

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	36
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	Présentation de Qparse	39
66	3.10	signature rythmique	41
67	3.11	Motif 4-4 binaire	42
68	3.12	Motif 4-4 jazz	42
69	3.13	forme rythmique 4-4 afro-latin	43
70	3.14	Simplification	44
71	3.15	44
72	4.1	Batterie électronique	48
73	4.2	Partition de référence	52
74	4.3	Motifs et gammes	56
75	4.4	Partition d'un forme rythmique en 4/4 binaire	57
76	4.5	Arbre de rythme — forme rythmique	58
77	4.6	Arbre de rythme — voix haute	58
78	4.7	Arbre de rythme — voix basse	59
79	4.8	59
80	4.9	60

81	4.10	60
82	4.11	60

83 **LISTE DES TABLEAUX**

84	1.1 speechToText vs AMT	13
85	3.1 Noms des instruments de la batterie	30
86	3.2 Pitches et instruments	36
87	3.3 Systèmes	40

INTRODUCTION GÉNÉRALE

89 QUOI?

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

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

111

112 POURQUOI?

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

119

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

122 COMMENT?

123 → Problématique :

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

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

134

135 → Méthodes :

136 → Contributions :

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

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

147

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

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

164

1. MuseScore3

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

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

178

179 PLAN

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

188

CONTEXTE

189

190

Sommaire

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

195

196

Introduction

199

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

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

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

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

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

1.1 Langues naturelles et musique en informatique

COMPUTER MUSIC

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

RIM

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

RIM et TAL

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

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

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

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

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

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

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

exemples / illustration de la proximité thématique?

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

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

TABLE 1.1 – speechToText vs AMT

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

286 1.2 La transcription automatique de la musique

287 1. OBJECTIF

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

299

300 2. APPLICATIONS

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

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

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

318

319 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

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

8. Ici, on parle de notation occidentale.

9. ethno-musicologie

10. par exemple par la constitution de corpus musicologiques

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

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

1.3 La transcription automatique de la batterie

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

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

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

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

cite méthode et école Agostini?

12. Roland TR-808

13. SQ-1

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

367 1.4 Les représentations de la musique

citer M. Müller FMP pour
 cette section ?

369 Les données audio

trop technique. ne pas re-
 pier wikipédia

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

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

375 Les données MIDI

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

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

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

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

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

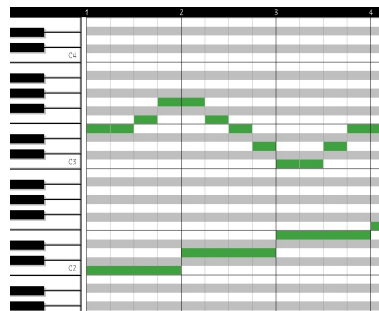


FIGURE 1.1 – Exemple événements avec durée

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

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

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

389 Ici, définir Tick et channel.

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

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

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

FIGURE 1.2 – Critère pour un évènement

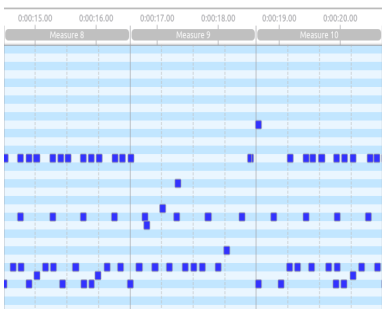


FIGURE 1.3 – Exemple évènements sans durée

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

402 **Les partitions**
403

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

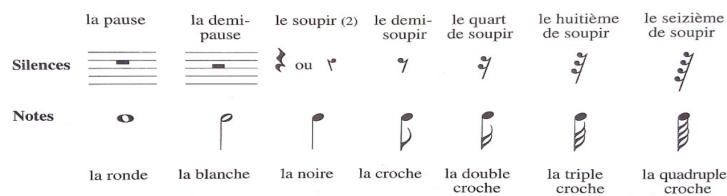


FIGURE 1.4 – Les silences



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

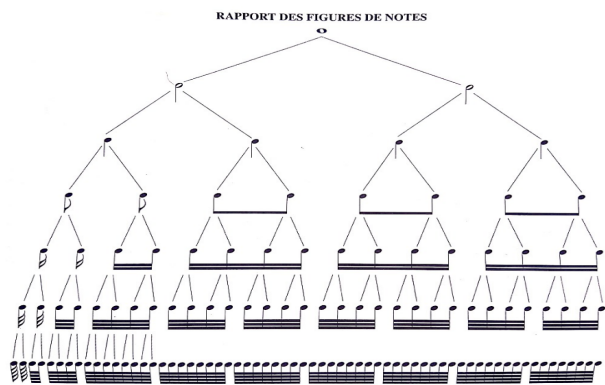


FIGURE 1.6 – Rapport des figures de notes
[1]

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

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



FIGURE 1.7 – Exemple de partition de piano

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

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

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

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

443 Les formats XML

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

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

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

453 appoggiatures

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

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

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

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

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

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

```

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

```

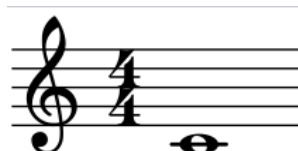


FIGURE 1.8 – MusicXML

signature rythmique

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

Conclusion

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

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

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

ÉTAT DE L'ART

Sommaire

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

Introduction

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

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

2.1 Monophonique et polyphonique

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

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

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

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

2.2 Audio vers MIDI

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

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

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

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

MIDI non-quantifié = performance (à expliquer)

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

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

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

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

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

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

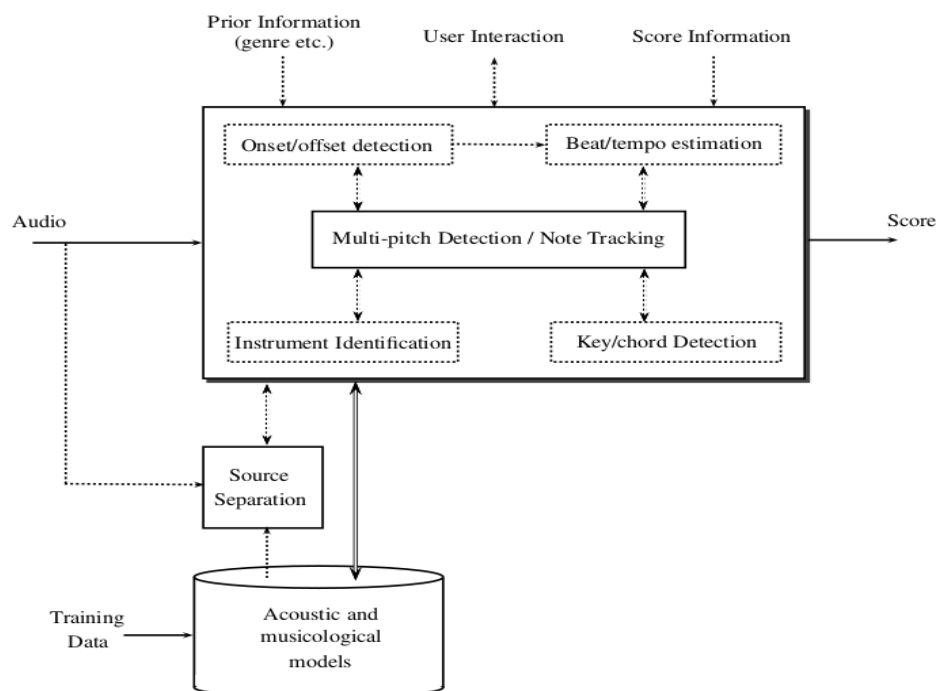


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

Les sous-systèmes et algorithmes optionnels sont présentés à l'aide de lignes pointillées. Les doubles flèches mettent en évidence les connexions entre les systèmes qui incluent la fusion d'informations et une communication plus interactive entre les systèmes.

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

haute fréquence, aigus?

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

classification des évène- 554
 ments? la phrase semble 555
 redondante

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

562 2.3 MIDI vers partition

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

ce n'est pas exactement 563
 cela. cf. proposition de des- 564
 cription + détaillée en com- 565
 mentaires

de manière conjointe 574

langage a priori 576

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

2.4 Approche linéaire et approche hiérarchique

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

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

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

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

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

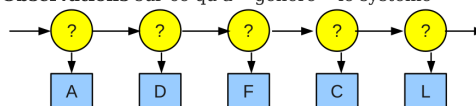


FIGURE 2.2 – HMM

593 *Source : Cours de Damien Nouvel*⁵

594

595

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

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

RT? 608
 609

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

le calcul d'équiv. 613
 614

citer thèse de David Rizo 617
 (Valencia) 618

619

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.s.net/fr/enseignement>

Example: *Summertime*

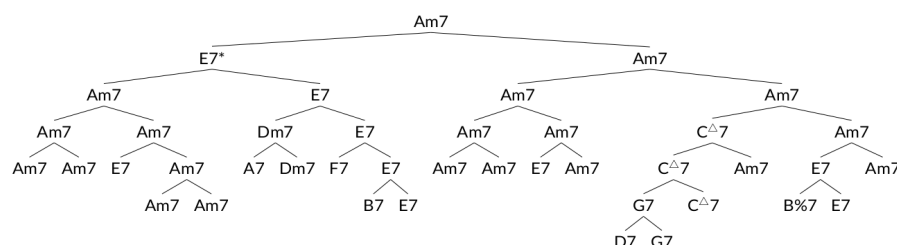


FIGURE 2.3 – arbre jazz

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 événements 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.

MÉTHODES

Sommaire

3.1	La notation de la batterie	29
3.2	Modélisation pour la transcription	36
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/>

668 Les hauteurs et les têtes de notes



FIGURE 3.1 – Les instruments de la batterie

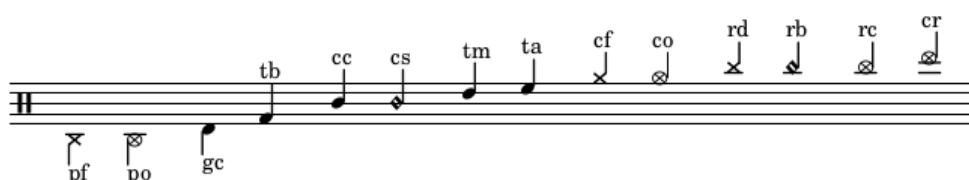


FIGURE 3.2 – Hauteur et têtes de notes

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

TABLE 3.1 – Noms des instruments de la batterie

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

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

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

Les durées

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

Par conséquent :

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

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

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

711 En résumé :

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

719

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

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



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

730 Un autre élément concernant la notation des durées en batterie est la né-
 731 cessité de faire ressortir la pulsation⁴ de la rendre visuelle. La première
 732 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 repaire visuel sur le temps.

En cas de nécessité de prolonger la durée d'une note au-delà de sa durée initial, et si cette note correspond à une ouverture de charley, on privilégiera donc la liaison.

Les silences

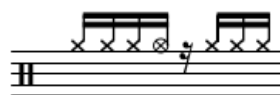
Les silences sont parfois utilisés pour noter les fermetures de charley (après une ouverture). Les fermetures du charley sont notées soit par un silence (correspondant à une fermeture de la pédale), soit par un écrasement de l'ouverture par un autre coup de charley fermé, au pied ou à la main.

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

pas très clair



Exemple 1



Exemple 2

FIGURE 3.4 – Silence joué

L'exemple 1 de la figure 3.4 montre ce qui est écrit dans les données MIDI et l'exemple 2 montre ce que le batteur doit penser en lisant la partition. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1 donnerait avec une partition comprenant plusieurs voix et plusieurs instruments jouant simultanément.

762 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte
 763 la note suivante pour l'écriture : - Si c'est un charley fermé joué à la main
 764 \Rightarrow la note sera cf;
 765 - Si c'est un charley fermé joué au pied \Rightarrow la note sera un silence.

766 Les équivalences rythmiques

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

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



FIGURE 3.5 – Équivalence

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

782 Les voix

783 Les voix⁵ désignent les différentes parties mélodiques constituant une
 784 composition musicale et destinées à être interprétées, simultanément ou
 785 successivement, par un ou plusieurs musiciens. En batterie, une voix
 786 est l'ensemble des instruments qui, à eux seuls, constituent une phrase
 787 rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures
 788 étant possibles pour un même rythme, on peut regrouper les instruments
 789 de la batterie par voix. Sur une portée de batterie, il existe le plus souvent
 1 ou 2 voix. Sur la figure 3.6, il faudra faire un choix entre les exemples
 1, 2 et 3 qui sont trois façons d'écrire le même rythme.
 Ce choix se fera en fonction des instruments joués, de la nature plus ou

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

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

une voix est caractérisée aussi par l'orientation des hampes?

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



FIGURE 3.6 – Séparation des voix

793 moins systématique de leurs phrasés, et des associations logiques entre
 794 les instruments dans la distribution des rythmes sur la batterie (voir la
 795 section 3.4).

796 Les accentuations et les ghost-notes

797 « Certaines notes dans une phrase musicale doivent, ainsi que les dif-
 798 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

799 La figure 3.7 ne prend en compte que les accents que nous avons estimés
 800 nécessaires (voir la section 3.2). Les accents sont marqués par le sym-
 801 bole « > ». Il est positionné au-dessus des notes représentant des cym-
 802 bales et en-dessous des notes représentant des toms ou la caisse claire.
 803 Ce choix a été fait pour la partition de la figure 4.2 car elle est plus li-
 804 sible ainsi, mais ces choix devront être adaptés en fonction des diffé-
 805 rents formes rythmiques reconnues (voir la section 3.4). Par exemple, pour les
 806 formes rythmiques jazz, les ligatures pour les toms et la caisse claire se-
 807 ront dirigés vers le bas, il faudra donc mettre les symboles d'accentuation
 808 correspondants au-dessus des têtes de notes.
 809

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

expliquer ce qu'est une *ghost-notes*

810

811

812

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

813

814

La dernière note de la figure 3.7 montre un exemple de ghost-notes. Le parenthésage a été choisi car il peut être utilisé sur n'importe quelle note sans changer la tête de note.

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. 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

815

816 Les flas

817 ICI, DESCRIPTION DES FLAS!

818 3.2 Modélisation pour la transcription

819 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

820 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 instru-

je ne comprend pas cette phrase.

821

822

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

ment, 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 mo-
derne, 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 né-
cessitent pas de faire des choix conditionnés par les limitations humaines
(2 pieds, 2 mains, et beaucoup plus d'instruments...)

La vitesse

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 :

- Vitesse inférieure à 40 : ghost-note ;
- Vitesse supérieure à 90 : accent ;
- Pas d'intention d'accent ni de ghost-note pour une vitesse entre 40
et 89 ;
- Les accents et les ghosts-notes ne sont significatifs ni pour les ins-
truments joués au pied, ni pour les cymbales crash.

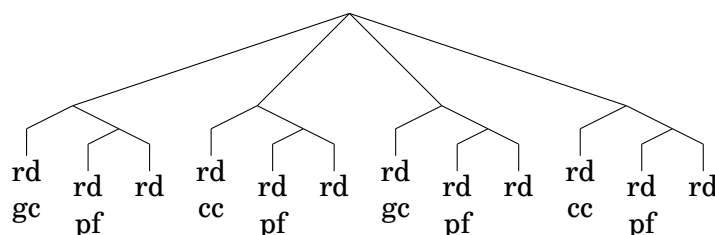
En effet, certaines vitesses en dessous de 40 étant détectées et ins-
crites dans les données MIDI sont dues au mouvement du talon du
batter qui bat la pulsation sans particulièrement jouer le charley.
Ce mouvement est perçu par le capteur de la batterie électronique
mais le charley n'est pas joué.

- Au final, nous avons relevé les ghost-notes et les accents pour la
caisse claire ainsi que les accents pour les toms et les cymbales
rythmiques (charley et ride).

Les arbres de rythmes

Les arbres de rythmes représentent un rythme unique dont les possibi-
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 :



854

citation lilypond

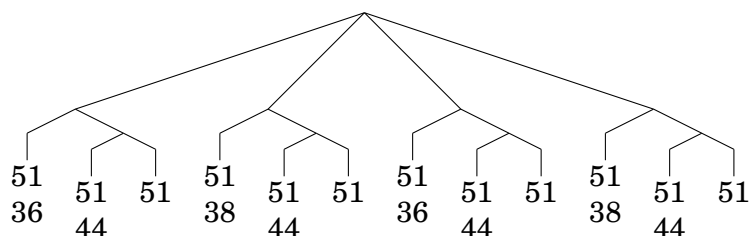
et l'analyse d'autres fichiers
MIDI ?

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

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

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

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



856
857 Chacun des trois exemples de la figure 3.6 est représenté par un des deux
858 arbres syntaxiques ci-dessus.

859 3.3 Qparse

La librairie Qparse⁶ 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 Foscari (PhD, CNAM) : construction de grammaire automatique à partir de corpus ; Evaluation.
- Clement Poncelet (Salzburg U.) : integration de la librairie Midifile pour les input MIDI.
- Philippe Rigaux (CNAM) : production de partition au format MEI et de modèle intermédiaire de partition en sortie.
- Masahiko Sakai (Nagoya U.) : mesure de la distance input/output pour la quantification et CMake framework ; évaluation.

Explication des différentes étapes de la figure 3.9⁷ :

— Input Qparse :

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

— Arbre de parsing :

Les données MIDI sont quantifiées, les notes de dates proches sont alignées et les relations entre les notes sont identifiées (accords, fla, etc. . .) ; un arbre de parsing global est créé ;

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

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

choisir titre plus explicite
par ex. analyse syntaxique
pour la transcription musicale

quantification rythmique
+ structuration de partition

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

Il effectue conjointement
des tâches de quantification
rythmique et d'inférence
de la structure de la 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 automate
d'arbres pondéré représentant
des préférences en terme de notation
musicale à produire.

ref. "Handbook of weighted
automata"

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

apprentissage

la figure 3.11 est trop compliquée.
rhythm grammar → automate d'arbres
pondéré. Parse Tree → arbre
syntaxique. qtz MIDI file : inutile.
Score Model → représentation
intermédiaire de partition. Score Model.

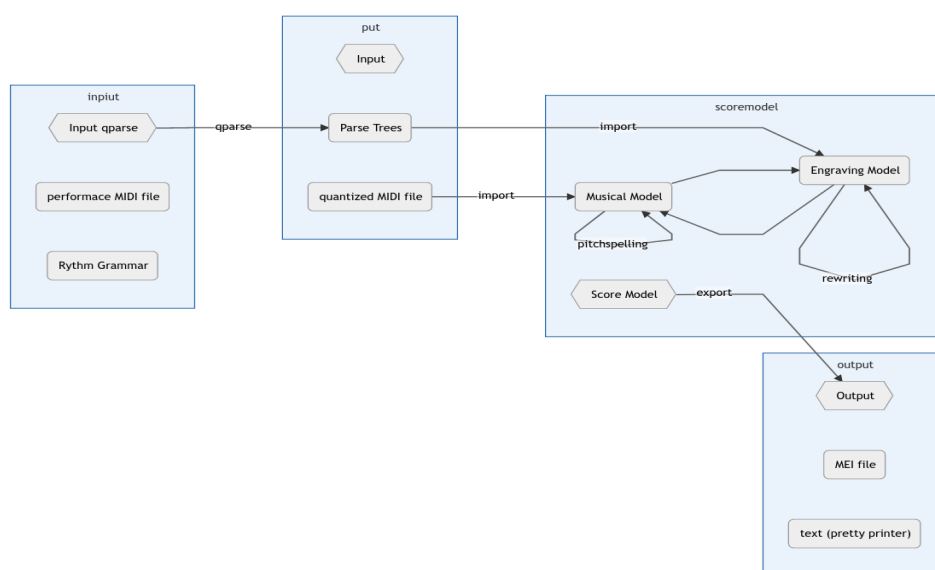


FIGURE 3.9 – Présentation de Qparse

- 887 — **Score Model :**
- 888 — Les instruments sont identifiés dans scoremodel/import/tableImporterDrum.cpp ;
- 889 — Réécriture 1 :
- 890 — séparation des voix \Rightarrow un arbre par voix \Rightarrow représentation intermédiaire (RI) ;
- 891 — Réécriture 2 :
- 892 — simplification de l'écriture de chaque voix dans la RI ;
- 893 — **Output :**
- 894 — export de la partition. Plusieurs formats sont possibles (xml, mei, lilypond,...).

898

899 Plusieurs enjeux :

- 900 — Problème du MIDI avec Qparse :
- 901 — ON-OFF en entrée \Rightarrow 1 seul symbole en sortie.
- 902 — Minimiser la distance entre le midi et la représentation en arbre.
- 903 — Un des problèmes de Qparse était qu'il était limité au monophonique.
- 904 — Quelles sont les limites du monophonique ?
- 905 — Impossibilité de traiter plusieurs voix et de reconnaître les accords.

3.4 Les forme rythmiques

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

Définitions

forme rythmique : motif + gamme/texte

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

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

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

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

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

TABLE 3.3 – Systèmes

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

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

Objectif des forme rythmiques

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

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

La partie *motif* des forme rythmiques sera utilisée pour la **définition des signature rythmiques**. Le *motif* et la gammes des forme rythmiques se-

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

946 Détection d'indication de mesure

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

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



Exemple 1



Exemple 2

FIGURE 3.10 – signature rythmique

La figure 3.10 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).

957 Choix d'une grammaire

Il faut prendre en compte l'existence potentielle de plusieurs grammaires dédiées chacune à un type de contenu MIDI. Le choix d'une grammaire pondérée doit être fait avant le parsing puisque Qparse prend en entrée un fichier MIDI et un fichier wta (grammaire). C'est pour cette raison que la signature rythmique doit être définie avant le choix de la grammaire. Pour les expériences effectuées avec le Groove MIDI Data Set, le style et l'indication de mesure sont récupérables par les noms des fichiers MIDI, mais il faudra par la suite les trouver automatiquement sans autres indications que les données MIDI elles-mêmes. Par conséquent, les motifs des formes rythmiques devront être recherchés sur l'input (*fichiers MIDI*) avant le lancement du parsing, afin de déterminer la signature rythmique en amont. Cette tâche devra probablement être effectuée en Machine Learning.

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

Groove MIDI Data Set pas présenté

méta-données

contenu

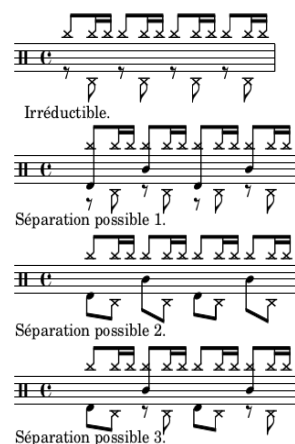
971 **Séparation des voix**

FIGURE 3.11 – Motif 4-4 binaire

les description ic sont as-972
sez techniques et difficile à
suivre. avant de détailler 973
des exemples, il faudrait
décrire les objectifs et le
principe de la procédure. 975

976 Ici, le forme rythmique est construit sur un modèle rock en 4/4 : after-beat
977 sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. La
978 forme rythmique est constituée par défaut du motif rd/pf/cc (voir 3.2) et
979 d'un gamme jouée à la grosse caisse. La première ligne de la figure 3.11
980 est appelée « Irréductible » car il n'y a pas d'autre choix pertinent pour
981 la répartition de la ride et du charley au pied. La troisième séparation
proposée est privilégiée car elle répartit selon 2 voix, une voix pour les
mains (rd + cc) et une voix pour les pieds (pf + gc). Ce choix paraît plus
équilibré car deux instruments sont utilisés par voix et plus logique pour
le lecteur puisque les mains sont en haut et les pieds en bas.

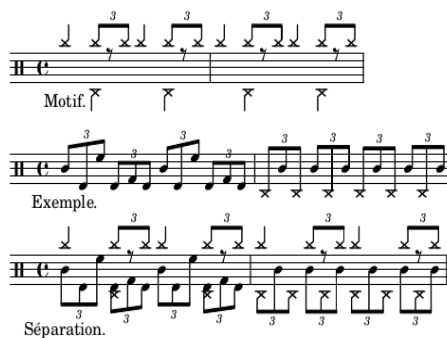


FIGURE 3.12 – Motif 4-4 jazz

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

quel exemple?

986

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



FIGURE 3.13 – forme rythmique 4-4 afro-latin

La figure 3.13 montre un exemple minimaliste de forme rythmique afro-latin [26]. Ce forme rythmique doit être écrit sur trois voix car la voix centrale est souvent plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas serait surchargé et peu lisible.

1008 Simplification de l'écriture

1009 Les explications qui suivent seront appuyé par une réécriture guidée par
1000 une forme rythmique dans la section 4.4.

1001 Les gammes qui accompagnent les motifs d'un forme rythmique étayent
1002 toutes les combinaisons d'un forme rythmique et elles permettent, combi-
1003 nées avec le motif d'un forme rythmique, de définir les règles de simplifi-
1004 cation propres à celui-ci.

1005 Voici les différentes étapes à suivre :

- 1006 — Pour chaque gamme du forme rythmique, faire un arbre de rythme
- 1007 représentant la gamme combinée avec le motif du forme ryth-
- 1008 mique ;
- 1009 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
- 1010 arbre de rythme par voix ;
- 1011 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
- 1012 nœuds qui nécessitent une simplification et écrire la règle.

1013 Certaines précisions concernant l'extraction de ces règles sont néces-
1014 saires. Il s'agit de précisions à propos de la durée, des silences et de la
1015 présence ou non d'ouverture de charley dans les instruments joués. Nous
1016 avons discuté de ces problèmes dans le chapitre 3.

1017 Voici quelques règles inhérentes à la simplification de l'écriture pour la
1018 batterie : Toutes les continuations (t) qui se trouvent en début de temps
1019 (figures 4.9, 4.11 et ??) sont transformées en silences (r) sauf si la note
1020 précédente est un charley ouvert ?

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

1021 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
 1022 2 notes au sein d'un même temps et favorise, une liaison si co, un point si
 1023 pas co et nécessaire, un simple ajustement de la figure de note si suffisant.
 1024

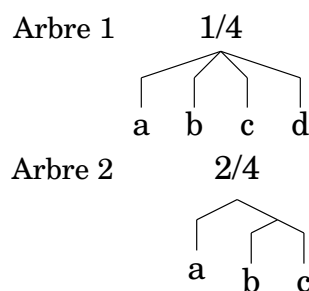


FIGURE 3.14 – Simplification

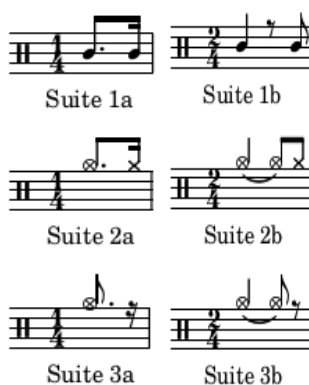


FIGURE 3.15

itemize

1025 Soit l'arbre 1 de la figure 3.14 dans lequel : a et d sont des instruments de
 1026 la batterie (x);
 1027 b et c sont des continuations (t);
 1028 Pour chacune des conditions suivantes, une suite de la figure 3.15 est
 1029 attribuée :
 1030 — Si a n'est pas un co :
 1031 ⇒ Suite 1a.
 1032 — Si a est un co :
 1033 — Si d est un cf :
 1034 ⇒ Suite 2a.
 1035 — Si d est un pf :
 1036 ⇒ Suite 3a : d deviens un silence (r).
 1037

1038 Soit l'arbre 2 de la figure 3.14 dans lequel :
1039 a et c sont des instruments de la batterie (x);
1040 b est une continuation (t); Pour chacune des conditions suivantes, une
1041 suite de la figure 3.15 est attribuée :
1042 — Si a n'est pas un co :
1043 ⇒ Suite 1b, b devient un silence.
1044 — Si a est un co :
1045 — Si c est un cf :
1046 ⇒ Suite 2b, b devient une liaison et c devient un cf.
1047 — Si c est un pf :
1048 ⇒ Suite 3b : b devient une liaison et c devient un silence.
1049
1050 *Rappel :*
1051 *cf = charley fermé joué à la main ;*
1052 *co = charley ouvert joué à la main ;*
1053 *pf = charley fermé joué au pied.*
1054

1055 Conclusion

1056 <dam>à développer un peu plus</dam> Nous avons formalisé une nota-
1057 tion de la batterie, modélisé cette notation pour la transcription de don-
1058 nées MIDI en partition, nous avons décrit Qparse.
1059 Enfin, nous avons exposé une approche de type dictionnaire (les « formes
1060 rythmiques ») pour détecter une signature rythmique, choisir une gram-
1061 maire pondérée appropriée et énoncer des règles de séparation des voix et
1062 de simplification de l'écriture.

1063

1064

EXPÉRIMENTATIONS

1065

Sommaire

1066	4.1	Le jeu de données	48
1067	4.2	Analyses et transcriptions manuelles	49
1068	4.3	Transcription polyphonique par parsing	53
1069	4.4	Réécriture guidée par une forme rythmique	56
1070	4.5	BILAN : résultats — évaluation — discussion	61

1071
1072
1073
1074

Introduction

1075

1076 Dans ce chapitre, nous présenterons le jeu de données et les analyse
1077 MIDI-Audio et transcriptions manuelles.

1078 Problématique :

1079 choix d'un outil de transcription ?

1080 passage au polyphonique indispensable pour la suite du travail et pour
1081 l'usage des formes rythmiques.

1082 Finir la chaîne de traitement indispensable pour obtenir des résultats
1083 chiffrés possible à évaluer.

1084 Nous présenterons mes trois contributions principales :

1085 — le code lilypond normalisé pour la transcription de la batterie avec
1086 la notation de type agostini.

1087 — les différentes étapes de résolution du passage au polyphonique.

1088 — l'expérimentation d'un forme rythmique implémentable qui devra
1089 être utilisé comme base de connaissances pour augmenter la ra-
1090 pidité et la qualité en sortie de Qparse et comme une méthode de
1091 création de nouvelles formes rythmiques.

1092 Enfin, nous finirons par une discussion sur les avancées réalisées dans ce
1093 travail, la pertinence des choix qui ont été faits et les moyens d'évaluer
1094 les résultats potentiels.

1095 4.1 Le jeu de données

1096 Nous avons utilisé le Groove MIDI Dataset¹ [27] (GMD) qui est un jeu de
 1097 données mis à disposition par Google sous la licence Creative Commons
 1098 Attribution 4.0 International (CC BY 4.0).

1099 Le GMD est composé de 13,6 heures de batterie sous forme de fichiers
 1100 MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000
 1101 mesures de batterie dans les styles les plus courants et avec différentes
 1102 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

1103

1104 Autres critères spécifiques au GMD :

- 1105 — Toutes les performances ont été jouées au métronome et à un tempo
 1106 choisi par le batteur.
- 1107 — 80% de la durée du GMD a été joué par des batteurs professionnels
 1108 qui ont pu improviser dans un large éventail de styles. Les don-
 1109 nées sont donc diversifiées en termes de styles et de qualités de jeu
 1110 (professionnel ou amateur).
- 1111 — Les batteurs avaient pour instruction de jouer des séquences de
 1112 plusieurs minutes ainsi que des fills²
- 1113 — Chaque performance est annotée d'un style (fourni par le batteur),
 1114 d'une signature rythmique et d'un tempo ainsi que d'une identifi-
 1115 cation anonyme du batteur.
- 1116 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
 1117 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
 1118 session du GMD.
- 1119 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
 1120 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.

1121 **Format des données**

1122 Le Roland TD-11 enregistre les données dans des fichiers MIDI et les
1123 divise en plusieurs pistes distinctes :

- 1124 — une pour le tempo et l'indication de mesure ;
- 1125 — une pour les changements de contrôle (position de la pédale de
1126 charley) ;
- 1127 — une pour les notes.

1128

1129 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1130 canal 9 (qui est le canal canonique pour la batterie).

1131 Pour simplifier le traitement de ces données, ces trois pistes ont été fu-
1132 sionnées en une seule piste qui a été mise sur le canal 9.

1133 **4.2 Analyses et transcriptions manuelles**

1134 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
1135 partir de fichiers MIDI et Audio du GMD.

1136 **Comparaisons de transcriptions**

1137 Pour les comparaisons de transcriptions, les transcriptions manuelles
1138 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les transcrip-
1139 tions automatiques (TA) ont toutes été générées par import d'un fichier
1140 MIDI dans MuseScore.

1141 **Exemple d'analyse 1**

Transcription manuelle ⇒ Transcription automatique



- 1142 — Erreur d'indication de mesure (3/4 au lieu de 4/4) ;
- 1143 — Les silences de la mesure 1 de la TA sont inutilement surchargés ;
- 1144 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
1145 premières notes (une double-croche et une croche) d'un triolet sur
1146 le temps 1 de la mesure 2 de la TA.

Transcription manuelle \Rightarrow Transcription automatique



1147 Exemple d'analyse 2

- 1148 — Les doubles croches ont été interprétées en quintolet
- 1149 — La deuxième double-croche est devenue une croche.

1150 Exemple d'analyse 3

Transcription manuelle \Rightarrow Transcription automatique



- 1151 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1152 — 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.
- 1153 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
- 1154 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.

1160 Exemple d'analyse 4

Transcription manuelle \Rightarrow Transcription automatique



- 1161
- 1162 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- 1163 manière excessivement complexe !

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

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

1164 **Exemple d'analyse 5 (flas)**

1165 Transcription manuelle



1166

1167 Transcription automatique

1168



1169

1170

- 1171 — Le premier fla est reconnu comme étant un triolet contenant une
- 1172 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1173 ornementée.
- 1174 — Le deuxième fla est reconnu comme étant un accord.
- 1175 — Les deux double en contre-temps sur le temps 4 de la TM sont mal
- 1176 quantifiées dans la TA.
- 1177 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1178 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1179 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1180 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1181 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1182 la noire (pf, voir le tableau 3.2) sur le temps 1 de la mesure 2 de
- 1183 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1184 temps 3 de la TA.
- 1185 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1186 exemple, le charley ouvert en contre-temps sur le temps 2 de la
- 1187 mesure 2 de la TM devrait avoir le même symbole sur la TA. Idem
- 1188 pour les cross-sticks.

1189 **Conclusion d'analyse**

1190 Ces analyses ont montré la difficulté pour un logiciel comme MuseScore
 1191 d'offrir une partition lisible. Les raisons sont le fait que les fichiers MIDI
 1192 ne sont pas encore quantifiés mais aussi qu'il n'y a pas de reconnaissance
 1193 de la forme du rythme impliquant sa position dans la mesure. Cette recon-
 1194 naissance pourrait permettre de rectifier les problèmes de signature ryth-
 1195 mique ainsi que les problèmes de décalage de temps. La reconnaissance
 1196 de la forme du rythme permettrait aussi de supprimer les aberrations du
 1197 type de celle de l'exemple d'analyse 4, puisque l'erreur sur cet exemple

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

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

Transcription de partition

FIGURE 4.2 – Partition de référence

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

Cette transcription a été entièrement faite avec Lilypond (voir le code lilypond sur le git https://github.com/MartinDigard/Stage_M2_Inria). 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>

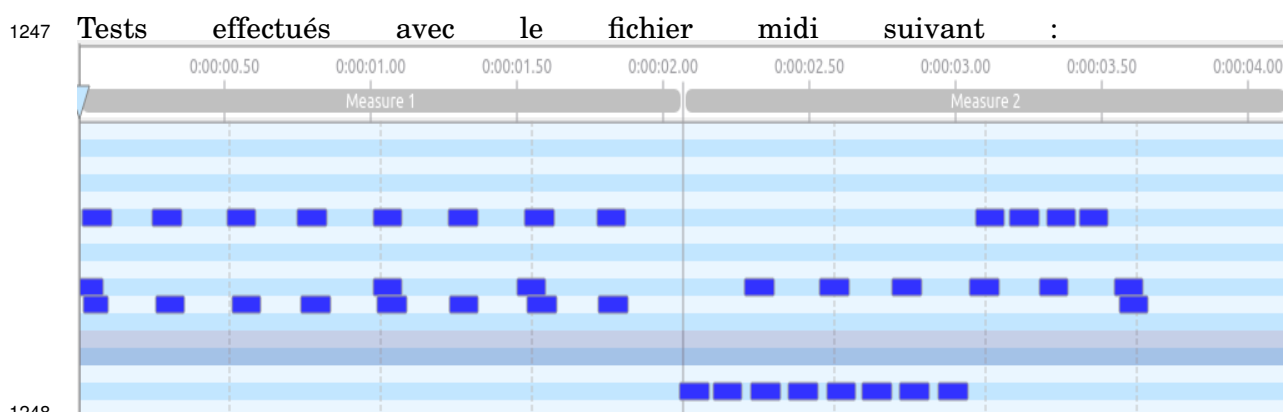
Les Jams

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.



1249 Un premier test convaincant est effectué avec la grammaire suivante :

1250

1251 <dam>incompréhensible ainsi, c'est dommage</dam> // bar level

1252 0 -> C0 1

1253 0 -> E1 1

1254 0 -> U4(1, 1, 1, 1) 1

1255

1256 // half bar level

1257 9 -> C0 1

1258 9 -> E1 1

1259

1260 // beat level

1261 1 -> C0 1

1262 1 -> E1 1

1263 1 -> T2(2, 2) 1

1264 1 -> T4(4, 4, 4, 4) 1

1265

1266 // croche level

1267 2 -> C0 1

1268 2 -> E1 1

1269

1270 // double level

1271 4 -> C0 1

1272 4 -> E1 1

1273 4 -> E2 1

1274 4 -> T2(6, 6) 1

1275

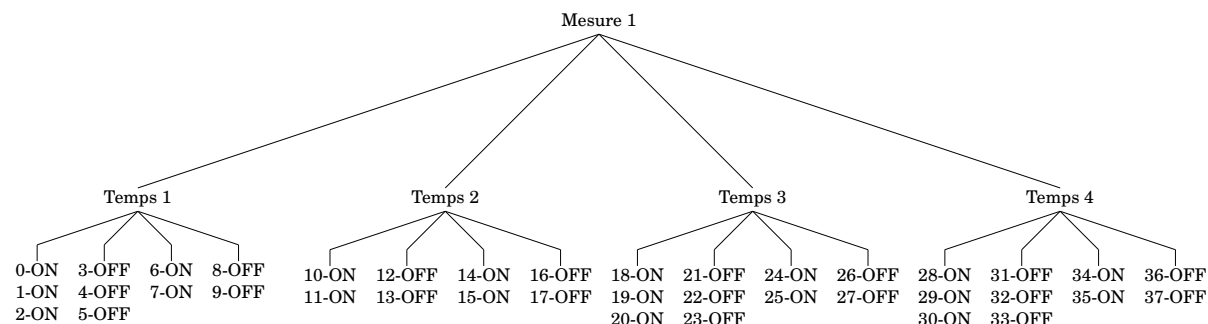
1276 // triple level

1277 6 -> E1 1

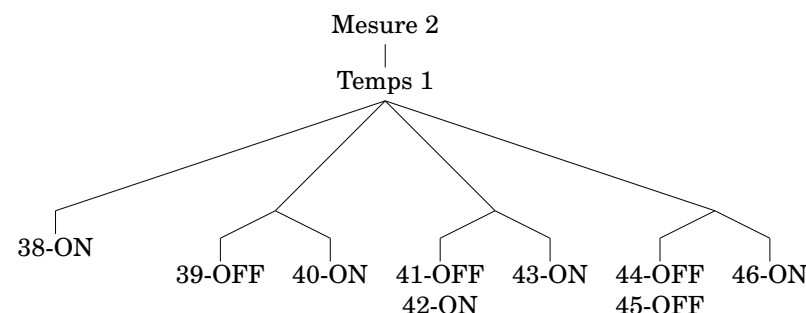
1278

1279 Cette grammaire sépare les ligatures par temps au niveau de la
1280 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.

1305 4.4 Réécriture guidée par une forme rythmique

1306 La démonstration qui suit est basée sur la partition de référence de la
 1307 figure 4.2 puisque la forme rythmique qui sera utilisée en est directement
 1308 extraite.

1309

1310 Nous allons montrer :

- 1311 — la composition de cette forme rythmique ;
- 1312 — son état finale, c'est à dire toutes les combinaisons entièrement
 1313 écrites en notation correcte sur partition ;
 1314 ⇒ cela constituera une référence pour la réécriture ;
- 1315 — un exemple de transformation de la forme rythmique en arbre de
 1316 rythme ;
- 1317 — l'application de la séparation des voix sur cet exemple basé sur la
 1318 référence citée précédemment (la forme rythmique en question) ;
 1319 ⇒ l'arbre de départ sera alors séparé en autant d'arbres qu'il y a
 1320 de voix (deux arbres pour cette forme rythmique) ;
- 1321 — les règles de simplification propres à la forme rythmique dont nous
 1322 parlons.

1323 L'objectif de cette démonstration est de montrer comment un jeu de plu-
 1324 sieurs formes rythmiques pourrait être implémenter dans le cadre d'une
 1325 approche dictionnaire.

1326 Motifs et gammes

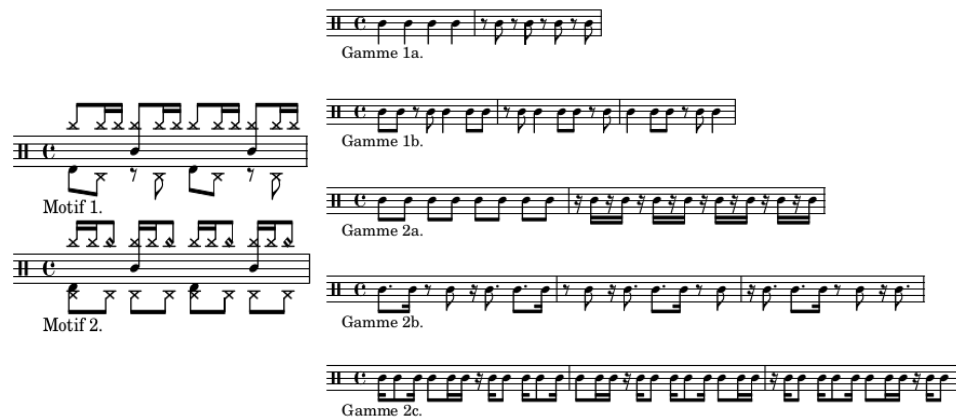


FIGURE 4.3 – Motifs et gammes

1327 Motifs

1328 À partir de la partition de référence, les deux motifs de la figure 4.3
 1329 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure

1330 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à
 1331 la mesure 28 avec des variations. Ces deux motifs sont très classiques et
 1332 pourront être détectés dans de nombreuses performances.
 1333

1334 **Gammes**

1335 Les gammes de la figure 4.3 étayent toutes les combinaisons d'un motif
 1336 en 4/4 binaires jusqu'aux doubles croches.
 1337 Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la pre-
 1338 mière ne contient que des noires et la deuxième que des croches en contre-
 1339 temps. Ces deux possibilités sont combinées de manière circulaire dans
 1340 les 3 mesures de la deuxième ligne.
 1341 Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures
 1342 dont la première ne contient que des croches et la deuxième que des
 1343 doubles-croches en contre-temps. Ces deux possibilités sont combinées de
 1344 manière circulaire dans les lignes 4 et 5 qui contiennent chacune 3 me-
 1345 sures.

1346 **Formes rythmiques — motifs et gammes combinés**

1347 Pour la suite de cette démonstration, je utiliserai le motif 1 de la figure
 1348 4.3.<dam>à commenter un peu plus, notamment pour dire si la combinai-
 son est faite automatiquement ou non</dam>

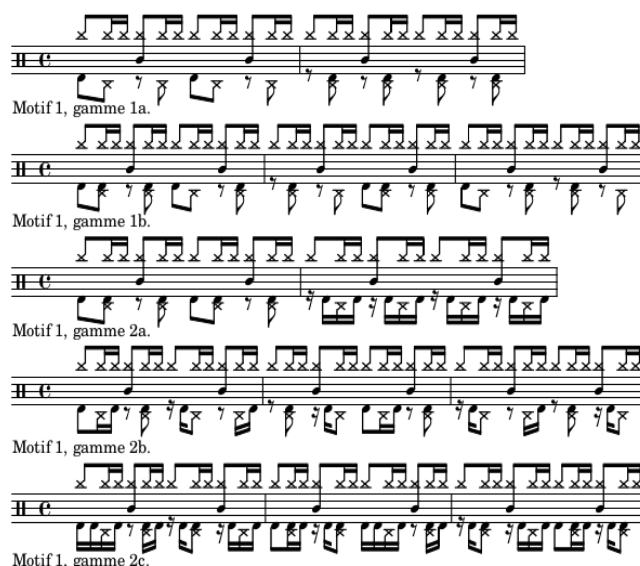


FIGURE 4.4 – Partition d'un forme rythmique en 4/4 binaire

1350 **Représentation de la forme rythmique en arbres de** 1351 **rythmes**

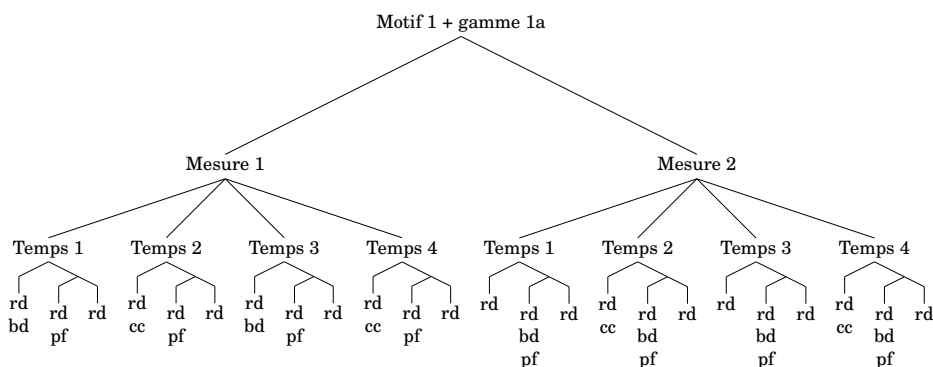


FIGURE 4.5 – Arbre de rythme — forme rythmique

1352 L'arbre de la figure 4.5 servira de base pour le suite de l'expérimentation.
1353 Comme indiqué à la racine de l'arbre, il représente la première ligne de la
1354 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,
1355 il manque des indications de notation telles que les voix spécifiques à
1356 chaque partie du rythme ainsi que les choix d'écriture pour les distances
1357 qui séparent les notes de chaque voix entre elles en termes de durée.

1358 **Réécriture — séparation des voix et simplification**

1359 **La séparation des voix**

1360 Ainsi l'arbre syntaxique de départ est divisé en autant d'instruments qui
1361 le constituent et les voix seront regroupées en suivant les règles du forme
rythmique.

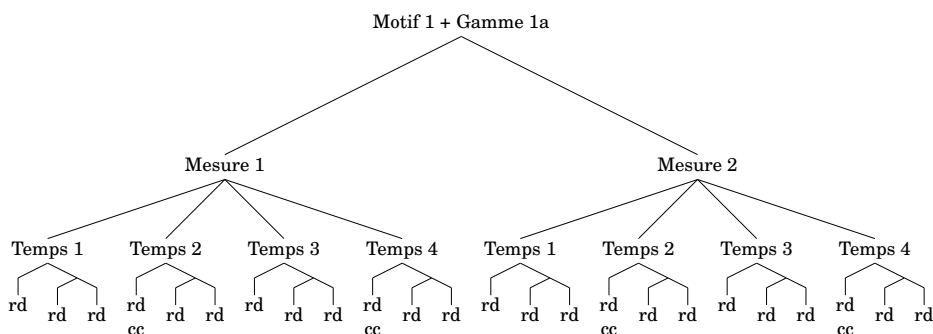


FIGURE 4.6 – Arbre de rythme — voix haute

1362 La voix haute (figure 4.6) regroupe la ride et la caisse-claire sur les liga-
1363 tures du haut.
1364

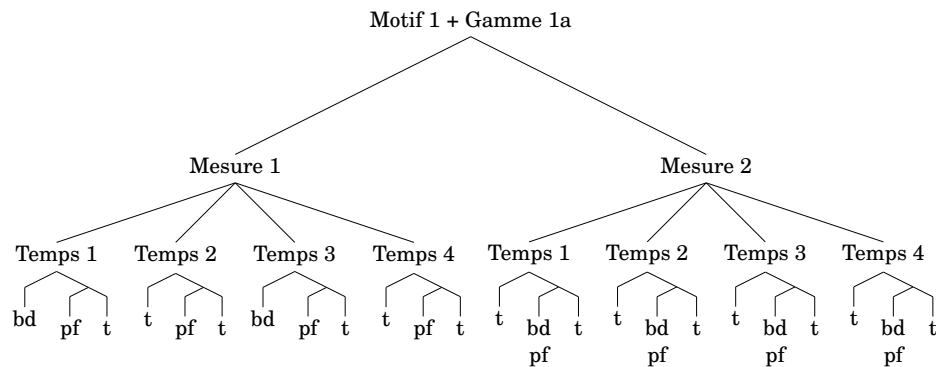


FIGURE 4.7 – Arbre de rythme — voix basse

1365 La voix basse (figure 4.7 regroupe la grosse-caisse et le charley au pied
1366 sur les ligatures du bas.

1367 **Les règles de simplifications**

1368 L'objectif des règles de simplifications est de réécrire les écarts de durées
1369 qui séparent les notes d'une manière appropriée pour la batterie et qui
1370 soit la plus simple possible. Les ligatures relient les notes d'un temps
1371 entre elles afin de rendre la pulsation visuelle).

1372

1373 Pour les figures ci-dessous :

1374 — x = une note ;

1375 — r = un silence ;

1376 — t = une continuation (point ou liaison)

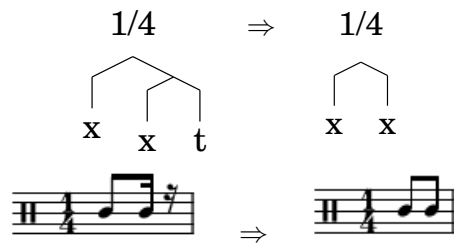


FIGURE 4.8

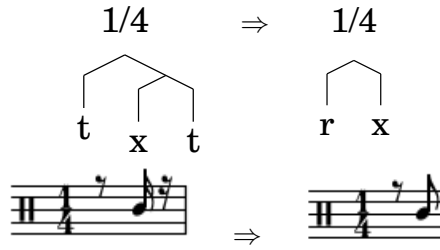


FIGURE 4.9

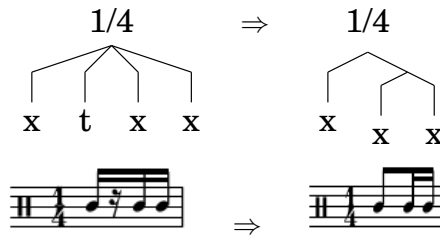


FIGURE 4.10

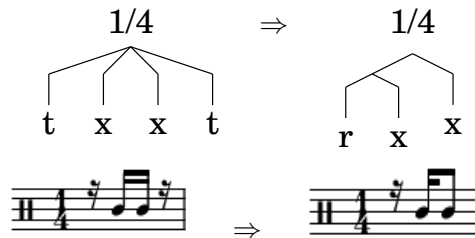


FIGURE 4.11

- 1377 Ces règles ont été tirées de l'ensemble des arbres de la forme rythmique.
 1378 Les arbres manquants seront mis en annexe.
 1379 Les règles remplacent par un silence les continuations (t) qui sont au dé-
 1380 but d'un temps. Cela est valable pour cette forme rythmique mais lorsqu'il
 1381 y a des ouvertures de charley, cela n'est pas toujours applicable.

1382 Conclusion sur cette réécriture guidée

- 1383 La méthode des formes rythmiques étant basée sur une approche diction-
 1384 naire, Le premier objectif de cette réécriture guidée est d'orienter la re-
 1385 cherche d'autres formes rythmiques par observation du jeu de données et
 1386 de montrer comment les construire pour agrandir la base de connaissance
 1387 de Qparse pour la transcription de la batterie.

1388 4.5 BILAN : résultats — évaluation — discussion

1389 Cette section regroupe les avancées qui ont été réalisées par rapport aux
1390 objectifs de départ ainsi qu’une réflexion sur le moyen d’évaluer les résul-
1391 tats de l’ADT avec Qparse. Nous avons amélioré le système de quantifi-
1392 cation de Qparse pour la batterie, notamment le passage à la polyphonie
1393 avec les Jams.

1394 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
1395 grammaires avec des fichiers MIDI courts.

1396 Puis, une sortie MEI a aussi été obtenue (encore à vérifier).

1397 Dans cette section, nous discuterons sur la pertinence de l’ensemble des
1398 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
1399 ont été faites ou non et nous tenterons d’en expliquer la ou les raisons.

1400 — Le choix de travailler avec Lilypond et non Verovio. Ce choix était
1401 motivé par la liberté totale concernant la notation de la batterie
1402 dont un et la disponibilité d’un set de notation de type Agostini.
1403 C’est la seule application qui me permettait d’écrire la notation de
1404 la batterie exactement comme je le souhaitais.

1405 — Avancé de la chaîne de traitement (nous sommes arrivés aux arbres
1406 de parsing, nous avons traité le polyphonique (identification des
1407 regroupements de notes⁵) ⇒ Quelques arbres ont été obtenus sur
1408 des exemples simples⁶)

1409 — 2 dimensions de le travail fourni :

1410 - La volonté de pousser un exemple simple jusqu’au bout de la
1411 chaîne pour obtenir des résultats et une évaluation sur au moins
1412 un exemple ; - La réalité du travail à fournir pour faire avancer sur
1413 la chaîne de traitement. ⇒ Une solution aurait été de considérer
1414 les arbres de parsing obtenus après le traitement du polyphonique
1415 comme un résultat local possible à évaluer au lieu d’attendre que la
1416 chaîne arrive jusqu’à la génération d’une partition mais cela n’était
1417 pas prioritaire pendant le stage.

1418 — Création d’un jeu de forme rythmique basique représentatif des
1419 différents styles à recouvrir. Ce jeu n’a pas pu être créé, car comme
1420 vu plus haut, je me suis focalisé sur un exemple pour pouvoir le
1421 vérifier entièrement et dans l’espoir de pouvoir le tester en fin de
1422 chaîne. **Évaluation** Matcher les motifs aurait été indispensable
1423 pour obtenir une quantité de résultats qui justifieraient une
1424 évaluation automatique permettant de faire des graphiques.

1425 L’évaluation fut entièrement manuelle car :

1426 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
1427 VS output) Pour l’évaluation, il aurait fallu produire un module.

5. fla ou accords entre autres...

6. exemple de 2 mesures, voir ...

1428 <dam>je ne sais pas si tu auras encore le temps de faire ça, sinon
1429 il faudra décrire comment tu aurais aimé évaluer, proprement et
1430 sans résultats chiffrés</dam> L'évaluation est-elle automatique ou
1431 manuelle ?
1432 Possibilité d'un export lilypond en arbre pour comparer l'output
1433 avec la transcription manuelle.
1434 Possibilité de transformer lilypond(output) et lilypond(ref) en
1435 ScoreModel ou MEI pour les comparer et faire des statistiques.
1436 Si transformés en MEI : diffscore de Francesco. Possibilité de
1437 transformer lilypond(output) et lilypond(ref) en MusicXML pour
1438 les comparer ou dans Music21. L'expérimentation peut-être consi-
1439 dérer comme une évaluation manuelle ? (magicien d'Oz)
1440 Lilypond vers MIDI + output vers MIDI \Rightarrow Comparaison des MIDI
1441 dumpés.
1442
1443 La transcription automatique de la batterie est un sujet passionnant mais
1444 difficile : Obtenir la totalité des éléments nécessaires pour le mémoire né-
1445 cessiterait plus de temps. Une base solide spécifique à la batterie a néan-
1446 moins été générée. Elle sera un bon point de départ pour les travaux fu-
1447 turs dont plusieurs propositions sont énoncés dans le présent document.

CONCLUSION GÉNÉRALE

1449 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1450 tion automatique de la batterie. Son objectif était de transcrire, à partir
1451 de leur représentation symbolique MIDI, des performances de batteur de
1452 différents niveaux et dans différents styles en partitions écrites.

1453 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1454 cessus de regroupement des événements MIDI qui nous a permis de faire
1455 la transition du monophonique vers le polyphonique. Une des données
1456 importante de ce processus était de différencier les nature des notes d'un
1457 accord, notamment de distinguer lorsque 2 notes constituent un accord
1458 ou un *fla*.

1459 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1460 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1461 tionnables en amont du parsing, soit par indication des noms des fichiers
1462 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1463 naire de patterns prédéfinis ⁷ qu'il serait pertinent de mettre en œuvre en
1464 machine learning.

1465 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1466 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1467 d'un système et pour la simplification grâce aux gammes du motif d'un
1468 système. Nous avons aussi montré comment, dans des travaux futurs, un
1469 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1470 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1471 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1472 la chaîne de traitement en terme de réécriture.

1473 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1474 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1475 entièrement automatisée de partitions de batterie et au problème général
1476 de l'AMT de bout en bout.[11]

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

BIBLIOGRAPHIE

- 1478 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1479 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1480 1996. – Cité pages 7, 18 et 35.
- 1481 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1482 et 15.
- 1483 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1484 page 12.
- 1485 [4] Gaël Richard et al. De fourier à la reconnaissance
1486 musicale. Available at [https://interstices.info/
1487 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1488 – Cité page 12.
- 1489 [5] Caroline Traube. Quelle place pour la science au sein de la musico-
1490 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1491 [6] Leonard Bernstein Office. The unanswered question : Six talks at
1492 harvard. Available at [https://leonardbernstein.com/about/
1493 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1494 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1495 entre les traitements de la musique et du langage. *La Lettre des
1496 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1497 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1498 Techniques de traitement automatique du langage naturel appli-
1499 quées aux représentations symboliques musicales. In *JIM 2021 -
1500 Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1501 Cité page 13.
- 1502 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-
1503 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1504 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1505 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1506 Cité page 13.
- 1507 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1508 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1509 lenges and future directions. *Journal of Intelligent Information Sys-*
1510 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 63.
- 1511 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1512 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1513 Cité pages 14 et 27.
- 1514 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-
1515 SURE, 1997. – Cité page 15.
- 1516 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1517 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1518 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1519 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1520 1483, 2018. – Cité pages 15, 23 et 27.
- 1521 [15] Moshekwa Malatji. Automatic music transcription for two instru-
1522 ments based variable q-transform and deep learning methods, 10
1523 2020. – Cité page 22.
- 1524 [16] Antti J. Eronen. Musical instrument recognition using ica-based
1525 transform of features and discriminatively trained hmms. *Seventh*
1526 *International Symposium on Signal Processing and Its Applications*,
1527 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1528 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1529 drum sound description for real-world music using template adap-
1530 tation and matching methods. *International Conference on Music*
1531 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1532 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1533 musical statistics as guides for audio-to-score piano transcription.
1534 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24 et 26.
- 1535 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1536 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1537 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1538 *Computation in Music*, volume Lecture Notes in Computer Science
1539 of *Proceedings of the Seventh International Conference on Mathema-*
1540 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1541 2019. Springer. – Cité pages 24 et 26.
- 1542 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1543 ring of rhythm structures. In *Proceedings of the First International*
1544 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1545 IEEE Computer Society. – Cité page 26.
- 1546 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1547 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1548 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1549 Cité page 26.
- 1550 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1551 equivalent rhythmic notations based on rhythm tree languages. In
1552 *Third International Conference on Technologies for Music Notation
1553 and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1554 pez Palma and Mike Solomon. – Cité page 26.
- 1555 [23] R. Marxer and J. Janer. Study of regularizations and constraints in
1556 nmf-based drums monaural separation. In *International Conference
1557 on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1558 02/09/2013 2013. – Cité page 27.
- 1559 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1560 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 29.
- 1561 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1562 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 29.
- 1563 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1564 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1565 94120 Fontenay-sous-Bois, 2000. – Cité pages 31 et 43.
- 1566 [27] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1567 Bamman. Learning to groove with inverse sequence transforma-
1568 tions. In *International Conference on Machine Learning (ICML)*,
1569 2019. – Cité page 48.

