

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 TAL et MIR	11
23	1.2 La transcription automatique de la musique	13
24	1.3 La transcription automatique de la batterie	15
25	1.4 Les représentations de la musique	16
26	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	23
30	2.4 Approche linéaire et approche hiérarchique	23
31	3 Méthodes	27
32	3.1 La notation de la batterie	27
33	3.2 Modélisation pour la transcription	34
34	3.3 Qparse	35
35	3.4 Les systèmes	37
36	4 Expérimentations	45
37	4.1 Le jeu de données	45
38	4.2 Analyse MIDI-Audio	47
39	4.3 Expérimentation théorique d'un système	51
40	4.4 Résultats et discussion	55
41	Conclusion générale	61
42	Bibliographie	63

LISTE DES FIGURES

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

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

81

LISTE DES TABLEAUX

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

INTRODUCTION GÉNÉRALE

86 QUOI?

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

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

108

109 POURQUOI?

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

116

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

119 COMMENT?

120 → Problématique :

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

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

131

132 → Méthodes :

133 → Contributions :

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

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

144

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

- 146 — une modélisation de la notation de la batterie (fusion de 3.1 et de
147 3.2) qui était jusqu'à présent inexistante.
- 148 — plusieurs transcriptions manuelles dans le but d'analyser les conte-
149 nus des fichiers MIDI et Audio (4.2) et de faire des comparaisons
150 de transcription avec des outils déjà existants¹.
- 151 — une partition de référence transcrite manuellement sur l'entièreté
152 d'une performance du jeu de données afin de repérer les éléments
153 importants pour la modélisation et de faire les liens entre les cri-
154 tères des données d'*input* avec l'écriture finale (4.2). Cette partition
155 avait aussi pour objectif d'effectuer des tests et des évaluations.
- 156 — le passage au polyphonique en théorie et en implémentation im-
157 pliquant la théorie sur la détection de l'identité de notes dans un
158 Jam² et l'implémentation de tests unitaires sur le traitement des
159 Jams (4.4).
- 160 — la création de grammaires pondérées spécifiques à la batterie (4.4)

161

1. MuseScore3

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

162 L'ensemble de ces sous-tâches a permis deux réalisations principales :
163 1) Obtenir des arbres de rythmes corrects en *output* de *qparse* avec des
164 exemples courts proches de la partition de référence.
165 2) La création d'une expérimentation théorique d'un système 4.3 dont
166 le but premier est de démontrer qu'elle est implémentable et applicable
167 à d'autres type de rythmes et dont le second objectif est de donner une
168 méthode de création d'un système à partir d'une partition.
169 Ces deux réalisations recouvrent une partie du chemin à parcourir
170 puisque pour effectuer des évaluations conséquentes sur résultat, la
171 chaîne de traitement doit être finie afin de pouvoir vérifier de manière
172 empirique que les systèmes, qui constituent ma contribution principale
173 pour ce mémoire, ont permis d'améliorer *qparse* pour la transcription
174 automatique de la batterie.

175

176 PLAN

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

CONTEXTE

Sommaire

1.1	TAL et MIR	11
1.2	La transcription automatique de la musique	13
1.3	La transcription automatique de la batterie	15
1.4	Les représentations de la musique	16

Introduction

La transcription automatique de la musique (AMT) est un défi ancien [2] et difficile qui n'est toujours pas résolu. Il a engendré une pluie de sous-tâches qui ont donné naissance au domaine de la recherche d'information musicale (MIR). Actuellement, de nombreux travaux de MIR font appel au traitement automatique des langues (TAL)¹.

cf. ismir.net

Dans ce chapitre, nous parlerons de l'informatique musicale, nous tenterons d'établir les liens existants entre le MIR et le TAL ainsi qu'entre les notions de langage musical et langue naturelle. Nous traiterons également de l'utilité et du problème de l'AMT et de la transcription automatique de la batterie (ADT).

du problème de l'AMT, de ses applications

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

1.1 TAL et MIR

L'informatique musicale [3] est une étude du traitement de la musique [4], en particulier des représentations musicales, de la transformée de Fourier pour la musique [5], de l'analyse de la structure de la musique

Je référerais plutôt à "Computer Music" : utilisation de méthodes numériques pour l'analyse et la synthèse de musique, qu'il s'agisse d'information audio ou symbolique (aide à l'écriture, transcription, base de partitions...) de musique

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

vaste champ de recherche pluridisciplinaire, à l'intersection de l'acoustique, du signal, de la synthèse sonore, de l'informatique, des sciences cognitives, des neurosciences, de la muscologie...

On peut citer : ...

? psycho-acoustique, neurosciences ?

sujet : la recherche et l'extraction d'information à partir de données musicales.

et de la reconnaissance des accords². D'autres sujets de recherche en informatique musicale comprennent la modélisation informatique de la musique, l'analyse informatique de la musique, la reconnaissance optique de la musique, les éditeurs audio numériques, les moteurs de recherche de musique en ligne, la recherche d'informations musicales et les questions cognitives dans la musique.

Le MIR³ apparaît vers le début des années 2000 [6]. C'est une science interdisciplinaire qui fait appel à de nombreux domaines comme la musicologie, l'analyse musicale, la psychologie, les sciences de l'information, le traitement du signal et les méthodes d'apprentissage automatisé en informatique. Cette discipline récente a notamment été soutenue par de grandes compagnies du web^{4 5 6} qui veulent développer des systèmes de recommandation de musique ou des moteurs de recherche dédiés au son et à la musique.

Is Music a Language?



Leonard Bernstein

Norton Lectures at Harvard, 1973

« The Unanswered Question: Six Talks at Harvard »

idea of music as a kind of universal language

notion of a worldwide, « inborn musical grammar »

cf. **Noam Chomsky** « Language and Mind »

theory of innate grammatical competence

ne pas inclure ce slide, car l'argument de Bernstein et Chomsky est insuffisant

Aborder la musique à travers le TAL nécessite une réflexion autour de la musique en tant que langage ainsi que la possibilité de comparer ce même

2. 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 successives

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

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

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

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

langage avec les langues naturelles. Quelques travaux en neurosciences ont abordé la question, notamment par observation des processus cognitifs et neuronaux que les systèmes de traitement de ces deux langages avaient en commun. Dans le travail de Poulin-Charronnat *et al.* [7], la musique est reconnue comme étant un système complexe spécifique à l'être humain dont une des similitudes avec les langues naturelles est l'émergence de régularités reconnues implicitement par le système cognitif. La question de la pertinence de l'analogie entre langues naturelles et langage musical a également été soulevée à l'occasion de projets de recherche en TAL. Keller *et al.* [8] ont exploré le potentiel de ces techniques à travers les plongements de mots et le mécanisme d'attention pour la modélisation de données musicales. La question du sens d'une phrase musicale apparaît, selon eux, à la fois comme une limite et un défi majeur pour l'étude de cette analogie.

D'autres travaux très récents, ont aussi été révélés lors de la *première conférence sur le NLP pour la musique et l'audio (NLP4MusA 2020)*. Lors de cette conférence, Jiang *et al.* [9] ont présenté leur implémentation d'un modèle de langage musical auto-attentif visant à améliorer le mécanisme d'attention par élément, déjà très largement utilisé dans les modèles de séquence modernes pour le texte et la musique.

Le domaine du TAL qui se rapproche le plus du MIR est la reconnaissance de la parole (Speech to text). En effet, la séparation des sources ont des approches similaires dans les deux domaines. De plus, il existe un lien entre partition musicale comme manière d'écrire la musique et texte comme manière d'écrire la parole.

on cite souvent la sémiotique (F. de Saussure) dans ce contexte.

exemples / illustration de la proximité thématique?

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

TABLE 1.1 – speechToText vs AMT

1.2 La transcription automatique de la musique

En musique, la transcription⁷ est la pratique consistant à noter un morceau ou un son qui n'était auparavant pas noté et/ou pas populaire en tant que musique écrite, par exemple, une improvisation de jazz ou une bande sonore de jeu vidéo. Lorsqu'un musicien est chargé de créer une partition à partir d'un enregistrement et qu'il écrit les notes qui composent le mor-

objectifs similaires sur le papier : speech-to-text, problèmes et applications aussi comparables : transcription, synthèse, séparation de sources... Mais information de nature différente cf. sous-tâches comme beat tracking et inférence de tempo en musique.

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

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

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

7. [https://en.wikipedia.org/wiki/Transcription_\(music\)](https://en.wikipedia.org/wiki/Transcription_(music))

ceau en notation musicale, on dit qu'il a créé une transcription musicale de cet enregistrement.

L'objectif de la transcription automatique de la musique (AMT) [10] est de convertir la performance d'un musicien en notation musicale - un peu comme la conversion de la parole en texte dans le traitement du langage naturel. L'AMT a des intérêt multiples, notamment pour la transcription de solos ou encore pour la constitution de corpus musicologiques, ou encore pour l'interprétation de la musique et l'analyse du contenu musical [11]. Par exemple, un grand nombre de fichiers audio et vidéo musicaux sont disponibles sur le Web, et pour la plupart d'entre eux, il est difficile de trouver les partitions musicales correspondantes, qui sont nécessaires pour pratiquer la musique, faire des reprises et effectuer une analyse musicale détaillée. Les partitions de musique classique sont facilement accessibles et il y a peu de demandes de nouvelles transcriptions. D'un point de vue pratique, des demandes beaucoup plus commerciales et académiques sont attendues dans le domaine de la musique populaire [11]. Les modèles grammaticaux qui représentent la structure hiérarchique des séquences d'accords se sont avérés très utiles dans les analyses récentes de l'harmonie du jazz [12]. Comme déjà évoqué précédemment, il s'agit d'un problème ancien et difficile. C'est un « graal » de l'informatique musicale. En 1976, H. C. Longuet-Higgins [2] évoquait déjà la représentation musicale en arbre syntaxique dans le but d'écrire automatiquement des partitions à partir de données audio en se basant sur un mimétisme psychologique de l'approche humaine. De même pour les chercheurs en audio James A. Moorer, Martin Piszczalski et Bernard Galler qui, en 1977⁸, ont utilisé leurs connaissances en ingénierie de l'audio et du numérique pour programmer un ordinateur afin de lui faire analyser un enregistrement musical numérique de manière à détecter les lignes mélodiques, les accords et les accents rythmiques des instruments à percussion.

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

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

- La séparation des sources à partir de l'audio.
- Le système de transcription :
 - Cœur du système :
 - ⇒ Algorithmes de détection des multi-pitches et de suivi des notes.

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

à l'instar de la 270

applications 271

préservation du patrimoine 273

e.g. musique de tradition 274

orale (ethno-musicologie) 275

citer un survey pour les 276

applications (pas [11]) 277

pas d'accord avec ça. pro-280
blème des partitions libres
de droit. 281

l'intérêt est aussi 282
d'avoir des partitions 283
au contenu exploitable 284
(texte ou XML) vs images
(pdf...) cf. par ex. cette 285
présentation d'OpenScore
à FOSDEM [https://ar-](https://archive.fosdem.org/2017/schedule/event/openscore/) 286
[chive.fosdem.org/2017/schedule/event/openscore/](https://archive.fosdem.org/2017/schedule/event/openscore/)
et mes transparents sur le 287
sujet 288

là on passe aux approches 289
scientifiques 290

quel rapport ? 291

la figure ne correspond pas 290
à ton travail. ici "score" = 291
MIDI performance. Tu peux 301
lister les sous-tâches en 302
section 2.2 302

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

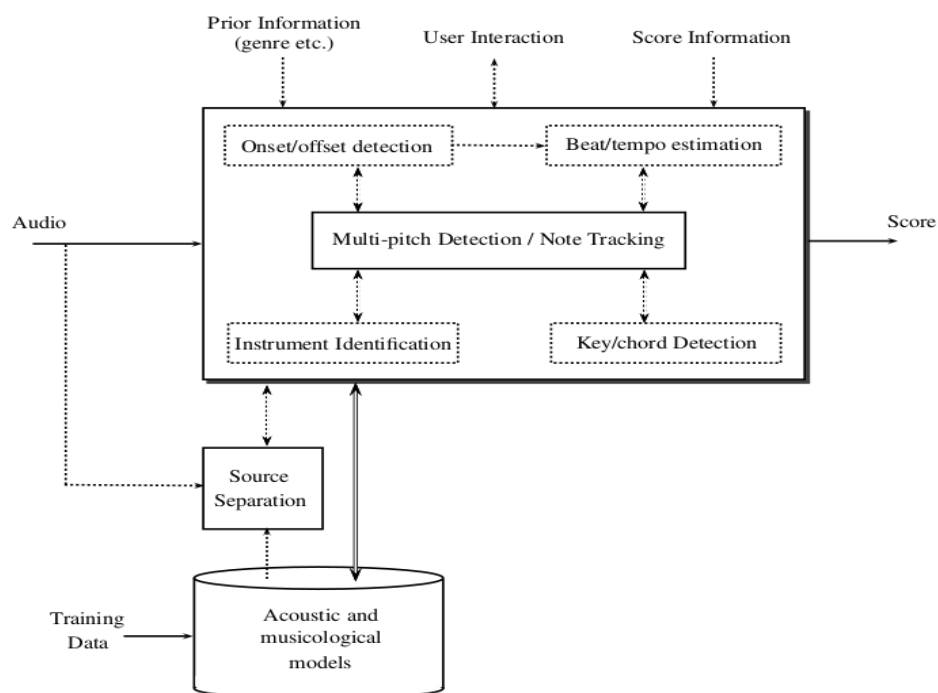


FIGURE 1.1 – Transcription automatique

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

318

éviter newpage

1.3 La transcription automatique de la batterie

319

320

tres bonne section

La batterie est un instrument récent qui s'est longtemps passé de partition. En effet pour un batteur, la qualité de lecteur lorsqu'elle était nécessaire, résidait essentiellement dans sa capacité à lire les partitions des

323

324 autres instrumentistes (par exemple, les grilles d'accords et la mélodie
325 du thème en jazz) afin d'improviser un accompagnement approprié que
326 personne ne pouvait écrire pour lui à sa place.

327 Les partitions de batterie sont arrivées par nécessité avec la pédagogie
328 et l'émergence d'écoles de batterie partout dans le monde. Un autre fac-
329 teur qui a contribué à l'expansion des partitions de batterie est l'émer-
330 gence de la musique assistée par ordinateur (MAO). En effet, l'usage de
331 boîtes à rythmes⁹ ou de séquenceurs¹⁰ permettant d'expérimenter soi-
332 même l'écriture de rythmes en les écoutant mixés avec d'autres instru-
333 ments sur des machines a permis aux compositeurs de s'émanciper de la
334 création d'un batteur en lui fournissant une partition contenant les par-
335 ties exactes qu'ils voulaient entendre sur leur musique.

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

340 Les applications de l'ADT seraient utiles, non seulement dans tous les
341 domaines musicaux contenant de la batterie dont certains manquent de
342 partitions, notamment les musiques d'improvisation (jazz, pop) [10], mais
343 aussi de manière plus générale dans le domaine du MIR : si les ordina-
344 teurs étaient capables d'analyser la partie de la batterie dans la musique
345 enregistrée, cela permettrait une variété de tâches de traitement de la
346 musique liées au rythme. En particulier, la détection et la classification
347 des événements sonores de la batterie par des méthodes informatiques
348 est considérée comme un problème de recherche important et stimulant
349 dans le domaine plus large de la recherche d'informations musicales [13].
350 L'ADT est un sujet de recherche crucial pour la compréhension des aspects
351 rythmiques de la musique, et a un impact potentiel sur des domaines plus
352 larges tels que l'éducation musicale et la production musicale.

353 1.4 Les représentations de la musique

355 Les données audio

356 Le fichier WAV¹¹ est une instance du Resource Interchange File Format
357 (RIFF) défini par IBM et Microsoft. Le format RIFF agit comme une "en-
358 veloppe" pour divers formats de codage audio. Bien qu'un fichier WAV
359 puisse contenir de l'audio compressé, le format audio WAV le plus cou-
360 rant est l'audio non compressé au format LPCM (linear pulse-code modu-
361 lation). Le LPCM est également le format de codage audio standard des

9. Roland TR-808

10. SQ-1

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

cite méthode et école Ago
tini?

citer [13] ici

ADT pas défini

"contenant" -> concernés
par

permettrait de faciliter

citer M. Müller FMP pou
cette section?

trop technique. ne pas re
pier wikipédia

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

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

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

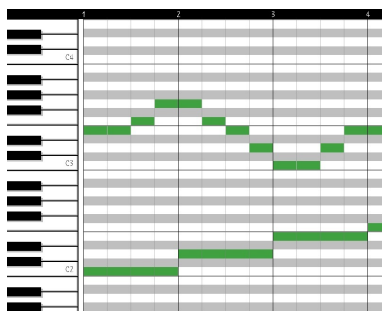
368 Les données MIDI

369 Le MIDI¹² (Musical Instrument Digital Interface) est une norme tech-
 370 nique qui décrit un protocole de communication, une interface numérique
 371 et des connecteurs électriques permettant de connecter une grande va-
 372 riété d'instruments de musique électroniques, d'ordinateurs et d'appareils
 373 audio connexes pour jouer, éditer et enregistrer de la musique.

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

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

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



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

FIGURE 1.2 – Exemple événements avec durée

377

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

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

FIGURE 1.3 – Critère pour un événement

379

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

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

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

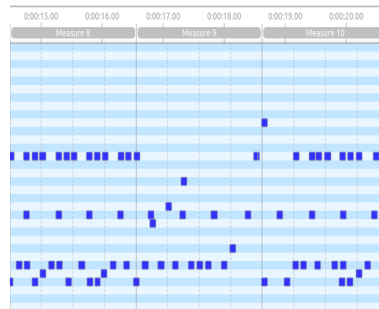


FIGURE 1.4 – Exemple évènements sans durée

384
 385 On observe que toutes les durées sont identiques.

386 Les partitions



FIGURE 1.5 – Exemple de partition de piano

387 Une partition de musique¹³ est un document qui porte la représentation
 388 systématique du langage musical sous forme écrite. Cette représentation
 389 est appelée transcription et elle sert à traduire les quatre caractéristiques
 390 du son musical :

- 391 — la hauteur ;
- 392 — la durée ;
- 393 — l'intensité ;
- 394 — le timbre.

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

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

- 397 — diachronique (succession des instants, ce qui constitue en musique
 398 la mélodie) ;
- 399

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

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

Le format MusicXML

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

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

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

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

standard W3C = MNX (en cours)

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



FIGURE 1.6 – MusicXML

Le figure 1.6¹⁴ représente un do en clef de sol de la durée d'une ronde sur une mesure en 4/4.

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

Conclusion

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

Nous avons également établi que le MIR est né de l'AMT qui est un problème ancien et très difficile et qu'il serait toujours très utile de le ré-

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

420 soudre (autant pour l'AMT que pour l'ADT).
421 Et enfin, nous avons décrit les représentations de la musique nécessaires
422 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	23
2.4	Approche linéaire et approche hiérarchique	23

Introduction

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

présenterons quelques travaux antérieurs

Nous aborderons le passage crucial du monophonique au polyphonique dans la transcription. Nous ferons un point sur les deux grandes parties de l'AMT 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 [10]. 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 du signal vers la génération du MIDI [14].

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

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

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

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

de signaux audio

MIDI non-quantifié = performance (à expliquer)

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

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

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

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

haute fréquence, aigus ?

classification des événements ? la phrase semble redondante

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

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

494 2.3 MIDI vers partition

495
 496 Le plus souvent, lorsque les articles abordent la transcription automa-
 497 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »
 498 (partition) désigne un ouput au format Music XML, ou simplement MIDI.
 499 Par exemple, dans [11], la chaîne de traitement va jusqu'à la génération
 500 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en
 501 extraire manuellement un fichier MusicXML contenant plusieurs voix.
 502 Seuls quelques travaux récents s'intéressent de près à la création d'outils
 503 permettant la génération de partition. Le problème de la conversion d'une
 504 séquence d'évènements musicaux symboliques en une partition musicale
 505 structurée est traité notamment dans [17]. Ce travail, qui vise à résoudre
 506 en une fois la quantification rythmique et la production de partition struc-
 507 turée, s'appuie tout au long du processus sur des grammaires génératives
 508 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-
 509 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été
 510 menées avec un ensemble de données composé d'extraits monophoniques ;
 511 il reste donc à traiter le passage au polyphonique, en couplant le problème
 512 de la séparation des voix avec la quantification du rythme.
 513 L'approche de [17] est fondée sur la conviction que la complexité de la
 514 structure musicale dépasse les modèles linéaires.

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

de manière conjointe

langage *a priori*

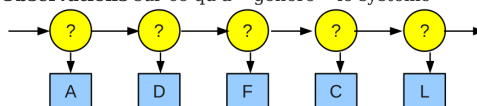
qui nécessite de traiter le
problème supplémentaire
de la séparation de voix. i.e.
pour la batterie on nveut
quantification + structu-
ration + séparation mais
seules les 2 premières sont
couplées dans l'approche de
tonn stage.

515 2.4 Approche linéaire et approche hiérarchique

516 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par
 517 exemple, Shibata *et al.* [11] ont utilisé le modèle de Markov caché (HMM)⁴
 518 pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux
 519 réseaux de neurones profonds, l'un pour la reconnaissance des *pitchs* et
 520 l'autre pour la reconnaissance de la *vélocité*. Pour la dernière couche, la
 521 probabilité est obtenue par une fonction sigmoïde. Ils construisent en-
 522 suite plusieurs HMM métriques étendus pour la musique polyphonique
 523 correspondant à des métriques possibles, puis ils calculent la probabilité
 524 maximale pour chaque modèle afin d'obtenir la métrique la plus probable.

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

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



- Applications : physique, reconnaissance de parole, traitement du langage, bio-informatique, finance, etc.

FIGURE 2.1 – HMM

525 *Source : Cours de Damien Nouvel*⁵

526

527

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

je ne comprend pas bien 535
 l'explication. le pb est plu- 536
 tot vue locale (déduction de 537
 la proba d'une durée à par- 538
 tir de la durée précédente, 539
 par ex. dans un HMM) vs
 vue globale, dans une hié-
 rarchie

RT? 540

techniques de réécriture 542
 appliquée à la déduction 543
 automatique, calcul symbolique 544

le calcul d'équiv. 545

546

547

548

citer thèse de David Rizo 549
 (Valencia) 550

551

le langage musical est lui-même structuré. En effet, l'usage d'arbres syn-
 taxiques est idéale pour représenter le langage musical. Une méthodolo-
 gie simple pour la description et l'affichage des structures musicales est
 présentée dans [18]. Les RT y sont évoqués comme permettant une co-
 hésion complète de la notation musicale traditionnelle avec des notations
 plus complexes. Jacquemard *et al.* [19] propose aussi une représentation
 formelle du rythme, inspirée de modèles théoriques antérieurs issus du
 domaine de la réécriture de termes. Ils démontrent aussi l'application des
 arbres de rythmes pour les équivalences rythmiques dans [20]. La réécri-
 ture d'arbres, dans un contexte de composition assistée par ordinateur,
 par exemple, pourrait permettre de suggérer à un utilisateur diverses
 notations possibles pour une valeur rythmique, avec des complexités dif-
 férentes.

La nécessité d'une approche hiérarchique pour la production automatique
 de partition est évoquée dans [17]. Les modèles de grammaire qui y sont
 exposés sont différents de modèles markoviens linéaires de précédents
 travaux.

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

Example: *Summertime*

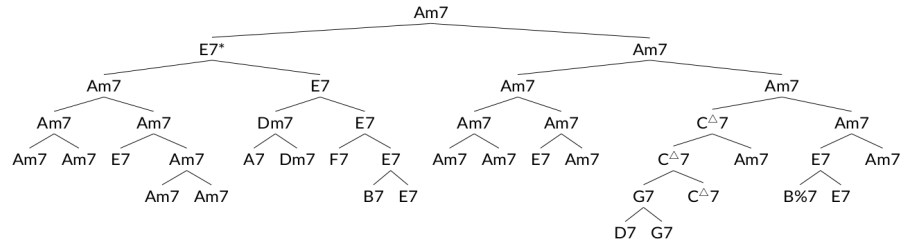


FIGURE 2.2 – arbre_jazz
Représentation arborescente d’une grille harmonique [12]

Conclusion

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

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

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

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

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

latex : enumerate

573

574

MÉTHODES

575

Sommaire

576	3.1	La notation de la batterie	27
577	3.2	Modélisation pour la transcription	34
578	3.3	Qparse	35
579	3.4	Les systèmes	37
580			
581			
582			

584 Introduction

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

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

592 Enfin, nous présenterons les systèmes.

plusieurs développeurs

systèmes, une représenta-
tion théorique qui permet...

593 3.1 La notation de la batterie



FIGURE 3.1

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

596 — La ronde, elle vaut 4 ;

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

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

4 temps

- 597 — La blanche, elle vaut 2 ;
- 598 — La noire, elle vaut 1 ;
- 599 — La croche, elle vaut 1/2.

plusieurs éléments 600

plutôt que wikipedia cite
Dannhauser ou autre ref
F.M. ou encore Gould 2011
Behind Bars 603

barre verticale liée à la tête
de note 605

haut ou bas 607

- 601 — Une tête de note :
Sa position sur la portée indique la hauteur de la note. La tête de note peut aussi indiquer une durée.
- 602 — 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)
- 603 — Un crochet : La durée d'une note est divisée par deux à chaque crochet ajouté à la hampe d'une figure de note.

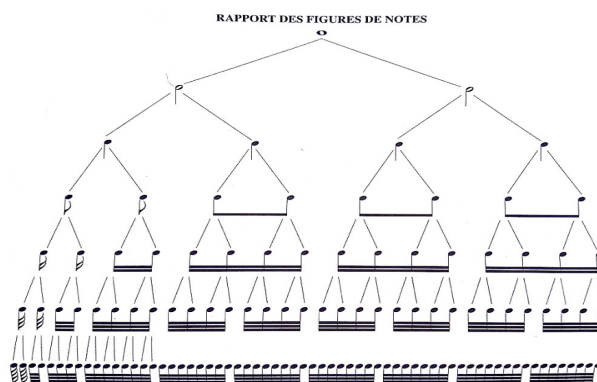


FIGURE 3.2 – Rapport des figures de notes
[1]

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

ce premier paragraphe (j'ai
qu'ici) est redondant avec
§1.4 (sub. partitions). dé-
placer en 1.4? cf. proposi-
tion plus loin 618

Les hauteurs et les têtes de notes

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

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

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

En effet, les hauteurs sur la portée représentent :

— La hauteur physique des instruments :

La caisse claire est centrale sur la portée et sur la batterie (au niveau de la ceinture, elle conditionne l'écart entre les pédales et aussi la position de tous les instruments basiques d'une batterie). Tout ce qui en-dessous de la caisse-claire sur la portée est en dessous de la caisse-claire sur la batterie (pédales, tom basse); Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est aussi sur la batterie.

— La hauteur des instruments en terme de fréquences :

Sauf pour le charley au pied et si l'on sépare en trois groupes (grosse-caisse, toms et cymbales), de bas en haut, les instruments vont du plus grave au plus aigu.

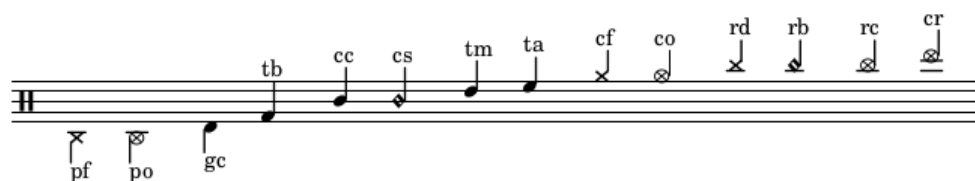


FIGURE 3.3 – Hauteur et têtes de notes

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

têtes de notes?

Les durées

Comme nous venons de la voir, la majorité des instruments de la batterie sont représentés par les têtes des notes. Par conséquent, les symboles rythmiques concernant la tête de note ne pourront pas être utilisés. Cela est valable aussi pour la présence ou non de la hampe puisque ce phénomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition blanche-ronde). L'usage des blanches existe dans certaines partitions de batterie [24] mais cela reste dans des cas très rares. Certains logiciels permettent de faire des blanches avec des symboles spécifiques à la batterie ou aux percussions mais leur lecture reste peu aisée et leur utilisation pour la batterie est rarissime.

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

- Le point;
- La liaison.

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

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

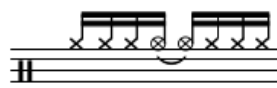
pour clarifier cela, tu pourrais décrire en 1.4 la notation conventionnelles (piano etc) et ici uniquement ce qui est spécifique à la batterie, en expliquant les différences.

expliquer comment, par ex. avec figure 3.4

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



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.4 – Point et liaison

= la position des temps 658

659

faire un "enumerate" 661

662

663

664

665

666

667

668

669

670

671

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

Exemple 1 : ouverture de charley quantifiée mais pas notes pas regroupées par temps.

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

Exemple 3 et exemple 4 : les deux exemples sont valables mais le deuxième est le plus souvent utilisé car plus intuitif (regroupement par temps).

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

672

Les silences

expliquer la notation (générale) des silences en §1.4? 673

674

quantifier = noter? ou quantifier la durée? 675

676

677

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

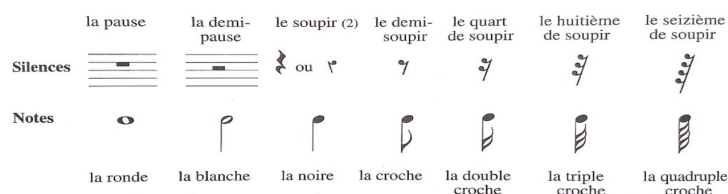


FIGURE 3.5 – Les silences

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

pas très clair

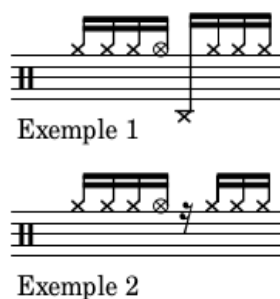


FIGURE 3.6 – Silence joué

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

itemize

cf?

691 Les équivalences rythmiques

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

phrase alambiquée... pour prolonger la durée?

seuls comptent les date de début de notes onsets.

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



FIGURE 3.7 – Équivalence

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

Les voix

Les voix³ désignent les différentes parties mélodiques constituant une composition musicale et destinées à être interprétées, simultanément ou successivement, par un ou plusieurs musiciens. En batterie, une voix est l'ensemble des instruments qui, à eux seuls, constituent une phrase rythmique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant possibles pour un même rythme, on peut regrouper les instruments de la batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou 2 voix. Sur la figure 3.8, il faudra faire un choix entre les exemples 1, 2 et 3 qui sont trois façons d'écrire le même rythme.

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

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

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

une voix est caractérisée aussi pas orientation des hampes?

Les accentuations et les ghost-notes

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

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

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

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



FIGURE 3.8 – Séparation des voix



FIGURE 3.9 – Les accents et les ghost-notes

730 ainsi, mais ces choix devront être adaptés en fonction des différents sys-
 731 tèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz,
 732 les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il
 733 faudra donc mettre les symboles d'accentuation correspondants au-dessus
 734 des têtes de notes.

735 La dernière note de la figure 3.9 montre un exemple de ghost-notes. Le
 736 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note
 737 sans changer la tête de note.

738 Pour les codes, on prend le code de la note et on ajoute un « a » pour un
 739 accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.9
 sont exposées en situation réelle dans la figure 3.10.

expliquer ce qu'est une
ghost-notes

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



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

3.2 Modélisation pour la transcription

Les pitches

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

TABLE 3.1 – Pitches et instruments

Il existe, pour de nombreux instruments de la batterie, plusieurs samples audio associés à des pitches. Pour cette première version, nous avons choisi de n’avoir qu’un code-instrument pour différentes variantes d’un instrument, c’est pourquoi 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. . .)

La vélocité

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

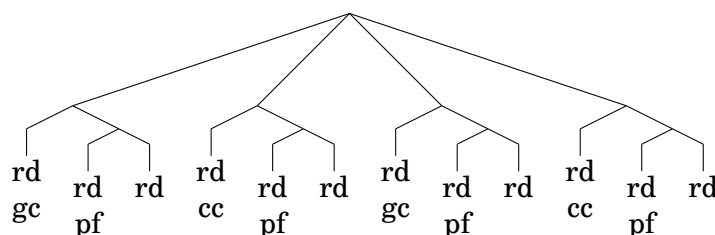
Cette transcription nous a mené aux observations suivantes :

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

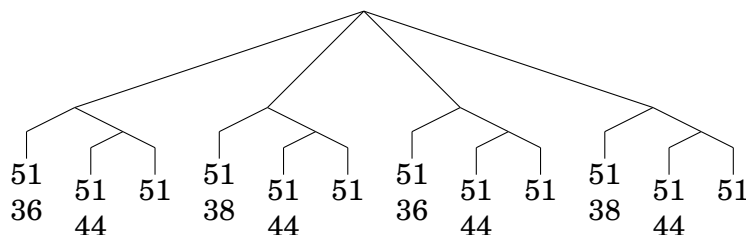
- Les accents et les ghosts-notes ne sont significatifs ni pour les instruments joués au pied, ni pour les cymbales crash.
- En effet, certaines vélocités en dessous de 40 étant détectées et inscrites dans les données MIDI sont dues au mouvement du talon du batteur qui bat la pulsation sans particulièrement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

Les arbres de rythmes

- Les arbres de rythmes représentent un rythme unique dont les possibilités de notation sur une partition sont théoriquement multiples.
- Voici une représentation de la figure 3.8 en arbre de rythmes avec les codes de chaque instrument :



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



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

3.3 Qparse

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

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

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

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

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

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

quantification rythmique + structuration de partition

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

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

grammaire \neq automate. 788
il faut choisir entre les 2 789
(pour la suite aussi)

apprentissage 793

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

Les principaux contributeurs sont :

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

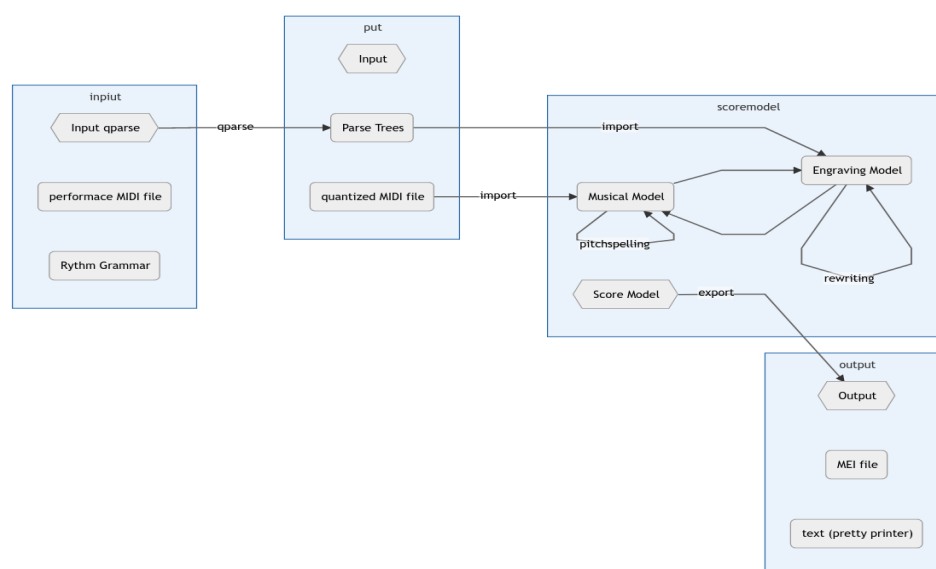


FIGURE 3.11 – Présentation de Qparse

la figure 3.11 est trop cor801
pliquée. rhythm grammar 802
→ automate d'arbres pon803
déré. Parse Tree → arbre804
syntaxique. qtz MIDI file :805
inutile. Score Model → r806
présentation intermédiaire807
de partition. Score Model
Engr. Model : inutile. gar-
der juste la fleche Rewri-
ting sur S.M.

Explication des différentes étapes de la figure 3.11⁵ :

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

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

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

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

Plusieurs enjeux :

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

3.4 Les systèmes

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

Définitions

Système : motif + gamme/texte

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

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

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

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

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

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

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

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

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

TABLE 3.2 – Sytèmes

850 dans 4 grandes catégories :

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

852 — 4/4 binaire

853 — 4/4 jazz

854 — 4/4 afro-cubain

855 **Objectif des systèmes**

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

857 — définir une métrique ;

858 — choisir une grammaire appropriée ;

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

860 tion.

bien. il faudrait expliquer
ça avant.

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

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

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

réécritures) seront extraites des voix séparées des systèmes.

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

867 **Détection d’indication de mesure**

868 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

869 temps et ses subdivisions.

métrique ≠ signature ryth-
mique (c’est plus général).
Il aurait fallu présenter
seulement la notation des
signatures rythmiques, par
exemple en 1.4



Exemple 1



Exemple 2

FIGURE 3.12 – Métrique

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

Choix d'une grammaire

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

Pour les expériences effectuées avec le Groove MIDI Data Set, le style et l'indication de mesure sont récupérables par les noms des fichiers MIDI, mais il faudra par la suite les trouver automatiquement sans autres indications que les données MIDI elles-mêmes. Par conséquent, les motifs des systèmes devront être recherchés sur l'input (*fichiers MIDI*) avant le lancement du parsing, afin de déterminer la métrique en amont. Cette tâche devra probablement être effectuée en Machine Learning.

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

Groove MIDI Data Set pas présenté

méta-données

contenu

Séparation des voix

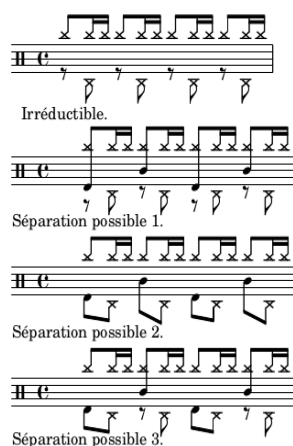


FIGURE 3.13 – Motif 4-4 binaire

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

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

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

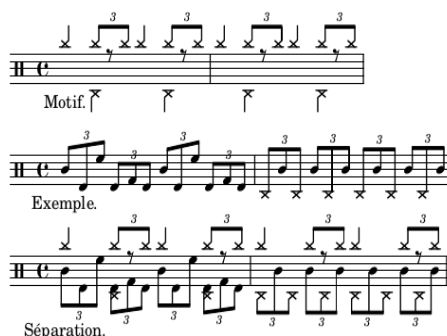


FIGURE 3.14 – Motif 4-4 jazz

902 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 903 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 904 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 905 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 906 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 907 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 908 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 909 et de caisse claire et son écriture séparée serait inutilement compliquée
 910 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 911 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de
 912 charley en bas.
 913

quel exemple?



FIGURE 3.15 – Système 4-4 afro-latin

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

Simplification de l'écriture

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

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

Voici les différentes étapes à suivre :

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

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

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

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

expérimentation théorique??

ce sont des figures et notations du chapitre suivant!

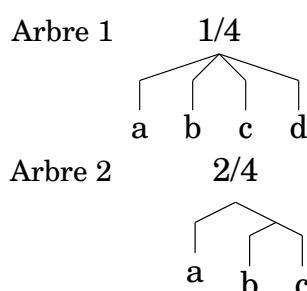


FIGURE 3.16 – Simplification

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

b et c sont des continuations (t) ;

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

- Si a n'est pas un co :

⇒ Suite 1a.

itemize

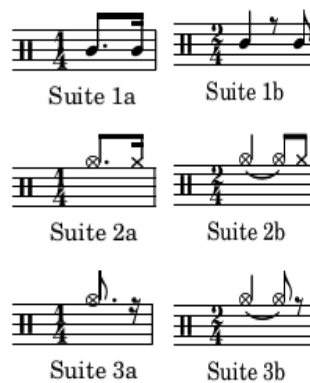


FIGURE 3.17

950 — Si a est un co :
 951 — Si d est un cf :
 952 ⇒ Suite 2a.
 953 — Si d est un pf :
 954 ⇒ Suite 3a : d deviens un silence (r).
 955

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

960 — Si a n'est pas un co :
 961 ⇒ Suite 1b, b devient un silence.
 962 — Si a est un co :
 963 — Si c est un cf :
 964 ⇒ Suite 2b, b devient une liaison et c devient un cf.
 965 — Si c est un pf :
 966 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
 967

968 *Rappel :*
 969 *cf = charley fermé joué à la main ;*
 970 *co = charley ouvert joué à la main ;*
 971 *pf = charley fermé joué au pied.*
 972

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

976 **Conclusion**

977 Nous avons formalisé une notation de la batterie, modélisé cette notation
978 pour la transcription de données MIDI en partition, nous avons décrit
979 Qparse.

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

EXPÉRIMENTATIONS

Sommaire

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

Introduction

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

4.1 Le jeu de données

Nous avons utilisé le Groove MIDI Dataset¹ [25] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).

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

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



FIGURE 4.1 – Batterie électronique

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

1012 Autres critères spécifiques au GMD :

- 1013 — Toutes les performances ont été jouées au métronome et à un tempo
- 1014 choisi par le batteur.
- 1015 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 1016 qui ont pu improviser dans un large éventail de styles. Les don-
- 1017 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 1018 (professionnel ou amateur).
- 1019 — Les batteurs avaient pour instruction de jouer des séquences de
- 1020 plusieurs minutes ainsi que des fills²
- 1021 — Chaque performance est annotée d'un style (fourni par le batteur),
- 1022 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 1023 nyme du batteur.
- 1024 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 1025 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 1026 session du GMD.
- 1027 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 1028 fichier MIDI.

1029 **Format des données**

enregistre les données dans
des fichiers MIDI

1030 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-
1031 tinctes :

- 1032 — une pour le tempo et l'indication de mesure ;
- 1033 — une pour les changements de contrôle (position de la pédale de
- 1034 charley) ;
- 1035 — une pour les notes.

1036

1037 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1038 canal 9 (qui est le canal canonique pour la batterie).

1039 Pour simplifier le traitement de ces données, ces trois pistes ont été
1040 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.

1041
 1042 « Control Changes The TD-11 also records control changes speci-
 1043 fying the position of the hi-hat pedal on each hit. We have preserved this
 1044 information under control 4. »
 1045 (<https://magenta.tensorflow.org/datasets/groove>)
 1046 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 1047 tions dans les fichiers MIDI.
 1048 L'utilisation de pretty_midi devient urgente!

1049 4.2 Analyse MIDI-Audio

1050
 1051 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
 1052 partir de fichiers MIDI et Audio du GMD.

= analyses et transcriptions
manuelles

1053 Comparaisons de transcriptions

1054 Pour les comparaisons de transcriptions, les transcriptions manuelles
 1055 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les trans-
 1056 criptions automatiques (TA) ont toutes été générées manuellement avec
 1057 MuseScore.

méthodologie tr. manuelle.
import MIDI pour MuseS-
core

1058 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1059 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1060 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1061 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1062 premières notes (une double-croche et une croche) d'un triolet sur
- 1063 le temps 1 de la mesure 2 de la TA.

1064 Exemple d'analyse 2

- 1065 — Les doubles croches ont été interprétées en quintolet
- 1066 — La deuxième double-croche est devenue une croche.

Transcription manuelle \Rightarrow Transcription automatique



Transcription manuelle \Rightarrow Transcription automatique

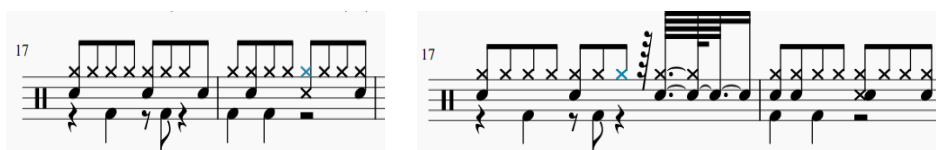


1067 Exemple d'analyse 3

- 1068 — Les grosses-caisses, les charleys et les caisses-claires ont été déca-
- 1069 lés d'un temps vers la droite.
- 1070 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1071 décalés d'une double croche vers la droite dans la TA.
- 1072 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1073 alors qu'elle appartenait à un triolet dans la TM.
- 1074 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1075 pas la TM.
- 1076

1077 Exemple d'analyse 4

Transcription manuelle \Rightarrow Transcription automatique



1078

- 1079 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- manière excessivement complexe!

conclusion sur ces exem

1081 Exemple avec des flas

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

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

1083 Transcription manuelle



1084

1085 Transcription automatique

1086



1087

1088

- 1089 — Le premier fla est reconnu comme étant un triolet contenant une
- 1090 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1091 ornementée.
- 1092 — Le deuxième fla est reconnu comme étant un accord.
- 1093 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1094 dans la TA.
- 1095 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1096 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1097 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1098 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1099 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1100 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1101 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1102 temps 3 de la TA.
- 1103 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1104 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1105 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1106 cross-sticks.

1107 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1108 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1109 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1110 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1111 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1112 [Inria](https://github.com/MartinDigard/Stage_M2_Inria)) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1113 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1114 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1115 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1116 et quatre).

1117 La transcription des données audio et MIDI contenues dans ces fichiers
 1118 a permis une analyse plus approfondie des critères à relever pour chaque
 1119 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 1121
 transcription manuelle?
 difficultés, durée? nb de
 passes... pourquoi LilyPon-
 det pas MuseScore?

4.3 Expérimentation théorique d'un système

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

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

Motifs et gammes



FIGURE 4.3 – Motifs et gammes

Motifs

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

Gammes

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

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

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

1146 dont la première ne contient que des croches et la deuxième que des
 1147 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1148 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1149 Systèmes — motifs et gammes combinés

1150 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1151 de la figure 4.3.

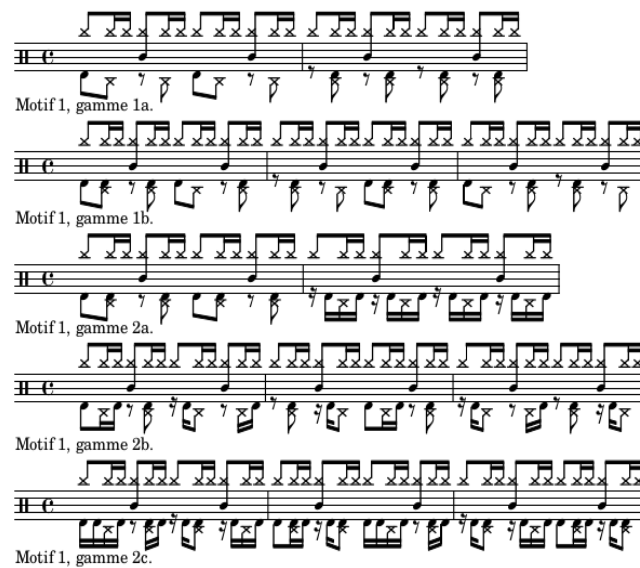


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1152

1153 Représentation du système en arbres de rythmes

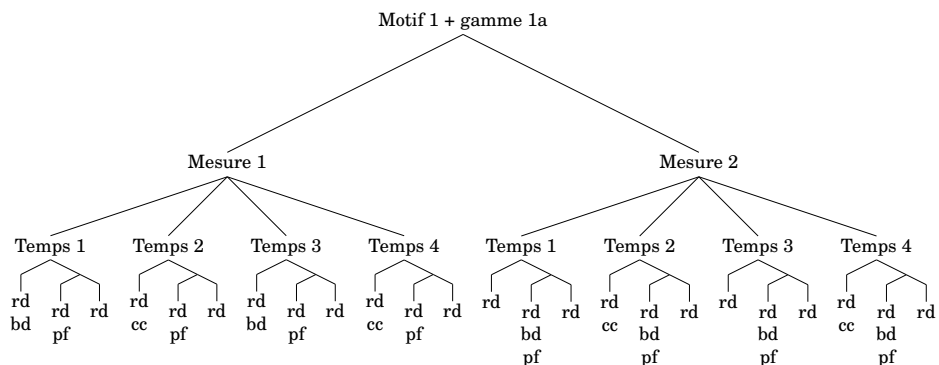


FIGURE 4.5 – Arbre de rythme — système

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.

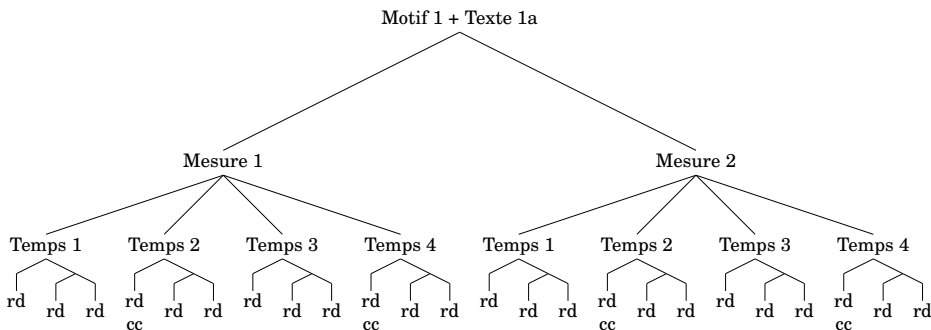


FIGURE 4.6 – Arbres de rythme — voix haute

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

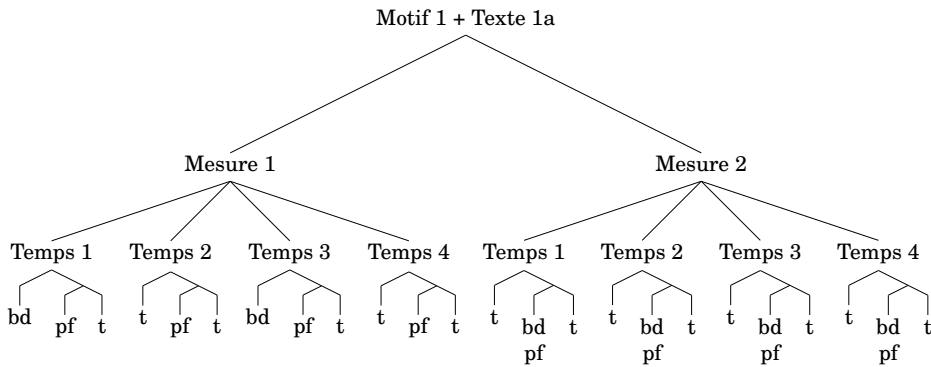


FIGURE 4.7 – Arbres de rythme — voix basse

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

1169 Les règles de simplifications

1170 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1171 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1172 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1173 entre elles (rendre la pulse visuelle).

1174

1175 Pour les figures ci-dessous :

1176 — x = une note ;

1177 — r = un silence ;

1178 — t = une continuation (point ou liaison)

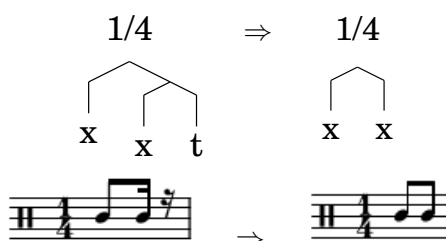


FIGURE 4.8

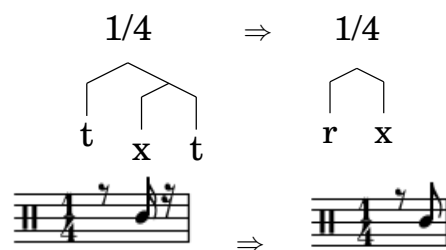


FIGURE 4.9

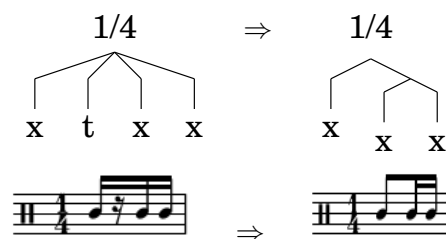


FIGURE 4.10

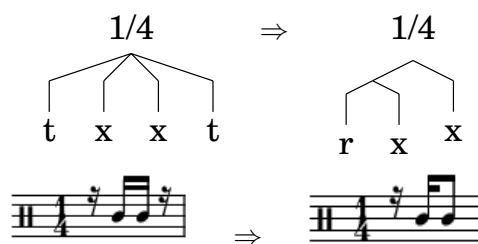


FIGURE 4.11

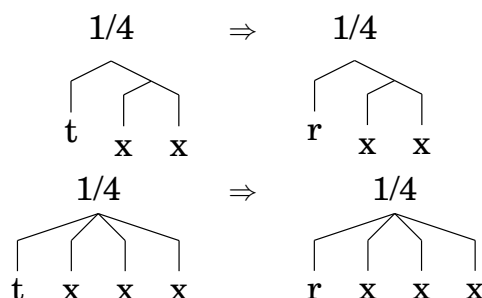


FIGURE 4.12

1179 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
1180 manquants seront mis en annexe.

1181 Les règles remplacent par un silence les continuations (t) qui sont au
1182 début d'un temps. Cela est valable pour ce système mais lorsqu'il y a des
1183 ouvertures de charley, cela n'est pas toujours applicable. Ce problème est
1184 évoqué de le chapitre 3.

1185 ⇒ **Objectif de cette expérimentation théorique :**

1186 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
1187 expérimentation théorique a pour but d'orienter la recherche d'autres
1188 systèmes par observation du jeu de données et de montrer comment les
1189 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

1190 4.4 Résultats et discussion

1191

1192 Cette section regroupe les avancées qui ont été réalisées par rapport aux
1193 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
1194 tats de l'ADT avec Qparse. Nous avons améliorer le système de quantifi-
1195 cation de Qparse pour la batterie, notamment le passage à la polyphonie
1196 avec les Jams.

1197 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
1198 grammaires avec des fichiers MIDI courts. Puis, une sortie MEI a été
1199 aussi été obtenu (encore à vérifier).

je vois 2 problématiques
et contrib. principales : 1)
transcription polyphonique
par parsing (verrou) : jams
etc 2) réécriture, pour sé-
paration en voix et sim-
plification, 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

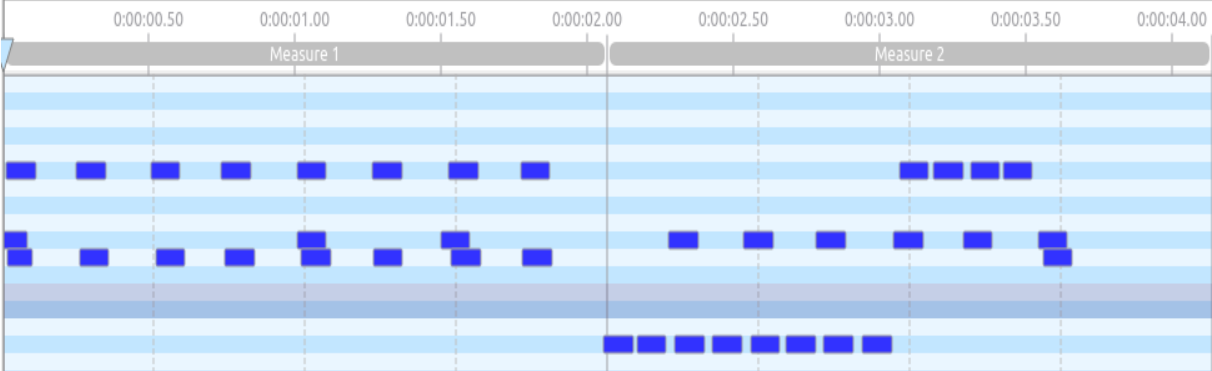
1200

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 :

1207

// bar level

1208

0 -> C0 1

1209

0 -> E1 1

1210

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

1211

1212

1213

// half bar level

1214

9 -> C0 1

1215

9 -> E1 1

1216

1217

// beat level

1218

1 -> C0 1

1219

1 -> E1 1

1220

1 -> T2(2, 2) 1

1221

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

1222

1223

// croche level

1224

2 -> C0 1

1225

2 -> E1 1

1226

1227

// double level

1228

4 -> C0 1

1229

4 -> E1 1

1230

4 -> E2 1

1231

4 -> T2(6, 6) 1

1232

1233 // triple level

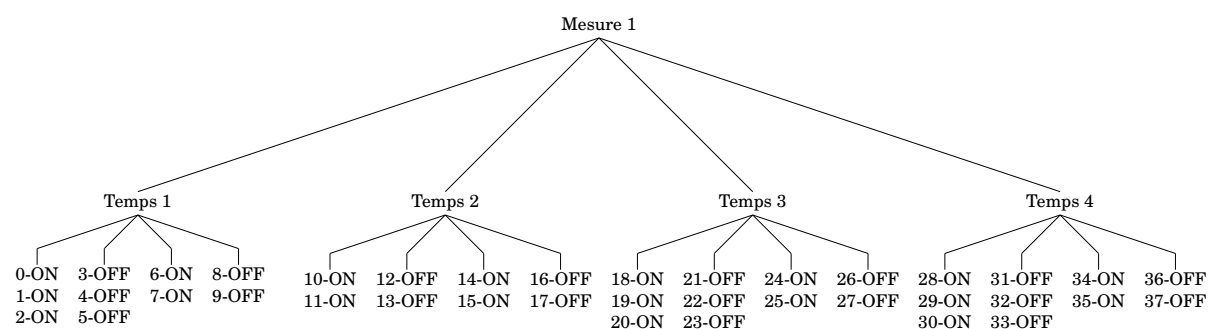
1234 6 -> E1 1

1235

1236 Cette grammaire sépare les ligatures par temps au niveau de la
 1237 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1238 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1239 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1240 il découpe correctement les mesures et les temps.

1241

1242



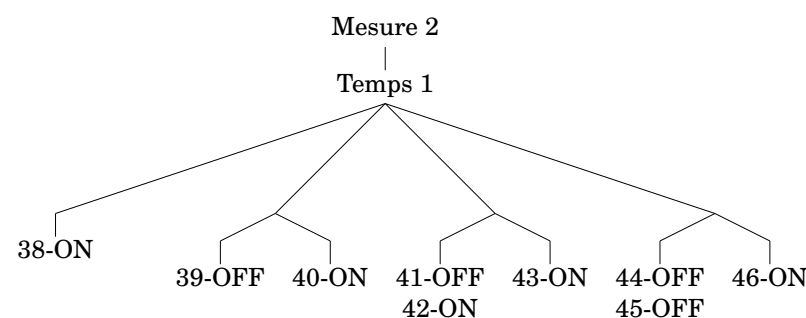
1243

1244

1245 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1246 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1247 quantification visibles dès le premier temps.

1248

1249



1250

1251

1252 Les Onsets sont correctement triés au niveau des doubles croches
 1253 mais certaines doubles croches sont inutilement subdivisées en triples
 1254 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1255 ci-dessus).

1256

1257 **2ème exemple :**

1258 Après une augmentation du poids des triples croches dans la grammaire
 1259 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1260 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1261 sont bien triés et aucuns ne sont subdivisés.

1262 **Évaluation**

1263 Pour l'évaluation, il aurait fallu produire un module.
 1264 L'évaluation est-elle automatique ou manuelle?
 1265 Possibilité d'un export lilypond en arbre pour comparer l'output avec la
 1266 transcription manuelle.
 1267 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel
 1268 ou MEI pour les comparer et faire des statistiques. Si transformés en
 1269 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)
 1270 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.
 1271 L'expérimentation peut-être considérer comme une évaluation manuelle?
 1272 (magicien d'Oz)
 1273 Lilypond vers MIDI + output vers MIDI \Rightarrow Comparaison des MIDI
 1274 dumpés.
 1275

1276 **Discussion**

1277 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
 1278 choix qui ont été faits. Nous ferons un bilan des différents avancés qui
 1279 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.
 1280 Écrire des règles de réécriture spécifique aux charley avec un système
 1281 approprié. Le jeu de système
 1282 — implémenter un pattern...
 1283 \Rightarrow manque de temps?
 1284
 1285 — La partie résultat est manquante car :
 1286 \Rightarrow Sujet très difficile ;
 1287 \Rightarrow Matcher les motifs peut être fait ultérieurement ;
 1288 Mais ce travail aurait été indispensable pour obtenir une quan-
 1289 tité de résultats qui justifieraient une évaluation automatique
 1290 permettant de faire des graphiques.
 1291
 1292 — L'évaluation fut entièrement manuelle car :
 1293 \Rightarrow Très dure automatiquement : il faut comparer 2 partitions (réf
 1294 VS output)
 1295 — Le ternaire jazz (voir expérience 2)
 1296 — Reconnaissance d'un motif sur le MIDI
 1297 Reconnaître un motif (système) sur une mesure de l'input (un fi-
 1298 chier midi représentant des données audios)
 1299 \Rightarrow Motif (système) reconnu : true ou false
 1300 Si true :
 1301 - Choisir la grammaire correspondante ;
 1302 - Parser le MIDI ;

- 1303 - Appliquer les règles de réécritures (Séparation des voix et simpli-
1304 fication)
- 1305 — Nous travaillerons aussi sur la détection de répétitions sur plu-
1306 sieurs mesures afin de pouvoir corriger des erreurs sur une des
1307 mesures qui aurait dû être identique aux autres mais qui présente
1308 des différences.
- 1309 — dans quelle catégorie mettre le shuffle ?
- 1310 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
1311 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
1312 à la batterie a été générée. Elle sera un bon point de départ pour les
1313 travaux futurs dont plusieurs propositions sont énoncés dans le présent
1314 document.

CONCLUSION GÉNÉRALE

1316 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1317 tion automatique de la batterie. Son objectif était de transcrire, à partir
1318 de leur représentation symbolique MIDI, des performances de batteur de
1319 différents niveaux et dans différents styles en partitions écrites.
1320 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1321 cessus de regroupement des événements MIDI qui nous a permis de faire
1322 la transition du monophonique vers le polyphonique. Une des données
1323 importante de ce processus était de différencier les nature des notes d'un
1324 accord, notamment de distinguer lorsque 2 notes constituent un accord
1325 ou un *fla*.
1326 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1327 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1328 tionnables en amont du parsing, soit par indication des noms des fichiers
1329 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1330 naire de patterns prédéfinis⁵ qu'il serait pertinent de mettre en œuvre en
1331 machine learning.
1332 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1333 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1334 d'un système et pour la simplification grâce aux gammes du motif d'un
1335 système. Nous avons aussi montré comment, dans des travaux futurs, un
1336 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1337 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1338 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1339 la chaîne de traitement en terme de réécriture.
1340 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1341 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1342 entièrement automatisée de partitions de batterie et au problème général
1343 de l'AMT de bout en bout.[10]

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

BIBLIOGRAPHIE

- [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition, 1996. – Cité pages 7, 28 et 32.
- [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11 et 14.
- [3] Wikipedia. Music informatics. Available at https://en.wikipedia.org/wiki/Music_informatics (2021/01/06). – Cité page 11.
- [4] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité page 11.
- [5] Gaël Richard et al. De fourier à la reconnaissance musicale. Available at <https://interstices.info/de-fourier-a-la-reconnaissance-musicale/> (2019/02/15). – Cité page 11.
- [6] Caroline Traube. Quelle place pour la science au sein de la musicologie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions entre les traitements de la musique et du langage. *La Lettre des Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- [8] Mikaela Keller, Kamil Aksebi, Lorenzo Moreira, and Louis Bigo. Techniques de traitement automatique du langage naturel appliquées aux représentations symboliques musicales. In *JIM 2021 - Journées d’Informatique Musicale*, Virtual, France, July 2021. – Cité page 13.
- [9] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering music relations with sequential attention. In *NLP4MUSA*, 2020. – Cité page 13.
- [10] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri. Automatic music transcription : Challenges and future directions. *Journal of Intelligent Information Systems*, 41, 12 2013. – Cité pages 14, 16, 21 et 61.

- 1376 [11] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1377 musical statistics as guides for audio-to-score piano transcription.
1378 *Information Sciences*, 566 :262–280, 2021. – Cité pages 14, 23 et 24.
- 1379 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1380 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1381 Cité pages 14 et 25.
- 1382 [13] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1383 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1384 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1385 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1386 1483, 2018. – Cité pages 16, 22 et 25.
- 1387 [14] Moshekwa Malatji. Automatic music transcription for two instru-
1388 ments based variable q-transform and deep learning methods, 10
1389 2020. – Cité page 22.
- 1390 [15] Antti J. Eronen. Musical instrument recognition using ica-based
1391 transform of features and discriminatively trained hmms. *Seventh*
1392 *International Symposium on Signal Processing and Its Applications*,
1393 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 22.
- 1394 [16] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1395 drum sound description for real-world music using template adap-
1396 tation and matching methods. *International Conference on Music*
1397 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 23.
- 1398 [17] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1399 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1400 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1401 *Computation in Music*, volume Lecture Notes in Computer Science
1402 of *Proceedings of the Seventh International Conference on Mathema-*
1403 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1404 2019. Springer. – Cité pages 23 et 24.
- 1405 [18] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1406 ring of rhythm structures. In *Proceedings of the First International*
1407 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1408 IEEE Computer Society. – Cité page 24.
- 1409 [19] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1410 Term Rewriting Based Structural Theory of Rhythm Notation. Re-
1411 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1412 Cité page 24.
- 1413 [20] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1414 equivalent rhythmic notations based on rhythm tree languages. In
1415 *Third International Conference on Technologies for Music Notation*

- 1416 *and Representation (TENOR)*, Coroña, Spain, May 2017. Helena Lo-
1417 pez Palma and Mike Solomon. – Cité page 24.
- 1418 [21] R. Marxer and J. Janer. Study of regularizations and constraints in
1419 nmf-based drums monaural separation. In *International Conference*
1420 *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1421 02/09/2013 2013. – Cité page 25.
- 1422 [22] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1423 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 28.
- 1424 [23] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1425 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 28.
- 1426 [24] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1427 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1428 94120 Fontenay-sous-Bois, 2000. – Cité pages 29 et 40.
- 1429 [25] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1430 Bamman. Learning to groove with inverse sequence transforma-
1431 tions. In *International Conference on Machine Learning (ICML)*,
1432 2019. – Cité page 45.

