

Décomposition en blocs autonomes comparables - Une proposition de description et d'annotation de structure pour le traitement automatique des morceaux de musique

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

▶ **To cite this version:**

Frédéric Bimbot, Olivier Le Blouch, Gabriel Sargent, Emmanuel Vincent. Décomposition en blocs autonomes comparables - Une proposition de description et d'annotation de structure pour le traitement automatique des morceaux de musique. Journées d'Informatique Musicale (JIM), May 2010, Rennes, France. inria-00553534

HAL Id: inria-00553534
https://inria.hal.science/inria-00553534

Submitted on 10 Jan 2021

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.

DÉCOMPOSITION EN BLOCS AUTONOMES COMPARABLES

UNE PROPOSITION DE DESCRIPTION ET D'ANNOTATION DE STRUCTURE POUR LE TRAITEMENT AUTOMATIQUE DE MORCEAUX DE MUSIQUE

Frédéric BIMBOT [1] frederic.bimbot@irisa.fr
Olivier LE BLOUCH [2] olivier.le_blouch@inria.fr
Gabriel SARGENT [2] gabriel.sargent@inria.fr
Emmanuel VINCENT [2] emmanuel.vincent@inria.fr

Equipe-Projet METISS

(1) IRISA, CNRS - UMR 6074 - (2) INRIA, Rennes Bretagne Atlantique
Campus Universitaire de Beaulieu, 35042 RENNES cedex, France

RÉSUMÉ

La structure musicale d'un morceau de musique est un concept auquel il est fréquemment fait référence en musicologie, mais pour lequel il n'existe pas de définition communément admise.

Ceci soulève un problème méthodologique lorsque l'on souhaite concevoir et évaluer un procédé d'inférence automatique de structure musicale mais aussi lorsqu'il s'agit de spécifier une procédure d'annotation manuelle permettant de produire de façon reproductible une description structurelle pour un morceau de musique.

Cet article présente une approche dite de *décomposition en blocs autonomes comparables*, fondée sur des principes inspirés du structuralisme et de la linguistique, et fournissant une méthodologie d'analyse à partir de critères simples et basés sur l'expérience musicale de l'annotateur.

Une évaluation préliminaire de la méthode proposée, appliquée à l'annotation auditive de 20 morceaux, a aboutit à une concordance entre annotateurs supérieure à 90 %, puis, ultimement, à une annotation consensuelle.

1. PRÉSENTATION

1.1. Préambule et contexte

Ces derniers mois, des travaux ont été initiés dans notre équipe dans le cadre du projet QUAERO [1], sur l'inférence de structure musicale par des algorithmes automatiques [2].

Les finalités de ces travaux sont multiples, notamment au niveau applicatif : la navigation rapide dans les morceaux de musique, la production automatique de résumés, la détection de couplets et de refrains, le remixage non-supervisé, etc… Mais ces travaux possèdent également un intérêt situé davantage en amont : l'utilisation d'informations structurelles pour la modélisation acoustique et musicale.

Cette thématique de recherche nous a amené à nous interroger préalablement sur la définition même du concept de structure musicale ainsi que sur la façon de spécifier une procédure d'annotation permettant de produire manuellement et de façon reproductible une description structurelle d'un morceau de musique.

Cet article présente le résultat d'une première étape de ces travaux, que nous exposons dans le but de pouvoir échanger des idées et confronter nos points de vue avec les communautés scientifiques des Sciences Pour l'Ingénieur et de la Musicologie

1.2. Cadre du travail

Les notions et la méthodologie présentées dans cet article sont destinées à être appliquées à ce que nous désignerons par la musique « conventionnelle », qui couvre notamment la très grande majorité de la musique populaire actuelle d'inspiration occidentale, mais aussi une fraction significative d'autres familles de musique, notamment une partie de la musique classique. Nous conservons cependant présent à l'esprit que plusieurs types de musique, notamment certains courants de la musique contemporaine, ne se prêtent pas aussi bien, voire pas du tout, à l'approche proposée.

L'entité considérée est le morceau de musique. Toutefois, on peut être amené à s'intéresser à la partie de morceau de musique (section), lorsque le morceau est composé de plusieurs grandes parties sans lien entre elles.

1.3. Objectifs

Le but de ce travail est de proposer une définition opérationnelle d'une certaine acception de la notion de structure musicale et de spécifier une procédure permettant l'annotation quasi-univoque de cette structure par différents annotateurs.

On observe en effet qu'il est fréquemment fait référence au concept de structure musicale dans les ouvrages de théorie musicale, en tant qu'agencement formel régissant la construction d'une œuvre.

Toutefois, plusieurs conceptions de la structure musicale coexistent et il n'existe pas de définition unique communément admise [3]. De plus, les définitions proposées reposent souvent sur le concept musicologique de « phrase musicale » qui est ambigu ou dénué de sens pour de nombreux morceaux.

Ceci soulève donc un problème méthodologique, dès lors que l'on souhaite évaluer un procédé automatique

d'inférence de structure, puisqu'il se pose alors la question de la référence à laquelle comparer le résultat de l'algorithme pour en mesurer les performances. Pour un état des lieux très complet sur la question, voir [4].

L'approche présentée dans cet article vise à définir et à spécifier une méthodologie d'annotation de la structure musicale d'un morceau qui possède les propriétés suivantes :

- *Reproductible* par différents annotateurs
- Basée sur *l'expérience musicale* auditive de l'annotateur (et non sur son expertise musicologique)
- *Déconnectée* de tout a priori algorithmique et de toute application particulière
- *Ne s'appuyant pas* sur des propriétés acoustiques absolues
- *Indépendant*e d'un quelconque *rôle musical* attribué aux éléments structurels (notamment « couplet », « refrain », etc…)
- Applicable à un *éventail de genres musicaux* le plus vaste possible

Au stade actuel de notre travail, nous nous sommes focalisés sur la question de la décomposition en éléments structurels et nous reportons à une phase ultérieure la question de l'étiquetage de ces éléments.

Nous présentons d'abord les fondements sur lesquels repose notre approche, puis nous spécifions un procédé visant à obtenir une décomposition structurelle basée sur ces fondements. Nous évaluons la cohérence de la décomposition obtenue par plusieurs annotateurs sur un ensemble de 20 morceaux de musique, l'annotation se faisant par écoute du morceau sous une forme sonore.

En dépit de son côté nécessairement un peu réducteur du fait de ses visées « opérationnelles », nous pensons que l'approche présentée ici est susceptible d'apporter un éclairage intéressant sur des concepts qui sont habituellement abordés sous l'angle de la création artistique ou de la perception, plutôt que sous celui de l'ingénierie.

1.4. Définitions préliminaires

On considère qu'un morceau possède *3 propriétés de référence* qui peuvent être constantes ou subir des modifications au cours du morceau :

- tonalité/modalité (hauteur et gamme de référence)
- tempo (vitesse d'exécution)
- timbre (instrumentation / texture sonore)

On considère également qu'un morceau de musique possède *4 niveaux d'organisation temporelle :*

- rythme (durée relative des notes)
- harmonie (progression des accords)
- mélodie (hauteur relative des notes)
- paroles (contenu linguistique)

L'ensemble forme 7 strates musicales dont on fait l'hypothèse qu'elles évoluent de façon indépendante au cours d'un morceau.

2. FONDEMENTS

2.1. Concept de structure musicale

L'approche que nous adoptons se place dans le cadre du *structuralisme*, initié par Ferdinand de Saussure dans le domaine de la Linguistique [5] et étendu ensuite à de nombreuses autres disciplines, notamment à la sémiologie musicale [6]. On peut en résumer ainsi le principe général : ce qui détermine la *structure* d'une entité, ce sont essentiellement les relations que ses constituants entretiennent les uns avec les autres au sein de l'entité, indépendamment de la forme et du sens de ces constituants.

Appliquée à notre objet d'étude, cette approche nous amène à considérer un morceau de musique comme le résultat de l'agencement d'un ensemble d'éléments constitutifs, selon un certain processus d'assemblage (dit *processus syntagmatique*). Les éléments constitutifs entretiennent également entre eux des relations qui permettent de les comparer (dites *relations paradigmatiques*) lesquelles s'expriment sous forme de relations d'équivalence. L'ensemble forme un *système* au sens structuraliste du terme, c'est-à-dire une « entité de dépendances internes », selon la définition de Hjelmslev [7].

Le morceau de musique apparaît alors comme une réalisation particulière (ou *observation*) issue de ce système et le problème d'inférence de structure musicale consiste à déterminer, à partir de cette unique observation, la délimitation des éléments constitutifs du morceau (segmentation ou, plus généralement *décomposition*) et l'attribution d'une classe d'équivalence à chacun d'entre eux (*étiquetage*).

Dans ce cadre général, la spécification d'un type particulier de structure musicale passe donc par la définition :

1. de la nature des éléments constitutifs considérés
2. de leur processus d'agencement
3. de la relation d'équivalence permettant la comparaison des éléments

2.2. Hypothèses de travail

Dans les travaux présentés ici, les éléments constitutifs sont supposés communs aux 4 niveaux d'organisation temporelle.

Ils sont limités dans le temps et ils sont agencés principalement par concaténation, avec toutefois possibilité de chevauchement (tuilage). On désigne ces éléments constitutifs par *blocs*.

Un bloc est *autonome*, c'est-à-dire doté d'une cohérence musicale propre. Il est défini par un *début*, une *durée* et une *taille*. La distinction entre durée et taille est explicitée en section 2.3.1.

Un bloc peut se décomposer en un *radical* (lui-même doté d'une cohérence musicale propre) et un ou plusieurs *affixes* (cf. 2.3).

Différentes relations d'équivalence exprimant la *comparabilité* entre blocs peuvent être considérées, notamment : l'isométrie, l'interchangeabilité, la similarité,…

Ainsi, dans le cadre de cet article, l'inférence de structure musicale consiste à effectuer une décomposition du morceau en *blocs autonomes comparables*.

La notion de *bloc autonome comparable* possède des similitudes avec la notion de « phrase musicale », sans toutefois pouvoir y être assimilée. Elle possède également une parenté avec la notion de *grouping structure*, développée dans [8].

2.3. Définitions spécifiques

On précise dans cette section quelques propriétés et définitions relatives aux segments musicaux d'un morceau, qui sont nécessaires pour spécifier la procédure de décomposition en blocs autonomes comparables.

2.3.1. Cohérence musicale

Pour plusieurs de ces définitions, on considère l'effet de transformations telles que la suppression, l'insertion ou la substitution de segments musicaux dans le morceau considéré et on évalue la *cohérence musicale* du résultat ainsi obtenu vis-à-vis de l'ensemble des strates définies au paragraphe 1.4.

La notion de cohérence musicale est délicate à définir. On peut tenter de la préciser en considérant par exemple que la cohérence musicale du morceau est préservée dès lors que la transformation effectuée ne crée pas de singularité morphologique vis-à-vis du morceau d'origine et est compatible avec son agencement syntagmatique : en d'autres termes, qu'on rencontre (ou qu'on pourrait rencontrer) ailleurs dans le même morceau, un passage semblable à celui résultant de la transformation effectuée.

Cette définition possède une part de subjectivité mais, en l'absence d'une formulation plus rigoureuse (qui reste à trouver), elle fournit un critère qui permet à un auditeur humain de décider si il considère que la transformation effectuée est admissible ou non.

2.3.2. Propriétés intrinsèques

Un segment musical est dit :

- *valide,* si il ne présente pas d'incohérence musicale majeure (sur aucune strate)
- *indépendant,* si son écoute isolée produit une impression de complétude chez l'auditeur.
- *itérable,* si son raccordement avec lui-même donne lieu à un segment valide.
- *autonome,* si il est indépendant ou itérable
- *incomplet,* si il n'est pas autonome
- *homogène,* si ses propriétés de référence sont constantes sur la durée du segment

La *taille* d'un segment musical est exprimée en nombre de *snaps*, celui-ci étant défini comme la période à laquelle on claque des doigts pour accompagner la musique à une périodicité approchant le plus possible la seconde.

Une portion (éventuellement non-connexe) de segment musical est qualifiée d'*affixe* s'il est possible de le supprimer tout en conservant un segment musical valide. Si l'affixe est situé en début (resp. en fin) de segment et si il est connexe, il est qualifié de *préfixe* (resp. *suffixe*), sinon, on le désigne comme *infixe* (éventuellement non-connexe).

2.3.3. Propriétés syntagmatiques

Un segment musical au sein d'un morceau est dit :

- *suppressible* si sa suppression du morceau n'entraine pas d'incohérence musicale majeure
- *insérable* à un endroit du morceau si son insertion n'entraîne pas d'incohérence musicale majeure

2.3.4. Propriétés paradigmatiques et relations d'équivalence

Un segment de taille *n'* est dit *réductible* à un autre segment de taille $n \leq n'$ si le segment de taille *n* peut s'obtenir à partir du segment de taille *n'* par suppression de *n'-n snaps*. Si le segment de taille *n* est valide, alors ce dernier constitue un *radical* de celui de taille *n'*.

Les principales relations d'équivalences (déjà évoquées plus haut dans la section 2.2) sont :

- *isométrie* : blocs ou segments de taille identique (isométrie absolue) ou réductibles à des radicaux de même taille (isométrie de radicaux).
- *interchangeabilité* : blocs ou segments qu'on peut interchanger dans un morceau sans introduire d'incohérence musicale majeure.
- *similarité* : blocs ou segments partageant des caractéristiques communes sur tout ou partie des strates musicales (cette ressemblance pouvant porter sur tout le bloc, ou seulement sur les radicaux).
- *isomorphisme :* blocs pouvant être obtenus l'un à partir de l'autre par transformation de leurs propriétés de référence.

2.4. Propriétés complémentaires

Par convention, on suppose que les frontières des blocs sont synchronisées avec les débuts de mesures (les anacrouses étant considérées comme faisant partie du bloc auquel elles se rattachent).

Les blocs ne sont pas nécessairement homogènes : leurs propriétés de référence peuvent varier au sein d'un bloc.

Les blocs d'un même morceau sont potentiellement répétés (à des variations près) ou isomorphes.

3. SPÉCIFICATIONS

La définition des blocs est ambiguë pour deux raisons.

D'une part, les notions d'*autonomie* et de *comparabilité* ne sont pas univoques ; elles peuvent s'avérer subjectives et

dépendantes du morceau étudié. D'autre part, à elles seules, ces notions ne mènent pas nécessairement à une structure unique

Pour lever ces ambiguïtés, on convient :

1) d'attribuer le même poids à toutes les strates sonores (tonalité, tempo, timbre, mélodie, rythme, harmonie, paroles) quel que soit le morceau étudié
2) d'ajouter des contraintes supplémentaires, dits *critères de décomposition*, explicités dans la sous-section suivante

On tente de formuler les critères de décomposition sans faire appel à des propriétés acoustiques absolues, ni au rôle musical des segments, de sorte à être le plus indépendant possible du genre musical auquel se rapporte le morceau.

3.1. Critères de décomposition

3.1.1. Test de suppressibilité

L'ensemble des blocs d'un morceau constitue un ensemble de segments interchangeables (au moins au sein d'une même classe d'équivalence). Il est toutefois difficile de tester cette propriété, à laquelle on préfère une propriété plus faible, mais plus pratique : la suppressibilité (définie en section 2.3.2).

Le test de suppressibilité vise à *identifier les principales hypothèses* plausibles de frontières de blocs. Cependant, tout segment suppressible ne constitue pas nécessairement un bloc.

3.1.2. Régularité de la structure

On fait l'hypothèse que tout morceau est bâti sur un ensemble minimal de *pulsations structurelles* qui régissent la taille des blocs, préférentiellement :

- une valeur (type I)
- deux valeurs (type II)
- une suite régulière de valeurs, dit « patron structurel » (type III)
- un ensemble limité de valeurs mais une absence de régularité structurelle (type IV)
- l'existence de blocs identifiables mais un large éventail de tailles différentes (type V)

On désignera par morceau de type 0 (ou indéterminable) tout morceau pour lequel on constate l'impossibilité de délimiter des frontières bien localisées de blocs autonomes.

On distingue :

- les *blocs réguliers*, majoritaires, dont la taille *n* est conforme à la pulsation structurelle (ou cohérente avec le patron structurel). Dans ce cas, le bloc sera annoté comme étant de taille *n*.
- les *blocs dérivés* de blocs réguliers, dont la taille est proche de l'une des pulsations structurelles et qui peuvent se décrire simplement :
  - soit par ajout d'un (ou plusieurs) suffixe(s), préfixe(s) ou infixe(s) à un bloc régulier (bloc dit *rallongé*)
  - soit par suppression en position finale, initiale ou quelconque d'une (ou plusieurs) portion(s) de bloc par rapport à un bloc régulier (bloc dit *tronqué*).

  Le *radical structurel* dont est supposé issu le bloc dérivé est généralement observé (à des variantes près) dans le même morceau (ou pourrait l'être sans sembler incohérent avec le reste du morceau).
- les *blocs irréguliers* dont la taille est différente et généralement éloignée de la (ou des) pulsation(s) structurelle(s) et/ou qui ne peuvent pas se rapporter clairement à un radical structurel. Il s'agit souvent de blocs de taille inférieure au moins de moitié à la pulsation structurelle (*raccord ou pontage*).

En outre, on autorise le *chevauchement de blocs* ou *tuilage*, c'est-à-dire un début de bloc anticipant la fin du précédent pour rendre compte de phénomènes de démarrage de blocs asynchrone avec la fin du précédent, que l'on peut observer dans certaines constructions de morceaux (notamment les canons ou les fugues).

A cet égard, il est important de distinguer le tuilage de l'anacrouse. Dans l'anacrouse, le contenu d'un ou plusieurs niveaux d'organisation temporelle débute avant la frontière de bloc selon un mécanisme d'*anticipation*. Au contraire, dans le tuilage, la frontière d'un nouveau bloc est observée alors que le bloc précédent n'est pas encore terminé, donnant lieu à la *superposition* des deux blocs et à leur coexistence sur tous les niveaux d'organisation temporelle pendant la durée de leur tuilage.

La contrainte de régularité vise à obtenir des *blocs comparables* au sein de chaque morceau.

La figure 1 présente un schéma de décomposition en blocs structurels dans le cas d'une décomposition parfaitement régulière de pulsation structurelle *n* puis illustre différentes configurations d'irrégularités pouvant être rencontrées.

3.1.3. Durée cible des blocs

En appliquant les deux critères précédents, on peut parfois aboutir à plusieurs solutions possibles composées de blocs suppressibles et présentant une bonne régularité. En général, ces solutions correspondent à des valeurs de pulsation structurelle qui sont multiples ou sous-multiples les unes des autres.

On départage ces situations par un critère de durée cible du bloc, basé sur la minimisation de ce que nous désignons par *contexte informatif prédominant.*

La figure 2 illustre une décomposition structurelle sous forme d'une représentation paradigmatique, mettant en évidence la correspondance existant entre portions homologues dans les différents blocs.

Si cette décomposition est exploitée pour effectuer une prédiction des propriétés musicales du morceau sur un court intervalle de temps, les portions les plus pertinentes pour y parvenir seront d'une part les portions appartenant au

même bloc structurel et d'autre part les portions homologues dans les autres blocs structurels.

Ce contexte informatif prédominant est distinct pour chaque portion du morceau mais il est déterminé uniquement par la structure et il constitue la partie prépondérante de l'entité de dépendances internes évoquée dans la section 2.1.

Si la longueur totale du morceau est égale à N et si la longueur typique des blocs est égale à n, alors le nombre de blocs est de l'ordre de N/n et le contexte informatif prédominant a une longueur cumulée d'environ :

$$C = n + (N/n) - 2 \qquad (1)$$

C est alors minimal lorsque $n = \sqrt{N}$.

Pour des morceaux de musique d'une durée typique de 4 minutes (soit $N = 240$, en prenant la seconde comme unité), la valeur de n minimisant C vaut environ 15.49 secondes.

Dans le travail présenté ici, on a retenu cette valeur (arrondie) comme valeur cible pour la pulsation structurelle ce que l'on exprime ainsi : au moins une des valeurs de pulsation structurelle doit être la plus proche possible, en échelle logarithmique, de la *durée cible* de 15 secondes. Avec ce critère, la décomposition structurelle obtenue tend à équilibrer le poids de l'axe syntagmatique et de l'axe paradigmatique dans leur contribution à la description de la structure du morceau.

Plus généralement, on pourra donner un poids relatif λ différent à chacun des deux termes de l'équation (1), conduisant à une fonction du type :

$$C(\lambda) = n + \lambda(N/n) - (\lambda+1) \qquad (2)$$

dont la minimisation (en n) induira des décompositions basées sur un équilibre différent.

Le recours à une durée cible permet de résoudre les situations ambiguës telle qu'un bloc composé de plusieurs sous-blocs identiques et favorise l'obtention de blocs comparables entre plusieurs morceaux.

Cette contrainte peut toutefois être relâchée lors de variations importantes du tempo à l'intérieur d'un bloc ou à l'échelle de plusieurs blocs successifs.

3.1.4. Détermination des affixes et des troncatures

Un bloc de taille $n+p$ supérieure à la pulsation structurelle n sera préférentiellement décomposé comme un radical de taille n accompagné d'un affixe p (notations $n+p$, $p+n$ et $n\&p$) si :

- Le radical est autonome et de préférence observé seul dans le morceau tel quel ou sous une forme similaire.
- L'affixe de taille p est suppressible mais pas autonome sans le radical de taille n.
- La valeur de p est plus petite que $n/2$

Un bloc de taille $n-p$ inférieure à la pulsation structurelle n sera préférentiellement décomposé comme un radical de taille n tronqué (ou privé) d'un sous-bloc de taille p (notations $n-p$, $-p+n$ et $n\backslash p$) , si :

- Le bloc peut être complété par un sous-bloc de taille p pour former un bloc régulier de taille n, autonome et de préférence observé dans le morceau tel quel ou sous une forme similaire.
- La valeur de p est plus petite que $n/2$

3.1.5. Recours au test d'interchangeabilité

Dans certaines situations ambiguës, notamment lorsque la pulsation structurelle a été fixée et que plusieurs segmentations compatibles avec cette pulsation structurelle mènent à des blocs suppressibles, on pourra avoir recours à un test d'interchangeabilité pour départager les différentes solutions possibles.

3.1.6. Situations particulières

Les tailles fractionnaires seront utilisées si l'annotateur estime que la taille d'un bloc correspond à un nombre non-entier de *snaps*.

Dans certains cas, l'annotateur jugera qu'il est impossible de définir un bloc, ou d'attribuer une taille à un bloc ou à un infixe. Ceci sera annoté par un symbole spécial (tableau 1).

Un symbole particulier sera utilisé dans les cas suivants : incertitude de l'annotateur sur la délimitation du bloc, incertitude sur la taille du bloc, taille indéterminable, segment de silence précédant le début ou suivant la fin du morceau, ou segment ne faisant pas partie du morceau.

3.2. Procédure de référence

A partir des critères énoncés ci-dessus, nous décrivons une procédure destinée à permettre à un annotateur humain de produire une décomposition structurelle conforme aux spécifications proposées.

1) *Déterminer une valeur de la pulsation structurelle*

   Déterminer une ou plusieurs hypothèses de taille n en s'intéressant aux parties structurellement les plus stables, notamment à distance de l'introduction et de la conclusion du morceau ou de la partie de morceau. Privilégier en priorité les hypothèses correspondant à des durées compatibles avec la durée de référence.

2) *Décomposer partiellement le morceau par rapport à la pulsation structurelle n*

   Identifier les hypothèses de début de blocs et déterminer les blocs suppressibles de taille n (réguliers) ou dérivant simplement (par extension ou troncature) de blocs réguliers de taille n (existants ou plausibles). Cette étape est également l'occasion de rechercher des tuilages.

3) *Compléter la décomposition*

   Compléter la décomposition partielle par des blocs irréguliers, notamment des « pontages » de taille n/2, n/4 ou des blocs complètement irréguliers. S'assurer à cette

étape que les blocs ainsi formés sont suppressibles (sinon, il convient de les rattacher à des blocs déjà existants).

4) *Evaluer la régularité de la décomposition obtenue selon les indices suivants*
   - forte majorité de blocs réguliers
   - répartition équilibrée entre plusieurs pulsations structurelles / corrélées à la macro-structure (couplet / refrain, sections, etc…)
   - patron structurel simple à décrire (répétitif, progressif, etc…)

   Pour évaluer la régularité de la structure, on ne tiendra pas compte de façon primordiale de la décomposition de l'introduction ou de la fin du morceau,. On n'attachera pas trop d'importance aux « pontages » (surtout si il s'agit de blocs dont la taille est égale à la demi-pulsation structurelle et qui jouent manifestement le rôle de transitions musicales) et on accordera une faible importance aux suffixes, aux préfixes et aux infixes (dans un ordre décroissant de régularité).

5) *Envisager d'autres solutions*

   Considérer la/les décomposition(s) qui pourraient être obtenues avec d'autres hypothèses de pulsation structurelle et évaluer si elles donnent un résultat préférable au précédent.

3.3. Conventions d'annotation

Le tableau 1 résume l'ensemble des conventions d'annotation adoptées pour décrire les tailles de blocs.

3.4. Tableau de bord et rapport d'annotation

Au cours (ou à l'issue) de l'étude d'un morceau, l'annotateur peut s'appuyer sur un « tableau de bord » généré automatiquement et qui synthétise les principales propriétés de l'annotation produite pour le morceau considéré, notamment :

- La distribution des tailles des blocs
- Les durées moyennes pour chaque taille (ainsi éventuellement les valeurs médianes, minimales, maximales …)
- La taille des tuilages
- Une représentation par automate minimal de la décomposition produite et différentes valeurs d'entropie correspondante (propre, croisée, …)

Par ailleurs, l'annotateur indique sur un compte-rendu d'annotation les informations suivantes :

- Type d'organisation structurelle (0, I, II, III, IV, V)
- Difficulté de l'annotation (sur une échelle de 1 à 5)
- Degré de confiance dans le résultat produit (sur une échelle de 1 à 5)
- Observations (c'est-à-dire toute information qu'il juge pertinente)

Le tableau 2 illustre des exemples de tableaux de bord pour l'annotation de 2 morceaux.

4. ÉVALUATION

4.1. Objectifs

Cette section présente les résultats d'un processus d'évaluation préliminaire visant à valider la pertinence de l'approche proposée. Il s'agit de mesurer la concordance des annotations produites par différents annotateurs sur un même ensemble de morceaux et sa progression en fonction de l'évolution des consignes.

4.2. Protocole d'évaluation

On sélectionne un ensemble de morceaux et on les fait annoter par plusieurs annotateurs, selon les spécifications présentées ci-dessus. On évalue la cohérence des segmentations en prenant les annotations deux à deux et en mesurant la précision *P*, le rappel *R* et la F-mesure d'une segmentation par rapport à l'autre.

P, *R* et *F* sont des métriques désormais classiques dans le domaine du traitement de l'information : la précision et le rappel entre deux segmentations correspondent au pourcentage de frontières de blocs concordantes, avec une tolérance temporelle $\pm\delta t$, en prenant l'une puis l'autre des annotations comme référence.

On calcule ensuite la F-mesure : $$F = 2\frac{PR}{P+R} \qquad (3)$$

pour chaque morceau, puis on moyenne ces valeurs sur l'ensemble du corpus.

Dans les expériences présentées dans cet article, le corpus de morceaux est composé de 20 titres dont la liste a été déterminée par l'IRCAM et qui ont servi d'ensemble de développement dans le cadre de la première phase du projet QUAERO pour la tâche 6.5 du projet CTC. La liste est donnée dans le tableau 3.

Quatre annotateurs ont participé à l'expérience. Aucun d'entre eux n'est musicologue, ni musicien professionnel. Toutefois, il convient de préciser qu'il s'agit des quatre auteurs de cet article, ce qui constitue un biais méthodologique qu'il faut prendre en compte dans l'interprétation des résultats.

La précision, le rappel et la F-mesure ont été calculés par le programme utilisé dans le cadre de la campagne d'évaluation QUAERO de Septembre 2009, réalisé par l'IRCAM et validé par l'IRIT [9]. La tolérance δt a été prise égale à 0,75 secondes.

4.3. Résultats

Le tableau 4 résume les résultats obtenus à l'issue d'une première session d'annotation.

Les annotations obtenues à l'issue de la première session ont été passées en revue collectivement pour discuter des discordances observées et notamment les imprécisions de

frontières et les situations où les spécifications n'avaient pas été respectées par tel ou tel annotateur. Ceci a donné lieu, après une période de « décantation » d'un mois environ, à une révision des annotations dont la concordance est résumée dans le tableau 5 ci-après. La figure 3 présente 3 exemples d'annotations multiples après révision.

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	Natalies 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

Tableau 3 : *Liste des morceaux utilisés pour les expériences décrites dans cet article.*

Annotateur	N°1	N°2	N°3	N°4
N°1	100	79.1	90.7	86.6
N°2	79.1	100	77.7	74.5
N°3	90.7	77.7	100	84.5
N°4	86.6	74.5	84.5	100

Tableau 4 : *Concordance des annotations à l'issue d'une session initiale, mesurée en terme de F-mesure (en %) sur 20 morceaux. La moyenne globale des concordances inter-annotateurs s'établit à 82.2 %.*

Annotateur	N°1	N°2	N°3	N°4
N°1	100	88.9	95.7	92.9
N°2	88.9	100	88.7	88.7
N°3	95.7	88.7	100	92.8
N°4	92.9	88.7	92.8	100

Tableau 5 : *Concordance des annotations à l'issue de la phase de révision, mesurée en terme de F-mesure (en %) sur 20 morceaux. La moyenne globale des concordances inter-annotateurs s'établit à 91.3 %.*

On observe, qu'après révision des annotations, la concordance inter-annotateur s'établit au-delà de 90 % ce qui valide globalement la cohérence du processus proposé, tout en laissant une petite marge d'amélioration potentielle. La figure 4 détaille la distribution des scores de concordance par morceaux. La médiane est à 95.8 %. Deux morceaux sont responsables de près de 4% des erreurs (sans eux, la concordance moyenne dépasserait 95 %).

Une part importante des discordances résiduelles observées provenant d'ambiguïtés de groupement de segments en nombre impair ayant une taille proche de la demi-pulsation structurelle, il a été décidé de traiter ces cas en adoptant un principe de « groupement par ressemblance ». Il s'agit d'un critère subsidiaire qui consiste, en cas de solutions multiples, à grouper dans un même bloc les segments qui se ressemblent et dissocier ceux qui sont dissemblables.

Ainsi, dans une phase ultime, une session de travail collective a permis d'aboutir à une décomposition consensuelle pour 19 des 20 morceaux, satisfaisant globalement les 4 annotateurs et l'ensemble des critères de décomposition. Le 20ème morceau (n°11) a été classé en catégorie 0 (frontières de blocs indéterminables).

5. CONCLUSIONS

Les travaux présentés dans cet article ont permis de spécifier une procédure d'annotation visant à produire manuellement et de façon raisonnablement reproductible une description structurelle d'un morceau de musique. Conformément aux objectifs visés, la procédure proposée ne s'appuie pas sur une expertise musicologique, ni sur une analyse du rôle musical des éléments constitutifs du morceau.

Une prochaine phase du travail consiste à décomposer selon ces principes une base de données musicales plus large (de l'ordre de 200 morceaux), ce qui permettra dans le même temps de conforter et d'affiner la méthodologie proposée.

Enfin, une procédure sera définie pour attribuer des « étiquettes » aux blocs obtenus, de sorte à enrichir les informations structurelles en tenant compte des similarités entre blocs. Une piste consiste à utiliser un jeu d'étiquettes par strate et aboutir ainsi à une description structurelle multi-dimensionnelle.

6. RÉFÉRENCES

[1] Projet QUAERO : http://www.quaero.org

[2] Gabriel Sargent, Frédéric Bimbot, Emmanuel Vincent : Un système de détection de rupture de timbre pour la description de la structure des morceaux de musique. JIM 2010.

[3] MIREX 2009 – Structural Segmentation : http://www.music-ir.org/mirex/2009

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

[5] F. de Saussure : Cours de Linguistique Générale. 1916.

[6] http://fr.wikipedia.org/wiki/Sémiologie_de_la_musique

[7] Louis Hjelmslev : Prolégomènes à une théorie du langage (1943).

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

[9] Quaero project : ID.CTC.12.R6.5.P2_Mars.pdf

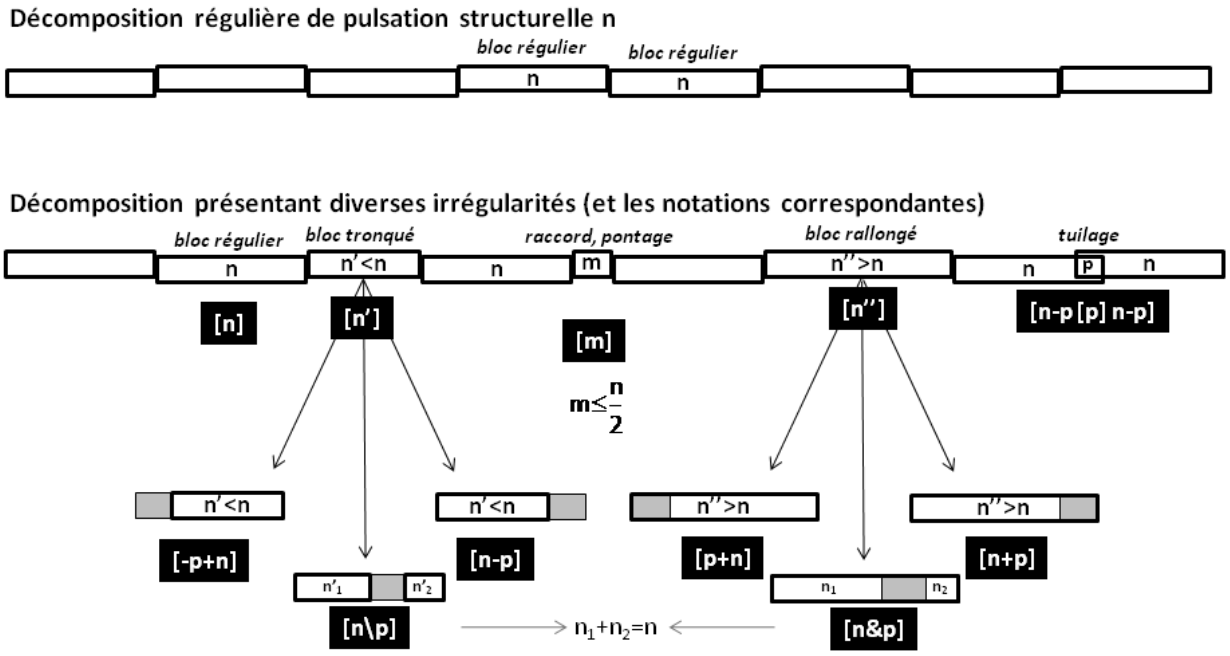


Figure 1 : *Décomposition en blocs structurels (vision syntagmatique)*
Illustration de diverses configurations de blocs irréguliers.

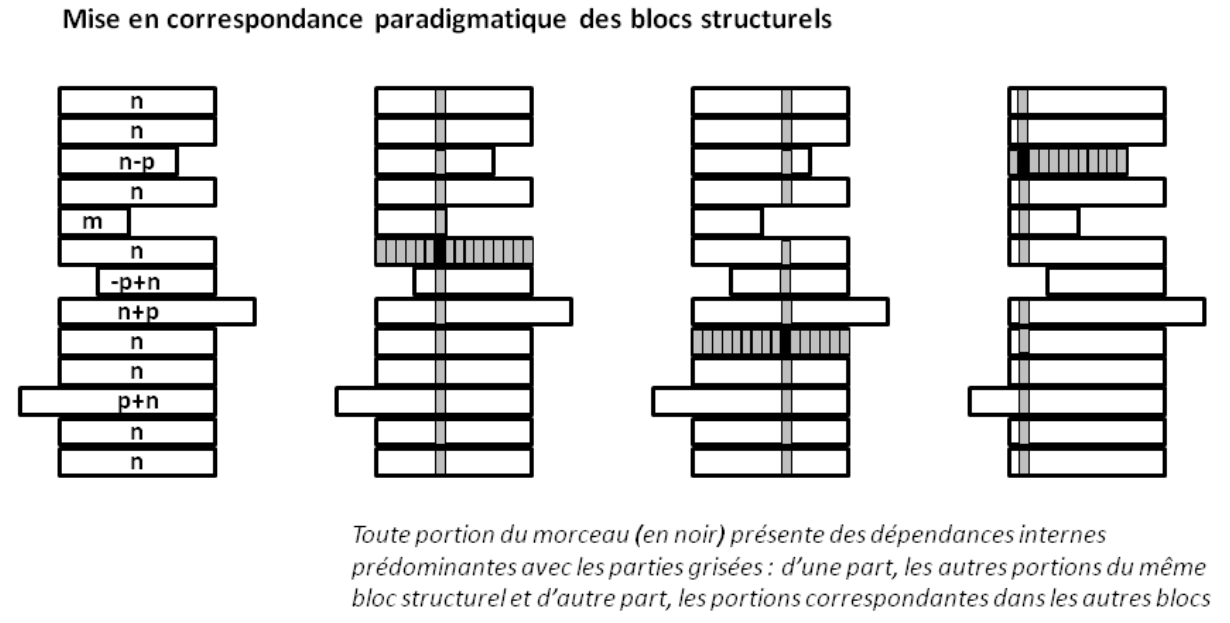


Figure 2 : *Décomposition en blocs structurels (vision paradigmatique)*
Illustration du contexte informatif prédominant.

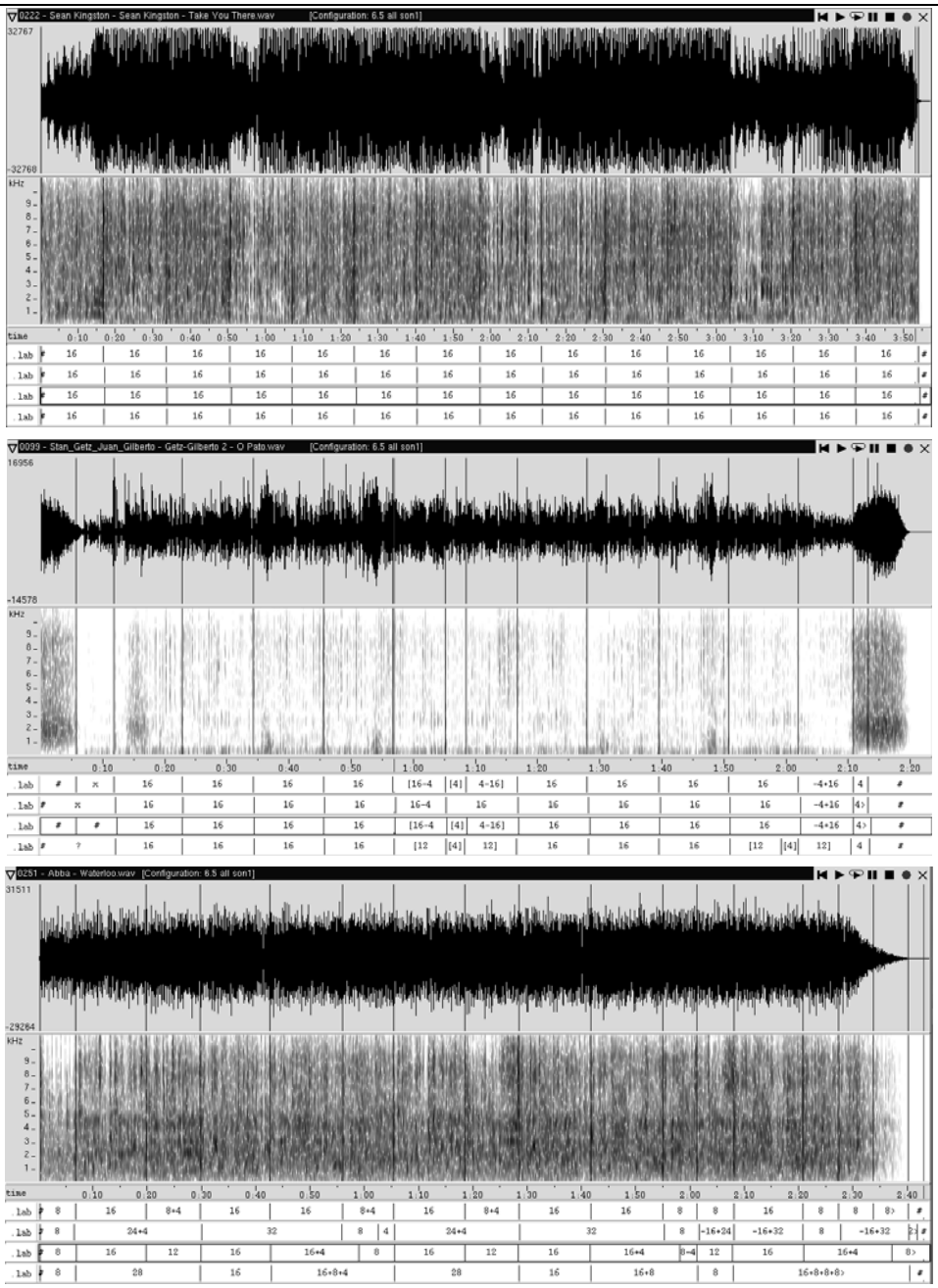


Figure 3 : *Exemples d'annotations obtenues pour 3 morceaux.*
De haut en bas : concordance totale (100 %), moyenne (93,7 %) et faible (71,0 %).
Visualisation avec le logiciel Wavesurfer.

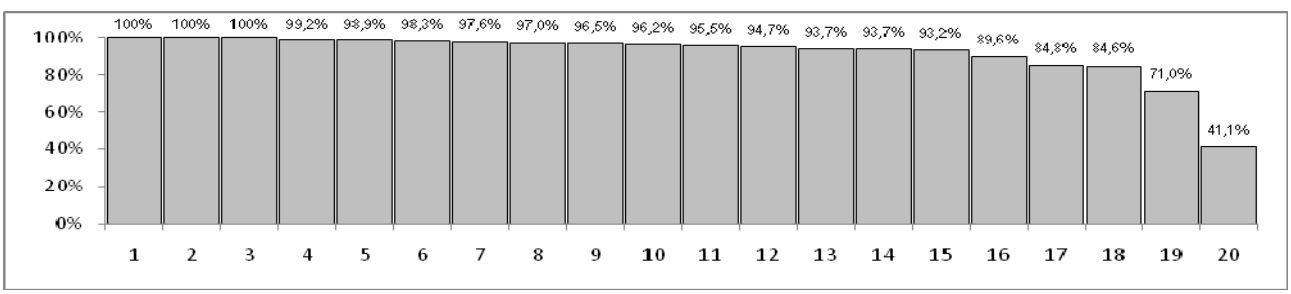


Figure 4 : *Taux de concordance entre annotateurs rangé par ordre décroissant sur les 20 morceaux étudiés.*

Situation	Notation	Exemple
Bloc de taille *n*, régulier	n	16
Bloc de taille *n-p*, obtenu par suppression de la fin d'un bloc régulier	n-p	16-4
Bloc de taille *n-p*, obtenu par suppression du début d'un bloc régulier	-p+n	-4+16
Bloc de taille *n-p*, obtenu par suppression de *p snaps* d'un bloc régulier	n\p	16\4
Bloc de taille *n+p*, obtenu par ajout d'un suffixe de taille *p*	n+p	16+4
Bloc de taille *n+p*, obtenu par ajout d'un préfixe de taille *p*	p+n	4+16
Bloc de taille *n+p*, obtenu par ajout d'un infixe de *p snaps*	n&p	16&4
Tuilage de 2 blocs de taille *n* avec chevauchement de taille *p*	[n-p [p] n-p]	[12 [4] 12]
Bloc irrégulier de taille *m*	m	8
Bloc initial du morceau, en fade-in	<n	<16
Bloc final du morceau, en fade-out	n>	16>
Bloc de taille fractionnaire	valeur décimale	7.5
Délimitation du bloc incertaine pour l'annotateur	??	??
Taille incertaine pour l'annotateur ou qu'il ne parvient pas à déterminer	n? ou ?	16? ou ?
Taille que l'annotateur considère comme indéterminable	x	x
Segment ne faisant pas partie du morceau	#	#

Tableau 1 : *Conventions d'annotation des tailles de blocs structurels.*

Référence du morceau : 13

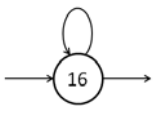
Durée totale : 234 s

Taille	Nombre	Taille x Nombre
16	14	224
16	**14**	**224**

Tuilages (taille totale)	**0**

	Taille	Durée
Pulsation structurelle T1	**16**	**16,7 s**

Automate minimal



Type d'organisation structurelle : I
Difficulté d'annotation : 1
Degré de confiance : 5
Observations : *R.A.S.*

Référence du morceau : 18

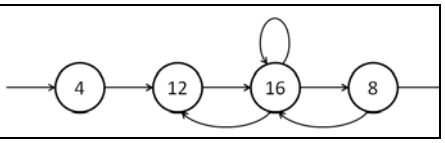
Durée totale : 207 s

Taille	Nombre	Taille x Nombre
16	8	128
12	4	48
8	3	24
4	1	4
12,75	**16**	**204**

Tuilages (taille totale)	**0**

	Taille	Durée
Pulsation structurelle T1	**16**	**16,2 s**
Pulsation structurelle T2	**12**	**12,1 s**

Automate minimal



Type d'organisation structurelle : II/III
Difficulté d'annotation : 2
Degré de confiance : 4
Observations : *Les blocs de 12 peuvent éventuellement être vus comme des blocs dérivés de type 16\4.*

Tableaux 2a et 2b : *Exemples de tableaux de bord résultant de l'annotation de 2 morceaux.*