

2 **Institut National des Langues et Civilisations**
3 **Orientales**

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

6 **MASTER**
7 **TRAITEMENT AUTOMATIQUE DES LANGUES**

8 *Parcours :*
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

12 *Directeur de mémoire :*
13 *Damien NOUVEL*

14 *Encadrant :*
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021

TABLE DES MATIÈRES

18	Liste des figures	4
19	Liste des tableaux	5
20	Introduction générale	7
21	1 Contexte	11
22	1.1 Langues naturelles et musique en informatique	12
23	1.2 La transcription automatique de la musique	14
24	1.3 La transcription automatique de la batterie	15
25	1.4 Les représentations de la musique	16
26	2 État de l'art	21
27	2.1 Monophonique et polyphonique	21
28	2.2 Audio vers MIDI	22
29	2.3 MIDI vers partition	24
30	2.4 Approche linéaire et approche hiérarchique	24
31	3 Méthodes	29
32	3.1 La notation de la batterie	29
33	3.2 Modélisation pour la transcription	37
34	3.3 Qparse	38
35	3.4 Les systèmes	40
36	4 Expérimentations	47
37	4.1 Le jeu de données	47
38	4.2 Analyse MIDI-Audio	49
39	4.3 Transcription polyphonique par parsing (?verrou?)	53
40	4.4 Expérimentation théorique d'un système	55
41	4.5 BILAN : résultats — évaluation — discussion	60
42	Conclusion générale	63
43	Bibliographie	65

LISTE DES FIGURES

45	1.1	Exemple évènements avec durée	16
46	1.2	Critère pour un évènement	17
47	1.3	Exemple évènements sans durée	17
48	1.4	Exemple de partition de piano	18
49	1.5	MusicXML	18
50	2.1	Transcription automatique <dam>remettre ici la citation de la	
51		capture d'écran avec la page</dam>	23
52	2.2	HMM	26
53	2.3	arbre_jazz	27
54	3.1	29
55	3.2	Rapport des figures de notes	30
56	3.3	Les instruments de la batterie	31
57	3.4	Hauteur et têtes de notes	31
58	3.5	Point et liaison	32
59	3.6	Les silences	33
60	3.7	Silence joué	34
61	3.8	Équivalence	35
62	3.9	Séparation des voix	35
63	3.10	Les accents et les ghost-notes	36
64	3.11	Exemple pour les accentuations et les ghost-notes	36
65	3.12	Présentation de Qparse	39
66	3.13	Métrique	41
67	3.14	Motif 4-4 binaire	42
68	3.15	Motif 4-4 jazz	43
69	3.16	Système 4-4 afro-latin	43
70	3.17	Simplification	44
71	3.18	45
72	4.1	Batterie électronique	48
73	4.2	Partition de référence	52
74	4.3	Motifs et gammes	55
75	4.4	Partition d'un système en 4/4 binaire	56
76	4.5	Arbre de rythme — système	57
77	4.6	Arbre de rythme — voix haute	57
78	4.7	Arbre de rythme — voix basse	58
79	4.8	58
80	4.9	58

81	4.10	59
82	4.11	59
83	4.12	59

LISTE DES TABLEAUX

85	1.1 speechToText vs AMT	13
86	3.1 Pitches et instruments	37
87	3.2 Systèmes	41

INTRODUCTION GÉNÉRALE

89 QUOI?

90 Ce mémoire de recherche, effectué en parallèle d'un stage à l'Inria dans
 91 le cadre du master de traitement automatique des langues de l'Inalco,
 92 contient une proposition originale ainsi que diverses contributions dans
 93 le domaine de la transcription automatique de la musique. Les travaux
 94 qui seront exposés ont tous pour objectif d'améliorer **qparse**, un outil de
 95 transcription automatique de la musique, et seront axés spécifiquement
 96 sur le cas de la batterie.

97 Nous parlerons de transcription musicale, en suivant des méthodes
 98 communes au domaine du traitement automatique des langues (TAL)
 99 plutôt que directement de langues naturelles, et nous parlerons aussi de
 100 génération automatique de partitions de musique à partir de données au-
 101 dio ou symboliques. En considérant que la musique à l'instar des langues
 102 naturelles est un moyen qui nous sert à exprimer nos ressentis sur le
 103 monde et les choses, ce travail reposera sur une citation de l'ouvrage
 104 de Danhauser [1] : « La musique s'écrit et se lit aussi facilement qu'on
 105 lit et écrit les paroles que nous prononçons. » L'exercice exposé dans ce
 106 mémoire nécessitera donc la manipulation d'un langage musical qui peut
 107 être analysé à l'aide de théories formelles et d'outils adéquats comme
 108 des grammaires (solfège, durées, nuances, volumes) et soulèvera des
 109 problématiques qui peuvent être résolues par l'utilisation de méthodes
 110 issues de l'informatique et de l'analyse des langues et des langages.

111

112 POURQUOI?

- 113 — sujet traité : la batterie
- 114 — intérêt spécifique de la génération de partition de batterie compa-
 115 rativement au autres instrument
- 116 — patrimoine
- 117 — rapidité de génération (musicien ou enseignement)
- 118 — ...

119

120 <flo>il faut revoir la fin, avec une description rapide du problème, de la
 121 méthode suivie et des contributions suivi d'un petit plan par parties.</flo>

122 COMMENT?

123 → Problématique :

124 L'écriture musicale offre de nombreuses possibilités pour la transcription

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

→ Méthodes :

→ Contributions :

liste des contributions : donner une échelle, un point de comparaison, du contexte, pour pouvoir mesurer l'importance de chaque contribution

La proposition principale de ce mémoire est basée sur la recherche de rythmes génériques sur l'*input*. Ces rythmes sont des *patterns* standards de batterie définis au préalable et accompagnés par les différentes combinaisons qui leur sont propres. On les nomme systèmes (voir sections 3.4, 4.4). L'objectif des systèmes est de fixer des choix le plus tôt possible afin de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces choix concernent notamment la métrique et les règles de réécriture.

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

- une modélisation de la notation de la batterie (fusion de 3.1 et de 3.2) qui était jusqu'à présent inexistante.
- plusieurs transcriptions manuelles dans le but d'analyser les contenus des fichiers MIDI et Audio (4.2) et de faire des comparaisons de transcription avec des outils déjà existants¹.
- une partition de référence transcrite manuellement sur l'entièreté d'une performance du jeu de données afin de repérer les éléments importants pour la modélisation et de faire les liens entre les critères des données d'*input* avec l'écriture finale (4.2). Cette partition avait aussi pour objectif d'effectuer des tests et des évaluations.
- le passage au polyphonique en théorie et en implémentation impliquant la théorie sur la détection de l'identité de notes dans un Jam² et l'implémentation de tests unitaires sur le traitement des Jams (4.3).
- la création de grammaires pondérées spécifiques à la batterie (4.3)

1. MuseScore3

2. groupe de notes rassemblées en raison d'un faible écart entre leur emplacements temporels

165 L'ensemble de ces sous-tâches a permis deux réalisations principales :
166 1) Obtenir des arbres de rythmes corrects en *output* de qparse avec des
167 exemples courts proches de la partition de référence.
168 2) La création d'une expérimentation théorique d'un système 4.4 dont
169 le but premier est de démontrer qu'elle est implémentable et applicable
170 à d'autres type de rythmes et dont le second objectif est de donner une
171 méthode de création d'un système à partir d'une partition.
172 Ces deux réalisations recouvrent une partie du chemin à parcourir
173 puisque pour effectuer des évaluations conséquentes sur résultat, la
174 chaîne de traitement doit être finie afin de pouvoir vérifier de manière
175 empirique que les systèmes, qui constituent ma contribution principale
176 pour ce mémoire, ont permis d'améliorer qparse pour la transcription
177 automatique de la batterie.

178

179 PLAN

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

188

CONTEXTE

189

190

Sommaire

191	1.1	Langues naturelles et musique en informatique	12
192	1.2	La transcription automatique de la musique	14
193	1.3	La transcription automatique de la batterie	15
194	1.4	Les représentations de la musique	16

195

196

Introduction

199

200 La transcription automatique de la musique (TAM) est un défi ancien [2]
 201 et difficile qui n'est toujours pas résolu de manière satisfaisante par les
 202 systèmes actuels. Il a engendré une grande variété de sous-tâches qui
 203 ont donné naissance au domaine de la recherche d'information musicale
 204 (RIM)¹. Actuellement, en raison de la nature séquentielle et symbolique
 205 des données musicales et du fait que les travaux en TAL sont assez avan-
 206 cés en analyse de données séquentielles ainsi qu'en traitement du signal,
 207 de nombreux travaux de RIM font appel au TAL. Certains de ces tra-
 208 vaux se concentrent notamment sur l'analyse des paroles de chansons².
 209 <moi>Mais d'autres traitent directement la musique + ref.</moi>

210 Dans ce chapitre, nous parlerons de l'informatique musicale, nous mon-
 211 trerons les liens existants entre le RIM et le TAL ainsi qu'entre les no-
 212 tions de langage musical et langue naturelle. Nous traiterons également
 213 du problème de l'AMT et de ses applications.

214 Enfin, nous décrirons les représentations de la musique qui sont néces-
 215 saires à la compréhension du présent travail.

1. <https://ismir.net/>

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

1.1 Langues naturelles et musique en informatique

COMPUTER MUSIC

L'informatique musicale ou *Computer Music* regroupe l'ensemble des méthodes permettant de créer ou d'analyser des données musicales à l'aide d'outils informatiques [3]. Ce domaine implique l'utilisation de méthodes numériques pour l'analyse et la synthèse de musique³, qu'il s'agisse d'informations audio, ou symboliques (aide à l'écriture, transcription, base de partitions...). Un exemple de tâche dans ce domaine pourrait être l'analyse de la structure de la musique et de la reconnaissance des accords⁴.

RIM

La RIM est née du domaine de l'informatique musicale et apparaît vers le début des années 2000 [5]. L'objectif de cette science est la recherche et l'extraction d'informations à partir de données musicales. Il s'agit d'un vaste champ de recherche pluridisciplinaire, à l'intersection de acoustique, signal, synthèse sonore, informatique, sciences cognitives, neurosciences, musicologie, psycho-acoustique, etc. Cette discipline récente a notamment été soutenue par de grandes entreprises technologiques^{5 6 7} qui veulent développer des systèmes de recommandation de musique ou des moteurs de recherche dédiés au son et à la musique.

RIM et TAL

Aborder la musique comme un langage avec des méthodes de TAL nécessite une réflexion autour de la musique en tant que langage ainsi que la possibilité de comparer ce même langage avec les langues naturelles. Léonard Bernstein [6] a donné une série de six conférences publiques à Harvard fondées en grande partie sur les théories linguistiques que Noam Chomsky a exposées dans son livre « Language and Mind ». Lors de la première conférence, qui a eu lieu le 9 octobre 1973, Bernstein a avoué être hanté par la notion d'une grammaire musicale mondiale innée et il analyse dans ses trois premières conférences, la musique en termes linguistiques (phonologie, syntaxe et sémantique). Quelques travaux en neurosciences ont également abordé ces questions, notamment par observation des processus cognitifs et neuronaux que les systèmes de trai-

3. Voir la transformée de Fourier pour la musique dans [4]

4. En musique, un accord est un ensemble de notes considéré comme formant un tout du point de vue de l'harmonie. Le plus souvent, ces notes sont jouées simultanément; mais les accords peuvent aussi s'exprimer par des notes successive

5. <https://research.deezer.com/>

6. <https://magenta.tensorflow.org/>

7. <https://research.atspotify.com/>

tement de ces deux productions humaines avaient en commun. Dans le travail de Poulin-Charronnat *et al.* [7], la musique est reconnue comme étant un système complexe spécifique à l'être humain dont une des similitudes avec les langues naturelles est l'émergence de régularités reconnues 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.* [8] ont exploré le potentiel de ces techniques à travers les plongements de mots et le mécanisme d'attention pour la modélisation de données musicales. La question de la sémantique d'une phrase musicale apparaît, selon eux, à la fois comme une limite et un défi majeur pour l'étude de cette analogie. Ces considérations nous rapproche de la sémiologie de F. de Saussure en tant que science générale des signes et dont la langue ne serait qu'un cas particulier, caractérisé par l'arbitrariété totale de ses unités [9].

exemples / illustration de la proximité thématique?

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.* [10] ont présenté leur implémentation d'un modèle de langage musical visant à améliorer le mécanisme d'attention par élément, déjà très largement utilisé dans les modèles de séquence modernes pour le texte et la musique. Le domaine du TAL qui se rapproche le plus du RIM est la reconnaissance de la parole (Speech to text). En effet, la séparation des sources ont des approches similaires dans les deux domaines. De plus, il existe un lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole. La transcription musicale étant la notation d'une œuvre musicale initialement non écrite, l'analogie avec l'écriture de la parole est aisée. Le tableau 1.1 montre des différences et des similitudes entre les deux domaines.

Domaines	Similitudes	Différences
Speech to text AMT	signal \Rightarrow phonèmes \Rightarrow texte signal \Rightarrow notes, accords \Rightarrow partition	données linéaires données structurées

TABLE 1.1 – speechToText vs AMT

Non seulement les objectifs sont similaires, mais les problèmes et les applications, eux aussi, sont comparables (transcription, synthèse, séparation de sources, ...). Il faut néanmoins relever que les informations sont traitées sont de nature différente (voir *mettre ref vers sous-tâches comme beat tracking et inférence de tempo en musique*).

286 1.2 La transcription automatique de la musique

287 1. OBJECTIF

288 Lorsqu'un musicien est chargé de créer une partition à partir d'un
289 enregistrement et qu'il écrit les notes qui composent le morceau en
290 notation musicale, on dit qu'il a créé une transcription musicale de cet
291 enregistrement. L'objectif de la TAM [11] est de convertir la performance
292 d'un musicien en notation musicale — à l'instar de la conversion de la
293 parole en texte dans le traitement du langage naturel. Cette définition
294 peut être comprise de deux manières différentes selon les articles scien-
295 tifiques : 1) Processus de conversion d'un enregistrement audio en une
296 notation pianoroll (une représentation bidimensionnelle des notes de
297 musique dans le temps) 2) Processus de conversion d'un enregistrement
298 en notation musicale commune⁸ (c'est-à-dire une partition).

299

300 2. APPLICATIONS

301 La TAM a des applications multiples [11] dont la plus directe est de don-
302 ner la possibilité à un musicien de générer la partition d'une improvisa-
303 tion en temps réel afin de pouvoir reproduire sa performance ultérieure-
304 ment. Une autre application notable est la préservation du patrimoine
305 par exemple dans les styles musicaux où il n'existe peu de partitions (le
306 jazz, la pop, les musiques de tradition orale⁹, ...). La TAM est aussi utile
307 pour la recherche et l'annotation automatique d'informations musicales,
308 pour l'analyse musicologique¹⁰ ou encore pour les systèmes musicaux in-
309 teractifs.

310 Un grand nombre de fichiers audio et vidéos musicaux sont disponibles
311 sur le Web, et pour la plupart d'entre eux, il est difficile de trouver les
312 partitions musicales correspondantes, qui sont pourtant nécessaires pour
313 pratiquer la musique, faire des reprises ou effectuer une analyse musicale
314 détaillée.

315 Mais l'intérêt de la TAM est aussi d'avoir des partitions au contenu
316 exploitable, avec des formats texte ou XML (entre autres...) dont les
317 données sont manipulables, contrairement à de simples images en pdf¹¹.

318

319 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

320 L'analyse de la structure hiérarchique des séquences d'accords par utili-
321 sation de modèles grammaticaux s'est avérée très utiles dans les analyses
322 récentes de l'harmonie du jazz [12]. Comme déjà évoqué précédemment, il
323 s'agit d'un problème ancien et difficile. C'est un « graal » de l'informatique

8. Ici, on parle de notation occidentale.

9. ethno-musicologie

10. par exemple par la constitution de corpus musicologiques

11. Voir <https://archive.fosdem.org/2017/schedule/event/openscore/> et
0_slides-Martin.pdf.

musicale. En 1976, H. C. Longuet-Higgins [2] évoquait déjà la représentation musicale en arbre syntaxique dans le but d'écrire automatiquement des partitions à partir de données audio en se basant sur un mimétisme psychologique de l'approche humaine. La tâche de la TAM comprend deux activités distinctes : 1) l'analyse et la représentation d'un morceau de musique ; 2) La génération d'une partition à partir de la représentation du morceau.

1.3 La transcription automatique de la batterie

La batterie est née au début du vingtième siècle [13]. C'est donc 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'écoles de batterie partout dans le monde. Un autre facteur qui a contribué à l'expansion des partitions de batterie est l'émergence de la musique assistée par ordinateur (MAO). En effet, l'usage de boîtes à rythmes¹² ou de séquenceurs¹³ permettant d'expérimenter soi-même l'écriture de rythmes en les écoutant mixés avec d'autres instruments sur des machines a permis aux compositeurs de s'émanciper de la création d'un batteur en lui fournissant une partition contenant les parties exactes qu'ils voulaient entendre sur leur musique.

La batterie a un statut à part dans l'univers de l'AMT puisqu'il s'agit d'instruments sans hauteur (du point de vue harmonique), d'événements sonores auxquels une durée est rarement attribuée et de notations spécifiques (symboles des têtes de notes) [14].

Les applications de la transcription automatique de la batterie (TAB) seraient utiles, non seulement dans tous les domaines musicaux concernés par la batterie dont certains manquent de partitions, notamment les musiques d'improvisation [11], mais aussi de manière plus générale dans le domaine de la RIM : si les ordinateurs étaient capables d'analyser la partie de la batterie dans la musique enregistrée, cela permettrait de faciliter de nombreuses tâches de traitement de la musique liées au rythme. 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 [14].

cite méthode et école Agostini?

12. Roland TR-808

13. SQ-1

363 La TAB est un sujet de recherche crucial pour la compréhension des as-
 364 pects rythmiques de la musique, et a potentiellement un fort impact sur
 365 des domaines plus larges tels que l'éducation musicale et la production
 366 musicale.

367 1.4 Les représentations de la musique

369 Les données audio

370 Le format de fichier WAV est une instance du *Resource Interchange File*
 371 *Format (RIFF)* défini par IBM et Microsoft. Le format RIFF agit comme
 372 une "enveloppe" pour divers formats de codage audio. Un fichier WAV
 373 peut contenir de l'audio compressé ou non compressé.

375 Les données MIDI

376 Le MIDI¹⁴ (Musical Instrument Digital Interface) est une norme tech-
 377 nique qui décrit un protocole de communication, une interface numérique
 378 et des connecteurs électriques permettant de connecter une grande va-
 379 riété d'instruments de musique électroniques, d'ordinateurs et d'appareils
 380 audio connexes pour jouer, éditer et enregistrer de la musique.
 381 Les données midi sont représentées sous forme de piano-roll. Chaque
 382 point sur la figure 1.1 est appelé « évènement MIDI » :

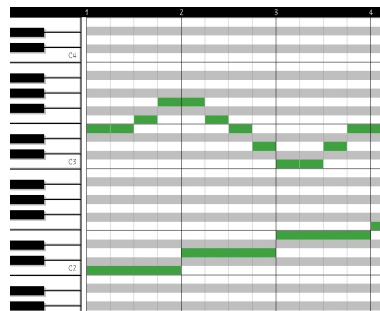


FIGURE 1.1 – Exemple évènements avec durée

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

Pour la batterie, les évènements sont considérés sans durée, nous ignore-
 rons 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.

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

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

citer M. Müller FMP pour cette section ?

trop technique. ne pas repier wikipédia

LPCM pas utile ici. parle juste échantillons et compression.

tu peux mentionner le format spectral (analyse harmonique) crucial en MIR audio.

ne pas copier wikipédia verbatim. source : midi.org MIDI est un protocole temps réel pour échanger des messages (événement) et un format de fichier.

fichier MIDI = séquence événements MIDI + dates (timestamp) performance musicale symbolique

donner ici les données des événements et expliquer ON/OFF (clavier)

il n'y a pas de durée d'évènement dans un MIDI file. la "durée" est une distance entre 2 événements ON et OFF (c'est important dans ton travail). le screenshot n'est pas utile, écrit plutôt une liste itemize

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

FIGURE 1.2 – Critère pour un évènement

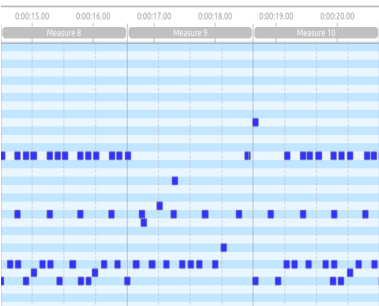


FIGURE 1.3 – Exemple évènements sans durée

392 On observe que toutes les durées sont identiques. <dam>je te suggère un
393 petit paragraphe ensuite, genre : "Le format MIDI, originellement une
394 norme technique, peut également être considéré comme une représenta-
395 tion musicale. Celle-ci peut effectivement être visualisée sous la forme
396 d'une partition ou jouée par l'ordinateur. Ce format historique, encore très
397 largement utilisé, est très important (mais aussi contraignant) dans le
398 cadre de notre travail, dans la mesure où de nombreux logiciels l'utilisent.
399 Pour la transcription musicale, il constitue une strate intermédiaire très
400 utile entre le signal audio (enregistrement) et la représentation musicale
401 lisible par un humain (partition)"</dam>

402 **Les partitions**

403 Une partition de musique¹⁵ est un document qui porte la représentation
404 systématique du langage musical sous forme écrite. Cette représentation
405 est appelée transcription et elle sert à traduire les quatre caractéristiques
406 du son musical :
407 — la hauteur ;
408 — la durée ;
409 — l'intensité ;

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



FIGURE 1.4 – Exemple de partition de piano

expliquer un peu plus av
exemple, ce serait mieux
d'avoir un ex. avec des
nuances, accents, appogia
tures...

414

415

explications sur l'aspect
structuré (hiérarchie) : les
mesures, les groupes ryht-
miques... c'est important
ici

418

— 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).

Les formats XML

Il existe plusieurs formats XML dédiés à la musique : MusicXML, MEI, MNX, ...

L'inconvénient de ces formats est qu'ils sont verbeux et ambigus, c'est pourquoi nous utilisons pour la transcription une représentation intermédiaire abstraite décrite plus loin.

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



FIGURE 1.5 – MusicXML

Le figure 1.5¹⁶ représente un do en clef de sol de la durée d'une ronde sur une mesure en 4/4 écrit au format MusicXML. Un des avantages de ce format est qu'il peut être converti aussi bien en données MIDI qu'en partition musicale, ce qui en fait une interface homme/machine.

16. Source images : <https://fr.wikipedia.org/wiki/MusicXML>

428 **Conclusion**

429 Dans ce chapitre, nous avons établi que la RIM s'intéresse de plus en plus
430 au TAL, et que, par ce biais, il y a des liens possibles entre le langage
431 musical et les langues naturelles, le plus proche étant probablement le
432 phénomène d'écriture des sons de l'un comme de l'autre.

433 Nous avons également établi que la RIM est née de la TAM qui est un
434 problème ancien et très difficile et qu'il serait toujours très utile de le
435 résoudre (autant pour la TAM que pour la TAB).

436 Et enfin, nous avons décrit les représentations de la musique nécessaires
437 à la compréhension du présent mémoire, allant du son jusqu'à l'écriture.

ÉTAT DE L'ART

Sommaire

2.1	Monophonique et polyphonique	21
2.2	Audio vers MIDI	22
2.3	MIDI vers partition	24
2.4	Approche linéaire et approche hiérarchique	24

Introduction

Dans ce chapitre, nous présenterons quelques travaux antérieurs dans le domaine de la transcription automatique de la musique et de la batterie afin de situer notre démarche.

Nous aborderons le passage crucial du monophonique au polyphonique dans la transcription. Nous ferons un point sur les deux grandes parties de la TAM de bout en bout : de l'audio vers le MIDI puis des données MIDI vers l'écriture d'une partition. Ensuite, nous discuterons des approches linéaires et des approches hiérarchiques.

2.1 Monophonique et polyphonique

Les premiers travaux en transcription ont été faits sur l'identification des instruments monophoniques¹ [11]. Actuellement, le problème de l'estimation automatique de la hauteur des signaux monophoniques peut être considéré comme résolu, mais dans la plupart des contextes musicaux, les instruments sont polyphoniques². L'estimation des hauteurs multiples

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

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

(détection multi-pitches ou F0 multiples) est le problème central de la création d'un système de transcription de musique polyphonique. Il s'agit de la détection de notes qui peuvent apparaître simultanément et être produites par plusieurs instruments différents. Ce défi est donc majeur pour la batterie puisque c'est un instrument qui est lui-même constitué de plusieurs instruments (caisse-claire, grosse-caisse, cymbales, toms, etc...). Le fort degré de chevauchement entre les durées ainsi qu'entre les fréquences complique l'identification des instruments polyphoniques. Cette tâche est étroitement liée à la séparation des sources et concerne aussi la séparation des voix. Les performances des systèmes actuels ne sont pas encore suffisantes pour permettre la création d'un système automatisé capable de transcrire de la musique polyphonique sans restrictions sur le degré de polyphonie ou le type d'instrument. Cette question reste donc encore ouverte.

2.2 Audio vers MIDI

Jusqu'à aujourd'hui, les recherches se sont majoritairement concentrées sur le traitement de signaux audio vers la génération du MIDI [15].

Cette partie englobe plusieurs sous-tâches dont la détection multi-pitches, la détection des onset et des offset, l'estimation du tempo, la quantification du rythme, la classification des genres musicaux, etc...

La figure 2.1 est une proposition de Benetos *et al.* [11] qui représente l'architecture générale d'un système de transcription musicale. On y observe plusieurs sous-tâches de la TAM :

- La séparation des sources à partir de l'audio.
- Le système de transcription :
 - Cœur du système :
 - ⇒ Algorithmes de détection des multi-pitches et de suivi des notes.
 - Quatres sous-tâches optionnelles accompagnent ces algorithmes :
 - identification de l'instrument ;
 - estimation de la tonalité et de l'accord ;
 - détection de l'apparition et du décalage ;
 - estimation du tempo et du rythme.
 - ça serait bien d'avoir une vision approximative des données : - identification de l'instrument : valeur symbolique prise dans une liste prédéfinie? - estimation de la tonalité et de l'accord : en note la gamme ou Hz? - détection de l'apparition et du décalage : mesure de temps / durée - estimation du tempo et du rythme :?
- Apprentissage sur des modèles accoustiques et musicologiques.

MIDI non-quantifié = performance (à expliquer)

en général tempo et quantification ne sont pas traités ici, le but est seulement la génération d'un MIDI non-quantifié

cela pourra être utile d'avoir une explication (ici ou en 1.4) sur la différence entre les timings de performance (dont le MIDI non-quantifié est un enregistrement symbolique) et les timing des partitions, avec 2 unités temporelles différentes (secondes et temps en relation par tempo.

classification des genres? ce n'est pas de la transcription! séparation des sources oui.

avant l'ADT, il faudrait dire 2 mots sur les techniques utilisées (cf. survey AMT Benetos et al.)

la figure ne correspond pas à ton travail, ici "score" = MIDI performance.

- 506 — *Optionnel* : Informations fournies de manière externe, soit fournie
 507 en amont (genre, instruments,...), soit par interaction avec un uti-
 508 lisateur (infos sur une partition incomplète).

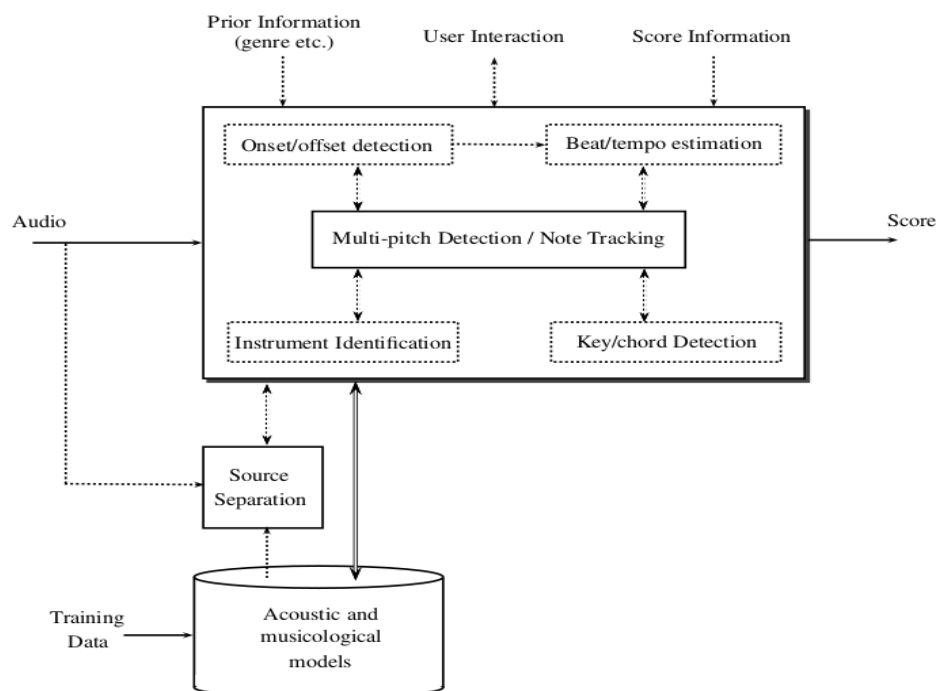


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

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.

509 En ADT [14], plusieurs stratégies de répartition pré/post-processing sont
 510 possibles pour la détection multi-pitches. Entamer la détection dès le pré-
 511 processing, en supprimant les features non-pertinentes pendant la sépa-
 512 ration des sources afin d'obtenir une meilleure détection des instruments
 513 de la batterie, est une démarche intuitive : supprimer la structure har-
 514 monique pour atténuer l'influence des instruments à hauteurs sur la dé-
 515 tecton grosse-caisse et caisse-claire en est un exemple. Mais certaines
 516 études montrent que des expériences similaires ont donné des résultats
 517 non-concluants et que la suppression des instruments à hauteurs peut
 518 avoir des effets néfastes sur les performances de l'ADT. En outre, les sys-
 519 tèmes d'ADT basés sur des réseaux de neurones récurrents (RNN) ou sur
 520 des factorisations matricielles non négative font la séparation des sources
 521 pendant l'optimisation, ce qui réduit la nécessité de la faire pendant le

haute fréquence, aigus?

522 pré-processing.
 523 Pour la reconnaissance des instruments, une approche possible [16] est
 524 de mettre un modèle probabiliste dans l'étape de la classification des évè-
 525 nements afin de classer les différents sons de la batterie. Cette méthode
 526 permet de se passer de samples audio isolés en modélisant la progression
 527 temporelle des *features*³ avec un modèle de markow caché (HMM). Les
 528 *features* sont transformés en représentations statistiques indépendantes.
 L'approche AdaMa [17] est une autre approche de la même catégorie ; elle
 commence par une estimation initiale des sons de la batterie qui sont ité-
 rativement raffinés pour correspondre à (pour matcher) l'enregistrement
 visé.

classification des évène- 525
 ments? la phrase semble 526
 redondante

pas clair... peut-être just 529
 mentionner les modèles 530
 probabilistes utilisés

533 2.3 MIDI vers partition

534 Le plus souvent, lorsque les articles abordent la transcription automa-
 535 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »
 536 (partition) désigne un ouput au format Music XML, ou simplement MIDI.
 537 Par exemple, dans [18], la chaîne de traitement va jusqu'à la génération
 538 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en
 539 extraire manuellement un fichier MusicXML contenant plusieurs voix.
 540 Seuls quelques travaux récents s'intéressent de près à la création d'outils
 541 permettant la génération de partition. Le problème de la conversion d'une
 542 séquence d'évènements musicaux symboliques en une partition musicale
 543 structurée est traité notamment dans [19]. Ce travail, qui vise à résoudre
 544 en une fois la quantification rythmique et la production de partition struc-
 545 turée, s'appuie tout au long du processus sur des grammaires génératives
 546 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-
 547 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été
 548 menées avec un ensemble de données composé d'extraits monophoniques ;
 549 il reste donc à traiter le passage au polyphonique, en couplant le problème
 de la séparation des voix avec la quantification du rythme.
 L'approche de [19] est fondée sur la conviction que la complexité de la
 structure musicale dépasse les modèles linéaires.

ce n'est pas exactement 534
 cela. cf. proposition de des- 535
 cription + détaillée en com- 536
 mentaires

de manière conjointe 545

langage a priori 547

qui nécessite de traiter le 550
 problème supplémentaire 551
 de la séparation de voix. i.e. 552
 pour la batterie on nveut 553
 quantification + structu- 554
 ration + séparation mais 555
 seules les 2 premières sont 556
 couplées dans l'approche de 557
 tonn stage.

554 2.4 Approche linéaire et approche hiérarchique

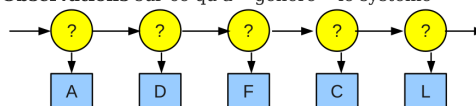
555 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par
 556 exemple, Shibata *et al.* [18] ont utilisé le modèle de Markov caché (HMM)⁴
 557 pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux

3. Features : caractéristiques individuelles mesurables d'un phénomène dans le do-
 maine de l'apprentissage automatique et de la reconnaissance des formes

4. https://fr.wikipedia.org/wiki/Modèle_de_Markov_caché
https://en.wikipedia.org/wiki/Hidden_Markov_model

558 réseaux de neurones profonds, l'un pour la reconnaissance des pitches et
559 l'autre pour la reconnaissance de la vélocité. Pour la dernière couche, la
560 probabilité est obtenue par une fonction sigmoïde. Ils construisent en-
561 suite plusieurs HMM métriques étendus pour la musique polyphonique
562 correspondant à des métriques possibles, puis ils calculent la probabilité
563 maximale pour chaque modèle afin d'obtenir la métrique la plus probable.

- Modèle de Markov **caché** :
 - **Hidden Markov Model (HMM) (Baum, 1965)**
 - Modélisation d'un processus stochastique « **génératif** » :
 - État du système : non connu
 - Connaissance pour chaque état des **probabilités** comme état initial, de **transition** entre états et de **génération** de symboles
 - **Observations** sur ce qu'a « généré » le système



- Applications : physique, reconnaissance de parole, traitement du langage, bio-informatique, finance, etc.

FIGURE 2.2 – HMM

564 *Source : Cours de Damien Nouvel*⁵

565

566

567 L'évaluation finale des résultats de [18] montre qu'il faut rediriger l'atten-
 568 tion vers les valeurs des notes, la séparation des voix et d'autres éléments
 569 délicats de la partition musicale qui sont significatifs pour l'exécution de
 570 la musique. Or, même si la quantification du rythme se fait le plus souvent
 571 par la manipulation de données linéaires allant notamment des *real time*
 572 *units* (secondes) vers les musical *time units* (temps, métrique, ...), de nom-
 573 breux travaux suggèrent d'utiliser une approche hiérarchique puisque le

je ne comprend pas bien 574
 l'explication. le pb est plu- 575
 tot vue locale (déduction 576
 la proba d'une durée à par- 577
 tir de la durée précédente, 578
 par ex. dans un HMM) vs 579
 vue globale, dans une hié-
 rarchie

RT? 579
 580

techniques de réécriture 581
 appliquée à la déduction 582
 automatique, calcul symbo-
 lique 583

le calcul d'équiv. 584
 585

citer thèse de David Rizo 588
 (Valencia) 589

586

587 La nécessité d'une approche hiérarchique pour la production automatique
 de partition est évoquée dans [19]. Les modèles de grammaire qui y sont
 exposés sont différents de modèles markoviens linéaires de précédents
 travaux.

5. <https://damien.nouvel.net/fr/enseignement>

612

613

MÉTHODES

614

Sommaire

615

616

617

618

619

620

621

622

3.1	La notation de la batterie	29
3.2	Modélisation pour la transcription	37
3.3	Qparse	38
3.4	Les systèmes	40

623

Introduction

624

625

626

627

628

629

630

631

Dans ce chapitre, nous expliquerons en détail les méthodes que nous avons employées pour l'ADT.

Pour commencer, nous exposerons une description de la notation de la batterie ainsi qu'une modélisation de celle-ci pour la représentation des données rythmiques en arbres syntaxiques. Nous poursuivrons avec une présentation de *qparse*¹, un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et au sein du laboratoire Cedric au CNAM.

Enfin, nous présenterons les systèmes.

plusieurs développeurs

systèmes, une représentation théorique qui permet...

632

3.1 La notation de la batterie



FIGURE 3.1

633

634

635

La figure 3.1 montre 4 figures de notes les plus courantes dont les noms et les durées sont respectivement, de gauche à droite :

— La ronde, elle vaut 4 ;

durées exprimées en unité de temps musicale, appelée le *temps*, cf. section...

4 temps

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

636 — La blanche, elle vaut 2 ;

637 — La noire, elle vaut 1 ;

638 — La croche, elle vaut 1/2.

plusieurs éléments

639 Une figure de note [1] de musique combine plusieurs critères ² :

— Une tête de note :

Sa position sur la portée indique la hauteur de la note. La tête de note peut aussi indiquer une durée.

— Une hampe :

Indicatrice d'appartenance à une voix en fonction de sa direction et indicatrice d'une durée représentée par sa présence ou non (blanche \neq ronde)

— Un crochet : La durée d'une note est divisée par deux à chaque crochet ajouté à la hampe d'une figure de note.

plutôt que wikipedia cite
Dannhauser ou autre ref
F.M. ou encore Gould 2011
Behind Bars

barre verticale liée à la tête
de note

haut ou bas

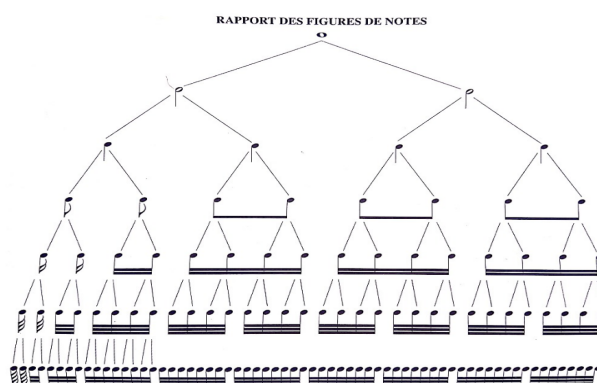


FIGURE 3.2 – Rapport des figures de notes
[1]

649 La figure 3.2 montre les rapports de durée entre les figures de notes. Plus
650 les durées sont longues, plus elles sont marquées par la tête de note (la
651 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de
652 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute
653 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.
654 Les notes à crochet (croche, double-croche, triple-croche...) peuvent être
655 reliées ou non par des ligatures (voir les 4 dernières lignes de la figure
656 3.2).

ce premier paragraphe (je
qu'ici) est redondant avec
§1.4 (sub. partitions). dé-
placer en 1.4? cf. proposi-
tion plus loin

Les hauteurs et les têtes de notes

658 Source figure 3.3 : [https://www.superprof.fr/blog/
659 composition-instrument-percussion/](https://www.superprof.fr/blog/composition-instrument-percussion/).

660 Pour la transcription, nous proposons une notation inspirée du recueil de
661 pièces pour batterie de J.-F. Juskowiak [24] et des méthodes de batterie

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



FIGURE 3.3 – Les instruments de la batterie

662 Agostini [25], car nous trouvons la position des éléments cohérente et in-
663 tuitive.

665 En effet, les hauteurs sur la portée représentent :

- 666 — La hauteur physique des instruments :
667 La caisse claire est centrale sur la portée et sur la batterie (au
668 niveau de la ceinture, elle conditionne l'écart entre les pédales et
669 aussi la position de tous les instruments basiques d'une batterie).
670 Tout ce qui en-dessous de la caisse-claire sur la portée est en
671 dessous de la caisse-claire sur la batterie (pédales, tom basse);
672 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est
673 aussi sur la batterie.
674
- 675 — La hauteur des instruments en terme de fréquences :
676 Sauf pour le charley au pied et si l'on sépare en trois groupes
677 (grosse-caisse, toms et cymbales), de bas en haut, les instruments
678 vont du plus grave au plus aigu.

pour aider, tu pourrais don-
ner une figure représentant
la batterie avec le nom des
instruments et abrégia-
tion.

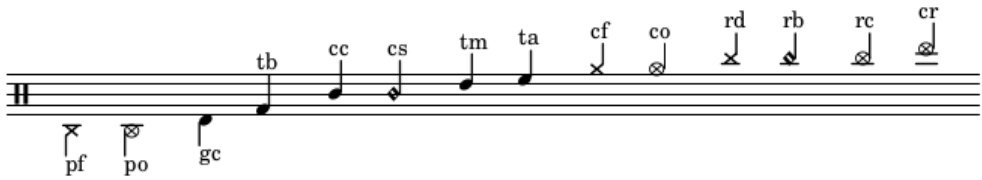


FIGURE 3.4 – Hauteur et têtes de notes

679 Les noms des instruments correspondant aux codes des notes de la figure
680 3.4 sont dans le tableau 3.1.

têtes de notes?

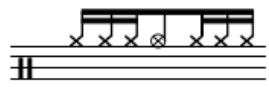
681 Les durées

682 Comme nous venons de la voir, la majorité des instruments de la batterie
 683 sont représentés par les têtes des notes. Par conséquent, les symboles
 684 rythmiques concernant la tête de note ne pourront pas être utilisés. Cela
 685 est valable aussi pour la présence ou non de la hampe puisque ce phé-
 686 nomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition
 687 blanche-ronde). L'usage des blanches existe dans certaines partitions de
 688 batterie [26] mais cela reste dans des cas très rares. Certains logiciels per-
 689 mettent de faire des blanches avec des symboles spécifiques à la batterie
 690 ou aux percussions mais leur lecture reste peu aisée et leur utilisation
 691 pour la batterie est rarissime.

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

- Le point ;
- La liaison.

695 Ces symboles ne seront utiles que pour l'écriture des ouvertures de char-
 696 ley. Le charley est le seul instrument de la batterie dont la durée est quan-
 697 tifiée (les cymbales attrapées à la main peuvent l'être aussi mais cela est
 698 très rare.)



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.5 – Point et liaison

= la position des temps 699

700

faire un "enumerate" 702

703

704

L'écriture de la batterie doit faire ressortir la pulsation. La première chose à prendre en compte pour analyser la figure 3.5 est donc la nécessité de regrouper les notes par temps à l'aide des ligatures.

Exemple 1 : ouverture de charley quantifiée mais pas notes pas regroupées par temps.

Exemple 2 : Ici, la liaison permet de regrouper les notes par temps en ob-
tenant le même rythme que dans l'exemple 1.
Exemple 3 et exemple 4 : les deux exemples sont valables mais le
deuxième est le plus souvent utilisé car plus intuitif (regroupement par
temps).
En cas de nécessité de prolonger la durée d'une note au-delà de sa durée
initial, et si cette note correspond à une ouverture de charley, on privilé-
giera la liaison.

Les silences

Les silences sont parfois utilisés pour quantifier les ouvertures de charley.
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.

expliquer la notation (géné-
rale) des silebces en §1.4?

quantifier = noter? ou
quantifier la durée?

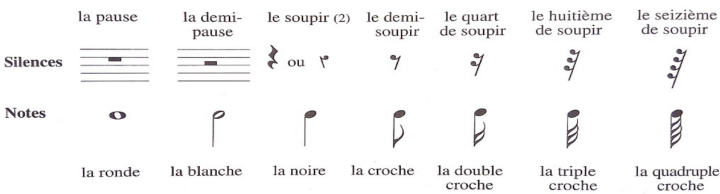


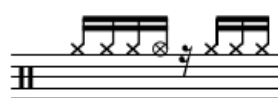
FIGURE 3.6 – Les silences

719 Physiquement, le charley est fermé par une pression du pied sur la pé-
 720 dale de charley. Dans les fichiers MIDI, cette pression est traduite par
 721 un charley joué au pied. Mais dans une vraie partition, cette écriture ne
 722 traduirait pas ce que le batteur doit penser.

pas très clair



Exemple 1



Exemple 2

FIGURE 3.7 – Silence joué

723 L'exemple 1 de la figure 3.7 montre ce qui est écrit dans les données MIDI
 724 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-
 725 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1
 726 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-
 727 truments jouant simultanément.

728 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte
 729 la note suivante pour l'écriture : - Si c'est un charley fermé joué à la main
 730 ⇒ la note sera cf;

itemize

cf?

731 - Si c'est un charley fermé joué au pied ⇒ la note sera un silence.

732 Les équivalences rythmiques

733 Pour les instruments mélodiques, la liaison et le point sont les deux seules
 734 possibilités en cas d'équivalence rythmique pour des notes dont la durée
 de l'une à l'autre est ininterrompue. Mais pour la batterie, à part dans
 le cas des ouvertures de charley (voir section 3.1), les durées des notes
 n'ont pas d'importance. L'usage des silences pour combler la distance ryth-
 mique entre deux notes devient donc possible.

phrase alambiquée... pou
prolonger la durée?seuls comptent les date
début de notes onsets.

739 Cela pris en compte, et étant donné que les indications de durée dans les
 740 têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide
 741 de silences sera privilégiée comme indication de durée sauf dans les cas
 742 où cela reste impossible. Ce choix à pour but de n'avoir qu'une manière
 743 d'écrire toutes les notes, que leurs têtes de notes soit modifiées ou non.

744 Sur la figure 3.8, théoriquement, il faudra choisir la notation de la
 745 deuxième mesure mais dans certains contextes, pour des raisons de lisi-
 746 bilité ou de surcharge, la version sans les silences de la troisième mesure
 747 pourra être choisie.



FIGURE 3.8 – Équivalence

Les voix

Les voix³ désignent les différentes parties mélodiques constituant une composition musicale et destinées à être interprétées, simultanément ou successivement, par un ou plusieurs musiciens. En batterie, une voix est l'ensemble des instruments qui, à eux seuls, constituent une phrase rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant possibles pour un même rythme, on peut regrouper les instruments de la batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou 2 voix. Sur la figure 3.9, il faudra faire un choix entre les exemples 1, 2 et 3 qui sont trois façons d'écrire le même rythme.

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

voix : citations possibles :
- "Joint Estimation of Note Values and Voices for Audio-to-Score Piano Transcription" Nakamura et al 2021 ou une des références de ce papier, par ex. [15] ou [16]. - ou thèse de Nicolas Guiomard-Kagan.

une voix est caractérisée aussi par orientation des hampes?



FIGURE 3.9 – Séparation des voix

Ce choix se fera en fonction des instruments joués, de la nature plus ou moins systématique de leurs phrasés, et des associations logiques entre les instruments dans la distribution des rythmes sur la batterie (voir la section 3.4).

3. [https://fr.wikipedia.org/wiki/Voix_\(polyphonie\)](https://fr.wikipedia.org/wiki/Voix_(polyphonie))

762 Les accentuations et les ghost-notes

763 « Certaines notes dans une phrase musicale doivent, ainsi que les dif-
764 férentes syllabes d'un mot, être accentuées avec plus ou moins de force,
porter une inflexion particulière. » [1]

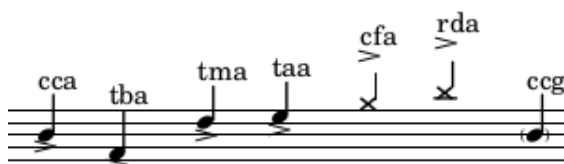


FIGURE 3.10 – Les accents et les ghost-notes

3.9 = liste des seuls "ins-
truments" qui peuvent être
accentués?

765 La figure 3.10 ne prend en compte que les accents que nous avons es-
766 timés nécessaires (voir la section 3.2). Les accents sont marqués par le
767 symbole « > ». Il est positionné au-dessus des notes représentant des cym-
768 bales et en-dessous des notes représentant des toms ou la caisse-claire.
769 Ce choix a été fait pour la partition de la figure 4.2 car elle est plus lisible
770 ainsi, mais ces choix devront être adaptés en fonction des différents sys-
771 tèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz,
772 les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il
773 faudra donc mettre les symboles d'accentuation correspondants au-dessus
774 des têtes de notes.
775

expliquer ce qu'est une
ghost-notes

776 La dernière note de la figure 3.10 montre un exemple de ghost-notes. Le
777 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note
778 sans changer la tête de note.

les codes de notes n'ont pas
encore été présentés...

779 Pour les codes, on prend le code de la note et on ajoute un « a » pour un
780 accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.10
sont exposées en situation réelle dans la figure 3.11.



FIGURE 3.11 – Exemple pour les accentuations et les ghost-notes

3.2 Modélisation pour la transcription

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

Il existe, pour de nombreux instruments de la batterie, plusieurs samples audio associés à des pitches. Pour cette première version, nous avons choisi de n’avoir qu’un code-instrument pour différentes variantes d’un instrument, c’est pourquoi certain code-instrument se voit attribuer plusieurs pitches dans le tableau 3.1.

Malgré le large panel de pitches disponible, il semblerait qu’aucun pitch ne désigne le charley ouvert joué au pied. Pourtant, dans la batterie moderne, plusieurs rythmes ne peuvent fournir le son du charley ouvert qu’avec le pied car les mains ne sont pas disponibles pour le jouer. Cela doit en partie être dû à l’utilisation des boîte à rythmes en MAO qui ne nécessitent pas de faire des choix conditionnés par les limitations humaines (2 pieds, 2 mains, et beaucoup plus d’instruments...)

je ne comprend pas cette phrase.

il s’agit juste d’une convention de codage des instruments de la batterie en événements MIDI... que l’on prend en entrée pour la transcription

La vélocité

La partition de la figure 4.2 a été transcrite manuellement avec lilypond par analyse des fichiers MIDI et audio correspondants.

Cette transcription nous a mené aux observations suivantes :

- Vélocité inférieure à 40 : ghost-note ;
- Vélocité supérieure à 90 : accent ;
- Pas d’intention d’accent ni de ghost-note pour une vélocité entre 40 et 89 ;

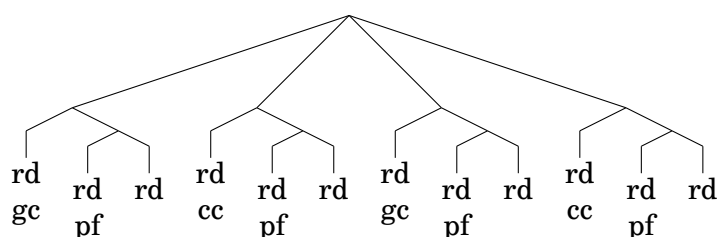
citation lilypond

et l’analyse d’autres fichiers MIDI ?

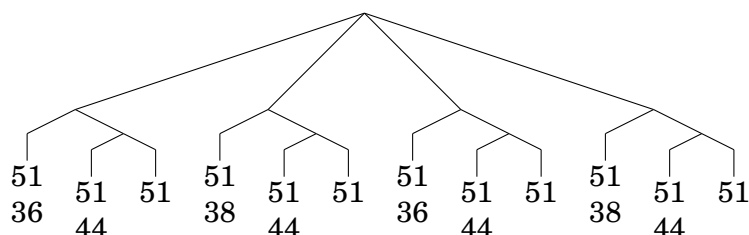
- 804 — Les accents et les ghosts-notes ne sont significatifs ni pour les ins-
 805 truments joués au pied, ni pour les cymbales crash.
 806 En effet, certaines vélocités en dessous de 40 étant détectées et ins-
 807 crites dans les données MIDI sont dues au mouvement du talon du
 808 batteur qui bat la pulsation sans particulièrement jouer le charley.
 809 Ce mouvement est perçu par le capteur de la batterie électronique
 810 mais le charley n'est pas joué.
 811 — Au final, nous avons relevé les ghost-notes et les accents pour la
 812 caisse-claire ainsi que les accents pour les toms et les cymbales
 813 rythmiques (charley et ride).

814 Les arbres de rythmes

815 Les arbres de rythmes représentent un rythme unique dont les possibili-
 816 tés de notation sur une partition sont théoriquement multiples.
 817 Voici une représentation de la figure 3.9 en arbre de rythmes avec les
 818 codes de chaque instrument :



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



821 Chacun des trois exemples de la figure 3.9 est représenté par un des deux
 822 arbres syntaxiques ci-dessus.

823 3.3 Qparse

824 La librairie Qparse⁴ implémente la quantification des rythmes basée
 825 sur des algorithmes d'analyse syntaxique pour les automates arbores-
 826

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

non c'est juste une repré-
 sentation du rythme, pas
 unique

expliquer le principe des
 RT : branchement = divi-
 sion d'intervalle temporel,
 feuilles = les événements
 musicaux commençant au
 début de l'intervalle). réfé-
 rences : - Laurson "Patch-
 work : A Visual Program-
 ming Language", 1996. -
 OpenMusic : visual pro-
 gramming environment for
 music composition, analysis
 and research, 2011.

Fig. 3.8, ex. 1, 2 ou 3?

choisir titre plus explicite
 par ex. analyse syntaxique
 pour la transcription musi-
 cale

quantification rythmique
 + structuration de partition

qparse est un outil pour la
 transcription musicale, qui,
 à partir d'une performance
 symbolique, séquentielle et
 non quantifiée, produit une
 partition structurée.

Il effectue conjointement
 des tâches de quantification
 rythmique et d'inférence
 de la structure de la parti-
 tion à l'aide de technique
 de parsing / analyse

cents pondérés. En prenant en entrée une performance musicale symbolique (séquence de notes avec dates et durées en temps réel, typiquement un fichier MIDI), et une grammaire hors-contexte pondérée décrivant un langage de rythmes préférés, il produit une partition musicale. Plusieurs formats de sortie sont possibles, dont XML, MEI.

grammaire \neq automate.
il faut choisir entre les 2
(pour la suite aussi)

Les principaux contributeurs sont :

- Florent Jacquemard (Inria) : développeur principal.
- Francesco Foscari (PhD, CNAM) : construction de grammaire automatique à partir de corpus ; Evaluation.
- Clement Poncelet (Salzburg U.) : integration de la librairie Midifile pour les input MIDI.
- Philippe Rigaux (CNAM) : production de partition au format MEI et de modèle intermédiaire de partition en sortie.
- Masahiko Sakai (Nagoya U.) : mesure de la distance input/output pour la quantification et CMake framework ; évaluation.

apprentissage

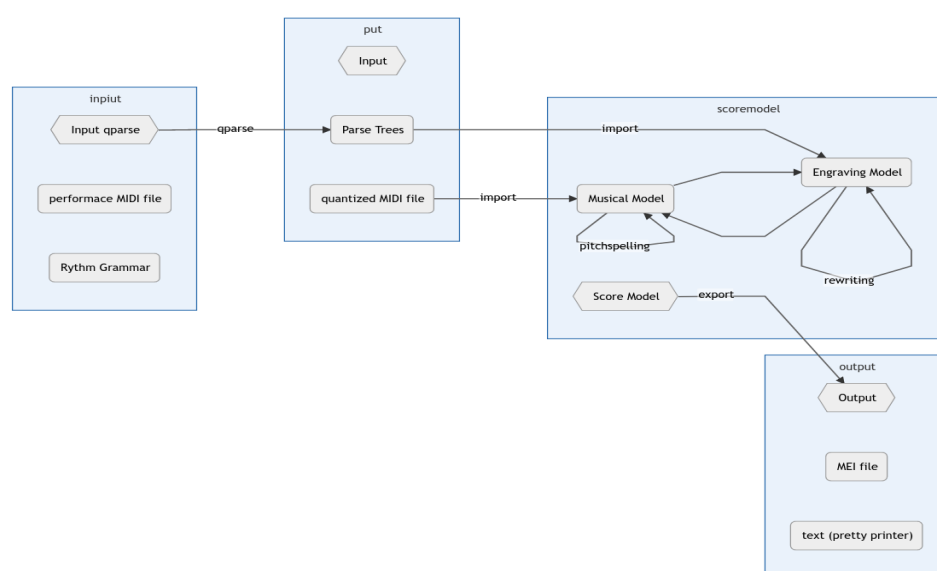


FIGURE 3.12 – Présentation de Qparse

Explication des différentes étapes de la figure 3.12⁵ :

- **Input Qparse** :
Un fichier MIDI (séquence d'événements datés (piano roll) accompagné d'un fichier contenant une grammaire pondérée) ;
- **Arbre de parsing** :
Les données MIDI sont quantifiées, les notes de dates proches sont

la figure 3.11 est trop compliquée. rhythm grammar → automate d'arbres pondéré. Parse Tree → arbre syntaxique. qtz MIDI file : inutile. Score Model → représentation intermédiaire de partition. Score Model, Engr. Model : inutile. garder juste la fleche Rewriting sur S.M.

5. <https://gitlab.inria.fr/qparse/qparselib/-/tree/distance/src/scoremodel>

- alignées et les relations entre les notes sont identifiées (accords, fla, etc. . .); un arbre de parsing global est créé ;
- **Score Model** :
 - Les instruments sont identifiés dans `scoremodel/import/tableImporterDrum.cpp` ;
 - Réécriture 1 :
 - séparation des voix \Rightarrow un arbre par voix \Rightarrow représentation intermédiaire (RI) ;
 - Réécriture 2 :
 - simplification de l'écriture de chaque voix dans la RI ;
 - **Output** :
 - export de la partition. Plusieurs formats sont possibles (xml, mei, lilypond, . . .).
- Plusieurs enjeux :
- Problème du MIDI avec Qparse :
 - ON-OFF en entrée \Rightarrow 1 seul symbole en sortie.
 - Minimiser la distance entre le midi et la représentation en arbre.
 - Un des problèmes de Qparse était qu'il était limité au monophonique.
 - Quelles sont les limites du monophonique ?
 - Impossibilité de traiter plusieurs voix et de reconnaître les accords.

3.4 Les systèmes

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

Définitions

Système : motif + gamme/texte

Motif : rythmes coordonnés joués avec 2 ou 3 membres en boucle (répartis sur 1 ou 2 voix)

Texte : rythme irrégulier joué avec un seul membre sur le motif (réparti sur 1 voix).

Gamme : 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.

Un ensemble de systèmes comprenant leur métrique et leurs règles spécifiques de réécriture sera nécessaire. Les systèmes devront être distribués

il faudrait expliquer là que le but est d'avoir des schemas types (= système) pour calculer la séparation en voix. = une heuristique pour éviter d'avoir à explorer une grande combinatoire. et que, une fois le système déterminé (ou sélectionné), la séparation se fait par réécriture du modèle (règles de projection simplification)

je ne comprend pas bien la définition de système : motif + gamme ou motif + gamme + texte ? la déf. des gammes n'est pas du tout claire.

est-ce que le motif est fixe et les gammes variables ? est-ce le motif qui détermine la métrique et les voix ?

métrique n'est pas définie, règles de réécriture non plus

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

TABLE 3.2 – Systèmes

dans 4 grandes catégories :
 Nous exposerons 3 systèmes afin d’illustrer les propos de cette section :
 — 4/4 binaire
 — 4/4 jazz
 — 4/4 afro-cubain

Objectif des systèmes

Les systèmes devront être matchés sur l’input MIDI afin de :
 — définir une métrique ;
 — choisir une grammaire appropriée ;
 — fournir les règles de réécriture (séparation des voix et simplification.

La partie *motif* des systèmes sera utilisée pour la **définition des métriques**. Le *motif* et la gammes des systèmes seront utilisés pour la **séparation des voix**. Les règles de **simplification** (les combinaisons de réécritures) seront extraites des voix séparées des systèmes.

Détection d’indication de mesure

La détection de la métrique est importante, non seulement pour connaître le nombre de temps par mesure ainsi que le nombre de subdivisions pour chacun de ces temps, mais aussi pour savoir comment écrire l’unité de temps et ses subdivisions.

bien, il faudrait expliquer ça avant.

pas exactement. les règles de projection et simplification font la séparation en voix : à partir d’un arbre syntaxique comme celui de 3.2, elles extraient 2 arbres, chacun contenant les événements d’une seule voix

métrique ≠ signature rythmique (c’est plus général). Il aurait fallu présenter rapidement la notation des signatures rythmiques, par exemple en 1.4



Exemple 1



Exemple 2

FIGURE 3.13 – Métrique

La figure 3.13 montre deux indications de mesure différentes. L'une (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des triolets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est traditionnellement écrit en binaire avec ou sans triolet (même si cette musique est dite ternaire alors que le rock ternaire sera plutôt écrit comme dans l'exemple 2).

Choix d'une grammaire

Il faut prendre en compte l'existence potentielle de plusieurs grammaires dédiées chacune à un type de contenu MIDI. Le choix d'une grammaire pondérée doit être fait avant le parsing puisque Qparse prend en entrée un fichier MIDI et un fichier wta (grammaire). C'est pour cette raison que la métrique doit être définie avant le choix de la grammaire.

Pour les expériences effectuées avec le Groove MIDI Data Set, le style et l'indication de mesure sont récupérables par les noms des fichiers MIDI, mais il faudra par la suite les trouver automatiquement sans autres indications que les données MIDI elles-mêmes. Par conséquent, les motifs des systèmes devront être recherchés sur l'input (*fichiers MIDI*) avant le lancement du parsing, afin de déterminer la métrique en amont. Cette tâche devra probablement être effectuée en Machine Learning.

Séparation des voix

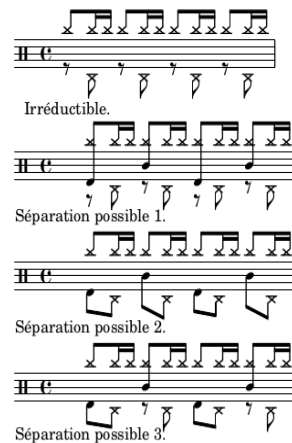


FIGURE 3.14 – Motif 4-4 binaire

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 rd/pf/cc (voir 3.1) et d'un texte joué à la grosse-caisse. La première ligne de la figure 3.14 est appelée « Irréductible

938 » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du
 939 charley au pied. La troisième séparation proposée est privilégiée car elle
 940 répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les
 941 pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont
 942 utilisés par voix et plus logique pour le lecteur puisque les mains sont en
 haut et les pieds en bas.

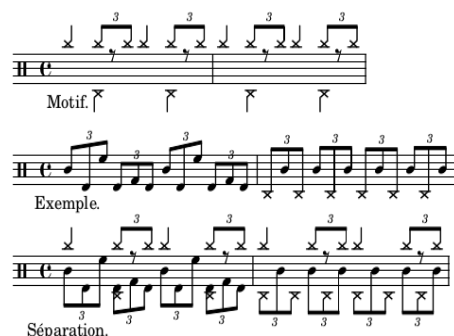


FIGURE 3.15 – Motif 4-4 jazz

943 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 944 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 945 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 946 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 947 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 948 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 949 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 950 et de caisse claire et son écriture séparée serait inutilement compliquée
 951 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 952 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de
 953 charley en bas.

quel exemple?



FIGURE 3.16 – Système 4-4 afro-latin

955 La figure 3.16 montre un exemple minimaliste de système afro-latin [26].
 956 Ce système doit être écrit sur trois voix car la voix centrale est souvent
 957 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas
 958 serait surchargé et peu lisible.

959 Simplification de l'écriture

960 Les explications qui suivent seront appuyé par une expérimentation théo-
 961 rique dans la section 4.4.

expérimentation théo-
 rique??

962 Les gammes qui accompagnent les motifs d'un système étayent toutes les
 963 combinaisons d'un système et elles permettent, combinées avec le motif
 964 d'un système, de définir les règles de simplification propres à celui-ci.

965 Voici les différentes étapes à suivre :

- 966 — Pour chaque gamme du système, faire un arbre de rythme repré-
 967 sentant la gamme combinée avec le motif du système ;
- 968 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
 969 arbre de rythme par voix ;
- 970 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
 971 nœuds qui nécessitent une simplification et écrire la règle.

972 Certaines précisions concernant l'extraction de ces règles sont néces-
 973 saires. Il s'agit de précisions à propos de la durée, des silences et de la
 974 présence ou non d'ouverture de charley dans les instruments joués. Nous
 975 avons discuté de ces problèmes dans le chapitre 3.

976 Voici quelques règles inhérentes à la simplification de l'écriture pour la
 977 batterie : Toutes les continuations (t) qui se trouvent en début de temps
 978 (figures 4.9, 4.11 et 4.12) sont transformées en silences (r) sauf si la note
 979 précédente est un charley ouvert ?

ce sont des figures et nota-
 tions du chapitre suivant!

980 Même si on favorise l'usage des silences pour l'écart entre les notes n'ap-
 981 partenant pas au même temps, on les supprime systématiquement pour
 982 2 notes au sein d'un même temps et favorise, une liaison si co, un point si
 983 pas co et nécessaire, un simple ajustement de la figure de note si suffisant.

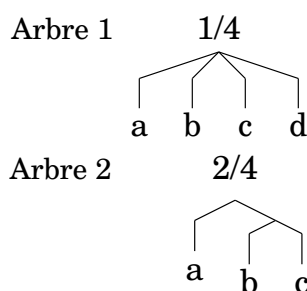


FIGURE 3.17 – Simplification

itemize

984 Soit l'arbre 1 de la figure 3.17 dans lequel : a et d sont des instruments de
 985 la batterie (x) ;

986 b et c sont des continuations (t) ;

987 Pour chacune des conditions suivantes, une suite de la figure 3.18 est
 988 attribuée :

- 989 — Si a n'est pas un co :
- 990 ⇒ Suite 1a.

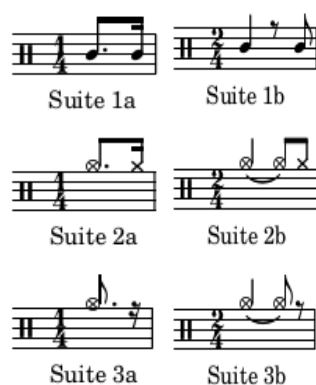


FIGURE 3.18

- 991 — Si a est un co :
 992 — Si d est un cf :
 993 ⇒ Suite 2a.
 994 — Si d est un pf :
 995 ⇒ Suite 3a : d deviens un silence (r).
 996

997 Soit l'arbre 2 de la figure 3.17 dans lequel :
 998 a et c sont des instruments de la batterie (x);
 999 b est une continuation (t); Pour chacune des conditions suivantes, une
 1000 suite de la figure 3.18 est attribuée :

- 1001 — Si a n'est pas un co :
 1002 ⇒ Suite 1b, b devient un silence.
 1003 — Si a est un co :
 1004 — Si c est un cf :
 1005 ⇒ Suite 2b, b devient une liaison et c devient un cf.
 1006 — Si c est un pf :
 1007 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
 1008

1009 *Rappel :*

- 1010 *cf* = charley fermé joué à la main ;
 1011 *co* = charley ouvert joué à la main ;
 1012 *pf* = charley fermé joué au pied.

1013

1014 **Problème : le cf et le co ne seront jamais sur la même voix que le**
 1015 **pf... Par conséquent, les règles concernant les charleys ouverts**
 1016 **doivent-elles être appliquées sur l'arbre de parsing de l'input?...**

1017 **Conclusion**

1018 Nous avons formalisé une notation de la batterie, modélisé cette notation
1019 pour la transcription de données MIDI en partition, nous avons décrit
1020 Qparse.

1021 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-
1022 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-
1023 propriée et énoncer des règles de séparation des voix et de simplification
1024 de l'écriture.

EXPÉRIMENTATIONS

Sommaire

1028	4.1	Le jeu de données	47
1029	4.2	Analyse MIDI-Audio	49
1030	4.3	Transcription polyphonique par parsing (?verrou?) . . .	53
1031	4.4	Expérimentation théorique d'un système	55
1032	4.5	BILAN : résultats — évaluation — discussion	60

Introduction

Dans ce chapitre, nous présenterons le jeu de données et les analyses audio-MIDI. Nous ferons ensuite l'expérimentation théorique d'un système implémentable qui devra être utilisé comme base de connaissances pour augmenter la rapidité et la qualité en sortie de Qparse. Nous présenterons ensuite les avancées réalisées dans ce travail et une réflexion sur les moyens de l'évaluer. Enfin, nous finirons par une discussion sur l'ensemble du travail réalisé.

4.1 Le jeu de données

Nous avons utilisé le Groove MIDI Dataset ¹ [27] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).

Le GMD est composé de 13,6 heures de batterie sous forme de fichiers MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000 mesures de batterie dans les styles les plus courants et avec différentes qualités de jeu. Tout le contenu a été joué par des humains sur la batterie électronique Roland TD-11 (figure 4.1).

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



FIGURE 4.1 – Batterie électronique

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

1054 Autres critères spécifiques au GMD :

- 1055 — Toutes les performances ont été jouées au métronome et à un tempo
- 1056 choisi par le batteur.
- 1057 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 1058 qui ont pu improviser dans un large éventail de styles. Les don-
- 1059 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 1060 (professionnel ou amateur).
- 1061 — Les batteurs avaient pour instruction de jouer des séquences de
- 1062 plusieurs minutes ainsi que des fills²
- 1063 — Chaque performance est annotée d'un style (fourni par le batteur),
- 1064 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 1065 nyme du batteur.
- 1066 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 1067 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 1068 session du GMD.
- 1069 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 1070 fichier MIDI.

1071 **Format des données**

enregistre les données dans
des fichiers MIDI

1072 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-
1073 tinctes :

- 1074 — une pour le tempo et l'indication de mesure ;
- 1075 — une pour les changements de contrôle (position de la pédale de
- 1076 charley) ;
- 1077 — une pour les notes.

1078

1079 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1080 canal 9 (qui est le canal canonique pour la batterie).

1081 Pour simplifier le traitement de ces données, ces trois pistes ont été
1082 fusionnées en une seule piste qui a été mise sur le canal 9.

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

1083
 1084 « Control Changes The TD-11 also records control changes speci-
 1085 fying the position of the hi-hat pedal on each hit. We have preserved this
 1086 information under control 4. »
 1087 (<https://magenta.tensorflow.org/datasets/groove>)
 1088 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 1089 tions dans les fichiers MIDI.
 1090 L'utilisation de pretty_midi devient urgente!

1091 4.2 Analyse MIDI-Audio

1092
 1093 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
 1094 partir de fichiers MIDI et Audio du GMD.

= analyses et transcriptions
manuelles

1095 Comparaisons de transcriptions

1096 Pour les comparaisons de transcriptions, les transcriptions manuelles
 1097 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les trans-
 1098 criptions automatiques (TA) ont toutes été générées manuellement avec
 1099 MuseScore.

méthodologie tr. manuelle.
import MIDI pour MuseS-
core

1100 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1101 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1102 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1103 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1104 premières notes (une double-croche et une croche) d'un triolet sur
- 1105 le temps 1 de la mesure 2 de la TA.

1106 Exemple d'analyse 2

- 1107 — Les doubles croches ont été interprétées en quintolet
- 1108 — La deuxième double-croche est devenue une croche.

Transcription manuelle \Rightarrow Transcription automatique



Transcription manuelle \Rightarrow Transcription automatique

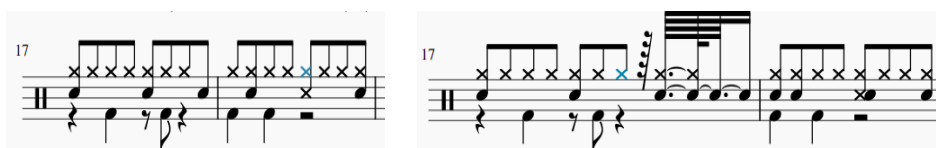


1109 Exemple d'analyse 3

- 1110 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1111
- 1112 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été décalés d'une double croche vers la droite dans la TA.
- 1113
- 1114 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
- 1115
- 1116 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.
- 1117
- 1118

1119 Exemple d'analyse 4

Transcription manuelle \Rightarrow Transcription automatique



1120

- 1121 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une manière excessivement complexe!

conclusion sur ces exemples

1123 Exemple avec des flas

sauf erreur, les "flas" ne sont pas définis. \rightarrow sections 1.4 (appogiatures) et 3.1 (flas)?

3. <http://lilypond.org/>
4. <https://musescore.com/>

1125 Transcription manuelle



1126

1127 Transcription automatique

1128



1129

1130

- 1131 — Le premier fla est reconnu comme étant un triolet contenant une
- 1132 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1133 ornementée.
- 1134 — Le deuxième fla est reconnu comme étant un accord.
- 1135 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1136 dans la TA.
- 1137 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1138 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1139 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1140 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1141 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1142 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1143 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1144 temps 3 de la TA.
- 1145 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1146 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1147 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1148 cross-sticks.

1149 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1150 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1151 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1152 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1153 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1154 *Inria*) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1155 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1156 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1157 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1158 et quatre).

1159 La transcription des données audio et MIDI contenues dans ces fichiers
 1160 a permis une analyse plus approfondie des critères à relever pour chaque
 1161 évènement MIDI et de la manière de les considérer dans un objectif de
 transcription en partition lisible pour un musicien (Voir la section 3.2).

des conclusions sur la 1163
 transcription manuelle?
 difficultés, durée? nb de
 passes... pourquoi LilyPon-
 det pas MuseScore?

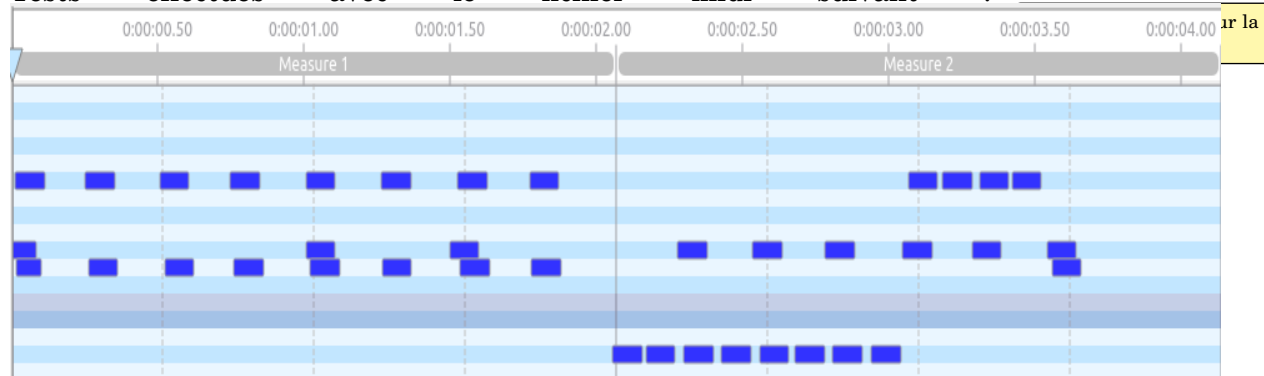
4.3 Transcription polyphonique par parsing (?verrou?)

Les Jams

Les Jams permettent de passer du monophonique au polyphonique.

Le 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

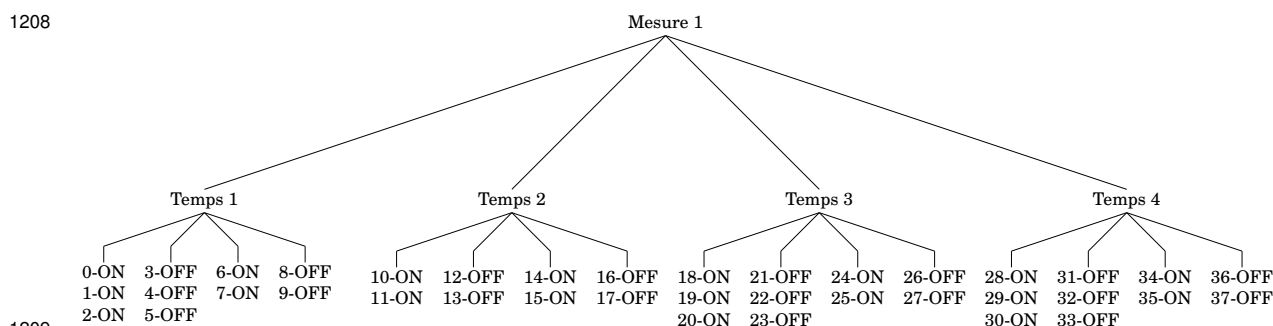
4 -> C0 1

4 -> E1 1

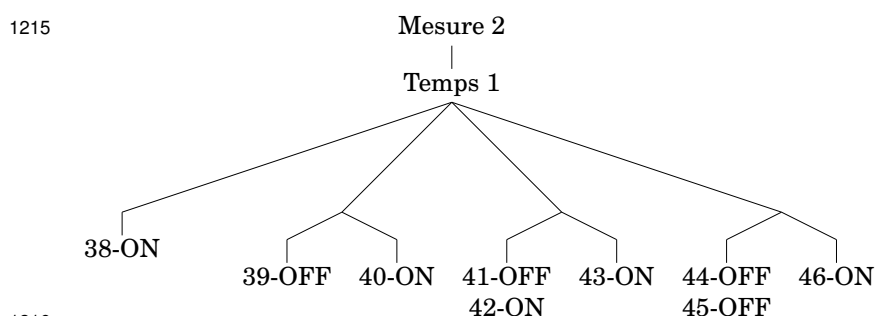
1196 4 -> E2 1
 1197 4 -> T2(6, 6) 1
 1198

1199 // triple level
 1200 6 -> E1 1
 1201

1202 Cette grammaire sépare les ligatures par temps au niveau de la
 1203 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1204 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1205 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1206 il découpe correctement les mesures et les temps.
 1207
 1208



1209
 1210 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1211 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1212 quantification visibles dès le premier temps.
 1213
 1214



1216
 1217 Les Onsets sont correctement triés au niveau des doubles croches
 1218 mais certaines doubles croches sont inutilement subdivisées en triples
 1219 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1220 ci-dessus).
 1221

1222 2ème exemple :

1223 Après une augmentation du poids des triples croches dans la grammaire
 1224

(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.

4.4 Expérimentation théorique d'un système

Cette expérimentation théorique, basée sur la partition de référence de la figure 4.2, montre le procédé de création d'un *système* et des règles qui en découlent (métrique, choix de grammaire, règles de séparation des voix et de simplification de l'écriture). Le *système* devra ensuite être implémenté pour appliquer des tests qui seront effectués, dans un premier temps, sur la partition de référence.

Le titre est contradictoire, et l'explication pas très claire

Motifs et gammes



FIGURE 4.3 – Motifs et gammes

Motifs

À partir de la partition de référence, les deux motifs de la figure 4.3 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à la mesure 28 avec des variations. Ces deux motifs sont très classiques et pourront être détectés dans de nombreuses performances.

Gammes

Les gammes de la figure 4.3 étayent toutes les combinaisons d'un motif en 4/4 binaires jusqu'aux doubles croches.

1247 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-
 1248 mière ne contient que des noires et la deuxième que des croches en l'air.
 1249 Ces deux possibilités sont combinées de manière circulaire dans les 3 me-
 1250 sures de la deuxième ligne.
 1251 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures
 1252 dont la première ne contient que des croches et la deuxième que des
 1253 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1254 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1255 **Systèmes — motifs et gammes combinés**

1256 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1257 de la figure 4.3.

FIGURE 4.4 – Partition d'un système en 4/4 binaire

1258

1259 **Représentation du système en arbres de rythmes**

1260 L'arbre de la figure 4.5 servira de base pour la suite de l'expérimentation.
 1261 Comme indiqué à la racine de l'arbre, il représente la première ligne de la
 1262 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,
 1263 il manque des indications de notation telles que les voix spécifiques à
 1264 chaque partie du rythme ainsi que les choix d'écriture pour les distances
 1265 qui séparent les notes de chaque voix entre elles en termes de durée.

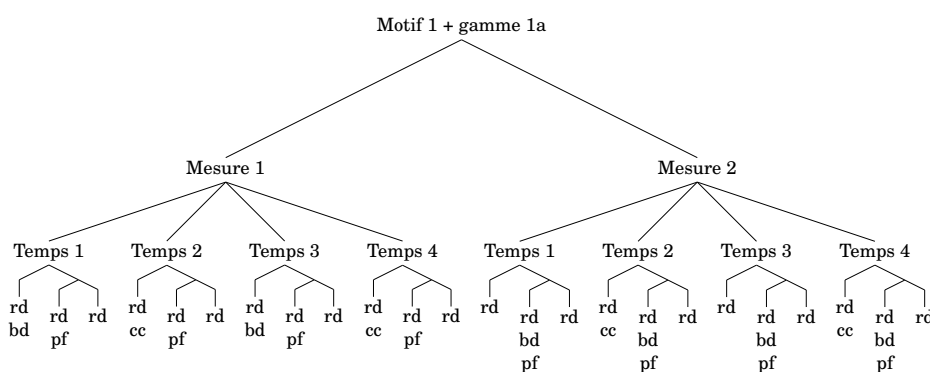


FIGURE 4.5 – Arbres de rythme — système

1266 Réécriture — séparation des voix et simplification

1267 La séparation des voix

1268 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui
 1269 le constituent et les voix seront regroupées en suivant les règles du sys-
 tème.

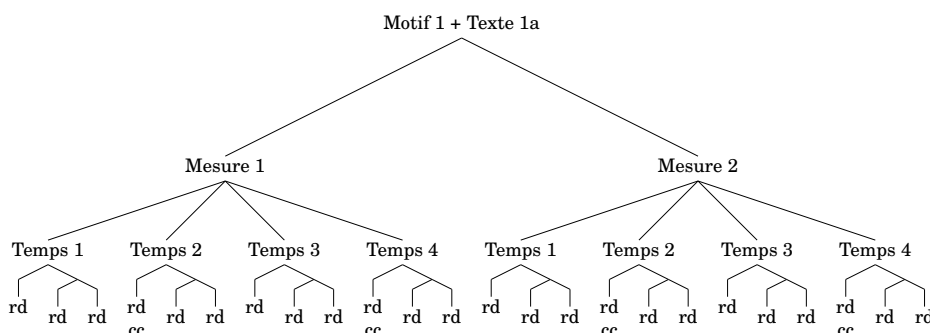


FIGURE 4.6 – Arbres de rythme — voix haute

1270
 1271 La voix haute regroupe la ride et la caisse-claire sur les ligatures du haut.
 1272
 1273 La voix basse regroupe la grosse-caisse et le charley au pied sur les liga-
 1274 tures du bas.

1275 Les règles de simplifications

1276 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1277 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1278 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1279 entre elles (rendre la pulse visuelle).

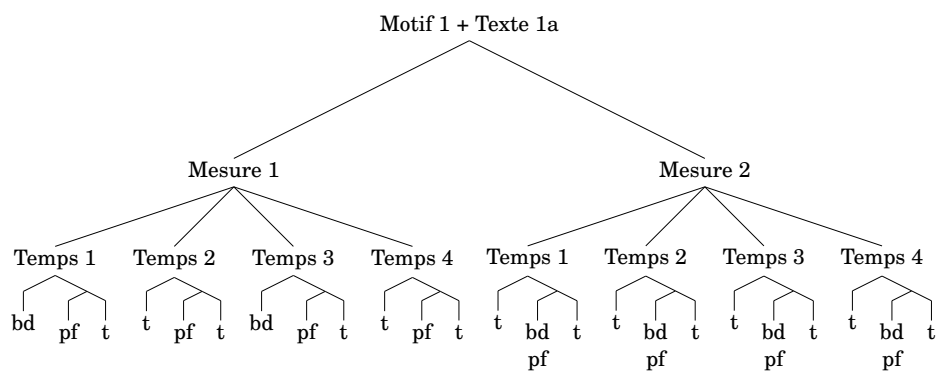


FIGURE 4.7 – Arbre de rythme — voix basse

1280

1281 Pour les figures ci-dessous :

1282 — x = une note ;

1283 — r = un silence ;

1284 — t = une continuation (point ou liaison)

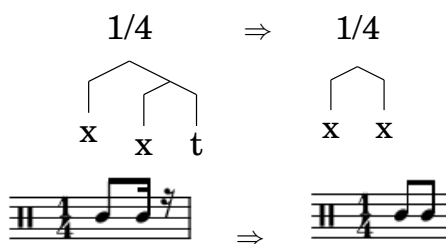


FIGURE 4.8

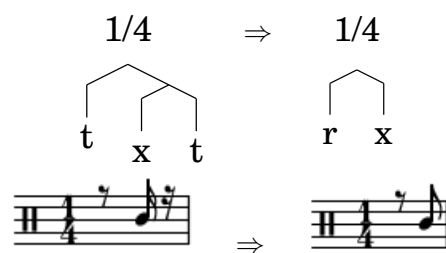


FIGURE 4.9

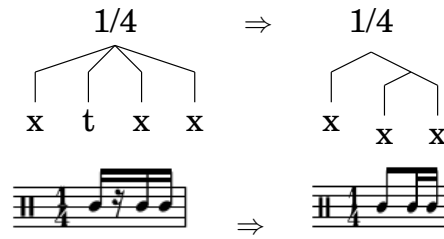


FIGURE 4.10

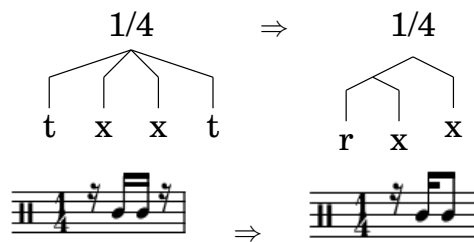


FIGURE 4.11

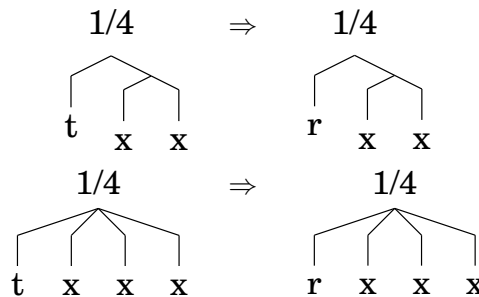


FIGURE 4.12

1285 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
 1286 manquants seront mis en annexe.

1287 Les règles remplacent par un silence les continuations (t) qui sont au
 1288 début d'un temps. Cela est valable pour ce système mais lorsqu'il y a des
 1289 ouvertures de charley, cela n'est pas toujours applicable. Ce problème est
 1290 évoqué de le chapitre 3.

1291 ⇒ **Objectif de cette expérimentation théorique :**

1292 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
 1293 expérimentation théorique a pour but d'orienter la recherche d'autres
 1294 systèmes par observation du jeu de données et de montrer comment les
 1295 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

4.5 BILAN : résultats — évaluation — discussion

Cette section regroupe les avancées qui ont été réalisées par rapport aux objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résultats de l'ADT avec Qparse. Nous avons amélioré le système de quantification de Qparse pour la batterie, notamment le passage à la polyphonie avec les Jams.

Nous avons pu obtenir des arbres de parsing corrects en améliorant les grammaires avec des fichiers MIDI courts. Puis, une sortie MEI a été aussi obtenue (encore à vérifier).

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

- Le choix de travailler avec Lilypond et non Verovio. Ce choix était motivé par la liberté totale concernant la notation de la batterie dont un et la disponibilité d'un set de notation de type Agostini. C'est la seule application qui me permettait d'écrire la notation de la batterie exactement comme je le souhaitais.
- Avancé de la chaîne de traitement (nous sommes arrivés aux arbres de parsing, nous avons traité le polyphonique (identification des regroupements de notes⁵) ⇒ Quelques arbres ont été obtenus sur des exemples simples⁶)
- 2 dimensions de le travail fourni :
 - La volonté de pousser un exemple simple jusqu'au bout de la chaîne pour obtenir des résultats et une évaluation sur au moins un exemple ;
 - La réalité du travail à fournir pour faire avancer sur la chaîne de traitement. ⇒ Une solution aurait été de considérer les arbres de parsing obtenus après le traitement du polyphonique comme un résultat local possible à évaluer au lieu d'attendre que la chaîne arrive jusqu'à la génération d'une partition mais cela n'était pas prioritaire pendant le stage.
- Création d'un jeu de système basique représentatif des différents styles à recouvrir. Ce jeu n'a pas pu être créé, car comme vu plus haut, je me suis focalisé sur un exemple pour pouvoir le vérifier entièrement et dans l'espoir de pouvoir le tester en fin de chaîne. **Évaluation** Matcher les motifs aurait été indispensable pour obtenir une quantité de résultats qui justifieraient une évaluation automatique permettant de faire des graphiques. L'évaluation fut entièrement manuelle car :
 - ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf VS output) Pour l'évaluation, il aurait fallu produire un module.

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

je vois 2 problématiques
et contrib. principales : 1)
transcription polyphonique
par parsing (verrou) : jams
etc 2) réécriture, pour sé-
paration en voix et sim-
plification, aidée (guidée)
par système. Ce serait bien
de présenter la contrib. 1
dans une section (comme
4.3 pour 2), avant d'aborder
résultats et discussion

1337 L'évaluation est-elle automatique ou manuelle?
1338 Possibilité d'un export lilypond en arbre pour comparer l'output
1339 avec la transcription manuelle.
1340 Possibilité de transformer lilypond(output) et lilypond(ref) en
1341 ScoreModel ou MEI pour les comparer et faire des statistiques.
1342 Si transformés en MEI : diffscore de Francesco. Possibilité de
1343 transformer lilypond(output) et lilypond(ref) en MusicXML pour
1344 les comparer ou dans Music21. L'expérimentation peut-être consi-
1345 dérer comme une évaluation manuelle? (magicien d'Oz)
1346 Lilypond vers MIDI + ouput vers MIDI \Rightarrow Comparaison des MIDI
1347 dumpés.
1348
1349 La transcription automatique de la batterie est un sujet passionnant mais
1350 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-
1351 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-
1352 moins été générée. Elle sera un bon point de départ pour les travaux fu-
1353 turs dont plusieurs propositions sont énoncés dans le présent document.

CONCLUSION GÉNÉRALE

1355 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1356 tion automatique de la batterie. Son objectif était de transcrire, à partir
1357 de leur représentation symbolique MIDI, des performances de batteur de
1358 différents niveaux et dans différents styles en partitions écrites.
1359 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1360 cessus de regroupement des événements MIDI qui nous a permis de faire
1361 la transition du monophonique vers le polyphonique. Une des données
1362 importante de ce processus était de différencier les nature des notes d'un
1363 accord, notamment de distinguer lorsque 2 notes constituent un accord
1364 ou un *fla*.
1365 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1366 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1367 tionnables en amont du parsing, soit par indication des noms des fichiers
1368 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1369 naire de patterns prédéfinis ⁷ qu'il serait pertinent de mettre en œuvre en
1370 machine learning.
1371 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1372 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1373 d'un système et pour la simplification grâce aux gammes du motif d'un
1374 système. Nous avons aussi montré comment, dans des travaux futurs, un
1375 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1376 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1377 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1378 la chaîne de traitement en terme de réécriture.
1379 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1380 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1381 entièrement automatisée de partitions de batterie et au problème général
1382 de l'AMT de bout en bout.[11]

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

BIBLIOGRAPHIE

- 1384 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1385 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1386 1996. – Cité pages 7, 30 et 36.
- 1387 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1388 et 15.
- 1389 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1390 page 12.
- 1391 [4] Gaël Richard et al. De fourier à la reconnaissance
1392 musicale. Available at [https://interstices.info/
1393 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1394 – Cité page 12.
- 1395 [5] Caroline Traube. Quelle place pour la science au sein de la musico-
1396 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1397 [6] Leonard Bernstein Office. The unanswered question : Six talks at
1398 harvard. Available at [https://leonardbernstein.com/about/
1399 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1400 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1401 entre les traitements de la musique et du langage. *La Lettre des
1402 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1403 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1404 Techniques de traitement automatique du langage naturel appli-
1405 quées aux représentations symboliques musicales. In *JIM 2021 -
1406 Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1407 Cité page 13.
- 1408 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-
1409 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1410 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1411 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1412 Cité page 13.
- 1413 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1414 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1415 lenges and future directions. *Journal of Intelligent Information Sys-*
1416 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 63.
- 1417 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1418 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1419 Cité pages 14 et 27.
- 1420 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-
1421 SURE, 1997. – Cité page 15.
- 1422 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1423 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1424 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1425 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1426 1483, 2018. – Cité pages 15, 23 et 27.
- 1427 [15] Moshekwa Malatji. Automatic music transcription for two instru-
1428 ments based variable q-transform and deep learning methods, 10
1429 2020. – Cité page 22.
- 1430 [16] Antti J. Eronen. Musical instrument recognition using ica-based
1431 transform of features and discriminatively trained hmms. *Seventh*
1432 *International Symposium on Signal Processing and Its Applications*,
1433 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1434 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1435 drum sound description for real-world music using template adap-
1436 tation and matching methods. *International Conference on Music*
1437 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1438 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1439 musical statistics as guides for audio-to-score piano transcription.
1440 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24 et 26.
- 1441 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1442 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1443 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1444 *Computation in Music*, volume Lecture Notes in Computer Science
1445 of *Proceedings of the Seventh International Conference on Mathema-*
1446 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1447 2019. Springer. – Cité pages 24 et 26.
- 1448 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1449 ring of rhythm structures. In *Proceedings of the First International*
1450 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1451 IEEE Computer Society. – Cité page 26.
- 1452 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1453 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1454 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1455 Cité page 26.
- 1456 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1457 equivalent rhythmic notations based on rhythm tree languages. In
1458 *Third International Conference on Technologies for Music Notation
1459 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1460 pez Palma and Mike Solomon. – Cité page 26.
- 1461 [23] R. Marxer and J. Janer. Study of regularizations and constraints in
1462 nmf-based drums monaural separation. In *International Conference
1463 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1464 02/09/2013 2013. – Cité page 27.
- 1465 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1466 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 30.
- 1467 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1468 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 31.
- 1469 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1470 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1471 94120 Fontenay-sous-Bois, 2000. – Cité pages 32 et 43.
- 1472 [27] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1473 Bamman. Learning to groove with inverse sequence transforma-
1474 tions. In *International Conference on Machine Learning (ICML)*,
1475 2019. – Cité page 47.

