifiés.

L'extension aux patrons avec insertion présente également cette difficulté mineure, à laquelle s'ajoute le besoin de définition d'un élément projeté pour les éléments insérés. Une solution consiste à restreindre l'historique tensoriel des éléments insérés à l'hypercube formé par l'hyperface dupliquée et sa copie : ce cas se ramène alors localement à un cas hypercubique régulier.

4. APPLICATION À LA SEGMENTATION STRUCTURELLE

Dans la section précédente, nous avons décrit un processus permettant d'évaluer la complexité d'une séquence donnée selon un patron tensoriel donné. Cette complexité n'est pas nécessairement représentative de la complexité de la séquence, toutefois en répétant cette opération pour tous les patrons tensoriels compatibles et en sélectionnant celui permettant la plus forte compression, une approximation plus juste peut être obtenue, à condition que l'éventail des patrons soit suffisamment riche. Le nombre de patrons compatibles pour une longueur donnée ne dépassant pas l'ordre de quelques centaines (dans le pire cas) pour des longueurs correspondant à l'échelle des segments musicaux, une telle exploration exhaustive reste tractable. Afin de prendre en compte la régularité plus grande des patrons hypercubiques, une pénalité, réglée expérimentalement, est appliquée pour chaque type d'altération appliquée au patron.

Afin de passer de cette mesure de complexité d'un segment individuel à l'inférence d'une segmentation d'un morceau entier, nous calculons, pour chaque paire de position de départ et de fin formant un segment ne dépassant pas une taille plafond de 64 éléments [5] la complexité de ce segment. Il en résulte un graphe dirigé représentant l'ensemble des segmentations possibles du morceau, chaque arc étant portant la complexité du segment correspondant. Ce graphe peut être parcouru avec un algorithme de recherche du plus court chemin afin d'obtenir une segmentation minimisant la somme des complexités de chaque segment.

5. EXPÉRIENCES

Nous évaluons cette méthode sur les chansons de musique pop de la base de données RWC-Pop, constituées de 100 chansons composées et interprétées dans le style des musiques du hit-parade aux États-Unis dans les années 80

5 . Ce seuil est fixé a priori au double de l'échelle ciblée pour la segmentation, 32 éléments dans notre cas. Le choix de cette échelle cible implique que les patrons utilisés sont généralement de dimension 5, avec occasionnellement des patrons de dimension 3,4 ou 6.

et au Japon dans les années 90 [8]. Comme point de référence, nous utilisons une autre méthode basée sur le principe de compression, mais utilisant des algorithmes génériques de factorisation de répétitions, détaillée dans [9, Ch3-4].

5.1. Séquences d'accords

À notre connaissance, sur RWC-Pop, aucune base d'annotation n'existe associant à chaque temps un accord. Afin de produire de telles annotations, nous avons recoupé automatiquement les annotations d'accords par plage temporelle de [10] avec les annotations de temps de [11]. Cette base d'annotations, que nous dénotons AUTO, présente toutefois de nombreuses erreurs liées à la quantification des temps (en particulier pour les changements d'accords à contretemps), et nous avons effectué une première passe manuelle de correction des données, produisant ainsi une nouvelle base dénotée CORRECTED. Une seconde passe, focalisée sur la cohérence des séquences ambigües dans les segments répétés, a permis d'obtenir une nouvelle révision dénotée MANUAL. La figure 9 présente un extrait d'une séquence, un exemple détaillé est étudié en fin d'article et le reste des données est disponible en ligne [6] .

```
[...] Gm Gm Gm Gm Am Am Am Am Bb Bb
Bb Bb C C C C Gm Gm Gm Gm Am Am Am Am
Bb Bb Bb Bb C C C C F F F F C C C C Bb
Bb Bb Bb C C C C F F F F C C C C Bb Bb
Bb Bb C C C C F F F F F F F F F F F F
F F F F Dm Dm Dm Dm Dm Dm Dm Dm Bb Bb
Bb Bb Bb Bb Bb Bb Dm Dm Dm Dm Dm Dm Dm
Dm Bb Bb Bb Bb Bb Bb Bb Bb C C C C C C
C C Bb Bb Bb Bb Bb Bb Bb Bb C C C C C
C C C Bb Bb Bb Bb C C C C Gm Gm Gm Gm
Am Am Am Am Bb Bb Bb Bb C C C C Gm Gm
Gm Gm Am Am Am Am Bb Bb Bb Bb C C C C
F F F F C C C C Bb Bb Bb Bb C C C C F
F F F C C C C Bb Bb Bb Bb C C C C F F
F F F F F F F F F F F F F F F F F F
F F F F F F F N N N N
```

Figure 9: Extrait (de 1'38 à la fin) de la séquence d'accords du morceau n°10 de RWC-Pop. On peut noter que l'échantillonnage au temps près est ici redondant (un accord par mesure aurait suffi), mais le suréchantillonnage ne change pas le résultat de l'algorithme.

5.2. Métriques

Afin d'obtenir des métriques faisant écho dans un cadre discret aux mesures utilisées pour la segmentation à partir de l'audio dans la campagne d'évaluation MIREX, nous avons choisi de mesurer la performance des algorithmes d'après la F-mesure, calculée d'une part exactement et d'autre part avec une tolérance de ± 3 temps (ce qui correspond approximativement aux intervalles de tolérance de 0.5 s et 3 s). Les segmentations de références sont issues de [12], adaptées manuellement à des valeurs de temps discrètes.

5.3. Sensibilité aux paramètres

En section 4 nous avons décrit une pénalisation des patrons modifiés pour refléter la complexité accrue des modèles utilisés pour expliquer la séquence. Concrètement, la complexité tensorielle d'un segment X est donnée par la formule

$$\mathcal{C}(X) = \min_P \mathcal{C}(X|P) + \mathcal{C}(P) \tag{1}$$

où $\mathcal{C}(X|P)$ est le nombre d'éléments discordants de X d'après le patron P et

$$\mathcal{C}(P) = \begin{cases} 0 & \text{si } P \text{ est hypercubique} \\ p_+ & \text{si } P \text{ est avec insertions} \\ p_- & \text{si } P \text{ est avec omission} \\ p_+ + p_- & \text{si } P \text{ est mixte} \end{cases} \tag{2}$$

donne la complexité du patron, où p_+ et p_- sont des paramètres dont nous cherchons à évaluer l'influence.

La table 1 expose les résultats obtenus pour différentes combinaisons de ces valeurs. On constate que l'influence de ces paramètres est modérée, à l'exception des cas où les pénalités sont nulles, et que les variations sont relativement lisses. Des expériences préliminaires de sous-échantillonnage confirment que les valeurs optimales sont peu fluctuantes.

Table 1: F-mesure moyenne (sans tolérance, en pourcentage) selon les valeurs des pénalités appliquées aux déformations du patron hypercubique.

p_+ \ p_-	0	0.01	1	2	3	4
0	43.7	48.6	47.9	48.4	48.0	47.6
0.01	48.4	49.8	52.7	54.1	53.1	52.1
1	47.8	52.3	55.5	55.6	55.6	55.2
2	50.2	54.1	56.4	59.5	58.7	56.9
3	50.2	54.6	56.7	59.3	**59.6**	58.6
4	49.7	54.6	56.9	58.4	57.5	57.9

5.4. Régularisation

Un examen manuel des segmentations révèle que les segments obtenus sont souvent pertinents, mais correspondent à une échelle de segmentation inférieure à celle considérée lors des annotations (de l'ordre de 32 temps par segment). Afin de calibrer l'algorithme pour cibler cette échelle (connue a priori), nous introduisons un terme de régularisation privilégiant les segments d'une longueur proche. Ce terme est de la forme suivante :

$$S_r(X) = \begin{cases} S(X) + r_+ \left(|X| - l_r\right) & \text{si } |X| \geq l_r \\ S(X) + r_- \left(l_r - |X|\right) & \text{si } |X| < l_r \end{cases} \tag{3}$$

6. https://gforge.inria.fr/projects/rwcpop-chordseq/

où $|X|$ est la taille de la séquence, l_r est la taille cible des segments et r_+ et r_- sont les coefficients de régularisation.

Table 2: F-mesure (sans tolérance, en pourcentage) selon les paramètres de régularisation. On utilise $p_+ = p_- = 2.25$, la paire de valeurs offrant les meilleurs résultats sans régularisation.

r_- \ r_+	0	0.0001	0.025	0.05	0.075	0.1	0.125	0.15
0	61.1	61.4	61.4	60.8	60.9	60.5	60.2	60.4
0.0001	62.8	62.8	62.5	61.9	61.7	61.2	61.3	61.1
0.025	62.2	62.2	62.4	62.4	62.6	62.7	61.7	61.7
0.05	65.5	65.5	65.3	65.5	66.2	66.1	65.9	64.8
0.075	66.2	66.2	66.7	67.0	66.9	66.7	66.6	66.3
0.1	67.8	67.9	67.0	66.9	66.7	66.7	66.8	66.5
0.125	67.8	67.9	67.1	66.5	66.4	66.1	66.4	66.3
0.15	**68.0**	67.8	67.5	67.5	66.5	66.1	65.9	66.0
0.175	67.4	67.4	67.6	67.2	66.8	66.5	65.9	65.7
0.2	65.9	65.9	66.1	66.5	66.6	65.8	65.3	65.0

La table 2 montre l'évolution de la F-mesure selon le réglage de ces paramètres. L'optimum est atteint pour une valeur de r_+ nulle, ce qui est cohérent avec le but de réduire la sur-segmentation. À l'exception d'un effet de seuil marqué entre les valeurs 0.025 et 0.05 pour r_-, l'influence de ces paramètres est encore modérée et lisse. À nouveau, des expériences préliminaires de sous-échantillonnage confirment les optimums obtenus.

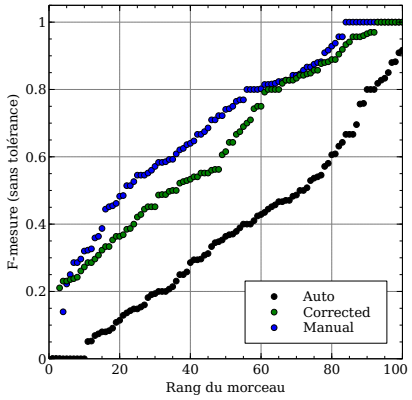
5.5. Choix de la base et variabilité

Les expériences ci-dessus étaient rapportées pour la F-mesure exacte, sur la base MANUAL, pour laquelle les données d'entrée sont de bonne qualité. Les bases AUTO et CORRECTED permettent de tester le comportement de l'algorithme lorsque les données sont à différents niveaux de dégradation (sans aller toutefois jusqu'à des annotations complètement automatiques). La figure 10 montre une vue transversale des performances par morceau de l'algorithme, selon la base et la version de la F-mesure utilisée.

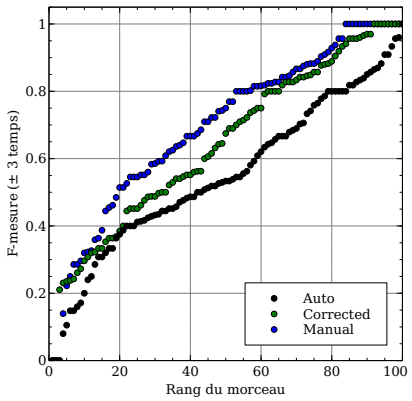
On observe que les différences entre la base MANUAL et la base CORRECTED sont assez modérées, mais qu'en revanche la base AUTO est nettement moins bien segmentée, en particulier dans le cas de la F-mesure exacte. Une autre propriété remarquable est que l'ajout d'une fenêtre de tolérance a une influence très faible sur les 2 meilleures bases, ce qui contraste avec ce que l'on peut généralement observer dans les campagnes MIREX sur la segmentation audio. Enfin, cette figure met en évidence la diversité de la qualité de la segmentation selon les morceaux : si presque la moitié de la base est correctement segmentée à 80% ou plus, 20% des morceaux ont moins de la moitié des frontières de segment correctement inférées.

5.6. Comparaison avec une autre approche

La table 3 rapporte les F-mesures obtenues pour les méthodes présentées dans cet article, comparées à une méthode de référence opérant sur le même type de données et utilisant des méthodes de compression de texte pour l'inférence de la structure [9, ch. 3-4]. Les méthodes à encodage tensoriel obtiennent des scores de segmentation nettement plus élevés.



(a) F-mesure sans tolérance



(b) F-mesure avec tolérance

Figure 10: F-mesure sur chacun des 100 morceaux de RWC-Pop selon l'annotation utilisée pour l'encodage tensoriel par analogie avec régularisation. Pour plus de lisibilité, les morceaux ont été ré-ordonnés par F-mesure croissante selon chaque annotation

5.7. Détail d'une segmentation

Le morceau numéro 24 de la base MANUAL est un morceau pour lequel la F-mesure entre la segmentation inférée est légèrement au dessus de la moyenne (F=0.71). La figure 11 montre la segmentation de référence et la segmentation obtenue par l'algorithme d'inférence.

On peut voir que les segmentations diffèrent à certains endroits :

— La frontière entre le deuxième segment et le troisième segment (0'19 à 0'41 dans l'audio) est placée 16 temps (un demi-segment) plus tard dans la segmentation inférée. En effet, sur la séquence d'accords, la première moitié du 3e segment forme un système non contrastif avec le deuxième segment (il s'agit d'une répétition à l'identique). L'indice principal qui permet de distinguer une frontière à cet endroit est le retour du chant dans l'instrumentation, mais cet indice n'est pas présent dans la séquence d'accords. On retrouve également la même

8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
4 (16) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
5 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab Ab
9.25 (40) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb
Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab
Ab Eb Eb Eb Eb Eb Eb Eb
9 (32) Ab Ab Ab Ab F# F# F# F# Eb Eb Eb Eb Eb Eb Eb Eb
Ab Ab Ab Ab F# F# F# F# Ab Ab Ab Ab N N N
8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
4 (16) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
9.25 (40) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb
Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab
Ab Eb Eb Eb Eb Eb Eb Eb
9 (32) Ab Ab Ab Ab F# F# F# F# Eb Eb Eb Eb Eb Eb Eb Eb
Ab Ab Ab Ab F# F# F# F# Ab Ab Ab Ab Ab N N N
8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
9 (32) Ab Ab Ab Ab F# F# F# F# Eb Eb Eb Eb Eb Eb Eb Eb
Ab Ab Ab Ab F# F# F# F# Ab Ab Ab Ab Ab N N N
12 (56) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Ab
Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm Fm
Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
4 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb N N N N

(a) Segmentation de référence. Score : 103.25

8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
7 (16) Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab
Ab
5 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab Ab
9.25 (40) Eb Eb Eb Eb Eb Eb Eb Eb Ab Ab Ab Ab F# F# F#
F# Eb Eb Eb Eb Eb Eb Eb Eb Ab Ab Ab Ab F# F# F# F# Ab
Ab Ab Ab Ab N N N
8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
7 (16) Eb Eb Eb Eb Eb Eb Ebm Ebm F# F# F# F# Ab Ab Ab
Ab
9.25 (40) Eb Eb Eb Eb Eb Eb Eb Eb Ab Ab Ab Ab F# F# F#
F# Eb Eb Eb Eb Eb Eb Eb Eb Ab Ab Ab Ab F# F# F# F# Ab
Ab Ab Ab Ab N N N
8.25 (40) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm Fm Fm
Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
9 (32) Ab Ab Ab Ab F# F# F# F# Eb Eb Eb Eb Eb Eb Eb Eb
Ab Ab Ab Ab F# F# F# F# Ab Ab Ab Ab Ab N N N
4 (32) Ab Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C Ab
Ab Ab Ab Bb Bb Bb Bb Gm Gm Gm Gm C C C C
8 (24) Fm Fm Fm Fm Bb Bb Bb Bb Gm Gm Gm Gm C C C C Fm
Fm Fm Fm Bb Bb N N
2 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm
4 (32) Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb Eb Eb Ebm
Ebm Eb Eb Eb Eb Eb Eb Ebm Ebm Eb Eb Eb Eb N N N N

(b) Segmentation inférée. Score: 95.25

Figure 11: Segmentations du morceau n°24 de RWC-Pop. Chaque segment est précédé par son score de complexité et sa taille. $(r_+, r_-, p_+, p_-) = (0, 0.125, 2.25, 3)$.

Table 3: F-mesure moyenne (en pourcentage, sans tolérance) pour les méthodes à encodage tensoriel et la méthode de référence

Encodage \ Base	AUTO	CORR	MANUAL
GDU [9]	30	44	52
Tensoriel	33	56	61
Tensoriel régularisé	37	63	69

configuration entre le 6e et le 7e segment.

— La frontière entre le 4e et le 5e segment est placée 8 temps plus tôt dans la segmentation inférée. Ces 8 temps en $E\flat$ (0'56 à 1'00) sont répétés dans le segment suivant et permettent de constituer un système relativement régulier avec le segment suivant. Dans la version audio du morceau, la dynamique et la présence d'un fill de batterie permettent d'écarter cette hypothèse, mais ces indices ne sont encore une fois pas présents dans la séquence d'accords.

— Le troisième segment en partant de la fin (3'01 à 3'28) a été divisé en deux sous-segments. On peut remarquer d'une part que le segment de la référence obtient le même score que les deux segments réunis, et d'autre part que cette sous-division du segment est également plausible d'après l'audio.

6. DISCUSSION ET CONCLUSION

Dans cet article, nous avons présenté une méthode de segmentation structurelle de séquences symboliques basée sur la compression individuelle de chaque segment, selon un modèle exploitant les relations multi-échelle entre leurs éléments. Nous avons expérimenté cette méthode sur des séquences d'accords de musique pop, comparant favorablement les résultats à ceux obtenus à l'aide d'une autre méthode.

Ces résultats, de l'ordre de 70% de F-mesure au temps près, sont proches des taux de concordance qui peuvent être observés entre annotateurs humains lors de campagnes d'annotation participative [13], ce qui montre un fort potentiel des méthodes à compression tensorielle, et plus généralement du paradigme d'inférence par compression.

Pour réaliser ce potentiel, plusieurs axes prometteurs consistent à réduire l'information défaussée aux diverses étapes de la modélisation :

— Les régularités à long terme, complètement ignorées par la présente méthode, sont fréquemment utilisés par des algorithmes d'inférence de structure (musicale ou autre) ; la complémentarité de ces approches suggère que des avancées sont possibles en les hybridant.

— Les modèles de relation entre accords utilisés sont une vue simpliste de l'harmonie, qui gagnerait à être raffinée en relâchant une partie des hypothèses émises en section 3. En particulier, le cas où la relation entre deux accords peut être interprétée de plusieurs façons (transposition chromatique ou dia-

tonique par exemple) semble être une piste ardue mais fructueuse.

— La réduction d'un morceau à une séquence d'accords est une simplification extrêmement brutale. Avec les modèles de relation appropriés, la méthode peut être appliquée à des mélodies, des rythmes, des activités d'instruments, etc. Individuellement, chacune de ces dimensions musicales est une simplification aussi brutale de la musique, mais leur utilisation jointe serait certainement bénéfique.

On peut enfin noter que l'information contenue dans la représentation utilisée par cette méthode dépasse la simple segmentation et inclut à la fois des estimations de la structure interne des segments et desquels de leurs éléments apportent des informations nouvelles.

7. REFERENCES

[1] Antoine Liutkus, Zafar Rafii, Roland Badeau, Bryan Pardo, and Gaël Richard. Adaptive Filtering for Music/Voice Separation Exploiting the Repeating Musical Structure. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 53–56. IEEE, 2012.

[2] Matt McVicar and Daniel P.W. Ellis. Leveraging Repetition for Improved Automatic Lyric Transcription in Popular Music. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3117–3121, 2014.

[3] François Pachet, Alexandre Papadopoulos, and Pierre Roy. Sampling Variations of Sequences for Structured Music Generation. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017.

[4] Jordan B.L. Smith. *Explaining Listener Differences in the Perception of Musical Structure*. PhD thesis, Queen Mary, University of London, UK, 2014.

[5] David Meredith. Analysis By Compression : Automatic Generation of Compact Geometric Encodings of Musical Objects. In *The Music Encoding Conference*, 2013.

[6] Corentin Louboutin and Frédéric Bimbot. Modeling the Multiscale Structure of Chord Sequences using Polytopic Graphs. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, oct 2017.

[7] Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, and Emmanuel Vincent. Semiotic Structure Labeling of Music Pieces : Concepts, Methods and Annotation Conventions. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, 2012.

[8] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC Music Database : Popular, Classical and Jazz Music Databases. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, volume 2, pages 287–288, 2002.

[9] Corentin Guichaoua. *Compression models and complexity criteria for the description and the inference of music structure*. Theses, Université Rennes 1, 2017.

[10] Taemin Cho. Manually Annotated Chord Data Set of US Pop Songs and Popular Music Collection of RWC Music Database, 2011.

[11] Masataka Goto. AIST Annotation for RWC Music Database. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, pages 359–360, 2006.

[12] Frédéric Bimbot, Gabriel Sargent, Emmanuel Deruty, Corentin Guichaoua, and Emmanuel Vincent. Semiotic Description of Music Structure : An Introduction to the Quaero/Metiss Structural Annotations. In *Audio Engineering Society Conference : 53rd International Conference : Semantic Audio*, London, United Kingdom, jan 2014.

[13] Cheng-i Wang and J. Mysore Gautham. Re-Visiting the Music Segmentation Problem With Crowdsourcing. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, pages 738–744, Suzhou, China, 2017.