Institut National des Langues et Civilisations Orientales

Département Textes, Informatique, Multilinguisme

Formes rythmiques pour la transcription automatique de la batterie

MASTER

TRAITEMENT AUTOMATIQUE DES LANGUES

Parcours :

Ingénierie Multilingue

par

Martin DIGARD

Directeur de mémoire :

Damien NOUVEL

Encadrant :

Florent JACQUEMARD

Année universitaire 2020-2021

TABLE DES MATIÈRES

Liste des figures

LISTE DES TABLEAUX

Introduction générale

Ce mémoire de recherche, effectué en parallèle d'un stage à l'Inria dans le cadre du master de traitement automatique des langues de l'Inalco, contient une proposition originale ainsi que diverses contributions dans le domaine de la transcription automatique de la musique. Les travaux qui seront exposés ont tous pour objectif d'améliorer qparse, un outil de transcription automatique de la musique, et seront axés spécifiquement sur le cas de la batterie.
Nous parlerons de transcription musicale, en suivant des méthodes communes au domaine du traitement automatique des langues (TAL) plutôt que directement de langues naturelles, et nous parlerons aussi de génération automatique de partitions de musique à partir de données audio ou symboliques. En considérant que la musique, à l'instar des langues naturelles, est un moyen qui nous sert à exprimer nos ressentis du monde et des choses, ce travail reposera sur une citation de l'ouvrage de Danhauser [1] : « La musique s'écrit et se lit aussi facilement qu'on lit et écrit les paroles que nous prononçons. »
L'exercice exposé dans ce mémoire nécessitera donc la manipulation d'un langage codifié (solfège, durées, nuances, volumes) qui peut être analysé à l'aide de théories formelles et d'outils adéquats comme des grammaires et soulèvera des problématiques qui peuvent être résolues par l'utilisation de méthodes issues de l'informatique et de l'analyse des langues et des langages.

L'écriture musicale offre de nombreuses possibilités pour la transcription d'un rythme donné. Le contexte musical ainsi que la lisibilité d'une partition pour un batteur entraîné conditionnent les choix d'écriture. Reconnaître la signature rythmique principale d'un rythme, la façon de regrouper les notes par des ligatures, ou simplement décider d'un usage pour une durée parmi les différentes continuations possibles (notes pointées, liaisons, silences, etc.) constituent autant de possibilités que de choix de représentation à réaliser. De plus, la batterie est dotée d'une écriture spécifique par rapport à la majorité des instruments.

La proposition principale de ce mémoire est basée sur la recherche de rythmes génériques sur l'*input*. Ces rythmes sont des *patterns* standards de batterie définis au préalable et accompagnés par les différentes

combinaisons qui leur sont propres. On les nomme formes rythmiques (voir section 3.5). L'objectif des formes rythmiques est de fixer des choix le plus tôt possible afin de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces choix concernent notamment la signature rythmique (voir section 3.5) et les règles de réécriture (voir section 3.4).

La proposition ci-dessus a nécessité plusieurs sous-tâches :

- une description de la notation de la batterie (voir 3.1 ainsi qu'une modélisation pour la transcription de la batterie 3.3) qui était jusqu'à présent inexistante ;
- l'écriture de script lilypond [1] pour toutes les figures de ce mémoire qui correspondent à des partitions de batterie (voir la section 3.2) ; tous les codes lilypond sont disponibles sur https://github.com/MartinDigard/Stage_M2_Inria ;
- plusieurs trancriptions manuelles dans le but d'analyser les contenus des fichiers MIDI (voir section 4.2) et de faire des comparaisons de transcriptions avec des outils déjà existants [2] ;
- une partition de référence transcrite manuellement sur l'entièreté d'une performance du jeu de données afin de repérer les éléments importants pour la modélisation et d'établir le rapport entre les données d'*input* et l'écriture finale (voir 3.11) ; cette partition avait aussi pour objectif d'effectuer des tests et des évaluations ;
- la création de grammaires hors-contexte pondérées spécifiques à la batterie (3.4) ;
- le passage au polyphonique (théorie et implémentation de tests unitaires) impliquant la dissociation d'évènements MIDI simultanés ainsi que leur identification (voir section 4.3.

L'ensemble de ces sous-tâches a permis deux réalisations principales :
1) L'obtention d'arbres de rythmes corrects en *output* de qparse avec des exemples courts proches de la partition de référence (voir section 3.4).
2) La création d'une réécriture guidée par une forme rythmique 4.4 dont le but premier est de démontrer qu'elle est implémentable et applicable à d'autres types de rythmes et dont le second objectif est de donner une méthode de création des formes rythmiques à partir d'une partition.

Ces deux réalisations recouvrent une partie du chemin à parcourir puisque pour effectuer des évaluations conséquentes sur résultats, la chaîne de traitement doit être finie afin de pouvoir vérifier de manière empirique que les formes rythmiques, qui constituent la proposition originale de ce mémoire, ont permis d'améliorer qparse pour la transcription automatique de la batterie.

1. http://lilypond.org/index.fr.html
2. MuseScore3

Nous présenterons le contexte (chapitre 1) suivi d'un état de l'art (chapitre 2) et nous définirons de manière générale le processus de transcription automatique de la musique pour enfin étayer les méthodes (chapitre 3) utilisées pour la transcription automatique de la batterie. Nous décrirons ensuite le corpus ainsi que les différentes expérimentations menées (chapitre 4). Nous conclurons par une discussion sur les résultats obtenus et les pistes d'améliorations futures à explorer. Les contributions apportées à l'outil qparse seront exposées dans les chapitres 3 et 4.

CHAPITRE 1

CONTEXTE

Sommaire

Introduction

La transcription automatique de la musique (TAM) est un défi ancien [2] et difficile qui n'est toujours pas résolu de manière satisfaisante par les systèmes actuels. Il a engendré une grande variété de sous-tâches qui ont donné naissance au domaine de la recherche d'informations musicales (RIM) [1]. Actuellement, en raison de la nature séquentielle et symbolique des données musicales et du fait que les travaux en TAL sont assez avancés en analyse de données séquentielles ainsi qu'en traitement du signal, de nombreux travaux en RIM font appel au TAL. Certains de ces travaux se concentrent notamment sur l'analyse des paroles de chansons [2].

Dans ce chapitre, nous traiterons de l'informatique musicale, nous montrerons les liens existants entre la RIM et le TAL ainsi qu'entre les notions de langage musical et de langues naturelles. Nous traiterons également du problème de la TAM et de ses applications. Enfin, nous décrirons les représentations de la musique qui sont nécessaires à la compréhension du présent travail.

1. https://ismir.net/
2. NLP4MuSA, the 2nd Workshop on Natural Language Processing for Music and Spoken Audio, co-located with ISMIR 2021.

1.1 Informatique musicale et langues naturelles

L'informatique musicale ou *Computer Music* regroupe l'ensemble des méthodes permettant de créer ou d'analyser des données musicales à l'aide d'outils informatiques [3]. Ce domaine implique l'utilisation de méthodes numériques pour l'analyse et la synthèse de la musique [4], qu'il s'agisse d'informations audio [3], ou symboliques [4]. Un exemple d'application dans ce domaine pourrait être l'analyse de la structure de la musique et de la reconnaissance des accords [5].
La RIM est née du domaine de l'informatique musicale et apparaît vers le début des années 2000 [5]. L'objectif de cette science est la recherche et l'extraction d'informations à partir de données musicales. Il s'agit d'un vaste champ de recherche pluridisciplinaire, à l'intersection des domaines de l'acoustique, du signal, de la synthèse sonore, de l'informatique, des sciences cognitives, des neurosciences, de la musicologie, de la psycho-acoustique,... Cette discipline récente a notamment été soutenue par de grandes entreprises technologiques [6] qui veulent développer des systèmes de recommandation de musique ou des moteurs de recherche dédiés au son et à la musique.

Aborder la musique comme un langage avec des méthodes de TAL nécessite une réflexion autour de la musique en tant que langage ainsi que la possibilité de comparer ce même langage avec les langues naturelles. Léonard Bernstein [6] a donné une série de six conférences publiques à Harvard fondées en grande partie sur les théories linguistiques que Noam Chomsky a exposées dans son livre *Language and Mind*. Lors de la première conférence, qui a eu lieu le 9 octobre 1973, Bernstein a avoué être hanté par la notion d'une grammaire musicale mondiale innée et il analyse dans ses trois premières conférences, la musique en termes linguistiques (phonologie, syntaxe et sémantique).
Quelques travaux en neurosciences ont également abordé ces questions, notamment par observation des processus cognitifs et neuronaux que les systèmes de traitement de ces deux productions humaines avaient en commun. Dans le travail de Poulin-Charronnat *et al.* [7], la musique est reconnue comme étant un système complexe spécifique à l'être humain dont une des similitudes avec les langues naturelles est l'émergence de régularités reconnues par le système cognitif.

3. Fichiers wav, formats spectraux,...
4. Fichier MIDI, aide à l'écriture, transcription, base de partitions...
5. En musique, un accord est un ensemble de notes considéré comme formant un tout du point de vue de l'harmonie. Le plus souvent, ces notes sont jouées simultanément ; mais les accords peuvent aussi s'exprimer par des notes successive
6. https://research.deezer.com/
https://magenta.tensorflow.org/
https://research.atspotify.com/

La question de la pertinence de l'analogie entre langues naturelles et langage musical a également été soulevée à l'occasion de projets de recherche en TAL. Keller *et al.* [8] ont exploré le potentiel de ces techniques à travers les plongements de mots et le mécanisme d'attention pour la modélisation de données musicales. La question de la sémantique d'une phrase musicale apparaît, selon eux, à la fois comme une limite et comme un défi majeur pour l'étude de cette analogie. Ces considérations nous rapprochent de la sémiologie de F. de Saussure en tant que science générale des signes et dont la langue ne serait qu'un cas particulier, caractérisé par l'arbitrariété totale de ses unités [9].
D'autres travaux très récents, ont aussi été révélés lors de la première conférence sur le NLP pour la musique et l'audio [7]. Lors de cette conférence, Jiang *et al.* [10] ont présenté leur implémentation d'un modèle de langage musical visant à améliorer le mécanisme d'attention par élément, déjà très largement utilisé dans les modèles de séquence modernes pour le texte et la musique.

Le domaine du TAL qui se rapproche le plus de la RIM est la reconnaissance de la parole (*Speech to text*). La transcription musicale étant la notation d'une œuvre musicale initialement non écrite, l'analogie avec l'écriture de la parole est aisée. De plus, ces deux domaines ont des manières similaires d'aborder la séparation des sources audio. Le tableau 1.1 montre certaines différences et similitudes de ce point de vu entre les deux disciplines.

Domaines	Similitudes	Différences
Speech to text	signal ⇒ phonèmes ⇒ texte	données linéaires [8]
AMT	signal ⇒ notes, accords ⇒ partition	données structurées [9]

TABLE 1.1 – *Speech to text* et transcription automatique musicale

Dans ce tableau, « données linéaires » signifie que les informations arrivent les unes après les autres dans un texte écrit et que la structure n'apparaît qu'à la fin de la lecture d'une phrase par exemple. « Données structurées » (ou hiérarchiques) signifie que les informations peuvent arriver simultanément et que la structure d'une phrase musicale est représentée visuellement dans une partition.
Non seulement les objectifs du *speech to text* et de la TAM sont similaires, mais les problèmes et les applications, eux aussi, sont comparables (transcription, synthèse, séparation de sources, ...). Il faut néanmoins relever que les informations sont traitées sont de nature différente puisque la mesure du temps, qui fait parti intégrante de l'information musicale, n'a aucune valeur pour des données textuelles.

7. NLP4MusA 2020

1.2 La transcription automatique de la musique

Lorsqu'un musicien est chargé de créer une partition à partir d'un enregistrement et qu'il écrit les notes qui composent le morceau en notation musicale, on dit qu'il a créé une transcription musicale de cet enregistrement. L'objectif de la TAM [11] est de convertir la performance d'un musicien en notation musicale — à l'instar de la conversion de la parole en texte dans le traitement des langues naturelles. Cette définition est interprétée, en fonction des articles scientifiques, de deux manières différentes :

1. processus de conversion d'un enregistrement audio en une notation pianoroll [10]
2. processus de conversion d'un enregistrement en notation musicale commune [11] (c'est-à-dire une partition).

La TAM a des applications multiples [11] dont la plus directe est de donner la possibilité à un musicien de générer la partition d'une improvisation en temps réel afin de pouvoir reproduire sa performance ultérieurement. Une autre application notable est la préservation du patrimoine, par exemple dans les styles musicaux où il existe peu de partitions (le jazz, la pop, les musiques de tradition orale [12] , ...). En outre, un grand nombre de fichiers audio et vidéos musicaux sont disponibles sur le Web, et pour la plupart d'entre eux, il est difficile de trouver les partitions musicales correspondantes, qui sont pourtant nécessaires pour pratiquer la musique, faire des reprises ou effectuer une analyse musicale détaillée.
La TAM est aussi utile pour la recherche et l'annotation automatique d'informations musicales, pour l'analyse musicologique [13] ou encore pour les systèmes musicaux interactifs. Cette discipline a également pour intérêt la génération de partitions dont les contenus sont exploitables, avec des formats texte ou XML (entre autres...) qui permettent de manipuler les données, contrairement à de simples images en pdf [14].

La transcription automatique de la musique est un problème ancien et difficile. C'est un « graal » de l'informatique musicale. En 1976, H. C. Longuet-Higgins [2] évoquait déjà la représentation musicale en arbre syntaxique dans le but d'écrire automatiquement des partitions à partir de données audio en se basant sur un mimétisme psychologique de l'approche humaine.

10. Une représentation bidimensionnelle des notes de musique dans le temps, comme un orgue de barbarie.
11. Ici, on parle de notation occidentale.
12. Ethno-musicologie
13. Par exemple par la constitution de corpus musicologiques
14. Voir https://archive.fosdem.org/2017/schedule/event/openscore/

La tâche de la TAM comprend actuellement deux activités distinctes :

1. l'analyse et la représentation d'un morceau de musique ;
2. La génération d'une partition à partir de cette représentation.

1.3 La transcription automatique de la batterie

La batterie est née au début du vingtième siècle [12]. Il s'agit donc d'un instrument récent qui s'est longtemps passé de partition. En effet pour un batteur, la qualité de lecteur lorsqu'elle était nécessaire, résidait essentiellement dans sa capacité à lire les partitions des autres instrumentistes (par exemple, les grilles d'accords et les mélodies des thèmes en jazz) afin d'improviser des accompagnements appropriés que personne ne pouvait écrire pour lui à sa place.
Les partitions de batterie sont arrivées par nécessité avec la pédagogie et l'émergence d'écoles de batterie [15] partout dans le monde. Un autre facteur qui a contribué à l'expansion des partitions de batterie est l'émergence de la musique assistée par ordinateur. En effet, l'usage de boîtes à rythmes [16] ou de séquenceurs [17] permettant d'expérimenter soi-même l'écriture de rythmes en les écoutant mixés avec d'autres instruments sur des machines a permis aux compositeurs de s'émanciper de la création d'un batteur en lui fournissant une partition contenant les parties exactes qu'ils voulaient entendre sur leur musique.

La batterie a un statut à part dans l'univers de la TAM puisqu'il s'agit d'instruments sans hauteur (du point de vue harmonique), d'événements sonores auxquels une durée est rarement attribuée et de notations spécifiques [13].
Les applications de la transcription automatique de la batterie (TAB) seraient utiles, non seulement dans tous les domaines musicaux concernés par la batterie dont certains manquent de partitions, notamment les musiques d'improvisation [11], mais aussi de manière plus générale dans le domaine de la RIM. En effet, si les ordinateurs étaient capables d'analyser la partie de la batterie dans la musique enregistrée, cela permettrait de faciliter de nombreuses tâches de traitement de la musique liées au rythme [13].

La TAB est un sujet de recherche crucial pour la compréhension des aspects rythmiques de la musique, et a potentiellement un fort impact sur

15. Les écoles de batterie Dante Agostini en sont un exemple très représentatif (https://www.danteagostini.com/fr/).
16. Roland TR-808
17. SQ-1

des domaines plus larges tels que l'éducation musicale et la production musicale.

1.4 Les représentations de la musique

Les partitions

Une partition de musique est un document qui porte la représentation systématique du langage musical sous forme écrite. Cette représentation est appelée transcription.

FIGURE 1.1 – Les figures de notes

La figure 1.1 montre 4 figures de notes les plus courantes dont les noms et les durées, exprimées en unité de temps musicale, appelée le « temps » sont respectivement, de gauche à droite :

- La ronde, elle vaut 4 temps ;
- La blanche, elle vaut 2 temps ;
- La noire, elle vaut 1 temps ;
- La croche, elle vaut 1/2 temps.

Une figure de note [1] de musique combine plusieurs éléments [14]

- Une tête de note :
  Sa position sur la portée indique la hauteur de la note. La tête de note peut aussi indiquer une durée.
- Une hampe :
  c'est une barre verticale liée à la tête de note. Elle est indicatrice d'une durée représentée par sa présence ou non (différence entre la ronde et la blanche)
- Un crochet : La durée d'une note est divisée par deux à chaque crochet ajouté à la hampe d'une figure de note.

La figure 1.2 montre les rapports de durée entre les figures de notes. Plus les durées sont longues, plus elles sont marquées par la tête de note ou la présence ou non de la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute un crochet à la hampe d'une figure de notes pour diviser sa durée par 2. Les notes à crochet (croches, doubles-croches, triples-croches,...) peuvent être reliées ou non par des ligatures (voir les 4 dernières lignes de la figure 1.2).

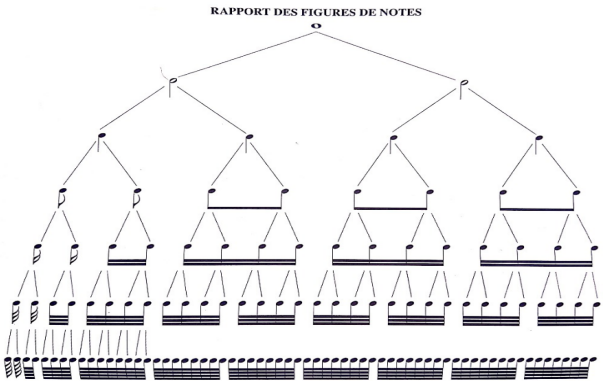


FIGURE 1.2 – Rapport des figures de notes
[1]

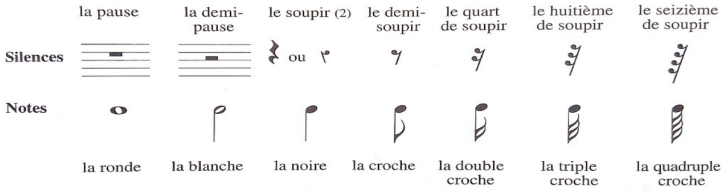


FIGURE 1.3 – Les silences
[1]

« Les silences sont des signes qui indiquent l'interruption du son. » [1]. La figure 1.3 montre pour chaque figure de note, la figure de silence qui lui correspond.

La transcription sert à traduire les trois caractéristiques du son musical :

- la hauteur, nombre de vibrations en un temps donné (son grave ou aigu — do, ré, mi,...) ;
- l'intensité, l'amplitude des vibrations (la force du son) ;
- le timbre, il permet de différencier deux instruments même s'ils jouent un son de même hauteur et de même intensité.

Ainsi que de leurs combinaisons appelées à former l'ossature de l'œuvre musicale dans son déroulement temporel, à la fois :

- diachronique (succession des instants, ce qui constitue en musique la mélodie) ;
- et synchronique (simultanéité des sons, c'est-à-dire l'harmonie).

Les partitions ont un aspect structuré (hiérarchique). Elles sont divisées en parties égales que l'on nomme « mesures ». Les mesures sont séparées par des barres verticales et sont elles-mêmes divisées implicitement en temps, qui sont eux-mêmes subdivisés. Le nombre de temps par mesure et le type de leurs subdivisions par défaut (binaire, ternaire,...) sont

déterminés par une fraction que l'on nomme « indication de mesure » ou « signature rythmique ».

— $\frac{4}{4} \rightarrow$ mesure à 4 temps binaire dont l'unité de temps est la noire ;

— $\frac{3}{4} \rightarrow$ mesure à 3 temps binaire dont l'unité de temps est la noire ;

— $\frac{6}{8} \rightarrow$ mesure à 2 temps ternaire dont l'unité de temps est la noire pointée [18].

La fraction de la signature rythmique est construite par rapport à la ronde. La fraction 4/4 signifie quatre quarts d'une ronde.

Pour les instruments mélodiques, un groupe de notes peut être organisé en voix, représentant des flots mélodiques joués en parallèle, avec une synchronisation plus ou moins stricte [15] [16].

Les données MIDI

Le format MIDI [19] est originellement une norme technique mais il peut également être considéré comme une représentation musicale. Celle-ci peut effectivement être enregistrée par des instruments compatibles, et visualisée ou jouée par un ordinateur. Ce format historique, encore très largement utilisé, est très important (mais aussi contraignant) dans le cadre de notre travail, dans la mesure où de nombreux logiciels l'utilisent. Pour la transcription musicale, il constitue une strate intermédiaire très utile entre le signal audio (enregistrement) et la représentation musicale lisible par un humain (partition).
Il s'agit d'un protocole en temps réel pour échanger des messages et un format de fichier. Un fichier MIDI est une séquence d'évènements datés et il représente une performance musicale symbolique.
Les données MIDI sont représentées sous forme de piano-roll. Chaque bande verte sur la figure 1.4 correspond à une note jouée par un instrument. Le début (*onset*) et la fin (*offset*) d'une bande verte sont appelés « évènement MIDI ». Il existe des évènements ON et des évènements OFF qui ont chacun une date sur la séquence MIDI. La durée d'une note est représentée par la distance entre les évènements ON et OFF qui lui correspondent.

Liste des élèments principaux d'un évènement MIDI :
— type de l'évènement, ON ou OFF ;

18. « Pointée » signifie que l'on rajoute à la durée d'une note la moitié de sa valeur. Une noire pointée vaut trois croches.
19. https ://www.midi.org/

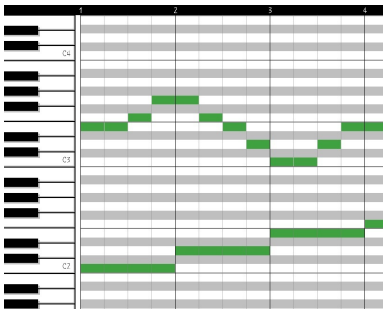
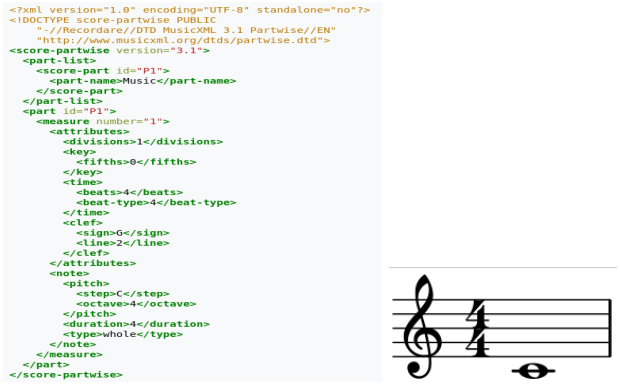


FIGURE 1.4 – Exemple d'évènements MIDI

— date de l'évènement, sa position sur la séquence ;
— pitch de l'évènement, à quelle note il correspond (à quelle instrument pour la batterie) ;
— velocité, volume sonore de l'évènement.

Les formats XML

Il existe plusieurs formats XML dédiés à la musique : MusicXML, MEI, MNX, ...
L'inconvénient de ces formats est qu'ils sont verbeux et ambigus, c'est pourquoi nous utilisons pour la transcription une représentation intermédiaire abstraite décrite plus loin.



```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE score-partwise PUBLIC
    "-//Recordare//DTD MusicXML 3.1 Partwise//EN"
    "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="3.1">
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>
  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key>
          <fifths>0</fifths>
        </key>
        <time>
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
        <clef>
          <sign>G</sign>
          <line>2</line>
        </clef>
      </attributes>
      <note>
        <pitch>
          <step>C</step>
          <octave>4</octave>
        </pitch>
        <duration>4</duration>
        <type>whole</type>
      </note>
    </measure>
  </part>
</score-partwise>
```

FIGURE 1.5 – Exemple de représentation MusicXML

Le figure 1.5 [20] représente un do en clef de sol de la durée d'une ronde sur une mesure en 4/4 écrit au format MusicXML. Un des avantages de ce format est qu'il peut être converti aussi bien en données MIDI qu'en

20. *Source images* : https://fr.wikipedia.org/wiki/MusicXML

partition musicale, ce qui en fait un bon format intermédiaire pour manipuler les données musicales et les échanger entre les programmes.

Conclusion

Dans ce chapitre, nous avons établi que la RIM a connu un fort développement ces dernières années, et qu'elle s'inspire de plus en plus de méthodes TAL. Par ce biais, nous avons déduit qu'il existait des liens possibles entre le traitement du langage musical et celui des langues naturelles, le plus proche étant probablement le phénomène de transcription (par analogie avec le *speech to text*). Nous avons contextualisé la transcription de la musique en générale et plus spécifiquement, de la batterie, en évoquant les applications possibles dans ses deux domaines. Enfin, nous avons présenté les représentations de la musique qui concernent directement ce mémoire.

CHAPITRE 2

ÉTAT DE L'ART

Sommaire

Introduction

Dans ce chapitre, nous présenterons quelques travaux antérieurs dans le domaine de la transcription automatique de la musique et de la batterie afin de situer notre démarche. Nous aborderons ensuite le passage crucial du monophonique au polyphonique dans la transcription. Puis nous décrirons les deux grandes parties de la TAM de bout en bout allant de l'audio vers l'écriture d'une partition en passant par le format MIDI. Enfin, nous ferons un point sur les approches linéaires et hiérarchiques afin de trouver un choix pertinent et équilibré à favoriser.

2.1 Du monophonique vers le polyphonique

Les premiers travaux en transcription ont été faits sur l'identification des instruments monophoniques [1] [11]. Actuellement, le problème de l'estimation automatique de la hauteur des signaux monophoniques peut être considéré comme résolu, mais dans la plupart des contextes musicaux, les instruments sont polyphoniques [2]. L'estimation de hauteurs multiples est le problème central de la création d'un système de transcription de

1. Instruments produisant une note à la fois, ou plusieurs notes de même durée en cas de monophonie par accord (flûte, clarinette, sax, hautbois, basson, trombone, trompette, cor, etc...)

2. guitare, piano, basse, violon, alto, violoncelle, contrebasse, glockenspiel, marimba, etc...

musique polyphonique. Tout signal audio musical peut être composé de plusieurs signaux, ceux-ci pouvant provenir de plusieurs instruments, ou d'un instrument dit polyphonique. Séparer, à partir du signal, les différentes sources audio (ou voix) afin de les représenter individuellement est une tâche difficile. La batterie, composée de plusieurs instruments (caisse claire, grosse caisse, cymbales, toms, etc...), est un cas typique d'intrument polyphonique pour lequel ce défi est majeur .

Les performances des systèmes actuels ne sont pas encore suffisantes pour permettre la création d'un système automatisé capable de transcrire de la musique polyphonique sans restrictions sur le degré de polyphonie ou le type d'instrument. Cette question reste donc encore ouverte.

2.2 De l'enregistrement audio vers le format MIDI

Jusqu'à aujourd'hui, les recherches se sont majoritairement concentrées sur le traitement de signaux audio vers la génération de contenus MIDI non-quantifiés (qui constituent la représentation symbolique d'une performance musicale) [17]. Cette tâche englobe plusieurs sous-tâches dont la séparation des sources audio, la détection multi-*pitchs* [3] et la détection des *onsets* et des *offsets*.

En transcription automatique de la batterie [13], plusieurs stratégies de répartition pré/post-*processing* sont possibles pour la détection multi-*pitchs*. La détection peut être entamée dès le pré-*processing*, en supprimant les *features* [4] non-pertinentes pendant la séparation des sources afin d'obtenir une meilleure détection des instruments de la batterie, par exemple en supprimant la structure harmonique pour atténuer l'influence des instruments à hauteurs sur la détection grosse caisse et caisse claire. Mais certaines études montrent que la suppression des instruments à hauteurs peut avoir des effets néfastes sur les performances de la transcription de batterie. En outre, les systèmes de TAB basés sur des réseaux de neurones récurrents ou sur des factorisations matricielles font la séparation des sources pendant l'optimisation, ce qui réduit la nécessité de la faire pendant le pré-processing. Pour la reconnaissance des instruments de la batterie, une autre approche possible est d'utiliser un modèle probabiliste pour classifier les différents sons de la batterie [18]. L'approche AdaMa [19], qui commence par une estimation initiale des sons de la bat-

3. La détection multi-*pitchs* est la détection des hauteurs simultanées pour les instruments polyphoniques. Il peut s'agir de notes d'un même instrument ou de plusieurs instruments différents.

4. Features : caractéristiques individuelles mesurables d'un phénomène dans le domaine de l'apprentissage automatique et de la reconnaissance des formes.

terie en les raffinant itérativement, est une autre approche de la même catégorie.

Ces méthodes visent toutes la génération d'un contenu MIDI non-quantifié qui est la représentation symbolique d'une performance musicale.

2.3 Du format MIDI vers une partition

Les approches mentionnées en section 2.2 produisent en sortie un fichier MIDI non-quantifié, qui est un format encore très éloigné d'une partition musicale. Un premier problème concerne les *timings* (dates et durées d'événements) qui doivent être alignés sur des positions temporelles correspondant à des valeurs exprimables avec la notation musicale (voir la différence entre contenu MIDI et musique écrite en section 1.4). On parle de quantification rhythmique.

Nakamura *et al.* 2016 présentent une approche de quantification rythmique avec un modèle probabiliste qui prend en entrée un fichier MIDI non quantifié et fourni en sortie un fichier MIDI quantifié. Shibata *et al.* 2021 [15] étendent ensuite l'approche à une transcription d'enregistrement audio vers un fichier MIDI quantifié. Ce dernier format, linéaire, ne correspond toutefois pas encore à une partition structurée, avec groupement rythmique hérarchique (voir la section 1.4). Dans ces travaux, la structuration des données en partition est déléguée à un éditeur de partitions (MuseScore), avec des résultats assez inégaux.

Seuls quelques travaux récents s'intéressent de près à la création d'outils permettant la génération de partition. Le problème de la conversion d'une séquence d'évènements musicaux symboliques en une partition musicale structurée est traité notamment dans [20]. Ce travail, qui vise à résoudre de manière conjointe la quantification rythmique et la production de partitions structurées, s'appuie tout au long du processus sur des grammaires génératives qui fournissent un modèle hiérarchique — langage *a priori* des partitions.

Les expériences parviennent à des résultats prometteurs, mais il faut relever qu'elles ont été menées avec un ensemble de données composé d'extraits monophoniques. Il reste donc à traiter le passage au polyphonique, qui nécessite de traiter le problème supplémentaire de la séparation de voix, en le couplant avec la quantification du rythme. L'approche de [20] est fondée sur la conviction que la complexité de la structure musicale dépasse les modèles linéaires.

2.4 De l'approche linéaire vers l'approche hiérarchique

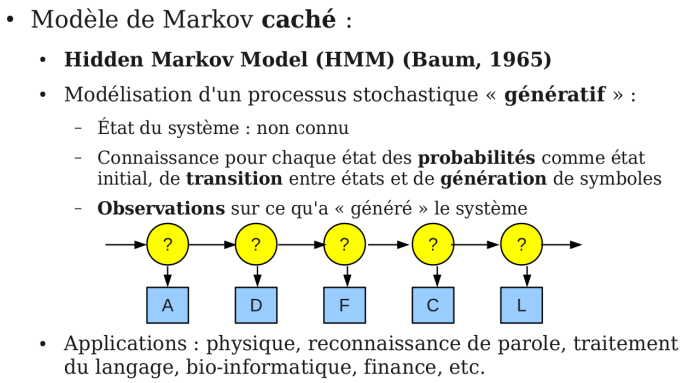


FIGURE 2.1 – Le modèle de Markov caché [5]

La figure 2.1 montre un exemple de modèle linéaire : le modèle de Markov caché (MMC). Ce type de modèle traite les évènements localement les uns après les autres. La figure mentionne entre autres que les MMC sont utilisés en reconnaissance de la parole. Il est intéressant de relever que la représentation écrite de la parole est linéaire puisque les évènements sont lus les uns après les autres avant que l'on puisse percevoir la structure d'une idée. Ce dernier point constitue une différence notable entre l'écriture de la parole et la notation musicale puisque cette dernière est explicitement structurée (voir le tableau 1.1).
Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par exemple, Shibata *et al.* [15] ont utilisé des MMC pour la reconnaissance des signatures rythmiques. Les auteurs utilisent d'abord deux réseaux de neurones profonds, l'un pour la reconnaissance des *pitchs* et l'autre pour la reconnaissance de la vélocité. Ils construisent ensuite plusieurs MMC étendus pour la musique polyphonique correspondant à des signatures rythmiques potentielles afin de trouver la signature la plus probable. L'évaluation finale des résultats de [15] montre qu'il faut rediriger l'attention vers les valeurs des notes, la séparation des voix et d'autres éléments délicats de la partition musicale qui sont significatifs pour son interprétation.
Même si la quantification du rythme se fait le plus souvent localement en manipulant des données linéaires, par déduction de la probabilité d'une durée à partir de la durée précédente (par exemple dans un MMC), de nombreux travaux suggèrent d'aborder le problème plus globalement en

5. Source : cours de Damien Nouvel https://damien.nouvels.net/fr/enseignement

utilisant une approche hiérarchique puisque le langage musical est lui-même structuré.

En effet, l'utilisation d'arbres syntaxiques semble appropriée pour représenter le langage musical. Une méthodologie simple pour la description et l'affichage des structures musicales est présentée dans [21]. Les arbres de rythmes y sont évoqués comme permettant une cohésion complète de la notation musicale traditionnelle avec des notations plus complexes. Jacquemard *et al.* [22] proposent aussi une représentation formelle du rythme, inspirée de modèles théoriques antérieurs issus du domaine des techniques de réécriture appliquées à la déduction automatique et au calcul symbolique. Ils montrent aussi qu'il est possible d'appliquer des arbres de rythmes pour le calcul d'équivalences rythmiques dans [23]. La réécriture d'arbres, dans un contexte de composition assistée par ordinateur, par exemple, pourrait permettre de suggérer à un utilisateur diverses notations possibles pour une valeur rythmique, avec des complexités différentes.

La nécessité d'une approche hiérarchique pour la production automatique de partition est évoquée dans [20]. Les modèles de grammaire qui y sont exposés sont différents des modèles markoviens linéaires de précédents travaux.

L'analyse de la structure hiérarchique des séquences d'accords par utilisation de modèles grammaticaux s'est avérée très utile dans les analyses récentes de l'harmonie du jazz [24].

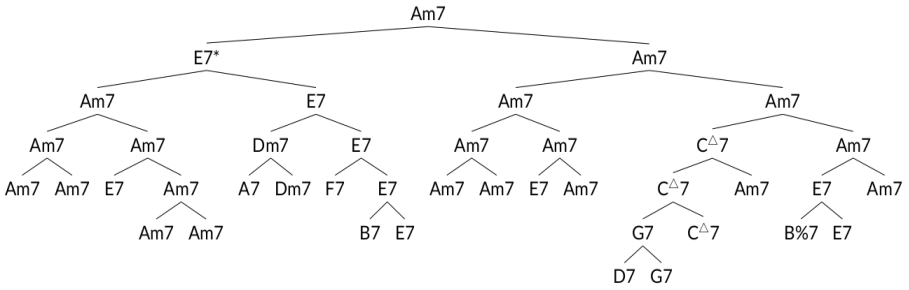


FIGURE 2.2 – Représentation arborescente d'une grille harmonique [24]

La figure 2.2 est une représentation dans un arbre syntaxique de la structure harmonique du standard de jazz *Summertime*. La racine de l'arbre est la tonalité du morceau. Entre la racine et les feuilles se trouve la structure harmonique qui défile en fonction des différentes parties du morceau et les feuilles représentent les accords joués.

Conclusion

La plupart des travaux déjà existants sur la TAB ont été énumérés par Wu *et al.* [13] qui, pour mieux comprendre la pratique des systèmes de TAB, se concentrent sur les méthodes basées sur la factorisation matricielle et celles utilisant des réseaux neuronaux récurrents. La majorité de ces recherches se concentre sur des méthodes de calcul pour la détection d'événements sonores de batterie à partir de signaux acoustiques ou sur la séparation entre les évènements sonores de batterie avec ceux des autres instruments dans un orchestre ou un groupe de musique [25], ainsi que sur l'extraction de caractéristiques de bas niveau telles que le classement des instruments et le moment de l'apparition du son. Très peu d'entre eux ont abordé la tâche de générer des partitions de batterie et, même quand le sujet est abordé, l'*output* final n'est souvent qu'un fichier MIDI non-quantifié et non une partition écrite.

En conclusion, il n'existe pas de formalisation de la notation de la batterie ni de réelle génération de partition finale, dont les enjeux principaux seraient :

1. le passage du monophonique au polyphonique, comprenant la distinction entre les sons simultanés et les appogiatures [6] ;
2. les choix d'écritures spécifiques à la batterie concernant la séparation des voix et les continuations.

6. Les appogiatures sont des ornements qui se placent devant une note principale et qui est jouée presque simultanément.

CHAPITRE 3

MÉTHODES

Sommaire

Introduction

Dans ce chapitre, nous détaillerons les méthodes que nous avons employées pour la TAB. Nous commencerons par donner une description de la notation de la batterie ainsi que des précisions sur lilypond, l'outil utilisé pour les transcriptions manuelles et les raisons qui ont motivé le choix de cet outil. Nous présenterons ensuite une modélisation de notation de la batterie pour sa représentation en arbres syntaxiques suivie d'une présentation de qparse [1], un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et plusieurs développeurs au sein du laboratoire Cedric au CNAM. Enfin, nous présenterons les formes rythmiques, une représentation théorique qui permet, par le biais de *patterns* accompagnés de leur combinaisons spécifiques associées, de détecter les signatures rythmiques de performances non-quantifiées et de restreindre les choix de règles à appliquer afin de simplifier les calculs de réécriture.

3.1 La notation de la batterie

Pour la transcription, j'ai choisi d'utiliser une notation inspirée du recueil de pièces pour batterie de J.-F. Juskowiak [26] et des méthodes de batterie

1. https://qparse.gitlabpages.inria.fr/

Dante Agostini [27], car je trouve la position des éléments cohérente et intuitive.

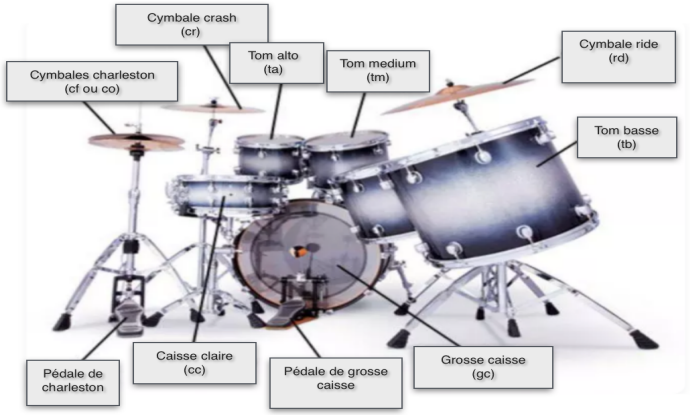


FIGURE 3.1 – Les instruments de la batterie

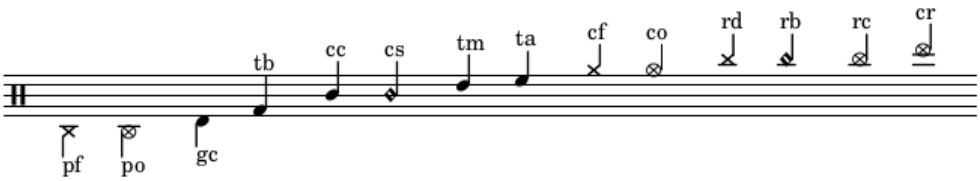


FIGURE 3.2 – Les hauteurs et têtes de notes

Noms figure 3.1	codes figure 3.2	référence
Pédale de charleston	pf ou po	charley fermé ou ouvert au pied
Grosse caisse	gc	grosse caisse
Tom basse	tb	tom basse
Caisse claire	cc	caisse claire
Tom médium	tm	tom médium
Tom alto	ta	tom alto
Cymbales charleston	cf ou co	charley fermé ou ouvert à la main
Cymbale ride	rd	ride
Cymbale crash	cr	crash

TABLE 3.1 – Les noms des instruments de la batterie

La figure 3.1 [2] montre une batterie standard avec tous les instruments habituellement présents sur une batterie et la figure 3.2 donne leur représentation sur une partition.

Le tableau 3.1 donne dans l'ordre :

1. les noms des instruments sur la figure 3.1 ;
2. leurs codes respectifs dans la figure 3.2 ;
3. les noms que j'utiliserai dans le présent document pour y référer.

Les figures 3.1, 3.2 et le tableau 3.1 peuvent aider à comprendre pourquoi je trouve la notation des méthodes Agostini cohérente et intuitive. En effet, les hauteurs sur la portée représentent :

1. La hauteur physique des instruments :
   La caisse claire est centrale sur la portée et sur la batterie (au niveau de la ceinture, elle conditionne l'écart entre les pédales et aussi la position de tous les instruments basiques d'une batterie).
   Tout ce qui est en-dessous de la caisse claire sur la portée est en dessous de la caisse claire sur la batterie (pédales, tom basse) ;
   Tout ce qui est au-dessus de la caisse claire sur la portée, l'est aussi sur la batterie.

2. La hauteur des instruments en terme de fréquences :
   Sauf pour le charley au pied et si on les sépare en trois groupes (grosse caisse, toms et cymbales), de bas en haut, les instruments vont du plus grave au plus aigu.

Les durées

Comme nous venons de le voir sur la figure 3.2, la majorité des instruments de la batterie sont représentés par les têtes des notes. De plus, le seul instrument dont le son peut être arrêté de manière quantifiée et dont la durée sonore nous intéresse est le charley [3].
Par conséquent :

1. les durées — sauf pour le charley — représenterons un écart temporel entre les notes et non une durée sonore et elles pourront donc être rallongées à l'aide de silences ;
2. les symboles rythmiques concernant les têtes de note ne pourront pas être utilisés pour exprimer les durées. Cela est valable aussi pour la présence ou non de la hampe puisque ce phénomène n'existe

2. Les noms des instruments ont été changés sur cette image qui provient du site https://www.superprof.fr/blog/composition-instrument-percussion/

3. Je ne prendrais pas en compte l'arrêt des cymbales à la main car ce phénomène n'existe pas dans les fichiers MIDI.

qu'avec les têtes de notes de type cercle-vide (opposition blanche-ronde). L'usage des blanches existe dans certaines partitions de batterie [28] mais cela reste dans des cas très rares. Certains logiciels permettent de faire des blanches avec des symboles spécifiques à la batterie ou aux percussions mais leur lecture reste peu aisée et leur utilisation pour la batterie est rarissime.

En résumé :

- toutes les notes ont une hampe ;
- une notes dont la hampe n'a pas de crochet est toujours une noire ;
- à part pour le charley ouvert, les durées n'expriment pas la durée d'un son mais une distance temporelle entre deux notes.
- à part pour le charley ouvert, la durée d'une note peut être prolongée par un silence (exemple : une noire + un soupir pour exprimer une blanche)

La durée d'une note peut être prolongée par divers symboles :

- le point : il rallonge la durée d'une note de la moitié de sa valeur ; dans l'exemple 3 de la figure 3.3, la deuxième note est une noire pointée, elle vaut donc la durée d'une noire + une croche (ou de trois croches) ;
- la liaison : elle rallonge la durée de la première note de la durée de la deuxième. La deuxième note de l'exemple 4 de la figure 3.3 est une croche qui est liée à une noire, sa durée est donc équivalente à celle d'une croche + une noire (ou de trois croches) ;
- les silences (sauf pour les ouvertures de charley).

Un autre élément concernant la notation des durées en batterie est la nécessité de faire ressortir la pulsation [4], de la rendre visuelle. La première chose à prendre en compte pour analyser la figure 3.3 est donc la nécessité de regrouper les notes par temps à l'aide des ligatures. Le deuxième point est de s'arranger pour qu'il y ait une indication visuelle au début de chaque temps.

- Exemple 1 : l'ouverture de charley est quantifiée mais les notes ne sont pas regroupées par temps.
- Exemple 2 : ici, la liaison permet de regrouper les notes par temps en obtenant le même rythme que dans l'exemple 1.
- Exemple 3 et exemple 4 : les deux exemples sont valables mais le deuxième est le plus souvent utilisé car la liaison donne un repaire visuel sur le temps.

4. La position des temps

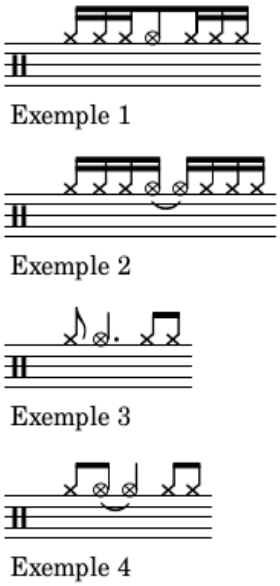


FIGURE 3.3 – Le point et la liaison

En cas de nécessité de prolonger la durée d'une note au-delà de son temps de départ (syncope) et si cette note ne correspond pas à une ouverture de charley, elle sera prolongée sur le temps suivant à l'aide de silences dont le premier sera positionné sur le temps. Si la note syncopée est une ouverture de charley, on privilégiera la liaison pour sa prolongation (exemple 4).

Les silences

Les silences sont parfois utilisés pour noter les fermetures de charley (après une ouverture). Les fermetures du charley sont notées soit par un silence (correspondant à une fermeture de la pédale), soit par un écrasement de l'ouverture par un autre coup de charley fermé, au pied ou à la main.

L'écriture littérale de contenu MIDI peut ressembler à l'exemple 1 de la figure 3.4. Sur cet exemple, le son de l'ouverture de charley est arrêté par une pression du pied sur la pédale et c'est ce que le batteur joue dans les faits. Mais il apparaît intuitivement que le but de la première note du deuxième temps n'est pas de générer un son de charley au pied mais uniquement de stopper l'ouverture. La notation de l'exemple 2 de la figure 3.4 serait donc préférable car elle représente mieux l'intention de ce rythme et elle n'empiète pas sur une potentielle voix basse qui pourrait le compléter (on évite une écriture surchargée).

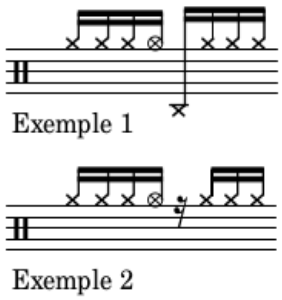


FIGURE 3.4 – Les silences en batterie

Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte la note suivante pour l'écriture :

1. si c'est un charley fermé joué à la main ⇒ la note sera un charley fermé joué à la main (cf) ;
2. si c'est un charley fermé joué au pied ⇒ la note sera un silence.

Les équivalences rythmiques

Pour les instruments mélodiques, dans le cas de notes dont la durée de l'une à l'autre est ininterrompue et si leur durée initiale est prolongée, seuls la liaison et le point permettent des notations équivalentes. Mais pour la batterie et à part dans le cas des ouvertures de charley (voir section 3.1), seules comptent des dates de début (*onsets*) : la durée du son n'a pas d'importance. L'usage des silences pour combler la distance rythmique entre deux notes devient donc possible.
Cela pris en compte, et étant donné que les indications de durée dans les têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide de silences sera privilégiée comme indication de durée sauf dans les cas où cela reste impossible. Ce choix à pour but de n'avoir qu'une manière d'écrire toutes les notes, quelles que soient leur tête de note (sauf pour le charley).

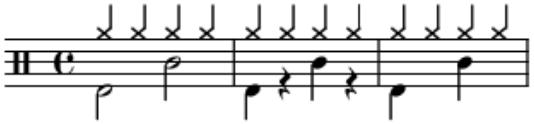


FIGURE 3.5 – Les équivalences rythmiques

Sur la figure 3.5, théoriquement, il faudra choisir la notation de la deuxième mesure mais dans certains contextes, pour des raisons de lisibilité ou de surcharge, la version sans les silences de la troisième mesure pourra être choisie.

Les voix

En batterie, une voix est théoriquement l'ensemble des instruments qui, à eux seuls, constituent une phrase rythmique. Mais en pratique, les instruments peuvent aussi être divisés par voix dans le but de ne pas surcharger la notation ou pour que leur disposition soit représentée sur la partition (voir section 3.1). Les voix sont caractérisées par l'orientation des hampes et plus précisément par les ligatures si les hampes sont dans la même direction (voir figure 3.21).

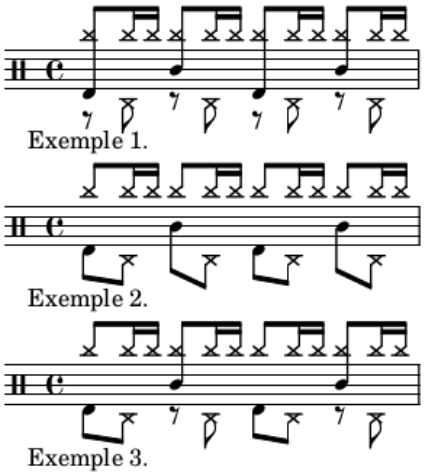


FIGURE 3.6 – La séparation des voix

Sur la figure 3.6, il faudra faire un choix entre les exemples 1, 2 et 3 qui sont trois façons équivalentes d'écrire le même rythme. Ce choix se fera en fonction des instruments joués, de la nature plus ou moins systématique de leurs phrasés, et des associations logiques entre les instruments dans la distribution des rythmes sur la batterie (voir la section 3.5).

Les accentuations et les ghost notes

« Certaines notes dans une phrase musicale doivent, ainsi que les différentes syllabes d'un mot, être accentuées avec plus ou moins de force, porter une inflexion particulière. » [1]
Théoriquement, tous les instruments peuvent être accentués (voir la section 3.3), mais la figure 3.7 représente ceux dont les accents sont presque toujours bien articulés dans le jeu standard des batteurs. En outre, les instruments qui ne sont pas représentés sur cette figure ne sont presque jamais accentués dans les partitions et ne sont pas présents de manière significative dans le jeu de données (voir section 4.1) utilisé dans ce travail.

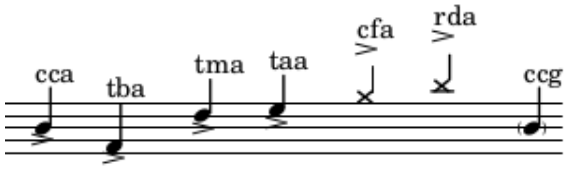


FIGURE 3.7 – Les accents et les ghost notes

Les accents sont marqués par le symbole « > ». Ils sont positionnés au-dessus des notes représentant des cymbales et en-dessous des notes représentant des toms ou la caisse claire. Ce choix a été fait pour la partition de la figure 3.11 car elle est plus lisible ainsi, mais ces choix devront être adaptés en fonction des différentes formes rythmiques reconnues (voir la section 3.5). Par exemple, pour les formes rythmiques jazz, les ligatures pour les toms et la caisse claire seront dirigées vers le bas, il faudra donc mettre les symboles d'accentuation correspondants au-dessus des têtes de notes.

La dernière note de la figure 3.7 montre un exemple de notation pour une ghost note jouée à la caisse claire. Une ghost note [29] est une note de faible volume sonore mais jouée fermement. Les ghost notes servent le plus souvent à donner le débit d'un rythme (ses subdivisions) pour le rendre plus dansant (lui donner plus de « groove » ou de « swing »). Le parenthésage a été choisi car il peut être utilisé sur n'importe quelle note sans changer la tête de note.

Toutes les notes de la figure 3.7 sont exposées en situation réelle dans la figure 3.8.

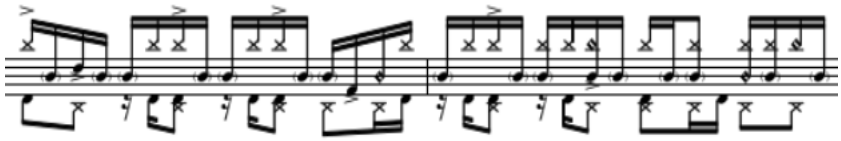


FIGURE 3.8 – Exemple pour les accentuations et les ghost notes

Les flas

Le fla est une appogiature qui consiste à jouer deux coups presque simultanés dont le premier est une ghost note et le deuxième une note normale ou accentuée.

FIGURE 3.9 – La notation du fla

3.2 La transcription manuelle avec lilypond

Mis à part les figures du chapitre 1 et certains exemples d'analyses de la section 4.2, toutes les partitions et figures de ce document ont été généré avec lilypond [5].

Présentation de lilypond

« LilyPond est un logiciel de gravure musicale, destiné à produire des partitions de qualité optimale. Ce projet apporte à l'édition musicale informatisée l'esthétique typographique de la gravure traditionnelle. LilyPond est un logiciel libre rattaché au projet GNU. » [6]

En raison de la grande liberté de choix que permet lilypond et du fait qu'une configuration pour la notation de type Agostini est disponible, je considère que lilypond est actuellement le meilleur choix pour transcrire de la batterie.

```
\version "2.20.0"
\language français

#(define mydrums '(
    (splashcymbal xcircle #f 9)
    (ridecymbal cross #f 7)
    (ridebell harmonic #f 7)
    (crashcymbal xcircle #f 7)
    (hihat cross #f 5)
    (openhihat xcircle #f 5)
    (hightom () #f 3)
    (lowmidtom () #f 2)
    (snare () #f 0)
    (sidestick harmonic #f 0)
    (lowfloortom () #f -3)
    (bassdrum () #f -5)
    (pedalhihat cross #f -7)
    (halfopenhihat xcircle #f -7)))
```

```
\include "../../0_drum_style_perso.ly"

up = \drummode {

  \override Script.Y-offset = #'-1.0

  % Mesure 1
  s s s s

  % Mesure 2
  <cyms cymc>8 cymr
  <sn cymr>16 \parenthesize sn cymr16 <\parenthesize sn cymr>
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <cymr ss> \parenthesize sn cymr16 <\parenthesize sn cymr>

  % Mesure 3
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <cymr ss> \parenthesize sn < \parenthesize sn cymr>16 cymr
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <cymr ss> \parenthesize sn cymr16 <\parenthesize sn cymr>

  % Mesure 4
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <sn cymr>8 < \parenthesize sn cymr>16 cymr
  cymr \parenthesize sn cymr16 cymr
  <sn cymr>16 \parenthesize sn cymr16 <\parenthesize sn cymr>

  % Mesure 5
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <sn cymr> \parenthesize sn < \parenthesize sn cymr>16 cymr
  cymr \parenthesize sn cymr16 cymr
  <ss cymr>8 <sn cymr>16_> cymr

  % Mesure 6
  cymr \parenthesize sn <cymr \parenthesize sn>16 cymr
   <ss cymr>8 cymr16 cymr
  cymr \parenthesize sn < \parenthesize sn cymr>16 cymr
  <sn-> cymr>16_> \parenthesize sn cymr16 <\parenthesize sn cymr>
```

FIGURE 3.10 – Extraits de code lilypond

5. http://lilypond.org/index.fr.html
6. Page d'accueil du site http://lilypond.org/index.fr.html

Sur la figure 3.10 :

— à gauche, une configuration aménagée pour la notation de type Agostini.

— à droite, le début de code mesure par mesure pour la voix haute d'une partition (la première ligne sert à prendre en compte le fichier de gauche).

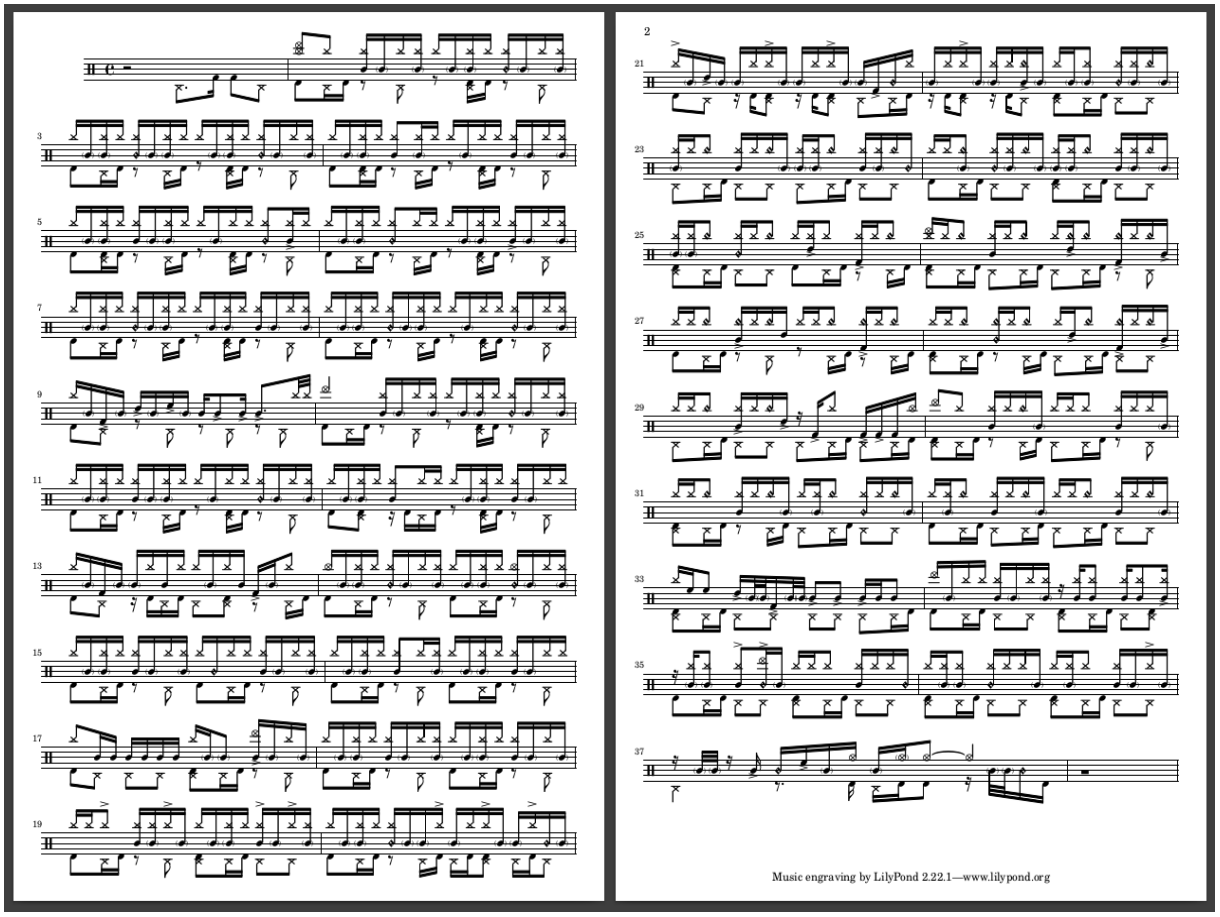


FIGURE 3.11 – Transcription de partition avec lilypond

La partition de la figure 3.11 est le résultat du code de la figure 3.10 (la totalité du code est sur https://github.com/MartinDigard/Stage_M2_Inria). Cette partition a été totalement transcrite manuellement avec lilypond par analyse des fichiers MIDI et audio correspondants.

— Difficultés principales : trouver une application permettant de choisir librement la notation de la batterie. Lylipond le permet mais beaucoup de recherches ont été nécessaires pour comprendre l'ensemble des fonctionnalités permettant de faire fonctionner la notation « Agostinienne » ainsi que les diverses subtilités de nota-

tions (accents, ghost notes, flas, ...).
lylipond reste néanmoins un choix très agréable, une fois ces difficultés surmontées.
— Écrire la partition de la figure 3.11 m'a pris beaucoup de temps car j'ai dû chercher comment écrire chaque nouvel évènement, mais les autres transcriptions ont été beaucoup plus rapide et très aisées.
— Même si cela représente un investissement au départ, je recommande lylipond pour écrire la batterie et je pense que c'est meilleur outil pour cette tâche pour le moment.
— Dans les autres logiciel d'édition de type musescore, la batterie est toujours confiné au système de notation américain.
— Pour une comparaison entre système de notation américain et le système de notation Agostini, voir section 4.2 est comparer les notations TM (Agostini) et TA (américain).

3.3 Modélisation pour la transcription

Les pitchs

Codes	Instruments	Pitchs
cf	charley-main-fermé	22, 42
co	charley-main-ouvert	26
pf	charley-pied-fermé	44
rd	ride	51
rb	ride-cloche (bell)	53
rc	ride-crash	59
cr	crash	55
cc	caisse claire	38, 40
cs	cross-stick	37
ta	tom-alto	48, 50
tm	tom-medium	45, 47
tb	tom-basse	43, 58
gc	grosse caisse	36

TABLE 3.2 – Les codes, l'identités et les pitchs des instruments

Le tableau 3.2 présente dans l'ordre, les codes des instruments, leur identité (instrument ou partie d'un instrument, joué avec les mains ou avec les pieds), le ou les pitchs qui lui sont associés.
Plusieurs pitchs peuvent parfois désigner le même instrument afin de pouvoir supporter des kits [7] de batterie plus larges (avec par exemple plu-

7. Les batteries électroniques permettent de choisir les sons que l'on veut donner à chaque instruments parmi des kits prédéfinis adaptés à différents style.

sieurs toms basses qui n'auraient pas tous exactement la même sonorité) ou simplement de styles différents (pour chaque kit standard, ce sont les mêmes intruments mais de styles différents)[8]. J'ai regroupé les pitchs des différents types d'un même instrument dans une seule ligne du tableau portant le nom du type de cet instrument. Ainsi, plusieurs toms basses différents dans les données MIDI deviennent tous un tom basse d'une batterie standard et la partition finale pourra être jouée sur n'importe quel kit de batterie standard.
Malgré le large panel de pitchs disponibles, il semblerait qu'aucun pitch ne désigne le charley ouvert joué au pied (« po » de la figure 3.2). Pourtant, dans la batterie moderne, plusieurs rythmes ne peuvent fournir le son du charley ouvert qu'avec le pied car les mains jouent autre chose en même temps. Cela doit en partie être dû à l'utilisation des boîtes à rythmes en musique assistée par ordinateur qui ne nécessitent pas de faire des choix conditionnés par les limitations humaines (2 pieds, 2 mains, et beaucoup plus d'instruments...)

La vélocité

La vélocité déterminera si les notes sont accentuées ou sont des ghost notes. Pour les codes, je propose d'ajouter un suffix (« a » pour accent et « g » pour ghost note) à la fin du code d'une note accentuée ou d'une ghost note. Les choix pour déterminer si les notes sont accentuées ou sont des ghost notes seront donnés dans la section 4.2.

Les arbres de rythmes

Les arbres de rythmes représentent un rythme dont les possibilités de notation sur une partition sont théoriquement multiples. Les branchements sont des divisions d'intervalles temporels et les feuilles sont des évènements musicaux commençant au début de l'intervalle [30] [31] .

La figure 3.12 est une représentation qui fonctionne avec les 3 exemples de la figure 3.6 en arbre de rythmes avec les codes de chaque instrument :

La figure 3.13 montre le même arbre dont les codes des instruments sont remplacés par leurs données MIDI respectives :

Chacun des trois exemples de la figure 3.6 sont représentés par chacun des deux arbres syntaxiques ci-dessus. On voit bien ici l'avantage de cette représentation pure des rythmes car elle permet de tester plusieurs notations équivalentes pour un même rythme.

8. Par exemple, les peaux des toms jazz raisonnent alors que celles des toms rock sont mates.

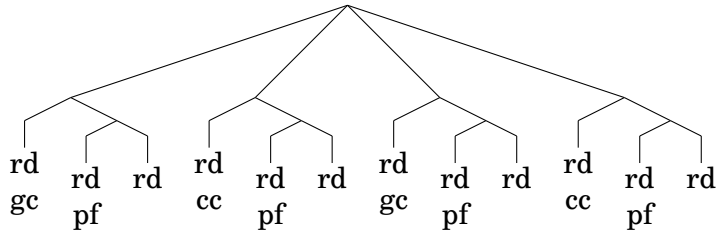


FIGURE 3.12 – Exemple d'un arbre de rythmes avec les codes des instruments

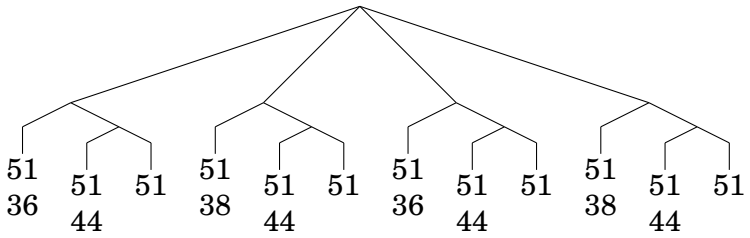


FIGURE 3.13 – Exemple d'un arbre de rythmes avec les pitchs des instruments

3.4 Analyse syntaxique pour la transcription

Comme le montre la figure 3.14, qparse [9] est un outil pour la transcription musicale qui produit une partition structurée à partir d'une performance symbolique séquentielle et non quantifiée, . Il effectue conjointement des tâches de quantification rhythmique et d'inférence de la structure de la partition à l'aide de techniques d'analyse syntaxique (*parsing*). Le but du *parsing* est en effet la structuration d'une représentation séquentielle en entrée (un mot fini), suivant un modèle de langage [32].
Dans le cas de qparse, le "mot" d'entrée est typiquement au format MIDI, et le modèle de langage est une grammaire d'arbres pondérés représentant des préférences en terme de notation musicale à produire [33]. basée sur des algorithmes d'analyse syntaxique pour les grammaires arborescentes pondérées. En prenant en entrée une performance musicale symbolique (séquence de notes avec dates et durées en temps réel, typiquement un fichier MIDI), et une grammaire hors-contexte pondérée décrivant un langage de rythmes préférés, il produit une partition musicale. Plusieurs formats de sortie sont possibles, dont les formats XML (MEI, MusicXML,. . .) ou lilypond.

Les principaux contributeurs sont :
— Florent Jacquemard (Inria) : développeur principal.

9. https://qparse.gitlabpages.inria.fr

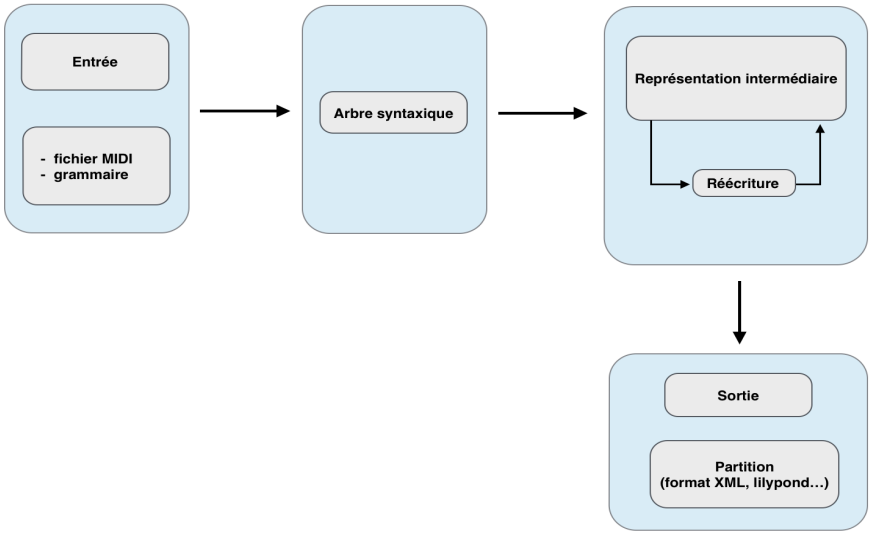


FIGURE 3.14 – Présentation de qparse

— Francesco Foscarin (PhD, CNAM) : apprentissage ; Evaluation.
— Clement Poncelet (Salzburg U.) : integration de la librairie Midifile pour les input MIDI.
— Philippe Rigaux (CNAM) : production de partition au format MEI et de modèle intermédiaire de partition en sortie.
— Masahiko Sakai (Nagoya U.) : mesure de la distance input/output pour la quantification et CMake framework ; évaluation.

Les enjeux

Un des problèmes de qparse était qu'il soit limité au monophonique. Il était donc impossible de traiter plusieurs voix et de reconnaître les accords. Ce qui est problématique pour la batterie étant donné que c'est un instrument dont l'essence est la coordination de plusieurs sons à la fois. Ce problème a été en partie résolu en regroupant les notes de faibles distances mutuelles dans des clusters appelés « Jam » dont nous parlerons plus loin. Un autre problème du MIDI avec qparse est le fait d'avoir deux symboles en entrée pour un seul généré en sortie. Ce problème est moins génant pour la batterie car nous avons pu ignorer tous *offsets*. La quantification nécessite d'ajuster les dates des notes une fois leur emplacement déduit tout en minimisant la distance entre le midi et la représentation en arbre. La grammaire pondérée sert à gérer cet équilibre.

La grammaire

Il s'agit d'une grammaire hors-contexte pondérée dont la structure agit comme un langage *a priori* de notation rythmique. Les règles de la grammaire et certaines valeurs de poids indiquent des rythmes et des écritures (pour les ligatures notamment) à favoriser.

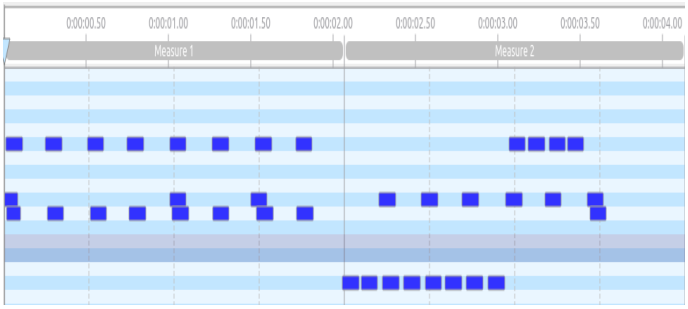


FIGURE 3.15 – Fichier MIDI pour les tests de grammaires

Les exemples suivants ont été écrits pour le fichier MIDI de la figure 3.15. Ce fichier contient les informations de trois mesures dont les deux premières contiennent des noires, des croches (et des doubles-croches pour la deuxième) et dont la troisième est vide. La signature rythmique est 4/4.

```
0 -> C0                 1
0 -> E1                 1
0 -> U4(1, 1, 1, 1)     1
1 -> C0                 1
1 -> E1                 1
1 -> T2(2, 2)           1
1 -> T4(4, 4, 4, 4)     1
2 -> C0                 1
2 -> E1                 1
4 -> C0                 1
4 -> E1                 1
4 -> E2                 1
```

FIGURE 3.16 – Exemple de grammaire

Sur la figure 3.16, chaque ligne est une règle. La colonne de gauche contient des symboles non-terminaux (0, 1, 2, 4). Cette grammaire sépare les temps par ligatures au niveau de la mesure. Puis elle autorise, au niveau du temps, les divisions par deux (croches) et par quatre (doubles-croches). Ces divisions seront reliées par des ligatures. Tous les poids sont

réglés sur 1 à titre expérimental. L'arbre de parsing en résultant (figure 3.17) est considéré comme « convaincant » car il découpe correctement les mesures et les temps.

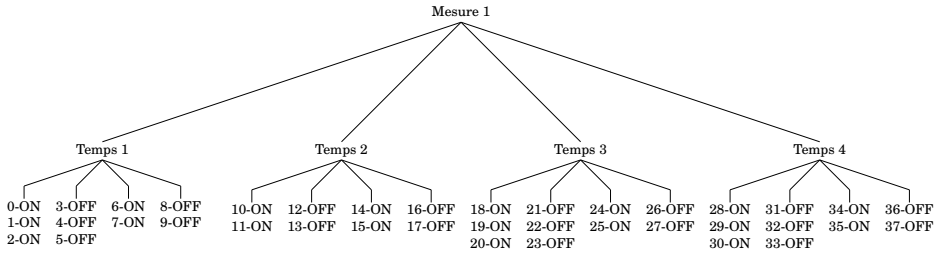


FIGURE 3.17 – Arbre en sortie de qparse

Le parsing

Les données MIDI sont quantifiées, les notes de dates proches sont alignées et les relations entre les notes sont identifiées (accords, fla, etc...) ; un arbre syntaxique global est créé (figure 3.17).

La représentation intermédiaire

- Les pitchs sont remplacés par les codes des instruments (voir tableau 3.1 ;
- réécriture 1 :
  séparation des voix ⇒ un arbre par voix ⇒ représentation intermédiaire (RI) ;
- réécriture 2 :
  simplification de l'écriture de chaque voix de la RI.

Cette procédure sera détaillée dans la section 4.4.

3.5 Les formes rythmiques

Il existe en batterie des motifs rythmiques répétés (joués en boucle). Ces motifs sont le résultat de la coordination de plusieurs instruments de la batterie, je les nommerai motifs dans la suite du document. Très souvent, un autre instrument est joué de manière indépendante sur le motif mais en respectant la cohérence rythmique du motif. Je nommerai gamme l'ensemble des combinaisons possibles pour cet autre instrument. La gamme d'un motif est toujours relative à sa signature rythmique. Enfin, j'ai nommé forme rythmique (FR) le couple motif-gamme.

Objectifs

Le but est d'avoir des schemas types (les formes rythmiques) pour calculer la séparation en voix. Cela constituerait une heuristique pour éviter d'avoir à explorer une grande combinatoire.
Un ensemble de formes rythmiques comprenant leur signatures rythmiques respectives et leurs règles spécifiques de réécriture sera nécessaire.
Une fois un système défini, la transcription à partir de l'entrée MIDI sera facilitée par la reconnaissance de la FR qui contraindra le processus, il premettra de :

— définir une signature rythmique ;
— choisir une grammaire appropriée ;
— fournir les règles de réécriture (séparation des voix et simplification.

Définitions

— motif : rythmes coordonnés joués avec deux ou trois instruments coordonnés en boucle (répartis sur 1 ou 2 voix) ;
— gamme : ensemble des combinaisons jouées par un autre instrument sur le motif (réparti sur 1 voix) ;
— forme rythmique : motif + gamme.

Les motifs sont fixes, contrairement aux gammes, qui regroupent l'ensemble des possibilités pouvant être rencontrées en situation réelle (sur une partition ou lors d'une performance de batterie).
Un motif détermine la signature rythmique d'une forme rythmique , et donc de sa gamme. Il détermine aussi avec quel instrument unique sera jouée la gamme et comment tous les instruments de la forme rythmique se répartiront en voix.

Les formes rythmiques devront être distribuées dans 4 grandes catégories [28] :

FR	SR [10]	Subdivisions	Possibles	nb voix
binaires	simple	doubles-croches	triolets, sextolets	2
jazz	simple	triolets	croches et doubles	2
ternaires	complexe	croches	duolets, quartelets	2
afros-cubains	simple	croches	-	3

TABLE 3.3 – Les formes rythmiques

Nous exposerons 3 formes rythmiques afin d'illustrer les propos de cette

10. SR du tableau signifie « signature rythmique »

section :
- 4/4 binaire
- 4/4 jazz
- 4/4 afro-cubain

Détection de la signature rythmique

La partie motif des FR sera utilisée pour la **définition des signature rythmiques**. La détection de la signature rythmique est importante, non seulement pour connaître le nombre de temps par mesure ainsi que le nombre de subdivisions pour chacun de ces temps, mais aussi pour savoir comment écrire l'unité de temps et ses subdivisions.

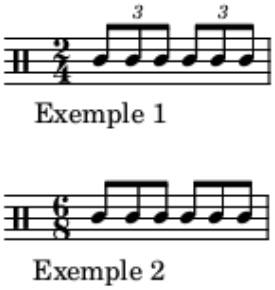


FIGURE 3.18 – Les signatures rythmiques

La figure 3.18 montre deux signatures rythmiques différentes. L'une (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des triolets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est traditionnellement écrit en binaire avec ou sans triolet (même si cette musique est dite ternaire alors que le rock ternaire sera plutôt écrit comme dans l'exemple 2).

Choix d'une grammaire

Il faut prendre en compte l'existence potentielle de plusieurs grammaires qui regroupent les contenus MIDI par signature rythmique. Le choix d'une grammaire pondérée doit être fait avant le parsing puisque qparse prend en entrée un fichier MIDI et un fichier wta (grammaire). C'est pour cette raison que la signature rythmique doit être définie avant le choix de la grammaire. Il faudrait les trouver automatiquement sans autres indications que les contenus MIDI. Par conséquent, les motifs des formes rythmiques devront être recherchés sur l'input (fichiers MIDI) avant le lancement du parsing, afin de déterminer la signature rythmique en amont. Cette tâche devra probablement être effectuée par utilisation d'apprentissage automatique.

Séparation des voix

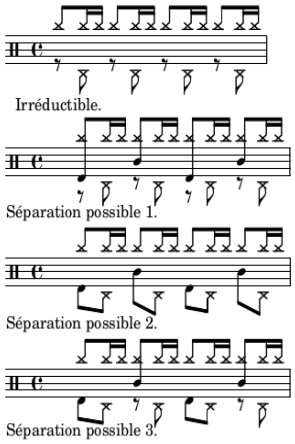


FIGURE 3.19 – Forme rythmique 4/4 binaire

Ici, la forme rythmique est construite sur un modèle rock avec une signature rythmique en 4/4.
La première ligne de la figure 3.19 est appelée « Irréductible » car il n'y a pas d'autre choix de séparation des voix pour la ride et le charley au pied. La troisième séparation proposée est privilégiée car elle répartit selon deux voix, une voix pour les mains (ride et caisse claire) et une voix pour les pieds (charley et grosse caisse). Ce choix paraît plus équilibré car deux instruments sont utilisés par voix (contrairement aux séparations possibles 1 et 2 de la figure 3.19) et plus logique pour le lecteur puisque les mains sont en haut et les pieds en bas.

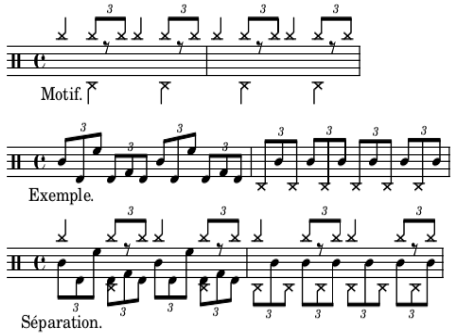


FIGURE 3.20 – Forme rythmique 4/4 jazz

Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré comme évident en jazz traditionnel. Ici, le parti pris est de tout écrire.
Dans la figure 3.20, les mesures 1 et 2 de la ligne « Exemple » combinées avec le motif de la première ligne, sont des cas typiques de la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus, la grosse caisse entre très souvent dans le flot des combinaisons de toms et de caisse claire et son écriture séparée serait inutilement compliquée et peu intuitive pour le lecteur. Le choix de séparation sera donc de laisser les cymbales jouées à la main en haut, et les toms, la caisse claire, la grosse caisse et la pédale de charley en bas.

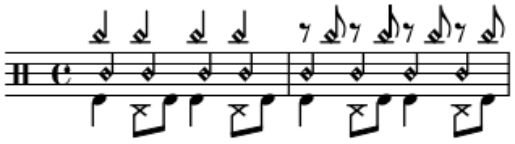


FIGURE 3.21 – Forme rythmique 4/4 afro-latin

La figure 3.21 montre un exemple minimaliste de forme rythmique afro-latin [28].
Cette forme rythmique doit être écrite sur trois voix car la voix centrale est souvent plus complexe que sur la figure et la mélanger avec le haut ou le bas serait surchargé et peu lisible.

Simplification de l'écriture

Les règles de simplification (les combinaisons de réécritures) seront extraites des voix séparées des formes rythmiques. Les explications qui suivent seront appuyées par une réécriture guidée dans la section 4.4.
Les gammes qui accompagnent les motifs étayent toutes les combinaisons d'un FR et elles permettent, combinées avec le motif, de définir ses propres règles de simplification.

Voici les différentes étapes à suivre :

- pour chaque gamme d'une forme rythmique, faire un arbre de rythmes représentant la gamme combinée avec le motif de la FR ;
- pour chaque arbre de rythmes obtenus, séparer les voix et faire un arbre de rythmes par voix ;
- pour chaque voix (arbres de rythmes) obtenus, extraire tous les nœuds qui nécessitent une simplification et écrire la règle.

Certaines précisions concernant l'extraction de ces règles sont nécessaires. Il s'agit de précisions à propos de la durée, des silences et de la présence ou non d'ouvertures de charley dans les instruments joués.

Nous avons discuté de ces problèmes plus haut dans ce chapitre.

Voici quelques règles inhérentes à la simplication de l'écriture pour la batterie :
Même si on favorise l'usage des silences pour l'écart entre les notes n'appartenant pas au même temps, on les remplace systèmatiquement par un point pour 2 notes au sein d'un même temps.

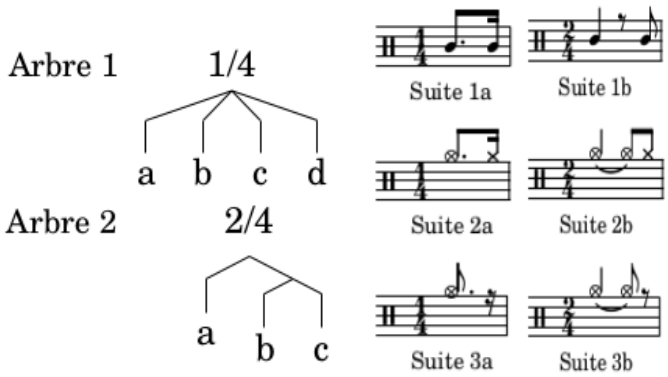


FIGURE 3.22 – Simplifications — arbres et notations possibles

Dans la figure 3.22, les « suites » sont des notations possibles relatives aux arbres 1 ou 2.

Rappel :
cf = charley fermé joué à la main ;
co = charley ouvert joué à la main ;
pf = charley fermé joué au pied.

Soit l'arbre 1 de la figure 3.22 dans lequel :
- a et d sont des instruments de la batterie (x) ;
- b et c sont des continuations (t).

Pour chacune des conditions suivantes, une suite de la figure 3.22 est attribuée :
- Si a n'est pas un co :
  ⇒ Suite 1a.
- Si a est un co :
  - Si d est un cf :
    ⇒ Suite 2a.
  - Si d est un pf :
    ⇒ Suite 3a : d deviens un silence (r).

Soit l'arbre 2 de la figure 3.22 dans lequel :
a et c sont des instruments de la batterie (x) ;

b est une continuation (t) ; Pour chacune des conditions suivantes, une suite de la figure 3.22 est attribuée :

- Si a n'est pas un co :
  ⇒ Suite 1b, b devient un silence.
- Si a est un co :
  - Si c est un cf :
    ⇒ Suite 2b, b devient une liaison et c devient un cf.
  - Si c est un pf :
    ⇒ Suite 3b : b deviens une liaison et c devient un silence.

Conclusion

Dans ce chapitre, nous avons formalisé une notation de la batterie inspirée des méthodes de batterie Dante Agostini, modélisé cette notation pour la transcription de données MIDI en partition. Nous avons ensuite parlé de l'outil utilisé pour les transcriptions manuelles en mettant en avant que cet outil devrait être utilisé pour la transcription de la batterie et en précisant que tous les codes lilypond pour la création des figures et partition de ce mémoire sont en accès libre sur github. Nous avons ensuite décrit qparse qui est l'outil que le travail de ce mémoire cherche à améliorer.
Enfin, nous avons exposé une approche de type dictionnaire (les « formes rythmiques ») pour détecter une signature rythmique, choisir une grammaire pondérée appropriée et énoncer des règles de séparation des voix et de simplification de l'écriture.

CHAPITRE 4

EXPÉRIMENTATIONS

Sommaire

Introduction

Dans ce chapitre, nous présenterons le jeu de données. Des analyses MIDI-Audio seront effectuées sur ce jeu de données par le biais de comparaisons de transcription et transcriptions manuelles avec lilypond. Nous aborderons aussi, le passage du monophonique au polyphonique, indispensable pour l'application des formes rythmiques dans la chaîne de traitement. Nous présenterons une réécriture guidée par une forme rythmique qui devra être utilisé comme base de connaissances pour augmenter la rapidité et la qualité en sortie de qparse et comme une méthode de création de nouvelles formes rythmiques. Enfin, nous discuterons sur l'ensemble des travaux finis, notamment les avancées réalisées dans ce travail.

4.1 Le jeu de données

Nous avons utilisé le Groove MIDI Dataset [1] [34] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).
Le GMD est composé de 13,6 heures de batterie sous forme de fichiers MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000

1. https://magenta.tensorflow.org/datasets/groove

mesures de batterie dans les styles les plus courants et avec différentes qualités de jeu. Tout le contenu a été joué par des humains sur la batterie électronique Roland TD-11 (figure 4.1). Autres critères spécifiques au

FIGURE 4.1 – Batterie électronique Roland TD-11 [2]

GMD :

- Toutes les performances ont été jouées au métronome et à un tempo choisi par le batteur.
- 80% de la durée du GMD a été joué par des batteurs professionnels qui ont pu improviser dans un large éventail de styles. Les données sont donc diversifiées en termes de styles et de qualités de jeu (professionnel ou amateur).
- Les batteurs avaient pour instruction de jouer des séquences de plusieurs minutes ainsi que des fills [3]
- Chaque performance est annotée d'un style (fourni par le batteur), d'une signature rythmique et d'un tempo ainsi que d'une identification anonyme du batteur.
- Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-session du GMD.
- Les sorties audio synthétisées ont été alignées à 2 ms près sur leur fichier MIDI.

Format des données

Le Roland TD-11 enregistre les données dans des fichiers MIDI et les divise en plusieurs pistes distinctes :

- une pour le tempo et l'indication de mesure ;
- une pour les changements de contrôle (position de la pédale de charley) ;
- une pour les notes.

2. *Source :* https://www.youtube.com/watch?v=BX1V_IE0g2c

3. Un *fill* est une séquence de relance dont la durée dépasse rarement 2 mesures. Il est souvent joué à la fin d'un cycle pour annoncer le suivant.

Les changements de contrôle sont placés sur le canal 0 et les notes sur le canal 9 (qui est le canal canonique pour la batterie).
Pour simplifier le traitement de ces données, ces trois pistes ont été fusionnées en une seule piste qui a été mise sur le canal 9.

4.2 Analyses et transcriptions manuelles

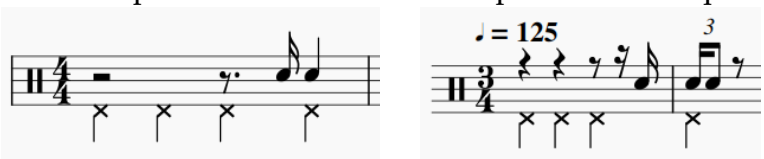
Ces analyses ont été faites dans le cadre de transcriptions manuelles à partir de fichiers MIDI et Audio du GMD.

Comparaisons de transcriptions

Pour les comparaisons de transcriptions, les transcriptions manuelles (TM) ont été éditées à l'aide de Lilypond [4] ou MuseScore [5] et les transcriptions automatiques (TA) ont toutes été générées par import d'un fichier MIDI dans MuseScore.

Exemple d'analyse 1

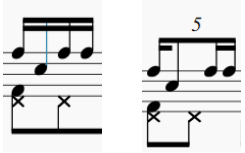
Transcription manuelle ⇒ Transcription automatique



— Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
— Les silences de la mesure 1 de la TA sont inutilement surchargés ;
— La noire du temps 4 de la mesure 1 de la TM est devenue les deux premières notes (une double-croche et une croche) d'un triolet sur le temps 1 de la mesure 2 de la TA.

Exemple d'analyse 2

Transcription manuelle ⇒ Transcription automatique

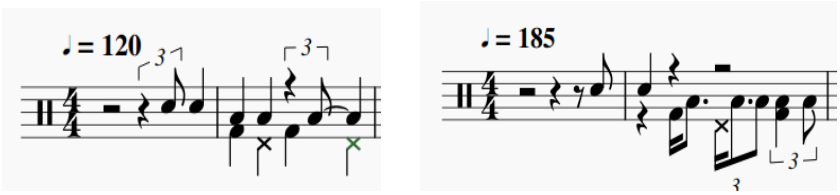


— Les doubles croches ont été interprétées en quintolet
— La deuxième double-croche est devenue une croche.

4. http://lilypond.org/
5. https://musescore.com/

Exemple d'analyse 3

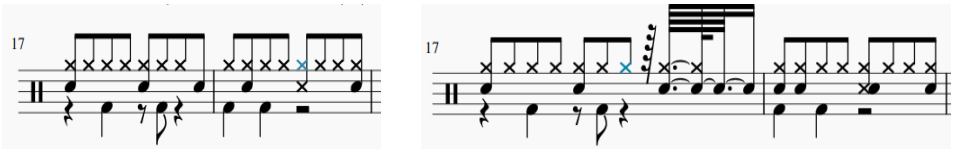
Transcription manuelle ⇒ Transcription automatique



— Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
— Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été décalés d'une double croche vers la droite dans la TA.
— La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
— Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.

Exemple d'analyse 4

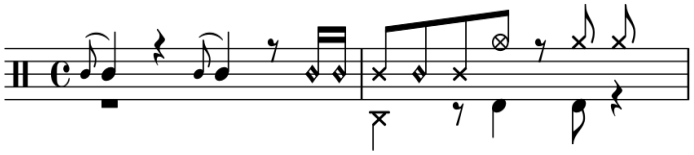
Transcription manuelle ⇒ Transcription automatique



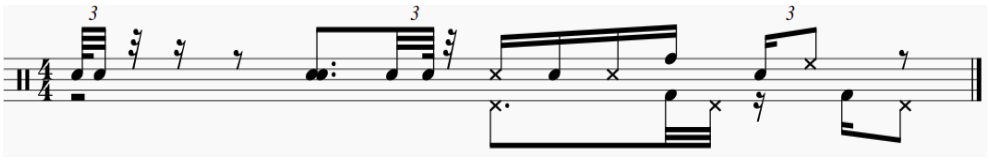
Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une manière excessivement complexe !

Exemple d'analyse 5 (flas)

Transcription manuelle



Transcription automatique



— Le premier fla est reconnu comme étant un triolet contenant une quadruple croche suivie d'une triple croche au lieu d'une seule note ornementée.
— Le deuxième fla est reconnu comme étant un accord.
— Les deux double en contre-temps sur le temps 4 de la TM sont mal quantifiée dans la TA.
— La TA ne reconnaît qu'une mesure quand la TM en transcrit deux. En effet, la TA a divisé par deux la durée des notes afin de les faire tenir dans une mesure à 4 temps dont les unités de temps sont les noires. Par exemple, le soupir du temps 2 de la TM devient un demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore, la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de la TM suivie d'un demi-soupir devient une croche pointée sur le temps 3 de la TA.
— Autre problème : certaines têtes de notes sont mal attribuées. Par exemple, le charley ouvert en contre-temps sur le temps 2 de la mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem pour les cross-sticks.

Conclusion d'analyse

Ces analyses ont montré la difficulté pour un logiciel comme MuseScore d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance de la forme du rythme impliquant sa position dans la mesure. Cette reconnaissance pourrait permettre de rectifier les problèmes de signature rythmique ainsi que les problèmes de décalage de temps. La reconnaissance de la forme du rythme permettrait aussi de supprimer les aberrations du type de celle de l'exemple d'analyse 4, puisque l'erreur sur cet exemple serait reconnue comme un élément qui ne rentre pas dans le cadre de la forme de rythme en question. La dernière raison qui rend le travail difficile est l'identification des flas, comment savoir si deux notes jouées très proches sont :

— séparées et rapides,
— mal jouées à l'unisson (accord),
— ou forment un fla ?

Transcription de partition

La figure 3.11 est la transcription manuelle des fichiers *004_jazz-funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.
Cette transcription a été entièrement faite avec Lilypond (voir le code lilypond sur le git https://github.com/MartinDigard/Stage_M2_Inria). Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est annoncé dans le GMD de style «jazz-funk» probablement en raison de la ride de type shabada rapide (le ternaire devient binaire avec la vitesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux et quatre).
La transcription manuelle de la partition de la figure 3.11 et l'analyse d'autre fichiers MIDI (voir section 4.2) m'ont mené aux observations suivantes :

- Vélocité inférieure à 40 : ghost-note ;
- Vélocité supérieure à 90 : accent ;
- Pas d'intention d'accent ni de ghost-note pour une vélocité entre 40 et 89 ;
- Les accents et les ghosts-notes ne sont significatifs ni pour les instruments joués au pied, ni pour les cymbales crash.
  En effet, certaines vélocités en dessous de 40 étant détectées et inscrites dans les données MIDI sont dues au mouvement du talon du batteur qui bat la pulsation sans particulièrement jouer le charley. Ce mouvement est perçu par le capteur de la batterie électronique mais le charley n'est pas joué.
- Au final, j'ai relevé les ghost-notes et les accents pour la caisse claire ainsi que les accents pour les toms et les cymbales rythmiques (charley et ride).

Conclusion sur les transcriptions manuelles

La transcription des données audio et MIDI contenues dans ces fichiers a permis une analyse plus approndie des critères à relever pour chaque évènement MIDI et de la manière de les considérer dans un objectif de transcription en partition lisible pour un musicien (Voir la section 3.3).

4.3 Transcription polyphonique par parsing

Les différentes étapes de résolution du passage au polyphonique. Les Jams permettent de passer du monophonique au polyphonique. On regroupe tous les évènements qu'on estime être à des dates suffisamment proches dans des clusters que nous avons appelés « Jam ». Un Jam représente une séquence de points successifs dans un segment d'entrée avec des informations supplémentaires sur leur rôle respectif, en supposant

qu'ils sont tous alignés à la même date lors de la quantification. Un Jam est caractérisé par l'index de son premier point dans le segment d'entrée et sa longueur (nombre de points). Une fois dans les Jams, les évènements sont identifiés selon les critères suivants :

— Ignored, les offsets sont ignorés ;
— Note, si c'est une note simple ;
— Rest, si c'est un silence ;
— GraceNote, si c'est un fla ;
— Error.

4.4 Réécriture guidée par une forme rythmique

La démonstration qui suit est basée sur la partition de référence de la figure 3.11 puisque la forme rythmique qui sera utilisée en est directement extraite.

Nous allons montrer :

— la composition de cette forme rythmique ;
— son état finale, c'est à dire toutes les combinaisons entièrement écrites en notation correcte sur partition ;
  ⇒ cela constituera une référence pour la réécriture ;
— un exemple de transformation de la forme rythmique en arbre de rythme ;
— l'application de la séparation des voix sur cet exemple basé sur la référence citée précédemment (la forme rythmique en question) ;
  ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a de voix (deux arbres pour cette forme rythmique) ;
— les règles de simplification propres à la forme rythmique dont nous parlons.

L'objectif de cette démonstration est de montrer comment un jeu de plusieurs formes rythmiques pourrait être implémenter dans le cadre d'une approche dictionnaire.

Motifs et gammes

Motifs

À partir de la partition de référence, les deux motifs de la figure 4.2 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à la mesure 28 avec des variations. Ces deux motifs sont très classiques et pourront être détectés dans de nombreuses performances.

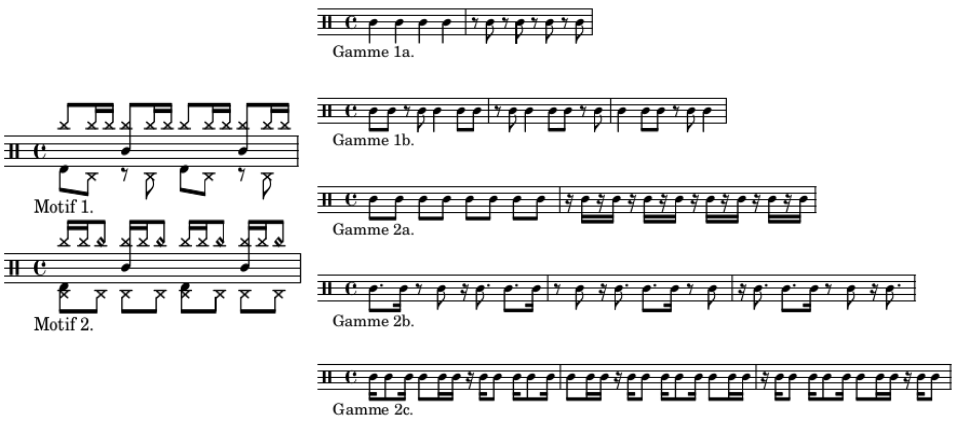


FIGURE 4.2 – Motifs et gammes

Gammes

Les gammes de la figure 4.2 étayent toutes les combinaisons d'un motif en 4/4 binaires jusqu'aux doubles croches.
Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la première ne contient que des noires et la deuxième que des croches en contre-temps. Ces deux possibilités sont combinées de manière circulaire dans les 3 mesures de la deuxième ligne.
Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures dont la première ne contient que des croches et la deuxième que des doubles-croches en contre-temps. Ces deux possibilités sont combinées de manière circulaire dans les lignes 4 et 5 qui contiennent chacunes 3 mesures.

Formes rythmiques — motifs et gammes combinés

Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure 4.2.<dam>à commenter un peu plus, notamment pour dire si la combinaison est faite automatiquement ou non</dam>

Représentation de la forme rythmique en arbres de rythmes

L'arbre de la figure 4.4 servira de base pour le suite de l'expérimentation. Comme indiqué à la racine de l'arbre, il représente la première ligne de la figure 4.3. Même si cet arbre représente parfaitement le rythme concerné, il manque des indications de notation telles que les voix spécifiques à chaque partie du rythme ainsi que les choix d'écriture pour les distances qui séparent les notes de chaque voix entre elles en termes de durée.

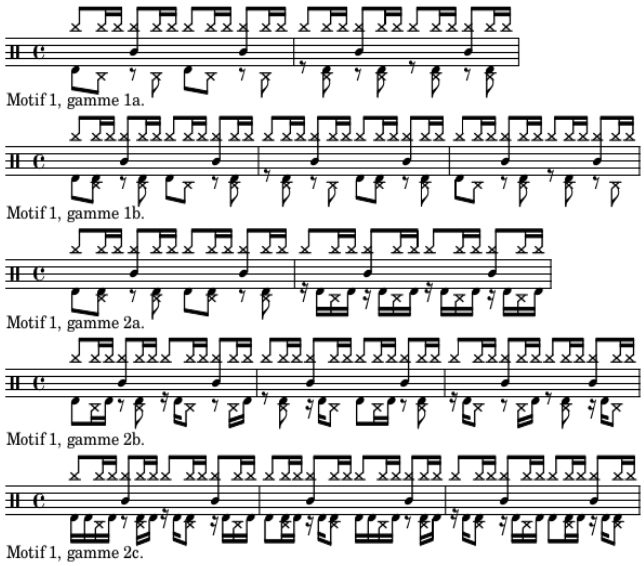


FIGURE 4.3 – Partition d'un forme rythmique en 4/4 binaire

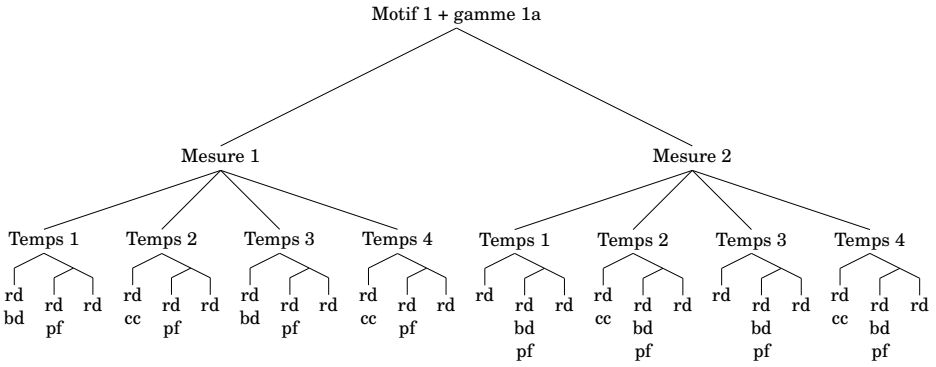


FIGURE 4.4 – Représentation arborescente d'une forme rythmique

Réécriture — séparation des voix et simplification

La séparation des voix

Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui le constituent et les voix seront regroupées en suivant les régles du forme rythmique.
La voix haute (figure 4.5) regroupe la ride et la caisse-claire sur les ligatures du haut.
La voix basse (figure 4.6 regroupe la grosse-caisse et le charley au pied sur les ligatures du bas.

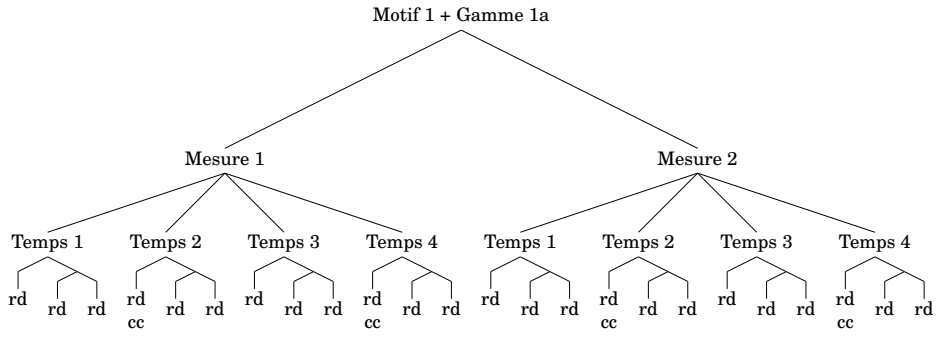


FIGURE 4.5 – Arbre de rythme — voix haute

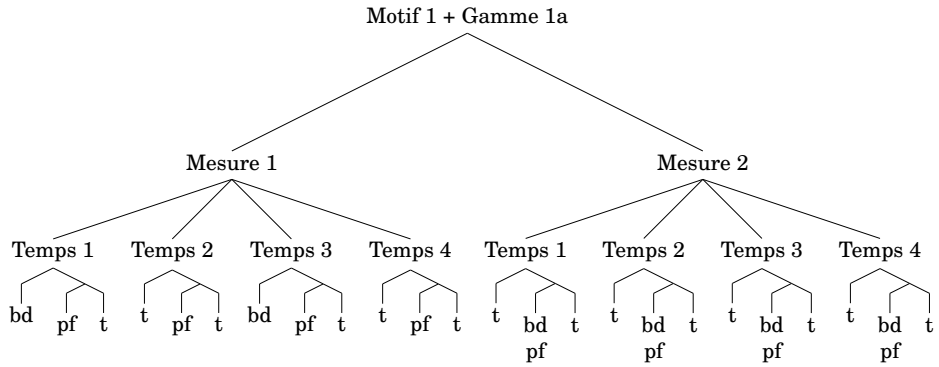


FIGURE 4.6 – Arbre de rythme — voix basse

Les règles de simplifications

L'objectif des règles de simplifications est de réécrire les écarts de durées qui séparent les notes d'une manière appropriée pour la batterie et qui soit la plus simple possible. Les ligatures relient les notes d'un temps entre elles afin de rendre la pulsation visuelle).

Pour les figures ci-dessous :

- x = une note ;
- r = un silence ;
- t = une continuation (point ou liaison)

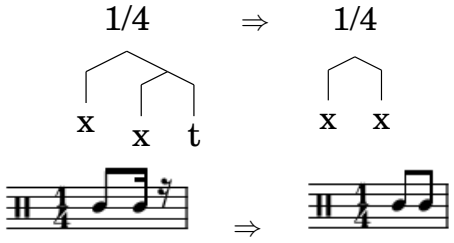


FIGURE 4.7 – Exemple de simplification 1

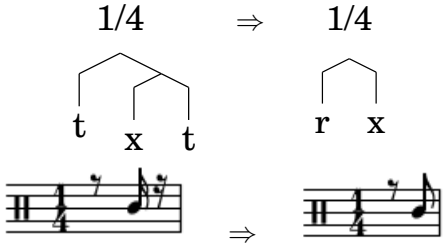


FIGURE 4.8 – Exemple de simplification 2

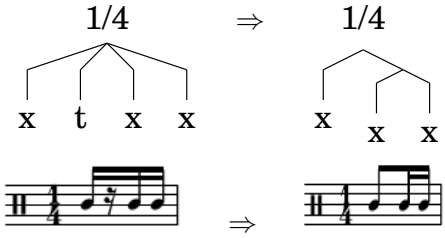


FIGURE 4.9 – Exemple de simplification 3

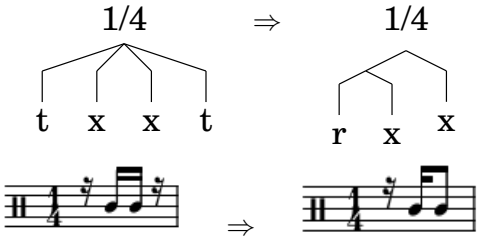


FIGURE 4.10 – Exemple de simplification 4

Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique. Les arbres manquants seront mis en annexe.

Les règles remplacent par un silence les continuations (t) qui sont au début d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il y a des ouvertures de charley, cela n'est pas toujours applicable.

Conclusion sur cette réécriture guidée

La méthode des formes rythmiques étant basée sur une approche dictionnaire, Le premier objectif de cette réécriture guidée est d'orienter la recherche d'autres formes rythmiques par observation du jeu de données et de montrer comment les construire pour agrandir la base de connaissance de Qparse pour la transcription de la batterie.

4.5 BILAN : résultats — évaluation — discussion

la pertinence des choix qui ont été faits et les moyens d'évaluer les résultats potentiels Cette section regroupe les avancées qui ont été réalisées par rapport aux objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résultats de l'ADT avec Qparse. Nous avons améliorer le système de quantification de Qparse pour la batterie, notamment le passage à la polyphonie avec les Jams.
Nous avons pu obtenir des arbres de parsing corrects en améliorant les grammaires avec des fichiers MIDI courts.
Puis, une sortie MEI a aussi été obtenu (encore à vérifier).
Dans cette section, nous discuterons sur la pertinence de l'ensemble des choix qui ont été faits. Nous ferons un bilan des différentes avancés qui ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.

- Le choix de travailler avec lilypond et non verovio. Ce choix était motivé par la liberté totale concernant la notation de la batterie dont un et la disponibilité d'un set de notation de type agostini. C'est la seule application qui me permettait d'écrire la notation de la batterie exactement comme je le souhaitais.
- Avancé de la chaîne de traitement (nous sommes arrivé au arbres de parsing, nous avons traité le polyphonique (identification des regroupements de notes [6]) ⇒ Quelques arbres ont été obtenus sur des exemples simples ([7])
- 2 dimensions de le travail fourni :
  - La volonté de pousser un exemple simple jusqu'au bout de la chaîne pour obtenir des résultats et une évaluation sur au moins un exemple ; - La réalité du travail à fournir pour faire avancer sur la chaîne de traitement. ⇒ Une solution aurait été de considérer les arbres de parsing obtenus après le traitement du polyphonique comme un résultat local possible à évaluer au lieu d'attendre que la chaîne arrive jusqu'à la génération d'une partition mais cela n'était pas prioritaire pendant le stage.
- Création d'un jeu de forme rythmique basique réprésentatif des différents styles à recouvrir. Ce jeu n'a pas pu être créé, car comme

6. fla ou accords entre autres...
7. exemple de 2 mesures, voir ...

vu plus haut, je me suis focalisé sur un exemple pour pouvoir le vérifier entièrement et dans l'espoir de pouvoir le tester en fin de chaîne. **Évaluation** Matcher les motifs aurait été indispensable pour obtenir une quantité de résultats qui justifieraient une évaluation automatique permettant de faire des graphiques.
L'évaluation fut entièrement manuelle car :
⇒ Très dure automatiquement : il faut comparer 2 partitions (réf VS output) Pour l'évaluation, il aurait fallu produire un module.
<dam>je ne sais pas si tu auras encore le temps de faire ça, sinon il faudra décrire comment tu aurais aimé évaluer, proprement et sans résultats chiffrés</dam> L'évaluation est-elle automatique ou manuelle ?
Possibilité d'un export lilypond en arbre pour comparer l'ouput avec la transcription manuelle.
Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel ou MEI pour les comparer et faire des statistiques. Si transformés en MEI : diffscore de Francesco. Possibilité de transformer lilypond(output) et lilypond(ref) en MusicXML pour les comparer ou dans Music21. L'expérimentation peut-être considérer comme une évaluation manuelle ? (magicien d'Oz)
Lilypond vers MIDI + ouput vers MIDI ⇒ Comparaison des MIDI dumpés.

La transcription automatique de la batterie est un sujet passionnant mais difficile : Obtenir la totalité des éléments nécessaires pour le mémoire nécessiterait plus de temps. Une base solide spécifique à la batterie a néanmoins été générée. Elle sera un bon point de départ pour les travaux futurs dont plusieurs propositions sont énoncés dans le présent document.

CONCLUSION GÉNÉRALE

Dans ce mémoire, nous avons traité de la problématique de la transcription automatique de la batterie. Son objectif était de transcrire, à partir de leur représentation symbolique MIDI, des performances de batteur de différents niveaux et dans différents styles en partitions écrites.
Nous avons avancé sur le parsing des données MIDI établissant un processus de regroupement des évènements MIDI qui nous a permis de faire la transition du monophonique vers le polyphonique. Une des données importante de ce processus était de différencier les nature des notes d'un *accord*, notamment de distinguer lorsque 2 notes constituent un *accord* ou un *fla*.
Nous avons établis des *grammaires pondérées* pour le parsing qui correspondent respectivement à des métriques spécifiques. Celles-ci étant sélectionnables en amont du parsing, soit par indication des noms des fichiers MIDI, soit par reconnaissance de la métrique avec une approche dictionnaire de patterns prédéfinis [8] qu'il serait pertinent de mettre en œuvre en machine learning.
Nous avons démontré que l'usage des *systèmes* élimine un grand nombre de calcul lors de la réécriture. Pour la séparation des voix grâce au motif d'un système et pour la simplification grâce aux gammes du motif d'un système. Nous avons aussi montré comment, dans des travaux futurs, un système dont le motif serait reconnu en amont dans un fichier MIDI pourrait prédéfinir le choix d'une grammaire par la reconnaissance d'une métrique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans la chaîne de traitement en terme de réécriture.
Il sera également intéressant d'étudier comment l'utilisation de LM peut améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération entièrement automatisée de partitions de batterie et au problème général de l'AMT de bout en bout.[11]

8. *Motifs* dans les *systèmes* de la présente proposition.

BIBLIOGRAPHIE

[1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition, 1996. – Cité pages 7, 16, 17 et 33.

[2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11 et 14.

[3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité page 12.

[4] Gaël Richard et al. De fourier à la reconnaissance musicale. Voir https://interstices.info/de-fourier-a-la-reconnaissance-musicale/ (2019/02/15). – Cité page 12.

[5] Caroline Traube. Quelle place pour la science au sein de la musicologie aujourd'hui ? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.

[6] Leonard Bernstein Office. The unanswered question : Six talks at harvard. Voir https://leonardbernstein.com/about/educator/norton-lectures (2021/01/01). – Cité page 12.

[7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions entre les traitements de la musique et du langage. *La Lettre des Neurosciences*, 58 :24–26, 2018. – Cité page 12.

[8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo. Techniques de traitement automatique du langage naturel appliquées aux représentations symboliques musicales. In *JIM 2021 - Journées d'Informatique Musicale*, Virtual, France, July 2021. – Cité page 13.

[9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémiologies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.

[10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering music relations with sequential attention. In *NLP4MUSA*, 2020. – Cité page 13.

[11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

lenges and future directions. *Journal of Intelligent Information Systems*, 41, 12 2013. – Cité pages 14, 15, 21 et 63.

[12] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE MESURE, 1997. – Cité page 15.

[13] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Gerhard Widmer, Jason Hockman, Meinard Müller, and Alexander Lerch. A review of automatic drum transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(9) :1457–1483, 2018. – Cité pages 15, 22 et 25.

[14] Elaine Gould. *Behind bars : the definitive guide to music notation*. Faber Music Ltd, 2016. – Cité page 16.

[15] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local musical statistics as guides for audio-to-score piano transcription. *Information Sciences*, 566 :262–280, 2021. – Cité pages 18, 23 et 24.

[16] Nicolas Guiomard-Kagan. *Traitement de la polyphonie pour l'analyse informatique de partitions musicales*. PhD thesis. – Cité page 18.

[17] Moshekwa Malatji. Automatic music transcription for two instruments based variable q-transform and deep learning methods, 10 2020. – Cité page 22.

[18] Antti J. Eronen. Musical instrument recognition using ica-based transform of features and discriminatively trained hmms. *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 22.

[19] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic drum sound description for real-world music using template adaptation and matching methods. *International Conference on Music Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 22.

[20] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Masahiko Sakai. A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring. In *MCM 2019 - Mathematics and Computation in Music*, volume Lecture Notes in Computer Science of *Proceedings of the Seventh International Conference on Mathematics and Computation in Music (MCM 2019)*, Madrid, Spain, June 2019. Springer. – Cité pages 23 et 25.

[21] C. Agon, K. Haddad, and G. Assayag. Representation and rendering of rhythm structures. In *Proceedings of the First International Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002. IEEE Computer Society. – Cité page 24.

[22] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A Term Rewriting Based Structural Theory of Rhythm Notation. Research report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. – Cité page 25.

[23] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating equivalent rhythmic notations based on rhythm tree languages. In *Third International Conference on Technologies for Music Notation and Representation (TENOR)*, Coroña, Spain, May 2017. Helena Lopez Palma and Mike Solomon. – Cité page 25.

[24] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. – Cité pages 4 et 25.

[25] R. Marxer and J. Janer. Study of regularizations and constraints in nmf-based drums monaural separation. In *International Conference on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland, 02/09/2013 2013. – Cité page 26.

[26] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 27.

[27] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 28.

[28] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publications, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins - 94120 Fontenay-sous-Bois, 2000. – Cité pages 30, 43 et 46.

[29] Frédéric Canet. La batterie... mot à mot! Voir https://rimshotetghostnote.fr/ (2021). – Cité page 34.

[30] M. Laurson. Patchwork : a visual programming language and some musical applications. 1996. – Cité page 38.

[31] Jean Bresson, Carlos Agon, and Gérard Assayag. Openmusic visual programming environment for music composition, analysis and research. – Cité page 38.

[32] Dick Grune and Ceriel JH Jacobs. Parsing techniques. *Monographs in Computer Science. Springer,*, page 13, 2007. – Cité page 39.

[33] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of weighted automata*. Springer Science & Business Media, 2009. – Cité page 39.

[34] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David Bamman. Learning to groove with inverse sequence transformations. In *International Conference on Machine Learning (ICML)*, 2019. – Cité page 49.