

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 Modélisation pour la transcription . . . . .	37
34	3.3 Qparse . . . . .	38
35	3.4 Les forme rythmiques . . . . .	40
36	<b>4 Expérimentations</b>	<b>47</b>
37	4.1 Le jeu de données . . . . .	48
38	4.2 Analyses et transcriptions manuelles . . . . .	49
39	4.3 Transcription polyphonique par parsing . . . . .	53
40	4.4 Réécriture guidée par une forme rythmique . . . . .	56
41	4.5 BILAN : résultats — évaluation — discussion . . . . .	61
42	<b>Conclusion générale</b>	<b>63</b>
43	<b>Bibliographie</b>	<b>65</b>

## LISTE DES FIGURES

45	1.1	Exemple évènements avec durée . . . . .	16
46	1.2	Critère pour un évènement . . . . .	17
47	1.3	Exemple évènements sans durée . . . . .	17
48	1.4	. . . . .	17
49	1.5	Rapport des figures de notes . . . . .	18
50	1.6	Exemple de partition de piano . . . . .	19
51	1.7	MusicXML . . . . .	19
52	2.1	Transcription automatique <dam>remettre ici la citation de la	
53		capture d'écran avec la page</dam> . . . . .	23
54	2.2	HMM . . . . .	26
55	2.3	arbre_jazz . . . . .	27
56	3.1	Les instruments de la batterie . . . . .	30
57	3.2	Hauteur et têtes de notes . . . . .	30
58	3.3	Point et liaison . . . . .	32
59	3.4	Les silences . . . . .	33
60	3.5	Silence joué . . . . .	34
61	3.6	Équivalence . . . . .	35
62	3.7	Séparation des voix . . . . .	35
63	3.8	Les accents et les ghost-notes . . . . .	36
64	3.9	Exemple pour les accentuations et les ghost-notes . . . . .	36
65	3.10	Présentation de Qparse . . . . .	39
66	3.11	signature rythmique . . . . .	42
67	3.12	Motif 4-4 binaire . . . . .	43
68	3.13	Motif 4-4 jazz . . . . .	43
69	3.14	forme rythmique 4-4 afro-latin . . . . .	43
70	3.15	Simplification . . . . .	44
71	3.16	. . . . .	45
72	4.1	Batterie électronique . . . . .	48
73	4.2	Partition de référence . . . . .	52
74	4.3	Motifs et gammes . . . . .	56
75	4.4	Partition d'un forme rythmique en 4/4 binaire . . . . .	57
76	4.5	Arbre de rythme — forme rythmique . . . . .	58
77	4.6	Arbre de rythme — voix haute . . . . .	58
78	4.7	Arbre de rythme — voix basse . . . . .	59
79	4.8	. . . . .	59
80	4.9	. . . . .	60

81	4.10 . . . . .	60
82	4.11 . . . . .	60

## 83 **LISTE DES TABLEAUX**

84	1.1 speechToText vs AMT . . . . .	13
85	3.1 Noms des instruments de la batterie . . . . .	30
86	3.2 Pitches et instruments . . . . .	37
87	3.3 Systèmes . . . . .	41



## INTRODUCTION GÉNÉRALE

### 89 QUOI?

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

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

111

### 112 POURQUOI?

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

119

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

### 122 COMMENT?

123 → Problématique :

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

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

134

135 → Méthodes :

136 → Contributions :

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

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

147

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

- 149 — une modélisation de la notation de la batterie (fusion de 3.1 et de  
150 3.2) qui était jusqu'à présent inexistante.
- 151 — plusieurs transcriptions manuelles dans le but d'analyser les conte-  
152 nus des fichiers MIDI et Audio (4.2) et de faire des comparaisons  
153 de transcription avec des outils déjà existants<sup>1</sup>.
- 154 — une partition de référence transcrite manuellement sur l'entièreté  
155 d'une performance du jeu de données afin de repérer les éléments  
156 importants pour la modélisation et de faire les liens entre les cri-  
157 tères des données d'*input* avec l'écriture finale (4.2). Cette partition  
158 avait aussi pour objectif d'effectuer des tests et des évaluations.
- 159 — le passage au polyphonique en théorie et en implémentation im-  
160 pliquant la théorie sur la détection de l'identité de notes dans un  
161 Jam<sup>2</sup> et l'implémentation de tests unitaires sur le traitement des  
162 Jams (4.3).
- 163 — la création de grammaires pondérées spécifiques à la batterie (4.3)

164

---

1. MuseScore3

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



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

178

## 179 PLAN

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



188

## CONTEXTE

189

190

### Sommaire

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

195

196

197

### Introduction

199

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

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

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

---

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

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

## 1.1 Langues naturelles et musique en informatique

### COMPUTER MUSIC

L'informatique musicale ou *Computer Music* regroupe l'ensemble des méthodes permettant de créer ou d'analyser des données musicales à l'aide d'outils informatiques [3]. Ce domaine implique l'utilisation de méthodes numériques pour l'analyse et la synthèse de musique<sup>3</sup>, qu'il s'agisse d'informations audio, ou symboliques (aide à l'écriture, transcription, base de partitions...). Un exemple de tâche dans ce domaine pourrait être l'analyse de la structure de la musique et de la reconnaissance des accords<sup>4</sup>.

### RIM

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

### RIM et TAL

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

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

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

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

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

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

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

exemples / illustration de la proximité thématique?

D'autres travaux très récents, ont aussi été révélés lors de la première conférence sur le NLP pour la musique et l'audio (NLP4MusA 2020). Lors de cette conférence, Jiang *et al.* [10] ont présenté leur implémentation d'un modèle de langage musical visant à améliorer le mécanisme d'attention par élément, déjà très largement utilisé dans les modèles de séquence modernes pour le texte et la musique. Le domaine du TAL qui se rapproche le plus du RIM est la reconnaissance de la parole (Speech to text). En effet, la séparation des sources ont des approches similaires dans les deux domaines. De plus, il existe un lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole. La transcription musicale étant la notation d'une œuvre musicale initialement non écrite, l'analogie avec l'écriture de la parole est aisée. Le tableau 1.1 montre des différences et des similitudes entre les deux domaines.

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

TABLE 1.1 – speechToText vs AMT

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

## 286 1.2 La transcription automatique de la musique

### 287 1. OBJECTIF

288 Lorsqu'un musicien est chargé de créer une partition à partir d'un  
289 enregistrement et qu'il écrit les notes qui composent le morceau en  
290 notation musicale, on dit qu'il a créé une transcription musicale de cet  
291 enregistrement. L'objectif de la TAM [11] est de convertir la performance  
292 d'un musicien en notation musicale — à l'instar de la conversion de la  
293 parole en texte dans le traitement du langage naturel. Cette définition  
294 peut être comprise de deux manières différentes selon les articles scien-  
295 tifiques : 1) Processus de conversion d'un enregistrement audio en une  
296 notation pianoroll (une représentation bidimensionnelle des notes de  
297 musique dans le temps) 2) Processus de conversion d'un enregistrement  
298 en notation musicale commune<sup>8</sup> (c'est-à-dire une partition).

299

### 300 2. APPLICATIONS

301 La TAM a des applications multiples [11] dont la plus directe est de don-  
302 ner la possibilité à un musicien de générer la partition d'une improvisa-  
303 tion en temps réel afin de pouvoir reproduire sa performance ultérieure-  
304 ment. Une autre application notable est la préservation du patrimoine  
305 par exemple dans les styles musicaux où il n'existe peu de partitions (le  
306 jazz, la pop, les musiques de tradition orale<sup>9</sup>, ...). La TAM est aussi utile  
307 pour la recherche et l'annotation automatique d'informations musicales,  
308 pour l'analyse musicologique<sup>10</sup> ou encore pour les systèmes musicaux in-  
309 teractifs.

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

315 Mais l'intérêt de la TAM est aussi d'avoir des partitions au contenu  
316 exploitable, avec des formats texte ou XML (entre autres...) dont les  
317 données sont manipulables, contrairement à de simples images en pdf<sup>11</sup>.

318

### 319 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

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

---

8. Ici, on parle de notation occidentale.

9. ethno-musicologie

10. par exemple par la constitution de corpus musicologiques

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

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

### 1.3 La transcription automatique de la batterie

La batterie est née au début du vingtième siècle [13]. C'est donc un instrument récent qui s'est longtemps passé de partition. En effet pour un batteur, la qualité de lecteur lorsqu'elle était nécessaire, résidait essentiellement dans sa capacité à lire les partitions des autres instrumentistes (par exemple, les grilles d'accords et la mélodie du thème en jazz) afin d'improviser un accompagnement approprié que personne ne pouvait écrire pour lui à sa place.

Les partitions de batterie sont arrivées par nécessité avec la pédagogie et l'émergence d'écoles de batterie partout dans le monde. Un autre facteur qui a contribué à l'expansion des partitions de batterie est l'émergence de la musique assistée par ordinateur (MAO). En effet, l'usage de boîtes à 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].

12. Roland TR-808

13. SQ-1

cite méthode et école Agostini?

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

## 367 1.4 Les représentations de la musique

citer M. Müller FMP pour  
 cette section ?

### 369 Les données audio

369 trop technique. ne pas re-  
 370 pier wikipédia

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

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

### 375 Les données MIDI

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

377 Le MIDI<sup>14</sup> (Musical Instrument Digital Interface) est une norme tech-  
 378 nique qui décrit un protocole de communication, une interface numérique  
 379 et des connecteurs électriques permettant de connecter une grande va-  
 380 riété d'instruments de musique électroniques, d'ordinateurs et d'appareils  
 381 audio connexes pour jouer, éditer et enregistrer de la musique.

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

382 Les données midi sont représentées sous forme de piano-roll. Chaque  
 383 point sur la figure 1.1 est appelé « événement MIDI » :

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

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

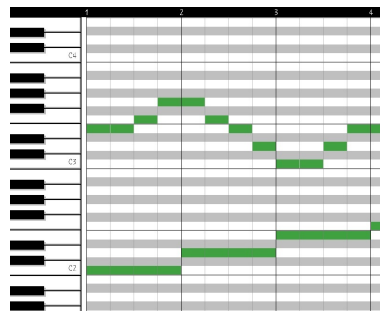


FIGURE 1.1 – Exemple événements avec durée

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

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

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

389 Ici, définir Tick et channel.

390 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

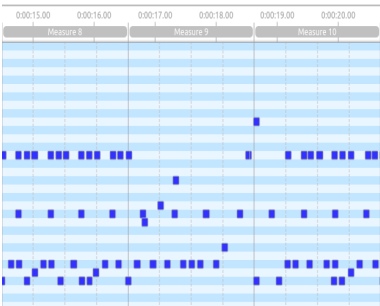


FIGURE 1.3 – Exemple évènements sans durée

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

402 **Les partitions**  
403



FIGURE 1.4

404 La figure 1.4 montre 4 figures de notes les plus courantes dont les noms  
405 et les durées sont respectivement, de gauche à droite :  
406 — La ronde, elle vaut 4 ;

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

407 — La blanche, elle vaut 2 ;

408 — La noire, elle vaut 1 ;

409 — La croche, elle vaut 1/2.

plusieurs éléments

410

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 :

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

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

barre verticale liée à la tête de note

415  
416

haut ou bas

417

418

419

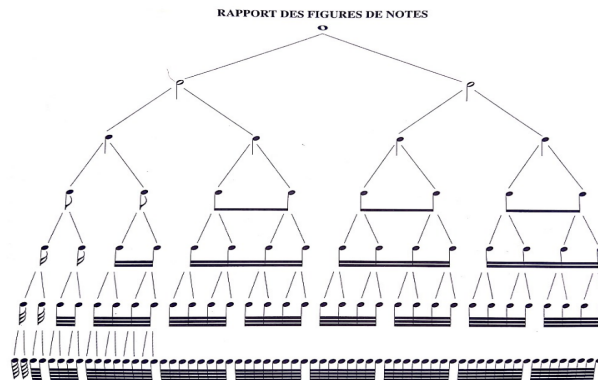


FIGURE 1.5 – Rapport des figures de notes

[1]

420 La figure 1.5 montre les rapports de durée entre les figures de notes. Plus  
 421 les durées sont longues, plus elles sont marquées par la tête de note (la  
 422 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de  
 423 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute  
 424 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.  
 425 Les notes à crochet (croche, double-croche, triple-croche...) peuvent être  
 426 reliées ou non par des ligatures (voir les 4 dernières lignes de la figure  
 1.5).

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

427

428

429

430

Une partition de musique <sup>16</sup> est un document qui porte la représentation  
 systématique du langage musical sous forme écrite. Cette représentation  
 est appelée transcription et elle sert à traduire les quatre caractéristiques  
 du son musical :

— la hauteur ;

— la durée ;

431

432

433

15. [https://fr.wikipedia.org/wiki/Note\\_de\\_musique](https://fr.wikipedia.org/wiki/Note_de_musique)

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



FIGURE 1.6 – Exemple de partition de piano

- 434 — l'intensité;  
 435 — le timbre.  
 437 Ainsi que de leurs combinaisons appelées à former l'ossature de l'œuvre  
 438 musicale dans son déroulement temporel, à la fois :  
 439 — diachronique (succession des instants, ce qui constitue en musique  
 440 la mélodie);  
 442 — et synchronique (simultanéité des sons, c'est-à-dire l'harmonie).

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

## 443 Les formats XML

- 444 Il existe plusieurs formats XML dédiés à la musique : MusicXML, MEI,  
 445 MNX, ...  
 446 L'inconvénient de ces formats est qu'ils sont verbeux et ambigus, c'est  
 447 pourquoi nous utilisons pour la transcription une représentation inter-  
 448 médiaire abstraite décrite plus loin.

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



FIGURE 1.7 – MusicXML

- 449 Le figure 1.7<sup>17</sup> représente un do en clef de sol de la durée d'une ronde  
 450 sur une mesure en 4/4 écrit au format MusicXML. Un des avantages de  
 451 ce format est qu'il peut être converti aussi bien en données MIDI qu'en  
 452 partition musicale, ce qui en fait une interface homme/machine.

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

453 **appogiatures**

454 <flo>Parler des appogiatures ici?</flo>

455 **signature rythmique**

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

457 **Conclusion**

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

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

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

## ÉTAT DE L'ART

## Sommaire

470	2.1	Monophonique et polyphonique . . . . .	21
471			
472	2.2	Audio vers MIDI . . . . .	22
473	2.3	MIDI vers partition . . . . .	24
474	2.4	Approche linéaire et approche hiérarchique . . . . .	24
475			
476			
477			

## 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.

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

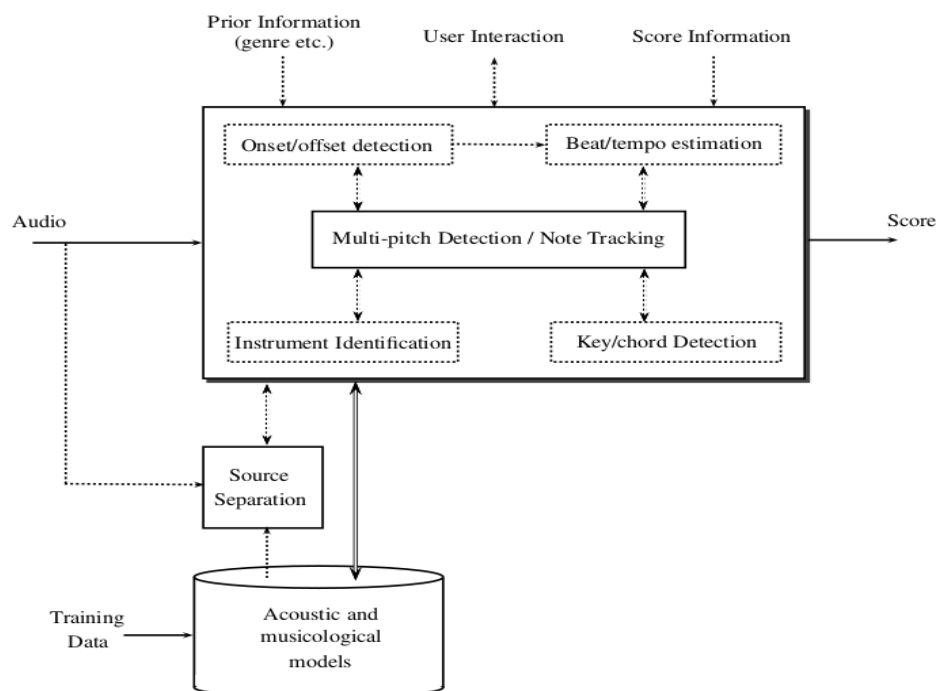


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

*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.*

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

haute fréquence, aigus?

551 pré-processing.  
 552 Pour la reconnaissance des instruments, une approche possible [16] est  
 553 de mettre un modèle probabiliste dans l'étape de la classification des évè-  
 554 nements afin de classer les différents sons de la batterie. Cette méthode  
 555 permet de se passer de samples audio isolés en modélisant la progression  
 556 temporelle des *features*<sup>3</sup> avec un modèle de markow caché (HMM). Les  
 557 *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- 554  
 ments? la phrase semble 555  
 redondante

pas clair... peut-être just 558  
 mentionner les modèles 559  
 probabilistes utilisés

## 562 2.3 MIDI vers partition

563 Le plus souvent, lorsque les articles abordent la transcription automa-  
 564 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »  
 565 (partition) désigne un ouput au format Music XML, ou simplement MIDI.  
 566 Par exemple, dans [18], la chaîne de traitement va jusqu'à la génération  
 567 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en  
 568 extraire manuellement un fichier MusicXML contenant plusieurs voix.  
 569 Seuls quelques travaux récents s'intéressent de près à la création d'outils  
 570 permettant la génération de partition. Le problème de la conversion d'une  
 571 séquence d'évènements musicaux symboliques en une partition musicale  
 572 structurée est traité notamment dans [19]. Ce travail, qui vise à résoudre  
 573 en une fois la quantification rythmique et la production de partition struc-  
 574 turée, s'appuie tout au long du processus sur des grammaires génératives  
 575 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-  
 576 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été  
 577 menées avec un ensemble de données composé d'extraits monophoniques ;  
 578 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 563  
 cela. cf. proposition de des- 564  
 cription + détaillée en com- 565  
 mentaires

de manière conjointe 574

langage a priori 576

qui nécessite de traiter le 579  
 problème supplémentaire 580  
 de la séparation de voix. i.e. 581  
 pour la batterie on nveut 582  
 quantification + structu- 583  
 ration + séparation mais 584  
 seules les 2 premières sont 585  
 couplées dans l'approche de 586  
 tonn stage.

## 2.4 Approche linéaire et approche hiérarchique

584 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par  
 585 exemple, Shibata *et al.* [18] ont utilisé le modèle de Markov caché (HMM)<sup>4</sup>  
 586 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)



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

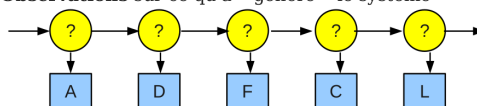


FIGURE 2.2 – HMM

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

594

595

596 L'évaluation finale des résultats de [18] montre qu'il faut rediriger l'atten-  
 597 tion vers les valeurs des notes, la séparation des voix et d'autres éléments  
 598 délicats de la partition musicale qui sont significatifs pour l'exécution de  
 599 la musique. Or, même si la quantification du rythme se fait le plus souvent  
 600 par la manipulation de données linéaires allant notamment des *real time*  
 601 *units* (secondes) vers les musical *time units* (temps, métrique, ...), de nom-  
 602 breux travaux suggèrent d'utiliser une approche hiérarchique puisque le  
 langage musical est lui-même structuré. En effet, l'usage d'arbres syn-  
 taxiques est idéale pour représenter le langage musical. Une méthodolo-  
 gie simple pour la description et l'affichage des structures musicales est  
 présentée dans [20]. Les RT y sont évoqués comme permettant une co-  
 hésion complète de la notation musicale traditionnelle avec des notations  
 plus complexes. Jacquemard *et al.* [21] propose aussi une représentation  
 formelle du rythme, inspirée de modèles théoriques antérieurs issus du  
 domaine de la réécriture de termes. Ils démontrent aussi l'application des  
 arbres de rythmes pour les équivalences rythmiques dans [22]. La réécri-  
 ture d'arbres, dans un contexte de composition assistée par ordinateur,  
 par exemple, pourrait permettre de suggérer à un utilisateur diverses  
 notations possibles pour une valeur rythmique, avec des complexités dif-  
 férentes.

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

je ne comprend pas bien 603  
 l'explication. le pb est plu-  
 604 tot vue locale (déduction de  
 la proba d'une durée à par-  
 605 tir de la durée précédente,  
 par ex. dans un HMM) vs  
 606 vue globale, dans une hié-  
 607 rarchie

RT? 608  
 609

techniques de réécriture 610  
 appliquée à la déduction 611  
 automatique, calcul symbo-  
 612 lique

le calcul d'équiv. 613  
 614

citer thèse de David Rizo 617  
 (Valencia) 618

619

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





## MÉTHODES

### Sommaire

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

### 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/>

## 668 Les hauteurs et les têtes de notes



FIGURE 3.1 – Les instruments de la batterie

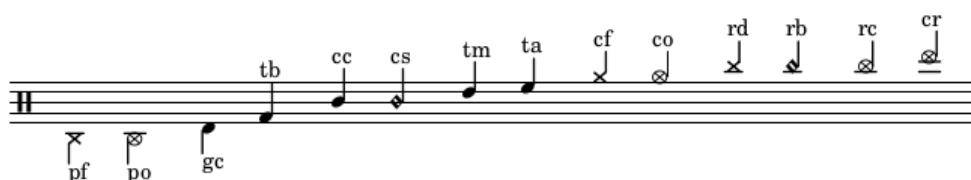


FIGURE 3.2 – Hauteur et têtes de notes

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

pour aider, tu pourrais dessiner une figure représentant la batterie avec le nom des instruments et abréviation.

La figure 3.1<sup>2</sup> montre une batterie standard avec tous les instruments habituellement présent sur une batterie. Le tableau 3.1 donne dans l'ordre :

1. les noms des instruments sur la figure 3.1 ;
2. leurs codes respectifs dans la figure 3.2 ;

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

- 674 3. les noms que j'utiliserai dans le présent document pour y référer.  
 675 En observant les figures 3.1, 3.1 et 3.2, le les hauteurs sur la portée re-  
 676 présentent :
- 677 — La hauteur physique des instruments :  
 678 La caisse claire est centrale sur la portée et sur la batterie (au  
 679 niveau de la ceinture, elle conditionne l'écart entre les pédales et  
 680 aussi la position de tous les instruments basiques d'une batterie).  
 681 Tout ce qui en-dessous de la caisse-claire sur la portée est en  
 682 dessous de la caisse-claire sur la batterie (pédales, tom basse);  
 683 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est  
 684 aussi sur la batterie.
  - 685 — La hauteur des instruments en terme de fréquences :  
 686 Sauf pour le charley au pied et si l'on sépare en trois groupes  
 687 (grosse-caisse, toms et cymbales), de bas en haut, les instruments  
 688 vont du plus grave au plus aigu.
- 689 Les noms des instruments correspondant aux codes des notes de la figure  
 690 ?? sont dans le tableau 3.2.

têtes de notes?

## 692 Les durées

693 Comme nous venons de la voir, la majorité des instruments de la batterie  
 694 sont représentés par les têtes des notes. De plus, le seul instrument dont  
 695 le son peut être arrêté de manière quantifiée et dont la durée sonore nous  
 696 intéresse est le charley<sup>3</sup>. Par conséquent : 1) les durées — sauf pour le  
 697 charley — représenteront un écart temporel entre les notes et non une  
 698 durée sonore et elles pourront donc être rallongée à l'aide de silences ;  
 699 2) les symboles rythmiques concernant les têtes de note ne pourront  
 700 pas être utilisés pour exprimer les durées. Cela est valable aussi pour  
 701 la présence ou non de la hampe puisque ce phénomène n'existe qu'avec  
 702 les têtes de notes de type cercle-vidé (opposition blanche-ronde). L'usage  
 703 des blanches existe dans certaines partitions de batterie [26] mais cela  
 704 reste dans des cas très rares. Certains logiciels permettent de faire des  
 705 blanches avec des symboles spécifiques à la batterie ou aux percussions  
 706 mais leur lecture reste peu aisée et leur utilisation pour la batterie est  
 707 rarissime.

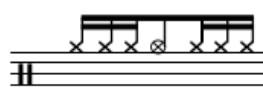
certaines têtes de notes  
vides alors que leur du-  
rée n'est pas celle des  
blanches? expliquer les dif-  
férences avec la notation  
conventionnelle cf 1.4

708  
 709 En résumé :

- 710 — toutes les notes ont une hampe ;
- 711 — une notes dont la hampe n'a pas de crochet est toujours une noire ;
- 712 — à part pour le charley ouvert, les durées n'expriment pas la durée
- 713 d'un son mais une distance temporelle entre deux notes.

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

- 714 — à part pour le charley ouvert, la durée d'une note peut être prolon-  
 715 gée par un silence (exemple : une noire + un soupir pour exprimer  
 716 une blanche)  
 717  
 718 La durée d'une note peut être prolongée par divers symboles :  
 719 — Le point (explication en-dessous de la figure 3.3 ;  
 720 — La liaison (explication en-dessous de la figure 3.3 ;  
 721 — les silences (pas pour les ouvertures de charley).



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

- 722 L'écriture de la batterie doit faire ressortir la pulsation<sup>4</sup>. La première  
 723 chose à prendre en compte pour analyser la figure 3.3 est donc la nécessité  
 724 de regrouper les notes par temps à l'aide des ligatures.  
 725 Exemple 1 : ouverture de charley quantifiée mais pas notes pas regrou-  
 726 pées par temps.  
 727 Exemple 2 : Ici, la liaison permet de regrouper les notes par temps en ob-  
 728 tenant le même rythme que dans l'exemple 1.  
 729 Exemple 3 et exemple 4 : les deux exemples sont valables mais le  
 730 deuxième est le plus souvent utilisé car plus intuitif (regroupement par  
 731 temps).  
 732 En cas de nécessité de prolonger la durée d'une note au-delà de sa durée  
 733 initial, et si cette note correspond à une ouverture de charley, on privilé-  
 734 giera la liaison.

---

4. La position des temps



Les silences

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

expliquer la notation (générale) des silebces en §1.4?

quantifier = noter? ou quantifier la durée?

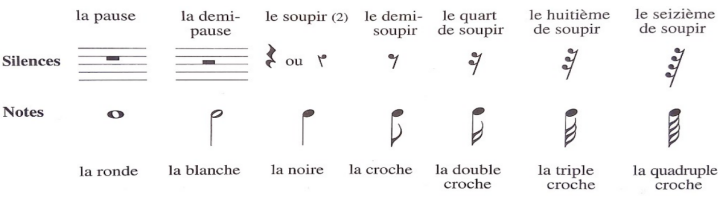


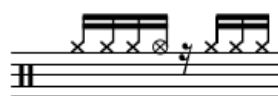
FIGURE 3.4 – Les silences

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

pas très clair



Exemple 1



Exemple 2

FIGURE 3.5 – Silence joué

745 L'exemple 1 de la figure 3.5 montre ce qui est écrit dans les données MIDI  
 746 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-  
 747 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1  
 748 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-  
 749 truments jouant simultanément.

750 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte  
 751 la note suivante pour l'écriture : - Si c'est un charley fermé joué à la main  
 752  $\Rightarrow$  la note sera cf;

itemize

753 - Si c'est un charley fermé joué au pied  $\Rightarrow$  la note sera un silence.

cf?

## 754 Les équivalences rythmiques

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

phrase alambiquée... pour  
prolonger la durée?seuls comptent les dates  
début de notes onsets.

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

766 Sur la figure 3.6, théoriquement, il faudra choisir la notation de la  
 767 deuxième mesure mais dans certains contextes, pour des raisons de lisi-  
 768 bilité ou de surcharge, la version sans les silences de la troisième mesure  
 769 pourra être choisie.



FIGURE 3.6 – Équivalence

## Les voix

Les voix<sup>5</sup> désignent les différentes parties mélodiques constituant une composition musicale et destinées à être interprétées, simultanément ou successivement, par un ou plusieurs musiciens. En batterie, une voix est l'ensemble des instruments qui, à eux seuls, constituent une phrase rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant possibles pour un même rythme, on peut regrouper les instruments de la batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou 2 voix. Sur la figure 3.7, il faudra faire un choix entre les exemples 1, 2 et 3 qui sont trois façons d'écrire le même rythme.

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

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

une voix est caractérisée aussi par orientation des hampes?



FIGURE 3.7 – Séparation des voix

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

5. [https://fr.wikipedia.org/wiki/Voix\\_\(polyphonie\)](https://fr.wikipedia.org/wiki/Voix_(polyphonie))

## 784 Les accentuations et les ghost-notes

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



FIGURE 3.8 – Les accents et les ghost-notes

787

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

788 La figure 3.8 ne prend en compte que les accents que nous avons estimés  
789 nécessaires (voir la section 3.2). Les accents sont marqués par le sym-  
790 bole « > ». Il est positionné au-dessus des notes représentant des cym-  
791 bales et en-dessous des notes représentant des toms ou la caisse-claire.  
792 Ce choix a été fait pour la partition de la figure 4.2 car elle est plus li-  
793 sible ainsi, mais ces choix devront être adaptés en fonction des différents  
794 forme rythmiques reconnus (voir la section 3.4). Par exemple, pour les  
795 forme rythmiques jazz, les ligatures pour les toms et la caisse-claire se-  
796 ront dirigés vers le bas, il faudra donc mettre les symboles d'accentuation  
797 correspondants au-dessus des têtes de notes.

expliquer ce qu'est une  
ghost-notes

798 La dernière note de la figure 3.8 montre un exemple de ghost-notes. Le  
799 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note  
sans changer la tête de note.

800

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

801 Pour les codes, on prend le code de la note et on ajoute un « a » pour un  
802 accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.8  
sont exposées en situation réelle dans la figure 3.9.



FIGURE 3.9 – Exemple pour les accentuations et les ghost-notes

803

## 804 Les flas

805 ICI, DESCRIPTION DES FLAS!

3.2 Modélisation pour la transcription

Les pitches

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

TABLE 3.2 – Pitches et instruments

Il existe, pour de nombreux instruments de la batterie, plusieurs samples audio associés à des pitches. Pour cette première version, nous avons choisi de n'avoir qu'un code-instrument pour différentes variantes d'un instrument, c'est pourquoi certain code-instrument se voit attribuer plusieurs pitches dans le tableau 3.2.

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

je ne comprend pas cette phrase.

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

La vélocité

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

Cette transcription nous a mené aux observations suivantes :

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

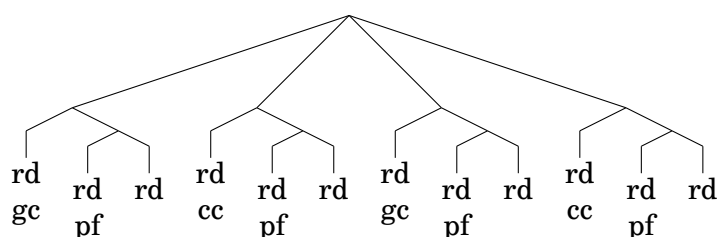
citation lilypond

et l'analyse d'autre fichiers MIDI?

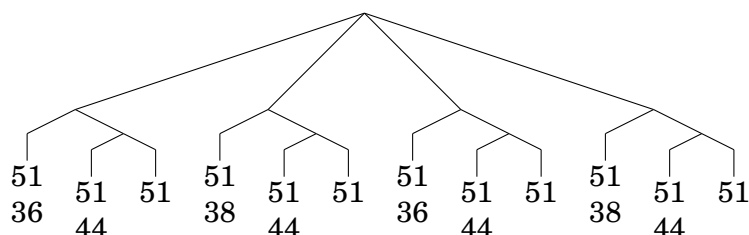
- 828 — Les accents et les ghosts-notes ne sont significatifs ni pour les ins-  
 829 truments joués au pied, ni pour les cymbales crash.  
 830 En effet, certaines vélocités en dessous de 40 étant détectées et ins-  
 831 crites dans les données MIDI sont dues au mouvement du talon du  
 832 batteur qui bat la pulsation sans particulièrement jouer le charley.  
 833 Ce mouvement est perçu par le capteur de la batterie électronique  
 834 mais le charley n'est pas joué.  
 835 — Au final, nous avons relevé les ghost-notes et les accents pour la  
 836 caisse-claire ainsi que les accents pour les toms et les cymbales  
 837 rythmiques (charley et ride).

### 838 Les arbres de rythmes

839 Les arbres de rythmes représentent un rythme unique dont les possibili-  
 840 tés de notation sur une partition sont théoriquement multiples.  
 841 Voici une représentation de la figure 3.7 en arbre de rythmes avec les  
 842 codes de chaque instrument :



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



844 Chacun des trois exemples de la figure 3.7 est représenté par un des deux  
 845 arbres syntaxiques ci-dessus.  
 846

## 847 3.3 Qparse

848 La librairie Qparse<sup>6</sup> implémente la quantification des rythmes basée  
 849 sur des algorithmes d'analyse syntaxique pour les automates arbores-  
 850

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

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

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

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

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

quantification rythmique  
 + structuration de partition

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

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

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

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

Les principaux contributeurs sont :

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

apprentissage

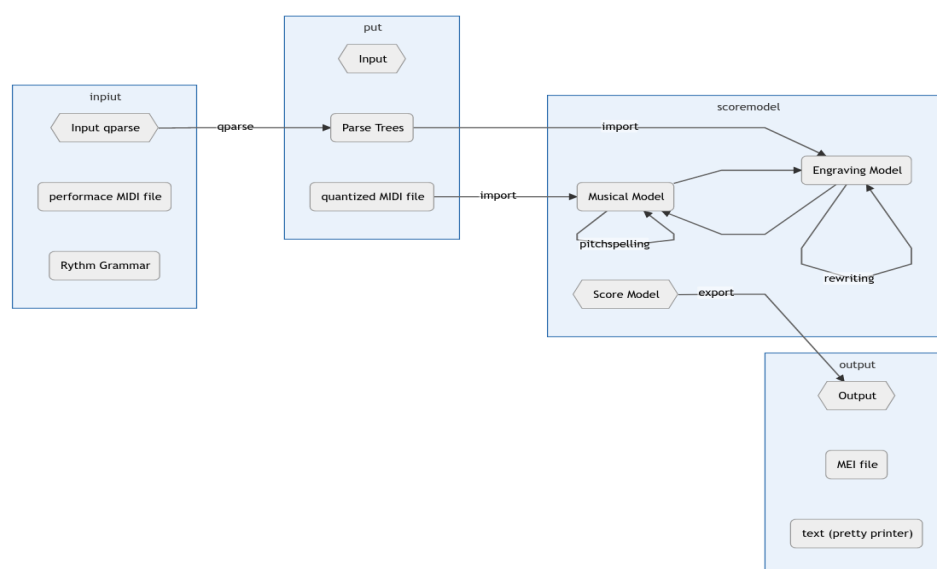


FIGURE 3.10 – Présentation de Qparse

Explication des différentes étapes de la figure 3.10<sup>7</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

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

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

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

### 3.4 Les 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 :

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 fait par réécriture du motif (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



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

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

- 918 — 4/4 binaire
- 919 — 4/4 jazz
- 920 — 4/4 afro-cubain

921 **Objectif des forme rythmiques**

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

- 923 — définir une signature rythmique ;
- 924 — choisir une grammaire appropriée ;
- 925 — fournir les règles de réécriture (séparation des voix et simplifica-  
926 tion.

928

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

bien. il faudrait expliquer  
ça avant.

934 **Détection d'indication de mesure**

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

939 La figure 3.11 montre deux indications de mesure différentes. L'une  
940 (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des trio-  
941 lets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est tra-  
942 ditionnellement écrit en binaire avec ou sans triolet (même si cette mu-  
943 sique est dite ternaire alors que le rock ternaire sera plutôt écrit comme  
944 dans l'exemple 2).

pas exactement. les règles  
de projection et simplifica-  
tion font la séparation en  
voix : à partir d'un arbre  
syntactique comme celui de  
3.2, elles extraient 2 arbres,  
chacun contenant les éven-  
nements d'une seule voix



Exemple 1



Exemple 2

FIGURE 3.11 – signature rythmique

### 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.

### Séparation des voix

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 3.12 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.

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 combinées avec le *motif* de la première ligne, sont des cas typiques de

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

Groove MIDI Data Set présenté

méta-données

contenu

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.

quel exemple?

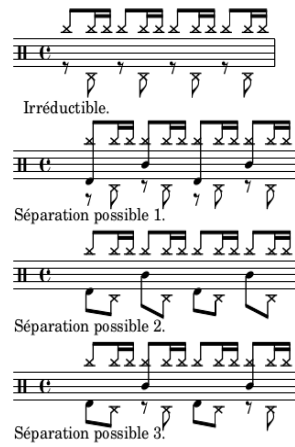


FIGURE 3.12 – Motif 4-4 binaire

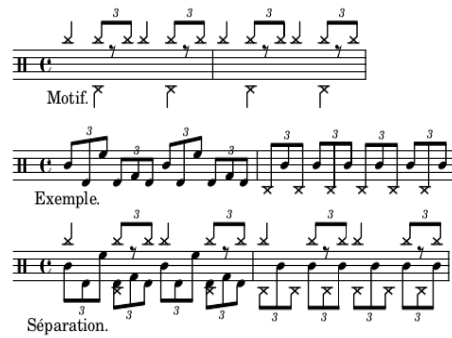


FIGURE 3.13 – Motif 4-4 jazz

la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus, la grosse caisse entre très souvent dans le flot des combinaisons de toms et de caisse claire et son écriture séparée serait inutilement compliquée et peu intuitive pour le lecteur. Le choix de séparation sera donc de laisser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de charley en bas.



FIGURE 3.14 – forme rythmique 4-4 afro-latin

La figure 3.14 montre un exemple minimaliste de forme rythmique afro-latin [26]. Ce forme rythmique doit être écrit sur trois voix car la voix

centrale est souvent plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas serait surchargé et peu lisible.

### Simplification de l'écriture

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

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

Voici les différentes étapes à suivre :

- Pour chaque gamme du forme rythmique, faire un arbre de rythme représentant la gamme combinée avec le motif du forme rythmique ;
- Pour chaque arbre de rythmes obtenus, séparer les voix et faire un arbre de rythme par voix ;
- Pour chaque voix (arbre de rythmes) obtenus, extraire tous les nœuds qui nécessitent une simplification et écrire la règle.

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

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

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 2 notes au sein d'un même temps et favorise, une liaison si co, un point si pas co et nécessaire, un simple ajustement de la figure de note si suffisant.

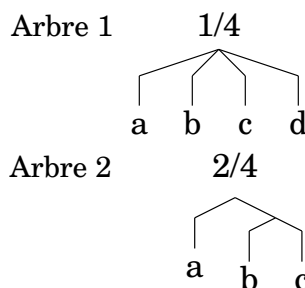


FIGURE 3.15 – Simplification

ce sont des figures et notations du chapitre suivant!

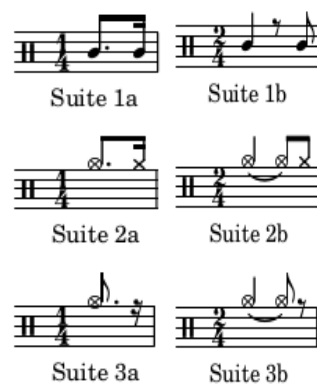


FIGURE 3.16

1013 Soit l'arbre 1 de la figure 3.15 dans lequel : a et d sont des instruments de  
 1014 la batterie (x);  
 1015 b et c sont des continuations (t);  
 1016 Pour chacune des conditions suivantes, une suite de la figure 3.16 est  
 1017 attribuée :

- 1018 — Si a n'est pas un co :
- 1019     ⇒ Suite 1a.
- 1020 — Si a est un co :
- 1021     — Si d est un cf :
- 1022         ⇒ Suite 2a.
- 1023     — Si d est un pf :
- 1024         ⇒ Suite 3a : d devient un silence (r).

1025

1026 Soit l'arbre 2 de la figure 3.15 dans lequel :  
 1027 a et c sont des instruments de la batterie (x);  
 1028 b est une continuation (t); Pour chacune des conditions suivantes, une  
 1029 suite de la figure 3.16 est attribuée :

- 1030 — Si a n'est pas un co :
- 1031     ⇒ Suite 1b, b devient un silence.
- 1032 — Si a est un co :
- 1033     — Si c est un cf :
- 1034         ⇒ Suite 2b, b devient une liaison et c devient un cf.
- 1035     — Si c est un pf :
- 1036         ⇒ Suite 3b : b devient une liaison et c devient un silence.

1037

1038 *Rappel :*

- 1039 *cf* = charley fermé joué à la main ;
- 1040 *co* = charley ouvert joué à la main ;
- 1041 *pf* = charley fermé joué au pied.

1042

## 1043 **Conclusion**

1044 <dam>à développer un peu plus</dam> Nous avons formalisé une nota-  
1045 tion de la batterie, modélisé cette notation pour la transcription de don-  
1046 nées MIDI en partition, nous avons décrit Qparse.  
1047 Enfin, nous avons exposé une approche de type dictionnaire (les « forme  
1048 rythmiques ») pour détecter une signature rythmique, choisir une gram-  
1049 maire pondérée appropriée et énoncer des règles de séparation des voix et  
1050 de simplification de l'écriture.

1051

1052

# EXPÉRIMENTATIONS

1053

## Sommaire

1054	4.1	Le jeu de données . . . . .	48
1055	4.2	Analyses et transcriptions manuelles . . . . .	49
1056	4.3	Transcription polyphonique par parsing . . . . .	53
1057	4.4	Réécriture guidée par une forme rythmique . . . . .	56
1058	4.5	BILAN : résultats — évaluation — discussion . . . . .	61
1059			
1060			
1061			
1062			

## 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.

## 1083 4.1 Le jeu de données

1084 Nous avons utilisé le Groove MIDI Dataset<sup>1</sup> [27] (GMD) qui est un jeu de  
 1085 données mis à disposition par Google sous la licence Creative Commons  
 1086 Attribution 4.0 International (CC BY 4.0).

1087 Le GMD est composé de 13,6 heures de batterie sous forme de fichiers  
 1088 MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000  
 1089 mesures de batterie dans les styles les plus courants et avec différentes  
 1090 qualités de jeu. Tout le contenu a été joué par des humains sur la batterie  
 électronique Roland TD-11 (figure 4.1).

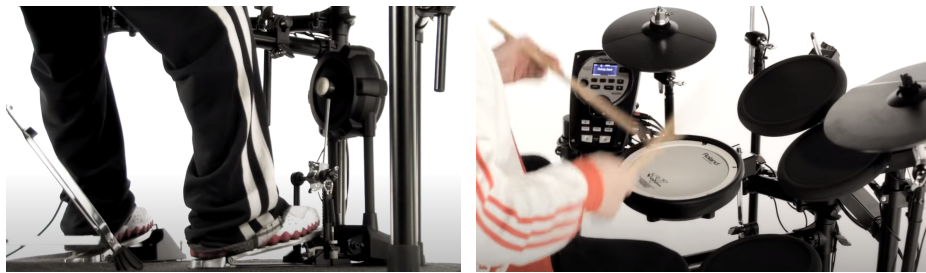


FIGURE 4.1 – Batterie électronique

Source : [https://www.youtube.com/watch?v=BX1V\\_IE0g2c](https://www.youtube.com/watch?v=BX1V_IE0g2c)

1091

1092 Autres critères spécifiques au GMD :

- 1093 — Toutes les performances ont été jouées au métronome et à un tempo  
 1094 choisi par le batteur.
- 1095 — 80% de la durée du GMD a été joué par des batteurs professionnels  
 1096 qui ont pu improviser dans un large éventail de styles. Les don-  
 1097 nées sont donc diversifiées en termes de styles et de qualités de jeu  
 1098 (professionnel ou amateur).
- 1099 — Les batteurs avaient pour instruction de jouer des séquences de  
 1100 plusieurs minutes ainsi que des fills<sup>2</sup>
- 1101 — Chaque performance est annotée d'un style (fourni par le batteur),  
 1102 d'une signature rythmique et d'un tempo ainsi que d'une identifi-  
 1103 cation anonyme du batteur.
- 1104 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10  
 1105 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-  
 1106 session du GMD.
- 1107 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur  
 1108 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.



### 1109 **Format des données**

1110 Le Roland TD-11 enregistre les données dans des fichiers MIDI et les  
1111 divise en plusieurs pistes distinctes :

- 1112 — une pour le tempo et l'indication de mesure ;
- 1113 — une pour les changements de contrôle (position de la pédale de  
1114 charley) ;
- 1115 — une pour les notes.

1116  
1117 Les changements de contrôle sont placés sur le canal 0 et les notes sur le  
1118 canal 9 (qui est le canal canonique pour la batterie).

1119 Pour simplifier le traitement de ces données, ces trois pistes ont été fu-  
1120 sionnées en une seule piste qui a été mise sur le canal 9.

## 1121 **4.2 Analyses et transcriptions manuelles**

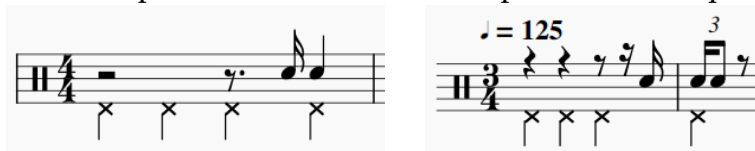
1122 Ces analyses ont été faites dans le cadre de transcriptions manuelles à  
1123 partir de fichiers MIDI et Audio du GMD.

### 1124 **Comparaisons de transcriptions**

1125 Pour les comparaisons de transcriptions, les transcriptions manuelles  
1126 (TM) ont été éditées à l'aide de Lilypond<sup>3</sup> ou MuseScore<sup>4</sup> et les transcrip-  
1127 tions automatiques (TA) ont toutes été générées par import d'un fichier  
1128 MIDI dans MuseScore.

### 1129 **Exemple d'analyse 1**

Transcription manuelle ⇒ Transcription automatique



- 1130 — Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
- 1131 — Les silences de la mesure 1 de la TA sont inutilement surchargés ;
- 1132 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux  
1133 premières notes (une double-croche et une croche) d'un triolet sur  
1134 le temps 1 de la mesure 2 de la TA.

Transcription manuelle  $\Rightarrow$  Transcription automatique



### 1135 Exemple d'analyse 2

- 1136 — Les doubles croches ont été interprétées en quintolet
- 1137 — La deuxième double-croche est devenue une croche.

### 1138 Exemple d'analyse 3

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 1139 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1140
- 1141 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1142 décalés d'une double croche vers la droite dans la TA.
- 1143 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1144 alors qu'elle appartenait à un triolet dans la TM.
- 1145 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1146 pas la TM.
- 1147

### 1148 Exemple d'analyse 4

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 1149
- 1150 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- 1151 manière excessivement complexe!

---

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

1152 **Exemple d'analyse 5 (flas)**

1153 Transcription manuelle



1154

1155 Transcription automatique

1156



1157

1158

- 1159 — Le premier fla est reconnu comme étant un triolet contenant une
- 1160 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1161 ornementée.
- 1162 — Le deuxième fla est reconnu comme étant un accord.
- 1163 — Les deux double en contre-temps sur le temps 4 de la TM sont mal
- 1164 quantifiées dans la TA.
- 1165 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1166 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1167 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1168 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1169 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1170 la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de
- 1171 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1172 temps 3 de la TA.
- 1173 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1174 exemple, le charley ouvert en contre-temps sur le temps 2 de la
- 1175 mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem
- 1176 pour les cross-sticks.

1177 **Conclusion d'analyse**

1178 Ces analyses ont montré la difficulté pour un logiciel comme MuseScore  
 1179 d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI  
 1180 ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance  
 1181 de la forme du rythme impliquant sa position dans la mesure. Cette recon-  
 1182 naissance pourrait permettre de rectifier les problèmes de signature ryth-  
 1183 mique ainsi que les problèmes de décalage de temps. La reconnaissance  
 1184 de la forme du rythme permettrait aussi de supprimer les aberrations du  
 1185 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

FIGURE 4.2 – Partition de référence

La figure 4.2 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).

### 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 évènement MIDI et de la manière de les considérer dans un objectif de transcription en partition lisible pour un musicien (Voir la section 3.2).

— difficultés principales : trouver une application permettant de choisir librement la notation de la batterie. Lylipond le permet mais beaucoup de recherches ont été nécessaires pour comprendre l'ensemble des fonctionnalités permettant de faire fonctionner la notation « agostinienne » ainsi que les diverses subtilités de notations (accents, ghost-notes, flas, ...).

lylipond reste néanmoins un choix très agréable, une fois ces difficultés surmontées.

— Écrire la partition de la figure 4.2 m'a pris beaucoup de temps car j'ai dû chercher comment écrire chaque nouvel évènement mais les autres transcriptions ont été beaucoup plus rapide et très aisées.

— Même si cela représente un investissement au départ, je recommande lylipond pour écrire la batterie et je pense que c'est meilleur outil pour cette tâche pour le moment. On peut configurer absolument tout.

— dans les autres logiciel d'édition de type musescore, la batterie est toujours confiné au système de notation américain.

— pour une comparaison entre système américain et système agostinien, voir section 4.2 est comparer les notations TM (agostinien) et TA (américain).

## 4.3 Transcription polyphonique par parsing

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

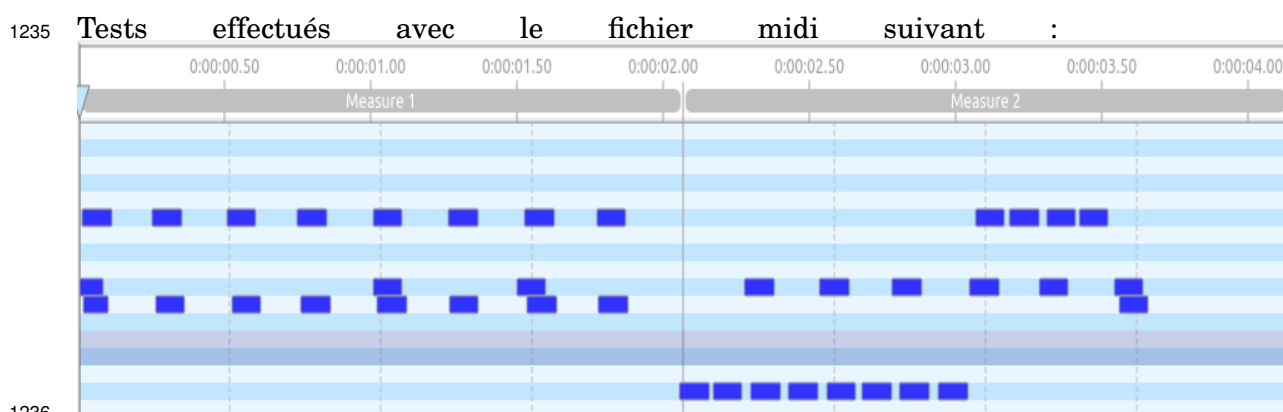
### Les Jams

Les Jams permettent de passer du monophonique au polyphonique.

### Le parsing

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

revenir (rapidement) sur la méthodologie suivie.



1237 Un premier test convaincant est effectué avec la grammaire suivante :

1238

1239 <dam>incompréhensible ainsi, c'est dommage</dam> // bar level

1240 0 -> C0 1

1241 0 -> E1 1

1242 0 -> U4(1, 1, 1, 1) 1

1243

1244 // half bar level

1245 9 -> C0 1

1246 9 -> E1 1

1247

1248 // beat level

1249 1 -> C0 1

1250 1 -> E1 1

1251 1 -> T2(2, 2) 1

1252 1 -> T4(4, 4, 4, 4) 1

1253

1254 // croche level

1255 2 -> C0 1

1256 2 -> E1 1

1257

1258 // double level

1259 4 -> C0 1

1260 4 -> E1 1

1261 4 -> E2 1

1262 4 -> T2(6, 6) 1

1263

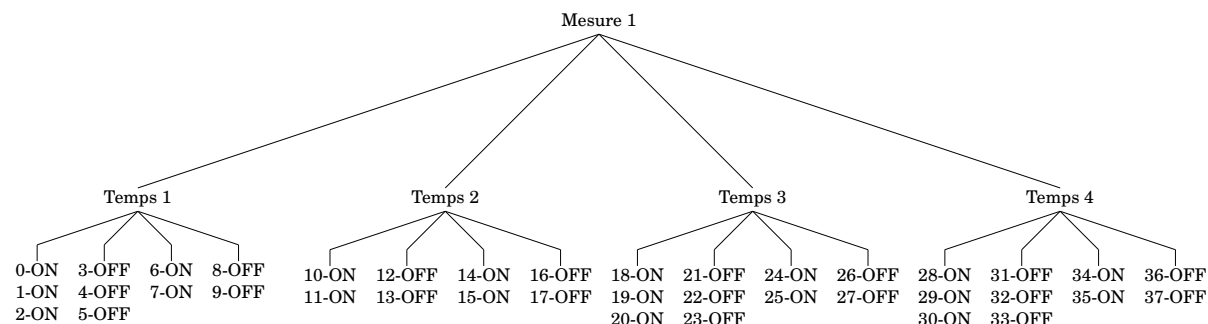
1264 // triple level

1265 6 -> E1 1

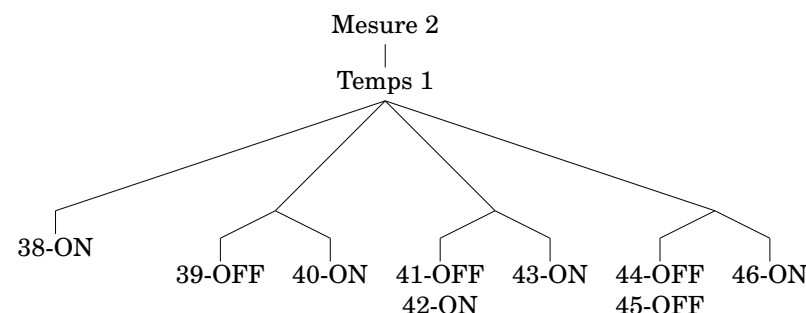
1266

1267 Cette grammaire sépare les ligatures par temps au niveau de la  
1268 mesure. Puis, au niveau du temps, elle autorise les divisions par deux

(croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1. L'arbre de parsing en résultant est considéré comme « convaincant » car il découpe correctement les mesures et les temps.



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



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

### 2ème exemple :

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

## 1293 4.4 Réécriture guidée par une forme rythmique

1294 La démonstration qui suit est basée sur la partition de référence de la  
 1295 figure 4.2 puisque la forme rythmique qui sera utilisée en est directement  
 1296 extraite.

1297

1298 Nous allons montrer :

- 1299 — la composition de cette forme rythmique ;
- 1300 — son état finale, c'est à dire toutes les combinaisons entièrement
- 1301 écrites en notation correcte sur partition ;
- 1302 ⇒ cela constituera une référence pour la réécriture ;
- 1303 — un exemple de transformation de la forme rythmique en arbre de
- 1304 rythme ;
- 1305 — l'application de la séparation des voix sur cet exemple basé sur la
- 1306 référence citée précédemment (la forme rythmique en question) ;
- 1307 ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a
- 1308 de voix (deux arbres pour cette forme rythmique) ;
- 1309 — les règles de simplification propres à la forme rythmique dont nous
- 1310 parlons.

1311 L'objectif de cette démonstration est de montrer comment un jeu de plu-  
 1312 sieurs formes rythmiques pourrait être implémenter dans le cadre d'une  
 1313 approche dictionnaire.

## 1314 Motifs et gammes

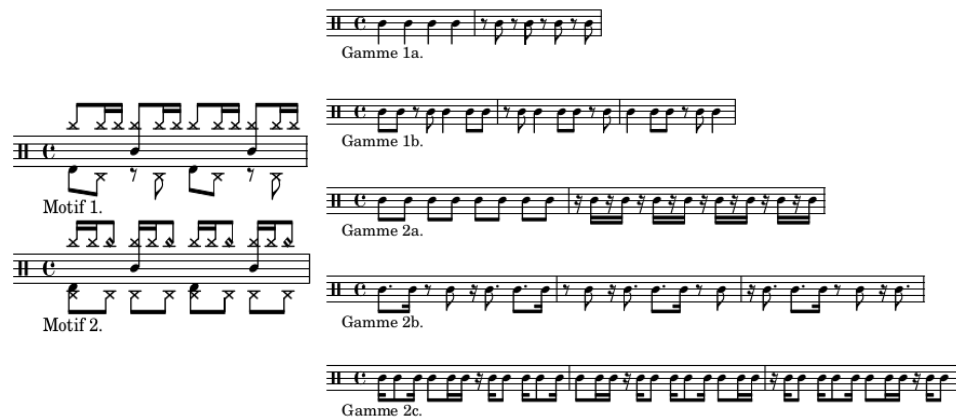


FIGURE 4.3 – Motifs et gammes

## 1315 Motifs

1316 À partir de la partition de référence, les deux motifs de la figure 4.3  
 1317 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure



1318 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à  
 1319 la mesure 28 avec des variations. Ces deux motifs sont très classiques et  
 1320 pourront être détectés dans de nombreuses performances.  
 1321

### 1322 **Gammes**

1323 Les gammes de la figure 4.3 étayent toutes les combinaisons d'un motif  
 1324 en 4/4 binaires jusqu'aux doubles croches.  
 1325 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-  
 1326 mière ne contient que des noires et la deuxième que des croches en contre-  
 1327 temps. Ces deux possibilités sont combinées de manière circulaire dans  
 1328 les 3 mesures de la deuxième ligne.  
 1329 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures  
 1330 dont la première ne contient que des croches et la deuxième que des  
 1331 doubles-croches en contre-temps. Ces deux possibilités sont combinées de  
 1332 manière circulaire dans les lignes 4 et 5 qui contiennent chacune 3 me-  
 1333 sures.

### 1334 **Formes rythmiques — motifs et gammes combinés**

1335 Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure  
 1336 4.3.<dam>à commenter un peu plus, notamment pour dire si la combinai-  
 son est faite automatiquement ou non</dam>

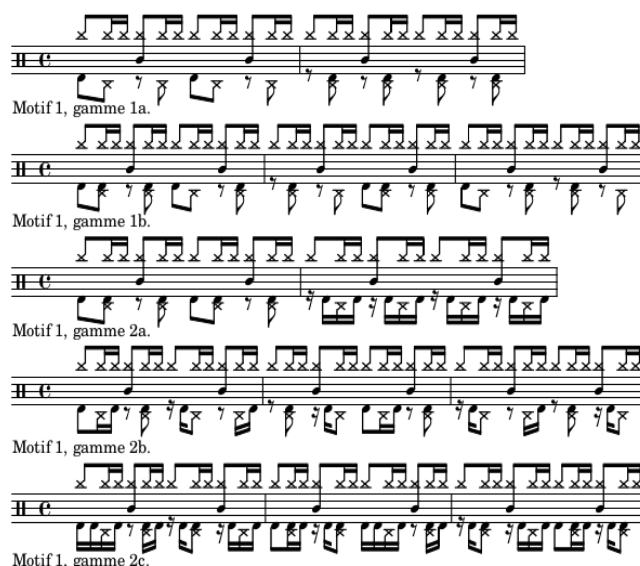


FIGURE 4.4 – Partition d'un forme rythmique en 4/4 binaire

## 1338 Représentation de la forme rythmique en arbres de 1339 rythmes

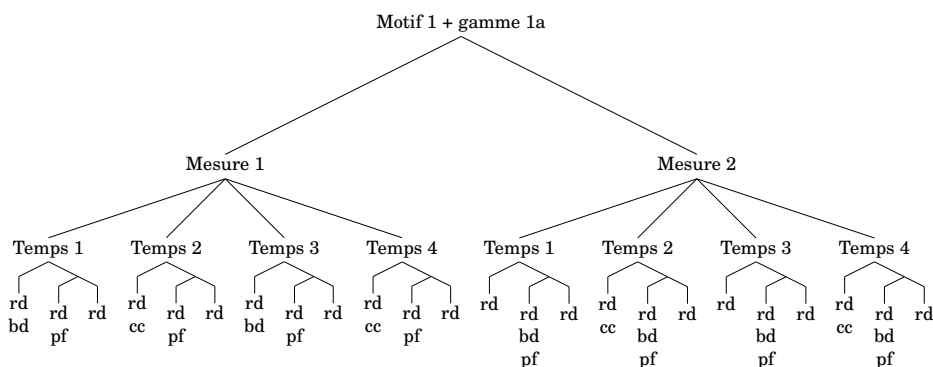


FIGURE 4.5 – Arbre de rythme — forme rythmique

1340 L'arbre de la figure 4.5 servira de base pour le suite de l'expérimentation.  
1341 Comme indiqué à la racine de l'arbre, il représente la première ligne de la  
1342 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,  
1343 il manque des indications de notation telles que les voix spécifiques à  
1344 chaque partie du rythme ainsi que les choix d'écriture pour les distances  
1345 qui séparent les notes de chaque voix entre elles en termes de durée.

## 1346 Réécriture — séparation des voix et simplification

### 1347 La séparation des voix

1348 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui  
1349 le constituent et les voix seront regroupées en suivant les règles du forme  
rythmique.

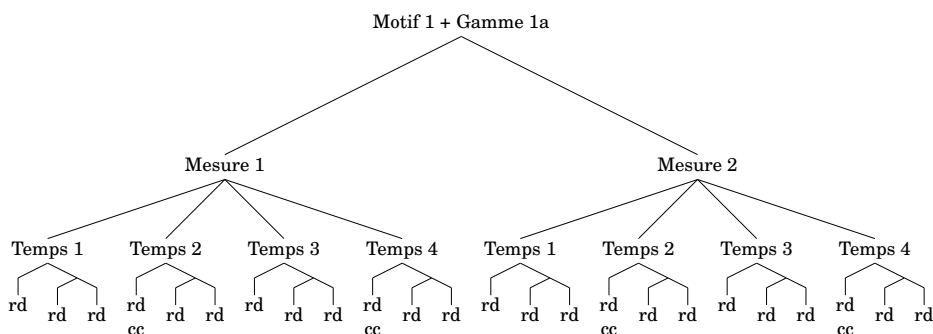


FIGURE 4.6 – Arbre de rythme — voix haute

1350  
1351 La voix haute (figure 4.6) regroupe la ride et la caisse-claire sur les liga-  
1352 tures du haut.

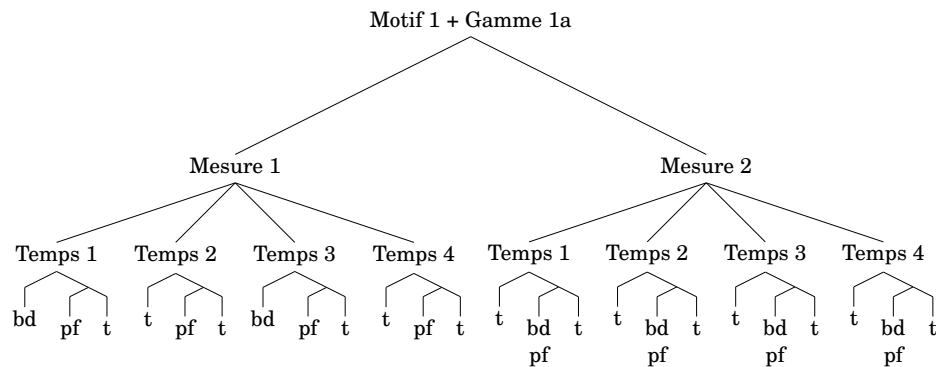


FIGURE 4.7 – Arbre de rythme — voix basse

1353 La voix basse (figure 4.7 regroupe la grosse-caisse et le charley au pied  
1354 sur les ligatures du bas.

1355 **Les règles de simplifications**

1356 L’objectif des règles de simplifications est de réécrire les écarts de durées  
1357 qui séparent les notes d’une manière appropriée pour la batterie et qui  
1358 soit la plus simple possible. Les ligatures relient les notes d’un temps  
1359 entre elles afin de rendre la pulsation visuelle).

1360

1361 Pour les figures ci-dessous :

- 1362 — x = une note ;  
1363 — r = un silence ;  
1364 — t = une continuation (point ou liaison)

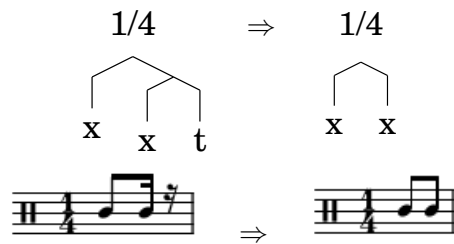


FIGURE 4.8

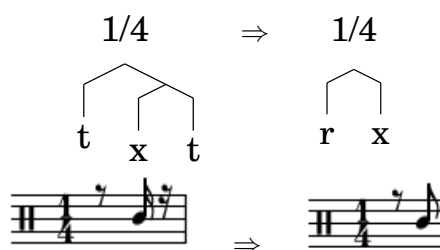


FIGURE 4.9

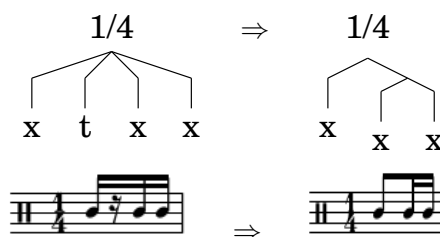


FIGURE 4.10

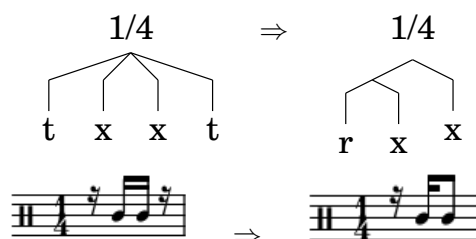


FIGURE 4.11

- 1365 Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique.  
 1366 Les arbres manquants seront mis en annexe.  
 1367 Les règles remplacent par un silence les continuations (t) qui sont au dé-  
 1368 but d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il  
 1369 y a des ouvertures de charley, cela n'est pas toujours applicable.

### 1370 Conclusion sur cette réécriture guidée

- 1371 La méthode des formes rythmiques étant basée sur une approche diction-  
 1372 naire, Le premier objectif de cette réécriture guidée est d'orienter la re-  
 1373 cherche d'autres formes rythmiques par observation du jeu de données et  
 1374 de montrer comment les construire pour agrandir la base de connaissance  
 1375 de Qparse pour la transcription de la batterie.

## 1376 4.5 BILAN : résultats — évaluation — discussion

1377 Cette section regroupe les avancées qui ont été réalisées par rapport aux  
1378 objectifs de départ ainsi qu’une réflexion sur le moyen d’évaluer les résul-  
1379 tats de l’ADT avec Qparse. Nous avons amélioré le système de quantifi-  
1380 cation de Qparse pour la batterie, notamment le passage à la polyphonie  
1381 avec les Jams.

1382 Nous avons pu obtenir des arbres de parsing corrects en améliorant les  
1383 grammaires avec des fichiers MIDI courts.

1384 Puis, une sortie MEI a aussi été obtenue (encore à vérifier).

1385 Dans cette section, nous discuterons sur la pertinence de l’ensemble des  
1386 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui  
1387 ont été faites ou non et nous tenterons d’en expliquer la ou les raisons.

1388 — Le choix de travailler avec Lilypond et non Verovio. Ce choix était  
1389 motivé par la liberté totale concernant la notation de la batterie  
1390 dont un et la disponibilité d’un set de notation de type Agostini.  
1391 C’est la seule application qui me permettait d’écrire la notation de  
1392 la batterie exactement comme je le souhaitais.

1393 — Avancé de la chaîne de traitement (nous sommes arrivés aux arbres  
1394 de parsing, nous avons traité le polyphonique (identification des  
1395 regroupements de notes<sup>5</sup>) ⇒ Quelques arbres ont été obtenus sur  
1396 des exemples simples<sup>6</sup>)

1397 — 2 dimensions de le travail fourni :

1398 - La volonté de pousser un exemple simple jusqu’au bout de la  
1399 chaîne pour obtenir des résultats et une évaluation sur au moins  
1400 un exemple ; - La réalité du travail à fournir pour faire avancer sur  
1401 la chaîne de traitement. ⇒ Une solution aurait été de considérer  
1402 les arbres de parsing obtenus après le traitement du polyphonique  
1403 comme un résultat local possible à évaluer au lieu d’attendre que la  
1404 chaîne arrive jusqu’à la génération d’une partition mais cela n’était  
1405 pas prioritaire pendant le stage.

1406 — Création d’un jeu de forme rythmique basique représentatif des  
1407 différents styles à recouvrir. Ce jeu n’a pas pu être créé, car comme  
1408 vu plus haut, je me suis focalisé sur un exemple pour pouvoir le  
1409 vérifier entièrement et dans l’espoir de pouvoir le tester en fin de  
1410 chaîne. **Évaluation** Matcher les motifs aurait été indispensable  
1411 pour obtenir une quantité de résultats qui justifieraient une  
1412 évaluation automatique permettant de faire des graphiques.

1413 L’évaluation fut entièrement manuelle car :

1414 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf  
1415 VS output) Pour l’évaluation, il aurait fallu produire un module.

---

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

1416 <dam>je ne sais pas si tu auras encore le temps de faire ça, sinon  
1417 il faudra décrire comment tu aurais aimé évaluer, proprement et  
1418 sans résultats chiffrés</dam> L'évaluation est-elle automatique ou  
1419 manuelle ?  
1420 Possibilité d'un export lilypond en arbre pour comparer l'ouput  
1421 avec la transcription manuelle.  
1422 Possibilité de transformer lilypond(output) et lilypond(ref) en  
1423 ScoreModel ou MEI pour les comparer et faire des statistiques.  
1424 Si transformés en MEI : diffscore de Francesco. Possibilité de  
1425 transformer lilypond(output) et lilypond(ref) en MusicXML pour  
1426 les comparer ou dans Music21. L'expérimentation peut-être consi-  
1427 dérer comme une évaluation manuelle ? (magicien d'Oz)  
1428 Lilypond vers MIDI + ouput vers MIDI  $\Rightarrow$  Comparaison des MIDI  
1429 dumpés.  
1430  
1431 La transcription automatique de la batterie est un sujet passionnant mais  
1432 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-  
1433 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-  
1434 moins été générée. Elle sera un bon point de départ pour les travaux fu-  
1435 turs dont plusieurs propositions sont énoncés dans le présent document.

## CONCLUSION GÉNÉRALE

1437 Dans ce mémoire, nous avons traité de la problématique de la transcrip-  
1438 tion automatique de la batterie. Son objectif était de transcrire, à partir  
1439 de leur représentation symbolique MIDI, des performances de batteur de  
1440 différents niveaux et dans différents styles en partitions écrites.

1441 Nous avons avancé sur le parsing des données MIDI établissant un pro-  
1442 cessus de regroupement des événements MIDI qui nous a permis de faire  
1443 la transition du monophonique vers le polyphonique. Une des données  
1444 importante de ce processus était de différencier les nature des notes d'un  
1445 accord, notamment de distinguer lorsque 2 notes constituent un accord  
1446 ou un *fla*.

1447 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-  
1448 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-  
1449 tionnables en amont du parsing, soit par indication des noms des fichiers  
1450 MIDI, soit par reconnaissance de la métrique avec une approche diction-  
1451 naire de patterns prédéfinis <sup>7</sup> qu'il serait pertinent de mettre en œuvre en  
1452 machine learning.

1453 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre  
1454 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif  
1455 d'un système et pour la simplification grâce aux gammes du motif d'un  
1456 système. Nous avons aussi montré comment, dans des travaux futurs, un  
1457 système dont le motif serait reconnu en amont dans un fichier MIDI pour-  
1458 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-  
1459 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans  
1460 la chaîne de traitement en terme de réécriture.

1461 Il sera également intéressant d'étudier comment l'utilisation de LM peut  
1462 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération  
1463 entièrement automatisée de partitions de batterie et au problème général  
1464 de l'AMT de bout en bout.[11]

---

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





## BIBLIOGRAPHIE

- 1466 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41  
1467 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,  
1468 1996. – Cité pages 7, 18 et 36.
- 1469 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11  
1470 et 15.
- 1471 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité  
1472 page 12.
- 1473 [4] Gaël Richard et al. De fourier à la reconnaissance  
1474 musicale. Available at [https://interstices.info/  
1475 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).  
1476 – Cité page 12.
- 1477 [5] Caroline Traube. Quelle place pour la science au sein de la musico-  
1478 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1479 [6] Leonard Bernstein Office. The unanswered question : Six talks at  
1480 harvard. Available at [https://leonardbernstein.com/about/  
1481 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1482 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions  
1483 entre les traitements de la musique et du langage. *La Lettre des  
1484 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1485 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.  
1486 Techniques de traitement automatique du langage naturel appli-  
1487 quées aux représentations symboliques musicales. In *JIM 2021 -  
1488 Journées d’Informatique Musicale*, Virtual, France, July 2021. –  
1489 Cité page 13.
- 1490 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-  
1491 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1492 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering  
1493 music relations with sequential attention. In *NLP4MUSA*, 2020. –  
1494 Cité page 13.
- 1495 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger  
1496 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1497 lenges and future directions. *Journal of Intelligent Information Sys-*  
1498 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 63.
- 1499 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J  
1500 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –  
1501 Cité pages 14 et 27.
- 1502 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-  
1503 SURE, 1997. – Cité page 15.
- 1504 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-  
1505 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander  
1506 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*  
1507 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–  
1508 1483, 2018. – Cité pages 15, 23 et 27.
- 1509 [15] Moshekwa Malatji. Automatic music transcription for two instru-  
1510 ments based variable q-transform and deep learning methods, 10  
1511 2020. – Cité page 22.
- 1512 [16] Antti J. Eronen. Musical instrument recognition using ica-based  
1513 transform of features and discriminatively trained hmms. *Seventh*  
1514 *International Symposium on Signal Processing and Its Applications*,  
1515 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1516 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic  
1517 drum sound description for real-world music using template adap-  
1518 tation and matching methods. *International Conference on Music*  
1519 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1520 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local  
1521 musical statistics as guides for audio-to-score piano transcription.  
1522 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24 et 26.
- 1523 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-  
1524 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-  
1525 tization and Score Structuring. In *MCM 2019 - Mathematics and*  
1526 *Computation in Music*, volume Lecture Notes in Computer Science  
1527 of *Proceedings of the Seventh International Conference on Mathema-*  
1528 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June  
1529 2019. Springer. – Cité pages 24 et 26.
- 1530 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-  
1531 ring of rhythm structures. In *Proceedings of the First International*  
1532 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.  
1533 IEEE Computer Society. – Cité page 26.
- 1534 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A  
1535 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1536 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –  
1537 Cité page 26.
- 1538 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating  
1539 equivalent rhythmic notations based on rhythm tree languages. In  
1540 *Third International Conference on Technologies for Music Notation  
1541 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-  
1542 pez Palma and Mike Solomon. – Cité page 26.
- 1543 [23] R. Marxer and J. Janer. Study of regularizations and constraints in  
1544 nmf-based drums monaural separation. In *International Conference  
1545 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,  
1546 02/09/2013 2013. – Cité page 27.
- 1547 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions  
1548 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 29.
- 1549 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue  
1550 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 29.
- 1551 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-  
1552 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -  
1553 94120 Fontenay-sous-Bois, 2000. – Cité pages 31 et 43.
- 1554 [27] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David  
1555 Bamman. Learning to groove with inverse sequence transforma-  
1556 tions. In *International Conference on Machine Learning (ICML)*,  
1557 2019. – Cité page 48.

