

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	<b>Liste des figures</b>	<b>4</b>
19	<b>Liste des tableaux</b>	<b>5</b>
20	<b>Introduction générale</b>	<b>7</b>
21	<b>1 Contexte</b>	<b>11</b>
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	<b>2 État de l'art</b>	<b>21</b>
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	<b>3 Méthodes</b>	<b>29</b>
32	3.1 La notation de la batterie . . . . .	29
33	3.2 La transcription manuelle . . . . .	37
34	3.3 Modélisation pour la transcription . . . . .	39
35	3.4 Analyse syntaxique pour la transcription musicale . . . . .	41
36	3.5 Les forme rythmiques . . . . .	43
37	<b>4 Expérimentations</b>	<b>49</b>
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	<b>Conclusion générale</b>	<b>65</b>
44	<b>Bibliographie</b>	<b>67</b>

## 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 . . . . .	42
70	3.13	signature rythmique . . . . .	44
71	3.14	Motif 4-4 binaire . . . . .	45
72	3.15	Motif 4-4 jazz . . . . .	45
73	3.16	forme rythmique 4-4 afro-latin . . . . .	46
74	3.17	Simplification . . . . .	47
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 :

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

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

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 <sup>1</sup>.

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 <sup>2</sup> 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)<sup>1</sup>. 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 chansons<sup>2</sup>. Mais d'autres traitent directement la musique + ref.

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 musique<sup>3</sup>, 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 accords<sup>4</sup>.

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<sup>5 6 7</sup> 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 $\Rightarrow$ phonèmes $\Rightarrow$ texte signal $\Rightarrow$ notes, accords $\Rightarrow$ partition	données linéaires données structurées

TABLE 1.1 – speechToText vs AMT

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

## 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 commune<sup>8</sup> (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 orale<sup>9</sup>, ...). La TAM est aussi utile  
311 pour la recherche et l'annotation automatique d'informations musicales,  
312 pour l'analyse musicologique<sup>10</sup> 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 pdf<sup>11</sup>.

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 rythmes<sup>12</sup> ou de séquenceurs<sup>13</sup> 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 MIDI<sup>14</sup> (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)

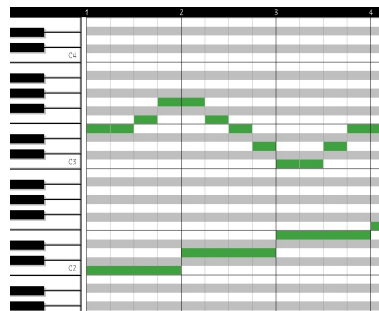


FIGURE 1.1 – Exemple événements avec durée

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

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

Ici, définir *Tick* et *channel*.

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

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



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

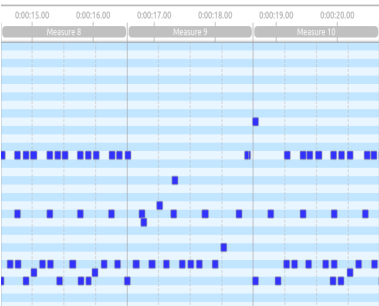


FIGURE 1.3 – Exemple évènements sans durée

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

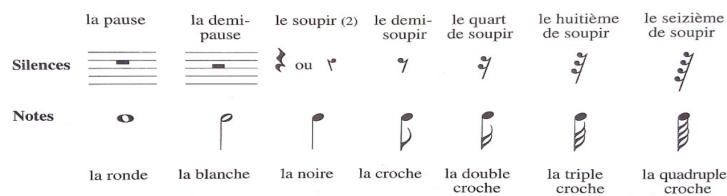


FIGURE 1.4 – Les silences



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 <sup>15</sup> :

- 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

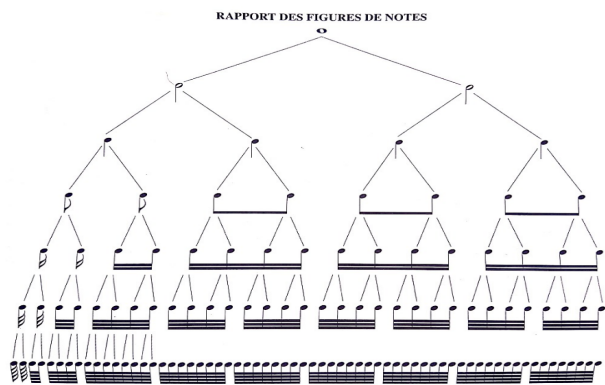


FIGURE 1.6 – Rapport des figures de notes

[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](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 musique<sup>16</sup> 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 1.8<sup>17</sup> 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>

```

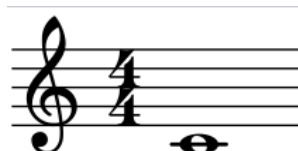


FIGURE 1.8 – MusicXML

## 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 monophoniques<sup>1</sup> [11]. Actuellement, le problème de l'estimation automatique de la hauteur des signaux monophoniques peut être considéré comme résolu, mais dans la plupart des contextes musicaux, les instruments sont polyphoniques<sup>2</sup>. 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).

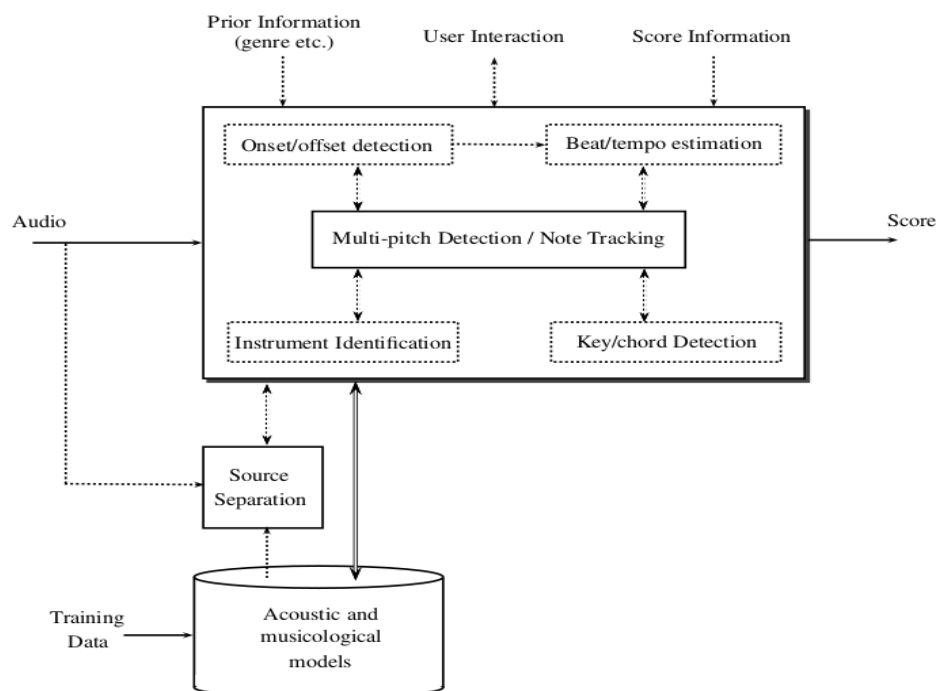


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

*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*<sup>3</sup> avec un modèle de markow caché (HMM). Les  
 561 *features* sont transformés en représentations statistiques indépendantes.  
 L'approche AdaMa [17] est une autre approche de la même catégorie ; elle  
 commence par une estimation initiale des sons de la batterie qui sont ité-  
 rativement raffinés pour correspondre à (pour matcher) l'enregistrement  
 visé.

classification des évène- 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  
 de la séparation des voix avec la quantification du rythme.  
 L'approche de [19] est fondée sur la conviction que la complexité de la  
 structure musicale dépasse les modèles linéaires.

ce n'est pas exactement 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)<sup>4</sup>  
 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://fr.wikipedia.org/wiki/Modèle_de_Markov_caché)  
[https://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](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

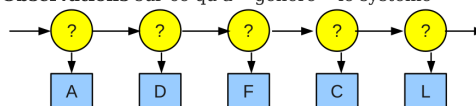


FIGURE 2.2 – HMM

597 *Source : Cours de Damien Nouvel*<sup>5</sup>

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

619

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

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





645

646

## MÉTHODES

647

## Sommaire

648	3.1	La notation de la batterie . . . . .	29
649	3.2	La transcription manuelle . . . . .	37
650	3.3	Modélisation pour la transcription . . . . .	39
651	3.4	Analyse syntaxique pour la transcription musicale . . .	41
652	3.5	Les forme rythmiques . . . . .	43

654

655

## 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 `qparse`<sup>1</sup>, 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

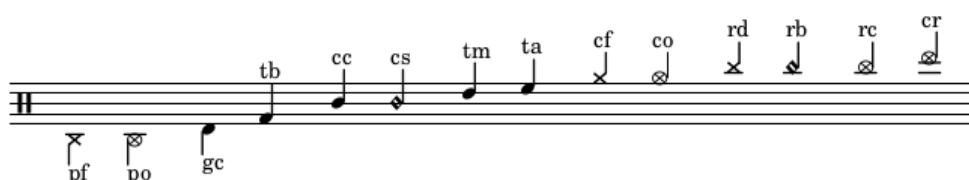


FIGURE 3.2 – Hauteur et têtes de notes

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 3.1<sup>2</sup> 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 charley<sup>3</sup>.

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 pulsation<sup>4</sup> 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.

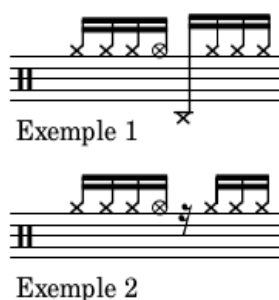


FIGURE 3.4 – Silence joué

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).

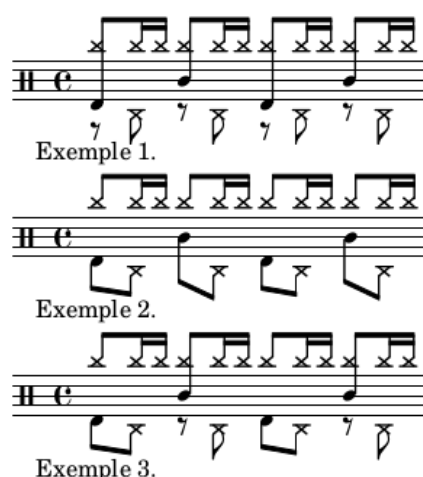


FIGURE 3.6 – Séparation des voix

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]

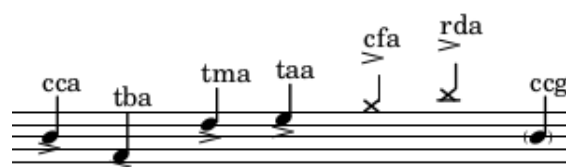


FIGURE 3.7 – Les accents et les ghost-notes

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 lilypond<sup>5</sup>.

### 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.

```

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

up = \drummode {

\override Script.Y-offset = #-1.0

% Mesure 1
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>

}

#(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)))

```

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 — difficultés principales : trouver une application permettant de choi-  
 869 sir librement la notation de la batterie. Lylipond le permet mais  
 870 beaucoup de recherches ont été nécessaires pour comprendre l'en-  
 871 semble des fonctionnalités permettant de faire fonctionner la nota-  
 872 tion « agostinienne » ainsi que les diverses subtilités de notations  
 873 (accents, ghost-notes, flas, ...).
- 874 lylipond reste néanmoins un choix très agréable, une fois ces diffi-  
 875 cultés surmontées.
- 876 — Écrire la partition de la figure 3.11 m'a pris beaucoup de temps car  
 877 j'ai dû chercher comment écrire chaque nouvel évènement mais les  
 878 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).

3.3 Modélisation pour la transcription

Les pitches

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

TABLE 3.2 – Codes, identités et pitches des instruments

Le tableau 3.2 présente dans l’ordre, les codes des instruments, leur identité (instrument ou parti d’un instrument — joué avec les mains ou avec les pieds), le ou les pitches qui lui sont associés. Plusieurs pitches peuvent parfois désigner le même instrument afin de pouvoir supporter des kits de batterie plus larges (avec par exemple plusieurs toms basses qui n’auraient pas tous exactement la même sonorité) ou simplement de styles différents (pour chaque kits standard, ce sont les mêmes instruments mais de styles différents)<sup>6</sup>. 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

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

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...)

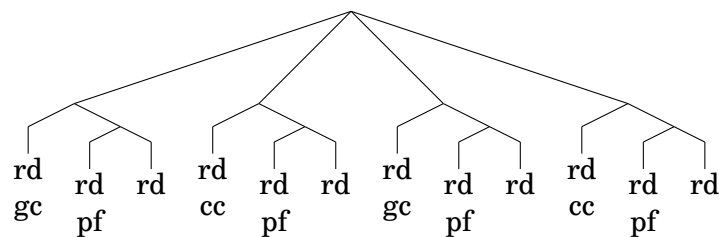
### La vitesse

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.

### Les arbres de rythmes

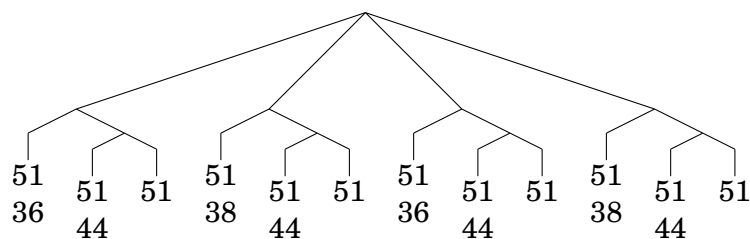
Les arbres de rythmes représentent un rythme dont les possibilités de notation sur une partition sont théoriquement multiples. Les branchements sont des divisions d'intervalle temporel, les feuilles sont des événements musicaux commençant au début de l'intervalle [29] [30] .

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



922

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



924



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

## 3.4 Analyse syntaxique pour la transcription musicale

Qparse<sup>7</sup> est un outil pour la transcription musicale, qui, à partir d'une performance symbolique, séquentielle et non quantifiée, produit une partition structurée. Il effectue conjointement des tâches de quantification rythmique et d'inférence de la structure de la partition à l'aide de technique de analyse syntaxique (*parsing*). Le but du *parsing* est en effet la structuration d'une représentation séquentielle en entrée (un mot fini), suivant un modèle de langage [31]. Dans le cas de qparse, le "mot" d'entrée est typiquement au format MIDI, et le modèle de langage est un automate d'arbres pondéré représentant des préférences en terme de notation musicale à produire [32]. 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.
- Francesco Foscari (PhD, CNAM) : apprentissage ; Evaluation.
- Clement Poncelet (Salzburg U.) : integration de la librairie Midifile pour les input MIDI.
- Philippe Rigaux (CNAM) : production de partition au format MEI et de modèle intermédiaire de partition en sortie.
- Masahiko Sakai (Nagoya U.) : mesure de la distance input/output pour la quantification et CMake framework ; évaluation.

Explication des différentes étapes de la figure 3.12<sup>8</sup> :

- **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

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

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

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

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

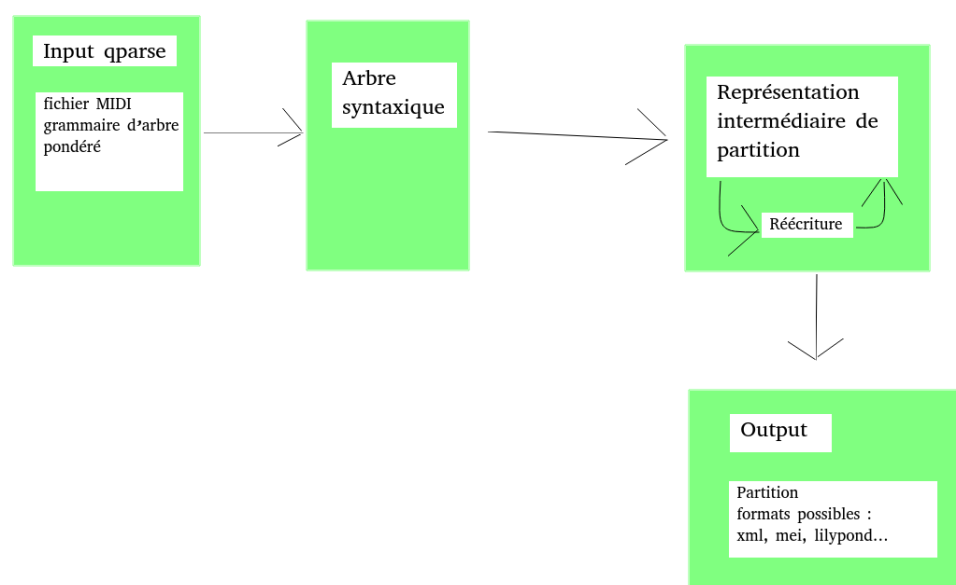


FIGURE 3.12 – Présentation de Qparse

alignées et les relations entre les notes sont identifiées (accords, fla, etc. . . ); un arbre de parsing global est créé ;

— **Score Model :**

— Les instruments sont identifiés dans `scoremodel/import/tableImporterDrum.cpp` ;

— Réécriture 1 :

séparation des voix  $\Rightarrow$  un arbre par voix  $\Rightarrow$  représentation intermédiaire (RI) ;

— Réécriture 2 :

simplification de l'écriture de chaque voix dans la RI ;

— **Output :**

export de la partition. Plusieurs formats sont possibles (xml, mei, lilypond, . . . ).

Plusieurs enjeux :

— Problème du MIDI avec Qparse :

ON-OFF en entrée  $\Rightarrow$  1 seul symbole en sortie.

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

— Un des problèmes de Qparse était qu'il était limité au monophonique.

Quelles sont les limites du monophonique ?

— Impossibilité de traiter plusieurs voix et de reconnaître les accords.

3.5 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 :

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

- Nous exposerons 3 forme rythmiques afin d'illustrer les propos de cette section :
- 4/4 binaire
  - 4/4 jazz
  - 4/4 afro-cubain

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 se-

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

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

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

signature rythmique n'est pas défini. règles de réécriture non plus

bien. il faudrait expliquer ça avant.

ront utilisés pour la **séparation des voix**. Les règles de **simplification** (les combinaisons de réécritures) seront extraites des voix séparées des forme rythmiques.

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

### 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.



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

### Choix d'une grammaire

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

le lien entre grammaire et signature rythmique n'est pas clair ici. Il aurait fallu expliquer le rôle des grammaires (automates) en 3.3(devenu 3.4)

Groove MIDI Data Set pas présenté

méta-données

contenu

Séparation des voix

The figure shows three musical staves, each with a treble clef and a common time signature 'C'. The first staff is labeled 'Irréductible.' and shows a sequence of eighth notes with 'x' marks above them. The second staff is labeled 'Séparation possible 1.' and shows the same sequence with some notes grouped. The third staff is labeled 'Séparation possible 2.' and shows a different grouping. The fourth staff is labeled 'Séparation possible 3.' and shows a third grouping.

FIGURE 3.14 – Motif 4-4 binaire

Ici, le forme rythmique est construit sur un modèle rock en 4/4 : after-beat sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. La forme rythmique est constituée par défaut du motif rd/pf/cc (voir 3.2) et d'un gamme jouée à la grosse caisse. La première ligne de la figure 3.14 est appelée « Irréductible » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du charley au pied. La troisième séparation proposée est privilégiée car elle répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont utilisés par voix et plus logique pour le lecteur puisque les mains sont en haut et les pieds en bas.

les description ic sont assez techniques et difficile à suivre. avant de détailler des exemples, il faudrait décrire les objectifs et le principe de la procédure.

The figure shows three musical staves, each with a treble clef and a common time signature 'C'. The first staff is labeled 'Motif.' and shows a sequence of eighth notes with 'x' marks above them. The second staff is labeled 'Exemple.' and shows the same sequence with some notes grouped. The third staff is labeled 'Séparation.' and shows a different grouping.

FIGURE 3.15 – Motif 4-4 jazz

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

quel exemple?

1065 combinées avec le *motif* de la première ligne, sont des cas typiques de la  
 1066 batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus, la  
 1067 grosse caisse entre très souvent dans le flot des combinaisons de toms et  
 1068 de caisse claire et son écriture séparée serait inutilement compliquée et  
 1069 peu intuitive pour le lecteur. Le choix de séparation sera donc de laisser  
 1070 les cymbales en haut et toms, caisse claire, grosse caisse et pédale de  
 1071 charley en bas.



FIGURE 3.16 – forme rythmique 4-4 afro-latin

1072 La figure 3.16 montre un exemple minimaliste de forme rythmique afro-  
 1073 latin [26]. Ce forme rythmique doit être écrit sur trois voix car la voix  
 1074 centrale est souvent plus complexe qu'ici (que des noirs) et la mélanger  
 1075 avec le haut ou le bas serait surchargé et peu lisible.

### 1076 Simplification de l'écriture

1077 Les explications qui suivent seront appuyé par une réécriture guidée par  
 1078 une forme rythmique dans la section 4.4.

1079 Les gammes qui accompagnent les motifs d'un forme rythmique étayent  
 1080 toutes les combinaisons d'un forme rythmique et elles permettent, combi-  
 1081 nées avec le motif d'un forme rythmique, de définir les règles de simplifi-  
 1082 cation propres à celui-ci.

1083 Voici les différentes étapes à suivre :

- 1084 — Pour chaque gamme du forme rythmique, faire un arbre de rythme
- 1085 représentant la gamme combinée avec le motif du forme ryth-
- 1086 mique ;
- 1087 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
- 1088 arbre de rythme par voix ;
- 1089 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
- 1090 nœuds qui nécessitent une simplification et écrire la règle.

1091 Certaines précisions concernant l'extraction de ces règles sont néces-  
 1092 saires. Il s'agit de précisions à propos de la durée, des silences et de la  
 1093 présence ou non d'ouverture de charley dans les instruments joués. Nous  
 1094 avons discuté de ces problèmes dans le chapitre 3.

1095 Voici quelques règles inhérentes à la simplification de l'écriture pour la  
 1096 batterie : Toutes les continuations (t) qui se trouvent en début de temps  
 1097 (figures 4.8, 4.10 et ??) sont transformées en silences (r) sauf si la note  
 précédente est un charley ouvert ?

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

1099 Même si on favorise l'usage des silences pour l'écart entre les notes n'appartenant pas au même temps, on les supprime systématiquement pour  
 1100 2 notes au sein d'un même temps et favorise, une liaison si co, un point si  
 1101 pas co et nécessaire, un simple ajustement de la figure de note si suffisant.  
 1102

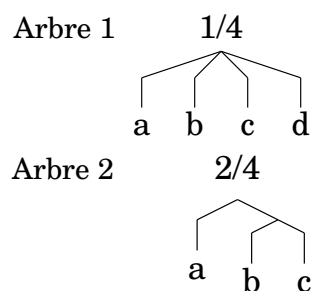


FIGURE 3.17 – Simplification

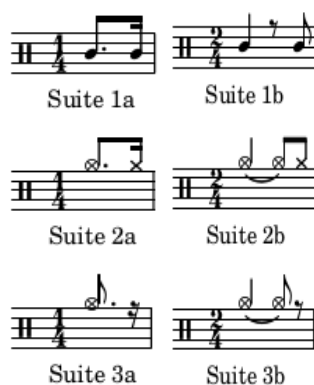


FIGURE 3.18

1103 Soit l'arbre 1 de la figure 3.17 dans lequel : a et d sont des instruments de  
 1104 la batterie (x);  
 1105 b et c sont des continuations (t);  
 1106 Pour chacune des conditions suivantes, une suite de la figure 3.18 est  
 1107 attribuée :

- 1108 — Si a n'est pas un co :  
 1109     ⇒ Suite 1a.
- 1110 — Si a est un co :  
 1111     — Si d est un cf :  
 1112         ⇒ Suite 2a.
- 1113     — Si d est un pf :  
 1114         ⇒ Suite 3a : d deviens un silence (r).

1115

itemize

1116 Soit l'arbre 2 de la figure 3.17 dans lequel :  
1117 a et c sont des instruments de la batterie (x);  
1118 b est une continuation (t); Pour chacune des conditions suivantes, une  
1119 suite de la figure 3.18 est attribuée :  
1120 — Si a n'est pas un co :  
1121     ⇒ Suite 1b, b devient un silence.  
1122 — Si a est un co :  
1123     — Si c est un cf :  
1124         ⇒ Suite 2b, b devient une liaison et c devient un cf.  
1125     — Si c est un pf :  
1126         ⇒ Suite 3b : b devient une liaison et c devient un silence.  
1127  
1128 *Rappel :*  
1129 *cf = charley fermé joué à la main ;*  
1130 *co = charley ouvert joué à la main ;*  
1131 *pf = charley fermé joué au pied.*  
1132

## 1133 Conclusion

1134 <dam>à développer un peu plus</dam> Nous avons formalisé une nota-  
1135 tion de la batterie, modélisé cette notation pour la transcription de don-  
1136 nées MIDI en partition, nous avons décrit Qparse.  
1137 Enfin, nous avons exposé une approche de type dictionnaire (les « forme  
1138 rythmiques ») pour détecter une signature rythmique, choisir une gram-  
1139 maire pondérée appropriée et énoncer des règles de séparation des voix et  
1140 de simplification de l'écriture.



1141

1142

## EXPÉRIMENTATIONS

1143

### Sommaire

1144	4.1	Le jeu de données . . . . .	50
1145	4.2	Analyses et transcriptions manuelles . . . . .	51
1146	4.3	Transcription polyphonique par parsing . . . . .	55
1147	4.4	Réécriture guidée par une forme rythmique . . . . .	58
1148	4.5	BILAN : résultats — évaluation — discussion . . . . .	63
1149			
1150			
1151			
1152			

### 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.

## 1173 4.1 Le jeu de données

1174 Nous avons utilisé le Groove MIDI Dataset<sup>1</sup> [33] (GMD) qui est un jeu de  
 1175 données mis à disposition par Google sous la licence Creative Commons  
 1176 Attribution 4.0 International (CC BY 4.0).

1177 Le GMD est composé de 13,6 heures de batterie sous forme de fichiers  
 1178 MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000  
 1179 mesures de batterie dans les styles les plus courants et avec différentes  
 1180 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](https://www.youtube.com/watch?v=BX1V_IE0g2c)

1181

1182 Autres critères spécifiques au GMD :

- 1183 — Toutes les performances ont été jouées au métronome et à un tempo  
 1184 choisi par le batteur.
- 1185 — 80% de la durée du GMD a été joué par des batteurs professionnels  
 1186 qui ont pu improviser dans un large éventail de styles. Les don-  
 1187 nées sont donc diversifiées en termes de styles et de qualités de jeu  
 1188 (professionnel ou amateur).
- 1189 — Les batteurs avaient pour instruction de jouer des séquences de  
 1190 plusieurs minutes ainsi que des fills<sup>2</sup>
- 1191 — Chaque performance est annotée d'un style (fourni par le batteur),  
 1192 d'une signature rythmique et d'un tempo ainsi que d'une identifi-  
 1193 cation anonyme du batteur.
- 1194 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10  
 1195 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-  
 1196 session du GMD.
- 1197 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur  
 1198 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.

### 1199 **Format des données**

1200 Le Roland TD-11 enregistre les données dans des fichiers MIDI et les  
1201 divise en plusieurs pistes distinctes :

- 1202 — une pour le tempo et l'indication de mesure ;
- 1203 — une pour les changements de contrôle (position de la pédale de  
1204 charley) ;
- 1205 — une pour les notes.

1206

1207 Les changements de contrôle sont placés sur le canal 0 et les notes sur le  
1208 canal 9 (qui est le canal canonique pour la batterie).

1209 Pour simplifier le traitement de ces données, ces trois pistes ont été fu-  
1210 sionnées en une seule piste qui a été mise sur le canal 9.

## 1211 **4.2 Analyses et transcriptions manuelles**

1212 Ces analyses ont été faites dans le cadre de transcriptions manuelles à  
1213 partir de fichiers MIDI et Audio du GMD.

### 1214 **Comparaisons de transcriptions**

1215 Pour les comparaisons de transcriptions, les transcriptions manuelles  
1216 (TM) ont été éditées à l'aide de Lilypond<sup>3</sup> ou MuseScore<sup>4</sup> et les transcrip-  
1217 tions automatiques (TA) ont toutes été générées par import d'un fichier  
1218 MIDI dans MuseScore.

### 1219 **Exemple d'analyse 1**

Transcription manuelle ⇒ Transcription automatique



- 1220 — Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
- 1221 — Les silences de la mesure 1 de la TA sont inutilement surchargés ;
- 1222 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux  
1223 premières notes (une double-croche et une croche) d'un triolet sur  
1224 le temps 1 de la mesure 2 de la TA.

Transcription manuelle  $\Rightarrow$  Transcription automatique



### 1225 Exemple d'analyse 2

- 1226 — Les doubles croches ont été interprétées en quintolet
- 1227 — La deuxième double-croche est devenue une croche.

### 1228 Exemple d'analyse 3

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 1229 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1230
- 1231 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1232 décalés d'une double croche vers la droite dans la TA.
- 1233 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1234 alors qu'elle appartenait à un triolet dans la TM.
- 1235 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1236 pas la TM.
- 1237

### 1238 Exemple d'analyse 4

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 1239
- 1240 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- 1241 manière excessivement complexe!

---

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

1242 **Exemple d'analyse 5 (flas)**

1243 Transcription manuelle



1244

1245 Transcription automatique

1246



1247

1248

- 1249 — Le premier fla est reconnu comme étant un triolet contenant une
- 1250 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1251 ornementée.
- 1252 — Le deuxième fla est reconnu comme étant un accord.
- 1253 — Les deux double en contre-temps sur le temps 4 de la TM sont mal
- 1254 quantifiées dans la TA.
- 1255 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1256 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1257 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1258 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1259 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1260 la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de
- 1261 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1262 temps 3 de la TA.
- 1263 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1264 exemple, le charley ouvert en contre-temps sur le temps 2 de la
- 1265 mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem
- 1266 pour les cross-sticks.

1267 **Conclusion d'analyse**

1268 Ces analyses ont montré la difficulté pour un logiciel comme MuseScore  
 1269 d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI  
 1270 ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance  
 1271 de la forme du rythme impliquant sa position dans la mesure. Cette recon-  
 1272 naissance pourrait permettre de rectifier les problèmes de signature ryth-  
 1273 mique ainsi que les problèmes de décalage de temps. La reconnaissance  
 1274 de la forme du rythme permettrait aussi de supprimer les aberrations du  
 1275 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](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 ghost-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).

### 4.3 Transcription polyphonique par parsing

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

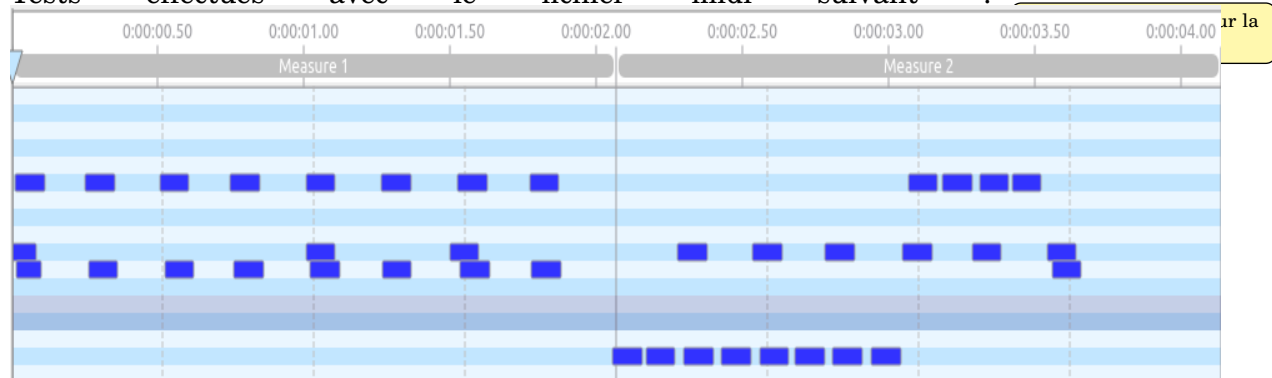
#### Les Jams

Les Jams permettent de passer du monophonique au polyphonique.

il faut revenir ici sur le parsing, et la notion d'alignement sur arbres syntaxiques pour définir les jams. illustrer avec les exemples précédents.

#### Le parsing

Tests effectués avec le fichier midi suivant :



Un premier test convaincant est effectué avec la grammaire suivante :

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

0 -> C0 1

0 -> E1 1

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

// half bar level

9 -> C0 1

9 -> E1 1

// beat level

1 -> C0 1

1 -> E1 1

1 -> T2(2, 2) 1

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

// croche level

2 -> C0 1

2 -> E1 1

```

1345 // double level
1346 4 -> C0 1
1347 4 -> E1 1
1348 4 -> E2 1
1349 4 -> T2(6, 6) 1
1350

```

```

1351 // triple level
1352 6 -> E1 1
1353

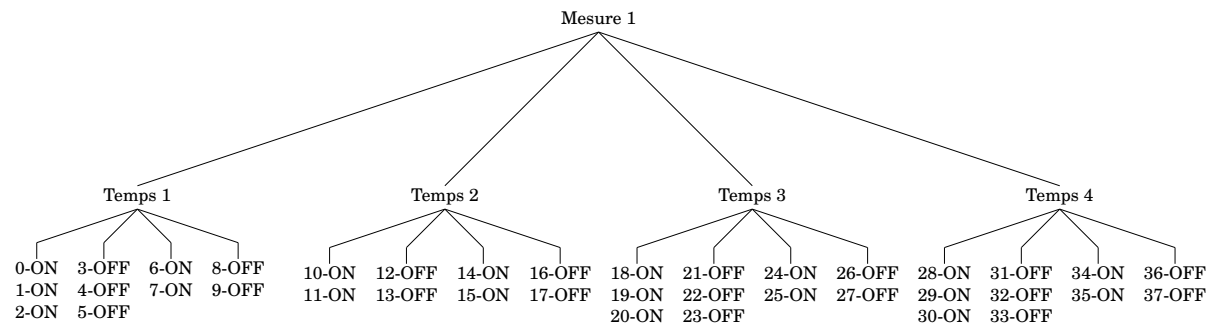
```

1354 Cette grammaire sépare les ligatures par temps au niveau de la  
 1355 mesure. Puis, au niveau du temps, elle autorise les divisions par deux  
 1356 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.  
 1357 L'arbre de parsing en résultant est considéré comme « convaincant » car  
 1358 il découpe correctement les mesures et les temps.

```

1359
1360

```



```

1361
1362

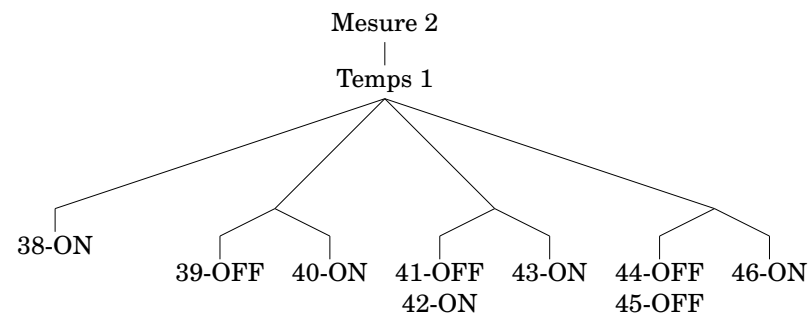
```

1363 Les temps de la première mesure du fichier MIDI sont bien quanti-  
 1364 fié mais ceux de la deuxième mesure présentent quelques défauts de  
 1365 quantification visibles dès le premier temps.

```

1366
1367

```



```

1368
1369

```

1370 Les Onsets sont correctement triés au niveau des doubles croches  
 1371 mais certaines doubles croches sont inutilement subdivisées en triples  
 1372 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps  
 1373 ci-dessus).



1374

1375 **2ème exemple :**

1376 Après une augmentation du poids des triples croches dans la grammaire  
1377 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à  
1378 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets  
1379 sont bien triés et aucuns ne sont subdivisés.

## 1380 4.4 Réécriture guidée par une forme rythmique

1381 La démonstration qui suit est basée sur la partition de référence de la  
 1382 figure 3.11 puisque la forme rythmique qui sera utilisée en est directe-  
 1383 ment extraite.

1384

1385 Nous allons montrer :

- 1386 — la composition de cette forme rythmique ;
- 1387 — son état finale, c'est à dire toutes les combinaisons entièrement
- 1388 écrites en notation correcte sur partition ;
- 1389 ⇒ cela constituera une référence pour la réécriture ;
- 1390 — un exemple de transformation de la forme rythmique en arbre de
- 1391 rythme ;
- 1392 — l'application de la séparation des voix sur cet exemple basé sur la
- 1393 référence citée précédemment (la forme rythmique en question) ;
- 1394 ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a
- 1395 de voix (deux arbres pour cette forme rythmique) ;
- 1396 — les règles de simplification propres à la forme rythmique dont nous
- 1397 parlons.

1398 L'objectif de cette démonstration est de montrer comment un jeu de plu-  
 1399 sieurs formes rythmiques pourrait être implémenter dans le cadre d'une  
 1400 approche dictionnaire.

## 1401 Motifs et gammes

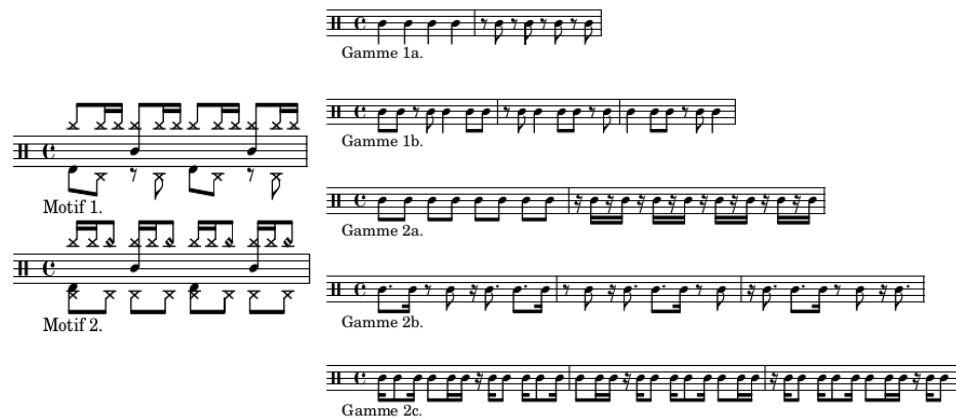


FIGURE 4.2 – Motifs et gammes

## 1402 Motifs

1403 À partir de la partition de référence, les deux motifs de la figure 4.2  
 1404 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure

1405 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à  
 1406 la mesure 28 avec des variations. Ces deux motifs sont très classiques et  
 1407 pourront être détectés dans de nombreuses performances.

1408

#### 1409 **Gammes**

1410 Les gammes de la figure 4.2 étayent toutes les combinaisons d'un motif  
 1411 en 4/4 binaires jusqu'aux doubles croches.

1412 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-  
 1413 mière ne contient que des noires et la deuxième que des croches en contre-  
 1414 temps. Ces deux possibilités sont combinées de manière circulaire dans  
 1415 les 3 mesures de la deuxième ligne.

1416 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures  
 1417 dont la première ne contient que des croches et la deuxième que des  
 1418 doubles-croches en contre-temps. Ces deux possibilités sont combinées de  
 1419 manière circulaire dans les lignes 4 et 5 qui contiennent chacune 3 me-  
 1420 sures.

#### 1421 **Formes rythmiques — motifs et gammes combinés**

1422 Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure  
 1423 4.2.<dam>à commenter un peu plus, notamment pour dire si la combinai-  
 son est faite automatiquement ou non</dam>

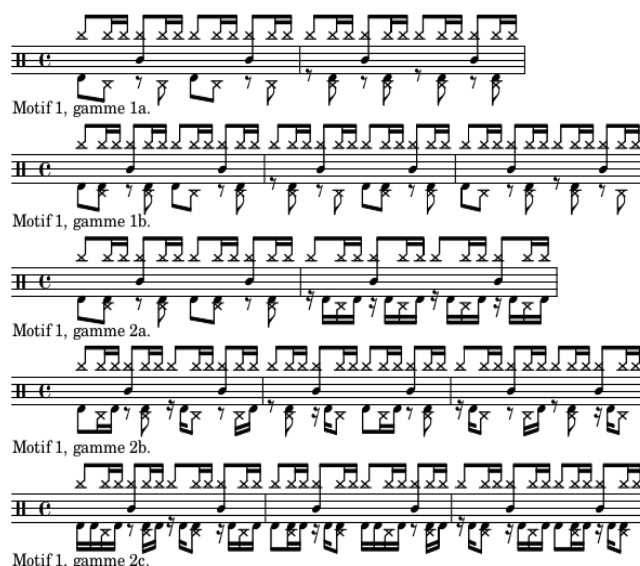


FIGURE 4.3 – Partition d'un forme rythmique en 4/4 binaire

1424

1425 **Représentation de la forme rythmique en arbres de**  
 1426 **rythmes**

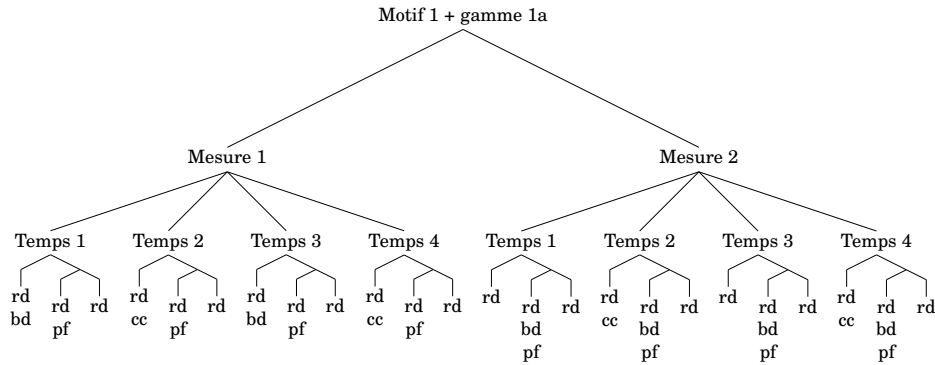


FIGURE 4.4 – Arbre de rythme — forme rythmique

1427 L'arbre de la figure 4.4 servira de base pour le suite de l'expérimentation.  
 1428 Comme indiqué à la racine de l'arbre, il représente la première ligne de la  
 1429 figure 4.3. Même si cet arbre représente parfaitement le rythme concerné,  
 1430 il manque des indications de notation telles que les voix spécifiques à  
 1431 chaque partie du rythme ainsi que les choix d'écriture pour les distances  
 1432 qui séparent les notes de chaque voix entre elles en termes de durée.

1433 **Réécriture — séparation des voix et simplification**

1434 **La séparation des voix**

1435 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui  
 1436 le constituent et les voix seront regroupées en suivant les règles du forme  
 rythmique.

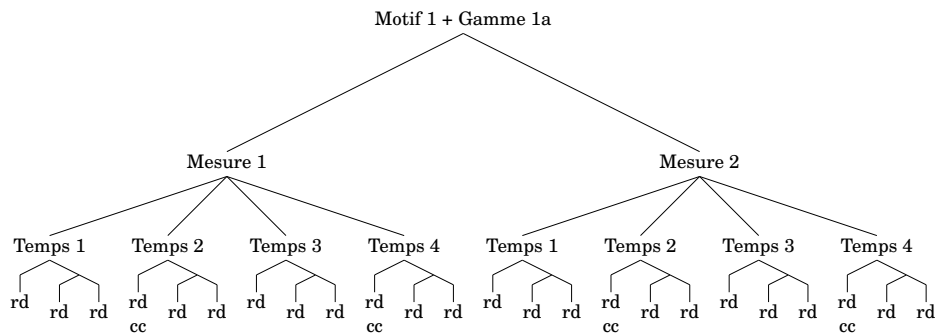


FIGURE 4.5 – Arbre de rythme — voix haute

1437  
 1438 La voix haute (figure 4.5) regroupe la ride et la caisse-claire sur les liga-  
 1439 tures du haut.

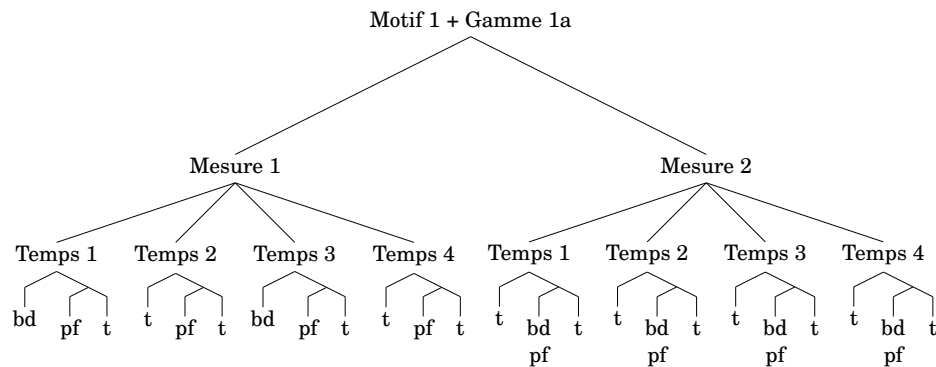


FIGURE 4.6 – Arbre de rythme — voix basse

1440 La voix basse (figure 4.6 regroupe la grosse-caisse et le charley au pied  
1441 sur les ligatures du bas.

1442 **Les règles de simplifications**

1443 L’objectif des règles de simplifications est de réécrire les écarts de durées  
1444 qui séparent les notes d’une manière appropriée pour la batterie et qui  
1445 soit la plus simple possible. Les ligatures relient les notes d’un temps  
1446 entre elles afin de rendre la pulsation visuelle).

1447

1448 Pour les figures ci-dessous :

- 1449 — x = une note ;  
1450 — r = un silence ;  
1451 — t = une continuation (point ou liaison)

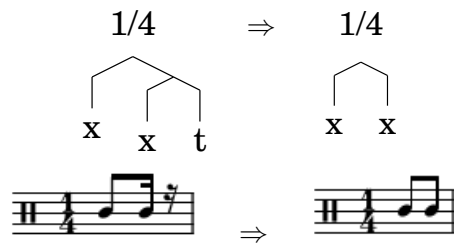


FIGURE 4.7

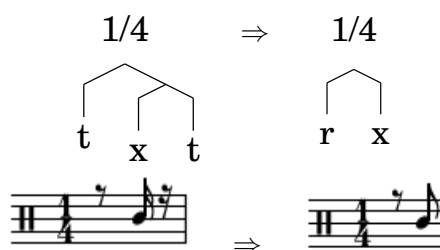


FIGURE 4.8

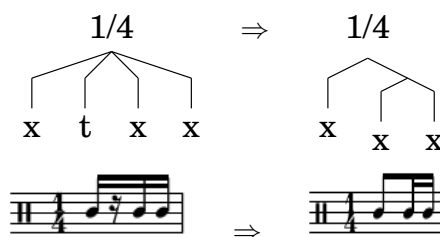


FIGURE 4.9

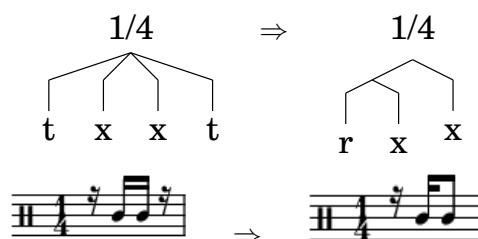


FIGURE 4.10

- 1452 Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique.  
 1453 Les arbres manquants seront mis en annexe.  
 1454 Les règles remplacent par un silence les continuations (t) qui sont au dé-  
 1455 but d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il  
 1456 y a des ouvertures de charley, cela n'est pas toujours applicable.

### 1457 Conclusion sur cette réécriture guidée

- 1458 La méthode des formes rythmiques étant basée sur une approche diction-  
 1459 naire, Le premier objectif de cette réécriture guidée est d'orienter la re-  
 1460 cherche d'autres formes rythmiques par observation du jeu de données et  
 1461 de montrer comment les construire pour agrandir la base de connaissance  
 1462 de Qparse pour la transcription de la batterie.

## 1463 4.5 BILAN : résultats — évaluation — discussion

1464 Cette section regroupe les avancées qui ont été réalisées par rapport aux  
 1465 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-  
 1466 tats de l'ADT avec Qparse. Nous avons amélioré le système de quantifi-  
 1467 cation de Qparse pour la batterie, notamment le passage à la polyphonie  
 1468 avec les Jams.

1469 Nous avons pu obtenir des arbres de parsing corrects en améliorant les  
 1470 grammaires avec des fichiers MIDI courts.

1471 Puis, une sortie MEI a aussi été obtenu (encore à vérifier).

1472 Dans cette section, nous discuterons sur la pertinence de l'ensemble des  
 1473 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui  
 1474 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.

1475 — Le choix de travailler avec Lilypond et non Verovio. Ce choix était  
 1476 motivé par la liberté totale concernant la notation de la batterie  
 1477 dont un et la disponibilité d'un set de notation de type Agostini.  
 1478 C'est la seule application qui me permettait d'écrire la notation de  
 1479 la batterie exactement comme je le souhaitais.

1480 — Avancé de la chaîne de traitement (nous sommes arrivés aux arbres  
 1481 de parsing, nous avons traité le polyphonique (identification des  
 1482 regroupements de notes<sup>5</sup>) ⇒ Quelques arbres ont été obtenus sur  
 1483 des exemples simples (<sup>6</sup>)

1484 — 2 dimensions de le travail fourni :

1485 - La volonté de pousser un exemple simple jusqu'au bout de la  
 1486 chaîne pour obtenir des résultats et une évaluation sur au moins  
 1487 un exemple ; - La réalité du travail à fournir pour faire avancer sur  
 1488 la chaîne de traitement. ⇒ Une solution aurait été de considérer  
 1489 les arbres de parsing obtenus après le traitement du polyphonique  
 1490 comme un résultat local possible à évaluer au lieu d'attendre que la  
 1491 chaîne arrive jusqu'à la génération d'une partition mais cela n'était  
 1492 pas prioritaire pendant le stage.

1493 — Création d'un jeu de forme rythmique basique représentatif des  
 1494 différents styles à recouvrir. Ce jeu n'a pas pu être créé, car comme  
 1495 vu plus haut, je me suis focalisé sur un exemple pour pouvoir le  
 1496 vérifier entièrement et dans l'espoir de pouvoir le tester en fin de  
 1497 chaîne. **Évaluation** Matcher les motifs aurait été indispensable  
 1498 pour obtenir une quantité de résultats qui justifieraient une  
 1499 évaluation automatique permettant de faire des graphiques.

1500 L'évaluation fut entièrement manuelle car :

1501 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf  
 1502 VS output) Pour l'évaluation, il aurait fallu produire un module.

---

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

1503 <dam>je ne sais pas si tu auras encore le temps de faire ça, sinon  
1504 il faudra décrire comment tu aurais aimé évaluer, proprement et  
1505 sans résultats chiffrés</dam> L'évaluation est-elle automatique ou  
1506 manuelle ?  
1507 Possibilité d'un export lilypond en arbre pour comparer l'ouput  
1508 avec la transcription manuelle.  
1509 Possibilité de transformer lilypond(output) et lilypond(ref) en  
1510 ScoreModel ou MEI pour les comparer et faire des statistiques.  
1511 Si transformés en MEI : diffscore de Francesco. Possibilité de  
1512 transformer lilypond(output) et lilypond(ref) en MusicXML pour  
1513 les comparer ou dans Music21. L'expérimentation peut-être consi-  
1514 dérer comme une évaluation manuelle ? (magicien d'Oz)  
1515 Lilypond vers MIDI + ouput vers MIDI  $\Rightarrow$  Comparaison des MIDI  
1516 dumpés.  
1517  
1518 La transcription automatique de la batterie est un sujet passionnant mais  
1519 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-  
1520 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-  
1521 moins été générée. Elle sera un bon point de départ pour les travaux fu-  
1522 turs dont plusieurs propositions sont énoncés dans le présent document.



## CONCLUSION GÉNÉRALE

1524 Dans ce mémoire, nous avons traité de la problématique de la transcrip-  
1525 tion automatique de la batterie. Son objectif était de transcrire, à partir  
1526 de leur représentation symbolique MIDI, des performances de batteur de  
1527 différents niveaux et dans différents styles en partitions écrites.  
1528 Nous avons avancé sur le parsing des données MIDI établissant un pro-  
1529 cessus de regroupement des événements MIDI qui nous a permis de faire  
1530 la transition du monophonique vers le polyphonique. Une des données  
1531 importante de ce processus était de différencier les nature des notes d'un  
1532 accord, notamment de distinguer lorsque 2 notes constituent un accord  
1533 ou un *fla*.  
1534 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-  
1535 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-  
1536 tionnables en amont du parsing, soit par indication des noms des fichiers  
1537 MIDI, soit par reconnaissance de la métrique avec une approche diction-  
1538 naire de patterns prédéfinis <sup>7</sup> qu'il serait pertinent de mettre en œuvre en  
1539 machine learning.  
1540 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre  
1541 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif  
1542 d'un système et pour la simplification grâce aux gammes du motif d'un  
1543 système. Nous avons aussi montré comment, dans des travaux futurs, un  
1544 système dont le motif serait reconnu en amont dans un fichier MIDI pour-  
1545 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-  
1546 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans  
1547 la chaîne de traitement en terme de réécriture.  
1548 Il sera également intéressant d'étudier comment l'utilisation de LM peut  
1549 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération  
1550 entièrement automatisée de partitions de batterie et au problème général  
1551 de l'AMT de bout en bout.[11]

---

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



## BIBLIOGRAPHIE

- 1553 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41  
1554 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,  
1555 1996. – Cité pages 7, 18 et 35.
- 1556 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11  
1557 et 15.
- 1558 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité  
1559 page 12.
- 1560 [4] Gaël Richard et al. De fourier à la reconnaissance  
1561 musicale. Available at [https://interstices.info/  
1562 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).  
1563 – Cité page 12.
- 1564 [5] Caroline Traube. Quelle place pour la science au sein de la musico-  
1565 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1566 [6] Leonard Bernstein Office. The unanswered question : Six talks at  
1567 harvard. Available at [https://leonardbernstein.com/about/  
1568 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1569 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions  
1570 entre les traitements de la musique et du langage. *La Lettre des  
1571 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1572 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.  
1573 Techniques de traitement automatique du langage naturel appli-  
1574 quées aux représentations symboliques musicales. In *JIM 2021 -  
1575 Journées d’Informatique Musicale*, Virtual, France, July 2021. –  
1576 Cité page 13.
- 1577 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-  
1578 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1579 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering  
1580 music relations with sequential attention. In *NLP4MUSA*, 2020. –  
1581 Cité page 13.
- 1582 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger  
1583 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1584 lenges and future directions. *Journal of Intelligent Information Sys-*  
1585 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 65.
- 1586 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J  
1587 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –  
1588 Cité pages 14 et 27.
- 1589 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-  
1590 SURE, 1997. – Cité page 15.
- 1591 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-  
1592 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander  
1593 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*  
1594 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–  
1595 1483, 2018. – Cité pages 15, 23 et 27.
- 1596 [15] Moshekwa Malatji. Automatic music transcription for two instru-  
1597 ments based variable q-transform and deep learning methods, 10  
1598 2020. – Cité page 22.
- 1599 [16] Antti J. Eronen. Musical instrument recognition using ica-based  
1600 transform of features and discriminatively trained hmms. *Seventh*  
1601 *International Symposium on Signal Processing and Its Applications*,  
1602 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1603 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic  
1604 drum sound description for real-world music using template adap-  
1605 tation and matching methods. *International Conference on Music*  
1606 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1607 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local  
1608 musical statistics as guides for audio-to-score piano transcription.  
1609 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24, 26 et 34.
- 1610 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-  
1611 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-  
1612 tization and Score Structuring. In *MCM 2019 - Mathematics and*  
1613 *Computation in Music*, volume Lecture Notes in Computer Science  
1614 of *Proceedings of the Seventh International Conference on Mathema-*  
1615 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June  
1616 2019. Springer. – Cité pages 24 et 26.
- 1617 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-  
1618 ring of rhythm structures. In *Proceedings of the First International*  
1619 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.  
1620 IEEE Computer Society. – Cité page 26.
- 1621 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A  
1622 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1623 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –  
1624 Cité page 26.
- 1625 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating  
1626 equivalent rhythmic notations based on rhythm tree languages. In  
1627 *Third International Conference on Technologies for Music Notation  
1628 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-  
1629 pez Palma and Mike Solomon. – Cité page 26.
- 1630 [23] R. Marxer and J. Janer. Study of regularizations and constraints in  
1631 nmf-based drums monaural separation. In *International Conference  
1632 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,  
1633 02/09/2013 2013. – Cité page 27.
- 1634 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions  
1635 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 29.
- 1636 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue  
1637 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 29.
- 1638 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-  
1639 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -  
1640 94120 Fontenay-sous-Bois, 2000. – Cité pages 31 et 46.
- 1641 [27] Nicolas Guiomard-Kagan. *Traitement de la polyphonie pour l'analyse  
1642 informatique de partitions musicales*. PhD thesis. – Cité page 34.
- 1643 [28] Frédéric Canet. La batterie... mot à mot! Available at [https://](https://rimshotetghostnote.fr/)  
1644 [rimshotetghostnote.fr/](https://rimshotetghostnote.fr/) (2021). – Cité page 36.
- 1645 [29] M. Laurson. Patchwork : a visual programming language and some  
1646 musical applications. 1996. – Cité page 40.
- 1647 [30] Jean Bresson, Carlos Agon, and Gérard Assayag. Openmusic visual  
1648 programming environment for music composition, analysis and re-  
1649 search. – Cité page 40.
- 1650 [31] Dick Grune and Criel JH Jacobs. Parsing techniques. *Monographs  
1651 in Computer Science*. Springer,, page 13, 2007. – Cité page 41.
- 1652 [32] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of  
1653 weighted automata*. Springer Science & Business Media, 2009. –  
1654 Cité page 41.
- 1655 [33] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David  
1656 Bamman. Learning to groove with inverse sequence transforma-  
1657 tions. In *International Conference on Machine Learning (ICML)*,  
1658 2019. – Cité page 50.

