

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

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

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

8 *Parcours :*  
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

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

14 *Encadrant :*  
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021



# TABLE DES MATIÈRES

18	<b>Liste des figures</b>	<b>4</b>
19	<b>Liste des tableaux</b>	<b>5</b>
20	<b>Introduction générale</b>	<b>7</b>
21	<b>1 Contexte</b>	<b>11</b>
22	1.1 Langues naturelles et musique en informatique . . . . .	12
23	1.2 La transcription automatique de la musique . . . . .	13
24	1.3 La transcription automatique de la batterie . . . . .	15
25	1.4 Les représentations de la musique . . . . .	16
26	<b>2 État de l'art</b>	<b>21</b>
27	2.1 Monophonique et polyphonique . . . . .	21
28	2.2 Audio vers MIDI . . . . .	22
29	2.3 MIDI vers partition . . . . .	23
30	2.4 Approche linéaire et approche hiérarchique . . . . .	23
31	<b>3 Méthodes</b>	<b>27</b>
32	3.1 La notation de la batterie . . . . .	27
33	3.2 Modélisation pour la transcription . . . . .	34
34	3.3 Qparse . . . . .	35
35	3.4 Les systèmes . . . . .	37
36	<b>4 Expérimentations</b>	<b>45</b>
37	4.1 Le jeu de données . . . . .	45
38	4.2 Analyse MIDI-Audio . . . . .	47
39	4.3 Expérimentation théorique d'un système . . . . .	51
40	4.4 Résultats et discussion . . . . .	55
41	<b>Conclusion générale</b>	<b>61</b>
42	<b>Bibliographie</b>	<b>63</b>

## LISTE DES FIGURES

44	1.1	Transcription automatique . . . . .	15
45	1.2	Exemple évènements avec durée . . . . .	17
46	1.3	Critère pour un évènement . . . . .	17
47	1.4	Exemple évènements sans durée . . . . .	18
48	1.5	Exemple de partition de piano . . . . .	18
49	1.6	MusicXML . . . . .	19
50	2.1	HMM . . . . .	24
51	2.2	arbre_jazz . . . . .	25
52	3.1	. . . . .	27
53	3.2	Rapport des figures de notes . . . . .	28
54	3.3	Hauteur et têtes de notes . . . . .	29
55	3.4	Point et liaison . . . . .	30
56	3.5	Les silences . . . . .	31
57	3.6	Silence joué . . . . .	31
58	3.7	Équivalence . . . . .	32
59	3.8	Séparation des voix . . . . .	33
60	3.9	Les accents et les ghost-notes . . . . .	33
61	3.10	Exemple pour les accentuations et les ghost-notes . . . . .	33
62	3.11	Présentation de Qparse . . . . .	36
63	3.12	Métrique . . . . .	38
64	3.13	Motif 4-4 binaire . . . . .	39
65	3.14	Motif 4-4 jazz . . . . .	40
66	3.15	Système 4-4 afro-latin . . . . .	40
67	3.16	Simplification . . . . .	41
68	3.17	. . . . .	42
69	4.1	Batterie électronique . . . . .	46
70	4.2	Partition de référence . . . . .	50
71	4.3	Motifs et gammes . . . . .	51
72	4.4	Partition d'un système en 4/4 binaire . . . . .	52
73	4.5	Arbre de rythme — système . . . . .	52
74	4.6	Arbre de rythme — voix haute . . . . .	53
75	4.7	Arbre de rythme — voix basse . . . . .	53
76	4.8	. . . . .	54
77	4.9	. . . . .	54
78	4.10	. . . . .	54
79	4.11	. . . . .	55

80	4.12 . . . . .	55
----	----------------	----

81

## LISTE DES TABLEAUX

82	1.1 speechToText vs AMT . . . . .	13
83	3.1 Pitches et instruments . . . . .	34
84	3.2 Systèmes . . . . .	38



## INTRODUCTION GÉNÉRALE

### 86 QUOI?

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

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

108

### 109 POURQUOI?

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

116

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

### 119 COMMENT?

120 → Problématique :

121 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.

131

132 → Méthodes :

133 → Contributions :

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

137 La proposition principale de ce mémoire est basée sur la recherche de  
138 rythmes génériques sur l'*input*. Ces rythmes sont des *patterns* standards  
139 de batterie définis au préalable et accompagnés par les différentes combi-  
140 naisons qui leur sont propres. On les nomme systèmes (voir sections 3.4,  
141 4.3). L'objectif des systèmes est de fixer des choix le plus tôt possible afin  
142 de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces  
143 choix concernent notamment la métrique et les règles de réécriture.

144

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

- 146 — une modélisation de la notation de la batterie (fusion de 3.1 et de  
147 3.2) qui était jusqu'à présent inexistante.
- 148 — plusieurs transcriptions manuelles dans le but d'analyser les conte-  
149 nus des fichiers MIDI et Audio (4.2) et de faire des comparaisons  
150 de transcription avec des outils déjà existants<sup>1</sup>.
- 151 — une partition de référence transcrite manuellement sur l'entièreté  
152 d'une performance du jeu de données afin de repérer les éléments  
153 importants pour la modélisation et de faire les liens entre les cri-  
154 tères des données d'*input* avec l'écriture finale (4.2). Cette partition  
155 avait aussi pour objectif d'effectuer des tests et des évaluations.
- 156 — le passage au polyphonique en théorie et en implémentation im-  
157 pliquant la théorie sur la détection de l'identité de notes dans un  
158 Jam<sup>2</sup> et l'implémentation de tests unitaires sur le traitement des  
159 Jams (4.4).
- 160 — la création de grammaires pondérées spécifiques à la batterie (4.4)

161

---

1. MuseScore3

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



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

1) Obtenir des arbres de rythmes corrects en *output* de *qparse* avec des exemples courts proches de la partition de référence.

2) La création d'une expérimentation théorique d'un système 4.3 dont le but premier est de démontrer qu'elle est implémentable et applicable à d'autres type de rythmes et dont le second objectif est de donner une méthode de création d'un système à partir d'une partition.

Ces deux réalisations recouvrent une partie du chemin à parcourir puisque pour effectuer des évaluations conséquentes sur résultat, la chaîne de traitement doit être finie afin de pouvoir vérifier de manière empirique que les systèmes, qui constituent ma contribution principale pour ce mémoire, ont permis d'améliorer *qparse* pour la transcription automatique de la batterie.

## PLAN

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



## CONTEXTE

## Sommaire

1.1	Langues naturelles et musique en informatique . . . . .	12
1.2	La transcription automatique de la musique . . . . .	13
1.3	La transcription automatique de la batterie . . . . .	15
1.4	Les représentations de la musique . . . . .	16

## Introduction

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

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

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

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

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

## 213 1.1 Langues naturelles et musique en 214 informatique

### 215 COMPUTER MUSIC

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

### 225 RIM

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

### 236 RIM et TAL

237 Aborder la musique comme un langage avec des méthodes de TAL nécessite une réflexion autour de la musique en tant que langage ainsi que  
238 la possibilité de comparer ce même langage avec les langues naturelles. Quelques travaux en neurosciences ont abordé ces questions, notamment  
239 par observation des processus cognitifs et neuronaux que les systèmes de traitement de ces deux productions humaines avaient en commun. Dans  
240 le travail de Poulin-Charronnat *et al.* [6], 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  
241 soulevée à l'occasion de projets de recherche en TAL. Keller *et al.* [7] ont

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/>

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 [8]. 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.* [9] 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 est la notation d'une œuvre musicale initialement non écrite.

exemples / illustration de la proximité thématique?

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

La tableau 1.1 montre des différences et des similitudes entre les deux domaines. objectifs similaires sur le papier : speech-to-text, problèmes et applications aussi comparables : transcription, synthèse, séparation de sources... Mais information de nature différente cf. sous-tâches comme beat tracking et inférence de tempo en musique.

## 1.2 La transcription automatique de la musique

En musique, la transcription<sup>8</sup> est la pratique consistant à noter un morceau ou un son qui n'était auparavant pas noté et/ou pas populaire en tant que musique écrite, par exemple, une improvisation de jazz ou une bande sonore de jeu vidéo. Lorsqu'un musicien est chargé de créer une partition à partir d'un enregistrement et qu'il écrit les notes qui composent le morceau en notation musicale, on dit qu'il a créé une transcription musicale de cet enregistrement.

il faut réorganiser cette partie : 1. objectif 2. applications 3. problèmes et méthodes scientifiques

pas très bien écrit. ne pas citer wikipedia mais article de survey

conversion d'une performance musicale en musique écrite, en général et notation occidentale

8. [https://en.wikipedia.org/wiki/Transcription\\_\(music\)](https://en.wikipedia.org/wiki/Transcription_(music))

L'objectif de la transcription automatique de la musique (AMT) [10] est de convertir la performance d'un musicien en notation musicale - un peu comme la conversion de la parole en texte dans le traitement du langage naturel. L'AMT a des intérêt multiples, notamment pour la transcription de solos ou encore pour la constitution de corpus musicologiques, ou encore pour l'interprétation de la musique et l'analyse du contenu musical [11]. Par exemple, un grand nombre de fichiers audio et vidéo musicaux sont disponibles sur le Web, et pour la plupart d'entre eux, il est difficile de trouver les partitions musicales correspondantes, qui sont nécessaires pour pratiquer la musique, faire des reprises et effectuer une analyse musicale détaillée. Les partitions de musique classique sont facilement accessibles et il y a peu de demandes de nouvelles transcriptions. D'un point de vue pratique, des demandes beaucoup plus commerciales et académiques sont attendues dans le domaine de la musique populaire [11]. Les modèles grammaticaux qui représentent la structure hiérarchique des séquences d'accords se sont avérés très utiles dans les analyses récentes de l'harmonie du jazz [12]. Comme déjà évoqué précédemment, il s'agit d'un problème ancien et difficile. C'est un « graal » de l'informatique 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. De même pour les chercheurs en audio James A. Moorer, Martin Piszczalski et Bernard Galler qui, en 1977<sup>9</sup>, ont utilisé leurs connaissances en ingénierie de l'audio et du numérique pour programmer un ordinateur afin de lui faire analyser un enregistrement musical numérique de manière à détecter les lignes mélodiques, les accords et les accents rythmiques des instruments à percussion.

La tâche de transcription automatique de la musique comprend deux activités distinctes : l'analyse d'un morceau de musique et l'impression d'une partition à partir de cette analyse.

La figure 1.1 est une proposition de Benetos *et al.* [10] qui représente l'architecture générale d'un système de transcription musicale. On y observe plusieurs sous-tâches de l'AMT :

- 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 :

9. [https://en.wikipedia.org/wiki/Transcription\\_\(music\)](https://en.wikipedia.org/wiki/Transcription_(music))

à l'instar de la

applications

préservation du patrimoine

e.g. musique de tradition orale (ethno-musicologie)

citer un survey pour les applications (pas [11])

pas d'accord avec ça. problème des partitions libres de droit.

l'intérêt est aussi d'avoir des partitions au contenu exploitable (texte ou XML) vs images (pdf...) cf. par ex. cette présentation d'OpenScore à FOSDEM <https://archive.fosdem.org/2017/schedule/event/openscore/> et mes transparents sur le sujet

là on passe aux approches scientifiques

quel rapport ?

la figure ne correspond pas à ton travail. ici "score" = MIDI performance. Tu peux lister les sous-tâches en section 2.2

- 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.
- Apprentissage sur des modèles acoustiques et musicologiques.
- *Optionnel* : Informations fournies de manière externe, soit fournie en amont (genre, instruments, ...), soit par interaction avec un utilisateur (infos sur une partition incomplète).

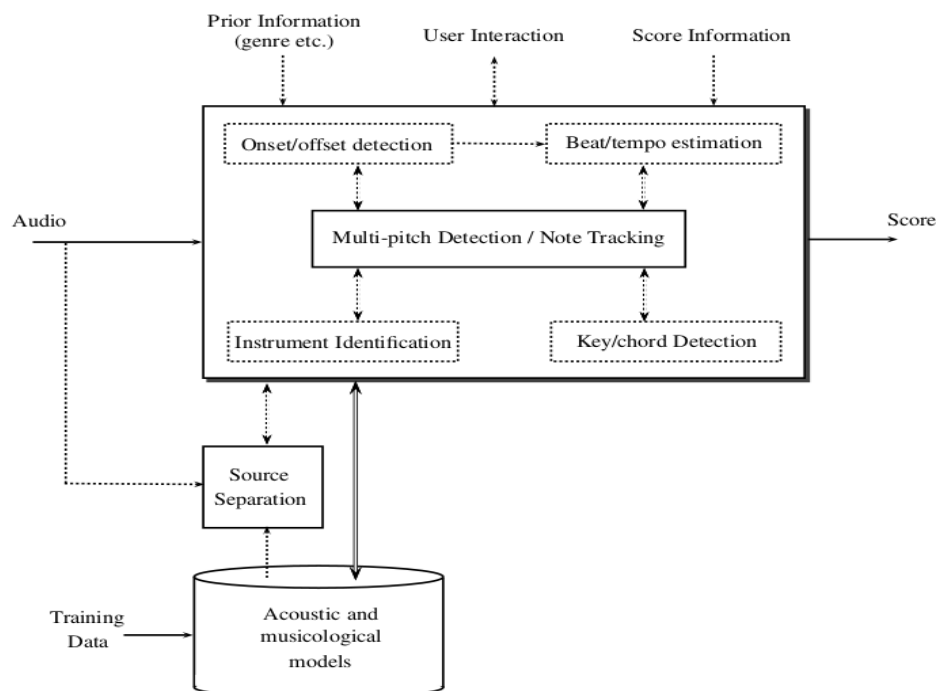


FIGURE 1.1 – Transcription automatique

*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.*

334

éviter newpage

### 1.3 La transcription automatique de la batterie

335

336

337

338

339

340

La batterie est 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

tres bonne section

341 du thème en jazz) afin d'improviser un accompagnement approprié que  
 342 personne ne pouvait écrire pour lui à sa place.

343 Les partitions de batterie sont arrivées par nécessité avec la pédagogie et  
 cite méthode et école Ago344  
 tini? 345 l'émergence d'écoles de batterie partout dans le monde. Un autre facteur  
 346 qui a contribué à l'expansion des partitions de batterie est l'émergence  
 347 de la musique assistée par ordinateur (MAO). En effet, l'usage de boîtes  
 348 à rythmes<sup>10</sup> ou de séquenceurs<sup>11</sup> permettant d'expérimenter soi-même  
 349 l'écriture de rythmes en les écoutant mixés avec d'autres instruments sur  
 350 des machines a permis aux compositeurs de s'émanciper de la création  
 351 d'un batteur en lui fournissant une partition contenant les parties exactes  
 qu'ils voulaient entendre sur leur musique.

352 La batterie a un statut à part dans l'univers de l'AMT puisqu'il s'agit  
 353 d'instruments sans hauteur (du point de vue harmonique), d'événements  
 354 sonores auxquels une durée est rarement attribuée et de notations spéci-  
 355 fiques (symboles des têtes de notes).

356 Les applications de l'ADT seraient utiles, non seulement dans tous les  
 ADT pas défini 357 domaines musicaux contenant de la batterie dont certains manquent de  
 358 partitions, notamment les musiques d'improvisation (jazz, pop) [10], mais  
 "contenant" -> concernés 359 aussi de manière plus générale dans le domaine du MIR : si les ordina-  
 par 360 teurs étaient capables d'analyser la partie de la batterie dans la musique  
 permettrait de faciliter 361 enregistrée, cela permettrait une variété de tâches de traitement de la  
 362 musique liées au rythme. En particulier, la détection et la classification  
 363 des événements sonores de la batterie par des méthodes informatiques  
 364 est considérée comme un problème de recherche important et stimulant  
 365 dans le domaine plus large de la recherche d'informations musicales [13].  
 366 L'ADT est un sujet de recherche crucial pour la compréhension des aspects  
 367 rythmiques de la musique, et a un impact potentiel sur des domaines plus  
 368 larges tels que l'éducation musicale et la production musicale.

## 369 1.4 Les représentations de la musique

370

### 371 Les données audio

372

373 Le fichier WAV<sup>12</sup> est une instance du Resource Interchange File Format  
 374 (RIFF) défini par IBM et Microsoft. Le format RIFF agit comme une "en-  
 375 veloppe" pour divers formats de codage audio. Bien qu'un fichier WAV  
 376 puisse contenir de l'audio compressé, le format audio WAV le plus cou-  
 377 rant est l'audio non compressé au format LPCM (linear pulse-code modu-  
 lation). Le LPCM est également le format de codage audio standard des

10. Roland TR-808

11. SQ-1

12. <https://en.wikipedia.org/wiki/WAV>



379 CD audio, qui stockent des données audio LPCM à deux canaux échan-  
 380 tillonnées à 44 100 Hz avec 16 bits par échantillon. Comme le LPCM n'est  
 381 pas compressé et conserve tous les échantillons d'une piste audio, les uti-  
 382 lisateurs professionnels ou les experts en audio peuvent utiliser le format  
 383 WAV avec l'audio LPCM pour obtenir une qualité audio maximale.

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

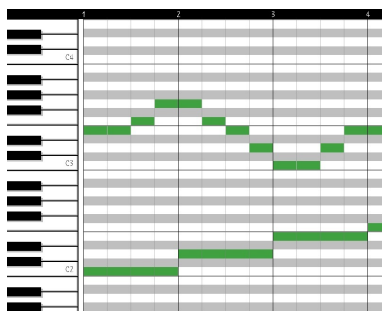
### 384 Les données MIDI

385 Le MIDI<sup>13</sup> (Musical Instrument Digital Interface) est une norme tech-  
 386 nique qui décrit un protocole de communication, une interface numérique  
 387 et des connecteurs électriques permettant de connecter une grande va-  
 388 riété d'instruments de musique électroniques, d'ordinateurs et d'appareils  
 389 audio connexes pour jouer, éditer et enregistrer de la musique.

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

392 Les données midi sont représentées sous forme de piano-roll. Chaque  
point sur la figure 1.2 est appelé « évènement MIDI » :

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



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

FIGURE 1.2 – Exemple évènements avec durée

393  
 394 Chaque évènement MIDI rassemble un ensemble d'informations sur la  
 hauteur, la durée, le volume, etc. . . :

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

FIGURE 1.3 – Critère pour un évènement

395

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

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

396 Pour la batterie, les évènements sont considérés sans durée, nous ignore-  
 397 rons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le  
 398 *channel* ne nous sera pas utile non plus.  
 399 *Ici, définir Tick et channel.*  
 Voici un exemple de piano-roll midi pour la batterie :

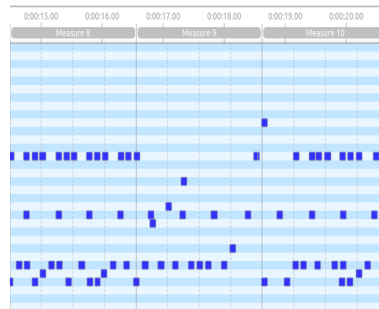


FIGURE 1.4 – Exemple évènements sans durée

400  
 401 On observe que toutes les durées sont identiques.

## 402 Les partitions



FIGURE 1.5 – Exemple de partition de piano

403 Une partition de musique<sup>14</sup> est un document qui porte la représentation  
 404 systématique du langage musical sous forme écrite. Cette représentation  
 405 est appelée transcription et elle sert à traduire les quatre caractéristiques  
 406 du son musical :  
 407 — la hauteur ;  
 408 — la durée ;  
 409 — l'intensité ;  
 — le timbre.

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

412 Ainsi que de leurs combinaisons appelées à former l'ossature de l'œuvre  
 413 musicale dans son déroulement temporel, à la fois :  
 414 — diachronique (succession des instants, ce qui constitue en musique  
 415 la mélodie) ;

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

— et synchronique (simultanéité des sons, c'est-à-dire l'harmonie).

## Le format MusicXML

MusicXML est un format de fichier basé sur XML pour représenter la notation musicale occidentale. Ce format est ouvert, entièrement documenté et peut être utilisé librement dans le cadre de l'accord de spécification finale de la communauté du W3C.

Un des avantages de ce format est qu'il peut être converti aussi bien en données MIDI qu'en partition musicale, ce qui en fait une interface homme/machine.

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

existe plusieurs formats XML : MusicXML, MEI, MNX, qui sont autant de schemas XML

standard W3C = MNX (en cours)

```
<?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 1.6 – MusicXML

Le figure 1.6<sup>15</sup> représente un do en clef de sol de la durée d'une ronde sur une mesure en 4/4.

inconvenient : format.s verbeux et ambigus. -> on utilise pour la transcription une représentation intermédiaire abstraite décrite plus loin.

## Conclusion

Dans ce chapitre, nous avons établi que le MIR 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 le MIR est né de l'AMT qui est un problème ancien et très difficile et qu'il serait toujours très utile de le ré-

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

436 soudre (autant pour l'AMT que pour l'ADT).  
437 Et enfin, nous avons décrit les représentations de la musique nécessaires  
438 à la compréhension du présent mémoire, allant du son jusqu'à l'écriture.

439

440

## ÉTAT DE L'ART

441

### Sommaire

442	2.1	Monophonique et polyphonique . . . . .	21
443	2.2	Audio vers MIDI . . . . .	22
444	2.3	MIDI vers partition . . . . .	23
445	2.4	Approche linéaire et approche hiérarchique . . . . .	23
446			
447			
448			

450

### Introduction

451 Dans ce chapitre, nous observerons les différentes avancées qui ont déjà  
 452 eu lieu dans le domaine de la transcription automatique de la musique et  
 453 de la batterie afin de situer notre démarche.

présenterons quelques tra-  
vaux antérieurs

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

459

### 2.1 Monophonique et polyphonique

460 Les premiers travaux en transcription ont été faits sur l'identification des  
 461 instruments monophoniques<sup>1</sup> [10]. Actuellement, le problème de l'esti-  
 462 mation automatique de la hauteur des signaux monophoniques peut être  
 463 considéré comme résolu, mais dans la plupart des contextes musicaux, les  
 464 instruments sont polyphoniques<sup>2</sup>. L'estimation des hauteurs multiples

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

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

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

## 2.2 Audio vers MIDI

Jusqu'à aujourd'hui, les recherches se sont majoritairement concentrées sur le traitement du signal vers la génération du MIDI [14].

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...

En ADT [13], plusieurs stratégies de répartition pré/post-processing sont possibles pour la détection multi-pitches. Entamer la détection dès le pré-processing, en supprimant les features non-pertinentes pendant la séparation des sources afin d'obtenir une meilleure détection des instruments de la batterie, est une démarche intuitive : supprimer la structure harmonique pour atténuer l'influence des instruments à hauteurs sur la détection grosse-caisse et caisse-claire en est un exemple. Mais certaines études montrent que des expériences similaires ont donné des résultats non-concluants et que la suppression des instruments à hauteurs peut avoir des effets néfastes sur les performances de l'ADT. En outre, les systèmes d'ADT basés sur des réseaux de neurones récurrents (RNN) ou sur des factorisations matricielles non négative font la séparation des sources pendant l'optimisation, ce qui réduit la nécessité de la faire pendant le pré-processing.

Pour la reconnaissance des instruments, une approche possible [15] est de mettre un modèle probabiliste dans l'étape de la classification des événements afin de classer les différents sons de la batterie. Cette méthode permet de se passer de samples audio isolés en modélisant la progression temporelle des *features*<sup>3</sup> avec un modèle de markow caché (HMM). Les

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

de signaux audio

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 le 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 2 mots sur les techniques utilisées (cf. survey AMT Benetos et al.)

haute fréquence, aigus ?

classification des événements ? la phrase semble redondante

505 *features* sont transformés en représentations statistiques indépendantes.  
 506 L'approche AdaMa [16] est une autre approche de la même catégorie ; elle  
 507 commence par une estimation initiale des sons de la batterie qui sont ité-  
 508 rativement raffinés pour correspondre à (pour matcher) l'enregistrement  
 509 visé.

pas clair... peut-être juste  
mentionner les modèles  
probabilistes utilisés

## 510 2.3 MIDI vers partition

511  
 512 Le plus souvent, lorsque les articles abordent la transcription automa-  
 513 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »  
 514 (partition) désigne un ouput au format Music XML, ou simplement MIDI.  
 515 Par exemple, dans [11], la chaîne de traitement va jusqu'à la génération  
 516 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en  
 517 extraire manuellement un fichier MusicXML contenant plusieurs voix.  
 518 Seuls quelques travaux récents s'intéressent de près à la création d'outils  
 519 permettant la génération de partition. Le problème de la conversion d'une  
 520 séquence d'événements musicaux symboliques en une partition musicale  
 521 structurée est traité notamment dans [17]. Ce travail, qui vise à résoudre  
 522 en une fois la quantification rythmique et la production de partition struc-  
 523 turée, s'appuie tout au long du processus sur des grammaires génératives  
 524 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-  
 525 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été  
 526 menées avec un ensemble de données composé d'extraits monophoniques ;  
 527 il reste donc à traiter le passage au polyphonique, en couplant le problème  
 528 de la séparation des voix avec la quantification du rythme.  
 529 L'approche de [17] est fondée sur la conviction que la complexité de la  
 530 structure musicale dépasse les modèles linéaires.

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

de manière conjointe

langage *a priori*

qui nécessite de traiter le  
problème supplémentaire  
de la séparation de voix. i.e.  
pour la batterie on n'vut  
quantification + structu-  
ration + séparation mais  
seules les 2 premières sont  
couplées dans l'approche de  
tonn stage.

## 531 2.4 Approche linéaire et approche hiérarchique

532 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par  
 533 exemple, Shibata *et al.* [11] ont utilisé le modèle de Markov caché (HMM)<sup>4</sup>  
 534 pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux  
 535 réseaux de neurones profonds, l'un pour la reconnaissance des *pitch*s et  
 536 l'autre pour la reconnaissance de la *vélocité*. Pour la dernière couche, la  
 537 probabilité est obtenue par une fonction sigmoïde. Ils construisent en-  
 538 suite plusieurs HMM métriques étendus pour la musique polyphonique  
 539 correspondant à des métriques possibles, puis ils calculent la probabilité  
 540 maximale pour chaque modèle afin d'obtenir la métrique la plus probable.

4. [https://fr.wikipedia.org/wiki/Modèle\\_de\\_Markov\\_caché](https://fr.wikipedia.org/wiki/Modèle_de_Markov_caché)  
[https://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](https://en.wikipedia.org/wiki/Hidden_Markov_model)

- 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

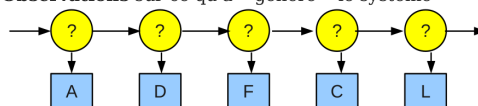


FIGURE 2.1 – HMM

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

542

543

544 L'évaluation finale des résultats de [11] montre qu'il faut rediriger l'atten-  
 545 tion vers les valeurs des notes, la séparation des voix et d'autres éléments  
 546 délicats de la partition musicale qui sont significatifs pour l'exécution de  
 547 la musique. Or, même si la quantification du rythme se fait le plus souvent  
 548 par la manipulation de données linéaires allant notamment des *real time*  
 549 *units* (secondes) vers les musical *time units* (temps, métrique, ...), de nom-  
 550 breux travaux suggèrent d'utiliser une approche hiérarchique puisque le

je ne comprend pas bien 551  
 l'explication. le pb est plu- 552  
 tot vue locale (déduction de 553  
 la proba d'une durée à par- 554  
 tir de la durée précédente, 555  
 par ex. dans un HMM) vs 556  
 vue globale, dans une hié- 557  
 rarchie

RT? 556  
 557

techniques de réécriture 558  
 appliquée à la déduction 559  
 automatique, calcul symbo- 560  
 lique 561

le calcul d'équiv. 561  
 562

563

564

citer thèse de David Rizo 565  
 (Valencia) 566

567

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 [18]. Les RT y sont évoqués comme permettant une co-  
 hésion complète de la notation musicale traditionnelle avec des notations  
 plus complexes. Jacquemard *et al.* [19] 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 [20]. 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 [17]. Les modèles de grammaire qui y sont  
 exposés sont différents de modèles markoviens linéaires de précédents  
 travaux.

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



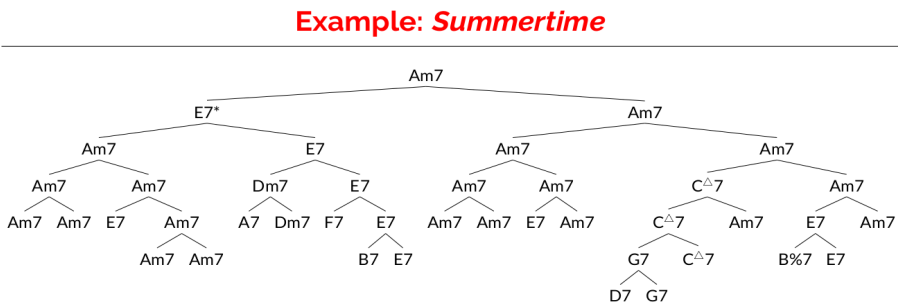


FIGURE 2.2 – arbre\_jazz  
Représentation arborescente d’une grille harmonique [12]

### Conclusion

La plupart des travaux déjà existants sur l’ADT ont été énumérés par Wu *et al.* [13] qui, pour mieux comprendre la pratique des systèmes d’ADT, se concentrent sur les méthodes basées sur la factorisation matricielle non négative et celles utilisant des réseaux neuronaux récurrents. La majorité de ces recherches se concentre sur des méthodes de calcul pour la détection d’événements sonores de batterie à partir de signaux acoustiques ou sur la séparation entre les évènements sonores de batterie avec ceux des autres instruments dans un orchestre ou un groupe de musique [21], ainsi que sur l’extraction de caractéristiques de bas niveau telles que la classe d’instrument et le moment de l’apparition du son. Très peu d’entre eux ont abordé la tâche de générer des partitions de batterie et, même quand le sujet est abordé, l’output final n’est souvent qu’un fichier MIDI ou MusicXML et non une partition écrite.

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

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

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

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

latex : enumerate



589

590

## MÉTHODES

591

## Sommaire

592

593

594

595

596

597

598

599

3.1	La notation de la batterie . . . . .	27
3.2	Modélisation pour la transcription . . . . .	34
3.3	Qparse . . . . .	35
3.4	Les systèmes . . . . .	37

600

## Introduction

601

602

603

604

605

606

607

608

Dans ce chapitre, nous expliquerons en détail les méthodes que nous avons employées pour l'ADT.

Pour commencer, nous exposerons une description de la notation de la batterie ainsi qu'une modélisation de celle-ci pour la représentation des données rythmiques en arbres syntaxiques. Nous poursuivrons avec une présentation de qparse<sup>1</sup>, un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et au sein du laboratoire Cedric au CNAM.

Enfin, nous présenterons les systèmes.

plusieurs développeurs

systèmes, une représentation théorique qui permet...

609

## 3.1 La notation de la batterie



FIGURE 3.1

610

611

612

La figure 3.1 montre 4 figures de notes les plus courantes dont les noms et les durées sont respectivement, de gauche à droite :

— La ronde, elle vaut 4 ;

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

4 temps

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

613 — La blanche, elle vaut 2 ;

614 — La noire, elle vaut 1 ;

615 — La croche, elle vaut 1/2.

plusieurs éléments

616 Une figure de note [1] de musique combine plusieurs critères <sup>2</sup> :

— Une tête de note :

Sa position sur la portée indique la hauteur de la note. La tête de note peut aussi indiquer une durée.

— Une hampe :

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

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

plutôt que wikipedia cite  
Dannhauser ou autre ref  
F.M. ou encore Gould 2011  
Behind Bars

barre verticale liée à la tête  
de note

haut ou bas

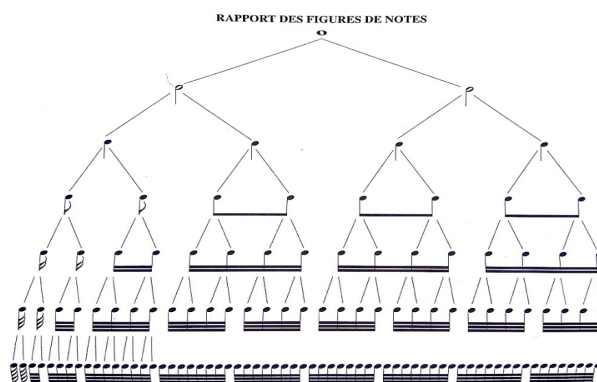


FIGURE 3.2 – Rapport des figures de notes  
[1]

626 La figure 3.2 montre les rapports de durée entre les figures de notes. Plus  
627 les durées sont longues, plus elles sont marquées par la tête de note (la  
628 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de  
629 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute  
630 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.  
631 Les notes à crochet (croche, double-croche, triple-croche...) peuvent être  
632 reliées ou non par des ligatures (voir les 4 dernières lignes de la figure  
633 3.2).

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

### Les hauteurs et les têtes de notes

635 Pour la transcription, nous proposons une notation inspirée du recueil de  
636 pièces pour batterie de J.-F. Juskowiak [22] et des méthodes de batterie  
637 Agostini [23], car nous trouvons la position des éléments cohérente et in-  
tuitive.

pour aider, tu pourrais dé-  
finir une figure représentant  
la batterie avec le nom des  
instruments et abbrévia-  
tion.

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

En effet, les hauteurs sur la portée représentent :

— 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.

— 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.

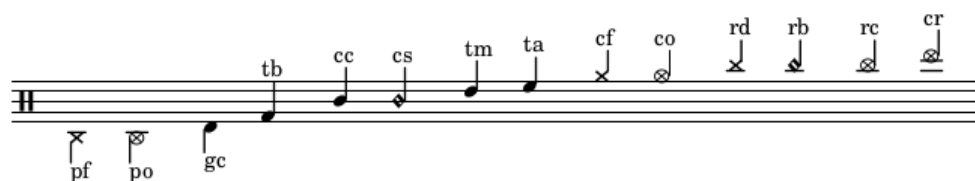


FIGURE 3.3 – Hauteur et têtes de notes

Les noms des instruments correspondant aux codes des notes de la figure 3.3 sont dans le tableau 3.1.

têtes de notes ?

## Les durées

Comme nous venons de la voir, la majorité des instruments de la batterie sont représentés par les têtes des notes. Par conséquent, les symboles rythmiques concernant la tête de note ne pourront pas être utilisés. 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 [24] 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.

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

- Le point ;
- La liaison.

Ces symboles ne seront utiles que pour l'écriture des ouvertures de charley. Le charley est le seul instrument de la batterie dont la durée est quan-

certaines têtes de notes vides alors que leur durée n'est pas celle des blanches ?

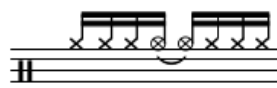
pour clarifier cela, tu pourrais décrire en 1.4 la notation conventionnelles (piano etc) et ici uniquement ce qui est spécifique à la batterie, en expliquant les différences.

expliquer comment, par ex. avec figure 3.4

672 tifiée (les cymbales attrapées à la main peuvent l'être aussi mais cela est  
673 très rare.)



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.4 – Point et liaison

= la position des temps 674

675

faire un "enumerate" 677

678

679

680

681

682

683

684

685

686

687

L'écriture de la batterie doit faire ressortir la pulsation. La première chose à prendre en compte pour analyser la figure 3.4 est donc la nécessité de regrouper les notes par temps à l'aide des ligatures.

Exemple 1 : ouverture de charley quantifiée mais pas notes 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 plus intuitif (regroupement par 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 la liaison.

688

## Les silences

expliquer la notation (générale) des silences en §1.4 ? 689

690

quantifier = noter ? ou quantifier la durée ? 691

692

693

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

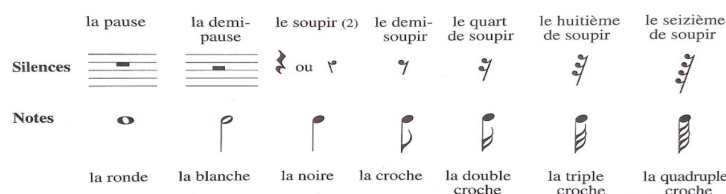


FIGURE 3.5 – Les silences

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

pas très clair

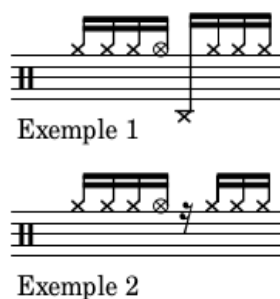


FIGURE 3.6 – Silence joué

698 L'exemple 1 de la figure 3.6 montre ce qui est écrit dans les données MIDI  
 699 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-  
 700 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1  
 701 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-  
 702 truments jouant simultanément.  
 703 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte  
 704 la note suivante pour l'écriture : - Si c'est un charley fermé joué à la main  
 705 ⇒ la note sera cf;  
 706 - Si c'est un charley fermé joué au pied ⇒ la note sera un silence.

itemize

cf?

## 707 Les équivalences rythmiques

708 Pour les instruments mélodiques, la liaison et le point sont les deux seules  
 709 possibilités en cas d'équivalence rythmique pour des notes dont la durée  
 710 de l'une à l'autre est ininterrompue. Mais pour la batterie, à part dans  
 711 le cas des ouvertures de charley (voir section 3.1), les durées des notes  
 712 n'ont pas d'importance. L'usage des silences pour combler la distance ryth-  
 713 mique entre deux notes devient donc possible.  
 714 Cela pris en compte, et étant donné que les indications de durée dans les  
 715 têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide

phrase alambiquée... pour prolonger la durée?

seuls comptent les date de début de notes onsets.

de silences sera privilégiée comme indication de durée sauf dans les cas où cela reste impossible. Ce choix à pour but de n'avoir qu'une manière d'écrire toutes les notes, que leurs têtes de notes soit modifiées ou non.



FIGURE 3.7 – Équivalence

Sur la figure 3.7, théoriquement, il faudra choisir la notation de la deuxième mesure mais dans certains contextes, pour des raisons de lisibilité ou de surcharge, la version sans les silences de la troisième mesure pourra être choisie.

### Les voix

Les voix<sup>3</sup> désignent les différentes parties mélodiques constituant une composition musicale et destinées à être interprétées, simultanément ou successivement, par un ou plusieurs musiciens. En batterie, une voix est l'ensemble des instruments qui, à eux seuls, constituent une phrase rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant possibles pour un même rythme, on peut regrouper les instruments de la batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou 2 voix. Sur la figure 3.8, 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 moins systématique de leurs phrasés, et des associations logiques entre les instruments dans la distribution des rythmes sur la batterie (voir la section 3.4).

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 pas orientation des hampes?

### Les accentuations et les ghost-notes

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

La figure 3.9 ne prend en compte que les accents que nous avons estimés nécessaires (voir la section 3.2). Les accents sont marqués par le symbole « > ». Il est positionné au-dessus des notes représentant des cymbales et en-dessous des notes représentant des toms ou la caisse-claire. Ce choix a été fait pour la partition de la figure 4.2 car elle est plus lisible

3.9 = liste des seuls "instruments" qui peuvent être accentués?

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



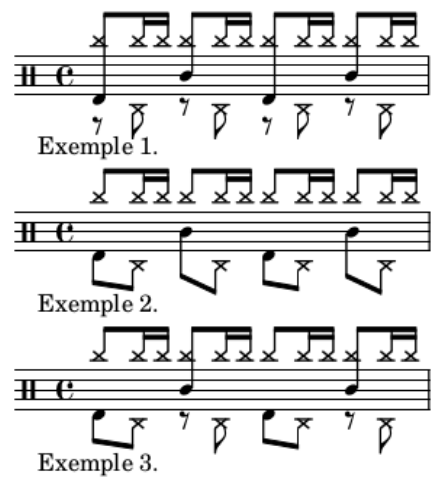


FIGURE 3.8 – Séparation des voix

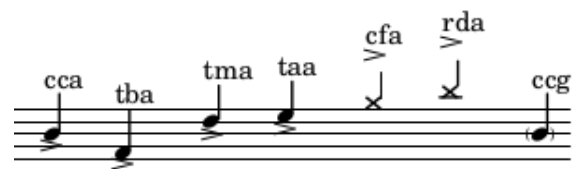


FIGURE 3.9 – Les accents et les ghost-notes

746 ainsi, mais ces choix devront être adaptés en fonction des différents sys-  
747 tèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz,  
748 les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il  
749 faudra donc mettre les symboles d'accentuation correspondants au-dessus  
750 des têtes de notes.  
751 La dernière note de la figure 3.9 montre un exemple de ghost-notes. Le  
752 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note  
753 sans changer la tête de note.  
754 Pour les codes, on prend le code de la note et on ajoute un « a » pour un  
755 accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.9  
sont exposées en situation réelle dans la figure 3.10.

expliquer ce qu'est une  
ghost-notes

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



FIGURE 3.10 – Exemple pour les accentuations et les ghost-notes

## 3.2 Modélisation pour la transcription

### Les pitches

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

TABLE 3.1 – Pitches et instruments

Il existe, pour de nombreux instruments de la batterie, plusieurs samples audio associés à des pitches. Pour cette première version, nous avons choisi de n'avoir qu'un code-instrument pour différentes variantes d'un instrument, c'est pourquoi certains code-instruments se voient attribuer plusieurs pitches dans le tableau 3.1.

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

### La vélocité

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

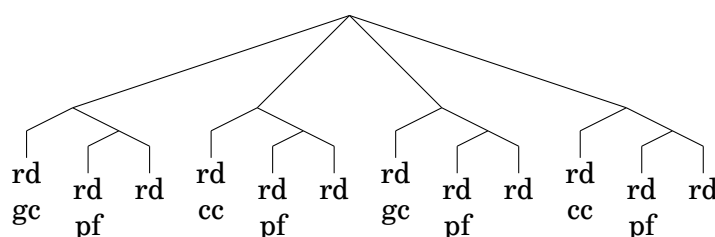
Cette transcription nous a mené aux observations suivantes :

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

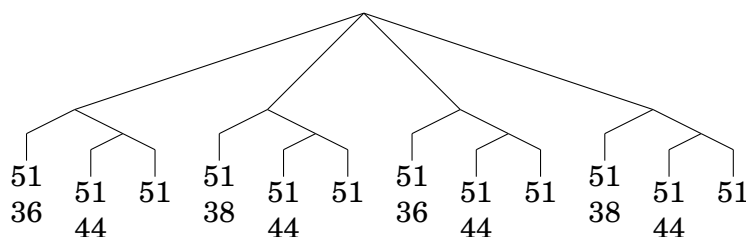
- Les accents et les ghosts-notes ne sont significatifs ni pour les instruments joués au pied, ni pour les cymbales crash.
- En effet, certaines vélocités en dessous de 40 étant détectées et inscrites dans les données MIDI sont dues au mouvement du talon du batteur qui bat la pulsation sans particulièrement jouer le charley. Ce mouvement est perçu par le capteur de la batterie électronique mais le charley n'est pas joué.
- Au final, 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 possibilités de notation sur une partition sont théoriquement multiples.
- Voici une représentation de la figure 3.8 en arbre de rythmes avec les codes de chaque instrument :



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



- Chacun des trois exemples de la figure 3.8 est représenté par un des deux arbres syntaxiques ci-dessus.

## 3.3 Qparse

- La librairie Qparse<sup>4</sup> implémente la quantification des rythmes basée sur des algorithmes d'analyse syntaxique pour les automates arbores-

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

non c'est juste une représentation du rythme, pas unique

expliquer le principe des RT : branchement = division d'intervalle temporel, feuilles = les événements musicaux commençant au début de l'intervalle). références : - Laurson "Patchwork : A Visual Programming Language", 1996. - OpenMusic : visual programming environment for music composition, analysis and research, 2011.

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

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

quantification rythmique + structuration de partition

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

Il effectue conjointement des tâches de quantification rythmique et d'inférence de la structure de la partition à l'aide de techniques de parsing / analyse

grammaire  $\neq$  automate. 804  
il faut choisir entre les 2  
(pour la suite aussi) 805

apprentissage 809

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

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.

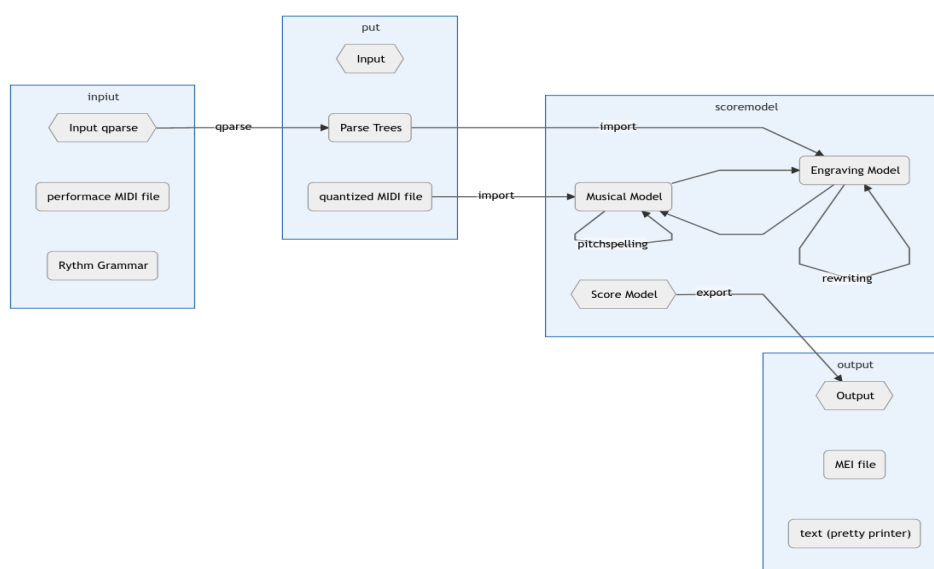


FIGURE 3.11 – Présentation de Qparse

la figure 3.11 est trop cor817  
pliquée. rhythm grammar  
→ automate d'arbres pon818  
déré. Parse Tree → arbre  
syntaxique. qtz MIDI file :  
inutile. Score Model → r820  
présentation intermédiaire  
de partition. Score Model821  
Engr. Model : inutile. gar-  
der juste la fleche Rewri-  
ting sur S.M. 822  
823

Explication des différentes étapes de la figure 3.11<sup>5</sup> :

- **Input Qparse** :  
Un fichier MIDI (séquence d'événements datés (piano roll) accompagné d'un fichier contenant une grammaire pondérée) ;
- **Arbre de parsing** :  
Les données MIDI sont quantifiées, les notes de dates proches sont

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

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

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

Plusieurs enjeux :

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

### 3.4 Les systèmes

Un système 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 système (gamme).

#### Définitions

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

Un ensemble de systèmes comprenant leur métrique et leurs règles spécifiques de réécriture sera nécessaire. Les systèmes devront être distribués

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

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

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

métrique n'est pas défini. règles de réécriture non plus

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

TABLE 3.2 – Sytèmes

866 dans 4 grandes catégories :

867 Nous exposerons 3 systèmes afin d'illustrer les propos de cette section :

868 — 4/4 binaire

869 — 4/4 jazz

870 — 4/4 afro-cubain

871 **Objectif des systèmes**

872 Les systèmes devront être matchés sur l'input MIDI afin de :

873 — définir une métrique ;

874 — choisir une grammaire appropriée ;

875 — fournir les règles de réécriture (séparation des voix et simplifica-

876 tion.

878 La partie *motif* des systèmes sera utilisée pour la **définition des mé-**

880 **triques**. Le *motif* et la gammes des systèmes seront utilisés pour la **sé-**

881 **paration des voix**. Les règles de **simplification** (les combinaisons de

réécritures) seront extraites des voix séparées des systèmes.

882 **Détection d'indication de mesure**

883 La détection de la métrique est importante, non seulement pour connaître

884 le nombre de temps par mesure ainsi que le nombre de subdivisions pour

885 chacun de ces temps, mais aussi pour savoir comment écrire l'unité de

temps et ses subdivisions.



Exemple 1



Exemple 2

FIGURE 3.12 – Métrique

bien. il faudrait expliquer ça avant.

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

métrique ≠ signature rythmique (c'est plus général). Il aurait fallu présenter simplement la notation des signatures rythmiques, par exemple en 1.4

La figure 3.12 montre deux indications de mesure différentes. L'une (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des triolets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est traditionnellement écrit en binaire avec ou sans triolet (même si cette musique est dite ternaire alors que le rock ternaire sera plutôt écrit comme dans l'exemple 2).

### Choix d'une grammaire

Il faut prendre en compte l'existence potentielle de plusieurs grammaires dédiées chacune à un type de contenu MIDI. Le choix d'une grammaire pondérée doit être fait avant le parsing puisque Qparse prend en entrée un fichier MIDI et un fichier wta (grammaire). C'est pour cette raison que la métrique 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 systèmes devront être recherchés sur l'input (*fichiers MIDI*) avant le lancement du parsing, afin de déterminer la métrique 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

### Séparation des voix

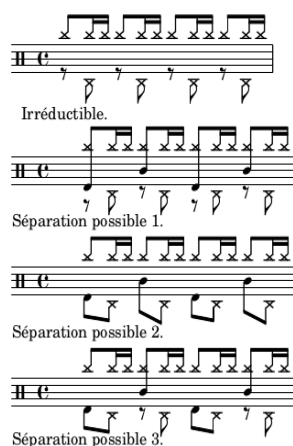


FIGURE 3.13 – Motif 4-4 binaire

Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système est constitué par défaut du motif rd/pf/cc (voir 3.1) et d'un texte joué à la grosse-caisse. La première ligne de la figure 3.13 est appelée « Irréductible

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

913 » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du  
 914 charley au pied. La troisième séparation proposée est privilégiée car elle  
 915 répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les  
 916 pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont  
 917 utilisés par voix et plus logique pour le lecteur puisque les mains sont en  
 haut et les pieds en bas.

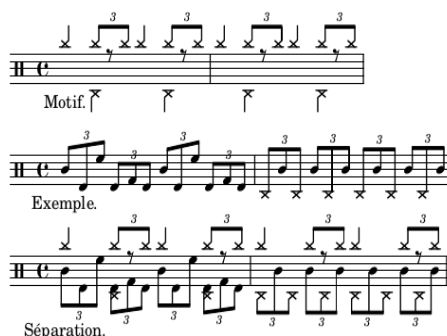


FIGURE 3.14 – Motif 4-4 jazz

918 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré  
 919 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :  
 920 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,  
 921 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et  
 922 2 combinées avec le *motif* de la première ligne, sont des cas typiques de  
 923 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,  
 924 la grosse caisse entre très souvent dans le flot des combinaisons de toms  
 925 et de caisse claire et son écriture séparée serait inutilement compliquée  
 926 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-  
 927 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de  
 928 charley en bas.  
 929

quel exemple?



FIGURE 3.15 – Système 4-4 afro-latin

930 La figure 3.15 montre un exemple minimaliste de système afro-latin [24].  
 931 Ce système doit être écrit sur trois voix car la voix centrale est souvent  
 932 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas  
 933 serait surchargé et peu lisible.



### Simplification de l'écriture

Les explications qui suivent seront appuyé par une expérimentation théorique dans la section 4.3.

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

Voici les différentes étapes à suivre :

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

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

Voici quelques règles inhérentes à la simplification de l'écriture pour la batterie : Toutes les continuations (t) qui se trouvent en début de temps (figures 4.9, 4.11 et 4.12) sont transformées en silences (r) sauf si la note précédente est un charley ouvert ?

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

expérimentation théorique??

ce sont des figures et notations du chapitre suivant!

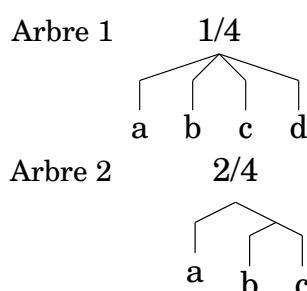


FIGURE 3.16 – Simplification

Soit l'arbre 1 de la figure 3.16 dans lequel : a et d sont des instruments de la batterie (x) ;

b et c sont des continuations (t) ;

Pour chacune des conditions suivantes, une suite de la figure 3.17 est attribuée :

- Si a n'est pas un co :

⇒ Suite 1a.

itemize

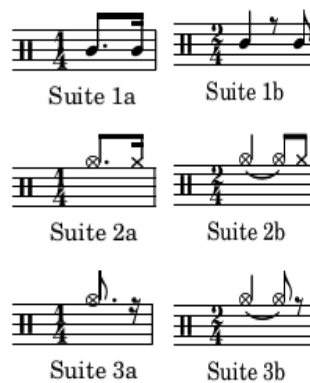


FIGURE 3.17

966 — Si a est un co :  
 967 — Si d est un cf :  
 968     ⇒ Suite 2a.  
 969 — Si d est un pf :  
 970     ⇒ Suite 3a : d deviens un silence (r).  
 971

972 Soit l'arbre 2 de la figure 3.16 dans lequel :  
 973 a et c sont des instruments de la batterie (x);  
 974 b est une continuation (t); Pour chacune des conditions suivantes, une  
 975 suite de la figure 3.17 est attribuée :

976 — Si a n'est pas un co :  
 977     ⇒ Suite 1b, b devient un silence.  
 978 — Si a est un co :  
 979     — Si c est un cf :  
 980         ⇒ Suite 2b, b devient une liaison et c devient un cf.  
 981     — Si c est un pf :  
 982         ⇒ Suite 3b : b deviens une liaison et c devient un silence.  
 983

984 *Rappel :*  
 985 *cf = charley fermé joué à la main ;*  
 986 *co = charley ouvert joué à la main ;*  
 987 *pf = charley fermé joué au pied.*  
 988

989 **Problème : le cf et le co ne seront jamais sur la même voix que le**  
 990 **pf... Par conséquent, les règles concernant les charleys ouverts**  
 991 **doivent-elles être appliquées sur l'arbre de parsing de l'input?...**

## 992 **Conclusion**

993 Nous avons formalisé une notation de la batterie, modélisé cette notation  
994 pour la transcription de données MIDI en partition, nous avons décrit  
995 Qparse.

996 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-  
997 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-  
998 propriée et énoncer des règles de séparation des voix et de simplification  
999 de l'écriture.



1000

1001

# EXPÉRIMENTATIONS

1002

## Sommaire

1003

1004

1005

1006

1007

1008

1009

1010

4.1	Le jeu de données . . . . .	45
4.2	Analyse MIDI-Audio . . . . .	47
4.3	Expérimentation théorique d'un système . . . . .	51
4.4	Résultats et discussion . . . . .	55

1011

## Introduction

1012

1013

1014

1015

1016

1017

1018

Dans ce chapitre, nous présenterons le jeu de données et les analyses audio-MIDI. Nous ferons ensuite l'expérimentation théorique d'un système implémentable qui devra être utilisé comme base de connaissances pour augmenter la rapidité et la qualité en sortie de Qparse. Nous présenterons ensuite les avancées réalisées dans ce travail et une réflexion sur les moyens de l'évaluer. Enfin, nous finirons par une discussion sur l'ensemble du travail réalisé.

1019

## 4.1 Le jeu de données

1020

1021

1022

1023

1024

1025

1026

1027

Nous avons utilisé le Groove MIDI Dataset<sup>1</sup> [25] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).

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

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



FIGURE 4.1 – Batterie électronique

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

1028 Autres critères spécifiques au GMD :

- 1029 — Toutes les performances ont été jouées au métronome et à un tempo
- 1030 choisi par le batteur.
- 1031 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 1032 qui ont pu improviser dans un large éventail de styles. Les don-
- 1033 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 1034 (professionnel ou amateur).
- 1035 — Les batteurs avaient pour instruction de jouer des séquences de
- 1036 plusieurs minutes ainsi que des fills<sup>2</sup>
- 1037 — Chaque performance est annotée d'un style (fourni par le batteur),
- 1038 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 1039 nyme du batteur.
- 1040 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 1041 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 1042 session du GMD.
- 1043 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 1044 fichier MIDI.

## 1045 **Format des données**

enregistre les données dans  
des fichiers MIDI

1046 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-  
1047 tinctes :

- 1048 — une pour le tempo et l'indication de mesure ;
- 1049 — une pour les changements de contrôle (position de la pédale de
- 1050 charley) ;
- 1051 — une pour les notes.

1052

1053 Les changements de contrôle sont placés sur le canal 0 et les notes sur le  
1054 canal 9 (qui est le canal canonique pour la batterie).

1055 Pour simplifier le traitement de ces données, ces trois pistes ont été  
1056 fusionnées en une seule piste qui a été mise sur le canal 9.

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.

1057  
 1058 « Control Changes The TD-11 also records control changes speci-  
 1059 fying the position of the hi-hat pedal on each hit. We have preserved this  
 1060 information under control 4. »  
 1061 (<https://magenta.tensorflow.org/datasets/groove>)  
 1062 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-  
 1063 tions dans les fichiers MIDI.  
 1064 L'utilisation de pretty\_midi devient urgente!

## 1065 4.2 Analyse MIDI-Audio

1066  
 1067 Ces analyses ont été faites dans le cadre de transcriptions manuelles à  
 1068 partir de fichiers MIDI et Audio du GMD.

= analyses et transcriptions  
manuelles

## 1069 Comparaisons de transcriptions

1070 Pour les comparaisons de transcriptions, les transcriptions manuelles  
 1071 (TM) ont été éditées à l'aide de Lilypond<sup>3</sup> ou MuseScore<sup>4</sup> et les trans-  
 1072 criptions automatiques (TA) ont toutes été générées manuellement avec  
 1073 MuseScore.

méthodologie tr. manuelle.  
import MIDI pour MuseS-  
core

## 1074 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1075 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1076 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1077 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1078 premières notes (une double-croche et une croche) d'un triolet sur
- 1079 le temps 1 de la mesure 2 de la TA.

## 1080 Exemple d'analyse 2

- 1081 — Les doubles croches ont été interprétées en quintolet
- 1082 — La deuxième double-croche est devenue une croche.

Transcription manuelle  $\Rightarrow$  Transcription automatique



Transcription manuelle  $\Rightarrow$  Transcription automatique



### 1083 Exemple d'analyse 3

- 1084 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
- 1085
- 1086 — 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.
- 1087
- 1088 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
- 1089
- 1090 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.
- 1091
- 1092

### 1093 Exemple d'analyse 4

Transcription manuelle  $\Rightarrow$  Transcription automatique



1094

- 1095 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une manière excessivement complexe!

conclusion sur ces exemples

### 1097 Exemple avec des flas

sauf erreur, les "flas" ne sont pas définis.  $\rightarrow$  sections 1.4 (appogiatures) et 3.1 (flas)?

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



## 1099 Transcription manuelle



1100

## 1101 Transcription automatique

1102



1103

1104

- 1105 — Le premier fla est reconnu comme étant un triolet contenant une
- 1106 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1107 ornementée.
- 1108 — Le deuxième fla est reconnu comme étant un accord.
- 1109 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1110 dans la TA.
- 1111 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1112 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1113 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1114 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1115 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1116 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1117 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1118 temps 3 de la TA.
- 1119 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1120 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1121 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1122 cross-sticks.

1123 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1124 La figure 4.2 est la transcription manuelle des fichiers *004\_jazz-*  
 1125 *funk\_116\_beat\_4-4.mid* et *004\_jazz-funk\_116\_beat\_4-4.wav* du GMD.

1126 Cette transcription a été entièrement faite avec Lilypond (voir le code  
 1127 lilypond sur le git [https://github.com/MartinDigard/Stage\\_M2\\_](https://github.com/MartinDigard/Stage_M2_Inria)  
 1128 *Inria*) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est  
 1129 annoncé dans le GMD de style «jazz-funk» probablement en raison de  
 1130 la ride de type shabada rapide (le ternaire devient binaire avec la vi-  
 1131 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux  
 1132 et quatre).

1133 La transcription des données audio et MIDI contenues dans ces fichiers  
 1134 a permis une analyse plus approfondie des critères à relever pour chaque  
 1135 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).

des conclusions sur la 1137  
 transcription manuelle?  
 difficultés, durée? nb de  
 passes... pourquoi LilyPon-  
 det pas MuseScore?

### 4.3 Expérimentation théorique d'un système

Cette expérimentation théorique, basée sur la partition de référence de la figure 4.2, montre le procédé de création d'un *système* et des règles qui en découlent (métrique, choix de grammaire, règles de séparation des voix et de simplification de l'écriture). Le *système* devra ensuite être implémenté pour appliquer des tests qui seront effectués, dans un premier temps, sur la partition de référence.

Le titre est contradictoire, et l'explication pas très claire

#### Motifs et gammes



FIGURE 4.3 – Motifs et gammes

#### Motifs

À partir de la partition de référence, les deux motifs de la figure 4.3 peuvent être systématisés. Le motif 1 est joué du début jusqu'à la mesure 18 avec des variations et des fills et le motif 2 est joué de la mesures 23 à la mesure 28 avec des variations. Ces deux motifs sont très classiques et pourront être détectés dans de nombreuses performances.

#### Gammes

Les gammes de la figure 4.3 étayent toutes les combinaisons d'un motif en 4/4 binaires jusqu'aux doubles croches.

Les lignes 1 et 2 traitent les croches. La ligne 1 a 2 mesures dont la première ne contient que des noires et la deuxième que des croches en l'air. Ces deux possibilités sont combinées de manière circulaire dans les 3 mesures de la deuxième ligne.

Les lignes 3, 4 et 5 traitent les doubles-croches. La ligne 3 a 2 mesures

1162 dont la première ne contient que des croches et la deuxième que des  
 1163 doubles-croches en l'air. Ces deux possibilités sont combinées de manière  
 1164 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

### 1165 Systèmes — motifs et gammes combinés

1166 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1  
 1167 de la figure 4.3.

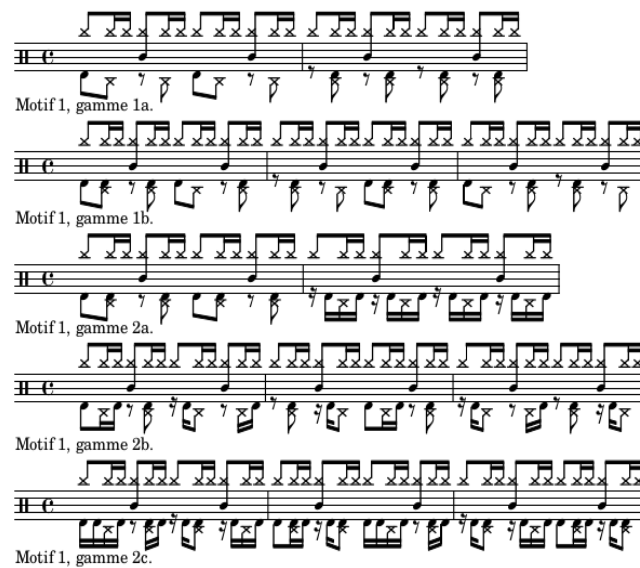


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1168

### 1169 Représentation du système en arbres de rythmes

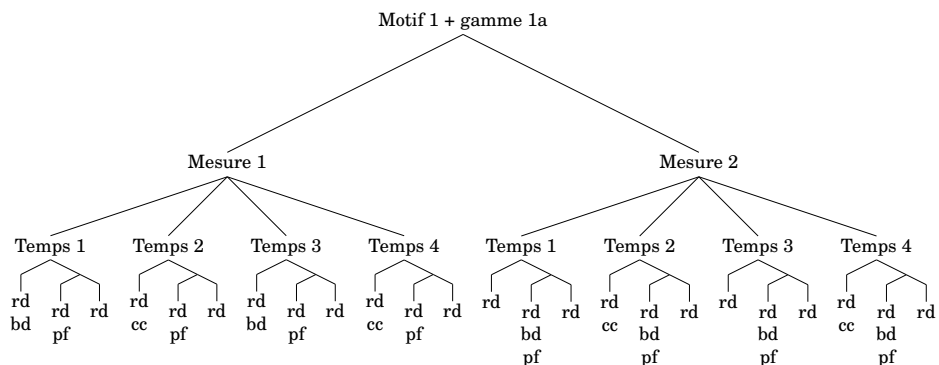


FIGURE 4.5 – Arbre de rythme — système

1170 L’arbre de la figure 4.5 servira de base pour le suite de l’expérimentation.  
1171 Comme indiqué à la racine de l’arbre, il représente la première ligne de la  
1172 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,  
1173 il manque des indications de notation telles que les voix spécifiques à  
1174 chaque partie du rythme ainsi que les choix d’écriture pour les distances  
1175 qui séparent les notes de chaque voix entre elles en termes de durée.

1176 **Réécriture — séparation des voix et simplification**

1177 **La séparation des voix**

1178 Ainsi l’arbre syntaxique de départ est divisé en autant d’instruments qui  
1179 le constituent et les voix seront regroupées en suivant les règles du sys-  
tème.

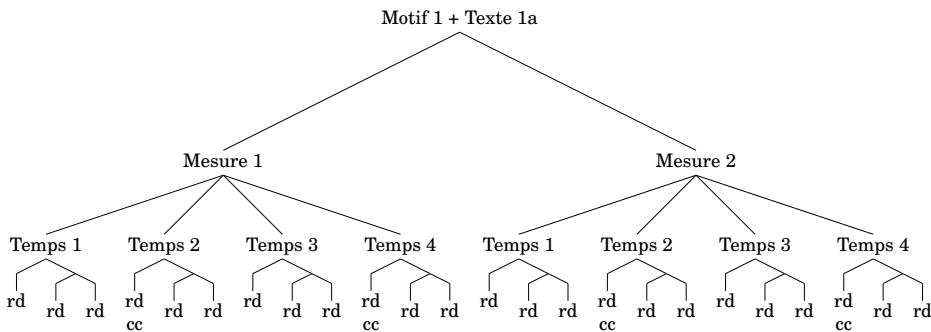


FIGURE 4.6 – Arbres de rythme — voix haute

1180  
1181 La voix haute regroupe la ride et la caisse-claire sur les ligatures du haut.

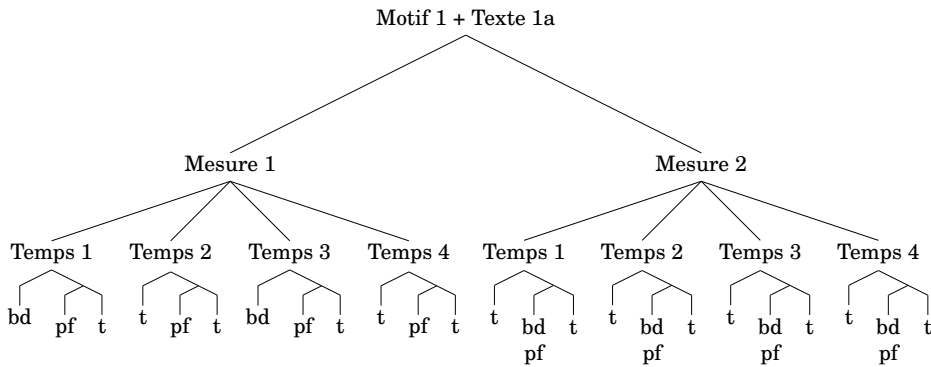


FIGURE 4.7 – Arbres de rythme — voix basse

1182  
1183 La voix basse regroupe la grosse-caisse et le charley au pied sur les liga-  
1184 tures du bas.

### 1185 Les règles de simplifications

1186 L'objectif des règles de simplifications est de réécrire les écarts de durées  
 1187 qui séparent les notes d'une manière appropriée pour la batterie et qui  
 1188 soit la plus simple possible. Les ligatures relient les notes d'un temps  
 1189 entre elles (rendre la pulse visuelle).

1190

1191 Pour les figures ci-dessous :

1192 — x = une note ;

1193 — r = un silence ;

1194 — t = une continuation (point ou liaison)

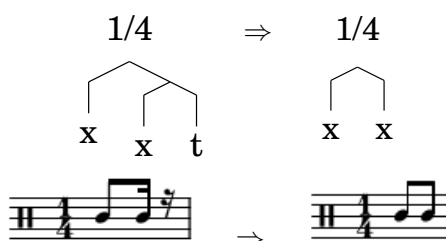


FIGURE 4.8

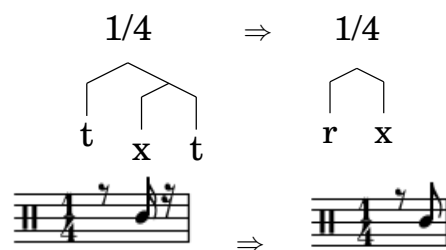


FIGURE 4.9

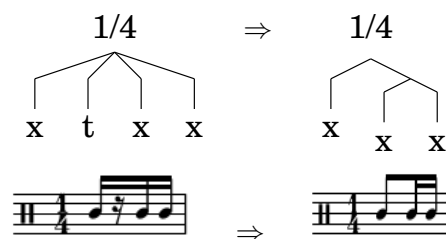


FIGURE 4.10

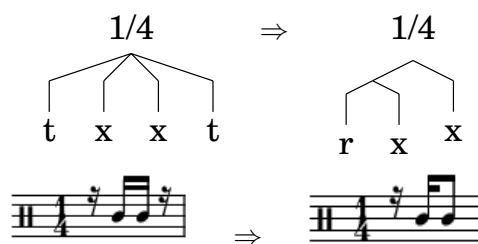


FIGURE 4.11

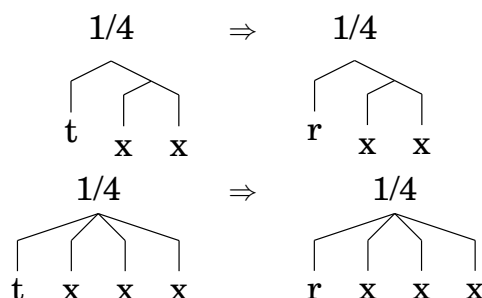


FIGURE 4.12

1195 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres  
1196 manquants seront mis en annexe.

1197 Les règles remplacent par un silence les continuations (t) qui sont au  
1198 début d'un temps. Cela est valable pour ce système mais lorsqu'il y a des  
1199 ouvertures de charley, cela n'est pas toujours applicable. Ce problème est  
1200 évoqué de le chapitre 3.

1201 ⇒ **Objectif de cette expérimentation théorique :**

1202 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette  
1203 expérimentation théorique a pour but d'orienter la recherche d'autres  
1204 systèmes par observation du jeu de données et de montrer comment les  
1205 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

## 1206 4.4 Résultats et discussion

1207

1208 Cette section regroupe les avancées qui ont été réalisées par rapport aux  
1209 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-  
1210 tats de l'ADT avec Qparse. Nous avons améliorer le système de quantifi-  
1211 cation de Qparse pour la batterie, notamment le passage à la polyphonie  
1212 avec les Jams.

1213 Nous avons pu obtenir des arbres de parsing corrects en améliorant les  
1214 grammaires avec des fichiers MIDI courts. Puis, une sortie MEI a été  
1215 aussi été obtenu (encore à vérifier).

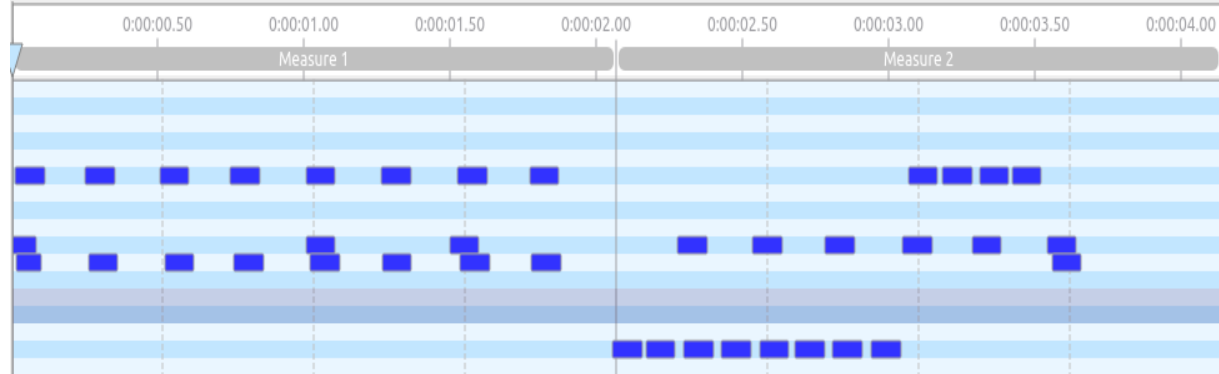
je vois 2 problématiques et contrib. principales : 1) transcription polyphonique par parsing (verrou) : jams etc 2) réécriture, pour séparation en voix et simplification, aidée (guidée) par système. Ce serait bien de présenter la contrib. 1 dans une section (comme 4.3 pour 2), avant d'aborder résultats et discussion

## 1216 Les Jams

Les Jams permettent de passer du monophonique au polyphonique.

### Le parsing

Tests effectués avec le fichier midi suivant :



Un premier test convaincant est effectué avec la grammaire suivante :

// bar level

0 -> C0 1

0 -> E1 1

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

// half bar level

9 -> C0 1

9 -> E1 1

// beat level

1 -> C0 1

1 -> E1 1

1 -> T2(2, 2) 1

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

// croche level

2 -> C0 1

2 -> E1 1

// double level

4 -> C0 1

4 -> E1 1

4 -> E2 1

4 -> T2(6, 6) 1



1249 // triple level

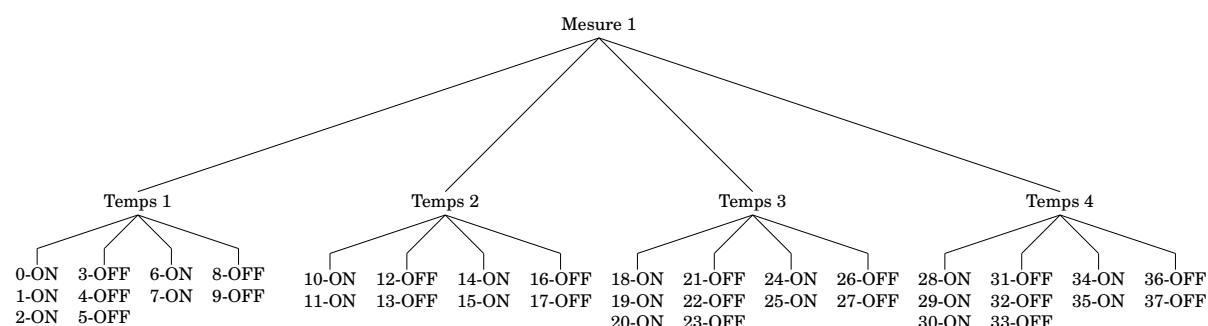
1250 6 -> E1 1

1251

1252 Cette grammaire sépare les ligatures par temps au niveau de la  
 1253 mesure. Puis, au niveau du temps, elle autorise les divisions par deux  
 1254 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.  
 1255 L'arbre de parsing en résultant est considéré comme « convaincant » car  
 1256 il découpe correctement les mesures et les temps.

1257

1258



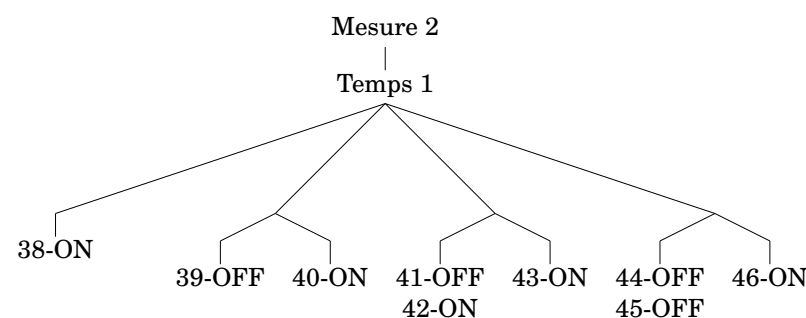
1259

1260

1261 Les temps de la première mesure du fichier MIDI sont bien quanti-  
 1262 fié mais ceux de la deuxième mesure présentent quelques défauts de  
 1263 quantification visibles dès le premier temps.

1264

1265



1266

1267

1268 Les Onsets sont correctement triés au niveau des doubles croches  
 1269 mais certaines doubles croches sont inutilement subdivisées en triples  
 1270 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps  
 1271 ci-dessus).

1272

### 1273 2ème exemple :

1274 Après une augmentation du poids des triples croches dans la grammaire  
 1275 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à  
 1276 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets  
 1277 sont bien triés et aucuns ne sont subdivisés.

## 1278 **Évaluation**

1279 Pour l'évaluation, il aurait fallu produire un module.  
 1280 L'évaluation est-elle automatique ou manuelle?  
 1281 Possibilité d'un export lilypond en arbre pour comparer l'output avec la  
 1282 transcription manuelle.  
 1283 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel  
 1284 ou MEI pour les comparer et faire des statistiques. Si transformés en  
 1285 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)  
 1286 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.  
 1287 L'expérimentation peut-être considérer comme une évaluation manuelle?  
 1288 (magicien d'Oz)  
 1289 Lilypond vers MIDI + output vers MIDI  $\Rightarrow$  Comparaison des MIDI  
 1290 dumpés.  
 1291

## 1292 **Discussion**

1293 Dans cette section, nous discuterons sur la pertinence de l'ensemble des  
 1294 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui  
 1295 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.  
 1296 Écrire des règles de réécriture spécifique aux charley avec un système  
 1297 approprié. Le jeu de système  
 1298 — implémenter un pattern...  
 1299  $\Rightarrow$  manque de temps?  
 1300  
 1301 — La partie résultat est manquante car :  
 1302  $\Rightarrow$  Sujet très difficile ;  
 1303  $\Rightarrow$  Matcher les motifs peut être fait ultérieurement ;  
 1304 Mais ce travail aurait été indispensable pour obtenir une quan-  
 1305 tité de résultats qui justifieraient une évaluation automatique  
 1306 permettant de faire des graphiques.  
 1307  
 1308 — L'évaluation fut entièrement manuelle car :  
 1309  $\Rightarrow$  Très dure automatiquement : il faut comparer 2 partitions (réf  
 1310 VS output)  
 1311 — Le ternaire jazz (voir expérience 2)  
 1312 — Reconnaissance d'un motif sur le MIDI  
 1313 Reconnaître un motif (système) sur une mesure de l'input (un fi-  
 1314 chier midi représentant des données audios)  
 1315  $\Rightarrow$  Motif (système) reconnu : true ou false  
 1316 Si true :  
 1317 - Choisir la grammaire correspondante ;  
 1318 - Parser le MIDI ;

- 1319       - Appliquer les règles de réécritures (Séparation des voix et simpli-  
1320       fication)
- 1321       — Nous travaillerons aussi sur la détection de répétitions sur plu-  
1322       sieurs mesures afin de pouvoir corriger des erreurs sur une des  
1323       mesures qui aurait dû être identique aux autres mais qui présente  
1324       des différences.
- 1325       — dans quelle catégorie mettre le shuffle ?
- 1326   Sujet passionnant mais difficile. Obtenir la totalité des critères pour le  
1327   mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique  
1328   à la batterie a été générée. Elle sera un bon point de départ pour les  
1329   travaux futurs dont plusieurs propositions sont énoncés dans le présent  
1330   document.



## CONCLUSION GÉNÉRALE

1332 Dans ce mémoire, nous avons traité de la problématique de la transcrip-  
1333 tion automatique de la batterie. Son objectif était de transcrire, à partir  
1334 de leur représentation symbolique MIDI, des performances de batteur de  
1335 différents niveaux et dans différents styles en partitions écrites.  
1336 Nous avons avancé sur le parsing des données MIDI établissant un pro-  
1337 cessus de regroupement des événements MIDI qui nous a permis de faire  
1338 la transition du monophonique vers le polyphonique. Une des données  
1339 importante de ce processus était de différencier les nature des notes d'un  
1340 accord, notamment de distinguer lorsque 2 notes constituent un accord  
1341 ou un *fla*.  
1342 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-  
1343 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-  
1344 tionnables en amont du parsing, soit par indication des noms des fichiers  
1345 MIDI, soit par reconnaissance de la métrique avec une approche diction-  
1346 naire de patterns prédéfinis<sup>5</sup> qu'il serait pertinent de mettre en œuvre en  
1347 machine learning.  
1348 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre  
1349 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif  
1350 d'un système et pour la simplification grâce aux gammes du motif d'un  
1351 système. Nous avons aussi montré comment, dans des travaux futurs, un  
1352 système dont le motif serait reconnu en amont dans un fichier MIDI pour-  
1353 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-  
1354 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans  
1355 la chaîne de traitement en terme de réécriture.  
1356 Il sera également intéressant d'étudier comment l'utilisation de LM peut  
1357 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération  
1358 entièrement automatisée de partitions de batterie et au problème général  
1359 de l'AMT de bout en bout.[10]

---

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



## BIBLIOGRAPHIE

- 1361 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41  
1362 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,  
1363 1996. – Cité pages 7, 28 et 32.
- 1364 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11  
1365 et 14.
- 1366 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité  
1367 page 12.
- 1368 [4] Gaël Richard et al. De fourier à la reconnaissance  
1369 musicale. Available at [https://interstices.info/  
1370 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).  
1371 – Cité page 12.
- 1372 [5] Caroline Traube. Quelle place pour la science au sein de la musico-  
1373 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1374 [6] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions  
1375 entre les traitements de la musique et du langage. *La Lettre des  
1376 Neurosciences*, 58 :24–26, 2018. – Cité page 12.
- 1377 [7] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.  
1378 Techniques de traitement automatique du langage naturel appli-  
1379 quées aux représentations symboliques musicales. In *JIM 2021 -  
1380 Journées d’Informatique Musicale*, Virtual, France, July 2021. –  
1381 Cité page 12.
- 1382 [8] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-  
1383 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1384 [9] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering  
1385 music relations with sequential attention. In *NLP4MUSA*, 2020. –  
1386 Cité page 13.
- 1387 [10] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger  
1388 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-  
1389 lenges and future directions. *Journal of Intelligent Information Sys-  
1390 tems*, 41, 12 2013. – Cité pages 14, 16, 21 et 61.

- 1391 [11] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local  
1392 musical statistics as guides for audio-to-score piano transcription.  
1393 *Information Sciences*, 566 :262–280, 2021. – Cité pages 14, 23 et 24.
- 1394 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J  
1395 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –  
1396 Cité pages 14 et 25.
- 1397 [13] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-  
1398 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander  
1399 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*  
1400 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–  
1401 1483, 2018. – Cité pages 16, 22 et 25.
- 1402 [14] Moshekwa Malatji. Automatic music transcription for two instru-  
1403 ments based variable q-transform and deep learning methods, 10  
1404 2020. – Cité page 22.
- 1405 [15] Antti J. Eronen. Musical instrument recognition using ica-based  
1406 transform of features and discriminatively trained hmms. *Seventh*  
1407 *International Symposium on Signal Processing and Its Applications*,  
1408 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 22.
- 1409 [16] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic  
1410 drum sound description for real-world music using template adap-  
1411 tation and matching methods. *International Conference on Music*  
1412 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 23.
- 1413 [17] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-  
1414 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-  
1415 tization and Score Structuring. In *MCM 2019 - Mathematics and*  
1416 *Computation in Music*, volume Lecture Notes in Computer Science  
1417 *of Proceedings of the Seventh International Conference on Mathema-*  
1418 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June  
1419 2019. Springer. – Cité pages 23 et 24.
- 1420 [18] C. Agon, K. Haddad, and G. Assayag. Representation and rende-  
1421 ring of rhythm structures. In *Proceedings of the First International*  
1422 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.  
1423 IEEE Computer Society. – Cité page 24.
- 1424 [19] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A  
1425 Term Rewriting Based Structural Theory of Rhythm Notation. Re-  
1426 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –  
1427 Cité page 24.
- 1428 [20] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating  
1429 equivalent rhythmic notations based on rhythm tree languages. In  
1430 *Third International Conference on Technologies for Music Notation*



- 1431        *and Representation (TENOR)*, Coroña, Spain, May 2017. Helena Lo-  
1432        pez Palma and Mike Solomon. – Cité page 24.
- 1433 [21] R. Marxer and J. Janer. Study of regularizations and constraints in  
1434        nmf-based drums monaural separation. In *International Conference*  
1435        *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,  
1436        02/09/2013 2013. – Cité page 25.
- 1437 [22] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions  
1438        Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 28.
- 1439 [23] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue  
1440        Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 28.
- 1441 [24] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-  
1442        tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -  
1443        94120 Fontenay-sous-Bois, 2000. – Cité pages 29 et 40.
- 1444 [25] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David  
1445        Bamman. Learning to groove with inverse sequence transforma-  
1446        tions. In *International Conference on Machine Learning (ICML)*,  
1447        2019. – Cité page 45.

