

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 La transcription manuelle	37
34	3.3 Modélisation pour la transcription	38
35	3.4 Qparse	40
36	3.5 Les forme rythmiques	42
37	4 Expérimentations	49
38	4.1 Le jeu de données	50
39	4.2 Analyses et transcriptions manuelles	51
40	4.3 Transcription polyphonique par parsing	55
41	4.4 Réécriture guidée par une forme rythmique	58
42	4.5 BILAN : résultats — évaluation — discussion	63
43	Conclusion générale	65
44	Bibliographie	67

LISTE DES FIGURES

46	1.1	Exemple évènements avec durée	16
47	1.2	Critère pour un évènement	17
48	1.3	Exemple évènements sans durée	17
49	1.4	Les silences	18
50	1.5	18
51	1.6	Rapport des figures de notes	18
52	1.7	Exemple de partition de piano	19
53	1.8	MusicXML	20
54	2.1	Transcription automatique <dam>remettre ici la citation de la	
55		capture d'écran avec la page</dam>	23
56	2.2	HMM	26
57	2.3	arbre_jazz	27
58	3.1	Les instruments de la batterie	30
59	3.2	Hauteur et têtes de notes	30
60	3.3	Point et liaison	32
61	3.4	Silence joué	33
62	3.5	Équivalence	34
63	3.6	Séparation des voix	35
64	3.7	Les accents et les ghost-notes	35
65	3.8	Exemple pour les accentuations et les ghost-notes	36
66	3.9	Définition du fla	36
67	3.10	lilypond — extraits de code	37
68	3.11	lilypond — transcription manuelle	38
69	3.12	Présentation de Qparse	41
70	3.13	signature rythmique	43
71	3.14	Motif 4-4 binaire	44
72	3.15	Motif 4-4 jazz	45
73	3.16	forme rythmique 4-4 afro-latin	45
74	3.17	Simplification	46
75	3.18	47
76	4.1	Batterie électronique	50
77	4.2	Motifs et gammes	58
78	4.3	Partition d'un forme rythmique en 4/4 binaire	59
79	4.4	Arbre de rythme — forme rythmique	60
80	4.5	Arbre de rythme — voix haute	60
81	4.6	Arbre de rythme — voix basse	61

82	4.7	61
83	4.8	62
84	4.9	62
85	4.10	62

86 **LISTE DES TABLEAUX**

87	1.1	speechToText vs AMT	13
88	3.1	Noms des instruments de la batterie	30
89	3.2	Codes, identités et pitchs des instruments	39
90	3.3	Sytèmes	43

INTRODUCTION GÉNÉRALE

92 QUOI?

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

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

114

115 POURQUOI?

- 116 — sujet traité : la batterie
- 117 — intérêt spécifique de la génération de partition de batterie compa-
 118 rativement au autres instrument
- 119 — patrimoine
- 120 — rapidité de génération (musicien ou enseignement)
- 121 — ...

122

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

125 COMMENT?

126 → Problématique :

127 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 <dam>que de choix de représentation à réaliser?</dam>. De plus, la batterie est dotée d'une écriture spécifique par rapport à la majorité des instruments.

137

138 → Méthodes :

139 → Contributions :

140 <louison>liste des contributions : donner une échelle, un point de compa-
141 raison, du contexte, pour pouvoir mesurer l'importance de chaque contri-
142 bution</louison>

143 La proposition principale de ce mémoire est basée sur la recherche de
144 rythmes génériques sur l'*input*. Ces rythmes sont des *patterns* standards
145 de batterie définis au préalable et accompagnés par les différentes combi-
146 naisons qui leur sont propres. On les nomme systèmes (voir sections 3.5,
147 ??). L'objectif des systèmes est de fixer des choix le plus tôt possible afin
148 de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces
149 choix concernent notamment la métrique et les règles de réécriture.

150

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

- 152 — une modélisation de la notation de la batterie (fusion de 3.1 et de
153 3.3) qui était jusqu'à présent inexistante.
- 154 — plusieurs transcriptions manuelles dans le but d'analyser les conte-
155 nus des fichiers MIDI et Audio (??) et de faire des comparaisons de
156 transcription avec des outils déjà existants ¹.
- 157 — une partition de référence transcrite manuellement sur l'entièreté
158 d'une performance du jeu de données afin de repérer les éléments
159 importants pour la modélisation et de faire les liens entre les cri-
160 tères des données d'*input* avec l'écriture finale (3.11). Cette par-
161 tition avait aussi pour objectif d'effectuer des tests et des évalua-
162 tions.
- 163 — le passage au polyphonique en théorie et en implémentation im-
164 pliquant la théorie sur la détection de l'identité de notes dans un
165 Jam ² et l'implémentation de tests unitaires sur le traitement des
166 Jams (4.3).
- 167 — 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

168

169 L'ensemble de ces sous-tâches a permis deux réalisations principales :

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

182

183 PLAN

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

CONTEXTE

Sommaire

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

Introduction

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

Dans ce chapitre, nous parlerons de l'informatique musicale, nous montrerons les liens existants entre le RIM et le TAL ainsi qu'entre les notions de langage musical et langue naturelle. Nous traiterons également du problème de l'AMT et de ses applications.

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

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

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

220 1.1 Langues naturelles et musique en 221 informatique

222 COMPUTER MUSIC

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

231

232 RIM

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

242

243 RIM et TAL

244 Aborder la musique comme un langage avec des méthodes de TAL né-
245 cessite une réflexion autour de la musique en tant que langage ainsi que
246 la possibilité de comparer ce même langage avec les langues naturelles.
247 Léonard Bernstein [6] a donné une série de six conférences publiques
248 à Harvard fondées en grande partie sur les théories linguistiques que
249 Noam Chomsky a exposées dans son livre « Language and Mind ». Lors
250 de la première conférence, qui a eu lieu le 9 octobre 1973, Bernstein a
251 avoué être hanté par la notion d'une grammaire musicale mondiale innée
252 et il analyse dans ses trois premières conférences, la musique en termes
253 linguistiques (phonologie, syntaxe et sémantique). Quelques travaux en
254 neurosciences ont également abordé ces questions, notamment par ob-
255 servation 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 ⇒ phonèmes ⇒ texte signal ⇒ notes, accords ⇒ 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*).

290 1.2 La transcription automatique de la musique

291 1. OBJECTIF

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

303

304 2. APPLICATIONS

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

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

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

322

323 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

324 L'analyse de la structure hiérarchique des séquences d'accords par utili-
325 sation de modèles grammaticaux s'est avérée très utiles dans les analyses
326 récentes de l'harmonie du jazz [12]. Comme déjà évoqué précédemment, il
327 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

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

371 1.4 Les représentations de la musique

citer M. Müller FMP pour
 cette section ?

372 Les données audio

373 trop technique. ne pas re-
 374 pier wikipédia

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

378 LPCM pas utile ici. parle
 379 juste échantillons et com-
 pression.

379 Les données MIDI

tu peux mentionner le for-
 mat spectral (analyse har-
 monique) crucial en MIR
 audio.

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

386 ne pas copier wikipédia ver-
 387 batim. 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)

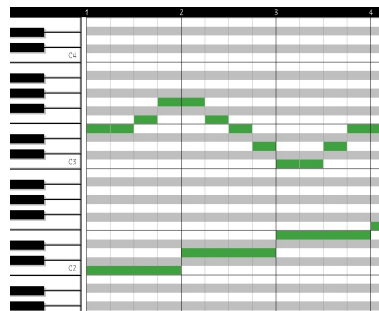


FIGURE 1.1 – Exemple événements avec durée

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

390 il n'y a pas de duration
 d'événement dans un MIDI
 file. la "durée" est une dis-
 tance entre 2 évènements
 ON et OFF (c'est important
 dans ton travail). le screen
 shot n'est pas utile, écrit
 plutôt une liste itemize

391 Pour la batterie, les événements sont considérés sans durée, nous ignore-
 392 rons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le
 393 *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>

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

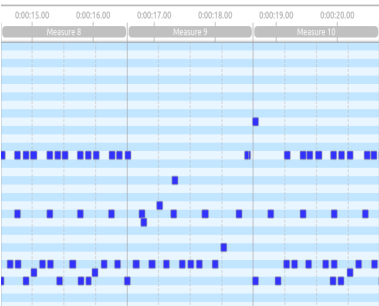


FIGURE 1.3 – Exemple évènements sans durée

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

Les partitions

Expliquer aussi la notation générale des silences

La figure 1.5 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 ;
- La blanche, elle vaut 2 ;
- La noire, elle vaut 1 ;
- La croche, elle vaut 1/2.

pour clarifier 3.1(sub les durées), décrire en 1.4 (ici) la notation conventionnelles (piano etc) et 3.1(sub les durées) uniquement ce qui est spécifique à la batterie, en expliquant les différences.

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

4 temps

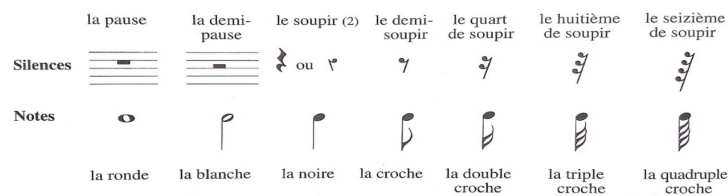


FIGURE 1.4 – Les silences



FIGURE 1.5

plutôt que wikipedia cite
Dannhauser ou autre ref.
F.M. ou encore Gould 2014
Behind Bars

Une figure de note [1] de musique combine plusieurs critères ¹⁵ :

- 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 :
barre verticale liée à la tête de note Indicatrice d'appartenance à une voix en fonction de sa direction (haut ou bas) et indicatrice d'une durée représentée par sa présence ou non (blanche ≠ ronde)
- Un crochet : La durée d'une note est divisée par deux à chaque crochet ajouté à la hampe d'une figure de note.

plusieurs éléments

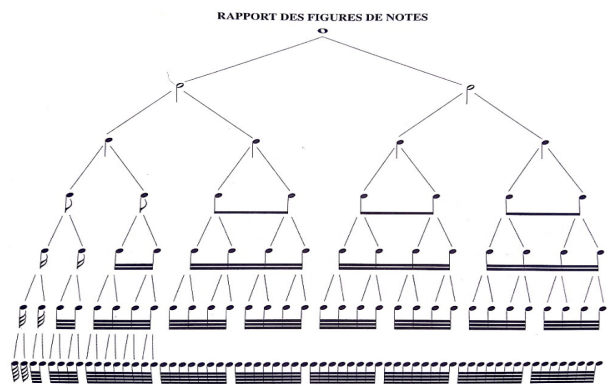


FIGURE 1.6 – Rapport des figures de notes
[1]

La figure 1.6 montre les rapports de durée entre les figures de notes. Plus les durées sont longues, plus elles sont marquées par la tête de note ou la présence ou non de la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute un crochet à la hampe d'une figure de notes

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

429 pour diviser sa durée par 2. Les notes à crochet (croche, double-croche,
430 triple-croche...) peuvent être reliées ou non par des ligatures (voir les 4
431 dernières lignes de la figure 1.6).



FIGURE 1.7 – Exemple de partition de piano

432 Une partition de musique¹⁶ est un document qui porte la représentation
433 systématique du langage musical sous forme écrite. Cette représentation
434 est appelée transcription et elle sert à traduire les quatre caractéristiques
435 du son musical :

- 436 — la hauteur ;
- 437 — la durée ;
- 438 — l'intensité ;
- 439 — le timbre.

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

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

447 Les formats XML

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

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

453 Le figure 1.8¹⁷ représente un do en clef de sol de la durée d'une ronde
454 sur une mesure en 4/4 écrit au format MusicXML. Un des avantages de
455 ce format est qu'il peut être converti aussi bien en données MIDI qu'en
456 partition musicale, ce qui en fait une interface homme/machine.

457 appoggiatures

458 <flo>Parler des appoggiatures ici?</flo>

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

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

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

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

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

```

<?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>

```

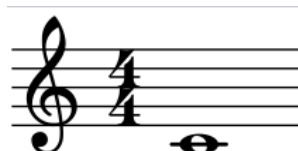


FIGURE 1.8 – MusicXML

signature rythmique

<flo>présenter rapidement la notation des signatures rythmiques</flo>

Conclusion

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

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

Et enfin, nous avons décrit les représentations de la musique nécessaires 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.

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

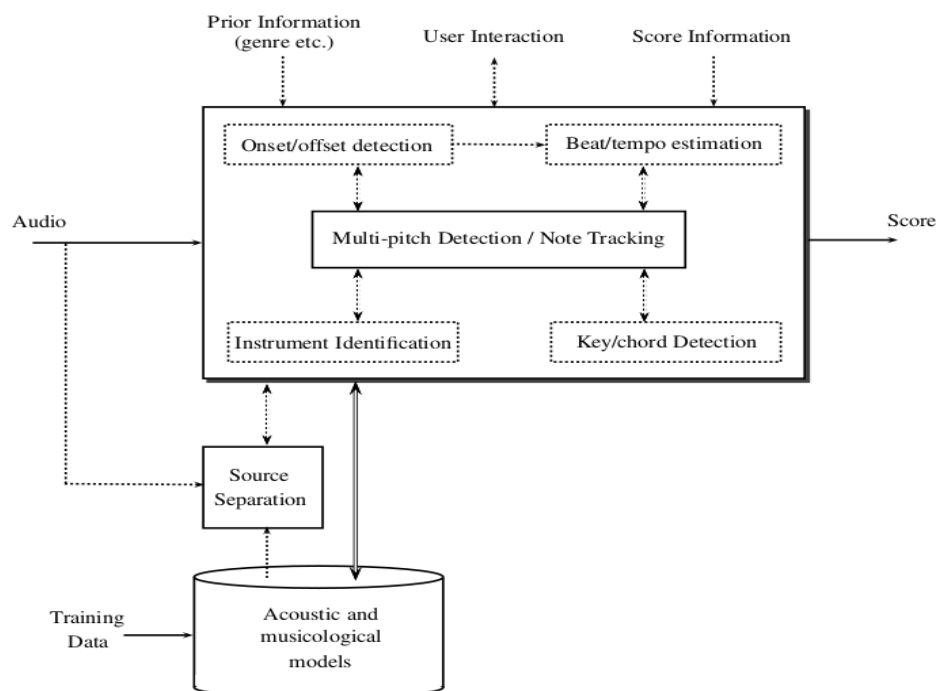


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

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.

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

haute fréquence, aigus?

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

classification des évène- 558
 ments? la phrase semble 559
 redondante

pas clair... peut-être just 562
 mentionner les modèles 563
 probabilistes utilisés

566 2.3 MIDI vers partition

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

ce n'est pas exactement 567
 cela. cf. proposition de des- 568
 cription + détaillée en com- 569
 mentaires

de manière conjointe 578

langage a priori 580

qui nécessite de traiter le 583
 problème supplémentaire 584
 de la séparation de voix. i.e. 585
 pour la batterie on nveut 586
 quantification + structu- 587
 ration + séparation mais 588
 seules les 2 premières sont 589
 couplées dans l'approche de 590
 tonn stage.

587 2.4 Approche linéaire et approche hiérarchique

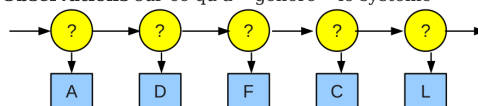
588 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par
 589 exemple, Shibata *et al.* [18] ont utilisé le modèle de Markov caché (HMM)⁴
 590 pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux

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

591 réseaux de neurones profonds, l'un pour la reconnaissance des pitches et
592 l'autre pour la reconnaissance de la vélocité. Pour la dernière couche, la
593 probabilité est obtenue par une fonction sigmoïde. Ils construisent en-
594 suite plusieurs HMM métriques étendus pour la musique polyphonique
595 correspondant à des métriques possibles, puis ils calculent la probabilité
596 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

597 *Source : Cours de Damien Nouvel*⁵

598

599

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

je ne comprend pas bien 607
 l'explication. le pb est plu- 608
 tot vue locale (déduction 609
 la proba d'une durée à par- 610
 tir de la durée précédente, 611
 par ex. dans un HMM) vs
 vue globale, dans une hié-
 rarchie

RT? 612
 613

techniques de réécriture 614
 appliquée à la déduction 615
 automatique, calcul symbo-
 lique 616

le calcul d'équiv. 617
 618

citer thèse de David Rizo 621
 (Valencia) 622

623

le langage musical est lui-même structuré. En effet, l'usage d'arbres syn-
 taxiques est idéale pour représenter le langage musical. Une méthodolo-
 gie simple pour la description et l'affichage des structures musicales est
 présentée dans [20]. Les RT y sont évoqués comme permettant une co-
 hésion complète de la notation musicale traditionnelle avec des notations
 plus complexes. Jacquemard *et al.* [21] propose aussi une représentation
 formelle du rythme, inspirée de modèles théoriques antérieurs issus du
 domaine de la réécriture de termes. Ils démontrent aussi l'application des
 arbres de rythmes pour les équivalences rythmiques dans [22]. La réécri-
 ture d'arbres, dans un contexte de composition assistée par ordinateur,
 par exemple, pourrait permettre de suggérer à un utilisateur diverses
 notations possibles pour une valeur rythmique, avec des complexités dif-
 férentes.

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.nouvels.net/fr/enseignement>

Example: *Summertime*

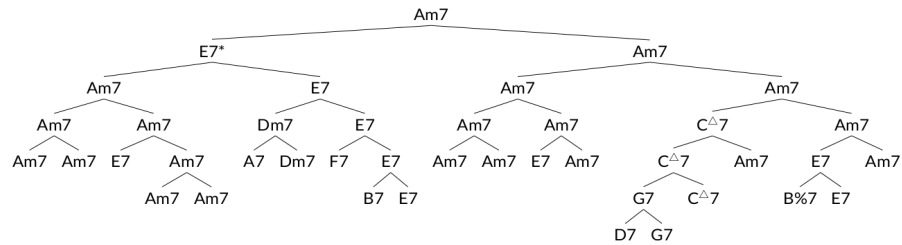


FIGURE 2.3 – arbre_jazz
Représentation arborescente d’une grille harmonique [12]

Conclusion

La plupart des travaux déjà existants sur l’ADT ont été énumérés par Wu *et al.* [14] qui, pour mieux comprendre la pratique des systèmes d’ADT, se concentrent sur les méthodes basées sur la factorisation matricielle non négative et celles utilisant des réseaux neuronaux récurrents. La majorité de ces recherches se concentre sur des méthodes de calcul pour la détection d’événements sonores de batterie à partir de signaux acoustiques ou sur la séparation entre les évènements sonores de batterie avec ceux des autres instruments dans un orchestre ou un groupe de musique [23], ainsi que sur l’extraction de caractéristiques de bas niveau telles que la classe d’instrument et le moment de l’apparition du son. Très peu d’entre eux ont abordé la tâche de générer des partitions de batterie et, même quand le sujet est abordé, l’output final n’est souvent qu’un fichier MIDI ou MusicXML et non une partition écrite.

Il n’existe pas de formalisation de la notation de la batterie ni de réelle génération de partition finale, dont les enjeux principaux seraient :

- 1) le passage du monophonique au polyphonique, comprenant la distinction entre les sons simultanés et les flas ou autres ornements ;
- 2) les choix d’écritures spécifiques à la batterie concernant la séparation des voix et les continuations.

à ma connaissance, aucun des travaux en nADT ne produit de partition XML

diff. pour production de partition (et 1 des obj. du stage) est...

latex : enumerate

MÉTHODES

Sommaire

3.1	La notation de la batterie	29
3.2	La transcription manuelle	37
3.3	Modélisation pour la transcription	38
3.4	Qparse	40
3.5	Les forme rythmiques	42

Introduction

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 qparse¹, un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et plusieurs développeurs au sein du laboratoire Cedric au CNAM.

Enfin, nous présenterons les forme rythmiques, <flo>une représentation théorique qui permet. . .</flo>

3.1 La notation de la batterie

Pour la transcription, j'ai choisi d'utiliser une notation inspirée du recueil de pièces pour batterie de J.-F. Juskowiak [24] et des méthodes de batterie Agostini [25], car je trouve la position des éléments cohérente et intuitive (voir section 3.1).

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

673 Les hauteurs et les têtes de notes



FIGURE 3.1 – Les instruments de la batterie

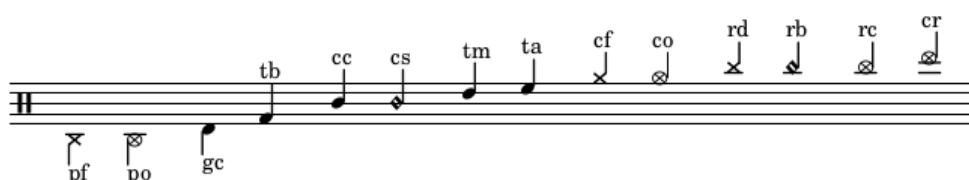


FIGURE 3.2 – Hauteur et têtes de notes

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

TABLE 3.1 – Noms des instruments de la batterie

674 La figure 3.1² montre une batterie standard avec tous les instruments
 675 habituellement présent sur une batterie et la figure 3.2 donne leur repré-
 676 sentation sur une partition.
 677 Le tableau 3.1 donne dans l'ordre :

2. Source : <https://www.superprof.fr/blog/composition-instrument-percussion/>

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

Les figures 3.1, 3.2 et le tableau 3.1 peuvent aider à comprendre pourquoi je trouve la notation agostinienne cohérente et intuitive.

En effet, les hauteurs sur la portée représentent :

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

Les durées

Comme nous venons de la voir, la majorité des instruments de la batterie sont représentés par les têtes des notes. De plus, le seul instrument dont le son peut être arrêté de manière quantifiée et dont la durée sonore nous intéresse est le charley³.

Par conséquent :

1. les durées — sauf pour le charley — représenteront un écart temporel entre les notes et non une durée sonore et elles pourront donc être rallongée à l'aide de silences ;
2. les symboles rythmiques concernant les têtes de note ne pourront pas être utilisés pour exprimer les durées. Cela est valable aussi pour la présence ou non de la hampe puisque ce phénomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition blanche-ronde). L'usage des blanches existe dans certaines partitions de batterie [26] mais cela reste dans des cas très rares. Certains logiciels permettent de faire des blanches avec des symboles spécifiques à la batterie ou aux percussions mais leur lecture reste peu aisée et leur utilisation pour la batterie est rarissime.

certaines têtes de notes vides alors que leur durée n'est pas celle des blanches ? expliquer les différences avec la notation conventionnelle cf 1.4

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

716 En résumé :

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

724

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

- 726 — Le point : il rallonge la durée d'une note de la moitié de sa valeur.
- 727 Dans la deuxième note de l'exemple 3 de la figure 3.3 est une noire
- 728 pointée, elle vaut donc la durée d'une noire + une croche (ou de trois
- 729 croche) ;
- 730 — La liaison : elle rallonge la durée de la première note de la durée de
- 731 la deuxième. La deuxième note de l'exemple 4 de la figure 3.3 est
- 732 une croche qui est liée à une noire, sa durée est donc équivalente à
- 733 celle d'une croche + une noire (ou de trois croches) ;
- 734 — les silences (pas pour les ouvertures de charley).



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

735 Un autre élément concernant la notation des durées en batterie est la né-
 736 cessité de faire ressortir la pulsation⁴ de la rendre visuelle. La première
 737 chose à prendre en compte pour analyser la figure 3.3 est donc la néces-

4. La position des temps

sité de regrouper les notes par temps à l'aide des ligatures. Le deuxième point est de s'arranger pour qu'il y ait une indication visuelle au début de chaque temps.

— Exemple 1 : l'ouverture de charley est quantifiée mais les notes ne sont pas regroupées par temps.

— Exemple 2 : Ici, la liaison permet de regrouper les notes par temps en obtenant le même rythme que dans l'exemple 1.

— Exemple 3 et exemple 4 : les deux exemples sont valables mais le deuxième est le plus souvent utilisé car la liaison donne un repère visuel sur le temps.

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

Les silences

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

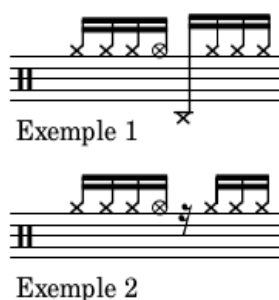


FIGURE 3.4 – Silence joué

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

767 rythme et elle n'empiète pas sur une potentielle voix basse qui pourrait le
768 compléter (on évite une écriture surchargée).

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

771 1. si c'est un charley fermé joué à la main \Rightarrow la note sera un charley
772 fermé joué à la main (cf) ;

773 2. si c'est un charley fermé joué au pied \Rightarrow la note sera un silence.

774 La deuxième règle sera soumise au cadre imposé par certaines formes
775 rythmiques pour lesquelles le charley joué au pied devra rester tel quel.

776 Les équivalences rythmiques

777 Pour les instruments mélodiques, dans le cas de notes dont la durée de
778 l'une à l'autre est ininterrompue et si leur durée initiale est prolongée,
779 seuls la liaison et le point permettent des notations équivalente. Mais
780 pour la batterie et à part dans le cas des ouvertures de charley (voir sec-
781 tion 3.1), seules comptent des dates de début (onsets) : la durée du son
782 n'a pas d'importance. L'usage des silences pour combler la distance ryth-
783 mique entre deux notes devient donc possible.

784 Cela pris en compte, et étant donné que les indications de durée dans les
785 têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide
786 de silences sera privilégiée comme indication de durée sauf dans les cas
787 où cela reste impossible. Ce choix à pour but de n'avoir qu'une manière
788 d'écrire toutes les notes, quelles que soient leur tête de note (sauf pour le
789 charley).



FIGURE 3.5 – Équivalence

790 Sur la figure 3.5, théoriquement, il faudra choisir la notation de la
791 deuxième mesure mais dans certains contextes, pour des raisons de lisi-
792 bilité ou de surcharge, la version sans les silences de la troisième mesure
793 pourra être choisie.

794 Les voix

795 Pour les instruments mélodiques, un groupe de notes peut être organisé
796 en *voix*, représentant des flots mélodiques joués en parallèle, avec une
797 synchronisation plus ou moins stricte [18] [27].

798 En batterie, une voix est théoriquement l'ensemble des instruments qui,
 799 à eux seuls, constituent une phrase rythmique. Mais en pratique, les ins-
 800 truments peuvent aussi être divisés par voix dans le but de ne pas sur-
 801 charger la notation ou pour que leur disposition soit représentée sur la
 802 partition (voir section 3.1). Les voix sont caractérisées par l'orientation
 803 des hampes et plus précisément par les ligatures si les hampes sont dans
 804 la même direction (voir figure 3.16).

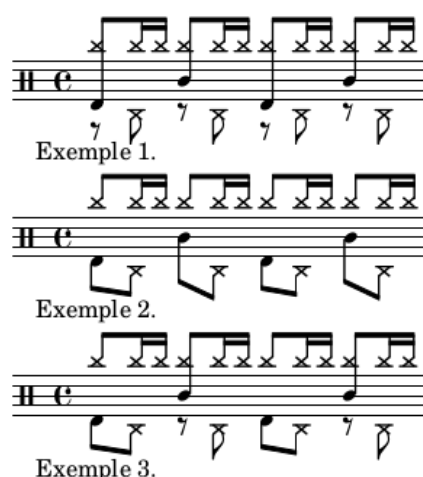


FIGURE 3.6 – Séparation des voix

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

810 Les accentuations et les ghost-notes

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



FIGURE 3.7 – Les accents et les ghost-notes

Théoriquement, tous les instruments peuvent être accentués (voir la section 3.3), mais la figure 3.7 représentent ceux dont les accents ne demandent pas un grand niveau de maîtrise et sont presque toujours bien articulés. En outre, les instruments qui ne sont pas représentés sur cette figure ne sont presque jamais accentués dans les partitions et ne sont pas présents de manière significative dans le GMD.

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

La dernière note de la figure 3.7 montre un exemple de notation pour une ghost note jouée à la caisse claire. Une ghost note [28] est une note de faible volume sonore mais jouée fermement. Les ghost notes servent le plus souvent à donner le débit d'un rythme (ses subdivisions) pour le rendre plus dansant (lui donner plus de « groove » ou de « swing »). Le parenthésage a été choisi car il peut être utilisé sur n'importe quelle note sans changer la tête de note.

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



FIGURE 3.8 – Exemple pour les accentuations et les ghost-notes

837

838 Les flas

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



FIGURE 3.9 – Définition du fla

841

3.2 La transcription manuelle

Mis à part les figures du chapitre 1 et certains exemples d'analyses de la section 4.2, toutes les partitions et figures de ce document ont généré avec lilypond⁵.

Présentation de lilypond

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

En raison de :

- notation agostini
- grande liberté de choix
- ...

Lilypond est actuellement le meilleur de logiciel de gravure musicale pour la batterie.

```

\version "2.20.0"
\language français

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

\include "...\lilypond_style_perso.ly"

up = \drummode {

\override Script.Y-offset = #-1.0

% Mesure 1
s s s s

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

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

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

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

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

```

FIGURE 3.10 – lilypond — extraits de code

Sur la figure 3.10 :

- à gauche : configuration aménagée pour la notation de type agostini.
- à droite : le début de code mesure par mesure pour la voix haute d'une partition (en haut du fichier, inclusion du fichier de config)

5. <http://lilypond.org/index.fr.html>

FIGURE 3.11 – lilypond — transcription manuelle

864 La partition de la figure 3.11 est le résultat du code de la figure 3.10 (la
 865 totalité du code est mis en annexe et est accessible sur le git). Cette parti-
 866 tion a été totalement transcrite manuellement avec lilypond par analyse
 867 des fichiers MIDI et audio correspondants.

868 3.3 Modélisation pour la transcription

869 Les pitches

870 Le tableau 3.2 présente dans l'ordre, les codes des instruments, leur iden-
 871 tité (instrument ou parti d'un instrument — joué avec les mains ou avec
 872 les pieds), le ou les pitches qui lui sont associés.
 873 Plusieurs pitches peuvent parfois désigner le même instrument afin de
 874 pouvoir supporter des kits de batterie plus larges (avec par exemple plu-
 875 sieurs toms basses qui n'auraient pas tous exactement la même sonorité)

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 3.2 – Codes, identités et pitches des instruments

ou simplement de styles différents (pour chaque kits standard, ce sont les mêmes instruments mais de styles différents)⁶. J'ai regroupé les pitches des différents types d'un même instrument dans une seule ligne du tableau portant le nom du type de cet instrument. Ainsi, plusieurs toms basses différents dans les données MIDI deviennent tous un tom basse d'une batterie standard et la partition finale pourra être jouée sur n'importe quel kit de batterie standard.

Malgré le large panel de pitches disponibles, il semblerait qu'aucun pitch ne désigne le charley ouvert joué au pied (« po » de la figure 3.2). Pourtant, dans la batterie moderne, plusieurs rythmes ne peuvent fournir le son du charley ouvert qu'avec le pied car les mains jouent autre chose en même temps. Cela doit en partie être dû à l'utilisation des boîtes à rythmes en MAO qui ne nécessitent pas de faire des choix conditionnés par les limitations humaines (2 pieds, 2 mains, et beaucoup plus d'instruments...)

La vitesse

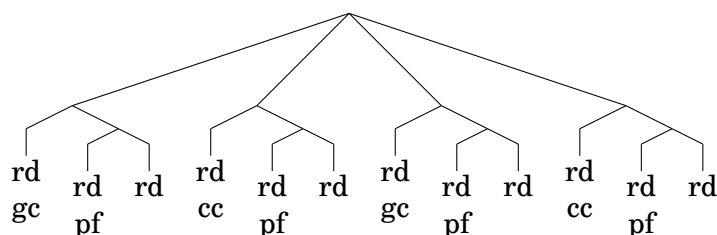
La vitesse déterminera si les notes sont accentuées ou sont des ghost notes. Pour les codes, je propose d'ajouter un suffix (« a » pour accent et « g » pour ghost note) à la fin du code d'une note accentuée ou d'une ghost note. Les choix pour déterminer si les notes sont accentuées ou sont des ghost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

6. Par exemple, les peaux des toms jazz raisonnent alors que les toms rock sont mat.

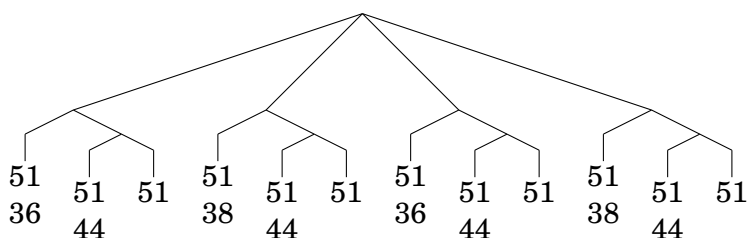
896 Les arbres de rythmes

897 Les arbres de rythmes représentent un rythme dont les possibilités de no-
898 tation sur une partition sont théoriquement multiples. Les branchements
899 sont des divisions d'intervalle temporel, les feuilles sont des évènements
900 musicaux commençant au début de l'intervalle [29] [30] .

901 Voici une représentation qui fonctionne avec les 3 exemples de la figure
3.6 en arbre de rythmes avec les codes de chaque instrument :



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



904 Chacun des trois exemples de la figure 3.6 est représenté par un des deux
905 arbres syntaxiques ci-dessus. <dam>complète un peu en précisant qu'on
906 voit bien ici l'avantage des arbres pour analyser ou construire la structure
907 (les phrases?) musicale</dam>

909 3.4 Qparse

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, qparse
à partir d'une performance
symbolique, séquentielle
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 syn-
taxique.

Le but du parsing/analyse
syntaxique est en effet la
structuration d'une repré-
sentation séquentielle en
entrée (un mot fini), sui-
vant un modèle de langage.

ref. Grune Jacobs "Parsing
..."

La librairie Qparse⁷ implémente la quantification des rythmes basée sur des algorithmes d'analyse syntaxique pour les automates arborescents 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.

Les principaux contributeurs sont :

— Florent Jacquemard (Inria) : développeur principal.

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

- 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

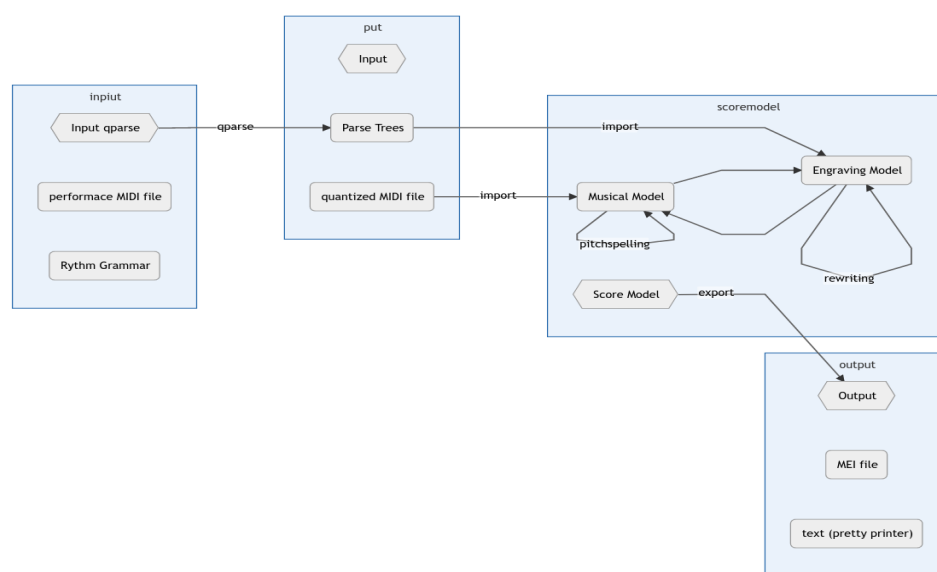


FIGURE 3.12 – Présentation de Qparse

Explication des différentes étapes de la figure 3.12⁸ :

- **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 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 ⇒ un arbre par voix ⇒ représentation in-

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.

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

- 942 termédiaire (RI);
- 943 — Réécriture 2 :
- 944 simplification de l'écriture de chaque voix dans la RI;
- 945 — **Output** :
- 946 export de la partition. Plusieurs formats sont possibles (xml, mei,
- 947 lilypond, ...).
- 948
- 949 Plusieurs enjeux :
- 950 — Problème du MIDI avec Qparse :
- 951 ON-OFF en entrée \Rightarrow 1 seul symbole en sortie.
- 952 — Minimiser la distance entre le midi et la représentation en arbre.
- 953 — Un des problèmes de Qparse était qu'il était limité au monopho-
- 954 nique.
- 955 Quelles sont les limites du monophonique?
- 956 — Impossibilité de traiter plusieurs voix et de reconnaître les ac-
- 957 cords.

3.5 Les forme rythmiques

Un forme rythmique 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 forme rythmique (gamme).

Définitions

forme rythmique : 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 forme rythmique considère l'ensemble des combinaisons que le batteur pourrait rencontrer en interprétant un texte rythmique à l'aide du forme rythmique.

Un ensemble de forme rythmiques comprenant leur signature rythmique et leurs règles spécifiques de réécriture sera nécessaire. Les forme rythmiques devront être distribués dans 4 grandes catégories :

Nous exposerons 3 forme rythmiques afin d'illustrer les propos de cette section :

- 980 — 4/4 binaire
- 981 — 4/4 jazz
- 982 — 4/4 afro-cubain

958 il faudrait expliquer là 959
que le but est d'avoir des 960
schemas types (= forme 961
rythmique) pour calculer 962
la séparation en voix. = 963
une heuristique pour évi- 964
ter d'avoir à explorer une 965
grande combinatoire. et 966
que, une fois le forme ryth-
mique déterminé (ou sé-
lectionné), la séparation se
fait par réécriture du mo-
dèle (règles de projection
simplification)

967 je ne comprend pas bien la 968
définition de forme ryth- 969
mique : motif + gamme ou 970
motif + gamme + texte? 971
déf. des gammes n'est pas
du tout claire.

972 est-ce que le motif est fixe 973
et les gammes variables? 974
est-ce le motif qui déter- 975
mine la signature ryth-
mique et les voix?

976 signature rythmique n'est 977
pas défini. règles de rééc-
ture non plus 978

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

TABLE 3.3 – Systèmes

Objectif des forme rythmiques

Les forme rythmiques devront être matchés sur l’input MIDI afin de :

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

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

bien. il faudrait expliquer ça avant.

Détection d’indication de mesure

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

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



Exemple 1



Exemple 2

FIGURE 3.13 – signature rythmique

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 mu-

1005 sique est dite ternaire alors que le rock ternaire sera plutôt écrit comme
1006 dans l'exemple 2).

1007 **Choix d'une grammaire**

1008 Il faut prendre en compte l'existence potentielle de plusieurs grammaires
1009 dédiées chacune à un type de contenu MIDI. Le choix d'une grammaire
1010 pondérée doit être fait avant le parsing puisque Qparse prend en entrée
1011 un fichier MIDI et un fichier wta (grammaire). C'est pour cette raison que
1012 la signature rythmique 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 in-
dications que les données MIDI elles-mêmes. Par conséquent, les motifs
des forme rythmiques devront être recherchés sur l'input (*fichiers MIDI*)
avant le lancement du parsing, afin de déterminer la signature rythmique
en amont. Cette tâche devra probablement être effectuée en Machine
Learning.

1021 **Séparation des voix**

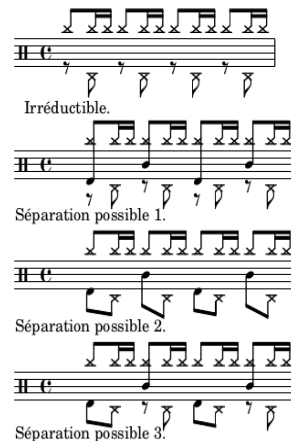


FIGURE 3.14 – Motif 4-4 binaire

1022 Ici, le forme rythmique est construit sur un modèle rock en 4/4 : after-beat
1023 sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. La
1024 forme rythmique est constituée par défaut du motif rd/pf/cc (voir 3.2) et
1025 d'un gamme jouée à la grosse caisse. La première ligne de la figure 3.14
1026 est appelée « Irréductible » car il n'y a pas d'autre choix pertinent pour
1027 la répartition de la ride et du charley au pied. La troisième séparation
1028 proposée est privilégiée car elle répartit selon 2 voix, une voix pour les
1029

1030 mains (rd + cc) et une voix pour les pieds (pf + gc). Ce choix paraît plus
1031 équilibré car deux instruments sont utilisés par voix et plus logique pour
le lecteur puisque les mains sont en haut et les pieds en bas.

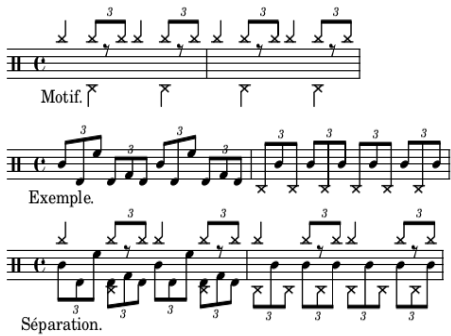


FIGURE 3.15 – Motif 4-4 jazz

1032
1033 Dans la plupart des méthodes, le charley n’est pas écrit car il est considéré
1034 comme évident en jazz traditionnel. Ce qui facilite grandement l’écriture :
1035 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici, le
1036 parti pris est de tout écrire. Dans l’exemple ci-dessus, les mesures 1 et 2
1037 combinées avec le *motif* de la première ligne, sont des cas typiques de la
1038 batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus, la
1039 grosse caisse entre très souvent dans le flot des combinaisons de toms et
1040 de caisse claire et son écriture séparée serait inutilement compliquée et
1041 peu intuitive pour le lecteur. Le choix de séparation sera donc de laisser
1042 les cymbales en haut et toms, caisse claire, grosse caisse et pédale de
1043 charley en bas.

quel exemple?



FIGURE 3.16 – forme rythmique 4-4 afro-latin

1044 La figure 3.16 montre un exemple minimaliste de forme rythmique afro-
1045 latin [26]. Ce forme rythmique doit être écrit sur trois voix car la voix
1046 centrale est souvent plus complexe qu’ici (que des noirs) et la mélanger
1047 avec le haut ou le bas serait surchargé et peu lisible.

1048 **Simplification de l’écriture**

1049 Les explications qui suivent seront appuyé par une réécriture guidée par
1050 une forme rythmique dans la section 4.4.

1051 Les gammes qui accompagnent les motifs d'un forme rythmique étayent
 1052 toutes les combinaisons d'un forme rythmique et elles permettent, combi-
 1053 nées avec le motif d'un forme rythmique, de définir les règles de simplifi-
 1054 cation propres à celui-ci.

1055 Voici les différentes étapes à suivre :

- 1056 — Pour chaque gamme du forme rythmique, faire un arbre de rythme
- 1057 représentant la gamme combinée avec le motif du forme ryth-
- 1058 mique ;
- 1059 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
- 1060 arbre de rythme par voix ;
- 1061 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
- 1062 nœuds qui nécessitent une simplification et écrire la règle.

1063 Certaines précisions concernant l'extraction de ces règles sont néces-
 1064 saires. Il s'agit de précisions à propos de la durée, des silences et de la
 1065 présence ou non d'ouverture de charley dans les instruments joués. Nous
 1066 avons discuté de ces problèmes dans le chapitre 3.

1067 Voici quelques règles inhérentes à la simplification de l'écriture pour la
 1068 batterie : Toutes les continuations (t) qui se trouvent en début de temps
 1069 (figures 4.8, 4.10 et ??) sont transformées en silences (r) sauf si la note
 précédente est un charley ouvert ?

1070 Même si on favorise l'usage des silences pour l'écart entre les notes n'appar-
 1071 tenant pas au même temps, on les supprime systématiquement pour
 1072 2 notes au sein d'un même temps et favorise, une liaison si co, un point si
 1073 pas co et nécessaire, un simple ajustement de la figure de note si suffisant.

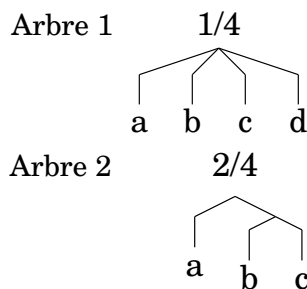


FIGURE 3.17 – Simplification

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

itemize

1075 Soit l'arbre 1 de la figure 3.17 dans lequel : a et d sont des instruments de
 1076 la batterie (x) ;

1077 b et c sont des continuations (t) ;

1078 Pour chacune des conditions suivantes, une suite de la figure 3.18 est
 1079 attribuée :

1080 — Si a n'est pas un co :

1081 ⇒ Suite 1a.

1082 — Si a est un co :

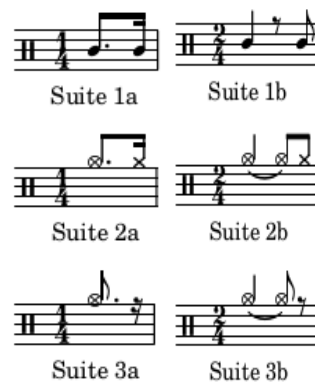


FIGURE 3.18

- 1083 — Si d est un cf :
 1084 ⇒ Suite 2a.
 1085 — Si d est un pf :
 1086 ⇒ Suite 3a : d devient un silence (r).
 1087
 1088 Soit l'arbre 2 de la figure 3.17 dans lequel :
 1089 a et c sont des instruments de la batterie (x);
 1090 b est une continuation (t); Pour chacune des conditions suivantes, une
 1091 suite de la figure 3.18 est attribuée :
 1092 — Si a n'est pas un co :
 1093 ⇒ Suite 1b, b devient un silence.
 1094 — Si a est un co :
 1095 — Si c est un cf :
 1096 ⇒ Suite 2b, b devient une liaison et c devient un cf.
 1097 — Si c est un pf :
 1098 ⇒ Suite 3b : b devient une liaison et c devient un silence.
 1099
 1100 *Rappel :*
 1101 *cf* = charley fermé joué à la main ;
 1102 *co* = charley ouvert joué à la main ;
 1103 *pf* = charley fermé joué au pied.
 1104

1105 Conclusion

1106 <dam>à développer un peu plus</dam> Nous avons formalisé une nota-
 1107 tion de la batterie, modélisé cette notation pour la transcription de don-
 1108 nées MIDI en partition, nous avons décrit Qparse.
 1109 Enfin, nous avons exposé une approche de type dictionnaire (les « forme
 1110 rythmiques ») pour détecter une signature rythmique, choisir une gram-

- ¹¹¹¹ maire pondérée appropriée et énoncer des règles de séparation des voix et
¹¹¹² de simplification de l'écriture.

1113

1114

EXPÉRIMENTATIONS

1115

Sommaire

1116	4.1	Le jeu de données	50
1117	4.2	Analyses et transcriptions manuelles	51
1118	4.3	Transcription polyphonique par parsing	55
1119	4.4	Réécriture guidée par une forme rythmique	58
1120	4.5	BILAN : résultats — évaluation — discussion	63
1121			
1122			
1123			
1124			

Introduction

Dans ce chapitre, nous présenterons le jeu de données et les analyse MIDI-Audio et transcriptions manuelles.

Problématique :

choix d'un outil de transcription ?

passage au polyphonique indispensable pour la suite du travail et pour l'usage des formes rythmiques.

Finir la chaîne de traitement indispensable pour obtenir des résultats chiffrés possible à évaluer.

Nous présenterons mes trois contributions principales :

- le code lilypond normalisé pour la transcription de la batterie avec la notation de type agostini.
- les différentes étapes de résolution du passage au polyphonique.
- l'expérimentation d'un forme rythmique implémentable qui devra être utilisé comme base de connaissances pour augmenter la rapidité et la qualité en sortie de Qparse et comme une méthode de création de nouvelles formes rythmiques.

Enfin, nous finirons par une discussion sur les avancées réalisées dans ce travail, la pertinence des choix qui ont été faits et les moyens d'évaluer les résultats potentiels.

1145 4.1 Le jeu de données

1146 Nous avons utilisé le Groove MIDI Dataset¹ [31] (GMD) qui est un jeu de
 1147 données mis à disposition par Google sous la licence Creative Commons
 1148 Attribution 4.0 International (CC BY 4.0).

1149 Le GMD est composé de 13,6 heures de batterie sous forme de fichiers
 1150 MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000
 1151 mesures de batterie dans les styles les plus courants et avec différentes
 1152 qualités de jeu. Tout le contenu a été joué par des humains sur la batterie
 électronique Roland TD-11 (figure 4.1).

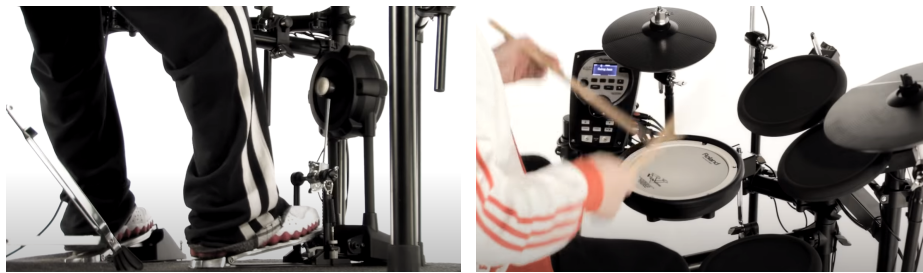


FIGURE 4.1 – Batterie électronique

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

1153

1154 Autres critères spécifiques au GMD :

- 1155 — Toutes les performances ont été jouées au métronome et à un tempo
 1156 choisi par le batteur.
- 1157 — 80% de la durée du GMD a été joué par des batteurs professionnels
 1158 qui ont pu improviser dans un large éventail de styles. Les don-
 1159 nées sont donc diversifiées en termes de styles et de qualités de jeu
 1160 (professionnel ou amateur).
- 1161 — Les batteurs avaient pour instruction de jouer des séquences de
 1162 plusieurs minutes ainsi que des fills²
- 1163 — Chaque performance est annotée d'un style (fourni par le batteur),
 1164 d'une signature rythmique et d'un tempo ainsi que d'une identifi-
 1165 cation anonyme du batteur.
- 1166 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
 1167 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
 1168 session du GMD.
- 1169 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
 1170 fichier MIDI.

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

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.

1171 **Format des données**

1172 Le Roland TD-11 enregistre les données dans des fichiers MIDI et les
1173 divise en plusieurs pistes distinctes :

- 1174 — une pour le tempo et l'indication de mesure ;
- 1175 — une pour les changements de contrôle (position de la pédale de
1176 charley) ;
- 1177 — une pour les notes.

1178

1179 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1180 canal 9 (qui est le canal canonique pour la batterie).

1181 Pour simplifier le traitement de ces données, ces trois pistes ont été fu-
1182 sionnées en une seule piste qui a été mise sur le canal 9.

1183 **4.2 Analyses et transcriptions manuelles**

1184 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
1185 partir de fichiers MIDI et Audio du GMD.

1186 **Comparaisons de transcriptions**

1187 Pour les comparaisons de transcriptions, les transcriptions manuelles
1188 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les transcrip-
1189 tions automatiques (TA) ont toutes été générées par import d'un fichier
1190 MIDI dans MuseScore.

1191 **Exemple d'analyse 1**

Transcription manuelle ⇒ Transcription automatique



- 1192 — Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
- 1193 — Les silences de la mesure 1 de la TA sont inutilement surchargés ;
- 1194 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
1195 premières notes (une double-croche et une croche) d'un triolet sur
1196 le temps 1 de la mesure 2 de la TA.

Transcription manuelle \Rightarrow Transcription automatique



1197 Exemple d'analyse 2

- 1198 — Les doubles croches ont été interprétées en quintolet
- 1199 — La deuxième double-croche est devenue une croche.

1200 Exemple d'analyse 3

Transcription manuelle \Rightarrow Transcription automatique



- 1201 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1202
- 1203 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1204 décalés d'une double croche vers la droite dans la TA.
- 1205 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1206 alors qu'elle appartenait à un triolet dans la TM.
- 1207 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1208 pas la TM.
- 1209

1210 Exemple d'analyse 4

Transcription manuelle \Rightarrow Transcription automatique



- 1211
- 1212 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- 1213 manière excessivement complexe !

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

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

1214 **Exemple d'analyse 5 (flas)**

1215 Transcription manuelle



1216

1217 Transcription automatique

1218



1219

1220

- 1221 — Le premier fla est reconnu comme étant un triolet contenant une
- 1222 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1223 ornementée.
- 1224 — Le deuxième fla est reconnu comme étant un accord.
- 1225 — Les deux double en contre-temps sur le temps 4 de la TM sont mal
- 1226 quantifiées dans la TA.
- 1227 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1228 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1229 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1230 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1231 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1232 la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de
- 1233 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1234 temps 3 de la TA.
- 1235 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1236 exemple, le charley ouvert en contre-temps sur le temps 2 de la
- 1237 mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem
- 1238 pour les cross-sticks.

1239 **Conclusion d'analyse**

1240 Ces analyses ont montré la difficulté pour un logiciel comme MuseScore
 1241 d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI
 1242 ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance
 1243 de la forme du rythme impliquant sa position dans la mesure. Cette recon-
 1244 naissance pourrait permettre de rectifier les problèmes de signature ryth-
 1245 mique ainsi que les problèmes de décalage de temps. La reconnaissance
 1246 de la forme du rythme permettrait aussi de supprimer les aberrations du
 1247 type de celle de l'exemple d'analyse 4, puisque l'erreur sur cet exemple

serait reconnue comme un élément qui ne rentre pas dans le cadre de la forme de rythme en question. La dernière raison qui rend le travail difficile est l'identification des flas, comment savoir si deux notes jouées très proches sont :

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

Transcription de partition

La figure 3.11 est la transcription manuelle des fichiers *004_jazz-funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

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

La transcription manuelle de la partition de la figure 3.11 et l'analyse d'autre fichiers MIDI (voir section 4.2) m'ont mené aux observations suivantes :

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

Conclusion sur les transcriptions manuelles

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

- 1287
- 1288
- 1289
- 1290
- 1291
- 1292
- 1293
- 1294
- 1295
- 1296
- 1297
- 1298
- 1299
- 1300
- 1301
- 1302
- 1303
- 1304
- 1305
- 1306
- difficultés principales : trouver une application permettant de choisir librement la notation de la batterie. Lylipond le permet mais beaucoup de recherches ont été nécessaires pour comprendre l'ensemble des fonctionnalités permettant de faire fonctionner la notation « agostinienne » ainsi que les diverses subtilités de notations (accents, ghost-notes, flas, ...).
 - lylipond reste néanmoins un choix très agréable, une fois ces difficultés surmontées.
 - Écrire la partition de la figure 3.11 m'a pris beaucoup de temps car j'ai dû chercher comment écrire chaque nouvel évènement mais les autres transcriptions ont été beaucoup plus rapide et très aisées.
 - Même si cela représente un investissement au départ, je recommande lylipond pour écrire la batterie et je pense que c'est meilleur outil pour cette tâche pour le moment. On peut configurer absolument tout.
 - dans les autres logiciel d'édition de type musescore, la batterie est toujours confiné au système de notation américain.
 - pour une comparaison entre système américain et système agostinien, voir section 4.2 est comparer les notations TM (agostinien) et TA (américain).

1307

4.3 Transcription polyphonique par parsing

1308

1309

<flo>Sujet de cette partie -> première problématique / contribution principale : transcription polyphonique par parsing (verrou) : jams etc</flo>

1310

Les Jams

1311

Les Jams permettent de passer du monophonique au polyphonique.

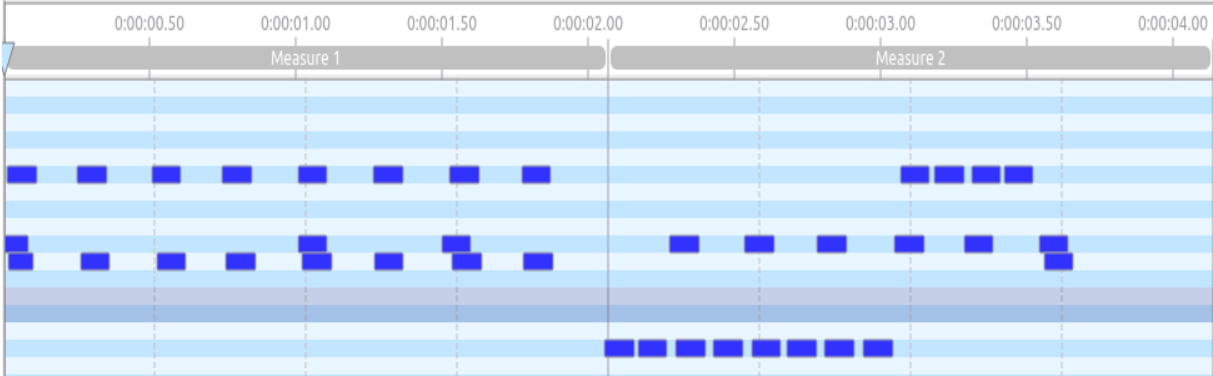
1312

1313

Le parsing

1314

Tests effectués avec le fichier midi suivant :



1315

1316 Un premier test convaincant est effectué avec la grammaire suivante :

1317

1318 <dam>incompréhensible ainsi, c'est dommage</dam> // bar level

1319 0 -> C0 1

1320 0 -> E1 1

1321 0 -> U4(1, 1, 1, 1) 1

1322

1323 // half bar level

1324 9 -> C0 1

1325 9 -> E1 1

1326

1327 // beat level

1328 1 -> C0 1

1329 1 -> E1 1

1330 1 -> T2(2, 2) 1

1331 1 -> T4(4, 4, 4, 4) 1

1332

1333 // croche level

1334 2 -> C0 1

1335 2 -> E1 1

1336

1337 // double level

1338 4 -> C0 1

1339 4 -> E1 1

1340 4 -> E2 1

1341 4 -> T2(6, 6) 1

1342

1343 // triple level

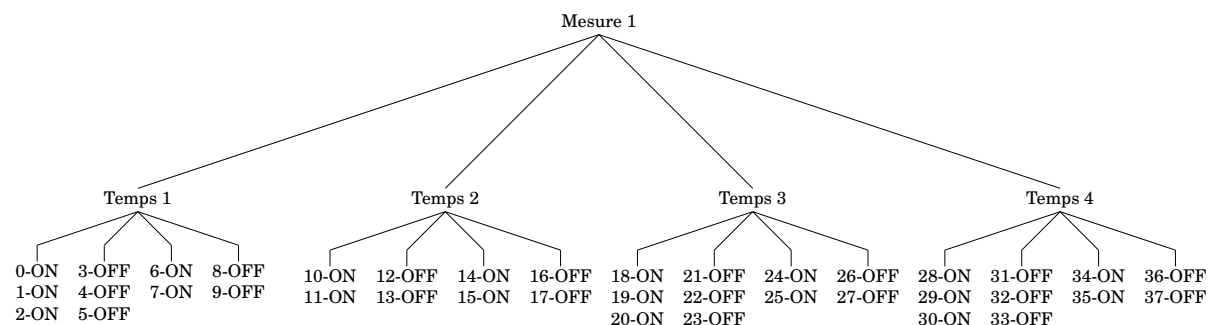
1344 6 -> E1 1

1345

1346 Cette grammaire sépare les ligatures par temps au niveau de la
1347 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
1348 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
1349 L'arbre de parsing en résultant est considéré comme « convaincant » car
1350 il découpe correctement les mesures et les temps.

1351

1352



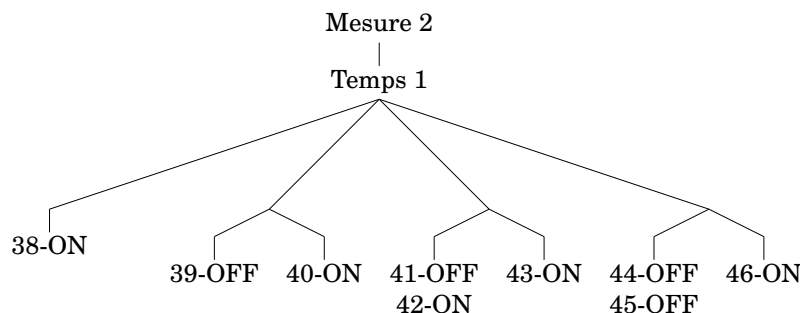
1353

1354

1355 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1356 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1357 quantification visibles dès le premier temps.

1358

1359



1360

1361

1362 Les Onsets sont correctement triés au niveau des doubles croches
 1363 mais certaines doubles croches sont inutilement subdivisées en triples
 1364 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1365 ci-dessus).

1366

1367 2ème exemple :

1368 Après une augmentation du poids des triples croches dans la grammaire
 1369 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1370 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1371 sont bien triés et aucuns ne sont subdivisés.

1372 4.4 Réécriture guidée par une forme rythmique

1373 La démonstration qui suit est basée sur la partition de référence de la
1374 figure 3.11 puisque la forme rythmique qui sera utilisée en est directe-
1375 ment extraite.

1376

1377 Nous allons montrer :

- 1378 — la composition de cette forme rythmique ;
- 1379 — son état finale, c'est à dire toutes les combinaisons entièrement
1380 écrites en notation correcte sur partition ;
1381 ⇒ cela constituera une référence pour la réécriture ;
- 1382 — un exemple de transformation de la forme rythmique en arbre de
1383 rythme ;
- 1384 — l'application de la séparation des voix sur cet exemple basé sur la
1385 référence citée précédemment (la forme rythmique en question) ;
1386 ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a
1387 de voix (deux arbres pour cette forme rythmique) ;
- 1388 — les règles de simplification propres à la forme rythmique dont nous
1389 parlons.

1390 L'objectif de cette démonstration est de montrer comment un jeu de plu-
1391 sieurs formes rythmiques pourrait être implémenter dans le cadre d'une
1392 approche dictionnaire.

1393 Motifs et gammes

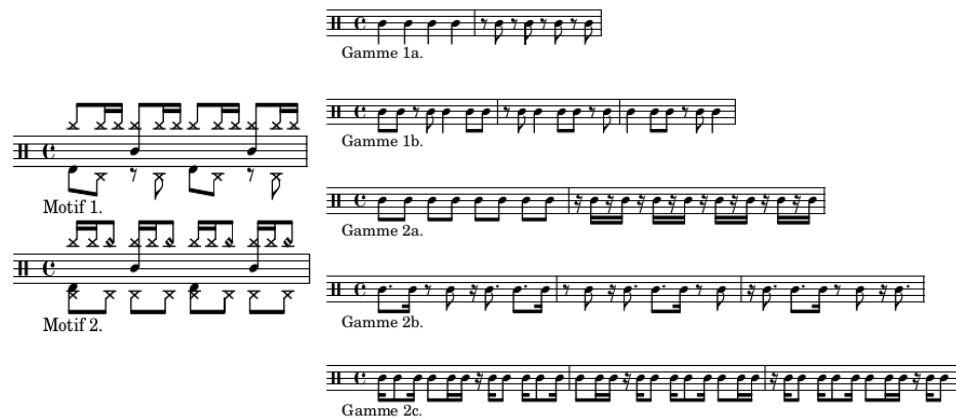


FIGURE 4.2 – Motifs et gammes

1394 Motifs

1395 À partir de la partition de référence, les deux motifs de la figure 4.2
1396 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure

1397 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à
 1398 la mesure 28 avec des variations. Ces deux motifs sont très classiques et
 1399 pourront être détectés dans de nombreuses performances.
 1400

1401 **Gammes**

1402 Les gammes de la figure 4.2 étayent toutes les combinaisons d'un motif
 1403 en 4/4 binaires jusqu'aux doubles croches.
 1404 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-
 1405 mière ne contient que des noires et la deuxième que des croches en contre-
 1406 temps. Ces deux possibilités sont combinées de manière circulaire dans
 1407 les 3 mesures de la deuxième ligne.
 1408 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures
 1409 dont la première ne contient que des croches et la deuxième que des
 1410 doubles-croches en contre-temps. Ces deux possibilités sont combinées de
 1411 manière circulaire dans les lignes 4 et 5 qui contiennent chacune 3 me-
 1412 sures.

1413 **Formes rythmiques — motifs et gammes combinés**

1414 Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure
 1415 4.2.<dam>à commenter un peu plus, notamment pour dire si la combinai-
 son est faite automatiquement ou non</dam>

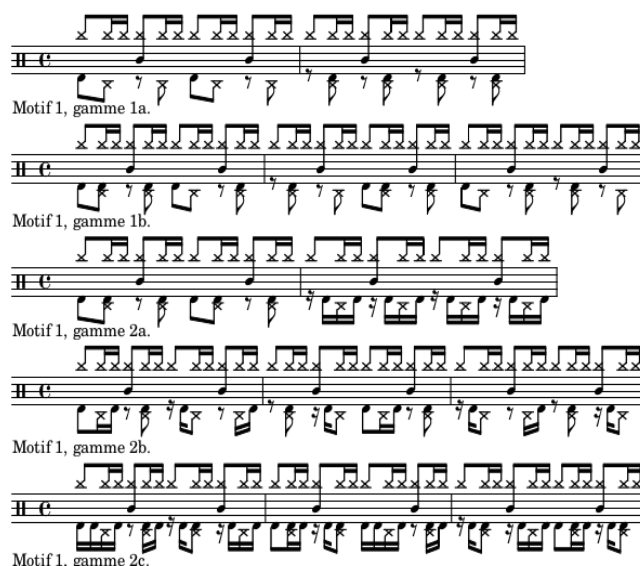


FIGURE 4.3 – Partition d'un forme rythmique en 4/4 binaire

1417 **Représentation de la forme rythmique en arbres de**
 1418 **rythmes**

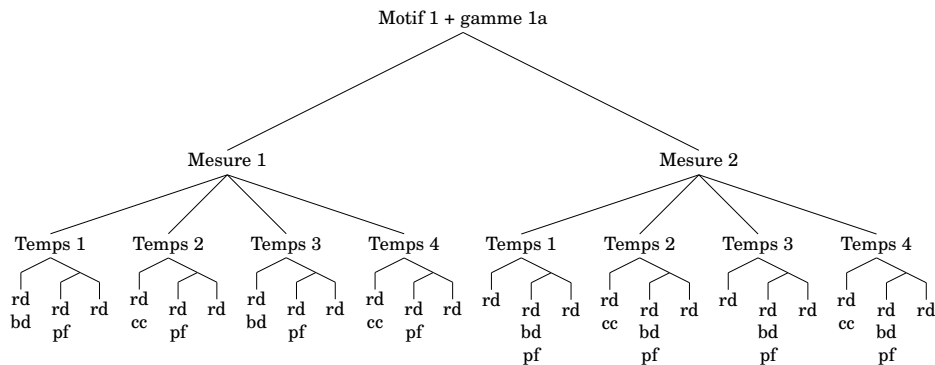


FIGURE 4.4 – Arbre de rythme — forme rythmique

1419 L'arbre de la figure 4.4 servira de base pour le suite de l'expérimentation.
 1420 Comme indiqué à la racine de l'arbre, il représente la première ligne de la
 1421 figure 4.3. Même si cet arbre représente parfaitement le rythme concerné,
 1422 il manque des indications de notation telles que les voix spécifiques à
 1423 chaque partie du rythme ainsi que les choix d'écriture pour les distances
 1424 qui séparent les notes de chaque voix entre elles en termes de durée.

1425 **Réécriture — séparation des voix et simplification**

1426 **La séparation des voix**

1427 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui
 1428 le constituent et les voix seront regroupées en suivant les règles du forme
 rythmique.

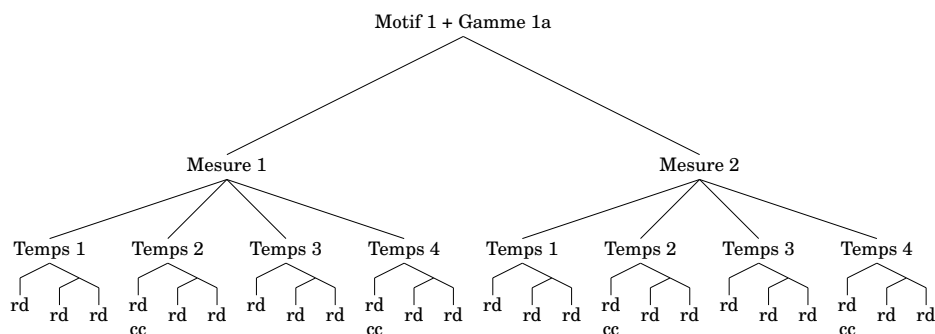


FIGURE 4.5 – Arbre de rythme — voix haute

1429
 1430 La voix haute (figure 4.5) regroupe la ride et la caisse-claire sur les liga-
 1431 tures du haut.

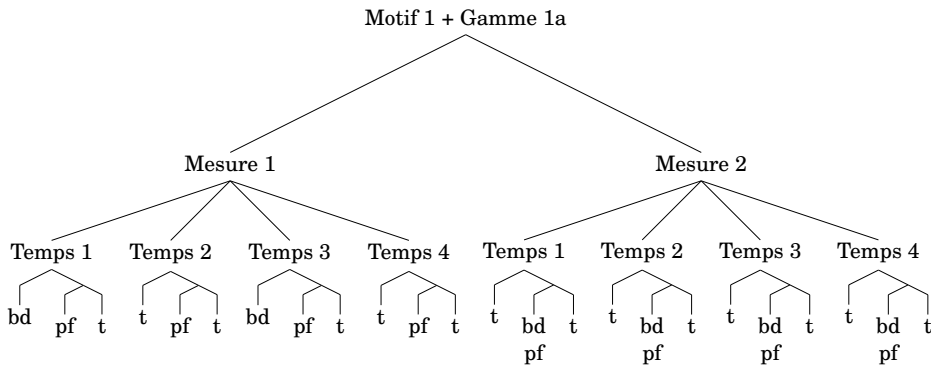


FIGURE 4.6 – Arbre de rythme — voix basse

1432 La voix basse (figure 4.6 regroupe la grosse-caisse et le charley au pied
1433 sur les ligatures du bas.

1434 **Les règles de simplifications**

1435 L’objectif des règles de simplifications est de réécrire les écarts de durées
1436 qui séparent les notes d’une manière appropriée pour la batterie et qui
1437 soit la plus simple possible. Les ligatures relient les notes d’un temps
1438 entre elles afin de rendre la pulsation visuelle).

1439

1440 Pour les figures ci-dessous :

- 1441 — x = une note ;
- 1442 — r = un silence ;
- 1443 — t = une continuation (point ou liaison)

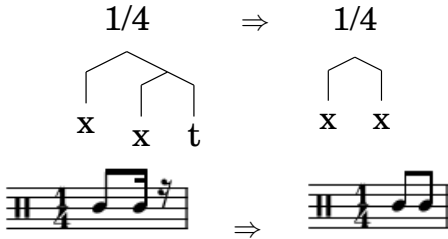


FIGURE 4.7

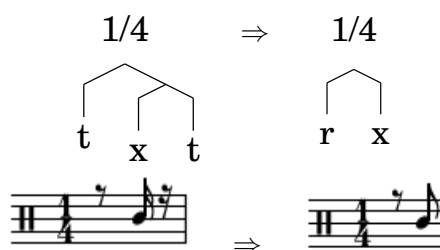


FIGURE 4.8

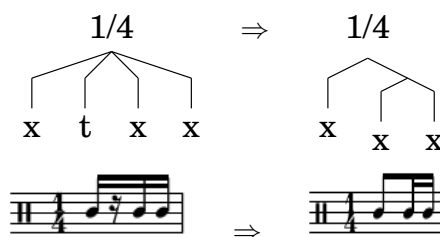


FIGURE 4.9

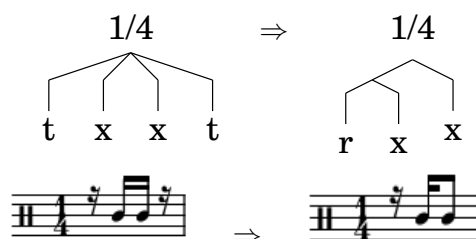


FIGURE 4.10

- 1444 Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique.
 1445 Les arbres manquants seront mis en annexe.
 1446 Les règles remplacent par un silence les continuations (t) qui sont au début d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il y a des ouvertures de charley, cela n'est pas toujours applicable.

1449 Conclusion sur cette réécriture guidée

- 1450 La méthode des formes rythmiques étant basée sur une approche dictionnaire, Le premier objectif de cette réécriture guidée est d'orienter la recherche d'autres formes rythmiques par observation du jeu de données et de montrer comment les construire pour agrandir la base de connaissance de Qparse pour la transcription de la batterie.

1455 4.5 BILAN : résultats — évaluation — discussion

1456 Cette section regroupe les avancées qui ont été réalisées par rapport aux
 1457 objectifs de départ ainsi qu’une réflexion sur le moyen d’évaluer les résul-
 1458 tats de l’ADT avec Qparse. Nous avons amélioré le système de quantifi-
 1459 cation de Qparse pour la batterie, notamment le passage à la polyphonie
 1460 avec les Jams.

1461 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
 1462 grammaires avec des fichiers MIDI courts.

1463 Puis, une sortie MEI a aussi été obtenue (encore à vérifier).

1464 Dans cette section, nous discuterons sur la pertinence de l’ensemble des
 1465 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
 1466 ont été faites ou non et nous tenterons d’en expliquer la ou les raisons.

1467 — Le choix de travailler avec Lilypond et non Verovio. Ce choix était
 1468 motivé par la liberté totale concernant la notation de la batterie
 1469 dont un et la disponibilité d’un set de notation de type Agostini.
 1470 C’est la seule application qui me permettait d’écrire la notation de
 1471 la batterie exactement comme je le souhaitais.

1472 — Avancé de la chaîne de traitement (nous sommes arrivés aux arbres
 1473 de parsing, nous avons traité le polyphonique (identification des
 1474 regroupements de notes⁵) ⇒ Quelques arbres ont été obtenus sur
 1475 des exemples simples⁶)

1476 — 2 dimensions de le travail fourni :

1477 - La volonté de pousser un exemple simple jusqu’au bout de la
 1478 chaîne pour obtenir des résultats et une évaluation sur au moins
 1479 un exemple ; - La réalité du travail à fournir pour faire avancer sur
 1480 la chaîne de traitement. ⇒ Une solution aurait été de considérer
 1481 les arbres de parsing obtenus après le traitement du polyphonique
 1482 comme un résultat local possible à évaluer au lieu d’attendre que la
 1483 chaîne arrive jusqu’à la génération d’une partition mais cela n’était
 1484 pas prioritaire pendant le stage.

1485 — Création d’un jeu de forme rythmique basique représentatif des
 1486 différents styles à recouvrir. Ce jeu n’a pas pu être créé, car comme
 1487 vu plus haut, je me suis focalisé sur un exemple pour pouvoir le
 1488 vérifier entièrement et dans l’espoir de pouvoir le tester en fin de
 1489 chaîne. **Évaluation** Matcher les motifs aurait été indispensable
 1490 pour obtenir une quantité de résultats qui justifieraient une
 1491 évaluation automatique permettant de faire des graphiques.

1492 L’évaluation fut entièrement manuelle car :

1493 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
 1494 VS output) Pour l’évaluation, il aurait fallu produire un module.

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

1495 <dam>je ne sais pas si tu auras encore le temps de faire ça, sinon
1496 il faudra décrire comment tu aurais aimé évaluer, proprement et
1497 sans résultats chiffrés</dam> L'évaluation est-elle automatique ou
1498 manuelle ?
1499 Possibilité d'un export lilypond en arbre pour comparer l'ouput
1500 avec la transcription manuelle.
1501 Possibilité de transformer lilypond(output) et lilypond(ref) en
1502 ScoreModel ou MEI pour les comparer et faire des statistiques.
1503 Si transformés en MEI : diffscore de Francesco. Possibilité de
1504 transformer lilypond(output) et lilypond(ref) en MusicXML pour
1505 les comparer ou dans Music21. L'expérimentation peut-être consi-
1506 dérer comme une évaluation manuelle ? (magicien d'Oz)
1507 Lilypond vers MIDI + ouput vers MIDI \Rightarrow Comparaison des MIDI
1508 dumpés.
1509
1510 La transcription automatique de la batterie est un sujet passionnant mais
1511 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-
1512 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-
1513 moins été générée. Elle sera un bon point de départ pour les travaux fu-
1514 turs dont plusieurs propositions sont énoncés dans le présent document.

CONCLUSION GÉNÉRALE

1516 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1517 tion automatique de la batterie. Son objectif était de transcrire, à partir
1518 de leur représentation symbolique MIDI, des performances de batteur de
1519 différents niveaux et dans différents styles en partitions écrites.
1520 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1521 cessus de regroupement des événements MIDI qui nous a permis de faire
1522 la transition du monophonique vers le polyphonique. Une des données
1523 importante de ce processus était de différencier les nature des notes d'un
1524 accord, notamment de distinguer lorsque 2 notes constituent un accord
1525 ou un *fla*.
1526 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1527 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1528 tionnables en amont du parsing, soit par indication des noms des fichiers
1529 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1530 naire de patterns prédéfinis ⁷ qu'il serait pertinent de mettre en œuvre en
1531 machine learning.
1532 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1533 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1534 d'un système et pour la simplification grâce aux gammes du motif d'un
1535 système. Nous avons aussi montré comment, dans des travaux futurs, un
1536 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1537 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1538 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1539 la chaîne de traitement en terme de réécriture.
1540 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1541 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1542 entièrement automatisée de partitions de batterie et au problème général
1543 de l'AMT de bout en bout.[11]

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

BIBLIOGRAPHIE

- 1545 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1546 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1547 1996. – Cité pages 7, 18 et 35.
- 1548 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1549 et 15.
- 1550 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1551 page 12.
- 1552 [4] Gaël Richard et al. De fourier à la reconnaissance
1553 musicale. Available at [https://interstices.info/
1554 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1555 – Cité page 12.
- 1556 [5] Caroline Traube. Quelle place pour la science au sein de la musico-
1557 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1558 [6] Leonard Bernstein Office. The unanswered question : Six talks at
1559 harvard. Available at [https://leonardbernstein.com/about/
1560 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1561 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1562 entre les traitements de la musique et du langage. *La Lettre des
1563 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1564 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1565 Techniques de traitement automatique du langage naturel appli-
1566 quées aux représentations symboliques musicales. In *JIM 2021 -
1567 Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1568 Cité page 13.
- 1569 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-
1570 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1571 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1572 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1573 Cité page 13.
- 1574 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1575 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1576 lenges and future directions. *Journal of Intelligent Information Sys-*
1577 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 65.
- 1578 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1579 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1580 Cité pages 14 et 27.
- 1581 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-
1582 SURE, 1997. – Cité page 15.
- 1583 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1584 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1585 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1586 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1587 1483, 2018. – Cité pages 15, 23 et 27.
- 1588 [15] Moshekwa Malatji. Automatic music transcription for two instru-
1589 ments based variable q-transform and deep learning methods, 10
1590 2020. – Cité page 22.
- 1591 [16] Antti J. Eronen. Musical instrument recognition using ica-based
1592 transform of features and discriminatively trained hmms. *Seventh*
1593 *International Symposium on Signal Processing and Its Applications*,
1594 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1595 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1596 drum sound description for real-world music using template adap-
1597 tation and matching methods. *International Conference on Music*
1598 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1599 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1600 musical statistics as guides for audio-to-score piano transcription.
1601 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24, 26 et 34.
- 1602 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1603 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1604 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1605 *Computation in Music*, volume Lecture Notes in Computer Science
1606 of *Proceedings of the Seventh International Conference on Mathema-*
1607 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1608 2019. Springer. – Cité pages 24 et 26.
- 1609 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1610 ring of rhythm structures. In *Proceedings of the First International*
1611 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1612 IEEE Computer Society. – Cité page 26.
- 1613 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1614 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1615 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1616 Cité page 26.
- 1617 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1618 equivalent rhythmic notations based on rhythm tree languages. In
1619 *Third International Conference on Technologies for Music Notation
1620 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1621 pez Palma and Mike Solomon. – Cité page 26.
- 1622 [23] R. Marxer and J. Janer. Study of regularizations and constraints in
1623 nmf-based drums monaural separation. In *International Conference
1624 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1625 02/09/2013 2013. – Cité page 27.
- 1626 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1627 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 29.
- 1628 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1629 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 29.
- 1630 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1631 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1632 94120 Fontenay-sous-Bois, 2000. – Cité pages 31 et 45.
- 1633 [27] Nicolas Guiomard-Kagan. *Traitement de la polyphonie pour l'analyse
1634 informatique de partitions musicales*. PhD thesis. – Cité page 34.
- 1635 [28] Frédéric Canet. La batterie... mot à mot! Available at [https://](https://rimshotetghostnote.fr/)
1636 rimshotetghostnote.fr/ (2021). – Cité page 36.
- 1637 [29] M. Laurson. Patchwork : a visual programming language and some
1638 musical applications. 1996. – Cité page 40.
- 1639 [30] Jean Bresson, Carlos Agon, and Gérard Assayag. Openmusic visual
1640 programming environment for music composition, analysis and re-
1641 search. – Cité page 40.
- 1642 [31] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1643 Bamman. Learning to groove with inverse sequence transforma-
1644 tions. In *International Conference on Machine Learning (ICML)*,
1645 2019. – Cité page 50.

