

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

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

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

8 *Parcours :*
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

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

14 *Encadrant :*
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021

TABLE DES MATIÈRES

18	Liste des figures	4
19	Liste des tableaux	5
20	Introduction générale	7
21	1 Contexte	11
22	1.1 Langues naturelles et musique en informatique	12
23	1.2 La transcription automatique de la musique	14
24	1.3 La transcription automatique de la batterie	15
25	1.4 Les représentations de la musique	16
26	2 État de l'art	21
27	2.1 Monophonique et polyphonique	21
28	2.2 Audio vers MIDI	22
29	2.3 MIDI vers partition	24
30	2.4 Approche linéaire et approche hiérarchique	24
31	3 Méthodes	29
32	3.1 La notation de la batterie	29
33	3.2 Modélisation pour la transcription	37
34	3.3 Qparse	38
35	3.4 Les systèmes	40
36	4 Expérimentations	47
37	4.1 Le jeu de données	47
38	4.2 Analyse MIDI-Audio	49
39	4.3 Expérimentation théorique d'un système	53
40	4.4 Résultats et discussion	57
41	Conclusion générale	63
42	Bibliographie	65

LISTE DES FIGURES

44	1.1	Exemple évènements avec durée	16
45	1.2	Critère pour un évènement	17
46	1.3	Exemple évènements sans durée	17
47	1.4	Exemple de partition de piano	18
48	1.5	MusicXML	18
49	2.1	Transcription automatique <dam>remettre ici la citation de la	
50		capture d'écran avec la page</dam>	23
51	2.2	HMM	26
52	2.3	arbre_jazz	27
53	3.1	29
54	3.2	Rapport des figures de notes	30
55	3.3	Les instruments de la batterie	31
56	3.4	Hauteur et têtes de notes	31
57	3.5	Point et liaison	32
58	3.6	Les silences	33
59	3.7	Silence joué	34
60	3.8	Équivalence	35
61	3.9	Séparation des voix	35
62	3.10	Les accents et les ghost-notes	36
63	3.11	Exemple pour les accentuations et les ghost-notes	36
64	3.12	Présentation de Qparse	39
65	3.13	Métrique	41
66	3.14	Motif 4-4 binaire	42
67	3.15	Motif 4-4 jazz	43
68	3.16	Système 4-4 afro-latin	43
69	3.17	Simplification	44
70	3.18	45
71	4.1	Batterie électronique	48
72	4.2	Partition de référence	52
73	4.3	Motifs et gammes	53
74	4.4	Partition d'un système en 4/4 binaire	54
75	4.5	Arbre de rythme — système	54
76	4.6	Arbre de rythme — voix haute	55
77	4.7	Arbre de rythme — voix basse	55
78	4.8	56
79	4.9	56

80	4.10	56
81	4.11	57
82	4.12	57

LISTE DES TABLEAUX

84	1.1 speechToText vs AMT	13
85	3.1 Pitches et instruments	37
86	3.2 Systèmes	41

INTRODUCTION GÉNÉRALE

88 QUOI?

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

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

110

111 POURQUOI?

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

118

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

121 COMMENT?

122 → Problématique :

123 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.

133

134 → Méthodes :

135 → Contributions :

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

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

146

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

- 148 — une modélisation de la notation de la batterie (fusion de 3.1 et de
149 3.2) qui était jusqu'à présent inexistante.
- 150 — plusieurs transcriptions manuelles dans le but d'analyser les conte-
151 nus des fichiers MIDI et Audio (4.2) et de faire des comparaisons
152 de transcription avec des outils déjà existants¹.
- 153 — une partition de référence transcrite manuellement sur l'entièreté
154 d'une performance du jeu de données afin de repérer les éléments
155 importants pour la modélisation et de faire les liens entre les cri-
156 tères des données d'*input* avec l'écriture finale (4.2). Cette partition
157 avait aussi pour objectif d'effectuer des tests et des évaluations.
- 158 — le passage au polyphonique en théorie et en implémentation im-
159 pliquant la théorie sur la détection de l'identité de notes dans un
160 Jam² et l'implémentation de tests unitaires sur le traitement des
161 Jams (4.4).
- 162 — la création de grammaires pondérées spécifiques à la batterie (4.4)

163

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	14
1.3	La transcription automatique de la batterie	15
1.4	Les représentations de la musique	16

Introduction

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

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

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

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

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

1.1 Langues naturelles et musique en informatique

COMPUTER MUSIC

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

RIM

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

RIM et TAL

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

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

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

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

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

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

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

exemples / illustration de la
proximité thématique?

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

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

TABLE 1.1 – speechToText vs AMT

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

285 1.2 La transcription automatique de la musique

286 1. OBJECTIF

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

298

299 2. APPLICATIONS

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

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

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

317

318 3. PROBLÈMES ET MÉTHODES SCIENTIFIQUES

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

8. Ici, on parle de notation occidentale.

9. ethno-musicologie

10. par exemple par la constitution de corpus musicologiques

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

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

1.3 La transcription automatique de la batterie

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

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

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

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

cite méthode et école Agostini?

12. Roland TR-808

13. SQ-1

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

1.4 Les représentations de la musique

Les données audio

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

Les données MIDI

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

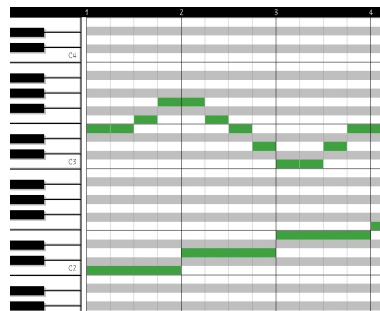


FIGURE 1.1 – Exemple événements avec durée

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

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

Ici, définir Tick et channel.

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

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

citer M. Müller FMP pour cette section ?

trop technique. ne pas repier wikipédia

LPCM pas utile ici. parle juste échantillons et compression.

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

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

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

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

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

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

FIGURE 1.2 – Critère pour un évènement

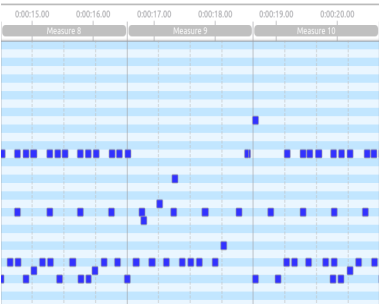


FIGURE 1.3 – Exemple évènements sans durée

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

401 **Les partitions**

402 Une partition de musique¹⁵ est un document qui porte la représentation
403 systématique du langage musical sous forme écrite. Cette représentation
404 est appelée transcription et elle sert à traduire les quatre caractéristiques
405 du son musical :
406 — la hauteur ;
407 — la durée ;
408 — l'intensité ;

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



FIGURE 1.4 – Exemple de partition de piano

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

413

414

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

417

— le timbre.

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

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

Les formats XML

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

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

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



FIGURE 1.5 – MusicXML

Le figure 1.5¹⁶ représente un do en clef de sol de la durée d'une ronde sur une mesure en 4/4 écrit au format MusicXML. 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.

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

427 **Conclusion**

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

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

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

ÉTAT DE L'ART

Sommaire

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

Introduction

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

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

2.1 Monophonique et polyphonique

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

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

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

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

2.2 Audio vers MIDI

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

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

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

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

MIDI non-quantifié = performance (à expliquer)

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

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

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

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

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

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

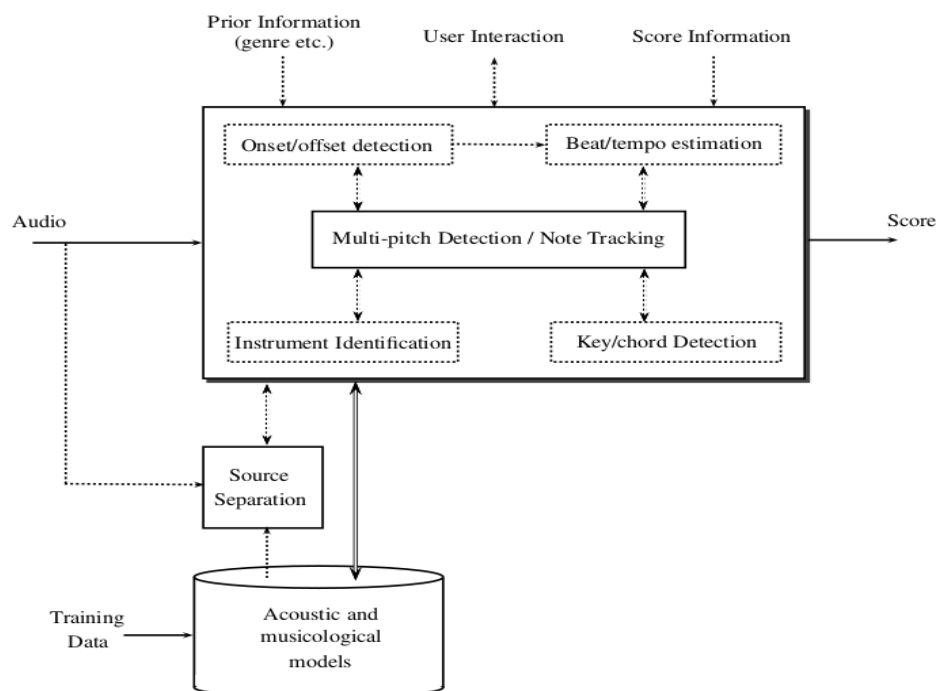


FIGURE 2.1 – Transcription automatique <dam>remettre ici la citation de la capture d'écran avec la page</dam>

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

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

haute fréquence, aigus?

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

classification des évène- 524
 ments? la phrase semble 525
 redondante

pas clair... peut-être just 528
 mentionner les modèles 529
 probabilistes utilisés

532 2.3 MIDI vers partition

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

ce n'est pas exactement 533
 cela. cf. proposition de des- 534
 cription + détaillée en com- 535
 mentaires

de manière conjointe 544

langage a priori 546

qui nécessite de traiter le 549
 problème supplémentaire 550
 de la séparation de voix. i.e. 551
 pour la batterie on nveut 552
 quantification + structu- 553
 ration + séparation mais 554
 seules les 2 premières sont 555
 couplées dans l'approche de 556
 tonn stage.

2.4 Approche linéaire et approche hiérarchique

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

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

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

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

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

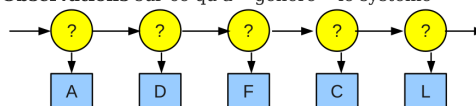


FIGURE 2.2 – HMM

563 *Source : Cours de Damien Nouvel*⁵

564

565

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

je ne comprend pas bien 573
 l'explication. le pb est plu- 574
 tot vue locale (déduction de 575
 la proba d'une durée à par- 576
 tir de la durée précédente, 577
 par ex. dans un HMM) vs 578
 vue globale, dans une hié- 579
 rarchie

RT? 578
 579

techniques de réécriture 580
 appliquée à la déduction 581
 automatique, calcul symbo- 582
 lique

le calcul d'équiv. 583

584

585

586

citer thèse de David Rizo 587
 (Valencia) 588

589

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

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

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

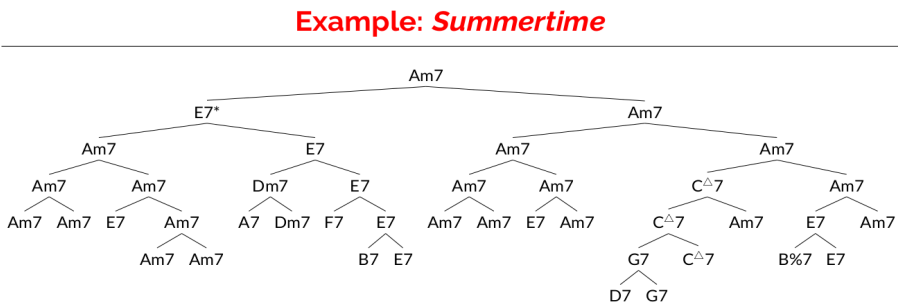


FIGURE 2.3 – arbre_jazz
Représentation arborescente d’une grille harmonique [12]

Conclusion

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

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

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

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

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

latex : enumerate

MÉTHODES

Sommaire

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

Introduction

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

Pour commencer, nous exposerons une description de la notation de la batterie ainsi qu'une modélisation de celle-ci pour la représentation des données rythmiques en arbres syntaxiques. Nous poursuivrons avec une présentation de qparse¹, un outil de transcription qui est développé à l'Inria, l'Université de Nagoya et au sein du laboratoire Cedric au CNAM.

Enfin, nous présenterons les systèmes.

plusieurs développeurs

systèmes, une représentation théorique qui permet...

3.1 La notation de la batterie



FIGURE 3.1

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/>

635 — La blanche, elle vaut 2 ;

636 — La noire, elle vaut 1 ;

637 — La croche, elle vaut 1/2.

plusieurs éléments

638 Une figure de note [1] de musique combine plusieurs critères ² :

— 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

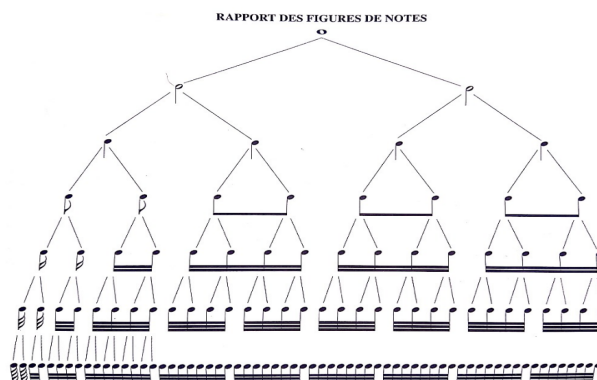


FIGURE 3.2 – Rapport des figures de notes
[1]

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

ce premier paragraphe (je
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

657 Source figure 3.3 : `z_images/3_methodes/0_notation_de_la_`
658 `batterie.`

659 Pour la transcription, nous proposons une notation inspirée du recueil de
660 pièces pour batterie de J.-F. Juskowiak [24] et des méthodes de batterie

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



FIGURE 3.3 – Les instruments de la batterie

661 Agostini [25], car nous trouvons la position des éléments cohérente et in-
662 tuitive.

664 En effet, les hauteurs sur la portée représentent :

- 665 — La hauteur physique des instruments :
666 La caisse claire est centrale sur la portée et sur la batterie (au
667 niveau de la ceinture, elle conditionne l'écart entre les pédales et
668 aussi la position de tous les instruments basiques d'une batterie).
669 Tout ce qui en-dessous de la caisse-claire sur la portée est en
670 dessous de la caisse-claire sur la batterie (pédales, tom basse);
671 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est
672 aussi sur la batterie.
673
- 674 — La hauteur des instruments en terme de fréquences :
675 Sauf pour le charley au pied et si l'on sépare en trois groupes
676 (grosse-caisse, toms et cymbales), de bas en haut, les instruments
677 vont du plus grave au plus aigu.

pour aider, tu pourrais don-
ner une figure représentant
la batterie avec le nom des
instruments et abrégia-
tion.

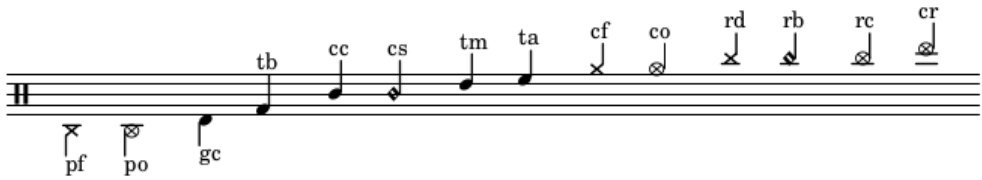


FIGURE 3.4 – Hauteur et têtes de notes

678 Les noms des instruments correspondant aux codes des notes de la figure
679 3.4 sont dans le tableau 3.1.

têtes de notes?

680 Les durées

681 Comme nous venons de la voir, la majorité des instruments de la batterie
 682 sont représentés par les têtes des notes. Par conséquent, les symboles
 683 rythmiques concernant la tête de note ne pourront pas être utilisés. Cela
 684 est valable aussi pour la présence ou non de la hampe puisque ce phé-
 685 nomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition
 686 blanche-ronde). L'usage des blanches existe dans certaines partitions de
 687 batterie [26] mais cela reste dans des cas très rares. Certains logiciels per-
 688 mettent de faire des blanches avec des symboles spécifiques à la batterie
 689 ou aux percussions mais leur lecture reste peu aisée et leur utilisation
 690 pour la batterie est rarissime.

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

- Le point ;
- La liaison.

694 Ces symboles ne seront utiles que pour l'écriture des ouvertures de char-
 695 ley. Le charley est le seul instrument de la batterie dont la durée est quan-
 696 tifiée (les cymbales attrapées à la main peuvent l'être aussi mais cela est
 697 très rare.)



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.5 – Point et liaison

= la position des temps 698

699

faire un "enumerate" 701

702

703

L'écriture de la batterie doit faire ressortir la pulsation. La première chose à prendre en compte pour analyser la figure 3.5 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 ob-
tenant 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 privilé-
giera la liaison.

Les silences

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

expliquer la notation (géné-
rale) des silebces en §1.4?

quantifier = noter? ou
quantifier la durée?

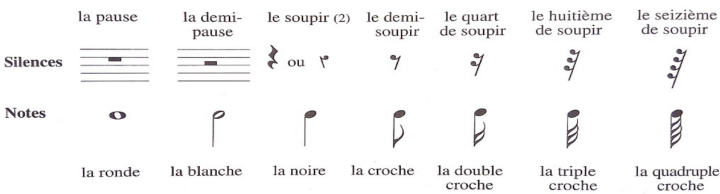


FIGURE 3.6 – Les silences

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

pas très clair



Exemple 1



Exemple 2

FIGURE 3.7 – Silence joué

722 L'exemple 1 de la figure 3.7 montre ce qui est écrit dans les données MIDI
 723 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-
 724 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1
 725 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-
 726 truments jouant simultanément.

727 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte
 728 la note suivante pour l'écriture : - Si c'est un charley fermé joué à la main
 729 ⇒ la note sera cf;

itemize

cf?

730 - Si c'est un charley fermé joué au pied ⇒ la note sera un silence.

731 Les équivalences rythmiques

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

phrase alambiquée... pou
prolonger la durée?seuls comptent les date
début de notes onsets.

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

743 Sur la figure 3.8, théoriquement, il faudra choisir la notation de la
 744 deuxième mesure mais dans certains contextes, pour des raisons de lisi-
 745 bilité ou de surcharge, la version sans les silences de la troisième mesure
 746 pourra être choisie.



FIGURE 3.8 – Équivalence

747 Les voix

748 Les voix³ désignent les différentes parties mélodiques constituant une
 749 composition musicale et destinées à être interprétées, simultanément ou
 750 successivement, par un ou plusieurs musiciens. En batterie, une voix
 751 est l'ensemble des instruments qui, à eux seuls, constituent une phrase
 752 rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures
 753 étant possibles pour un même rythme, on peut regrouper les instruments
 754 de la batterie par voix. Sur une portée de batterie, il existe le plus souvent
 755 1 ou 2 voix. Sur la figure 3.9, il faudra faire un choix entre les exemples
 1, 2 et 3 qui sont trois façons d'écrire le même rythme.

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

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

une voix est caractérisée aussi par orientation des hampes?



FIGURE 3.9 – Séparation des voix

756 Ce choix se fera en fonction des instruments joués, de la nature plus ou
 757 moins systématique de leurs phrasés, et des associations logiques entre
 758 les instruments dans la distribution des rythmes sur la batterie (voir la
 759 section 3.4).
 760

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

761 Les accentuations et les ghost-notes

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

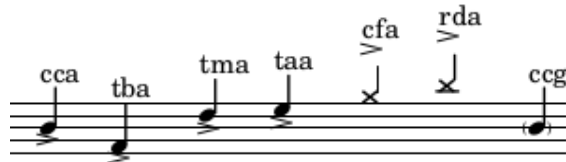


FIGURE 3.10 – Les accents et les ghost-notes

764

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

765 La figure 3.10 ne prend en compte que les accents que nous avons es-
766 timés nécessaires (voir la section 3.2). Les accents sont marqués par le
767 symbole « > ». Il est positionné au-dessus des notes représentant des cym-
768 bales et en-dessous des notes représentant des toms ou la caisse-claire.
769 Ce choix a été fait pour la partition de la figure 4.2 car elle est plus lisible
770 ainsi, mais ces choix devront être adaptés en fonction des différents sys-
771 tèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz,
772 les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il
773 faudra donc mettre les symboles d'accentuation correspondants au-dessus
774 des têtes de notes.

775

expliquer ce qu'est une
ghost-notes

776 La dernière note de la figure 3.10 montre un exemple de ghost-notes. Le
777 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note
sans changer la tête de note.

778

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

779 Pour les codes, on prend le code de la note et on ajoute un « a » pour un
accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.10
sont exposées en situation réelle dans la figure 3.11.



FIGURE 3.11 – Exemple pour les accentuations et les ghost-notes

780

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 certain code-instrument se voit 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...)

je ne comprend pas cette phrase.

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

La vélocité

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

Cette transcription nous a mené aux observations suivantes :

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

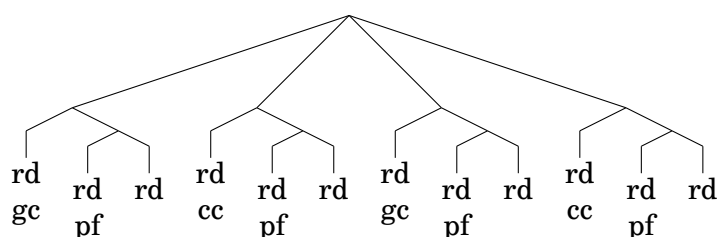
citation lilypond

et l’analyse d’autres fichiers MIDI ?

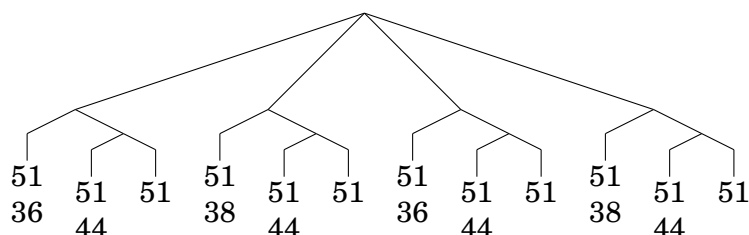
- 803 — Les accents et les ghosts-notes ne sont significatifs ni pour les ins-
 804 truments joués au pied, ni pour les cymbales crash.
 805 En effet, certaines vélocités en dessous de 40 étant détectées et ins-
 806 crites dans les données MIDI sont dues au mouvement du talon du
 807 batteur qui bat la pulsation sans particulièrement jouer le charley.
 808 Ce mouvement est perçu par le capteur de la batterie électronique
 809 mais le charley n'est pas joué.
 810 — Au final, nous avons relevé les ghost-notes et les accents pour la
 811 caisse-claire ainsi que les accents pour les toms et les cymbales
 812 rythmiques (charley et ride).

813 Les arbres de rythmes

814 Les arbres de rythmes représentent un rythme unique dont les possibili-
 815 tés de notation sur une partition sont théoriquement multiples.
 816 Voici une représentation de la figure 3.9 en arbre de rythmes avec les
 817 codes de chaque instrument :



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



819 Chacun des trois exemples de la figure 3.9 est représenté par un des deux
 820 arbres syntaxiques ci-dessus.
 821

822 3.3 Qparse

823 La librairie Qparse⁴ implémente la quantification des rythmes basée
 824 sur des algorithmes d'analyse syntaxique pour les automates arbores-
 825

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

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

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

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

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

quantification rythmique
 + structuration de partition

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

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

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

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

Les principaux contributeurs sont :

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

apprentissage

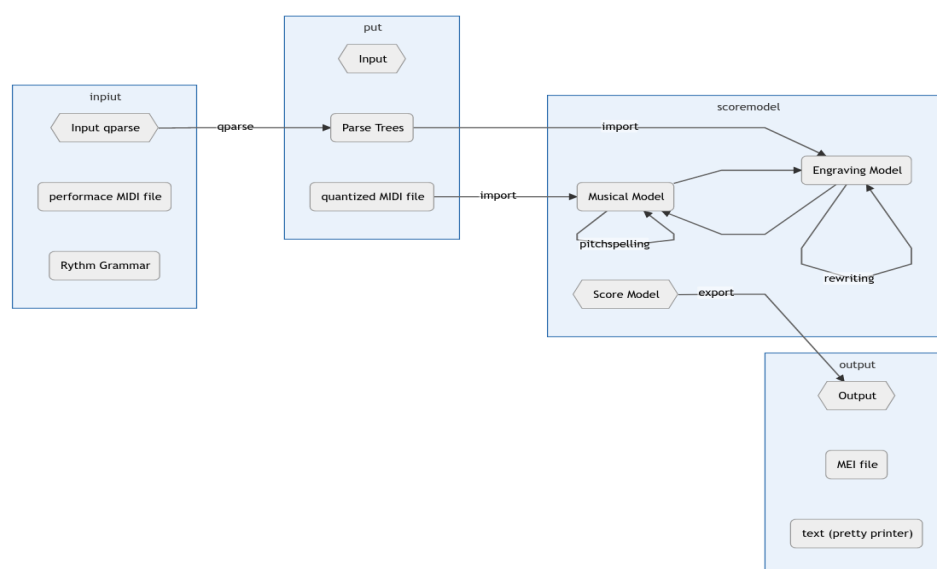


FIGURE 3.12 – Présentation de Qparse

Explication des différentes étapes de la figure 3.12⁵ :

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

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

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 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 définie, 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	duolets, quartelets	2
afros-cubains	simple	croches	-	3

TABLE 3.2 – Systèmes

dans 4 grandes catégories :

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

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

Objectif des systèmes

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

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

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

Détection d’indication de mesure

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

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 rapidement la notation des signatures rythmiques, par exemple en 1.4



Exemple 1



Exemple 2

FIGURE 3.13 – Métrique

La figure 3.13 montre deux indications de mesure différentes. L'une (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des triolets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est traditionnellement écrit en binaire avec ou sans triolet (même si cette musique est dite ternaire alors que le rock ternaire sera plutôt écrit comme dans l'exemple 2).

Choix d'une grammaire

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

Séparation des voix

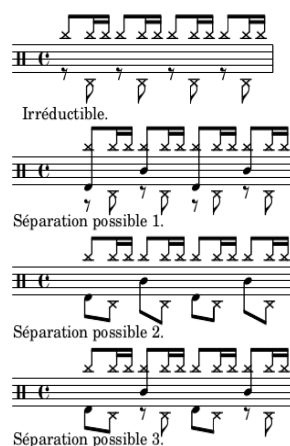


FIGURE 3.14 – Motif 4-4 binaire

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 3.14 est appelée « Irréductible

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

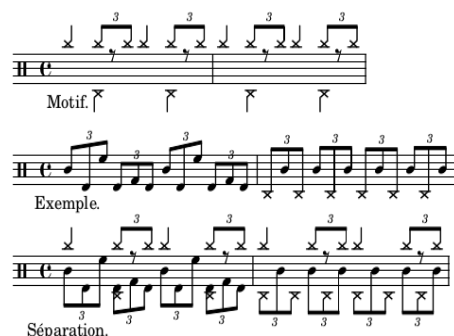


FIGURE 3.15 – Motif 4-4 jazz

942 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 943 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 944 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 945 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 946 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 947 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 948 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 949 et de caisse claire et son écriture séparée serait inutilement compliquée
 950 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 951 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de
 952 charley en bas.

quel exemple?



FIGURE 3.16 – Système 4-4 afro-latin

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

958 Simplification de l'écriture

959 Les explications qui suivent seront appuyé par une expérimentation théo-
 960 rique dans la section 4.3.

expérimentation théo-
 rique??

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

964 Voici les différentes étapes à suivre :

- 965 — Pour chaque gamme du système, faire un arbre de rythme repré-
 966 sentant la gamme combinée avec le motif du système ;
- 967 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
 968 arbre de rythme par voix ;
- 969 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
 970 nœuds qui nécessitent une simplification et écrire la règle.

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

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

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

978 Même si on favorise l'usage des silences pour l'écart entre les notes n'ap-
 979 partenant pas au même temps, on les supprime systématiquement pour
 980 2 notes au sein d'un même temps et favorise, une liaison si co, un point si
 981 pas co et nécessaire, un simple ajustement de la figure de note si suffisant.
 982

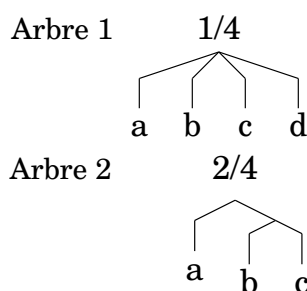


FIGURE 3.17 – Simplification

itemize

983 Soit l'arbre 1 de la figure 3.17 dans lequel : a et d sont des instruments de
 984 la batterie (x) ;

985 b et c sont des continuations (t) ;

986 Pour chacune des conditions suivantes, une suite de la figure 3.18 est
 987 attribuée :

- 988 — Si a n'est pas un co :
- 989 ⇒ Suite 1a.

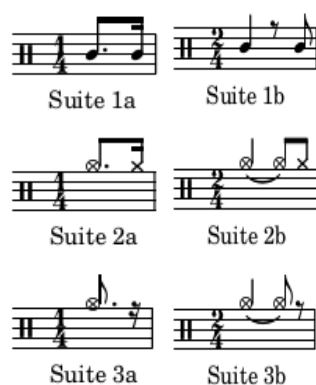


FIGURE 3.18

- 990 — Si a est un co :
 991 — Si d est un cf :
 992 ⇒ Suite 2a.
 993 — Si d est un pf :
 994 ⇒ Suite 3a : d deviens un silence (r).
 995

996 Soit l'arbre 2 de la figure 3.17 dans lequel :
 997 a et c sont des instruments de la batterie (x);
 998 b est une continuation (t); Pour chacune des conditions suivantes, une
 999 suite de la figure 3.18 est attribuée :

- 1000 — Si a n'est pas un co :
 1001 ⇒ Suite 1b, b devient un silence.
 1002 — Si a est un co :
 1003 — Si c est un cf :
 1004 ⇒ Suite 2b, b devient une liaison et c devient un cf.
 1005 — Si c est un pf :
 1006 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
 1007

1008 *Rappel :*

- 1009 *cf* = charley fermé joué à la main ;
 1010 *co* = charley ouvert joué à la main ;
 1011 *pf* = charley fermé joué au pied.

1012

1013 **Problème : le cf et le co ne seront jamais sur la même voix que le**
 1014 **pf... Par conséquent, les règles concernant les charleys ouverts**
 1015 **doivent-elles être appliquées sur l'arbre de parsing de l'input?...**

1016 **Conclusion**

1017 Nous avons formalisé une notation de la batterie, modélisé cette notation
1018 pour la transcription de données MIDI en partition, nous avons décrit
1019 Qparse.

1020 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-
1021 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-
1022 propriée et énoncer des règles de séparation des voix et de simplification
1023 de l'écriture.

1024

1025

EXPÉRIMENTATIONS

1026

Sommaire

1027	4.1	Le jeu de données	47
1028	4.2	Analyse MIDI-Audio	49
1029	4.3	Expérimentation théorique d'un système	53
1030	4.4	Résultats et discussion	57
1031			
1032			
1033			

1035

Introduction

1036 Dans ce chapitre, nous présenterons le jeu de données et les analyses
 1037 audio-MIDI. Nous ferons ensuite l'expérimentation théorique d'un sys-
 1038 *tème* implémentable qui devra être utilisé comme base de connaissances
 1039 pour augmenter la rapidité et la qualité en sortie de Qparse. Nous pré-
 1040 senterons ensuite les avancées réalisées dans ce travail et une réflexion
 1041 sur les moyens de l'évaluer. Enfin, nous finirons par une discussion sur
 1042 l'ensemble du travail réalisé.

1043

4.1 Le jeu de données

1044 Nous avons utilisé le Groove MIDI Dataset ¹ [27] (GMD) qui est un jeu de
 1045 données mis à disposition par Google sous la licence Creative Commons
 1046 Attribution 4.0 International (CC BY 4.0).
 1047 Le GMD est composé de 13,6 heures de batterie sous forme de fichiers
 1048 MIDI et audio alignés. Il contient 1150 fichiers MIDI et plus de 22 000
 1049 mesures de batterie dans les styles les plus courants et avec différentes
 1050 qualités de jeu. Tout le contenu a été joué par des humains sur la batterie
 1051 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

1052 Autres critères spécifiques au GMD :

- 1053 — Toutes les performances ont été jouées au métronome et à un tempo
- 1054 choisi par le batteur.
- 1055 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 1056 qui ont pu improviser dans un large éventail de styles. Les don-
- 1057 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 1058 (professionnel ou amateur).
- 1059 — Les batteurs avaient pour instruction de jouer des séquences de
- 1060 plusieurs minutes ainsi que des fills²
- 1061 — Chaque performance est annotée d'un style (fourni par le batteur),
- 1062 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 1063 nyme du batteur.
- 1064 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 1065 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 1066 session du GMD.
- 1067 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 1068 fichier MIDI.

1069 **Format des données**

enregistre les données dans
des fichiers MIDI

1070 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-
1071 tinctes :

- 1072 — une pour le tempo et l'indication de mesure ;
- 1073 — une pour les changements de contrôle (position de la pédale de
- 1074 charley) ;
- 1075 — une pour les notes.

1076

1077 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1078 canal 9 (qui est le canal canonique pour la batterie).

1079 Pour simplifier le traitement de ces données, ces trois pistes ont été
1080 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.

1081
 1082 « Control Changes The TD-11 also records control changes speci-
 1083 fying the position of the hi-hat pedal on each hit. We have preserved this
 1084 information under control 4. »
 1085 (<https://magenta.tensorflow.org/datasets/groove>)
 1086 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 1087 tions dans les fichiers MIDI.
 1088 L'utilisation de pretty_midi devient urgente!

1089 4.2 Analyse MIDI-Audio

1090
 1091 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
 1092 partir de fichiers MIDI et Audio du GMD.

= analyses et transcriptions
manuelles

1093 Comparaisons de transcriptions

1094 Pour les comparaisons de transcriptions, les transcriptions manuelles
 1095 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les trans-
 1096 criptions automatiques (TA) ont toutes été générées manuellement avec
 1097 MuseScore.

méthodologie tr. manuelle.
import MIDI pour MuseS-
core

1098 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1099 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1100 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1101 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1102 premières notes (une double-croche et une croche) d'un triolet sur
- 1103 le temps 1 de la mesure 2 de la TA.

1104 Exemple d'analyse 2

- 1105 — Les doubles croches ont été interprétées en quintolet
- 1106 — La deuxième double-croche est devenue une croche.

Transcription manuelle ⇒ Transcription automatique



Transcription manuelle ⇒ Transcription automatique

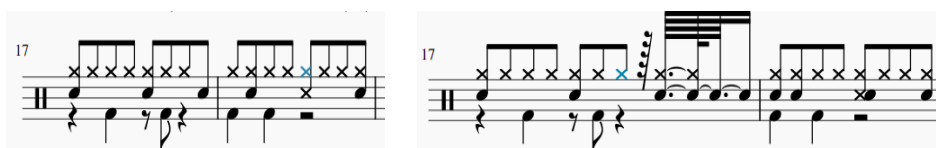


1107 Exemple d'analyse 3

- 1108 — Les grosses-caisses, les charleys et les caisses-claires ont été déca-
- 1109 lés d'un temps vers la droite.
- 1110 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1111 décalés d'une double croche vers la droite dans la TA.
- 1112 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1113 alors qu'elle appartenait à un triolet dans la TM.
- 1114 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1115 pas la TM.
- 1116

1117 Exemple d'analyse 4

Transcription manuelle ⇒ Transcription automatique



1118

- 1119 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- manière excessivement complexe!

conclusion sur ces exemples

1121 Exemple avec des flas

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

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

1123 Transcription manuelle



1124

1125 Transcription automatique

1126



1127

1128

- 1129 — Le premier fla est reconnu comme étant un triolet contenant une
- 1130 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1131 ornementée.
- 1132 — Le deuxième fla est reconnu comme étant un accord.
- 1133 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1134 dans la TA.
- 1135 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1136 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1137 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1138 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1139 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1140 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1141 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1142 temps 3 de la TA.
- 1143 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1144 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1145 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1146 cross-sticks.

1147 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1148 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1149 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1150 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1151 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1152 [Inria](https://github.com/MartinDigard/Stage_M2_Inria)) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1153 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1154 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1155 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1156 et quatre).

1157 La transcription des données audio et MIDI contenues dans ces fichiers
 1158 a permis une analyse plus approfondie des critères à relever pour chaque
 1159 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 1161
 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 mesure 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

1186 dont la première ne contient que des croches et la deuxième que des
 1187 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1188 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1189 Systèmes — motifs et gammes combinés

1190 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1191 de la figure 4.3.

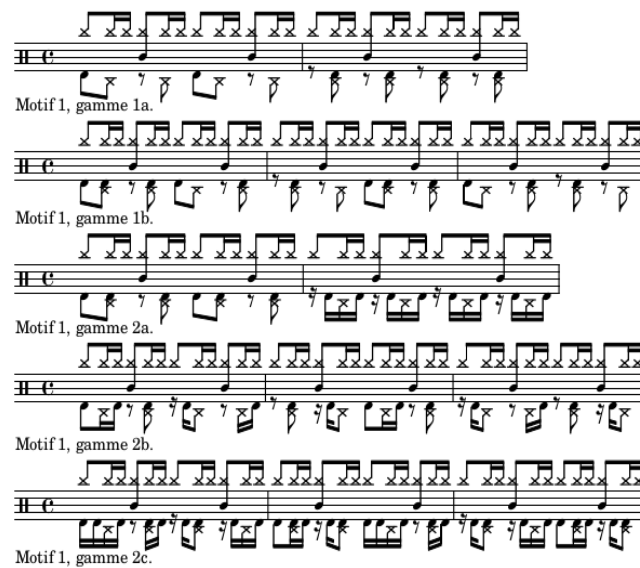


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1192

1193 Représentation du système en arbres de rythmes

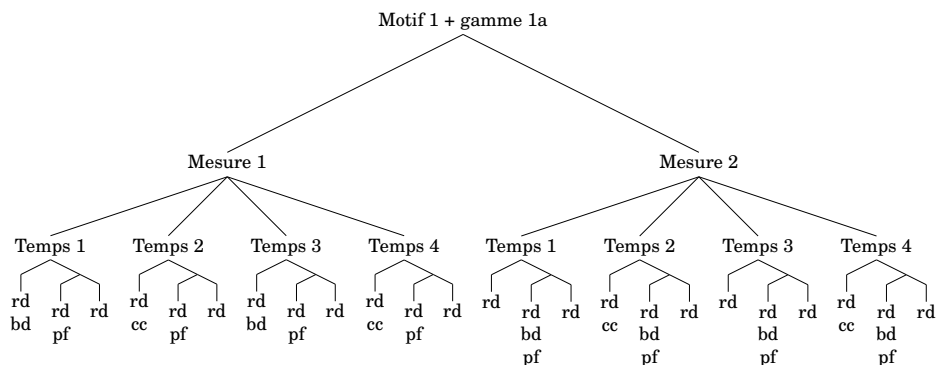


FIGURE 4.5 – Arbre de rythme — système

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

Réécriture — séparation des voix et simplification

La séparation des voix

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

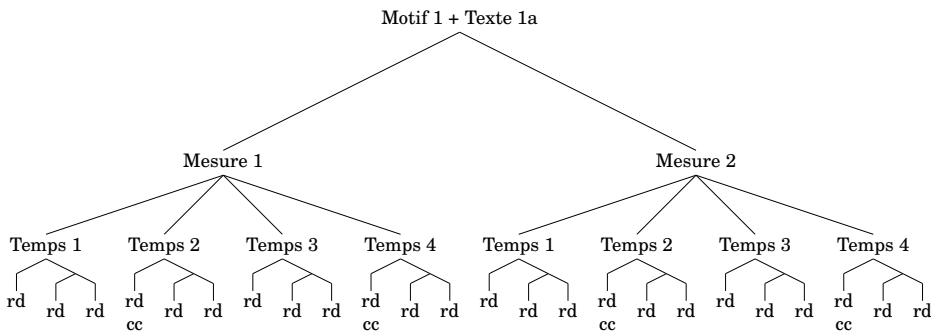


FIGURE 4.6 – Arbres de rythme — voix haute

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

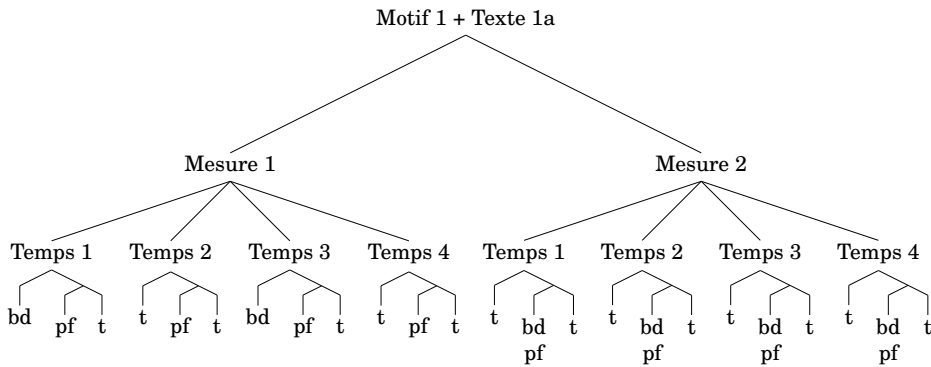


FIGURE 4.7 – Arbres de rythme — voix basse

La voix basse regroupe la grosse-caisse et le charley au pied sur les ligatures du bas.

1209 Les règles de simplifications

1210 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1211 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1212 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1213 entre elles (rendre la pulse visuelle).

1214

1215 Pour les figures ci-dessous :

1216 — x = une note ;

1217 — r = un silence ;

1218 — t = une continuation (point ou liaison)

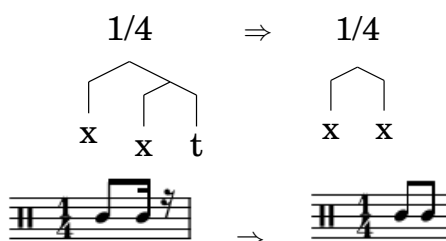


FIGURE 4.8

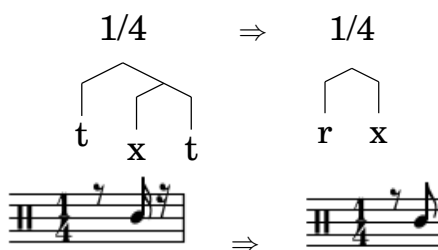


FIGURE 4.9

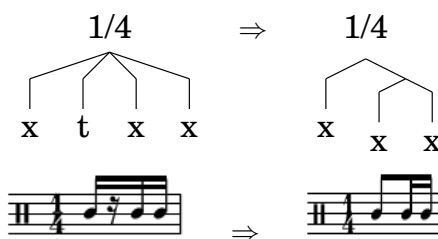


FIGURE 4.10

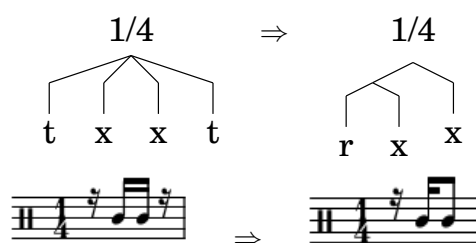


FIGURE 4.11

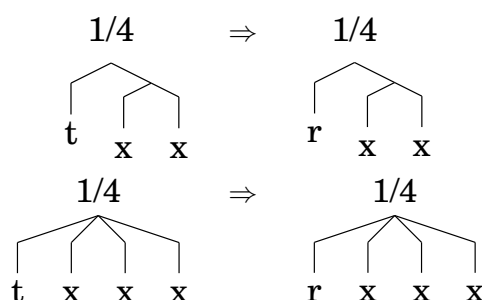


FIGURE 4.12

1219 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
1220 manquants seront mis en annexe.

1221 Les règles remplacent par un silence les continuations (t) qui sont au
1222 début d'un temps. Cela est valable pour ce système mais lorsqu'il y a des
1223 ouvertures de charley, cela n'est pas toujours applicable. Ce problème est
1224 évoqué de le chapitre 3.

1225 ⇒ **Objectif de cette expérimentation théorique :**

1226 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
1227 expérimentation théorique a pour but d'orienter la recherche d'autres
1228 systèmes par observation du jeu de données et de montrer comment les
1229 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

1230 4.4 Résultats et discussion

1231

1232 Cette section regroupe les avancées qui ont été réalisées par rapport aux
1233 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
1234 tats de l'ADT avec Qparse. Nous avons améliorer le système de quantifi-
1235 cation de Qparse pour la batterie, notamment le passage à la polyphonie
1236 avec les Jams.

1237 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
1238 grammaires avec des fichiers MIDI courts. Puis, une sortie MEI a été
1239 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

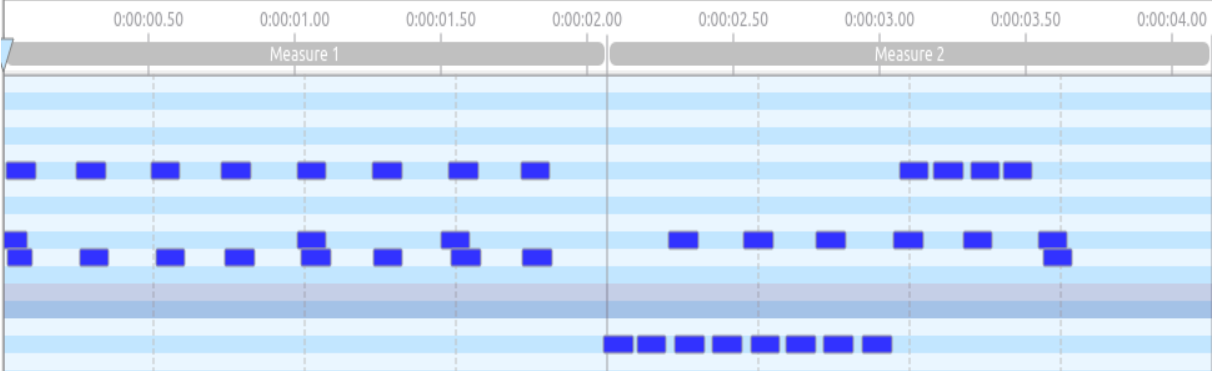
1240

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 :

1247

// bar level

1248

0 -> C0 1

1249

0 -> E1 1

1250

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

1251

1252

// half bar level

1253

9 -> C0 1

1254

9 -> E1 1

1255

1256

// beat level

1257

1 -> C0 1

1258

1 -> E1 1

1259

1 -> T2(2, 2) 1

1260

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

1261

1262

// croche level

1263

2 -> C0 1

1264

2 -> E1 1

1265

1266

// double level

1267

4 -> C0 1

1268

4 -> E1 1

1269

4 -> E2 1

1270

4 -> T2(6, 6) 1

1271

1272

1273 // triple level

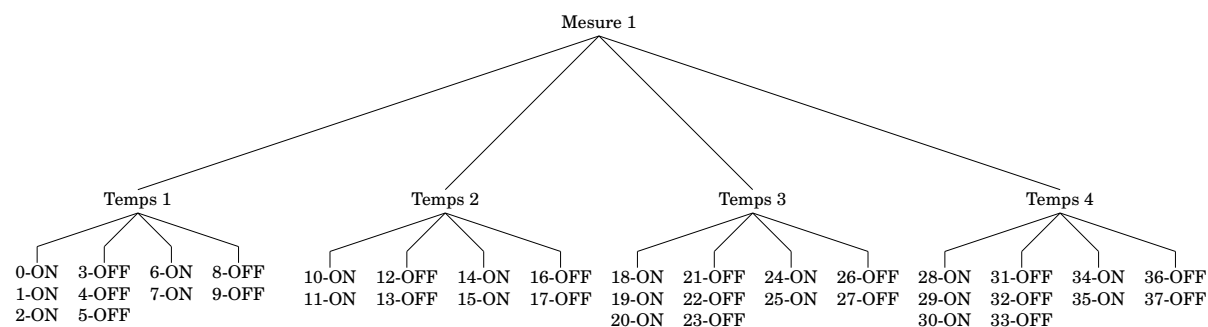
1274 6 -> E1 1

1275

1276 Cette grammaire sépare les ligatures par temps au niveau de la
 1277 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1278 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1279 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1280 il découpe correctement les mesures et les temps.

1281

1282



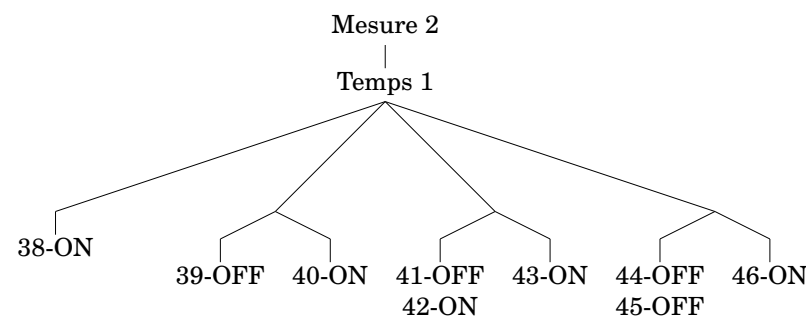
1283

1284

1285 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1286 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1287 quantification visibles dès le premier temps.

1288

1289



1290

1291

1292 Les Onsets sont correctement triés au niveau des doubles croches
 1293 mais certaines doubles croches sont inutilement subdivisées en triples
 1294 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1295 ci-dessus).

1296

1297 **2ème exemple :**

1298 Après une augmentation du poids des triples croches dans la grammaire
 1299 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1300 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1301 sont bien triés et aucuns ne sont subdivisés.

1302 **Évaluation**

1303 Pour l'évaluation, il aurait fallu produire un module.
1304 L'évaluation est-elle automatique ou manuelle?
1305 Possibilité d'un export lilypond en arbre pour comparer l'output avec la
1306 transcription manuelle.
1307 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel
1308 ou MEI pour les comparer et faire des statistiques. Si transformés en
1309 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)
1310 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.
1311 L'expérimentation peut-être considérer comme une évaluation manuelle?
1312 (magicien d'Oz)
1313 Lilypond vers MIDI + output vers MIDI \Rightarrow Comparaison des MIDI
1314 dumpés.
1315

1316 **Discussion**

1317 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
1318 choix qui ont été faits. Nous ferons un bilan des différents avancés qui
1319 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.
1320 Écrire des règles de réécriture spécifique aux charley avec un système
1321 approprié. Le jeu de système
1322 — implémenter un pattern...
1323 \Rightarrow manque de temps?
1324
1325 — La partie résultat est manquante car :
1326 \Rightarrow Sujet très difficile ;
1327 \Rightarrow Matcher les motifs peut être fait ultérieurement ;
1328 Mais ce travail aurait été indispensable pour obtenir une quan-
1329 tité de résultats qui justifieraient une évaluation automatique
1330 permettant de faire des graphiques.
1331
1332 — L'évaluation fut entièrement manuelle car :
1333 \Rightarrow Très dure automatiquement : il faut comparer 2 partitions (réf
1334 VS output)
1335 — Le ternaire jazz (voir expérience 2)
1336 — Reconnaissance d'un motif sur le MIDI
1337 Reconnaître un motif (système) sur une mesure de l'input (un fi-
1338 chier midi représentant des données audios)
1339 \Rightarrow Motif (système) reconnu : true ou false
1340 Si true :
1341 - Choisir la grammaire correspondante ;
1342 - Parser le MIDI ;

- 1343 - Appliquer les règles de réécritures (Séparation des voix et simpli-
1344 fication)
- 1345 — Nous travaillerons aussi sur la détection de répétitions sur plu-
1346 sieurs mesures afin de pouvoir corriger des erreurs sur une des
1347 mesures qui aurait dû être identique aux autres mais qui présente
1348 des différences.
- 1349 — dans quelle catégorie mettre le shuffle ?
- 1350 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
1351 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
1352 à la batterie a été générée. Elle sera un bon point de départ pour les
1353 travaux futurs dont plusieurs propositions sont énoncés dans le présent
1354 document.

CONCLUSION GÉNÉRALE

Dans ce mémoire, nous avons traité de la problématique de la transcription automatique de la batterie. Son objectif était de transcrire, à partir de leur représentation symbolique MIDI, des performances de batteur de différents niveaux et dans différents styles en partitions écrites. Nous avons avancé sur le parsing des données MIDI établissant un processus de regroupement des événements MIDI qui nous a permis de faire la transition du monophonique vers le polyphonique. Une des données importante de ce processus était de différencier les nature des notes d'un *accord*, notamment de distinguer lorsque 2 notes constituent un *accord* ou un *fla*. Nous avons établis des *grammaires pondérées* pour le parsing qui correspondent respectivement à des métriques spécifiques. Celles-ci étant sélectionnables en amont du parsing, soit par indication des noms des fichiers MIDI, soit par reconnaissance de la métrique avec une approche dictionnaire de patterns prédéfinis⁵ qu'il serait pertinent de mettre en œuvre en machine learning. Nous avons démontré que l'usage des *systèmes* élimine un grand nombre de calcul lors de la réécriture. Pour la séparation des voix grâce au motif d'un système et pour la simplification grâce aux gammes du motif d'un système. Nous avons aussi montré comment, dans des travaux futurs, un système dont le motif serait reconnu en amont dans un fichier MIDI pourrait prédéfinir le choix d'une grammaire par la reconnaissance d'une métrique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans la chaîne de traitement en terme de réécriture. Il sera également intéressant d'étudier comment l'utilisation de LM peut améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération entièrement automatisée de partitions de batterie et au problème général de l'AMT de bout en bout.[11]

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

BIBLIOGRAPHIE

- 1385 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1386 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1387 1996. – Cité pages 7, 30 et 36.
- 1388 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1389 et 15.
- 1390 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1391 page 12.
- 1392 [4] Gaël Richard et al. De fourier à la reconnaissance
1393 musicale. Available at [https://interstices.info/
1394 de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1395 – Cité page 12.
- 1396 [5] Caroline Traube. Quelle place pour la science au sein de la musico-
1397 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1398 [6] Leonard Bernstein Office. The unanswered question : Six talks at
1399 harvard. Available at [https://leonardbernstein.com/about/
1400 educator/norton-lectures](https://leonardbernstein.com/about/educator/norton-lectures) (2021/01/01). – Cité page 12.
- 1401 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1402 entre les traitements de la musique et du langage. *La Lettre des
1403 Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1404 [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1405 Techniques de traitement automatique du langage naturel appli-
1406 quées aux représentations symboliques musicales. In *JIM 2021 -
1407 Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1408 Cité page 13.
- 1409 [9] Peter Wunderli. Ferdinand de saussure : La sémiologie et les sémio-
1410 logies. *Semiotica*, 2017(217) :135–146, 2017. – Cité page 13.
- 1411 [10] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1412 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1413 Cité page 13.
- 1414 [11] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1415 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-

- 1416 lenges and future directions. *Journal of Intelligent Information Sys-*
 1417 *tems*, 41, 12 2013. – Cité pages 14, 15, 21, 22 et 63.
- 1418 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
 1419 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
 1420 Cité pages 14 et 27.
- 1421 [13] Georges Paczynski. *Une histoire de la batterie de jazz*. OUTRE ME-
 1422 SURE, 1997. – Cité page 15.
- 1423 [14] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
 1424 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
 1425 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
 1426 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
 1427 1483, 2018. – Cité pages 15, 23 et 27.
- 1428 [15] Moshekwa Malatji. Automatic music transcription for two instru-
 1429 ments based variable q-transform and deep learning methods, 10
 1430 2020. – Cité page 22.
- 1431 [16] Antti J. Eronen. Musical instrument recognition using ica-based
 1432 transform of features and discriminatively trained hmms. *Seventh*
 1433 *International Symposium on Signal Processing and Its Applications*,
 1434 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 24.
- 1435 [17] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
 1436 drum sound description for real-world music using template adap-
 1437 tation and matching methods. *International Conference on Music*
 1438 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 24.
- 1439 [18] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
 1440 musical statistics as guides for audio-to-score piano transcription.
 1441 *Information Sciences*, 566 :262–280, 2021. – Cité pages 24 et 26.
- 1442 [19] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
 1443 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
 1444 tization and Score Structuring. In *MCM 2019 - Mathematics and*
 1445 *Computation in Music*, volume Lecture Notes in Computer Science
 1446 of *Proceedings of the Seventh International Conference on Mathema-*
 1447 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
 1448 2019. Springer. – Cité pages 24 et 26.
- 1449 [20] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
 1450 ring of rhythm structures. In *Proceedings of the First International*
 1451 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
 1452 IEEE Computer Society. – Cité page 26.
- 1453 [21] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
 1454 Term Rewriting Based Structural Theory of Rhythm Notation. Re-

- 1455 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1456 Cité page 26.
- 1457 [22] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1458 equivalent rhythmic notations based on rhythm tree languages. In
1459 *Third International Conference on Technologies for Music Notation*
1460 *and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1461 pez Palma and Mike Solomon. – Cité page 26.
- 1462 [23] R. Marxer and J. Janer. Study of regularizations and constraints in
1463 nmf-based drums monaural separation. In *International Conference*
1464 *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1465 02/09/2013 2013. – Cité page 27.
- 1466 [24] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1467 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 30.
- 1468 [25] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1469 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 31.
- 1470 [26] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1471 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1472 94120 Fontenay-sous-Bois, 2000. – Cité pages 32 et 43.
- 1473 [27] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1474 Bamman. Learning to groove with inverse sequence transforma-
1475 tions. In *International Conference on Machine Learning (ICML)*,
1476 2019. – Cité page 47.

