
**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 Introduction	9
1.2 Musique, TAL et RIM	9
1.3 La transcription automatique	10
1.4 Conclusion	12
2 État de l'art	13
2.1 Introduction	13
2.2 Audio, MIDI et partition	13
2.3 Approche hiérarchique VS approches linéaires	13
2.4 Du monophonique vers le polyphonique	13
2.5 Conclusion	13
3 Méthodes et contributions	15
3.1 Introduction	15
3.2 Qparse	16
3.3 Représentations, systèmes et réécriture	16
3.4 Conclusion	27
4 Expérimentations et résultats	29
4.1 Introduction	29
4.2 Corpus et expérimentations	29
4.3 Résultats et évaluation	49
4.4 Conclusion	49
5 Discussion	51
5.1 Introduction	51
5.2 Contenu	51
5.3 Conclusion	51
Conclusion générale	53

Bibliographie

55

LISTE DES FIGURES

LISTE DES TABLEAUX

3.1	Pitchs et instruments	20
3.2	Vélocité et nuances	20

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 d'amélioration de Qparse, un outil de transcription et d'écriture 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. 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.

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'accord et la mélodie du thème de 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 car l'autonomie musiciens pour écouter le résultats de leur compositions sans nécessairement avoir les musiciens pour jouer leur a permis d'écrire ce qu'il voulait entendre pour leur futur batteur.

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 entraînent 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.

Nous proposons de rechercher des rythmes génériques (*motifs*) en amont dans la chaîne de traitement. Les *motifs* sont prédéfinis avec des combinaisons possibles (*gammes*) qui leur sont associées. Ces *motifs* et leur *gammes* respectives sont appelés *systèmes*. L'usage des *systèmes* a pour objectif de fixer des choix le plus tôt possible dans la chaîne de traitement afin de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces choix concernent notamment la métrique, la séparation des voix ainsi

que les règles de réécriture.

Nous dresserons dans une première partie, 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 présente proposition. Dans une seconde partie, Le corpus sera présenté 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	Introduction	9
1.2	Musique, TAL et RIM	9
1.3	La transcription automatique	10
1.4	Conclusion	12

1.1 Introduction

Dans ce chapitre, nous présenterons le rapport possible entre la musique et le TAL, en considérant les notions de langage musical et langue naturelle, le lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole.

Nous aborderons aussi les applications des techniques TAL pour les différents traitements associés à la musique. Dans ce chapitre, nous parlerons de l'implication du TAL dans la musique et nous ferons un tour d'horizon des domaines de la recherche d'information musicale.

Nous présenterons ensuite le problème de la transcription automatique de la musique.

1.2 Musique, TAL et RIM

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. [1], 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. [2] 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 musicale apparaît, selon eux, à la fois comme une limite et un défi majeur pour l'étude de cette analogie.

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. [3] 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.

1.3 La transcription automatique

L'objectif de la transcription automatique de la musique (AMT) [4] 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. Bien que l'AMT soit un domaine de recherche en plein essor dans lequel plusieurs approches différentes sont encore activement étudiées, les performances des systèmes actuels ne sont pas encore suffisantes pour certaines applications qui exigent un haut degré de précision [4]. 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 [5]. Seuls quelques travaux récents [6] s'intéressent de près à la création d'outils permettant la génération de partition.

Définition

Le terme « transcription musicale automatique » a été utilisé pour la première fois par les chercheurs en audio James A. Moorer, Martin Piszczalski et Bernard Galler en 1977. Grâce à leurs connaissances en ingénierie audio numérique, ces chercheurs pensaient qu'un ordinateur pouvait être programmé pour analyser un enregistrement numérique de musique de manière à détecter les hauteurs des lignes mélodiques et des motifs d'accords, ainsi que les accents rythmiques des instruments à percussion.

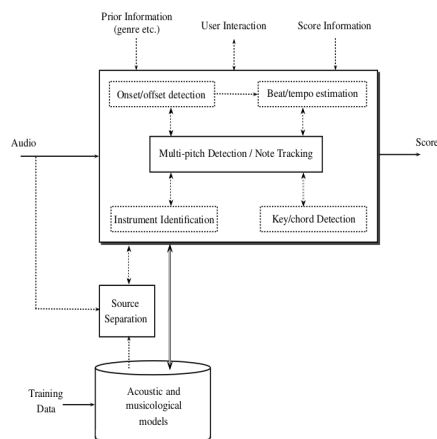
La tâche de transcription automatique de la musique comprend deux activités distinctes : l'analyse d'un morceau de musique et l'impression

d'une partition à partir de cette analyse.

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

Architecture

La figure suivante, qui est une proposition de Benetos et Al. [4], représente l'architecture générale d'un système de transcription musicale.



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.

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) [4]. 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. [7] 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 [7]. 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 [8], 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.

1.4 Conclusion

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

ÉTAT DE L'ART

Sommaire

2.1	Introduction	13
2.2	Audio, MIDI et partition	13
2.3	Approche hiérarchique VS approches linéaires	13
2.4	Du monophonique vers le polyphonique	13
2.5	Conclusion	13

2.1 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.

2.2 Audio, MIDI et partition

2.3 Approche hiérarchique VS approches linéaires

2.4 Du monophonique vers le polyphonique

2.5 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 ET CONTRIBUTIONS

Sommaire

3.1	Introduction	15
3.2	Qparse	16
3.3	Représentations, systèmes et réécriture	16
3.4	Conclusion	27

3.1 Introduction

Dans ce chapitre, nous expliquerons en détails les méthodes employées pour adapter à la batterie le processus décrit dans le chapitre précédent.

3.2 Qparse

En entrée : midi (séquence d'événements datés (piano roll) accompagné d'une grammaire pondérée)

⇒ parsing

⇒ global parsing tree

⇒ RI (Représentation Intermédiaire) arbres locaux par instruments

⇒ Sortie (xml, mei, lilypond, ...)

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

3.3 Représentations, systèmes et réécriture

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.)

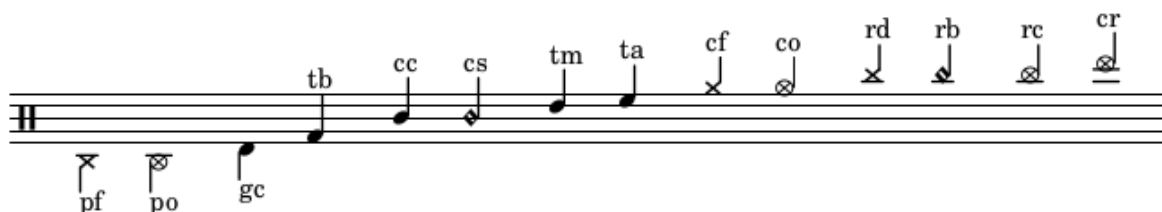
source : https://fr.wikipedia.org/wiki/Note_de_musique

Hauteurs et têtes de notes pour la batterie

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).



partition 1

Les nuances



partition 2

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

Les durées

Basé sur [9] et sur [10]

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

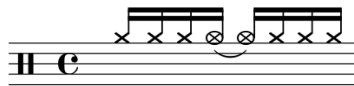
Les cymbales-crash et les ouvertures de charley constituent les seuls cas qui excluent cette option. Le charley car ses ouvertures/fermetures sont presque toujours quantifiées et les cymbales-crash car elles peuvent être arrêtées à la main de manière quantifié aussi mais ce cas est très rare, nous allons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

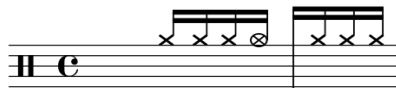
1. Certains logiciels le permettent mais leur lecture reste peu aisée

Exemples :

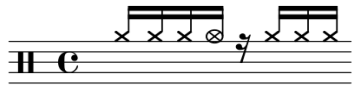
A1 — Évènement MIDI.



A2 — Réécriture.



B1 — Évènement MIDI.



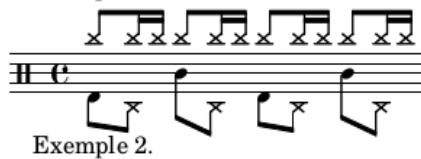
B2 — Réécriture.

La séparation des voix

Plusieurs écritures sont possibles pour un même rythme.



Exemple 1.



Exemple 2.



Exemple 3.

partition 3

Sur la partition 3, 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.

La représentation numérique de la batterie

Représentation symbolique (MIDI)

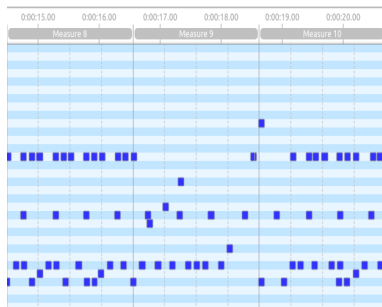
MIDI (acronyme de Musical Instrument Digital Interface) est une norme technique qui décrit un protocole de communication, une interface numérique et des connecteurs électriques permettant de connecter une grande variété d'instruments de musique électroniques, d'ordinateurs et d'appareils audio connexes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

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

Les données MIDI

Voir : <https://www.midi.org/>

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



Les évènements MIDI rassemble un ensemble d'informations sur la hauteur, la durée, le volume :

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

Pour la batterie, nous ne considérerons pas les offsets (« Off Event »).

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

Les pitches

Pas de charley pied ouvert. . .

La vélocité

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

TABLE 3.2 – Vélocité et nuances

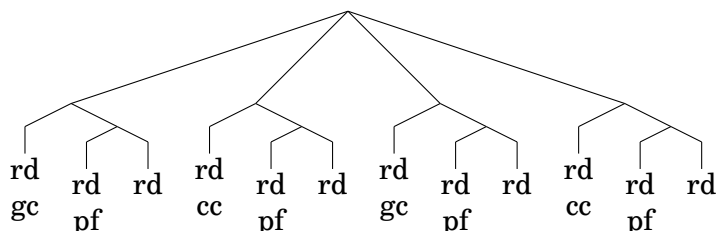
Nous ne prendrons en compte la vélocité que pour la cc, les toms et les cymbales jouées aux mains. Les nuances de grosse caisse et charley aux pieds sont le plus souvent insignifiantes, elles ne sont marquées sur le figure qu'à titre indicatif. Si la vélocité est en dessous de 40, il s'agit de ghost-notes : 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 vélocité va de 40 à 89, on considèrera le volume comme normal et aucun symbole supplémentaire ne sera ajouté à la note.

Les dilemmes

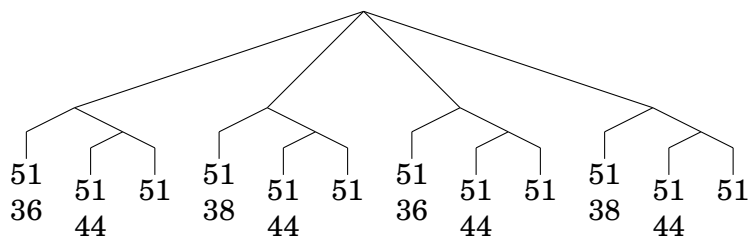
Le charley de pitch 46 est considéré comme le charley ouvert joué à la main sur le haut de la cymbale mais souvent, ça correspond au geste « tranche-olive » de la baguette lorsque le batteur accentue avec la tranche et joue moins fort avec l'olive sur le plat de la cymbale. Je vais dans un premier temps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

Représentations en arbres

Voici une représentation de la *partition 3* 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. (Voir *partition 3*).

Les systèmes

Définition

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.

Utilité

- Séparation des voix
- Définir une métrique
- Conditionner des règles spécifiques de réécriture

Créer un ensemble de systèmes :

- 4/4 binaire FAIT
- jazz vs ternaire(12/8) EN COURS...
- afro-cubain
- 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**.

Pour la séparation des voix

Motif 4-4 binaire

The notation shows a 4/4 measure divided into four eighth notes. The first line shows the full motif with cymbal symbols (ride, crash, cymbal, crash) above the notes. The second line, labeled 'Irréductible.', shows the same motif with the cymbal symbols removed. The third line, labeled 'Séparation possible 1.', shows the motif with the first two notes grouped together and the last two notes grouped together. The fourth line, labeled 'Séparation possible 2.', shows the motif with the first note and the last note grouped together, and the second and third notes grouped together. The fifth line, labeled 'Séparation possible 3.', shows the motif with the first note and the third note grouped together, and the second and fourth notes grouped together.

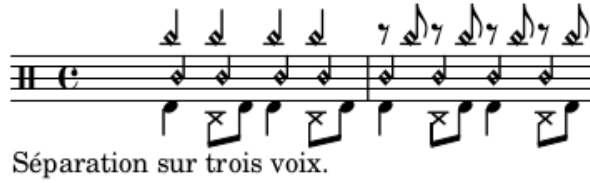
Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système est constitué par défaut du motif ride/ch-pf/cc et d'un texte joué à la grosse-caisse. La troisième séparation proposée est privilégiée car elle répartit 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

The notation shows a 4/4 measure divided into four eighth notes. The first line shows the motif with triplets of eighth notes above the notes. The second line, labeled 'Exemple.', shows the motif with the first two notes grouped together and the last two notes grouped together. The third line, labeled 'Séparation.', shows the motif with the first note and the third note grouped together, and the second and fourth notes grouped together.

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.

La réécriture des événements MIDI pour la batterie

Exemples à écrire en arbre :

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

Les règles de réécriture

$$\begin{array}{c} \frac{2}{8} \\ \diagup \quad \diagdown \\ \text{x} \quad \text{tie} \end{array}$$

$$\begin{array}{c} \frac{2}{8} \\ | \\ \text{x} \end{array}$$

$$\begin{array}{c} \frac{1}{4} \\ \diagup \quad \diagdown \\ \text{x} \quad \text{tie} \end{array}$$

$$\begin{array}{c} \frac{1}{4} \\ \diagup \quad \diagdown \\ \text{x} \quad \text{r} \end{array}$$

3.4 Conclusion

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

EXPÉRIMENTATIONS ET RÉSULTATS

Sommaire

4.1	Introduction	29
4.2	Corpus et expérimentations	29
4.3	Résultats et évaluation	49
4.4	Conclusion	49

4.1 Introduction

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

4.2 Corpus et expérimentations

Le corpus

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

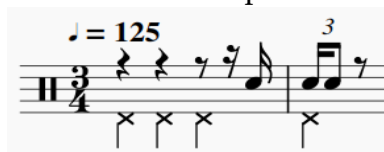
Détails (entre autres tensorflow avec le dataset) à : <https://magenta.tensorflow.org/datasets/groove#license>
écouter le dataset groove

Les expérimentations

Comparaisons de transcriptions

drummer_01/session3 — 10_rock-folk_90_beat_4-4

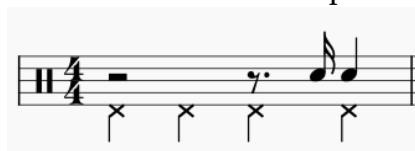
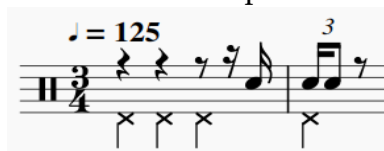
Fichier midi vers partition avec musescore ⇒ Transcription manuelle



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

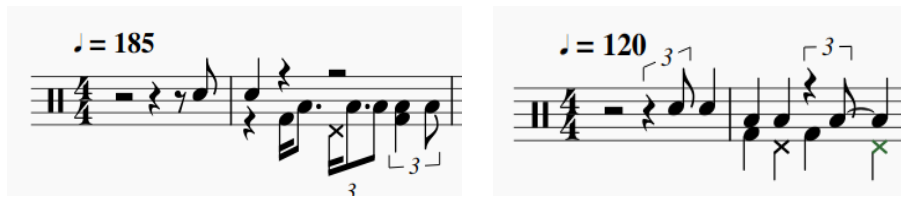
Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- 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

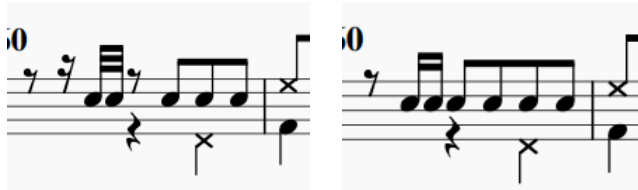
Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- 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

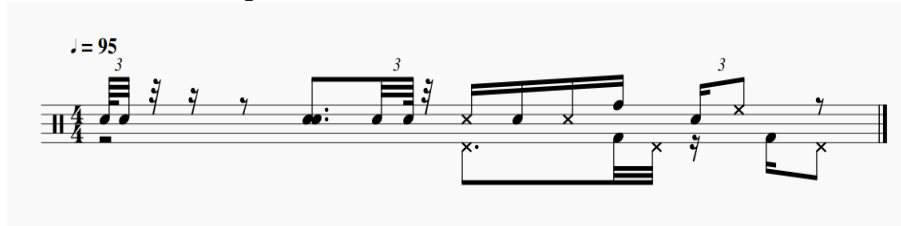
Fichier midi vers partition avec musescore ⇒ Transcription manuelle



- 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 :

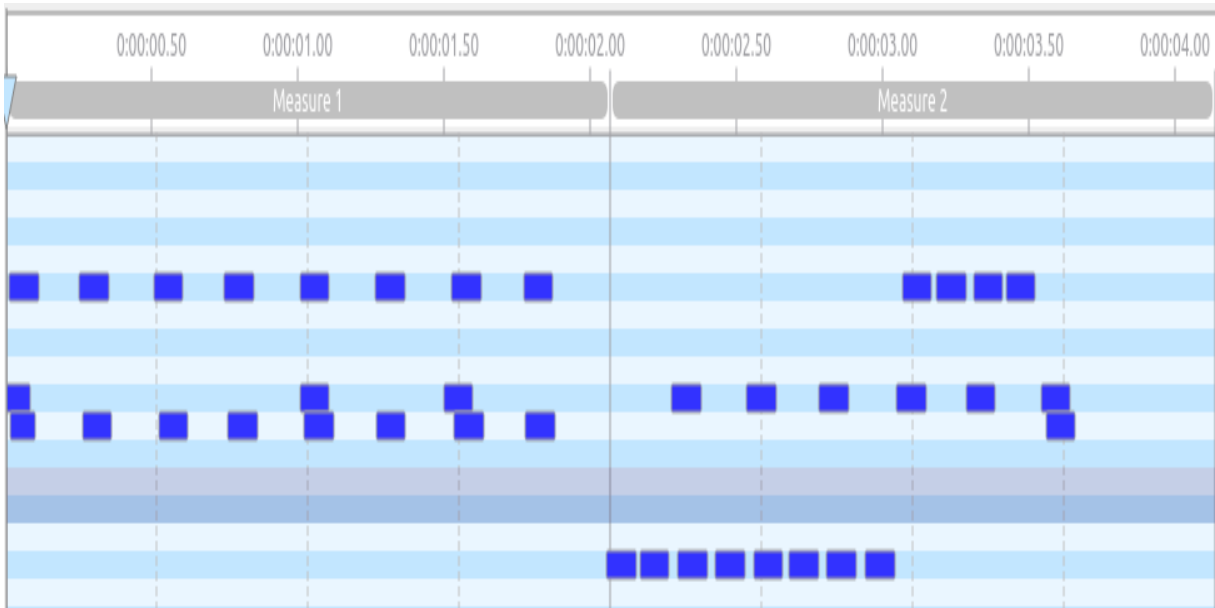


Transcription manuelle :



Les tests unitaires

Le parsing avec squant



Détails des expérimentations de bout en bout

Expérience 1 - 4/4 binaire
Partition de référence pour l'ouput

Music engraving by LilyPond 2.22.1—www.lilypond.org

Systèmes recherchés Textes :

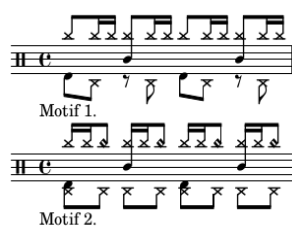
Débit croches.



Débit doubles-croches.



Motifs :



Systèmes résultants :

Motif 1, texte 1a.

Motif 1, texte 1b.

Motif 1, texte 2a.

Motif 1, texte 2b.

Motif 1, texte 2c.

Motif 2, texte 1a.

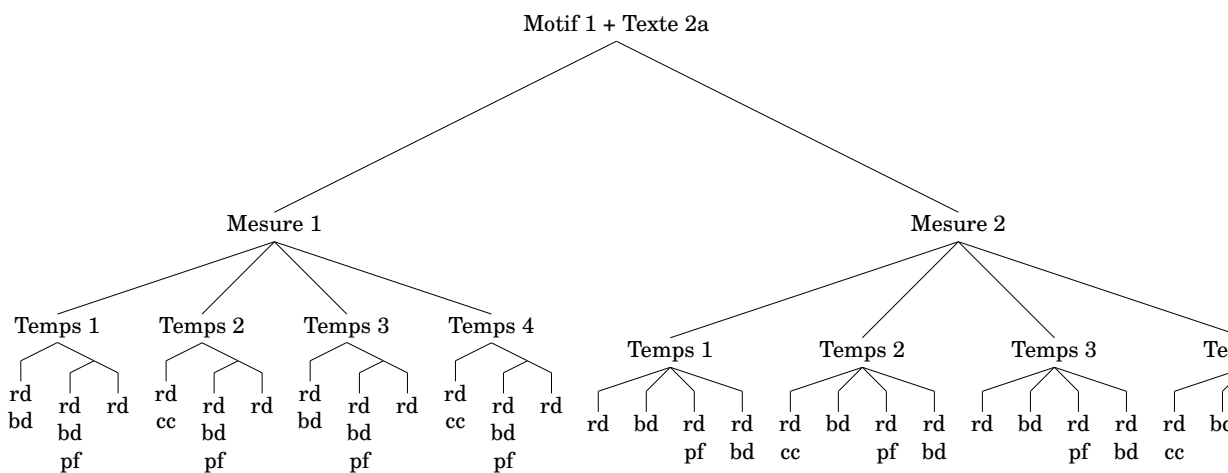
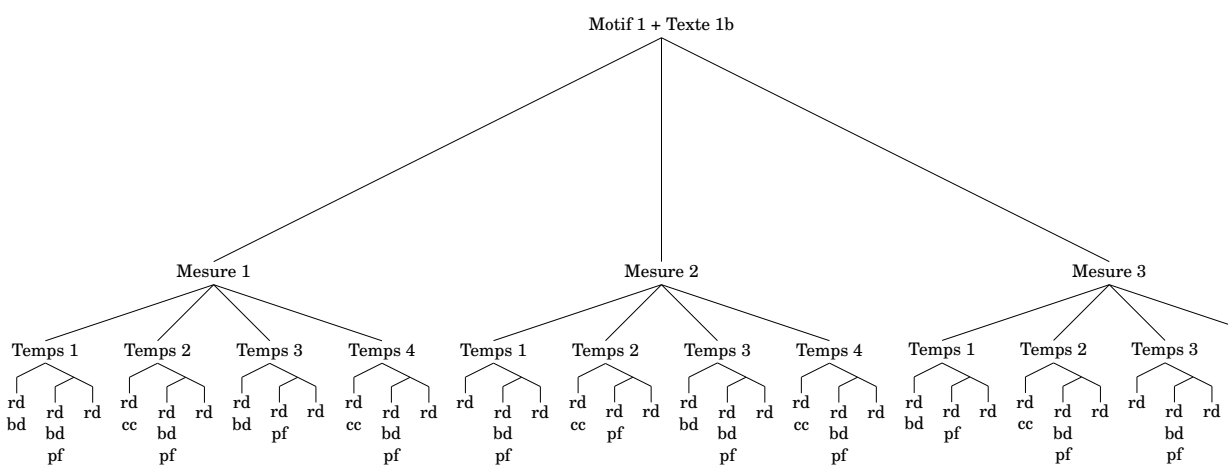
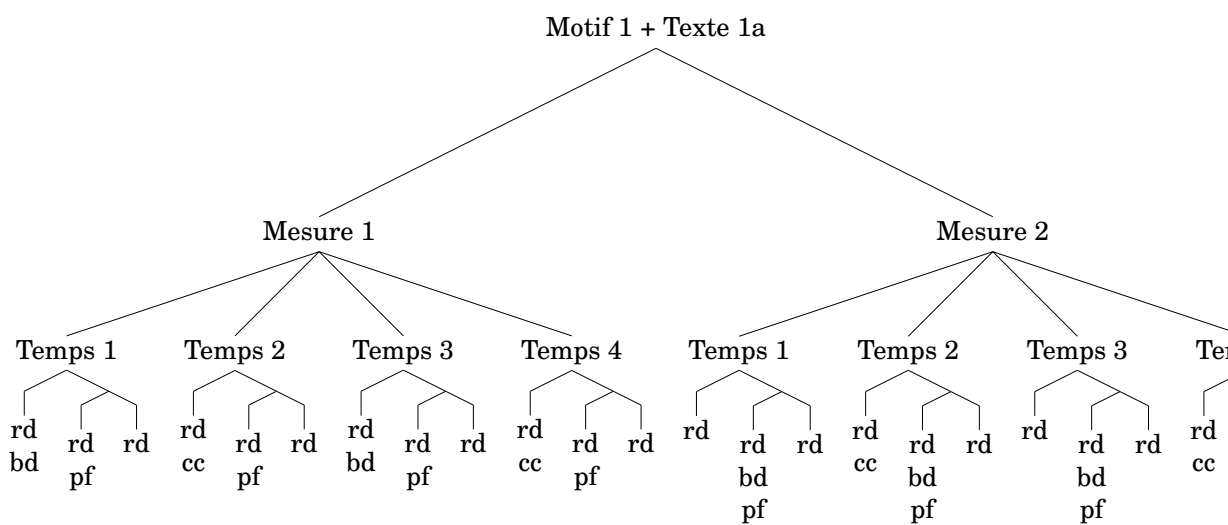
Motif 2, texte 1b.

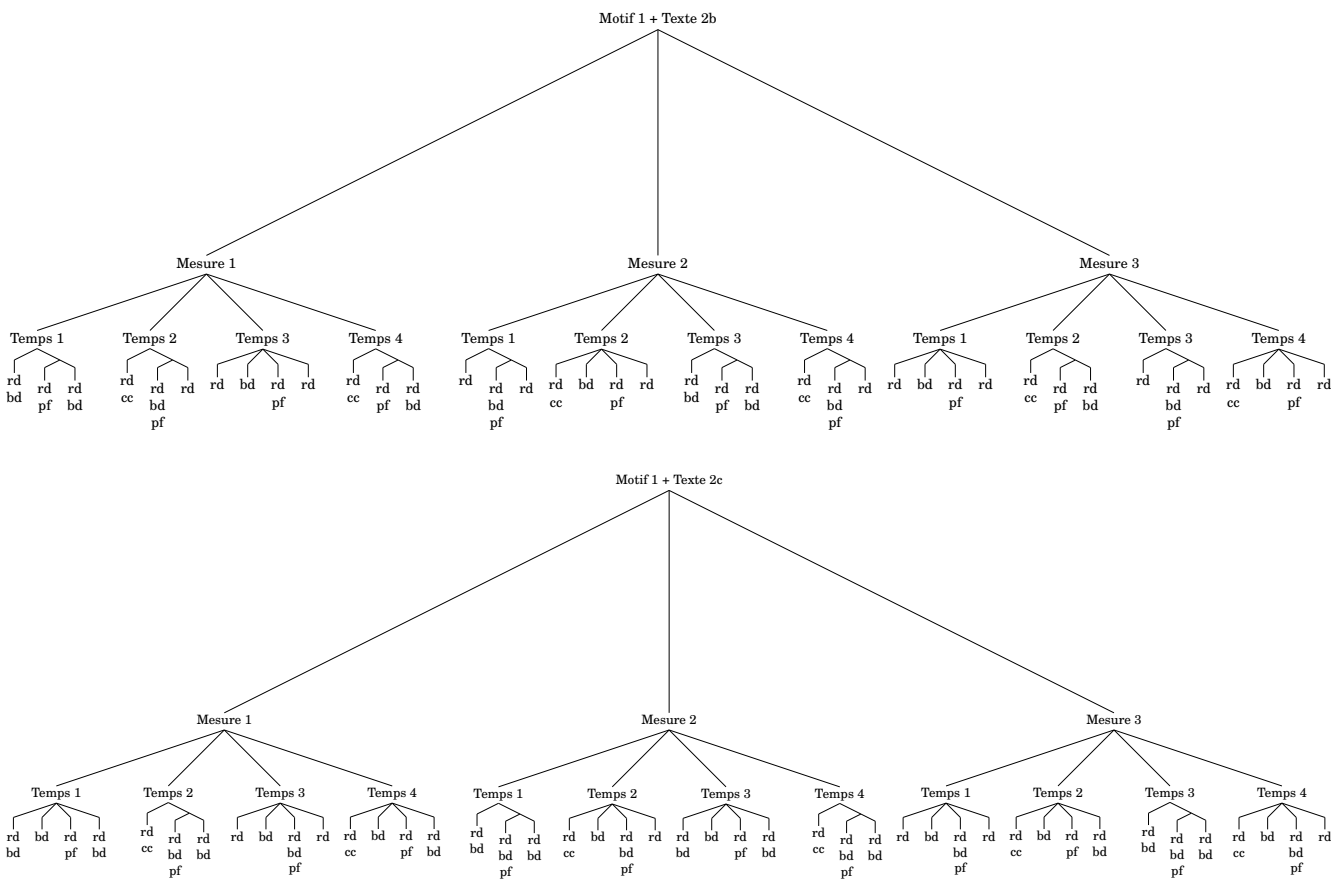
Motif 2, texte 2a.

Motif 2, texte 2b.

Motif 2, texte 2c.

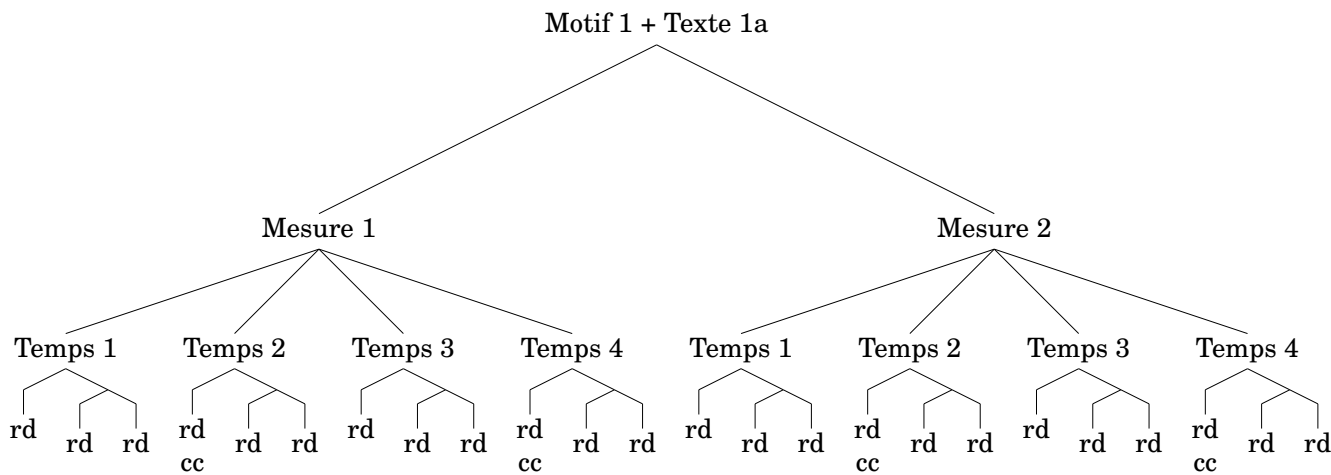
Représentation des systèmes en arbres de rythmes



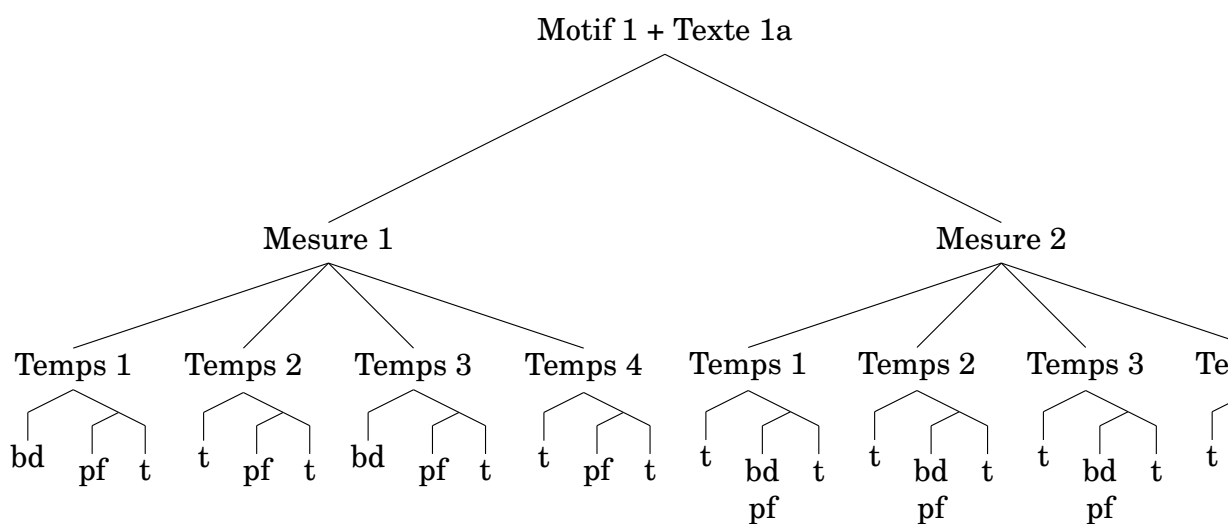


Séparation des voix Motif 1 + Texte 1a

Voix haute

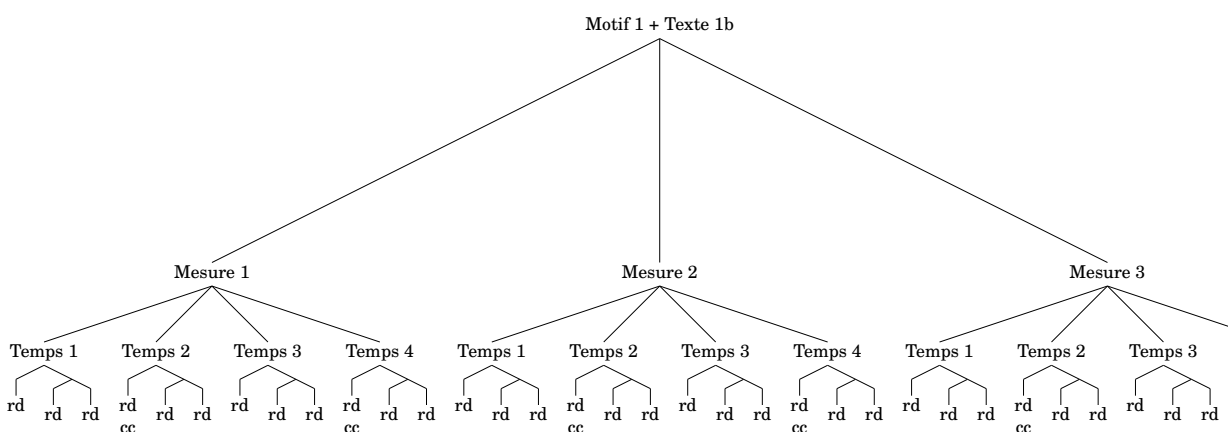


Voix basse

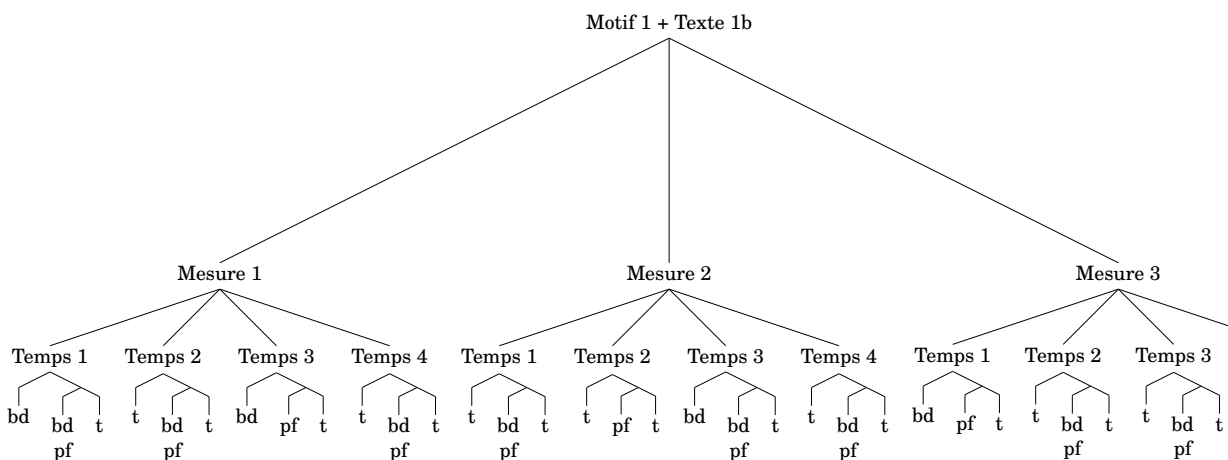


Motif 1 + Texte 1b

Voix haute

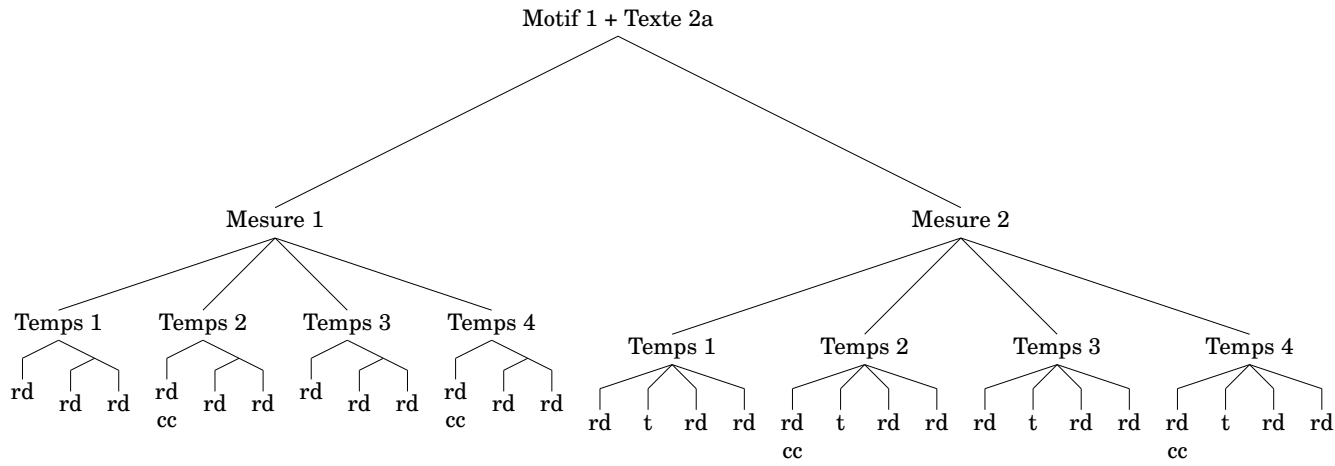


Voix basse

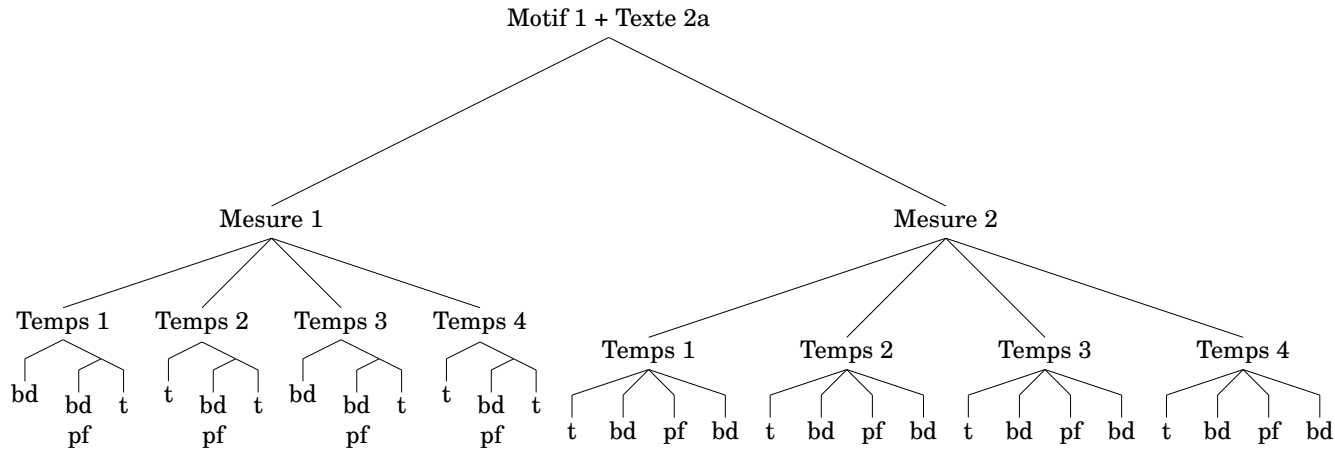


Motif 1 + Texte 2a

Voix haute

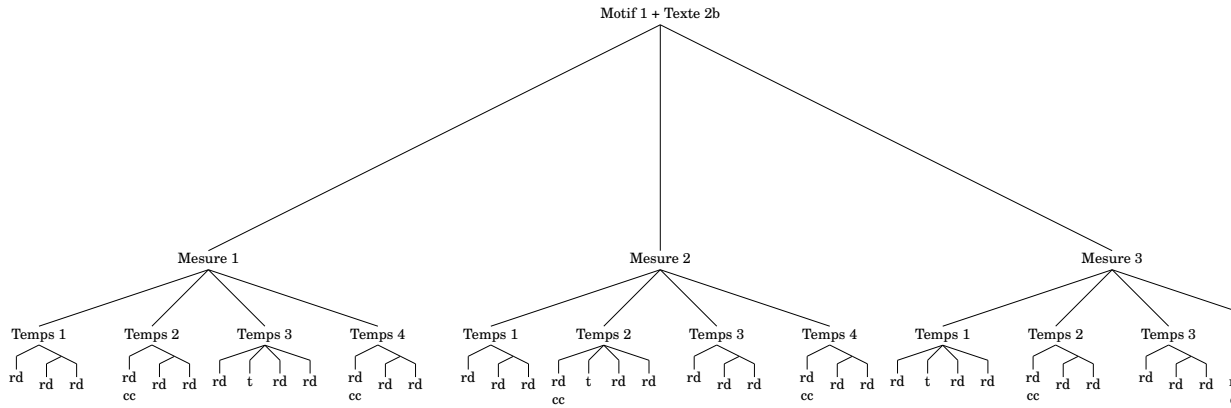


Voix basse

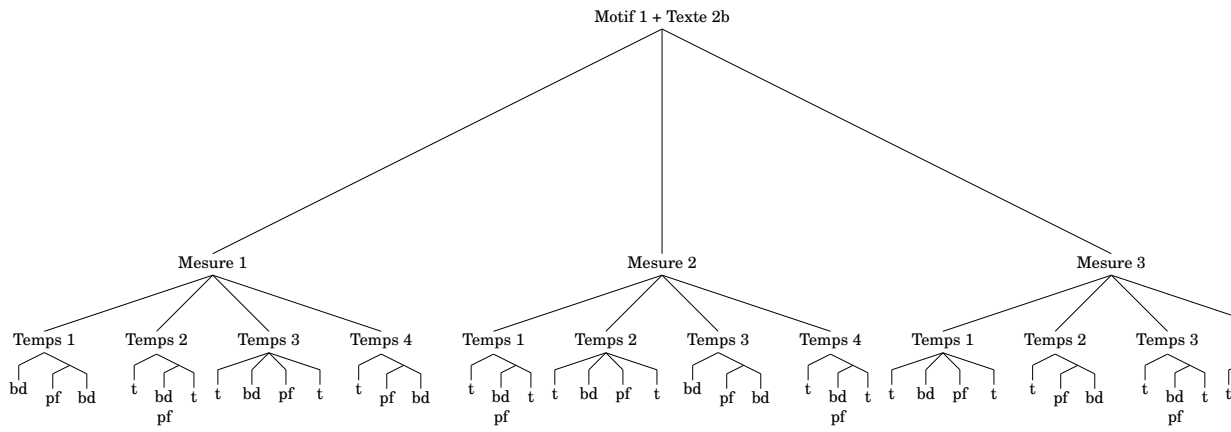


Motif 1 + Texte 2b

Voix haute

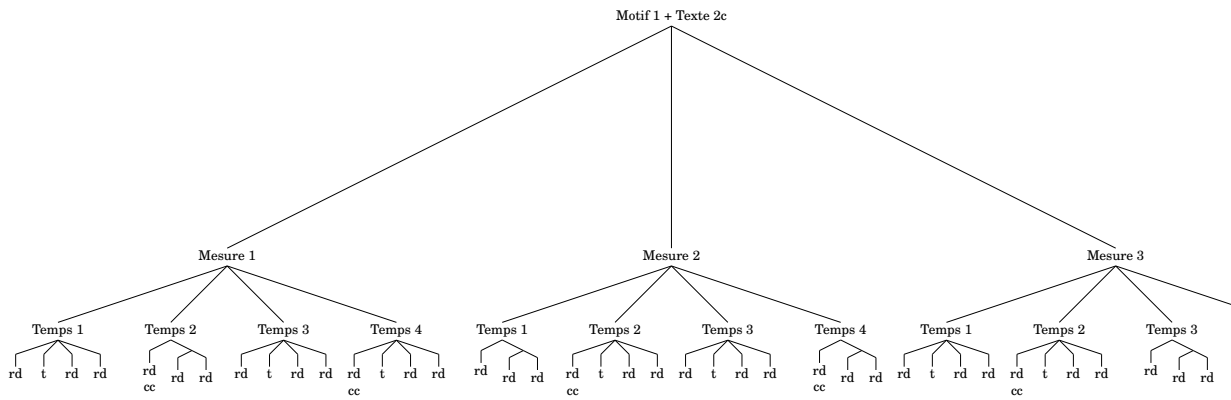


Voix basse

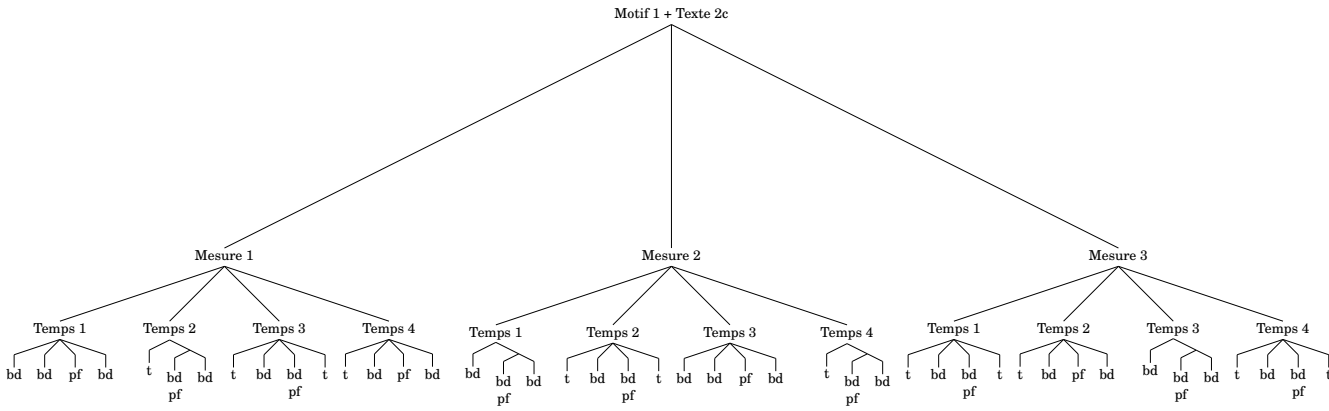


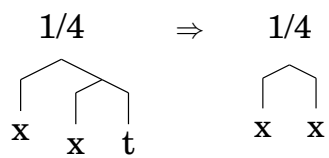
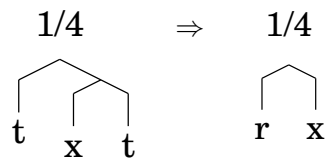
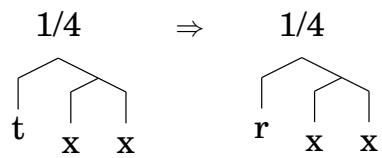
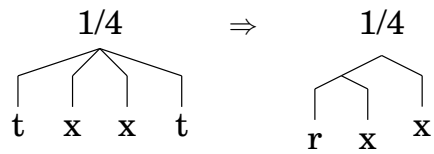
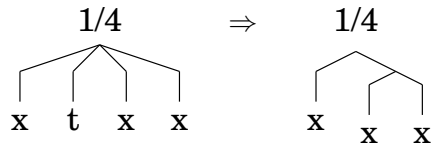
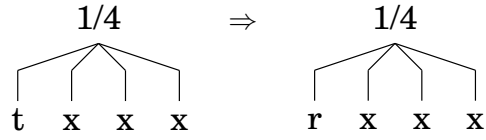
Motif 1 + Texte 2c

Voix haute



Voix basse



Règles de réécriture pour le 4/4 binaire

Expérience 2**Partition de référence pour l'output**

The image displays a musical score for three staves, enclosed in a rectangular frame. The top staff is a treble clef with a common time signature (C). It contains a sequence of notes with triplets and slurs. The middle staff is a bass clef with a common time signature (C). It contains a sequence of notes with triplets and slurs. The bottom staff is a bass clef with a common time signature (C). It contains a sequence of notes with triplets and slurs. The score is a reference partition for an output.

En cours...

squant : parsing du fichier midi

squant lit le midi

grammaire wta qui détermine le poids

La distance est automatiquement déterminée par squant

distance à l'input

complexité de la notation

On veut minimiser le coût et la distance \Rightarrow Trouver un compromis.

`./build/squant2 -h`

Essayer de lire un fichier midi avec squant2 lire mesure par mesure
Regarder les wta(grammaire)

Quelques tests de lecture midi

Les 4 messages suivants sont présents dans tous les tests qui suivent
[info] schema file: test/schema/schema-01.wta (??? weight model option)
[warning] no declaration MAX_GRACE in grammar file test/schema/schema-01.wta
[warning] no declaration TIMESIG in grammar file test/schema/schema-01.wta
[warning] MIDIfile has not joined tracks

`./build/squant2 -v 4 -a test/schema/schema-01.wta -m 004_jazz-funk_116_beat_4-4.mid -config ./params.ini`

[error] at least one of the options -bars or -barsec mandatory

`./build/squant2 -verbosity 4 -schema test/schema/schema-01.wta -midi 004_jazz-funk_116_beat_4-4.mid -config ./params.ini -barsec 3`

squant2: /home/martin/qparselib/src/schemata/SymbLabel.cpp:44: static void SymbLabel::Abandon (core dumped)

Tester squant2 avec les fichiers midi du corpus du gitlab

La commande suivante : `build/squant2 -v 5 -a ./test/schema/schema-03-R.wta -m ./corpus-master_qparselib/103-SaintSaens-elephant/perf/103_FJ.mid -config ./params.ini -mono -barsec 3.0 -ts 3/4`

Donne :

(1) $3(\bullet, \bar{2} : 2(\bullet,),)$

(2) $3(\bullet, \bar{2} : 2(\bullet, \bullet),)$

(3) $3(\bullet, \bar{2} : 2(\bullet, \circ),)$

Pour comprendre les grammaires :

Regarder les fichiers wta commentés.

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

A_Parse-based_Framework_for_Coupled_RhythmQuantization_and_Score_Structuring.pdf
Réfléchir au coût de notation (grace notes, etc.)

cluster.md**Contributions****Chaîne de traitement**

- 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 :
 - Séparer les voix (*Règles établis par le système*)
 - Simplifier l'écriture de chaque voix (*Règles établis par le système*)

Contribution sur la branch « distance » dans :

- qparselib/notes/cluster.md
- qparselib/src/segment/import/ :
DrumCode.hpp et.cpp

4.3 Résultats et évaluation

Références pour l'évaluation

1 - Transcription manuelle à partir de fichier midi et/ou wav d'une partition contenant des systèmes. Écriture des systèmes contenues dans la partition (arbres, séparation des voix, réécriture)

Comparaison d'arbres

Trouver un moyen de comparer l'arbre obtenu automatiquement de l'arbre de la transcription manuelle.

4.4 Conclusion

Conclusion de ce chapitre.

DISCUSSION

Sommaire

5.1	Introduction	51
5.2	Contenu	51
5.3	Conclusion	51

5.1 Introduction

Dans ce chapitre...

5.2 Contenu

Une section dans ce chapitre...

5.3 Conclusion

Conclusion de ce chapitre.

CONCLUSION GÉNÉRALE

Conclusion : la conclusion globale du mémoire.

Dans ce mémoire, nous avons traité de la problématique...

L'intégration de ces LM avec l'état de l'art des modèles acoustiques (AM) et des méthodes pour les tâches de traitement du signal ci-dessus. Cela nécessite la prise en compte du contexte musical et des informations musicales de haut niveau des LM en plus des caractéristiques acoustiques de bas niveau ci-dessus.

En outre, certaines expériences seront menées sur la base d'ensembles de données publiques, afin d'évaluer l'approche intégrée. Elles devraient couvrir certains cas de motifs rythmiques complexes se chevauchant.

Au-delà de l'intégration de modèles, 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. [7]

BIBLIOGRAPHIE

- [1] 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 9.
- [2] 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.
- [3] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering music relations with sequential attention. In *NLP4MUSA*, 2020. – Cité page 10.
- [4] 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 10, 11 et 12.
- [5] Moshekwa Malatji. Automatic music transcription for two instruments based variable q-transform and deep learning methods, 10 2020. – Cité page 10.
- [6] 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 page 10.
- [7] 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 12 et 53.
- [8] 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 12.
- [9] 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 17.
- [10] 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 17.