
**Institut National des Langues et Civilisations
Orientales**

Département Textes, Informatique, Multilinguisme

Titre du mémoire

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	5
Liste des tableaux	5
Introduction générale	7
1 Contexte	9
1.1 Informatique musicale	10
1.2 TAL et MIR	10
1.3 La transcription automatique de la musique	11
1.4 La transcription automatique de la batterie	13
1.5 Les représentations de la musique	15
2 État de l'art	17
2.1 Monophonique et Polyphonique	17
2.2 Audio vers MIDI	18
2.3 MIDI vers partition	18
3 Méthodes	19
3.1 La notation de la batterie	20
3.2 Modélisation pour la transcription	23
3.3 Représentation symbolique (MIDI)	25
3.4 Qparse	26
3.5 Les systèmes	27
4 Expérimentations	33
4.1 Corpus	33
4.2 Analyse MIDI-Audio	35
4.3 Expérimentation théorique d'un système	40
4.4 Développement	44
5 Discussion	49
5.1 Travaux réalisés	49
5.2 Travaux non-réalisés	49
5.3 Travaux futures	50
Conclusion générale	51

Bibliographie

53

LISTE DES FIGURES

1.1	Transcription automatique	12
3.1	Hauteur et têtes de notes	20
3.2	Nuances	20
3.3	Durées	22
3.4	Séparation des voix	22
3.5	Exemple évènements avec durée	25
3.6	Critère pour un évènement	25
3.7	Exemple évènements sans durée	26
4.1	référence 1	40
4.2	Motifs	41

LISTE DES TABLEAUX

3.1	Pitches et instruments	23
3.2	Vélocité et nuances	23
3.3	Sytèmes	28

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 ayant toutes pour objectif d'améliorer **qparse**, un outil de transcription automatique de la musique sur sa capacité à transcrire la batterie. Nous ne parlerons donc pas directement de langues naturelles, mais de l'écriture automatique de partitions de musique à partir de données audios ou symbolique. Cette exercice nécessitera la manipulation d'un langage musical codifié avec une grammaire (solfège, durées, nuances, volumes) et soulèvera des problématiques concernées par les techniques du traitement automatique des langues.

Voici la proposition de ce mémoire ainsi que les contributions apportées lors du stage :

- Proposition principale : les systèmes (3.5) :
Recherche de rythmes génériques en amont dans la chaîne de traitement.
⇒ L'objectif 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 métrique et les règles de réécriture.
- Une description de la notation de la batterie (3.1)
- Une modélisation de la transcription de la batterie (3.2)
- Analyse MIDI-Audio (4.2)
- Théorie et test unitaire pour le passage au polyphonique (4.4.2)
- Création de grammaires pondérée pour la batterie (4.4.3)
- Contribution sur la branch « distance » dans :
 - `qparselib/notes/cluster.md`
 - `qparselib/src/segment/import/` :
`DrumCode` hpp et cpp

Nous présenterons le contexte général suivi d'un état de l'art et nous définirons de manière générale le processus de transcription automatique de la musique pour enfin étayer les méthodes utilisées pour la transcription automatique de la batterie et nous présenterons les principales contributions apportées à l'outil **qparse**. Nous décrirons ensuite le corpus ainsi que les différentes expérimentations menées. Nous concluerons par une

discussion sur les résultats obtenus et les pistes d'améliorations futures à explorer.

CONTEXTE

Sommaire

1.1	Informatique musicale	10
1.2	TAL et MIR	10
1.2.1	Langages	10
1.2.2	Transcription et TAL	11
1.3	La transcription automatique de la musique	11
1.3.1	La transcription musicale	11
1.3.2	La transcription automatique	11
1.3.3	Le processus général	12
1.3.4	Utilité de l'AMT	13
1.4	La transcription automatique de la batterie	13
1.5	Les représentations de la musique	15
1.5.1	Audio	15
1.5.2	MIDI	15
1.5.3	Partition	15

Introduction

La transcription automatique de la musique (AMT) est un défi ancien [1] et difficile qui n'est toujours pas résolu. Il a engendré une pluie de sous-tâches qui ont donné naissance au domaine de la recherche d'information musicale (MIR). Actuellement, de nombreux travaux de MIR font appel au traitement automatique des langues (TAL)¹.

Dans ce chapitre, nous parlerons de l'informatique musicale, nous tenterons d'établir les liens existants entre le MIR et le TAL ainsi qu'entre les notions de langage musical et langue naturelle. Nous traiterons également de l'utilité et du problème de l'AMT et de la transcription automatique de la batterie (ADT).

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

Enfin, nous décrirons les représentations de la musique qui sont nécessaires à la compréhension du présent travail.

1.1 Informatique musicale

L'informatique musicale² est une étude du traitement de la musique [2], en particulier des représentations musicales, de la transformée de Fourier pour la musique³, de l'analyse de la structure de la musique et de la reconnaissance des accords. D'autres sujets de recherche en informatique musicale comprennent la modélisation informatique de la musique (symbolique, distribuée, etc.), l'analyse informatique de la musique, la reconnaissance optique de la musique, les éditeurs audio numériques, les moteurs de recherche de musique en ligne, la recherche d'informations musicales et les questions cognitives dans la musique.

Le MIR apparaît vers le début des années 2000 [3]. C'est une science interdisciplinaire qui fait appel à de nombreux domaines comme la musicologie, l'analyse musicale, la psychologie, les sciences de l'information, le traitement du signal que les méthodes d'apprentissage automatisé en informatique et qui a pour but de les catégoriser. Cette nouvelle discipline est notamment soutenu par de grandes compagnies du web qui veulent développer des systèmes de recommandation de musique ou des moteurs de recherche dédiés au son et à la musique.

1.2 TAL et MIR

1.2.1 Langages

Aborder la musique à travers le 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. Quelques travaux en neuroscience ont abordé la question, notamment par observation des processus cognitifs et neuronaux que les systèmes de traitement de ces deux langages avaient en communs. Dans le travail de Poulin-Charronnat et al. [4], 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 implicitement par le système cognitif. 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. [5] 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 du sens d'une phrase musi-

2. https://en.wikipedia.org/wiki/Music_informatics

3. <https://interstices.info/de-fourier-a-la-reconnaissance-musicale/>

cale apparaît, selon eux, à la fois comme une limite et un défi majeur pour l'étude de cette analogie.

Ici, Digression sur la musicologie calculatoire vs linguistique computationnelle ?

D'autres travaux très récents, ont aussi été révélés lors de la *première conférence sur le NLP pour la musique et l'audio (NLP4MusA 2020)*. Lors de cette conférence, Jiang et al. [6] ont présenté leur implémentation d'un modèle de langage musical auto-attentif 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.

1.2.2 Transcription et TAL

Lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole.

1.3 La transcription automatique de la musique

1.3.1 La transcription musicale

En musique, la transcription⁴ est la pratique consistant à noter un morceau ou un son qui n'était auparavant pas noté et/ou pas populaire en tant que musique écrite, par exemple, une improvisation de jazz ou une bande sonore de jeu vidéo. 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.

1.3.2 La transcription automatique

L'objectif de la transcription automatique de la musique (AMT) [7] est de convertir la performance d'un musicien en notation musicale - un peu comme la conversion de la parole en texte dans le traitement du langage naturel. Comme déjà évoqué précédemment, il s'agit d'un problème ancien et difficile. C'est un « graal » de l'informatique musicale.

En 1976, H. C. Longuet-Higgins [1] 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.

De même pour les chercheurs en audio James A. Moorer, Martin Piszczalski et Bernard Galler⁵ qui, en 1977, ont utilisé leurs connaissances

4. [https://en.wikipedia.org/wiki/Transcription_\(music\)](https://en.wikipedia.org/wiki/Transcription_(music))

5. [https://en.wikipedia.org/wiki/Transcription_\(music\)](https://en.wikipedia.org/wiki/Transcription_(music))

en ingénierie de l'audio et du numérique pour programmer un ordinateur afin de lui faire analyser un enregistrement musical numérique de manière à détecter les lignes mélodiques, les accords et les accents rythmiques des instruments à percussion.

1.3.3 Le processus général

La tâche de transcription automatique de la musique comprend deux activités distinctes : l'analyse d'un morceau de musique et l'impression d'une partition à partir de cette analyse. La figure suivante, qui est une proposition de Benetos et Al. [7], représente l'architecture générale d'un système de transcription musicale.

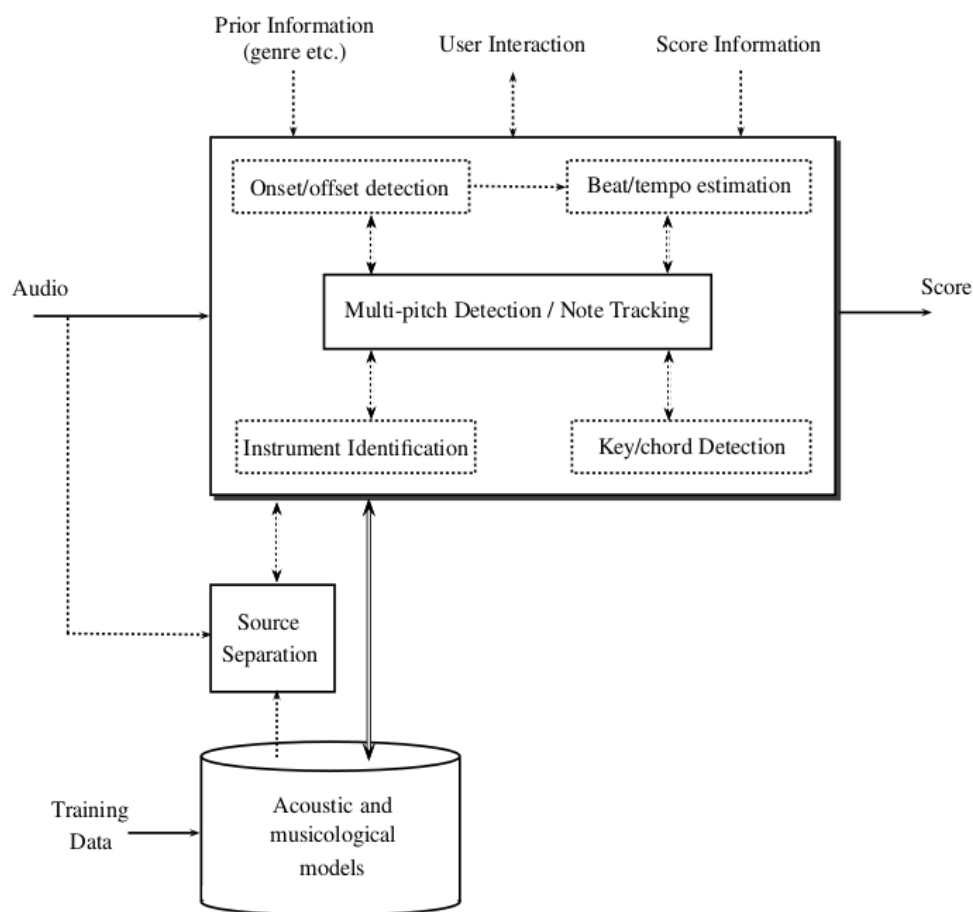


FIGURE 1.1 – Transcription automatique

Les sous-systèmes et algorithmes optionnels sont présentés à l'aide de lignes pointillées. Les doubles flèches mettent en évidence les connexions

entre les systèmes qui incluent la fusion d'informations et une communication plus interactive entre les systèmes.

1.3.4 Utilité de l'AMT

Transcrire des solos, intérêt \Rightarrow constitution de corpus musicologique.

Voir l'intro de [7]

Les performances des systèmes actuels ne sont pas encore suffisantes pour certaines applications qui exigent un haut degré de précision [7]. Même si les applications typiques de l'AMT comprennent l'estimation de la multi-tonalité, la classification des genres musicaux, la détection du début et de la fin des notes de musique, l'estimation du tempo, le suivi du rythme et la transcription de la musique, la plupart des travaux se sont concentrés sur le traitement du signal vers la génération du midi [8]. Seuls quelques travaux récents [9] s'intéressent de près à la création d'outils permettant la génération de partition.

1.4 La transcription automatique de la batterie

La batterie est 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 la mélodie du thème en jazz) afin d'improviser un accompagnement approprié 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'école de batterie partout dans le monde. La musique assistée par ordinateur (MAO), a elle aussi largement contribué à l'expansion des partitions de batterie puisque les compositeurs pouvaient utiliser des boîte à rythmes ou des séquenceurs pour écouter leurs productions et ainsi écrire une partition pour un batteur en s'émancipant de sa présence.

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'écritures. Reconnaître la métrique principale d'un rythme, la façon de regrouper les notes par les 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 difficultés.

Au cœur du système se trouvent les algorithmes de détection des multi-pitches et de suivi des notes. Quatre sous-tâches de transcription liées à la détection des hauteurs multiples et au suivi des notes apparaissent comme des algorithmes facultatifs du système (cases en pointillé) qui peuvent être intégrés dans un système de transcription. Il s'agit de l'identification de l'instrument, de l'estimation de la tonalité et de l'accord, de

la détection de l'apparition et du décalage, et de l'estimation du tempo et du rythme. La séparation des sources, un problème indépendant mais lié, pourrait être traitée par un système séparé qui pourrait informer et interagir avec le système de transcription en général, et plus spécifiquement avec le sous-système d'identification des instruments. En option, des informations peuvent également être fournies de manière externe au système de transcription. Elles peuvent être données sous forme d'informations préalables (c'est-à-dire le genre, l'instrumentation, etc.), via l'interaction de l'utilisateur ou en fournissant des informations à partir d'une partition préexistante partiellement correcte ou incomplète. Enfin, les données de formation peuvent être utilisées pour apprendre des modèles acoustiques et musicologiques qui, par la suite, informent le système de transcription et interagissent avec lui. avec le système de transcription.

Les applications de l'AMT ont aussi de la valeur dans les domaines oraux ou d'improvisation qui manquent de partition (jazz, pop) [7]. Les applications de l'ADT seraient utiles pour ces styles de musiques puisque la batterie y est amplement représentée. Un grand nombre de travaux ont déjà été menés dans le domaine de l'ADT. La plupart ont été énumérés par Wu et al. [10] qui, pour mieux comprendre la pratique des systèmes d'ADT, se concentrent sur les méthodes basées sur la factorisation matricielle non négative et celles utilisant des réseaux neuronaux récurrents.

La batterie a un statut à part dans l'univers de l'AMT puisqu'il s'agit d'instruments sans hauteur, d'événements auxquels une durée est rarement attribuée et de notations spécifiques (par exemple sur les têtes de notes). Si les ordinateurs étaient capables d'analyser la partie de la batterie dans la musique enregistrée, cela permettrait une variété de tâches de traitement de la musique liées au rythme. En particulier, la détection et la classification des événements sonores de la batterie par des méthodes informatiques est considérée comme un problème de recherche important et stimulant dans le domaine plus large de la recherche d'informations musicales [10]. Cependant, la plupart des travaux déjà entrepris se concentrent sur des méthodes de calcul pour la détection d'événements sonores de batterie à partir de signaux acoustiques ou sur la séparation entre les événements sonores de batterie avec ceux des autres instruments dans un orchestre ou un groupe de musique [11], ainsi que sur l'extraction de caractéristiques de bas niveau telles que la classe d'instrument et le moment de l'apparition du son. Très peu d'entre eux ont abordé la tâche de générer des partitions de batterie.

1.5 Les représentations de la musique

1.5.1 Audio

1.5.2 MIDI

1.5.3 Partition

Édition de partition.

Mettre une image de partition ici Une partition de musique⁶ est un document qui porte la représentation systématique du langage musical sous forme écrite. Cette représentation est appelée transcription et elle sert à traduire les quatre caractéristiques du son musical :

- la hauteur ;
- la durée ;
- l'intensité ;
- le timbre.

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).

Conclusion

Dans le cas, de l'ADT, l'architecture reste la même mais de nombreuses choses seront à affiner, notamment pour les questions de continuation ainsi que celle des ghost-notes et des accents.

6. [https://fr.wikipedia.org/wiki/Partition_\(musique\)](https://fr.wikipedia.org/wiki/Partition_(musique))

ÉTAT DE L'ART

Sommaire

2.1	Monophonique et Polyphonique	17
2.2	Audio vers MIDI	18
2.3	MIDI vers partition	18
2.3.1	Approche linéaire	18
2.3.2	Approche hiérarchique	18

Introduction

ainsi que les différentes avancées qui ont déjà eues lieux dans le domaine de la transcription de la musique. Et enfin, les avancées en terme de transcription automatique de la batterie. Actuellement, des modèles polyvalent qui n'arrivent pas à récupérer toute la richesse des sons sont utilisés.

2.1 Monophonique et Polyphonique

C'est le problème de la séparation des voix. Les premiers travaux ont été fait sur l'identification des instruments monophoniques (une seule note à la fois, ou plusieurs notes de même durée en cas de monophonie par accord). 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 polyphonique.

Le fort degrés de chevauchement entre les durées ainsi qu'entre les fréquences rendent l'identification des instruments polyphoniques difficile. Cette tâche est étroitement liées à la séparation des sources.

La création d'un système automatisé capable de transcrire de la musique polyphonique sans restrictions sur le degré de polyphonie ou le type d'ins-

trument reste encore ouverte. Un des principaux enjeux de ce problème est la détection des hauteurs de son multiples (F0 multiples) [7].

2.2 Audio vers MIDI

Voir : [7] - Multi-pitch détection and note tracking

- Détection of onsets and offsets - Instrument recognition
- Extraction of rhythmic information (tempo, beat, and musical timing)
- Estimation of pitch and harmony (key, chords and pitch spelling)

2.3 MIDI vers partition

2.3.1 Approche linéaire

nakamura [12]

2.3.2 Approche hiérarchique

[9] évoque la nécessité d'une approche hiérarchique pour la production automatique de partition même si la quantification du rythme se fait le plus souvent par la manipulation de données linéaires :

- rtu (real time units : secondes) vers mtu (musical time units : temps, métrique,...)

Dans [9], les modèles de grammaire exposés sont différents de modèles markoviens linéaires de précédent travaux. [13] [14]

Conclusion

Nous avons décidé de compléter le travail qui concerne la batterie en commençant par l'endroit le moins pratiqué, à savoir la transcription en partition pour à l'avenir réaliser la chaîne de bout en bout : de l'audio jusqu'à l'écriture de partition.

MÉTHODES

Sommaire

3.1	La notation de la batterie	20
3.1.1	Hauteurs et têtes de notes	20
3.1.2	Nuances	20
3.1.3	Durées	21
3.1.4	Voix	22
3.2	Modélisation pour la transcription	23
3.2.1	Les pitches	23
3.2.2	La vélocité	23
3.2.3	Les arbres de rythmes	24
3.3	Représentation symbolique (MIDI)	25
3.4	Qparse	26
3.4.1	La grammaire pondérée	26
3.4.2	Le parsing	27
3.4.3	La réécriture	27
3.5	Les systèmes	27

Introduction

Dans ce chapitre, nous expliquerons en détails les méthodes que nous avons employées pour l'ADT. Nous commencerons par une description de qparse et des arbres de rythmes. Nous proposerons ensuite une modélisation comprenant une description de la notation de la batterie mise en relation avec les informations MIDI, ceci ayant pour objectif le parsing des données MIDI en arbre syntaxique. Enfin, nous démontrerons un modèle théorique de pattern (implémentable) qui devra être utilisé comme base de connaissance pour obtenir un système plus rapide et une meilleure qualité en sortie.

3.1 La notation de la batterie

Les 3 parties d'une note en général :

- durée
- hampe
- tête de note (peut aussi indiquer la durée mais en batterie on évitera les blanches, etc.)

source : https://fr.wikipedia.org/wiki/Note_de_musique

3.1.1 Hauteurs et têtes de notes

Pour la transcriptions, nous proposons de choisir la base Agostini. La caisse claire centrale sur la portée est aussi centrale sur la batterie est elle est un élément qui conditionne la position des jambes (écart entre les pédales, etc.) ainsi que l'organisation des éléments en hauteur (toms, cymbales, etc.). On pensera en terme de symétrie la répartition des éléments par rapport au point central que constitue la caisse claire.

Cette symétrie s'opère en trois dimensions :

- Les hauteurs en terme de fréquences ;
- La hauteur physique des éléments :
Du bas vers le haut : pédales, toms et caisse, cymbales
- L'ergonomie, qui hiérarchise l'importance des éléments sur la portée (caisse claire au centre, hh-pied et ride sont aux deux extrémités).

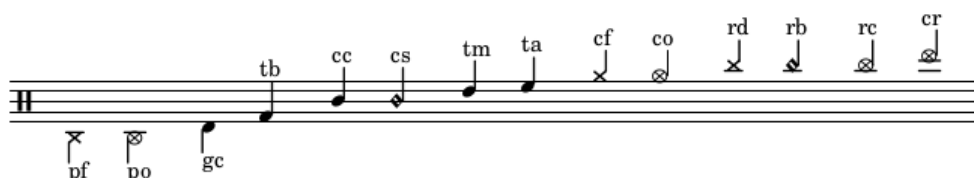


FIGURE 3.1 – Hauteur et têtes de notes

3.1.2 Nuances



FIGURE 3.2 – Nuances

Bien expliquer les accents, remplacer p et f par g et a
⇒ nuance VS articulation

3.1.3 Durées

Basé sur [15] et sur [16]

Pour la plupart des instruments mélodiques, la liaison et le point sont les deux seules possibilités en cas d'équivalence rythmique pour des notes dont la durée de l'une à l'autre est ininterrompue. Mais puisque les durées des notes n'ont pas d'importance en batterie, l'usage des silences pour combler la distance rythmique entre deux notes devient possible.

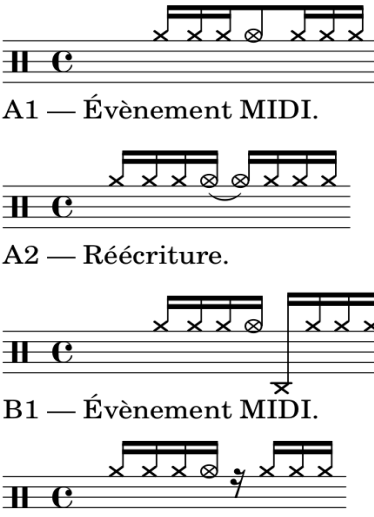
Ceci pris en compte, et étant donné que les indications de durée dans les têtes de notes ne sont pas pratique en batterie (les symboles « x » des cymbales ne peuvent pas porter d'indication de durée dans la tête de notes¹), l'écriture à l'aide de silences sera privilégiée comme indication de durée sauf dans les cas où cela reste impossible. Ce choix à pour but de n'avoir qu'une manière d'écrire toutes notes, que leurs têtes de notes soit modifiées ou non.

Exemple blanche vs noire + soupir

Les cymbales-crash et les ouvertures de charley constituent les seuls cas qui excluent cette option. Le charley car ses ouvertures/fermetures sont presque toujours quantifiées et les cymbales-crash car elles peuvent être arrêtées à la main de manière quantifié aussi mais ce cas est très rare, nous allons donc nous concentrer sur les ouvertures de charley et considérer les crashes comme des événements sans durée.

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.

1. Certains logiciels le permettent mais leur lecture reste peu aisée



A1 — Évènement MIDI.

A2 — Réécriture.

B1 — Évènement MIDI.

B2 — Réécriture.

FIGURE 3.3 – Durées

3.1.4 Voix

Plusieurs écritures sont possibles pour un même rythme :



Exemple 1.

Exemple 2.

Exemple 3.

FIGURE 3.4 – Séparation des voix

Sur la figure 3.4, il faudra faire un choix entre les exemples 1, 2 et 3 qui sont trois façon d'écrire la même chose. Ce choix se fera en fonction de

la lisibilité, de quelles instruments auront des phrasés plus ou moins chargé et/ou variés, auquel cas on les mettra dans une seule voix afin de ne pas charger la partition, etc. Ainsi l'arbre syntaxique de départ sera divisé en autant d'instruments qui le constituent et les voix seront regroupées de manière cohérentes.

3.2 Modélisation pour la transcription

3.2.1 Les pitches

Codes	Instruments	Pitches
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 3.1 – Pitches et instruments

Pas de charley pied ouvert. . .

3.2.2 La vélocité

Codes	Instruments	Pitches	Vélocité
cop	charley-main-ouvert	46	?

TABLE 3.2 – Vélocité et nuances

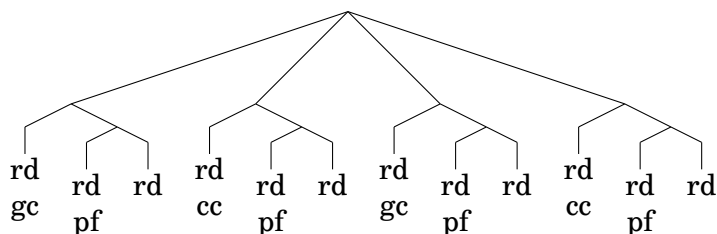
Nous ne prendrons en compte la vélocité que pour la cc, les toms et les cymbales jouées aux mains. Les nuances de grosse caisse et charley aux pieds sont le plus souvent insignifiantes, elles ne sont marquées sur le figure qu'à titre indicatif. Si la vélocité est en dessous de 40, il s'agit de ghost-notes : la tête de note devra être entouré de parenthèses et le suffixe *p* (*piano*) devra être ajouté au codes de l'instrument. (Voir ccp ci-dessus.) Si la vélocité est au dessus de 90, il s'agit de notes accentuées :

le symbole « > » et le suffixe *f* (*forte*) devra être ajouté au codes de l'instrument. (Voir ccf ci-dessus.) Lorsque la vitesse va de 40 à 89, on considérera le volume comme normal et aucun symbole supplémentaire ne sera ajouté à la note.

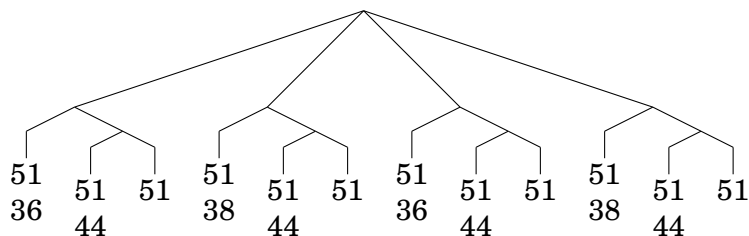
Le charley de pitch 46 est considéré comme le charley ouvert joué à la main sur le haut de la cymbale mais souvent, ça correspond au geste « tranche-olive » de la baguette lorsque le batteur accentue avec la tranche et joue moins fort avec l'olive sur le plat de la cymbale. Je vais dans un premier temps considérer le pitch comme **charley-main-ouvert-piano** (ghost-note)

3.2.3 Les arbres de rythmes

Voici une représentation de la figure 3.4 en arbre de rythme avec les codes de chaque instrument :



Ci-dessous, le même arbre dont les codes des instruments sont remplacés par leurs données midi respectives :



Cet arbre représente un rythme unique dont les possibilités de notation sur une partition sont théoriquement multiples. Les trois exemples de la figure 3.4 peuvent-être représentés par les arbres ci-dessus.

3.3 Représentation symbolique (MIDI)

MIDI (Musical Instrument Digital Interface) est une norme technique qui décrit un protocole de communication, une interface numérique et des connecteurs électriques permettant de connecter une grande variété d'instruments de musique électroniques, d'ordinateurs et d'appareils audio connexes pour jouer, éditer et enregistrer de la musique.²

Les données midi sont représentées sous forme de piano-roll. Chaque points sur la figure suivante est appelé « évènement midi » :

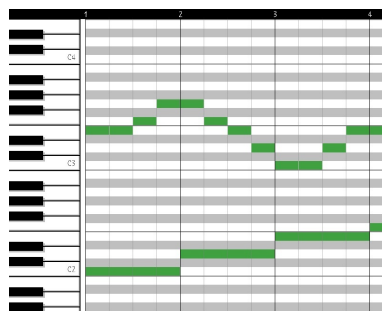


FIGURE 3.5 – Exemple évènements avec durée

Chaque évènement MIDI rassemble un ensemble d'informations sur la hauteur, la durée, le volume, etc. . . :

Protocol	Event
Property	Value
Type	Note On/Off Event
On Tick	15812
Off Tick	15905
Duration	93
Note	45
Velocity	89
Channel	9

FIGURE 3.6 – Critère pour un évènement

Pour la batterie, les évènements sont considérés sans durée, nous ignorons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le channel ne nous sera pas utile non-plus. *Ici, définir Tick et channel.*

2. <https://en.wikipedia.org/wiki/MIDI>

Voici un exemple de piano-roll midi pour la batterie :

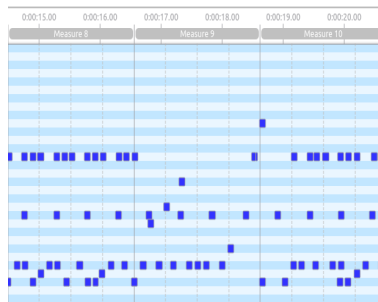


FIGURE 3.7 – Exemple évènements sans durée

On observe que toutes les durées sont identiques.

3.4 Qparse

Mettre ici un schéma de la chaîne de traitement de qparse (workflow)

Pb du MIDI avec qparse entrée ON-OFF \Rightarrow sortie 1 seul symbole.

Qparse produit une partition musicale en prenant en entrée une performance musicale symbolique (par exemple un fichier MIDI) et un automate à arbre pondéré décrivant un langage de rythmes préférés (grammaire pondérée). La quantification des rythmes est basée sur des algorithmes d'analyse syntaxique applicables sur des automates arborescents.³ En entrée : midi (séquence d'événements datés (piano roll) accompagné d'une grammaire pondérée)

\Rightarrow parsing

\Rightarrow global parsing tree

\Rightarrow RI (Représentation Intermédiaire) arbres locaux par instruments

\Rightarrow Sortie (xml, mei, lilypond, ...)

Minimiser la distance entre le midi et la représentation en arbre.

Un des problèmes de Qparse était qu'il était limité au monophonique. \Rightarrow Expliquer ici les limites du monophonique...

3.4.1 La grammaire pondérée

La grammaire pondérée qui accompagne le MIDI en input est une grammaire hors-contexte pondérée. Chaque règle comporte un poids qui sert à favoriser certains rythmes plutôt que d'autres.⁴

3. <https://qparse.gitlabpages.inria.fr>

4. <https://qparse.gitlabpages.inria.fr/docs/scientific/>

3.4.2 Le parsing

Le parsing du midi donné en input crée une représentation symbolique sous forme d'arbre de rythme.

Ici \Rightarrow exemple avec :

3bars_fill_groove-016.mid \Rightarrow arbre

3.4.3 La réécriture

Séparation des voix

Simplification

Ici, description basique des règles de réécriture

3.5 Les systèmes

Un système est la combinaison d'un ou plusieurs éléments qui jouent un rythme en boucle (motif) et d'un autre élément qui joue un texte rythmique variable mais respectant les règles propre au système (gamme).

Système = motif + gamme/texte

motif = rythmes coordonnés joués avec 2 ou 3 membres en boucle (réparti sur 1 ou 2 voix)

gamme/texte = rythme irrégulier joué avec un seul membre sur le motif (Réparti sur 1 voix). La gamme d'un système considère l'ensemble des combinaisons que le batteur pourrait rencontrer en interprétant un texte rythmique à l'aide du système.

Nous partirons de propositions génériques de systèmes (environs trois systèmes dans différents styles de batterie) que nous tenterons de détecter dans le jeu de données groove.

Quatre systèmes standards :

- binaire
- ternaire (shuffle, afro, rock)
- jazz
- afro-cubain

Nous travaillerons aussi sur la détection de répétitions sur plusieurs mesures afin de pouvoir corriger des erreurs sur une des mesures qui aurait dû être identique aux autres mais qui présente des différences.

Intérêt des systèmes

Détection d'indication de mesure et choix de grammaire pondérée

Il faut prendre en compte l'existence potentielle de plusieurs grammaires (*un fichier wta par grammaire*) chacune dédiée à un type de contenu MIDI. Le choix d'une grammaire pondérée doit être fait avant le parsing puisque qparse prends en entrée un fichier MIDI et un fichier wta.

Pour les expériences effectuées avec le Groove MIDI Data Set, le style et l'indication de mesure sont récupérables par les noms de fichiers MIDI, mais il faudra par la suite les trouver automatiquement sans autres indications que les données MIDI elles-mêmes.

En conséquence, les motifs des systèmes devront être recherchés sur l'input (*fichiers MIDI*) avant le lancement du parsing, afin de déterminer la métrique en amont en vu de sélectionner la grammaire pondérée (*fichier wta*) adéquate pour le parsing. Nous pensons que cette tâche devrait être effectuée en Machine Learning. **Les systèmes devront être matchés sur l'input MIDI**

- Définir une métrique ;
- Réécriture : séparation des voix ;
- Réécriture : Set de règles spécifiques de simplification.

Il faudra créer un ensemble de systèmes comprenant leurs règles spécifiques de réécriture (séparation des voix et simplifications). 3 grandes catégories :

Systèmes	Métriques	Subdivisions	Possibles	nb voix
binaires	simple	doubles-croches	triolet, sextolet	2
jazz	simple	triolet	croches et doubles-croches	2
ternaires	complexe	croches	duolet, quartelet	2
afros-cubains	simple	croches	-	3

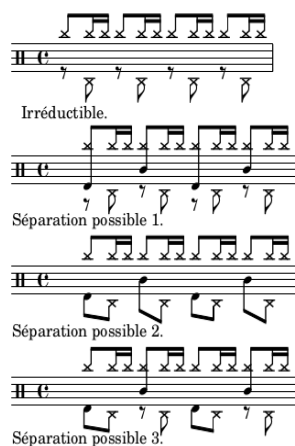
TABLE 3.3 – Systèmes

- ternaire (mesures complexes, principalement croches, noire pointée, duolets et quartelets possibles)
- afro-cubain (mesure)
- Tout transcrire avec lilypond et en arbres d'analyse syntaxique.
- Créer les arbres de voix séparées.
- Écrire les règles de réécriture.
- Créer les arbres de voix séparées simplifiés (rewriting).

Pour la **séparation des voix** et la **définition des métriques**, nous nous intéresserons principalement à la partie *motif* des systèmes qui seront présentés. La partie *texte* nous intéressera plus pour les **combinaisons de réécritures**.

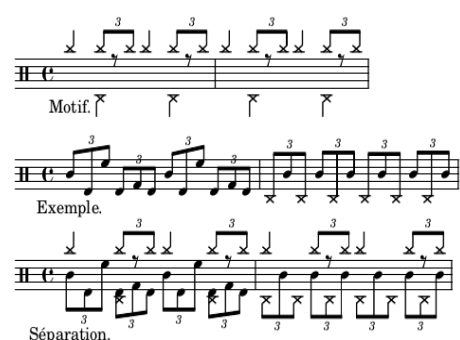
Réécriture — Pour la séparation des voix

Motif 4-4 binaire



Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système est constitué par défaut du motif ride/ch-pf/cc et d'un texte joué à la grosse-caisse. La troisième séparation proposée est privilégiée car elle répartit selon 2 voix, une voix pour les mains (ride + cc) et une voix pour les pieds (ch-pf + gc). Ce choix paraît plus équilibré car deux instruments sont utilisés par voix et plus logique pour le lecteur puisque les mains sont en haut et les pieds en bas.

Motif 4-4 jazz



Dans la plupart des méthodes, le charley n'est pas écrit car considéré comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture : la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici, le partie prit et de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et 2 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 en haut et toms, caisse-claire, grosse-caisse et pédale de charley en bas.

Système 4-4 afro-cubain



Pour la reconnaissance de la métrique

12/8 vs 4/4 ternaire

Motif 12/8

Pour les règles de réécriture

Les textes qui accompagnent les motifs étayent toutes les combinaisons d'un systèmes.

Exemples à écrire en arbre :

- SI (pas pf) ET (note sur un temps suivie de note en l'air) :
⇒ (Temps1 : Note pertinente) + (Temps2 : Silence pertinent + Note pertinente.)
- Si (po ou co) déborde sur le temps suivant :
⇒ Liaison car marchera dans tous les cas même la où le point ne marchera pas (voir A2).
- Une blanche sera écrite noir + soupir.

Conclusion

Bilan sur les différentes méthodes employées et la contribution que cela représente.

EXPÉRIMENTATIONS

Sommaire

4.1	Corpus	33
4.2	Analyse MIDI-Audio	35
4.2.1	Comparaisons de transcriptions	36
4.2.2	Transcription de partition	40
4.3	Expérimentation théorique d'un système	40
4.3.1	Motifs	40
4.3.2	Gammes	41
4.3.3	Systèmes	42
4.3.4	Démonstration	43
4.4	Développement	44
4.4.1	DrumCode	44
4.4.2	Tests unitaires sur les Jams	44
4.4.3	Parsing	45

Introduction

Dans ce chapitre, nous présenterons le corpus, les expérimentations et les différents choix effectués pour les tests.

4.1 Corpus

Différence de qualité de jeu

groove MIDI dataset

<https://magenta.tensorflow.org/datasets/groove>



Des batteurs pro ont été engagés pour jouer sur un roland td-11

The Groove MIDI Dataset (GMD), has several attributes that distinguish it from existing ones:

- The dataset contains about 13.6 hours, 1,150 MIDI files, and over 22,000 measures of drumming.
- Each performance was played along with a metronome set at a specific tempo by the drummer.
- The data includes performances by a total of 10 drummers, with more than 80% of duration coming from hired professionals. The professionals were able to improvise in a wide range of styles, resulting in a diverse dataset.
- The drummers were instructed to play a mix of long sequences (several minutes of continuous playing) and short beats and fills.
- Each performance is annotated with a genre (provided by the drummer), tempo, and anonymized drummer ID.
- Most of the performances are in 4/4 time, with a few examples from other time signatures.
- Four drummers were asked to record the same set of 10 beats in their own style. These are included in the test set split, labeled `eval-session/groove1-10`.
- In addition to the MIDI recordings that are the primary source of data for the experiments in this work, we captured the synthesized audio outputs of the drum set and aligned them to within 2ms of the corresponding MIDI files.

Les métadatas :

The metadata file (`info.csv`) has the following fields for every MIDI/WAV pair:

Field	Description
drummer	An anonymous string ID for the drummer of the performance.
session	A string ID for the recording session (unique per drummer).
id	A unique string ID for the performance.
style	A string style for the performance formatted as "<primary>/<secondary>". The primary style comes from the Genre List below.
bpm	An integer tempo in beats per minute for the performance.
beat_type	Either "beat" or "fill"
time_signature	The time signature for the performance formatted as "<numerator>-<denominator>".
midi_filename	Relative path to the MIDI file.
audio_filename	Relative path to the WAV file (if present).
duration	The float duration in seconds (of the MIDI).
split	The predefined split the performance is a part of. One of "train", "validation", or "test".

Genre List: afrobeat, afrocuban, blues, country, dance, funk, gospel, highlife, hiphop, jazz, latin, middleeastern, neworleans, pop, punk, reggae, rock, soul

A train/validation/test split configuration is provided for easier comparison of model accuracy on various tasks.

Split	Beats	Fills	Measures (approx.)	Hits	Duration (minutes)
Train	378	519	17752	357618	648.5
Validation	48	76	2269	44044	82.2
Test	77	52	2193	43832	84.3
Total	503	647	22214	445494	815.0

Détails (entre autres tensorflow avec le dataset) à : <https://magenta.tensorflow.org/datasets/groove#license>
écouter le dataset groove

4.2 Analyse MIDI-Audio

Ces analyses ont été faites essentiellement dans le cadre de transcription manuelles à partir de fichiers MIDI et Audio du Groove MIDI Data Set. Les partitions manuelles sont éditées à l'aide de lilypond¹. Les transcriptions automatiques sont générées avec MuseScore².

1. <http://lilypond.org/>

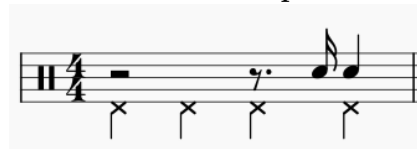
2. <https://musescore.com/>

4.2.1 Comparaisons de transcriptions

Transcription manuelle VS transcription automatique :

drummer_01/session3 — 10_rock-folk_90_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



drum-

mer_01/session3 — 10_rock-folk_90_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- Erreur d'indication de mesure ;
- Mauvaise transcription d'une noire.

La noire du 4ème temps se retrouve sur le premier temps de la mesure suivante et elle se transforme en un triolet de double croches dont seules les deux premières seraient jouées.

drummer_01/session3 — 10_rock-folk_90_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- Erreur de quantification : les doubles croches ont été interprétées en quintolet ;

drummer_01/session3 — 2_jazz-swing_185_beat_4-4

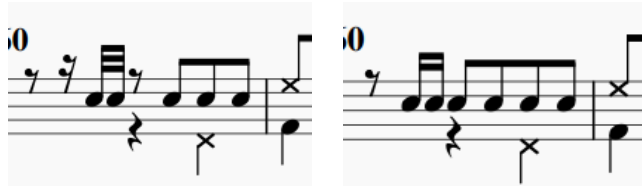
Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- L'indication de mesure est correcte mais tout a été décalé d'un temps car la première noire sur la caisse claire est jouée sur le 4ème temps et non sur le premier temps de la deuxième mesure comme l'indique la transcription de musescore.
- Les toms basses des 1er et 2ème temps de la mesure musescore auraient dû être sur les temps et non décalés d'une double croche vers la droite.

drummer_01/session1 — 1_funk_80_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- On dirait que lorsque certaines notes sont proches, elles se resserrent et suppriment celles qui aurait dû être sur le temps.
- Erreur d'indication de mesure ;
- Mauvaise transcription d'une noire.

La noire du 4ème temps se retrouve sur le premier temps de la mesure suivante et elle se transforme en un triolet de double croches dont seules les deux premières seraient jouées.

drummer_01/session3 — 10_rock-folk_90_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- Erreur de quantification : les doubles croches ont été interprétées en quintolet ;

drummer_01/session3 — 2_jazz-swing_185_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- L'indication de mesure est correcte mais tout a été décalé d'un temps car la première noire sur la caisse claire est jouée sur le 4ème temps et non sur le premier temps de la deuxième mesure comme l'indique la transcription de musescore.
- Les toms basses des 1er et 2ème temps de la mesure musescore auraient dû être sur les temps et non décalés d'une double croche vers la droite.

drummer_01/session1 — 1_funk_80_beat_4-4

Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- On dirait que lorsque certaines notes sont proches, elles se resserrent et suppriment celles qui aurait dû être sur le temps.

Exemple avec des flas

Fichier midi vers partition avec musescore :



Transcription manuelle :



MuseScore donne un aperçu de l'état de l'art pour la transcription de la

batterie.

4.2.2 Transcription de partition

Music engraving by LilyPond 2.22.1—www.lilypond.org

FIGURE 4.1 – référence 1

Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est annoncé dans le groove-dataset 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 de cette partition a occasionné plusieurs remarque :
- vélocité, place des accents, etc. . .

4.3 Expérimentation théorique d'un système

4.3.1 Motifs

Les motifs 1 et 2 peuvent être extraits de la figure 3.8. Ces deux motifs sont très classiques et seront réutilisables aussi dans d'autres contextes.

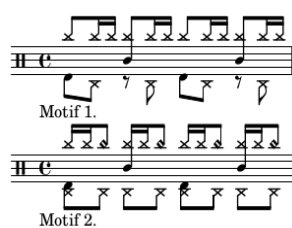


FIGURE 4.2 – Motifs

Le motif 1 est joué jusqu'à la mesure 18 avec des variations et des breaks.
Le motif 2 est joué des mesures 23 à 28.

4.3.2 Gammes

Débit croches.



Débit doubles-croches.



4.3.3 Systèmes

The image displays musical notation for two motifs, each with three text variants. Each variant consists of a melody line (treble clef, C major key signature) and a bass line (bass clef, C major key signature). The notation includes various rhythmic values and rests, with some notes marked with 'x' to indicate specific articulation or emphasis.

Motif 1:

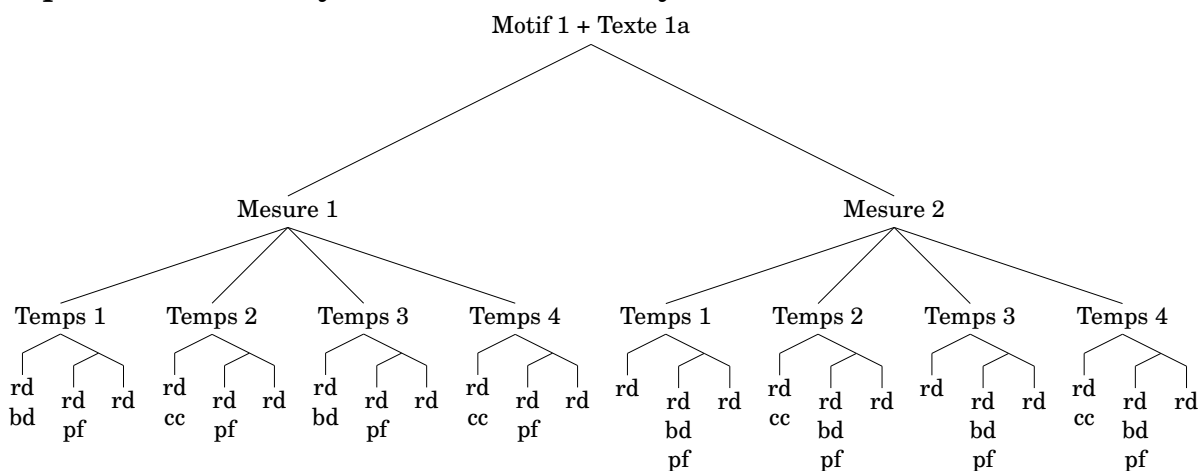
- Motif 1, texte 1a.
- Motif 1, texte 1b.
- Motif 1, texte 2a.
- Motif 1, texte 2b.
- Motif 1, texte 2c.

Motif 2:

- Motif 2, texte 1a.
- Motif 2, texte 1b.
- Motif 2, texte 2a.
- Motif 2, texte 2b.
- Motif 2, texte 2c.

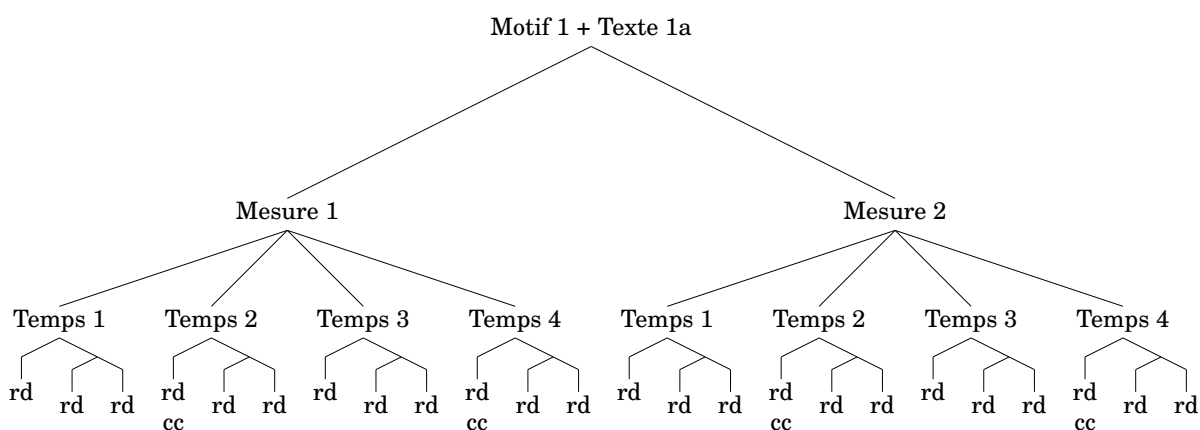
4.3.4 Démonstration

Représentation des systèmes en arbres de rythmes

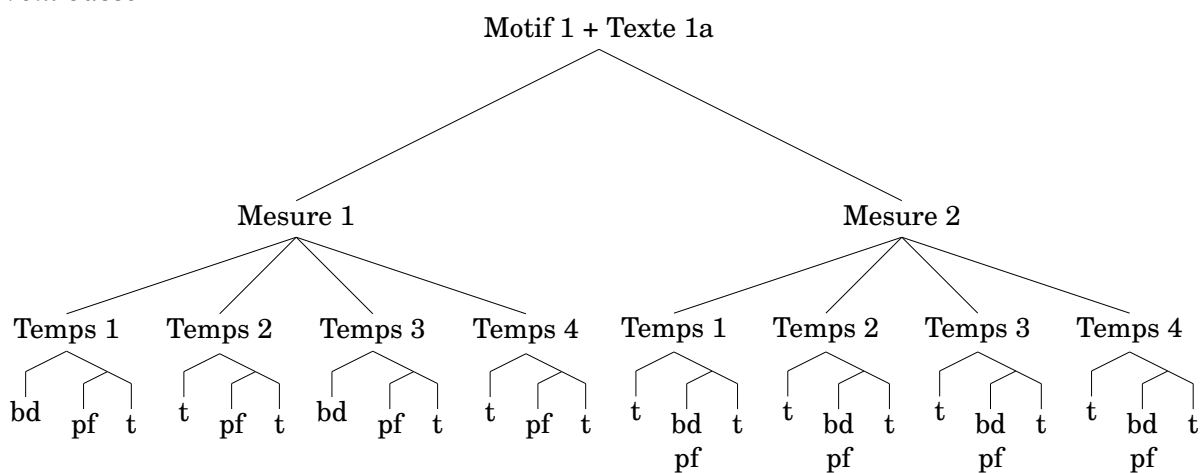


Réécriture Règles établies par le système Séparation des voix

Voix haute

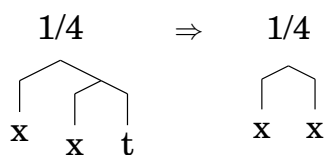
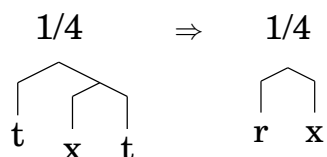
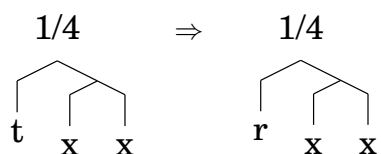
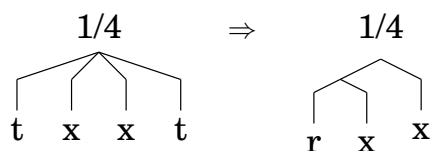
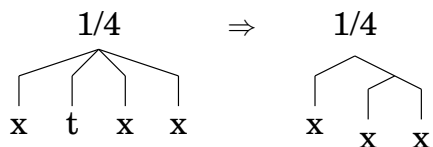
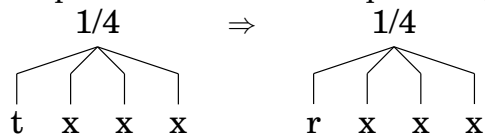


Voix basse



Règles de simplifications pour le 4/4 binaire

Simplifier l'écriture de chaque voix (*Règles établis par le système*)



4.4 Développement

4.4.1 DrumCode

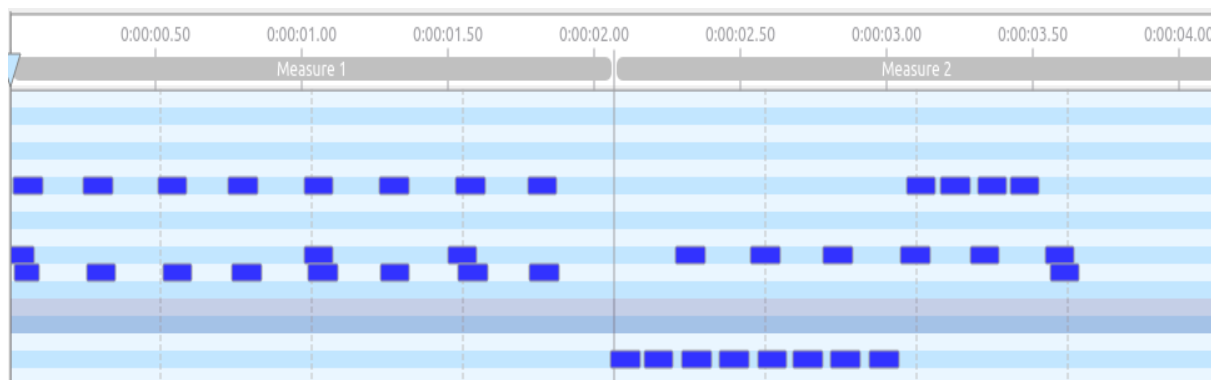
Adaptation de la modélisation pour la transcription en cpp.

4.4.2 Tests unitaires sur les Jams

Les Jams permettent de passer du monophonique au polyphonique.

4.4.3 Parsing

Tests effectués avec le fichier midi suivant :



Un premier test convaincant est effectué avec la grammaire suivante :

```
// bar level
0 -> C0 1
0 -> E1 1
0 -> U4(1, 1, 1, 1) 1
```

```
// half bar level
9 -> C0 1
9 -> E1 1
```

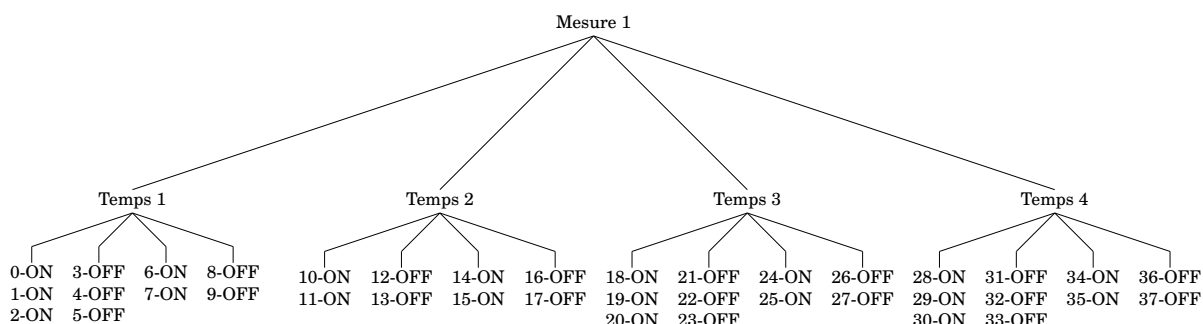
```
// beat level
1 -> C0 1
1 -> E1 1
1 -> T2(2, 2) 1
1 -> T4(4, 4, 4, 4) 1
```

```
// croche level
2 -> C0 1
2 -> E1 1
```

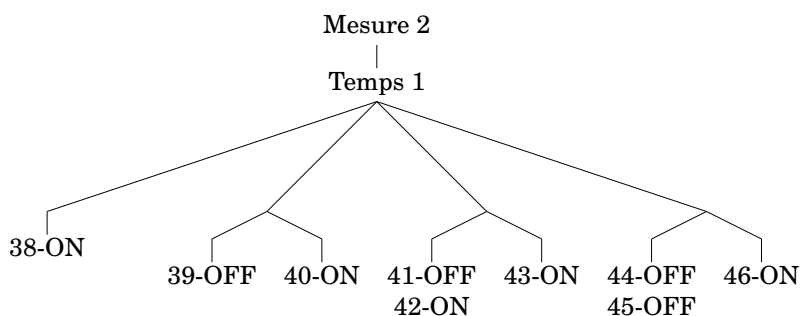
```
// double level
4 -> C0 1
4 -> E1 1
4 -> E2 1
4 -> T2(6, 6) 1
```

```
// triple level
6 -> E1 1
```

Cette grammaire sépare les ligatures par temps au niveau de la mesure. Puis, au niveau du temps, elle autorise les divisions par deux (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1. L'arbre de parsing en résultant est considéré comme « convaincant » car il découpe correctement les mesures et les temps.



Les temps de la première mesure du fichier MIDI sont bien quantifié mais ceux de la deuxième mesure présentent quelques défauts de quantification visibles dès le premier temps.



Les Onsets sont correctement triés au niveau des doubles croches mais certaines doubles croches sont inutilement subdivisées en triples croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps ci-dessus).

2ème tentative :

Après une augmentation du poids des triples croches dans la grammaire (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets sont bien triés et aucuns ne sont subdivisés.

Conclusion

Conclusion de ce chapitre.

DISCUSSION

Sommaire

5.1	Travaux réalisés	49
5.1.1	Développer la notation	49
5.1.2	La modélisation	49
5.1.3	Le jeu de système	49
5.2	Travaux non-réalisés	49
5.3	Travaux futures	50

Introduction

Dans ce chapitre, nous discuterons sur la pertinence de l'ensemble des choix qui ont été faits. Nous ferons un bilan des différentes avancées qui ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.

5.1 Travaux réalisés

Faire une auto-critique des travaux réalisés.

5.1.1 Développer la notation

5.1.2 La modélisation

5.1.3 Le jeu de système

5.2 Travaux non-réalisés

Expliquer pourquoi ces travaux n'ont pas pu être réalisés.

- implémenter un pattern. . .
⇒ manque de temps ?

- La partie résultat est manquante car :
 - ⇒ Sujet très difficile ;
 - ⇒ Matcher les motifs peut être fait ultérieurement ;
Mais ce travail 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)

5.3 Travaux futures

- Le ternaire jazz (voir expérience 2)
- Reconnaissance d'un motif sur le MIDI
Reconnaître un motif (système) sur une mesure de l'input (un fichier midi représentant des données audios)
 - ⇒ Motif (système) reconnu : true ou false
 - Si true :
 - Choisir la grammaire correspondante ;
 - Parser le MIDI ;
 - Appliquer les règles de réécritures (Séparation des voix et simplification)

Conclusion

Sujet passionnant mais difficile. Obtenir la totalité des critères pour le mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique à la batterie a 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 événements 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 ¹ 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.[10]

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

BIBLIOGRAPHIE

- [1] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 9 et 11.
- [2] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité page 10.
- [3] Caroline Traube. Quelle place pour la science au sein de la musicologie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 10.
- [4] 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 10.
- [5] 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 10.
- [6] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering music relations with sequential attention. In *NLP4MUSA*, 2020. – Cité page 11.
- [7] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri. Automatic music transcription : Challenges and future directions. *Journal of Intelligent Information Systems*, 41, 12 2013. – Cité pages 11, 12, 13, 14 et 18.
- [8] Moshekwa Malatji. Automatic music transcription for two instruments based variable q-transform and deep learning methods, 10 2020. – Cité page 13.
- [9] 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 13 et 18.

- [10] 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 14 et 51.
- [11] 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 14.
- [12] 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 page 18.
- [13] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. – Cité page 18.
- [14] Martin Rohrmeier. Towards a formalisation of musical rhythm. In *Proceedings of the 21st Int. Society for Music Information Retrieval Conf*, 2020. – Cité page 18.
- [15] 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 21.
- [16] 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)*, Coruña, Spain, May 2017. Helena Lopez Palma and Mike Solomon. – Cité page 21.