

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

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

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

8 *Parcours :*
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

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

14 *Encadrant :*
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021

TABLE DES MATIÈRES

18	Liste des figures	4
19	Liste des tableaux	5
20	Introduction générale	7
21	1 Contexte	11
22	1.1 Langues naturelles et musique en informatique	12
23	1.2 La transcription automatique de la musique	14
24	1.3 La transcription automatique de la batterie	15
25	1.4 Les représentations de la musique	16
26	2 État de l'art	21
27	2.1 Monophonique et polyphonique	21
28	2.2 Audio vers MIDI	22
29	2.3 MIDI vers partition	24
30	2.4 Approche linéaire et approche hiérarchique	24
31	3 Méthodes	29
32	3.1 La notation de la batterie	29
33	3.2 Modélisation pour la transcription	37
34	3.3 Qparse	38
35	3.4 Les forme rythmiques	40
36	4 Expérimentations	47
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	Conclusion générale	63
43	Bibliographie	65

LISTE DES FIGURES

45	1.1	Exemple évènements avec durée	16
46	1.2	Critère pour un évènement	17
47	1.3	Exemple évènements sans durée	17
48	1.4	Les silences	18
49	1.5	18
50	1.6	Rapport des figures de notes	18
51	1.7	Exemple de partition de piano	19
52	1.8	MusicXML	20
53	2.1	Transcription automatique <dam>remettre ici la citation de la	
54		capture d'écran avec la page</dam>	23
55	2.2	HMM	26
56	2.3	arbre_jazz	27
57	3.1	Les instruments de la batterie	30
58	3.2	Hauteur et têtes de notes	30
59	3.3	Point et liaison	32
60	3.4	Silence joué	33
61	3.5	Équivalence	34
62	3.6	Séparation des voix	35
63	3.7	Les accents et les ghost-notes	35
64	3.8	Exemple pour les accentuations et les ghost-notes	36
65	3.9	Définition du fla	36
66	3.10	Présentation de Qparse	39
67	3.11	signature rythmique	42
68	3.12	Motif 4-4 binaire	43
69	3.13	Motif 4-4 jazz	43
70	3.14	forme rythmique 4-4 afro-latin	43
71	3.15	Simplification	44
72	3.16	45
73	4.1	Batterie électronique	48
74	4.2	Partition de référence	52
75	4.3	Motifs et gammes	56
76	4.4	Partition d'un forme rythmique en 4/4 binaire	57
77	4.5	Arbre de rythme — forme rythmique	58
78	4.6	Arbre de rythme — voix haute	58
79	4.7	Arbre de rythme — voix basse	59
80	4.8	59

81	4.9	60
82	4.10	60
83	4.11	60

84

LISTE DES TABLEAUX

85	1.1	speechToText vs AMT	13
86	3.1	Noms des instruments de la batterie	30
87	3.2	Pitchs et instruments	37
88	3.3	Sytèmes	41

INTRODUCTION GÉNÉRALE

90 QUOI?

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

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

112

113 POURQUOI?

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

120

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

123 COMMENT?

124 → Problématique :

125 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.

135

136 → Méthodes :

137 → Contributions :

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

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

148

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

- 150 — une modélisation de la notation de la batterie (fusion de 3.1 et de
151 3.2) qui était jusqu'à présent inexistante.
- 152 — plusieurs transcriptions manuelles dans le but d'analyser les conte-
153 nus des fichiers MIDI et Audio (4.2) et de faire des comparaisons
154 de transcription avec des outils déjà existants¹.
- 155 — une partition de référence transcrite manuellement sur l'entièreté
156 d'une performance du jeu de données afin de repérer les éléments
157 importants pour la modélisation et de faire les liens entre les cri-
158 tères des données d'*input* avec l'écriture finale (4.2). Cette partition
159 avait aussi pour objectif d'effectuer des tests et des évaluations.
- 160 — le passage au polyphonique en théorie et en implémentation im-
161 pliquant la théorie sur la détection de l'identité de notes dans un
162 Jam² et l'implémentation de tests unitaires sur le traitement des
163 Jams (4.3).
- 164 — la création de grammaires pondérées spécifiques à la batterie (4.3)

165

1. MuseScore3

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

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

179

180 PLAN

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

CONTEXTE

Sommaire

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

Introduction

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

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

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

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

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

217 1.1 Langues naturelles et musique en 218 informatique

219 COMPUTER MUSIC

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

228 RIM

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

238 RIM et TAL

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

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

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

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

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

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

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

exemples / illustration de la proximité thématique?

D'autres travaux très récents, ont aussi été révélés lors de la *première conférence sur le NLP pour la musique et l'audio (NLP4MusA 2020)*. Lors de cette conférence, Jiang *et al.* [10] ont présenté leur implémentation d'un modèle de langage musical visant à améliorer le mécanisme d'attention par élément, déjà très largement utilisé dans les modèles de séquence modernes pour le texte et la musique.

Le domaine du TAL qui se rapproche le plus du RIM est la reconnaissance de la parole (Speech to text). En effet, la séparation des sources ont des approches similaires dans les deux domaines. De plus, il existe un lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole. La transcription musicale étant la notation d'une œuvre musicale initialement non écrite, l'analogie avec l'écriture de la parole est aisée. Le tableau 1.1 montre des différences et des similitudes entre les deux domaines.

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

TABLE 1.1 – speechToText vs AMT

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

287 1.2 La transcription automatique de la musique

288 1. OBJECTIF

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

300

301 2. APPLICATIONS

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

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

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

319

320 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

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

8. Ici, on parle de notation occidentale.

9. ethno-musicologie

10. par exemple par la constitution de corpus musicologiques

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

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

1.3 La transcription automatique de la batterie

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

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

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

Les applications de la transcription automatique de la batterie (TAB) seraient utiles, non seulement dans tous les domaines musicaux concernés par la batterie dont certains manquent de partitions, notamment les musiques d'improvisation [11], mais aussi de manière plus générale dans le domaine de la RIM : si les ordinateurs étaient capables d'analyser la partie de la batterie dans la musique enregistrée, cela permettrait de faciliter de nombreuses tâches de traitement de la musique liées au rythme. En particulier, la détection et la classification des événements sonores de la batterie par des méthodes informatiques est considérée comme un problème de recherche important et stimulant dans le domaine plus large de la recherche d'informations musicales [14].

cite méthode et école Agostini?

12. Roland TR-808

13. SQ-1

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

368 1.4 Les représentations de la musique

370 Les données audio

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.

376 Les données MIDI

377 Le MIDI¹⁴ (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.
 382 Les données midi sont représentées sous forme de piano-roll.
 383 Chaque point sur la figure 1.1 est appelé « événement MIDI » :

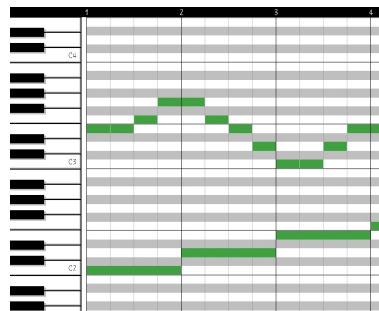


FIGURE 1.1 – Exemple événements avec durée

385 Chaque événement MIDI rassemble un ensemble d'informations sur la
 386 hauteur, la durée, le volume, etc... :
 387 Pour la batterie, les événements sont considérés sans durée, nous ignore-
 388 rons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le
 389 *channel* ne nous sera pas utile non plus.
 390 Ici, définir Tick et channel.
 391 Voici un exemple de piano-roll midi pour la batterie :

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

citer M. Müller FMP pour cette section ?

trop technique. ne pas repier wikipédia

LPCM pas utile ici. parle juste échantillons et compression.

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

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

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

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

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

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

FIGURE 1.2 – Critère pour un évènement

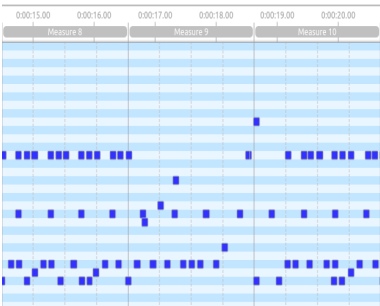


FIGURE 1.3 – Exemple évènements sans durée

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

403 **Les partitions**
404

405 Expliquer aussi la notation générale des silences
406 La figure 1.5 montre 4 figures de notes les plus courantes dont les noms
407 et les durées sont respectivement, de gauche à droite :
408 — La ronde, elle vaut 4 ;
409 — La blanche, elle vaut 2 ;
410 — La noire, elle vaut 1 ;
411 — La croche, elle vaut 1/2.

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

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

4 temps

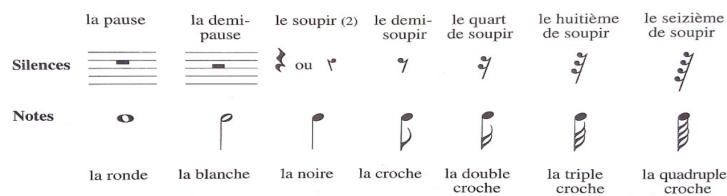


FIGURE 1.4 – Les silences



FIGURE 1.5

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

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

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

plusieurs éléments

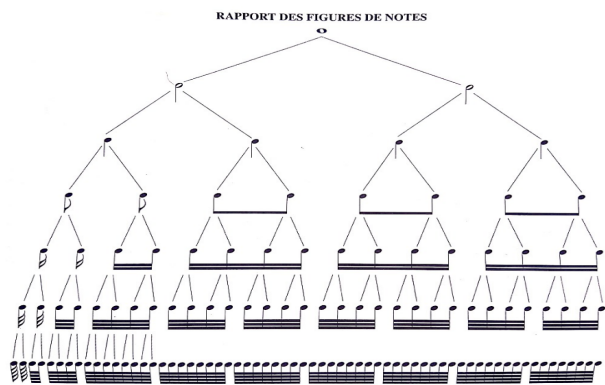


FIGURE 1.6 – Rapport des figures de notes
[1]

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

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

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



FIGURE 1.7 – Exemple de partition de piano

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

- 433 — la hauteur ;
- 434 — la durée ;
- 435 — l'intensité ;
- 436 — le timbre.

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

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

444 Les formats XML

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

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

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

454 appoggiatures

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

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

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

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

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

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

```

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

```

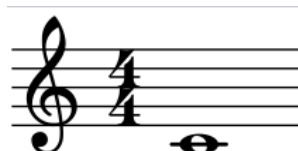


FIGURE 1.8 – MusicXML

signature rythmique

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

Conclusion

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

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

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

ÉTAT DE L'ART

Sommaire

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

Introduction

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

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

2.1 Monophonique et polyphonique

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

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

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

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

2.2 Audio vers MIDI

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

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

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

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

MIDI non-quantifié = performance (à expliquer)

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

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

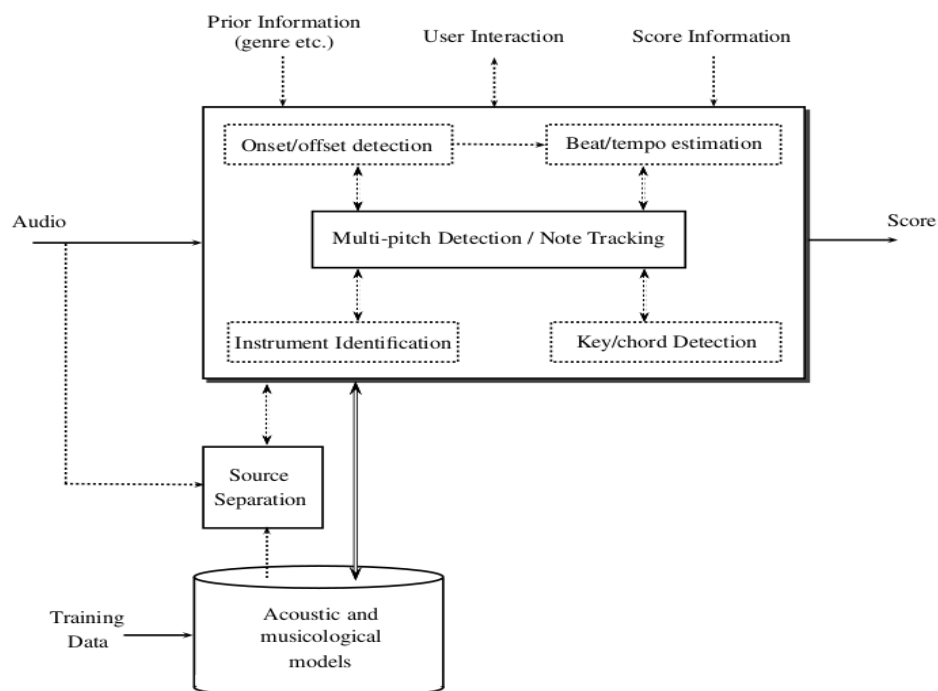


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

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.

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

haute fréquence, aigus?

552 pré-processing.
 553 Pour la reconnaissance des instruments, une approche possible [16] est
 554 de mettre un modèle probabiliste dans l'étape de la classification des évè-
 555 nements afin de classer les différents sons de la batterie. Cette méthode
 556 permet de se passer de samples audio isolés en modélisant la progression
 557 temporelle des *features*³ avec un modèle de markow caché (HMM). Les
 558 *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- 555
 ments? la phrase semble 556
 redondante

pas clair... peut-être just 559
 mentionner les modèles 560
 probabilistes utilisés

563 2.3 MIDI vers partition

564 Le plus souvent, lorsque les articles abordent la transcription automa-
 565 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »
 566 (partition) désigne un ouput au format Music XML, ou simplement MIDI.
 567 Par exemple, dans [18], la chaîne de traitement va jusqu'à la génération
 568 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en
 569 extraire manuellement un fichier MusicXML contenant plusieurs voix.
 570 Seuls quelques travaux récents s'intéressent de près à la création d'outils
 571 permettant la génération de partition. Le problème de la conversion d'une
 572 séquence d'évènements musicaux symboliques en une partition musicale
 573 structurée est traité notamment dans [19]. Ce travail, qui vise à résoudre
 574 en une fois la quantification rythmique et la production de partition struc-
 575 turée, s'appuie tout au long du processus sur des grammaires génératives
 576 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-
 577 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été
 578 menées avec un ensemble de données composé d'extraits monophoniques ;
 579 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 564
 cela. cf. proposition de des- 565
 cription + détaillée en com- 566
 mentaires

de manière conjointe 575

langage a priori 577

qui nécessite de traiter le 580
 problème supplémentaire 581
 de la séparation de voix. i.e. 582
 pour la batterie on nveut 583
 quantification + structu- 584
 ration + séparation mais 585
 seules les 2 premières sont 586
 couplées dans l'approche de 587
 tonn stage.

584 2.4 Approche linéaire et approche hiérarchique

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

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

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

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

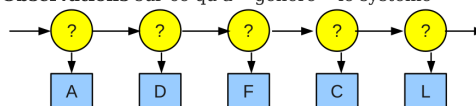


FIGURE 2.2 – HMM

594 *Source : Cours de Damien Nouvel*⁵

595

596

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

je ne comprend pas bien 604
 l'explication. le pb est plu-
 tot vue locale (déduction 605
 la proba d'une durée à par-
 tir de la durée précédente,
 par ex. dans un HMM) vs 606
 vue globale, dans une hié-
 rarchie 607
 608

RT? 609
 610

techniques de réécriture 611
 appliquée à la déduction
 automatique, calcul symbo- 612
 lique 613

le calcul d'équiv. 614
 615

citer thèse de David Rizo 618
 (Valencia) 619

620

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

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

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

Example: *Summertime*

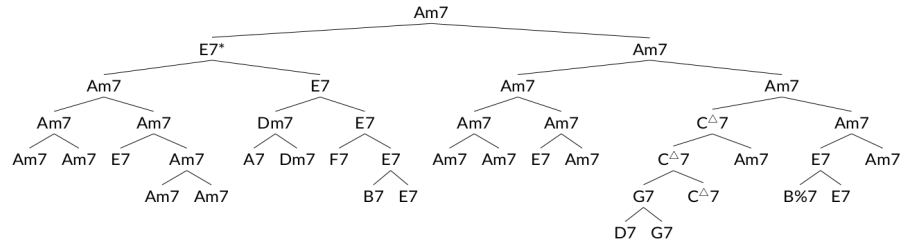


FIGURE 2.3 – arbre_jazz
Représentation arborescente d’une grille harmonique [12]

Conclusion

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

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

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

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

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

latex : enumerate

MÉTHODES

Sommaire

3.1	La notation de la batterie	29
3.2	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 qparse¹, un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et plusieurs développeurs au sein du laboratoire Cedric au CNAM.

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

3.1 La notation de la batterie

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

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

669 Les hauteurs et les têtes de notes



FIGURE 3.1 – Les instruments de la batterie

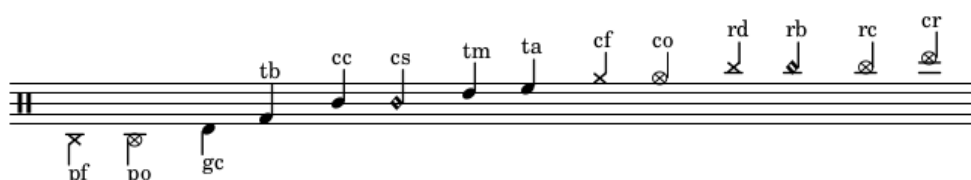


FIGURE 3.2 – Hauteur et têtes de notes

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

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

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

- 674 1. les noms des instruments sur la figure 3.1 ;
675 2. leurs codes respectifs dans la figure 3.2 ;
676 3. les noms que j'utiliserai dans le présent document pour y référer.
- 677 Les figures 3.1, 3.2 et le tableau 3.1 peuvent aider à comprendre pourquoi
678 je trouve la notation agostinienne cohérente et intuitive.
679 En effet, les hauteurs sur la portée représentent :
- 680 1. La hauteur physique des instruments :
681 La caisse claire est centrale sur la portée et sur la batterie (au
682 niveau de la ceinture, elle conditionne l'écart entre les pédales et
683 aussi la position de tous les instruments basiques d'une batterie).
684 Tout ce qui en-dessous de la caisse claire sur la portée est en dessous
685 de la caisse claire sur la batterie (pédales, tom basse) ;
686 Tout ce qui est au-dessus de la caisse claire sur la portée, l'est aussi
687 sur la batterie.
688
- 689 2. La hauteur des instruments en terme de fréquences :
690 Sauf pour le charley au pied et si l'on sépare en trois groupes (grosse
691 caisse, toms et cymbales), de bas en haut, les instruments vont du
692 plus grave au plus aigu.

693 Les durées

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

698 Par conséquent :

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

711

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

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

712 En résumé :

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

720

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

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



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

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

4. La position des temps

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

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

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

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

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

Les silences

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

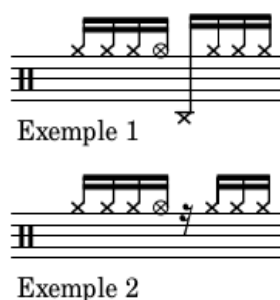


FIGURE 3.4 – Silence joué

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

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

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

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

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

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

772 Les équivalences rythmiques

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

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



FIGURE 3.5 – Équivalence

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

790 Les voix

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

794 En batterie, une voix est théoriquement l'ensemble des instruments qui,
 795 à eux seuls, constituent une phrase rythmique. Mais en pratique, les ins-
 796 truments peuvent aussi être divisés par voix dans le but de ne pas sur-
 797 charger la notation ou pour que leur disposition soit représentée sur la
 798 partition (voir section 3.1). Les voix sont caractérisées par l'orientation
 799 des hampes et plus précisément par les ligatures si les hampes sont dans
 800 la même direction (voir figure 3.14).



FIGURE 3.6 – Séparation des voix

801 Sur la figure 3.6, il faudra faire un choix entre les exemples 1, 2 et 3 qui
 802 sont trois façons équivalentes d'écrire le même rythme. Ce choix se fera en
 803 fonction des instruments joués, de la nature plus ou moins systématique
 804 de leurs phrasés, et des associations logiques entre les instruments dans
 805 la distribution des rythmes sur la batterie (voir la section 3.4).

806 Les accentuations et les ghost-notes

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

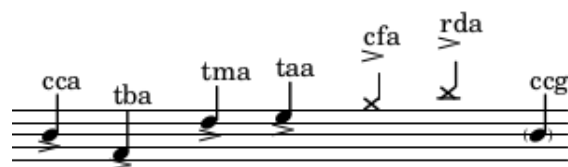


FIGURE 3.7 – Les accents et les ghost-notes

Théoriquement, tous les instruments peuvent être accentués (voir la section 3.2), mais la figure 3.7 représentent ceux dont les accents ne demandent pas un grand niveau de maîtrise et sont presque toujours bien articulés. En outre, les instruments qui ne sont pas représentés sur cette figure ne sont presque jamais accentués dans les partitions et ne sont pas présents de manière significative dans le GMD.

Les accents sont marqués par le symbole « > ». Ils sont positionnés au-dessus des notes représentant des cymbales et en-dessous des notes représentant des toms ou la caisse claire. Ce choix a été fait pour la partition de la figure 4.2 car elle est plus lisible ainsi, mais ces choix devront être adaptés en fonction des différentes formes rythmiques reconnues (voir la section 3.4). Par exemple, pour les formes rythmiques jazz, les ligatures pour les toms et la caisse claire seront dirigées vers le bas, il faudra donc mettre les symboles d'accentuation correspondants au-dessus des têtes de notes.

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

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



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

833

834 Les flas

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



FIGURE 3.9 – Définition du fla

837

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

Pour les codes, on prend le code de la note et on ajoute un « a » pour un accent et un « g » pour une ghost-note.

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. . .)

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

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 ;

citation lilypond

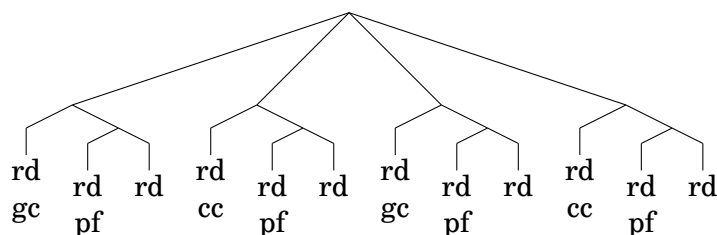
et l’analyse d’autres fichiers MIDI ?

- 860 — Pas d'intention d'accent ni de ghost-note pour une vitesse entre 40
- 861 et 89 ;
- 862 — Les accents et les ghosts-notes ne sont significatifs ni pour les ins-
- 863 truments joués au pied, ni pour les cymbales crash.
- 864 En effet, certaines vitesses en dessous de 40 étant détectées et ins-
- 865 crites dans les données MIDI sont dues au mouvement du talon du
- 866 batteur qui bat la pulsation sans particulièrement jouer le charley.
- 867 Ce mouvement est perçu par le capteur de la batterie électronique
- 868 mais le charley n'est pas joué.
- 869 — Au final, nous avons relevé les ghost-notes et les accents pour la
- 870 caisse claire ainsi que les accents pour les toms et les cymbales
- 871 rythmiques (charley et ride).

872 Les arbres de rythmes

873 Les arbres de rythmes représentent un rythme unique dont les possibili-
tés de notation sur une partition sont théoriquement multiples.

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



non c'est juste une repré-874
sentation du rythme, pas
unique 875

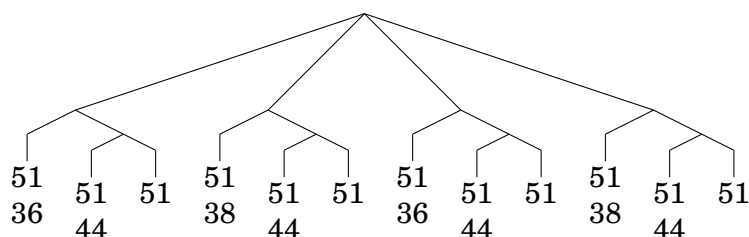
expliquer le principe des
RT : branchement = divi-
sion d'intervalle temporel,
feuilles = les événements
musicaux commençant au
début de l'intervalle). réf-
érences : - Laurson "Patch-
work : A Visual Program-
ming Language", 1996. -
OpenMusic : visual pro-
gramming environment for
music composition,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

876

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

877

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



878

879 Chacun des trois exemples de la figure 3.6 est représenté par un des deux
880 arbres syntaxiques ci-dessus.

881 3.3 Qparse

choisir titre plus explicite882
par ex. analyse syntaxique
pour la transcription musi-
cale

La librairie Qparse⁵ implémente la quantification des rythmes basée sur des algorithmes d'analyse syntaxique pour les automates arborescents pondérés. En prenant en entrée une performance musicale symbolique (séquence de notes avec dates et durées en temps réel, typiquement un fichier MIDI), et une grammaire hors-contexte pondérée décrivant un langage de rythmes préférés, il produit une partition musicale. Plusieurs formats de sortie sont possibles, dont XML, MEI.

Les principaux contributeurs sont :

- Florent Jacquemard (Inria) : développeur principal.
- Francesco Foscarin (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.

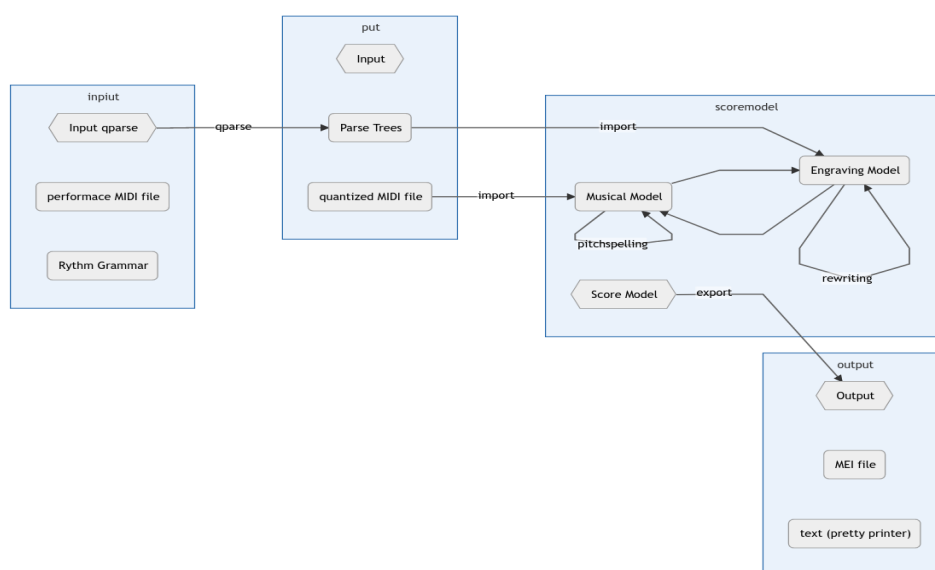


FIGURE 3.10 – Présentation de Qparse

Explication des différentes étapes de la figure 3.10⁶ :

— **Input Qparse :**

Un fichier MIDI (séquence d'événements datés (piano roll) accompagné d'un fichier contenant une grammaire pondérée) ;

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

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

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 partition à l'aide de technique de parsing / analyse syntaxique.

Le but du parsing/analyse syntaxique est en effet la structuration d'une représentation séquentielle en entrée (un mot fini), suivant un modèle de langage.

ref. Grune Jacobs "Parsing techniques" Springer 2007

dans le cas de qparse, le "mot" d'entrée est typiquement au format MIDI, et le modèle de langage est un un automate d'arbres pondéré représentant des préférences en terme de notation musicale à produire.

ref. "Handbook of weighted automata"

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

apprentissage

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

- 905 — **Arbre de parsing :**
- 906 Les données MIDI sont quantifiées, les notes de dates proches sont
- 907 alignées et les relations entre les notes sont identifiées (accords, fla,
- 908 etc. . .); un arbre de parsing global est créé;
- 909 — **Score Model :**
- 910 — Les instruments sont identifiés dans `scoremodel/import/tableImporterDrum.cpp`;
- 911
- 912 — Réécriture 1 :
- 913 séparation des voix \Rightarrow un arbre par voix \Rightarrow représentation in-
- 914 termédiaire (RI);
- 915 — Réécriture 2 :
- 916 simplification de l'écriture de chaque voix dans la RI;
- 917 — **Output :**
- 918 export de la partition. Plusieurs formats sont possibles (xml, mei,
- 919 lilypond, . . .).
- 920
- 921 Plusieurs enjeux :
- 922 — Problème du MIDI avec Qparse :
- 923 ON-OFF en entrée \Rightarrow 1 seul symbole en sortie.
- 924 — Minimiser la distance entre le midi et la représentation en arbre.
- 925 — Un des problèmes de Qparse était qu'il était limité au monopho-
- 926 nique.
- 927 Quelles sont les limites du monophonique?
- 928 — Impossibilité de traiter plusieurs voix et de reconnaître les ac-
- 929 cords.

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.

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

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 :

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

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

- 4/4 binaire
- 4/4 jazz
- 4/4 afro-cubain

Objectif des forme rythmiques

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

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

bien. il faudrait expliquer ça avant.

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

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

Détection d'indication de mesure

La détection de la signature rythmique est importante, non seulement pour connaître le nombre de temps par mesure ainsi que le nombre de subdivisions pour chacun de ces temps, mais aussi pour savoir comment écrire l'unité de temps et ses subdivisions.
La figure 3.11 montre deux indications de mesure différentes. L'une (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des triolets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est traditionnellement écrit en binaire avec ou sans triolet (même si cette musique est dite ternaire alors que le rock ternaire sera plutôt écrit comme dans l'exemple 2).



Exemple 1



Exemple 2

FIGURE 3.11 – signature rythmique

979 **Choix d'une grammaire**

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

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

Groove MIDI Data Set pas
présenté

méta-données

contenu

985 Pour les expériences effectuées avec le Groove MIDI Data Set, le style et
 986 l'indication de mesure sont récupérables par les noms des fichiers MIDI,
 987 mais il faudra par la suite les trouver automatiquement sans autres in-
 988 dications que les données MIDI elles-mêmes. Par conséquent, les motifs
 989 des forme rythmiques devront être recherchés sur l'input (*fichiers MIDI*)
 990 avant le lancement du parsing, afin de déterminer la signature rythmique
 991 en amont. Cette tâche devra probablement être effectuée en Machine
 992 Learning.

993 **Séparation des voix**

994 Ici, la forme rythmique est construit sur un modèle rock en 4/4 : after-beat
 995 sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. La
 996 forme rythmique est constituée par défaut du motif rd/pf/cc (voir 3.2) et
 997 d'un gamme jouée à la grosse caisse. La première ligne de la figure 3.12
 998 est appelée « Irréductible » car il n'y a pas d'autre choix pertinent pour
 999 la répartition de la ride et du charley au pied. La troisième séparation
 1000 proposée est privilégiée car elle répartit selon 2 voix, une voix pour les
 1001 mains (rd + cc) et une voix pour les pieds (pf + gc). Ce choix paraît plus
 1002 équilibré car deux instruments sont utilisés par voix et plus logique pour
 1003 le lecteur puisque les mains sont en haut et les pieds en bas.
 1004 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré

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

1005 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 1006 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici, le
 1007 parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et 2
 1008 combinées avec le *motif* de la première ligne, sont des cas typiques de la

quel exemple?

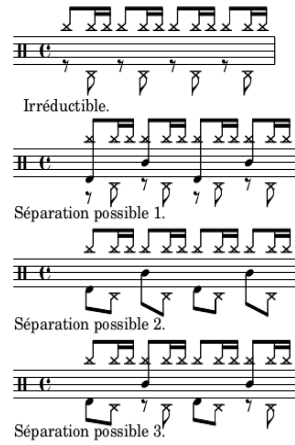


FIGURE 3.12 – Motif 4-4 binaire

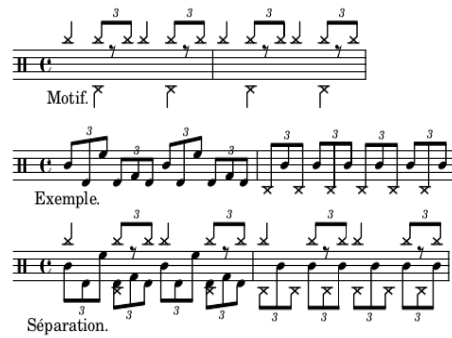


FIGURE 3.13 – Motif 4-4 jazz

1010 batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus, la
 1011 grosse caisse entre très souvent dans le flot des combinaisons de toms et
 1012 de caisse claire et son écriture séparée serait inutilement compliquée et
 1013 peu intuitive pour le lecteur. Le choix de séparation sera donc de laisser
 1014 les cymbales en haut et toms, caisse claire, grosse caisse et pédale de
 1015 charley en bas.



FIGURE 3.14 – forme rythmique 4-4 afro-latin

1016 La figure 3.14 montre un exemple minimaliste de forme rythmique afro-
 1017 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.

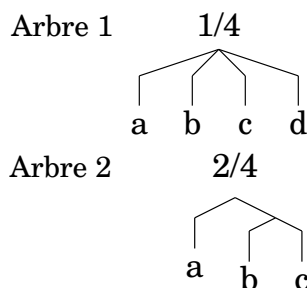


FIGURE 3.15 – Simplification

ce sont des figures et notations du chapitre suivant!

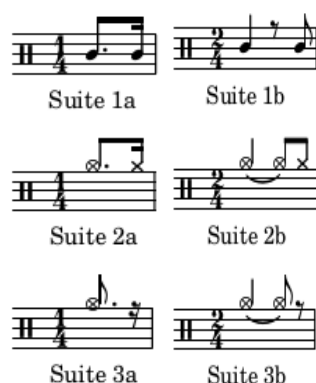


FIGURE 3.16

1047 Soit l'arbre 1 de la figure 3.15 dans lequel : a et d sont des instruments de
 1048 la batterie (x);
 1049 b et c sont des continuations (t);
 1050 Pour chacune des conditions suivantes, une suite de la figure 3.16 est
 1051 attribuée :

- 1052 — Si a n'est pas un co :
- 1053 ⇒ Suite 1a.
- 1054 — Si a est un co :
- 1055 — Si d est un cf :
- 1056 ⇒ Suite 2a.
- 1057 — Si d est un pf :
- 1058 ⇒ Suite 3a : d devient un silence (r).

1059
 1060 Soit l'arbre 2 de la figure 3.15 dans lequel :
 1061 a et c sont des instruments de la batterie (x);
 1062 b est une continuation (t); Pour chacune des conditions suivantes, une
 1063 suite de la figure 3.16 est attribuée :

- 1064 — Si a n'est pas un co :
- 1065 ⇒ Suite 1b, b devient un silence.
- 1066 — Si a est un co :
- 1067 — Si c est un cf :
- 1068 ⇒ Suite 2b, b devient une liaison et c devient un cf.
- 1069 — Si c est un pf :
- 1070 ⇒ Suite 3b : b devient une liaison et c devient un silence.

1071
 1072 *Rappel :*

- 1073 *cf* = charley fermé joué à la main ;
- 1074 *co* = charley ouvert joué à la main ;
- 1075 *pf* = charley fermé joué au pied.

1076

1077 **Conclusion**

1078 <dam>à développer un peu plus</dam> Nous avons formalisé une nota-
1079 tion de la batterie, modélisé cette notation pour la transcription de don-
1080 nées MIDI en partition, nous avons décrit Qparse.

1081 Enfin, nous avons exposé une approche de type dictionnaire (les « forme
1082 rythmiques ») pour détecter une signature rythmique, choisir une gram-
1083 maire pondérée appropriée et énoncer des règles de séparation des voix et
1084 de simplification de l'écriture.

1085

1086

EXPÉRIMENTATIONS

1087

Sommaire

1088	4.1	Le jeu de données	48
1089	4.2	Analyses et transcriptions manuelles	49
1090	4.3	Transcription polyphonique par parsing	53
1091	4.4	Réécriture guidée par une forme rythmique	56
1092	4.5	BILAN : résultats — évaluation — discussion	61
1093			
1094			
1095			

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.

4.1 Le jeu de données

Nous avons utilisé le Groove MIDI Dataset¹ [29] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).

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

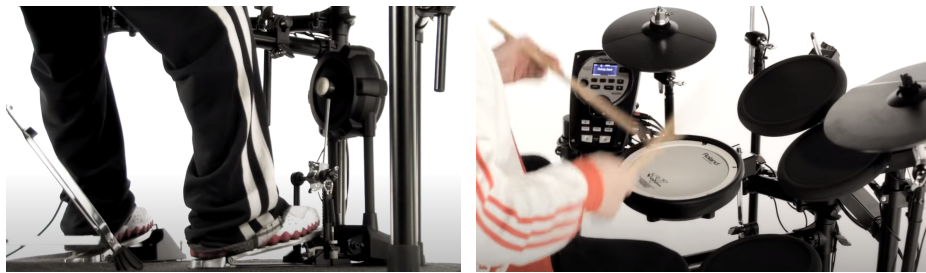


FIGURE 4.1 – Batterie électronique

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

1125

1126 Autres critères spécifiques au GMD :

- 1127 — Toutes les performances ont été jouées au métronome et à un tempo
1128 choisi par le batteur.
- 1129 — 80% de la durée du GMD a été joué par des batteurs professionnels
1130 qui ont pu improviser dans un large éventail de styles. Les don-
1131 nées sont donc diversifiées en termes de styles et de qualités de jeu
1132 (professionnel ou amateur).
- 1133 — Les batteurs avaient pour instruction de jouer des séquences de
1134 plusieurs minutes ainsi que des fills²
- 1135 — Chaque performance est annotée d'un style (fourni par le batteur),
1136 d'une signature rythmique et d'un tempo ainsi que d'une identifi-
1137 cation anonyme du batteur.
- 1138 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
1139 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
1140 session du GMD.
- 1141 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
1142 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.

1143 **Format des données**

1144 Le Roland TD-11 enregistre les données dans des fichiers MIDI et les
1145 divise en plusieurs pistes distinctes :

- 1146 — une pour le tempo et l'indication de mesure ;
- 1147 — une pour les changements de contrôle (position de la pédale de
- 1148 charley) ;
- 1149 — une pour les notes.

1150

1151 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1152 canal 9 (qui est le canal canonique pour la batterie).

1153 Pour simplifier le traitement de ces données, ces trois pistes ont été fu-
1154 sionnées en une seule piste qui a été mise sur le canal 9.

1155 **4.2 Analyses et transcriptions manuelles**

1156 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
1157 partir de fichiers MIDI et Audio du GMD.

1158 **Comparaisons de transcriptions**

1159 Pour les comparaisons de transcriptions, les transcriptions manuelles
1160 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les transcrip-
1161 tions automatiques (TA) ont toutes été générées par import d'un fichier
1162 MIDI dans MuseScore.

1163 **Exemple d'analyse 1**

Transcription manuelle ⇒ Transcription automatique



- 1164 — Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
- 1165 — Les silences de la mesure 1 de la TA sont inutilement surchargés ;
- 1166 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1167 premières notes (une double-croche et une croche) d'un triolet sur
- 1168 le temps 1 de la mesure 2 de la TA.

Transcription manuelle \Rightarrow Transcription automatique



1169 Exemple d'analyse 2

- 1170 — Les doubles croches ont été interprétées en quintolet
- 1171 — La deuxième double-croche est devenue une croche.

1172 Exemple d'analyse 3

Transcription manuelle \Rightarrow Transcription automatique

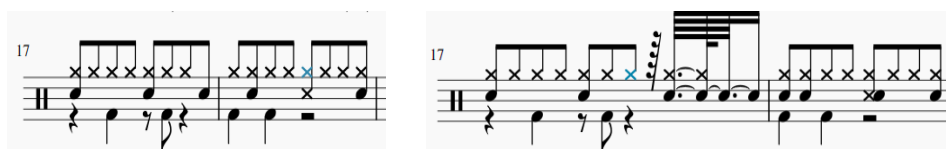


- 1173 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1174 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été décalés d'une double croche vers la droite dans la TA.
- 1175 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
- 1176 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.

1180
1181

1182 Exemple d'analyse 4

Transcription manuelle \Rightarrow Transcription automatique



1183

- 1184 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une manière excessivement complexe!
- 1185

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

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

1186 **Exemple d'analyse 5 (flas)**

1187 Transcription manuelle



1188

1189 Transcription automatique

1190



1191

1192

- 1193 — Le premier fla est reconnu comme étant un triolet contenant une
- 1194 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1195 ornementée.
- 1196 — Le deuxième fla est reconnu comme étant un accord.
- 1197 — Les deux double en contre-temps sur le temps 4 de la TM sont mal
- 1198 quantifiées dans la TA.
- 1199 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1200 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1201 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1202 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1203 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1204 la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de
- 1205 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1206 temps 3 de la TA.
- 1207 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1208 exemple, le charley ouvert en contre-temps sur le temps 2 de la
- 1209 mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem
- 1210 pour les cross-sticks.

1211 **Conclusion d'analyse**

1212 Ces analyses ont montré la difficulté pour un logiciel comme MuseScore
 1213 d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI
 1214 ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance
 1215 de la forme du rythme impliquant sa position dans la mesure. Cette recon-
 1216 naissance pourrait permettre de rectifier les problèmes de signature ryth-
 1217 mique ainsi que les problèmes de décalage de temps. La reconnaissance
 1218 de la forme du rythme permettrait aussi de supprimer les aberrations du
 1219 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

Music engraving by LilyPond 2.22.1—www.lilypond.org

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). 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>

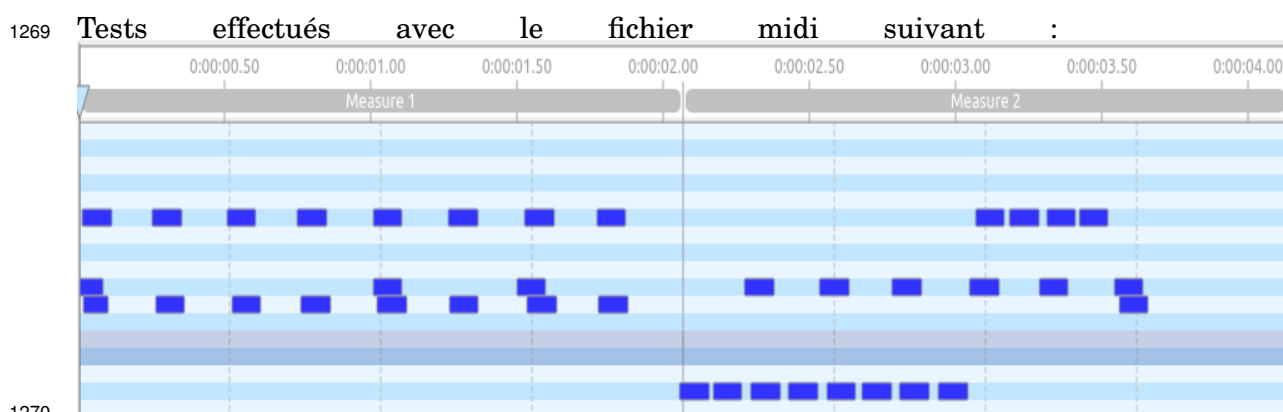
Les Jams

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.



1271 Un premier test convaincant est effectué avec la grammaire suivante :

1272

1273 <dam>incompréhensible ainsi, c'est dommage</dam> // bar level

1274 0 -> C0 1

1275 0 -> E1 1

1276 0 -> U4(1, 1, 1, 1) 1

1277

1278 // half bar level

1279 9 -> C0 1

1280 9 -> E1 1

1281

1282 // beat level

1283 1 -> C0 1

1284 1 -> E1 1

1285 1 -> T2(2, 2) 1

1286 1 -> T4(4, 4, 4, 4) 1

1287

1288 // croche level

1289 2 -> C0 1

1290 2 -> E1 1

1291

1292 // double level

1293 4 -> C0 1

1294 4 -> E1 1

1295 4 -> E2 1

1296 4 -> T2(6, 6) 1

1297

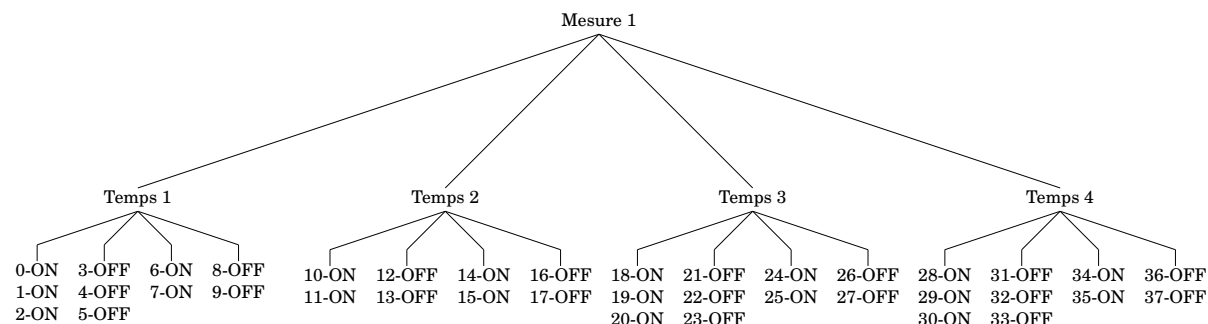
1298 // triple level

1299 6 -> E1 1

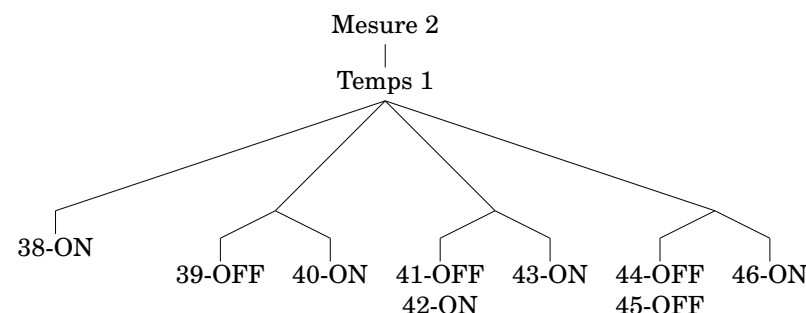
1300

1301 Cette grammaire sépare les ligatures par temps au niveau de la
1302 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.

1327 4.4 Réécriture guidée par une forme rythmique

1328 La démonstration qui suit est basée sur la partition de référence de la
 1329 figure 4.2 puisque la forme rythmique qui sera utilisée en est directement
 1330 extraite.

1331

1332 Nous allons montrer :

- 1333 — la composition de cette forme rythmique ;
- 1334 — son état finale, c'est à dire toutes les combinaisons entièrement
- 1335 écrites en notation correcte sur partition ;
- 1336 ⇒ cela constituera une référence pour la réécriture ;
- 1337 — un exemple de transformation de la forme rythmique en arbre de
- 1338 rythme ;
- 1339 — l'application de la séparation des voix sur cet exemple basé sur la
- 1340 référence citée précédemment (la forme rythmique en question) ;
- 1341 ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a
- 1342 de voix (deux arbres pour cette forme rythmique) ;
- 1343 — les règles de simplification propres à la forme rythmique dont nous
- 1344 parlons.

1345 L'objectif de cette démonstration est de montrer comment un jeu de plu-
 1346 sieurs formes rythmiques pourrait être implémenter dans le cadre d'une
 1347 approche dictionnaire.

1348 Motifs et gammes

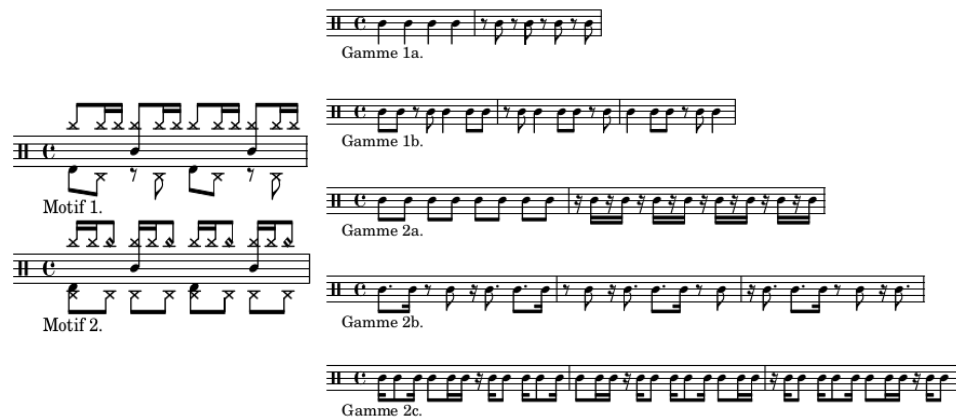


FIGURE 4.3 – Motifs et gammes

1349 Motifs

1350 À partir de la partition de référence, les deux motifs de la figure 4.3
 1351 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure

1352 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à
 1353 la mesure 28 avec des variations. Ces deux motifs sont très classiques et
 1354 pourront être détectés dans de nombreuses performances.
 1355

1356 **Gammes**

1357 Les gammes de la figure 4.3 étayent toutes les combinaisons d'un motif
 1358 en 4/4 binaires jusqu'aux doubles croches.
 1359 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-
 1360 mière ne contient que des noires et la deuxième que des croches en contre-
 1361 temps. Ces deux possibilités sont combinées de manière circulaire dans
 1362 les 3 mesures de la deuxième ligne.
 1363 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures
 1364 dont la première ne contient que des croches et la deuxième que des
 1365 doubles-croches en contre-temps. Ces deux possibilités sont combinées de
 1366 manière circulaire dans les lignes 4 et 5 qui contiennent chacune 3 me-
 1367 sures.

1368 **Formes rythmiques — motifs et gammes combinés**

1369 Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure
 1370 4.3.<dam>à commenter un peu plus, notamment pour dire si la combinai-
 son est faite automatiquement ou non</dam>

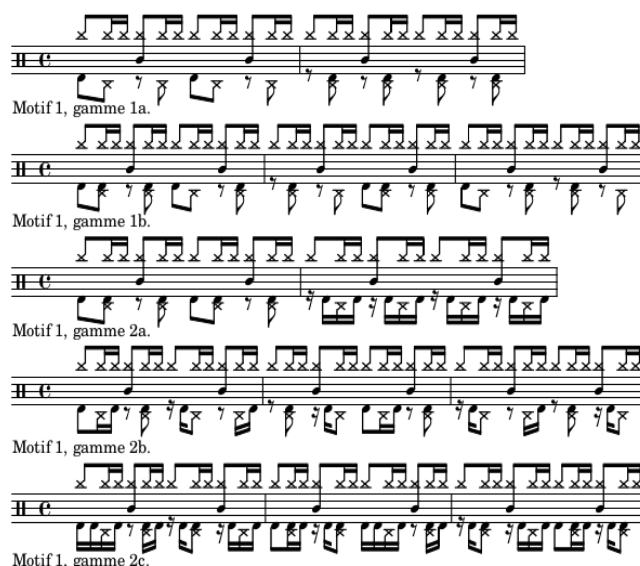


FIGURE 4.4 – Partition d'un forme rythmique en 4/4 binaire

1372 Représentation de la forme rythmique en arbres de 1373 rythmes

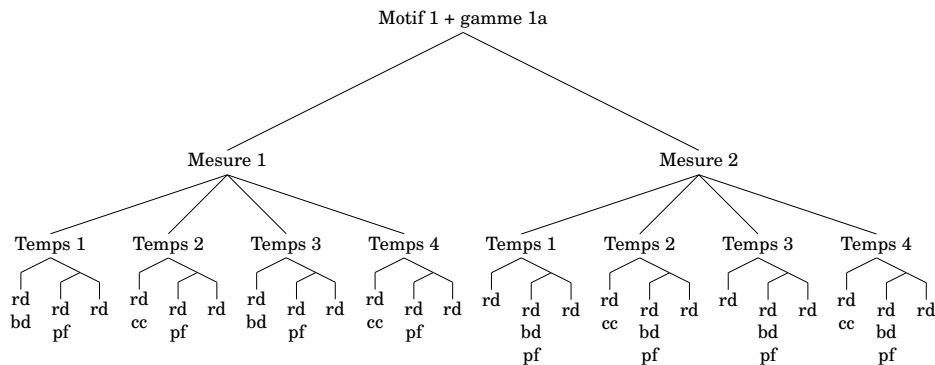


FIGURE 4.5 – Arbre de rythme — forme rythmique

1374 L'arbre de la figure 4.5 servira de base pour le suite de l'expérimentation.
 1375 Comme indiqué à la racine de l'arbre, il représente la première ligne de la
 1376 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,
 1377 il manque des indications de notation telles que les voix spécifiques à
 1378 chaque partie du rythme ainsi que les choix d'écriture pour les distances
 1379 qui séparent les notes de chaque voix entre elles en termes de durée.

1380 Réécriture — séparation des voix et simplification

1381 La séparation des voix

1382 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui
 1383 le constituent et les voix seront regroupées en suivant les règles du forme
 rythmique.

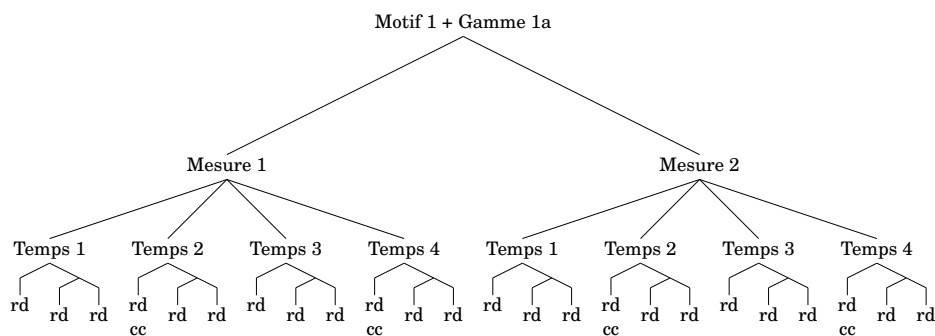


FIGURE 4.6 – Arbre de rythme — voix haute

1384 La voix haute (figure 4.6) regroupe la ride et la caisse-claire sur les liga-
 1385 tures du haut.
 1386

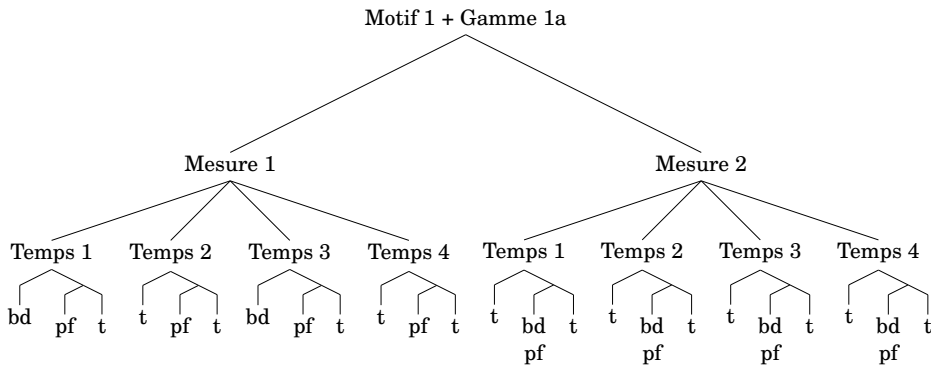


FIGURE 4.7 – Arbre de rythme — voix basse

1387 La voix basse (figure 4.7 regroupe la grosse-caisse et le charley au pied
1388 sur les ligatures du bas.

1389 **Les règles de simplifications**

1390 L’objectif des règles de simplifications est de réécrire les écarts de durées
1391 qui séparent les notes d’une manière appropriée pour la batterie et qui
1392 soit la plus simple possible. Les ligatures relient les notes d’un temps
1393 entre elles afin de rendre la pulsation visuelle).

1394

1395 Pour les figures ci-dessous :

- 1396 — x = une note ;
1397 — r = un silence ;
1398 — t = une continuation (point ou liaison)

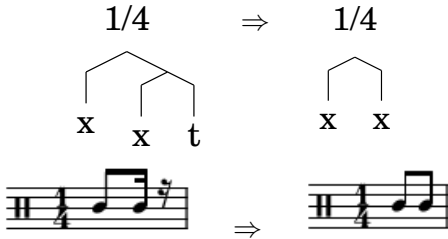


FIGURE 4.8

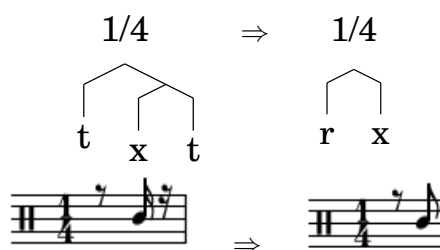


FIGURE 4.9

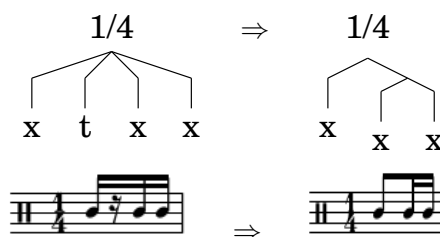


FIGURE 4.10

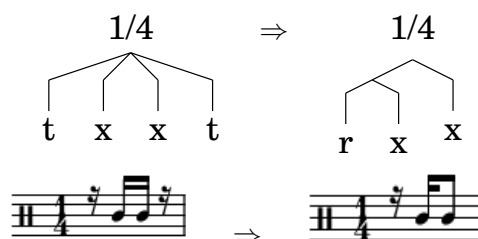


FIGURE 4.11

- 1399 Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique.
 1400 Les arbres manquants seront mis en annexe.
 1401 Les règles remplacent par un silence les continuations (t) qui sont au dé-
 1402 but d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il
 1403 y a des ouvertures de charley, cela n'est pas toujours applicable.

1404 Conclusion sur cette réécriture guidée

- 1405 La méthode des formes rythmiques étant basée sur une approche diction-
 1406 naire, Le premier objectif de cette réécriture guidée est d'orienter la re-
 1407 cherche d'autres formes rythmiques par observation du jeu de données et
 1408 de montrer comment les construire pour agrandir la base de connaissance
 1409 de Qparse pour la transcription de la batterie.

1410 4.5 BILAN : résultats — évaluation — discussion

1411 Cette section regroupe les avancées qui ont été réalisées par rapport aux
 1412 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
 1413 tats de l'ADT avec Qparse. Nous avons amélioré le système de quantifi-
 1414 cation de Qparse pour la batterie, notamment le passage à la polyphonie
 1415 avec les Jams.

1416 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
 1417 grammaires avec des fichiers MIDI courts.

1418 Puis, une sortie MEI a aussi été obtenue (encore à vérifier).

1419 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
 1420 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
 1421 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.

1422 — Le choix de travailler avec Lilypond et non Verovio. Ce choix était
 1423 motivé par la liberté totale concernant la notation de la batterie
 1424 dont un et la disponibilité d'un set de notation de type Agostini.
 1425 C'est la seule application qui me permettait d'écrire la notation de
 1426 la batterie exactement comme je le souhaitais.

1427 — Avancé de la chaîne de traitement (nous sommes arrivés aux arbres
 1428 de parsing, nous avons traité le polyphonique (identification des
 1429 regroupements de notes⁵) ⇒ Quelques arbres ont été obtenus sur
 1430 des exemples simples⁶)

1431 — 2 dimensions de le travail fourni :

1432 - La volonté de pousser un exemple simple jusqu'au bout de la
 1433 chaîne pour obtenir des résultats et une évaluation sur au moins
 1434 un exemple ; - La réalité du travail à fournir pour faire avancer sur
 1435 la chaîne de traitement. ⇒ Une solution aurait été de considérer
 1436 les arbres de parsing obtenus après le traitement du polyphonique
 1437 comme un résultat local possible à évaluer au lieu d'attendre que la
 1438 chaîne arrive jusqu'à la génération d'une partition mais cela n'était
 1439 pas prioritaire pendant le stage.

1440 — Création d'un jeu de forme rythmique basique représentatif des
 1441 différents styles à recouvrir. Ce jeu n'a pas pu être créé, car comme
 1442 vu plus haut, je me suis focalisé sur un exemple pour pouvoir le
 1443 vérifier entièrement et dans l'espoir de pouvoir le tester en fin de
 1444 chaîne. **Évaluation** Matcher les motifs aurait été indispensable
 1445 pour obtenir une quantité de résultats qui justifieraient une
 1446 évaluation automatique permettant de faire des graphiques.

1447 L'évaluation fut entièrement manuelle car :

1448 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
 1449 VS output) Pour l'évaluation, il aurait fallu produire un module.

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

1450 <dam>je ne sais pas si tu auras encore le temps de faire ça, sinon
1451 il faudra décrire comment tu aurais aimé évaluer, proprement et
1452 sans résultats chiffrés</dam> L'évaluation est-elle automatique ou
1453 manuelle ?
1454 Possibilité d'un export lilypond en arbre pour comparer l'output
1455 avec la transcription manuelle.
1456 Possibilité de transformer lilypond(output) et lilypond(ref) en
1457 ScoreModel ou MEI pour les comparer et faire des statistiques.
1458 Si transformés en MEI : diffscore de Francesco. Possibilité de
1459 transformer lilypond(output) et lilypond(ref) en MusicXML pour
1460 les comparer ou dans Music21. L'expérimentation peut-être consi-
1461 dérer comme une évaluation manuelle ? (magicien d'Oz)
1462 Lilypond vers MIDI + output vers MIDI \Rightarrow Comparaison des MIDI
1463 dumpés.
1464
1465 La transcription automatique de la batterie est un sujet passionnant mais
1466 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-
1467 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-
1468 moins été générée. Elle sera un bon point de départ pour les travaux fu-
1469 turs dont plusieurs propositions sont énoncés dans le présent document.

CONCLUSION GÉNÉRALE

1471 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1472 tion automatique de la batterie. Son objectif était de transcrire, à partir
1473 de leur représentation symbolique MIDI, des performances de batteur de
1474 différents niveaux et dans différents styles en partitions écrites.

1475 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1476 cessus de regroupement des événements MIDI qui nous a permis de faire
1477 la transition du monophonique vers le polyphonique. Une des données
1478 importante de ce processus était de différencier les nature des notes d'un
1479 accord, notamment de distinguer lorsque 2 notes constituent un accord
1480 ou un *fla*.

1481 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1482 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1483 tionnables en amont du parsing, soit par indication des noms des fichiers
1484 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1485 naire de patterns prédéfinis ⁷ qu'il serait pertinent de mettre en œuvre en
1486 machine learning.

1487 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1488 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1489 d'un système et pour la simplification grâce aux gammes du motif d'un
1490 système. Nous avons aussi montré comment, dans des travaux futurs, un
1491 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1492 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1493 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1494 la chaîne de traitement en terme de réécriture.

1495 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1496 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1497 entièrement automatisée de partitions de batterie et au problème général
1498 de l'AMT de bout en bout.[11]

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

BIBLIOGRAPHIE

- 1500 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1501 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1502 1996. – Cité pages 7, 18 et 35.
- 1503 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1504 et 15.
- 1505 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1506 page 12.
- 1507 [4] Gaël Richard et al. De fourier à la reconnaissance
1508 musicale. Available at [https://interstices.info/
1509 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1510 – Cité page 12.
- 1511 [5] Caroline Traube. Quelle place pour la science au sein de la musico-
1512 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1513 [6] Leonard Bernstein Office. The unanswered question : Six talks at
1514 harvard. Available at [https://leonardbernstein.com/about/
1515 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1516 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1517 entre les traitements de la musique et du langage. *La Lettre des
1518 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1519 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1520 Techniques de traitement automatique du langage naturel appli-
1521 quées aux représentations symboliques musicales. In *JIM 2021 -
1522 Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1523 Cité page 13.
- 1524 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-
1525 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1526 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1527 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1528 Cité page 13.
- 1529 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1530 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1531 lenges and future directions. *Journal of Intelligent Information Sys-*
1532 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 63.
- 1533 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1534 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1535 Cité pages 14 et 27.
- 1536 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-
1537 SURE, 1997. – Cité page 15.
- 1538 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1539 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1540 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1541 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1542 1483, 2018. – Cité pages 15, 23 et 27.
- 1543 [15] Moshekwa Malatji. Automatic music transcription for two instru-
1544 ments based variable q-transform and deep learning methods, 10
1545 2020. – Cité page 22.
- 1546 [16] Antti J. Eronen. Musical instrument recognition using ica-based
1547 transform of features and discriminatively trained hmms. *Seventh*
1548 *International Symposium on Signal Processing and Its Applications*,
1549 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1550 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1551 drum sound description for real-world music using template adap-
1552 tation and matching methods. *International Conference on Music*
1553 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1554 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1555 musical statistics as guides for audio-to-score piano transcription.
1556 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24, 26 et 34.
- 1557 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1558 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1559 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1560 *Computation in Music*, volume Lecture Notes in Computer Science
1561 of *Proceedings of the Seventh International Conference on Mathema-*
1562 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1563 2019. Springer. – Cité pages 24 et 26.
- 1564 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1565 ring of rhythm structures. In *Proceedings of the First International*
1566 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1567 IEEE Computer Society. – Cité page 26.
- 1568 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1569 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1570 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1571 Cité page 26.
- 1572 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1573 equivalent rhythmic notations based on rhythm tree languages. In
1574 *Third International Conference on Technologies for Music Notation
1575 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1576 pez Palma and Mike Solomon. – Cité page 26.
- 1577 [23] R. Marxer and J. Janer. Study of regularizations and constraints in
1578 nmf-based drums monaural separation. In *International Conference
1579 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1580 02/09/2013 2013. – Cité page 27.
- 1581 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1582 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 29.
- 1583 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1584 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 29.
- 1585 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1586 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1587 94120 Fontenay-sous-Bois, 2000. – Cité pages 31 et 43.
- 1588 [27] Nicolas Guiomard-Kagan. *Traitement de la polyphonie pour l'analyse
1589 informatique de partitions musicales*. PhD thesis. – Cité page 34.
- 1590 [28] Frédéric Canet. La batterie... mot à mot! Available at [https://](https://rimshotetghostnote.fr/)
1591 rimshotetghostnote.fr/ (2021). – Cité page 36.
- 1592 [29] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1593 Bamman. Learning to groove with inverse sequence transforma-
1594 tions. In *International Conference on Machine Learning (ICML)*,
1595 2019. – Cité page 48.

