

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
98 audio ou symboliques. En considérant que la musique et les langues na-
99 turelles sont deux moyens qui nous servent à exprimer nos ressentis sur
100 le 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 codifié
104 par une grammaire (solfège, durées, nuances, volumes) et soulèvera des
105 problématiques concernées par les techniques du TAL.

106

107 POURQUOI?

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

114

116 COMMENT?

117 → Problématique :

118 L'écriture musicale offre de nombreuses possibilités pour la transcription
119 d'un rythme donné. Le contexte musical ainsi que la lisibilité d'une
120 partition pour un batteur entraîné conditionnent les choix d'écriture.
121 Reconnaître la métrique principale d'un rythme, la façon de regrouper les
122 notes par les ligatures, ou simplement décider d'un usage pour une durée

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

123 parmi les différentes continuations possibles (notes pointées, liaisons,
124 silences, etc.) constituent autant de possibilités que de difficultés. De
125 plus, la batterie est dotée d'une écriture spécifique par rapport à la
126 majorité des instruments.

127

128 → Méthodes :

129 → Contributions :

130 Il a donc fallu modéliser plusieurs critères concernant sa représentation.
131 Cette modélisation étaient jusqu'à présent inexistante.

132 La proposition principale de ce mémoire est basée sur la recherche de
133 rythmes génériques sur l'*input*. Ces rythmes sont des *patterns* standards
134 de batterie définis au préalable et accompagnés par les différentes combi-
135 naisons qui leur sont propres. On les nomme systèmes (voir sections 3.4,
136 4.3). L'objectif des systèmes est de fixer des choix le plus tôt possible afin
137 de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces
138 choix concernent notamment la métrique et les règles de réécriture.

139 La proposition ci-dessus a nécessité plusieurs sous-tâches dont une des-
140 cription de la notation de la batterie (3.1) ainsi qu'une modélisation pour
141 la transcription de la batterie (3.2).

142 Plusieurs transcriptions manuelles ont été effectuées afin d'analyser les
143 contenus des fichiers MIDI et Audio (4.2) et de faire des comparaisons
144 de transcription avec des outils déjà existants¹. Une partition entière a
145 aussi était transcrite manuellement afin de repérer les éléments impor-
146 tants pour la modélisation et faire les liens entre les critères des données
147 d'input avec 4.2 L'ensemble de ces sous-tâches a permis la création expé-
148 rimentation théorique d'un système 4.3.

149 - Une fois proposition élaboré et verrouillé

150 - construire la chaîne jusqu'au bout

151 - nous avons pu la pousser jusqu'à la polyphonique en théorie et en implé-
152 mentation, ma contribution sur ce sujet étant la théorie sur la détection
153 de l'identité de notes dans un cluster (accord), l'implémentation de tests
154 unitaires sur les Jams (4.4) et la création de grammaires pondérées spé-
155 cifiques à la batterie (4.4)

156 Les codes cpp sur la drum ont pu être construit en parti grâce aux
157 travaux réalisés dans ce mémoire de recherche.

158

159 PLAN

160 Nous présenterons le contexte (chapitre 1) suivi d'un état de l'art (chapitre
161 2) et nous définirons de manière générale le processus de transcription
162 automatique de la musique pour enfin étayer les méthodes (chapitre 3)
163 utilisées pour la transcription automatique de la batterie. Nous décrirons
164 ensuite le corpus ainsi que les différentes expérimentations menées (cha-

1. MuseScore3

165 pitre 4). Nous concluerons par une discussion sur les résultats obtenus et
166 les pistes d'améliorations futures à explorer. Les contributions apportées
167 à l'outil qparse seront exposées dans les chapitres 3 et 4.

CONTEXTE

Sommaire

171	1.1	TAL et MIR	11
172	1.2	La transcription automatique de la musique	13
173	1.3	La transcription automatique de la batterie	15
174	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

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

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

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 psychologie...

On peut citer : ...

? psycho-acoustique, neurosciences ?

sujet : la recherche et 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 il ne suffit pas à expliquer la relation entre la musique et le langage. Il faut aussi citer Bernstein et Chomsky.

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))

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

251 L'objectif de la transcription automatique de la musique (AMT) [10] est
252 de convertir la performance d'un musicien en notation musicale - un peu

à l'instar de la 253

applications 254

préservation du patrimoine 255

e.g. musique de tradition 257

orale (ethno-musicologie) 258

citer un survey pour les 259

applications (pas [11]) 260

261

262

pas d'accord avec ça. pro-263

blème des partitions libres 264

de droit. 265

l'intérêt est aussi 266

d'avoir des partitions 267

au contenu exploitable 268

(texte ou XML) vs images 269

(pdf...) cf. par ex. cette 270

présentation d'OpenScore 271

à FOSDEM https://ar- 272

chive.fosdem.org/2017/schedule/event/openscore/ 273

et mes transparents sur le 274

sujet 275

là on passe aux approches 276

scientifiques 277

quel rapport ? 278

279

280

281

282

la figure ne correspond pas 283

à ton travail. ici "score" = 284

MIDI performance. Tu peux 285

lister les sous-tâches en 286

section 2.2 287

288

289

290

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))

- 291 Quatres sous-tâches optionnelles accompagnent ces algo-
 292 rithmes :
 293 — identification de l'instrument ;
 294 — estimation de la tonalité et de l'accord ;
 295 — détection de l'apparition et du décalage ;
 296 — estimation du tempo et du rythme.
 297 — Apprentissage sur des modèles acoustiques et musicologiques.
 298 — *Optionnel* : Informations fournies de manière externe, soit fournie
 299 en amont (genre, instruments,...), soit par interaction avec un uti-
 300 lisateur (infos sur une partition incomplète).

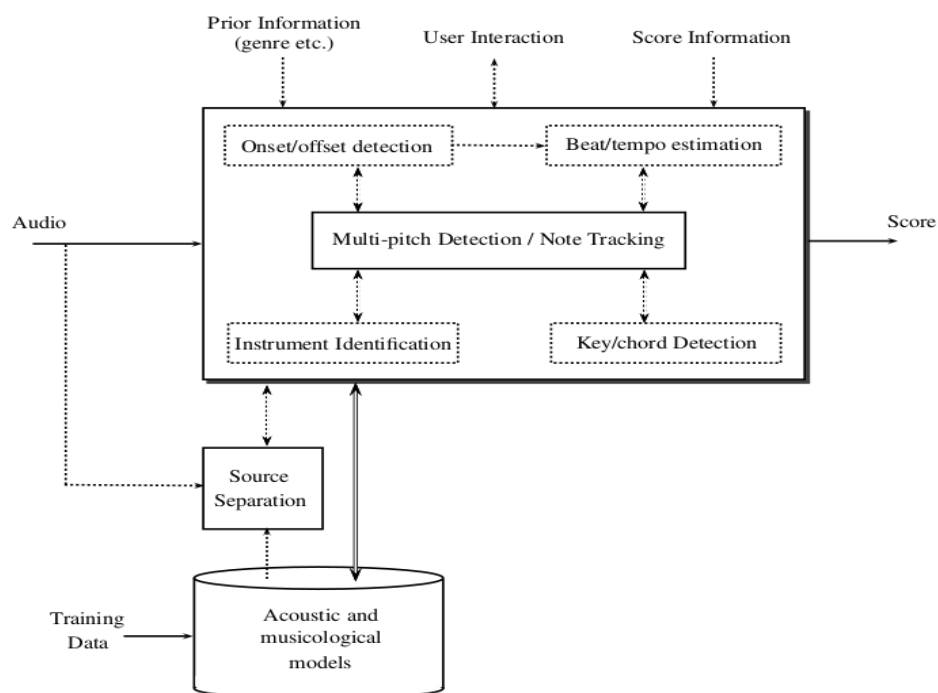


FIGURE 1.1 – Transcription automatique

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.

301

éviter newpage

302 1.3 La transcription automatique de la batterie

303

tres bonne section

304 La batterie est un instrument récent qui s'est longtemps passé de parti-
 305 tion. En effet pour un batteur, la qualité de lecteur lorsqu'elle était néces-
 306 saire, résidait essentiellement dans sa capacité à lire les partitions des

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

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

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

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

336 1.4 Les représentations de la musique

338 Les données audio

339 Le fichier WAV¹¹ est une instance du Resource Interchange File Format
340 (RIFF) défini par IBM et Microsoft. Le format RIFF agit comme une "en-
341 veloppe" pour divers formats de codage audio. Bien qu'un fichier WAV
342 puisse contenir de l'audio compressé, le format audio WAV le plus cou-
343 rant est l'audio non compressé au format LPCM (linear pulse-code modu-
344 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.

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

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

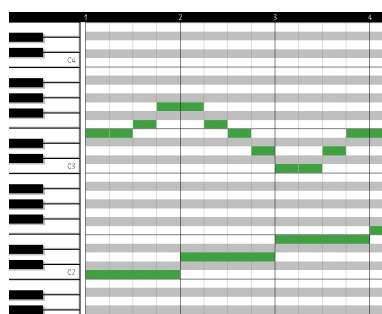
351 Les données MIDI

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

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

359 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

360

361 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

362

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

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

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

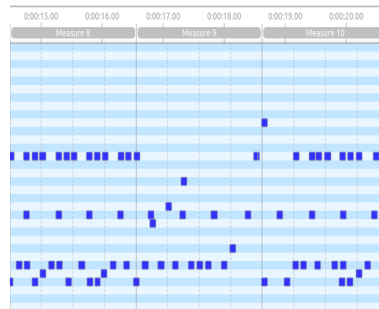


FIGURE 1.4 – Exemple évènements sans durée

367
 368 On observe que toutes les durées sont identiques.

369 Les partitions



FIGURE 1.5 – Exemple de partition de piano

370 Une partition de musique¹³ est un document qui porte la représentation
 371 systématique du langage musical sous forme écrite. Cette représentation
 372 est appelée transcription et elle sert à traduire les quatre caractéristiques
 373 du son musical :
 374 — la hauteur ;
 375 — la durée ;
 376 — l'intensité ;
 377 — le timbre.

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

378 Ainsi que de leurs combinaisons appelées à former l'ossature de l'œuvre
 379 musicale dans son déroulement temporel, à la fois :
 380 — diachronique (succession des instants, ce qui constitue en musique
 381 la mélodie) ;
 382

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>

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

ÉTAT DE L'ART

Sommaire

409	2.1	Monophonique et polyphonique	21
410			
411	2.2	Audio vers MIDI	22
412	2.3	MIDI vers partition	23
413	2.4	Approche linéaire et approche hiérarchique	23
414			
416			

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

haute fréquence, aigus?

classification des événements? la phrase semble redondante

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

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

477 2.3 MIDI vers partition

478
 479 Le plus souvent, lorsque les articles abordent la transcription automa-
 480 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »
 481 (partition) désigne un ouput au format Music XML, ou simplement MIDI.
 482 Par exemple, dans [11], la chaîne de traitement va jusqu'à la génération
 483 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en
 484 extraire manuellement un fichier MusicXML contenant plusieurs voix.
 485 Seuls quelques travaux récents s'intéressent de près à la création d'outils
 486 permettant la génération de partition. Le problème de la conversion d'une
 487 séquence d'évènements musicaux symboliques en une partition musicale
 488 structurée est traité notamment dans [17]. Ce travail, qui vise à résoudre
 489 en une fois la quantification rythmique et la production de partition struc-
 490 turée, s'appuie tout au long du processus sur des grammaires génératives
 491 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-
 492 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été
 493 menées avec un ensemble de données composé d'extraits monophoniques ;
 494 il reste donc à traiter le passage au polyphonique, en couplant le problème
 495 de la séparation des voix avec la quantification du rythme.
 496 L'approche de [17] est fondée sur la conviction que la complexité de la
 497 structure musicale dépasse les modèles linéaires.

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

de manière conjointe

langage *a priori*

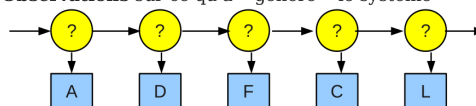
qui nécessite de traiter le
problème supplémentaire
de la séparation de voix. i.e.
pour la batterie on n'event
quantification + structu-
ration + séparation mais
seules les 2 premières sont
couplées dans l'approche de
tonn stage.

498 2.4 Approche linéaire et approche hiérarchique

499 Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par
 500 exemple, Shibata *et al.* [11] ont utilisé le modèle de Markov caché (HMM)⁴
 501 pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux
 502 réseaux de neurones profonds, l'un pour la reconnaissance des pitches et
 503 l'autre pour la reconnaissance de la vélocité. Pour la dernière couche, la
 504 probabilité est obtenue par une fonction sigmoïde. Ils construisent en-
 505 suite plusieurs HMM métriques étendus pour la musique polyphonique
 506 correspondant à des métriques possibles, puis ils calculent la probabilité
 507 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

508 *Source : Cours de Damien Nouvel*⁵

509

510

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

je ne comprend pas bien 518
 l'explication. le pb est plu- 519
 tot vue locale (déduction de 520
 la proba d'une durée à par- 521
 tir de la durée précédente, 522
 par ex. dans un HMM) vs
 vue globale, dans une hié-
 rarchie

RT? 523

techniques de réécriture 525
 appliquée à la déduction 526
 automatique, calcul symbo-
 lique 527

le calcul d'équiv. 528

529

530

citer thèse de David Rizo 532
 (Valencia) 533

534

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

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

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

Example: *Summertime*

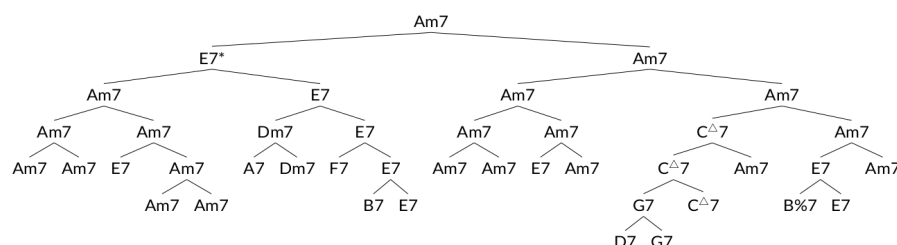


FIGURE 2.2 – arbre_jazz
Représentation arborescente d'une grille harmonique [12]

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 événements 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

MÉTHODES

Sommaire

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

Introduction

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

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

Enfin, nous présenterons les systèmes.

plusieurs développeurs

systèmes, une représentation théorique qui permet...

3.1 La notation de la batterie



FIGURE 3.1

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

— La ronde, elle vaut 4 ;

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

4 temps

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

580 — La blanche, elle vaut 2 ;

581 — La noire, elle vaut 1 ;

582 — La croche, elle vaut 1/2.

plusieurs éléments

583

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

— Une tête de note :

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

— Une hampe :

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

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

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

584

585

586

587

barre verticale liée à la tête
de note

588

589

haut ou bas

590

591

592

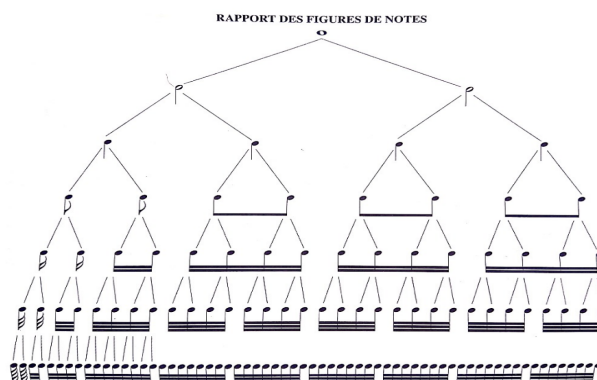


FIGURE 3.2 – Rapport des figures de notes
[1]

593 La figure 3.2 montre les rapports de durée entre les figures de notes. Plus
594 les durées sont longues, plus elles sont marquées par la tête de note (la
595 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de
596 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute
597 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.
598 Les notes à crochet (croche, double-croche, triple-croche...) peuvent être
599 reliées ou non par des ligatures (voir les 4 dernières lignes de la figure
600 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

600

601

Les hauteurs et les têtes de notes

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

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

606

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

En effet, les hauteurs sur la portée représentent :

— La hauteur physique des instruments :

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

— La hauteur des instruments en terme de fréquences :

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

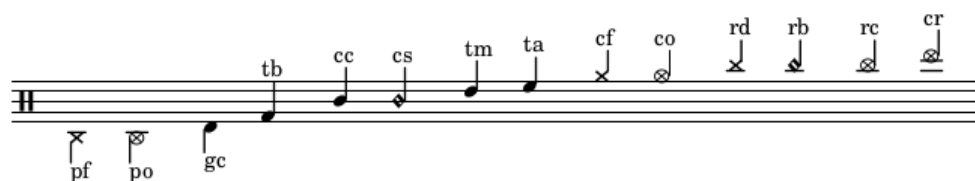


FIGURE 3.3 – Hauteur et têtes de notes

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

têtes de notes?

Les durées

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

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

- Le point;
- La liaison.

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

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

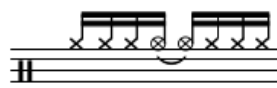
pour clarifier cela, tu pourrais décrire en 1.4 la notation conventionnelles (piano etc) et ici uniquement ce qui est spécifique à la batterie, en expliquant les différences.

expliquer comment, par ex. avec figure 3.4

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



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.4 – Point et liaison

= la position des temps 641

642

faire un "enumerate" 644

645

646

647

648

649

650

651

652

653

654

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.

655

Les silences

expliquer la notation (générale) des silences en §1.4 ? 656

657

quantifier = noter ? ou quantifier la durée ? 658

659

660

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.

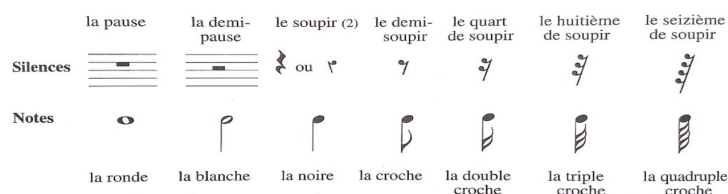


FIGURE 3.5 – Les silences

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

pas très clair

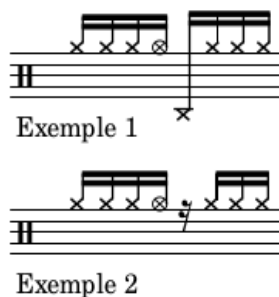


FIGURE 3.6 – Silence joué

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

itemize

cf?

674 Les équivalences rythmiques

675 Pour les instruments mélodiques, la liaison et le point sont les deux seules
 676 possibilités en cas d'équivalence rythmique pour des notes dont la durée
 677 de l'une à l'autre est ininterrompue. Mais pour la batterie, à part dans
 678 le cas des ouvertures de charley (voir section 3.1), les durées des notes
 679 n'ont pas d'importance. L'usage des silences pour combler la distance ryth-
 680 mique entre deux notes devient donc possible.
 681 Cela pris en compte, et étant donné que les indications de durée dans les
 682 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 a 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 flots mélodiques joués en parallèle, avec une synchronisation plus ou moins stricte.

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

une voix est caractérisée aussi pas orientation des hampes?

Les accentuations et les ghost-notes

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

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

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

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

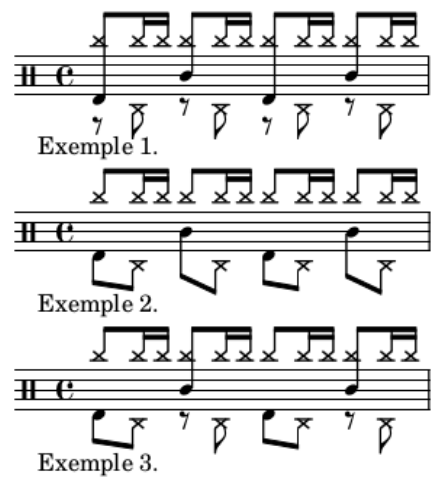


FIGURE 3.8 – Séparation des voix



FIGURE 3.9 – Les accents et les ghost-notes

713 ainsi, mais ces choix devront être adaptés en fonction des différents sys-
714 tèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz,
715 les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il
716 faudra donc mettre les symboles d'accentuation correspondants au-dessus
717 des têtes de notes.
718 La dernière note de la figure 3.9 montre un exemple de ghost-notes. Le
719 parenthésage a été choisi car il peut être utilisé sur n'importe quelle note
720 sans changer la tête de note.
721 Pour les codes, on prend le code de la note et on ajoute un « a » pour un
722 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.

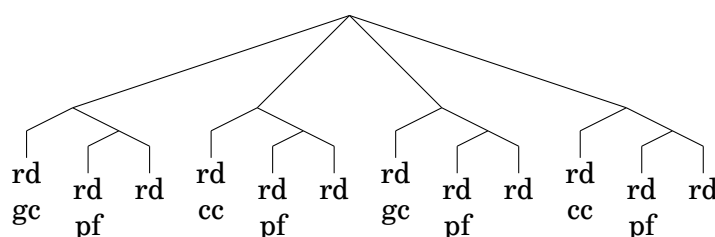
Cette transcription nous a mené aux observations suivantes :

- 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 ;

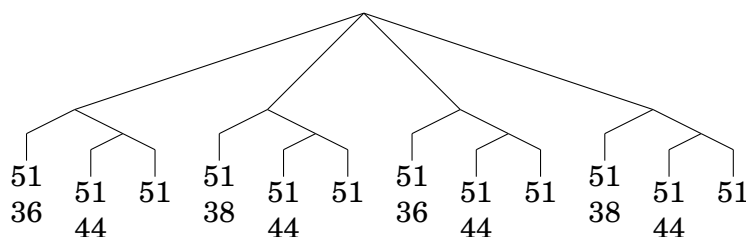
- Les accents et les ghosts-notes ne sont significatifs ni pour les instruments joués au pied, ni pour les cymbales crash.
- En effet, certaines vélocités en dessous de 40 étant détectées et inscrites dans les données MIDI sont dues au mouvement du talon du batteur qui bat la pulsation sans particulièrement 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. 771
il faut choisir entre les 2 772
(pour la suite aussi)

apprentissage 776

769 cents pondérés. En prenant en entrée une performance musicale symbo-
770 lique (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.

773 Les principaux contributeurs sont :

- 774 — Florent Jacquemard (Inria) : développeur principal.
- 775 — Francesco Foscari (PhD, CNAM) : construction de grammaire au-
776 tomatisée à partir de corpus ; Evaluation.
- 777 — Clement Poncelet (Salzburg U.) : integration de la librairie Midifile
778 pour les input MIDI.
- 779 — Philippe Rigaux (CNAM) : production de partition au format MEI
780 et de modèle intermédiaire de partition en sortie.
- 781 — Masahiko Sakai (Nagoya U.) : mesure de la distance input/output
782 pour la quantification et CMake framework ; évaluation.
- 783

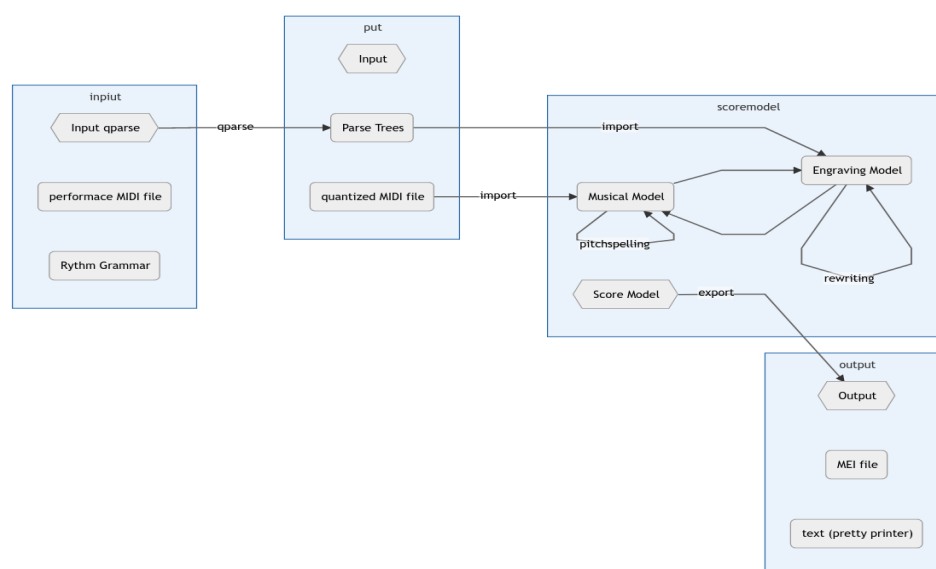


FIGURE 3.11 – Présentation de Qparse

la figure 3.11 est trop com-784
pliquée. rhythm grammar
→ automate d'arbres pon-785
déré. Parse Tree → arbre
syntaxique. qtz MIDI file :
inutile. Score Model → r-787
présentation intermédiaire
de partition. Score Model
Engr. Model : inutile. gar-
der juste la fleche Rewri-
ting sur S.M. 789

Explication des différentes étapes de la figure 3.11⁵ :

- **Input Qparse** :
Un fichier MIDI (séquence d'événements datés (piano roll) accom-
pagné 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	duolets, quartelets	2
afros-cubains	simple	croches	-	3

TABLE 3.2 – Sytèmes

833 dans 4 grandes catégories :

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

835 — 4/4 binaire

836 — 4/4 jazz

837 — 4/4 afro-cubain

838 **Objectif des systèmes**

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

840 — définir une métrique ;

841 — choisir une grammaire appropriée ;

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

843 tion.

845 bien, il faudrait expliquer ça avant.

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

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

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

réécritures) seront extraites des voix séparées des systèmes.

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

850 **Détection d’indication de mesure**

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

852 métrique ≠ signature rythmique (c’est plus général). Il aurait fallu présenter simplement 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

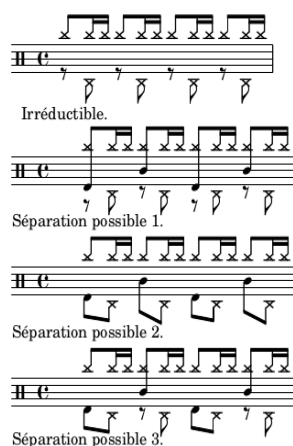


FIGURE 3.13 – Motif 4-4 binaire

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.

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

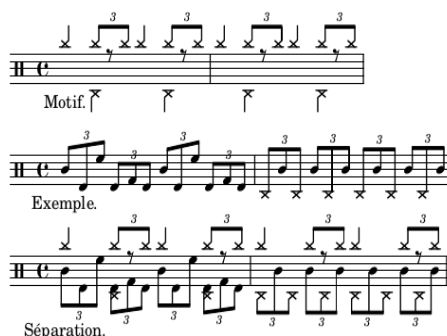


FIGURE 3.14 – Motif 4-4 jazz

885 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 886 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 887 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 888 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 889 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 890 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 891 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 892 et de caisse claire et son écriture séparée serait inutilement compliquée
 893 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 894 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de
 895 charley en bas.
 896

quel exemple?



FIGURE 3.15 – Système 4-4 afro-latin

897 La figure 3.15 montre un exemple minimaliste de système afro-latin [24].
 898 Ce système doit être écrit sur trois voix car la voix centrale est souvent
 899 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas
 900 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!

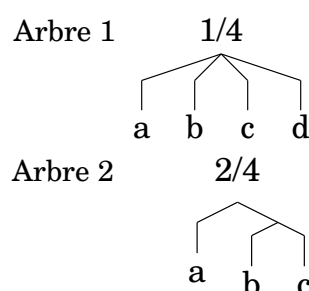


FIGURE 3.16 – Simplification

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

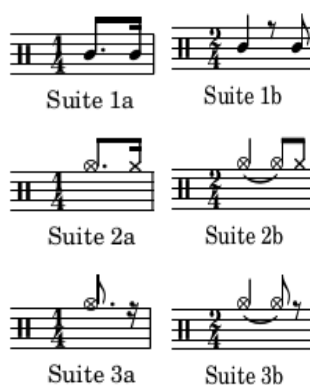


FIGURE 3.17

933 — Si a est un co :
 934 — Si d est un cf :
 935 ⇒ Suite 2a.
 936 — Si d est un pf :
 937 ⇒ Suite 3a : d deviens un silence (r).
 938

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

943 — Si a n'est pas un co :
 944 ⇒ Suite 1b, b devient un silence.
 945 — Si a est un co :
 946 — Si c est un cf :
 947 ⇒ Suite 2b, b devient une liaison et c devient un cf.
 948 — Si c est un pf :
 949 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
 950

951 *Rappel :*
 952 *cf = charley fermé joué à la main ;*
 953 *co = charley ouvert joué à la main ;*
 954 *pf = charley fermé joué au pied.*
 955

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

959 **Conclusion**

960 Nous avons formalisé une notation de la batterie, modélisé cette notation
961 pour la transcription de données MIDI en partition, nous avons décrit
962 Qparse.

963 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-
964 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-
965 propriée et énoncer des règles de séparation des voix et de simplification
966 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

995 Autres critères spécifiques au GMD :

- 996 — Toutes les performances ont été jouées au métronome et à un tempo
- 997 choisi par le batteur.
- 998 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 999 qui ont pu improviser dans un large éventail de styles. Les don-
- 1000 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 1001 (professionnel ou amateur).
- 1002 — Les batteurs avaient pour instruction de jouer des séquences de
- 1003 plusieurs minutes ainsi que des fills²
- 1004 — Chaque performance est annotée d'un style (fourni par le batteur),
- 1005 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 1006 nyme du batteur.
- 1007 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 1008 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 1009 session du GMD.
- 1010 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 1011 fichier MIDI.

1012 **Format des données**

enregistre les données dans
des fichiers MIDI

1013 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-
1014 tinctes :

- 1015 — une pour le tempo et l'indication de mesure ;
- 1016 — une pour les changements de contrôle (position de la pédale de
- 1017 charley) ;
- 1018 — une pour les notes.

1019

1020 Les changements de contrôle sont placés sur le canal 0 et les notes sur le
1021 canal 9 (qui est le canal canonique pour la batterie).

1022 Pour simplifier le traitement de ces données, ces trois pistes ont été
1023 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.

1024
 1025 « Control Changes The TD-11 also records control changes speci-
 1026 fying the position of the hi-hat pedal on each hit. We have preserved this
 1027 information under control 4. »
 1028 (<https://magenta.tensorflow.org/datasets/groove>)
 1029 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 1030 tions dans les fichiers MIDI.
 1031 L'utilisation de pretty_midi devient urgente!

1032 4.2 Analyse MIDI-Audio

1033

1034 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
 1035 partir de fichiers MIDI et Audio du GMD.

= analyses et transcriptions
manuelles

1036 Comparaisons de transcriptions

1037 Pour les comparaisons de transcriptions, les transcriptions manuelles
 1038 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les trans-
 1039 criptions automatiques (TA) ont toutes été générées manuellement avec
 1040 MuseScore.

méthodologie tr. manuelle.
import MIDI pour MuseS-
core

1041 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1042 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1043 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1044 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1045 premières notes (une double-croche et une croche) d'un triolet sur
- 1046 le temps 1 de la mesure 2 de la TA.

1047 Exemple d'analyse 2

- 1048 — Les doubles croches ont été interprétées en quintolet
- 1049 — La deuxième double-croche est devenue une croche.

Transcription manuelle ⇒ Transcription automatique



Transcription manuelle ⇒ Transcription automatique



1050 Exemple d'analyse 3

- 1051 — Les grosses-caisses, les charleys et les caisses-claires ont été déca-
- 1052 lés d'un temps vers la droite.
- 1053 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été
- 1054 décalés d'une double croche vers la droite dans la TA.
- 1055 — La première caisse-claire de la mesure 1 devient binaire dans la TA
- 1056 alors qu'elle appartenait à un triolet dans la TM.
- 1057 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe
- 1058 pas la TM.
- 1059

1060 Exemple d'analyse 4

Transcription manuelle ⇒ Transcription automatique



1061

- 1062 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
- 1063 manière excessivement complexe!

conclusion sur ces exemples

1064 Exemple avec des flas

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

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

1066 Transcription manuelle



1067

1068 Transcription automatique

1069



1070

1071

- 1072 — Le premier fla est reconnu comme étant un triolet contenant une
- 1073 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1074 ornementée.
- 1075 — Le deuxième fla est reconnu comme étant un accord.
- 1076 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1077 dans la TA.
- 1078 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1079 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1080 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1081 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1082 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1083 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1084 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1085 temps 3 de la TA.
- 1086 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1087 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1088 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1089 cross-sticks.

1090 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1091 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1092 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1093 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1094 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1095 *Inria*) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1096 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1097 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1098 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1099 et quatre).

1100 La transcription des données audio et MIDI contenues dans ces fichiers
 1101 a permis une analyse plus approfondie des critères à relever pour chaque
 1102 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 1104
 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

1129 dont la première ne contient que des croches et la deuxième que des
 1130 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1131 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1132 Systèmes — motifs et gammes combinés

1133 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1134 de la figure 4.3.

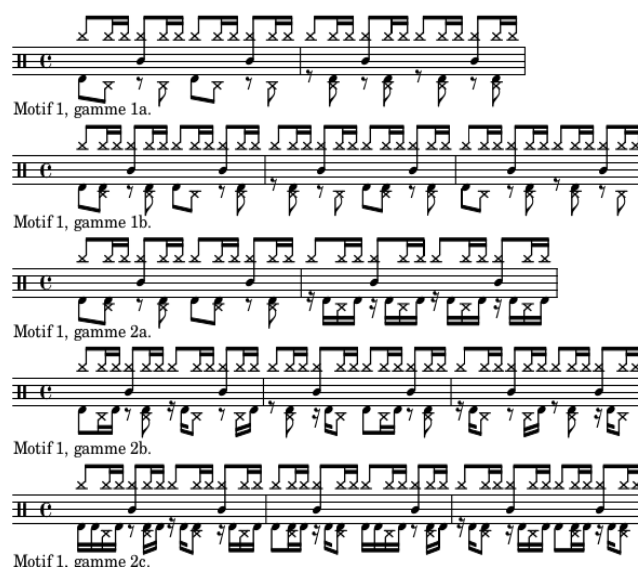


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1135

1136 Représentation du système en arbres de rythmes

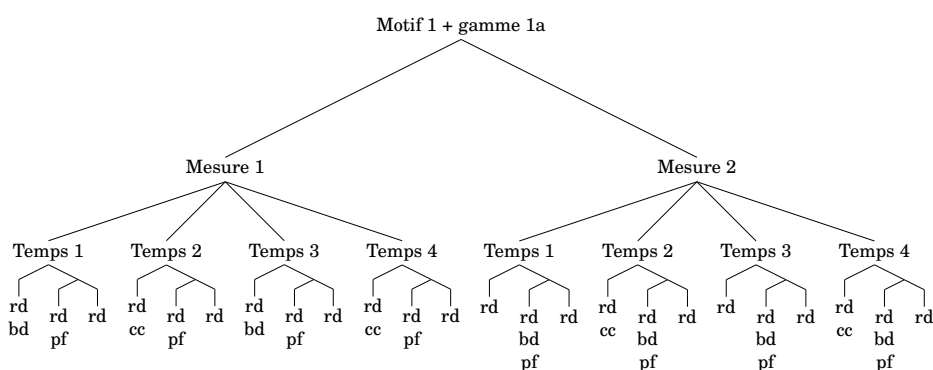


FIGURE 4.5 – Arbre de rythme — système

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.

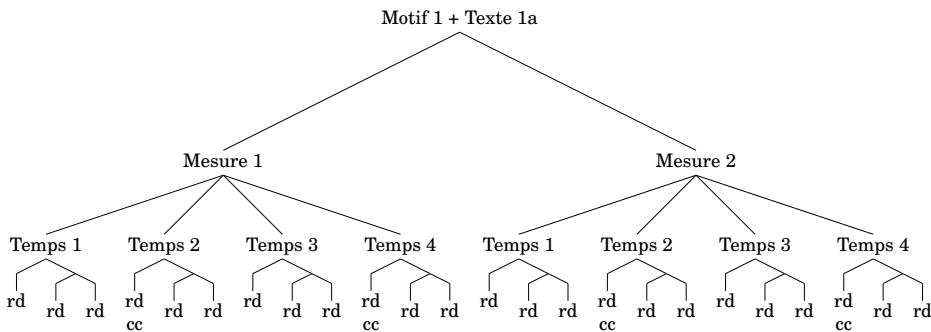


FIGURE 4.6 – Arbres de rythme — voix haute

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

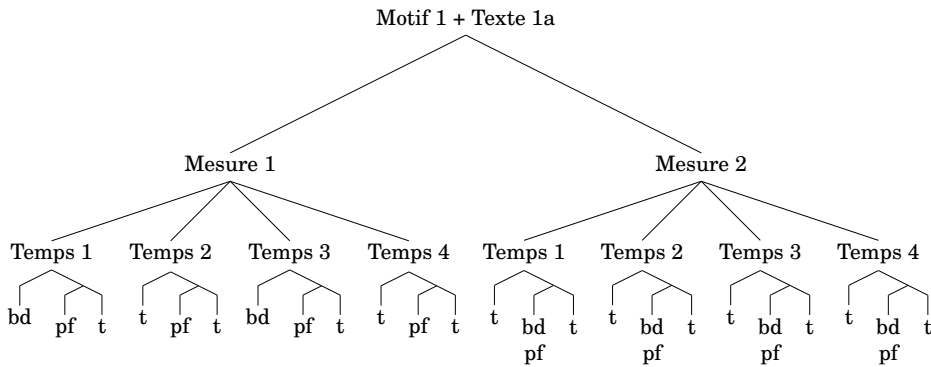


FIGURE 4.7 – Arbres de rythme — voix basse

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

1152 Les règles de simplifications

1153 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1154 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1155 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1156 entre elles (rendre la pulse visuelle).

1157

1158 Pour les figures ci-dessous :

1159 — x = une note ;

1160 — r = un silence ;

1161 — t = une continuation (point ou liaison)

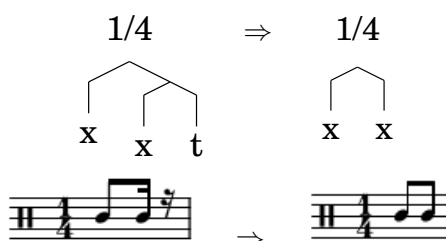


FIGURE 4.8

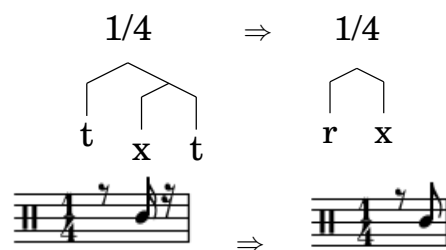


FIGURE 4.9

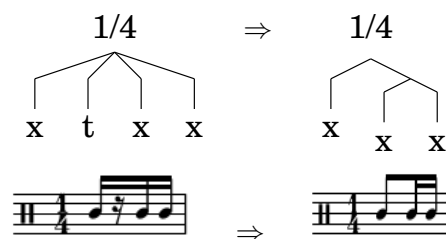


FIGURE 4.10

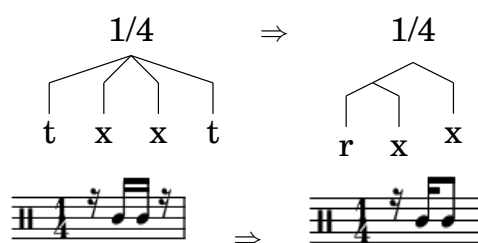


FIGURE 4.11

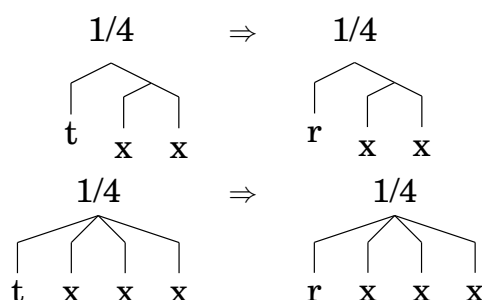


FIGURE 4.12

1162 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
1163 manquants seront mis en annexe.

1164 Les règles remplacent par un silence les continuations (t) qui sont au
1165 début d'un temps. Cela est valable pour ce système mais lorsqu'il y a des
1166 ouvertures de charley, cela n'est pas toujours applicable. Ce problème est
1167 évoqué de le chapitre 3.

1168 ⇒ **Objectif de cette expérimentation théorique :**

1169 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
1170 expérimentation théorique a pour but d'orienter la recherche d'autres
1171 systèmes par observation du jeu de données et de montrer comment les
1172 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

1173 4.4 Résultats et discussion

1174

1175 Cette section regroupe les avancées qui ont été réalisées par rapport aux
1176 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
1177 tats de l'ADT avec Qparse. Nous avons améliorer le système de quantifi-
1178 cation de Qparse pour la batterie, notamment le passage à la polyphonie
1179 avec les Jams.

1180 Nous avons pu obtenir des arbres de parsing corrects en améliorant les
1181 grammaires avec des fichiers MIDI courts. Puis, une sortie MEI a été
1182 aussi été obtenu (encore à vérifier).

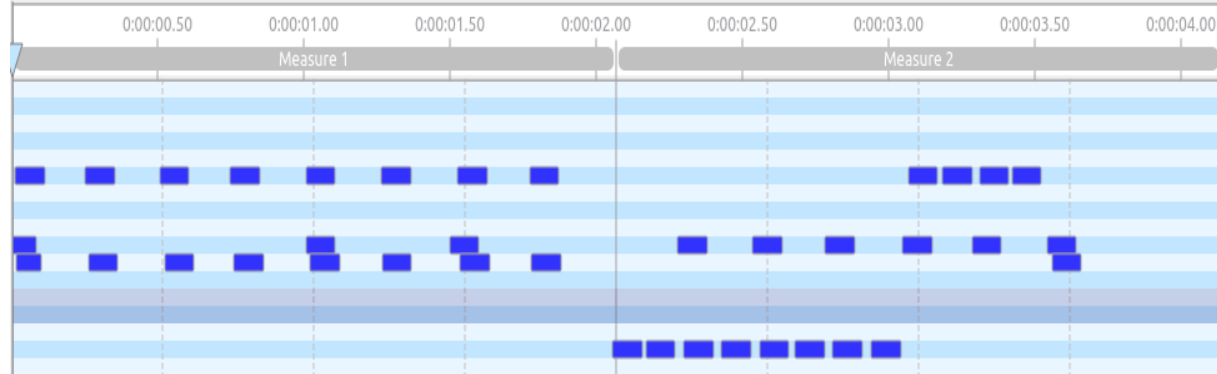
je vois 2 problématiques
et contrib. principales : 1)
transcription polyphonique
par parsing (verrou) : jams
etc 2) réécriture, pour sé-
paration en voix et sim-
plification, aidée (guidée)
par système. Ce serait bien
de présenter la contrib. 1
dans une section (comme
4.3 pour 2), avant d'aborder
résultats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

1183 Les Jams

Les Jams permettent de passer du monophonique au polyphonique.

Le parsing

Tests effectués avec le fichier midi suivant :



Un premier test convaincant est effectué avec la grammaire suivante :

```
// bar level
0 -> C0 1
0 -> E1 1
0 -> U4(1, 1, 1, 1) 1

// half bar level
9 -> C0 1
9 -> E1 1

// beat level
1 -> C0 1
1 -> E1 1
1 -> T2(2, 2) 1
1 -> T4(4, 4, 4, 4) 1

// croche level
2 -> C0 1
2 -> E1 1

// double level
4 -> C0 1
4 -> E1 1
4 -> E2 1
4 -> T2(6, 6) 1
```

il faut revenir ici sur le parsing, et la notion d'alignement sur arbres syntaxiques pour définir les jams. illustrer avec les exemples précédents.

revenir (rapidement) sur la méthodologie suivie.

1216 // triple level

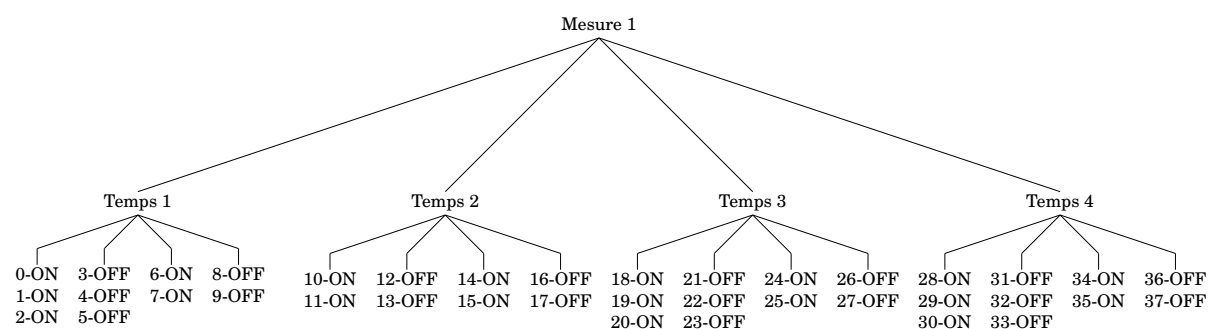
1217 6 -> E1 1

1218

1219 Cette grammaire sépare les ligatures par temps au niveau de la
 1220 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1221 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1222 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1223 il découpe correctement les mesures et les temps.

1224

1225



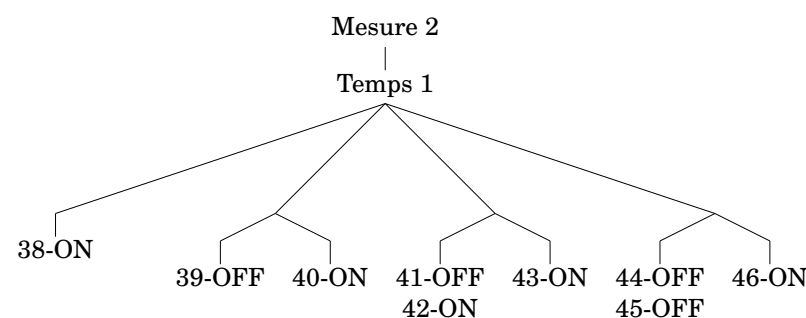
1226

1227

1228 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1229 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1230 quantification visibles dès le premier temps.

1231

1232



1233

1234

1235 Les Onsets sont correctement triés au niveau des doubles croches
 1236 mais certaines doubles croches sont inutilement subdivisées en triples
 1237 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1238 ci-dessus).

1239

1240 2ème exemple :

1241 Après une augmentation du poids des triples croches dans la grammaire
 1242 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1243 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1244 sont bien triés et aucuns ne sont subdivisés.

1245 **Évaluation**

1246 Pour l'évaluation, il aurait fallu produire un module.
1247 L'évaluation est-elle automatique ou manuelle?
1248 Possibilité d'un export lilypond en arbre pour comparer l'output avec la
1249 transcription manuelle.
1250 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel
1251 ou MEI pour les comparer et faire des statistiques. Si transformés en
1252 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)
1253 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.
1254 L'expérimentation peut-être considérer comme une évaluation manuelle?
1255 (magicien d'Oz)
1256 Lilypond vers MIDI + output vers MIDI \Rightarrow Comparaison des MIDI
1257 dumpés.
1258

1259 **Discussion**

1260 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
1261 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
1262 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.
1263 Écrire des règles de réécriture spécifique aux charley avec un système
1264 approprié. Le jeu de système
1265 — implémenter un pattern...
1266 \Rightarrow manque de temps?
1267
1268 — La partie résultat est manquante car :
1269 \Rightarrow Sujet très difficile ;
1270 \Rightarrow Matcher les motifs peut être fait ultérieurement ;
1271 Mais ce travail aurait été indispensable pour obtenir une quan-
1272 tité de résultats qui justifieraient une évaluation automatique
1273 permettant de faire des graphiques.
1274
1275 — L'évaluation fut entièrement manuelle car :
1276 \Rightarrow Très dure automatiquement : il faut comparer 2 partitions (réf
1277 VS output)
1278 — Le ternaire jazz (voir expérience 2)
1279 — Reconnaissance d'un motif sur le MIDI
1280 Reconnaître un motif (système) sur une mesure de l'input (un fi-
1281 chier midi représentant des données audios)
1282 \Rightarrow Motif (système) reconnu : true ou false
1283 Si true :
1284 - Choisir la grammaire correspondante ;
1285 - Parser le MIDI ;

- 1286 - Appliquer les règles de réécritures (Séparation des voix et simpli-
1287 fication)
- 1288 — Nous travaillerons aussi sur la détection de répétitions sur plu-
1289 sieurs mesures afin de pouvoir corriger des erreurs sur une des
1290 mesures qui aurait dû être identique aux autres mais qui présente
1291 des différences.
- 1292 — dans quelle catégorie mettre le shuffle ?
- 1293 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
1294 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
1295 à la batterie a été générée. Elle sera un bon point de départ pour les
1296 travaux futurs dont plusieurs propositions sont énoncés dans le présent
1297 document.

CONCLUSION GÉNÉRALE

1299 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1300 tion automatique de la batterie. Son objectif était de transcrire, à partir
1301 de leur représentation symbolique MIDI, des performances de batteur de
1302 différents niveaux et dans différents styles en partitions écrites.
1303 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1304 cessus de regroupement des événements MIDI qui nous a permis de faire
1305 la transition du monophonique vers le polyphonique. Une des données
1306 importante de ce processus était de différencier les nature des notes d'un
1307 accord, notamment de distinguer lorsque 2 notes constituent un *accord*
1308 ou un *fla*.
1309 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1310 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1311 tionnables en amont du parsing, soit par indication des noms des fichiers
1312 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1313 naire de patterns prédéfinis⁵ qu'il serait pertinent de mettre en œuvre en
1314 machine learning.
1315 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1316 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1317 d'un système et pour la simplification grâce aux gammes du motif d'un
1318 système. Nous avons aussi montré comment, dans des travaux futurs, un
1319 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1320 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1321 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1322 la chaîne de traitement en terme de réécriture.
1323 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1324 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1325 entièrement automatisée de partitions de batterie et au problème général
1326 de l'AMT de bout en bout.[10]

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

BIBLIOGRAPHIE

- 1328 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1329 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1330 1996. – Cité pages 7, 28 et 32.
- 1331 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 11
1332 et 14.
- 1333 [3] Wikipedia. Music informatics. Available at [https://en.](https://en.wikipedia.org/wiki/Music_informatics)
1334 [wikipedia.org/wiki/Music_informatics](https://en.wikipedia.org/wiki/Music_informatics) (2021/01/06). – Cité
1335 page 11.
- 1336 [4] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1337 page 11.
- 1338 [5] Gaël Richard et al. De fourier à la reconnaissance
1339 musicale. Available at [https://interstices.info/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/)
1340 [de-fourier-a-la-reconnaissance-musicale/](https://interstices.info/de-fourier-a-la-reconnaissance-musicale/) (2019/02/15).
1341 – Cité page 11.
- 1342 [6] Caroline Traube. Quelle place pour la science au sein de la musico-
1343 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 12.
- 1344 [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1345 entre les traitements de la musique et du langage. *La Lettre des*
1346 *Neurosciences*, 58 :24–26, 2018. – Cité page 13.
- 1347 [8] Mikaela Keller, Kamil Aksebi, Lorenzo Moreira, and Louis Bigo.
1348 Techniques de traitement automatique du langage naturel appli-
1349 quées aux représentations symboliques musicales. In *JIM 2021 -*
1350 *Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1351 Cité page 13.
- 1352 [9] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1353 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1354 Cité page 13.
- 1355 [10] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1356 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-
1357 lenges and future directions. *Journal of Intelligent Information Sys-*
1358 *tems*, 41, 12 2013. – Cité pages 14, 16, 21 et 61.

- 1359 [11] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1360 musical statistics as guides for audio-to-score piano transcription.
1361 *Information Sciences*, 566 :262–280, 2021. – Cité pages 14, 23 et 24.
- 1362 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1363 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1364 Cité pages 14 et 25.
- 1365 [13] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1366 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1367 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1368 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1369 1483, 2018. – Cité pages 16, 22 et 25.
- 1370 [14] Moshekwa Malatji. Automatic music transcription for two instru-
1371 ments based variable q-transform and deep learning methods, 10
1372 2020. – Cité page 22.
- 1373 [15] Antti J. Eronen. Musical instrument recognition using ica-based
1374 transform of features and discriminatively trained hmms. *Seventh*
1375 *International Symposium on Signal Processing and Its Applications*,
1376 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 22.
- 1377 [16] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1378 drum sound description for real-world music using template adap-
1379 tation and matching methods. *International Conference on Music*
1380 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 23.
- 1381 [17] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1382 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1383 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1384 *Computation in Music*, volume Lecture Notes in Computer Science
1385 of *Proceedings of the Seventh International Conference on Mathema-*
1386 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1387 2019. Springer. – Cité pages 23 et 24.
- 1388 [18] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1389 ring of rhythm structures. In *Proceedings of the First International*
1390 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1391 IEEE Computer Society. – Cité page 24.
- 1392 [19] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1393 Term Rewriting Based Structural Theory of Rhythm Notation. Re-
1394 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1395 Cité page 24.
- 1396 [20] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1397 equivalent rhythmic notations based on rhythm tree languages. In
1398 *Third International Conference on Technologies for Music Notation*

- 1399 *and Representation (TENOR)*, Coroña, Spain, May 2017. Helena Lo-
1400 pez Palma and Mike Solomon. – Cité page 24.
- 1401 [21] R. Marxer and J. Janer. Study of regularizations and constraints in
1402 nmf-based drums monaural separation. In *International Conference*
1403 *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1404 02/09/2013 2013. – Cité page 25.
- 1405 [22] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1406 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 28.
- 1407 [23] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1408 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 28.
- 1409 [24] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1410 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1411 94120 Fontenay-sous-Bois, 2000. – Cité pages 29 et 40.
- 1412 [25] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1413 Bamman. Learning to groove with inverse sequence transforma-
1414 tions. In *International Conference on Machine Learning (ICML)*,
1415 2019. – Cité page 45.

