

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

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

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

8 *Parcours :*  
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

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

14 *Encadrant :*  
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021



# TABLE DES MATIÈRES

18	<b>Liste des figures</b>	<b>4</b>
19	<b>Liste des tableaux</b>	<b>5</b>
20	<b>Introduction générale</b>	<b>7</b>
21	<b>1 Contexte</b>	<b>9</b>
22	1.1 TAL et MIR . . . . .	9
23	1.2 La transcription automatique de la musique . . . . .	11
24	1.3 La transcription automatique de la batterie . . . . .	14
25	1.4 Les représentations de la musique . . . . .	14
26	<b>2 État de l'art</b>	<b>19</b>
27	2.1 Monophonique et polyphonique . . . . .	19
28	2.2 Audio vers MIDI . . . . .	20
29	2.3 MIDI vers partition . . . . .	21
30	2.4 Approche linéaire et approche hiérarchique . . . . .	21
31	<b>3 Méthodes</b>	<b>25</b>
32	3.1 La notation de la batterie . . . . .	25
33	3.2 Modélisation pour la transcription . . . . .	32
34	3.3 Qparse . . . . .	34
35	3.4 Les systèmes . . . . .	35
36	<b>4 Expérimentations</b>	<b>43</b>
37	4.1 Le jeu de données . . . . .	43
38	4.2 Analyse MIDI-Audio . . . . .	45
39	4.3 Expérimentation théorique d'un système . . . . .	49
40	4.4 Résultats et discussion . . . . .	54
41	<b>Conclusion générale</b>	<b>59</b>
42	<b>Bibliographie</b>	<b>61</b>

## LISTE DES FIGURES

44	1.1	Transcription automatique . . . . .	13
45	1.2	Exemple évènements avec durée . . . . .	15
46	1.3	Critère pour un évènement . . . . .	16
47	1.4	Exemple évènements sans durée . . . . .	16
48	1.5	Exemple de partition de piano . . . . .	16
49	1.6	MusicXML . . . . .	17
50	2.1	HMM . . . . .	22
51	2.2	arbre_jazz . . . . .	23
52	3.1	Rapport des figures de notes . . . . .	26
53	3.2	Hauteur et têtes de notes . . . . .	27
54	3.3	Point et liaison . . . . .	28
55	3.4	Les silences . . . . .	28
56	3.5	Silence joué . . . . .	29
57	3.6	Équivalence . . . . .	30
58	3.7	Séparation des voix . . . . .	30
59	3.8	Les accents et les ghost-notes . . . . .	31
60	3.9	Exemple pour les accentuations et les ghost-notes . . . . .	31
61	3.10	Présentation de Qparse . . . . .	34
62	3.11	Métrique . . . . .	37
63	3.12	Motif 4-4 binaire . . . . .	38
64	3.13	Motif 4-4 jazz . . . . .	38
65	3.14	Système 4-4 afro-latin . . . . .	39
66	3.15	Simplification . . . . .	40
67	3.16	. . . . .	40
68	4.1	Batterie électronique . . . . .	44
69	4.2	Partition de référence . . . . .	48
70	4.3	Motifs et gammes . . . . .	49
71	4.4	Partition d'un système en 4/4 binaire . . . . .	50
72	4.5	Arbre de rythme — système . . . . .	50
73	4.6	Arbre de rythme — voix haute . . . . .	51
74	4.7	Arbre de rythme — voix basse . . . . .	51
75	4.8	. . . . .	52
76	4.9	. . . . .	52
77	4.10	. . . . .	52
78	4.11	. . . . .	53

79	4.12 . . . . .	53
----	----------------	----

80

## LISTE DES TABLEAUX

81	3.1 Pitches et instruments . . . . .	32
82	3.2 Systèmes . . . . .	36



## INTRODUCTION GÉNÉRALE

Ce mémoire de recherche, effectué en parallèle d'un stage à l'Inria dans le cadre du master de traitement automatique des langues de l'Inalco, contient une proposition originale ainsi que diverses contributions ayant toutes pour objectif d'améliorer **qparse**, un outil de transcription automatique de la musique sur sa capacité à transcrire la batterie. Nous ne parlerons donc pas directement de langues naturelles, mais de l'écriture automatique de partitions de musique à partir de données audios ou symboliques. Cette exercice nécessitera la manipulation d'un langage musical codifié avec une grammaire (solfège, durées, nuances, volumes) et soulèvera des problématiques concernées par les techniques du traitement automatique des langues.

L'écriture musicale offre de nombreuses possibilités pour la transcription d'un rythme donné. Le contexte musical ainsi que la lisibilité d'une partition pour un batteur entraîné conditionnent les choix d'écritures. Reconnaître la métrique principale d'un rythme, la façon de regrouper les notes par les ligatures, ou simplement décider d'un usage pour une durée parmi les différentes continuations possibles (notes pointées, liaisons, silences, etc.) constituent autant de possibilités que de difficultés.

102

Voici la proposition de ce mémoire ainsi que les contributions apportées lors du stage :

- Proposition principale : les systèmes (3.4, 4.3) :  
Recherche de rythmes génériques en amont dans la chaîne de traitement.  
⇒ L'objectif de fixer des choix le plus tôt possible afin de simplifier le reste des calculs en éliminant une partie d'entre eux. Ces choix concernent notamment la métrique et les règles de réécriture.
- Une description de la notation de la batterie (3.1)
- Une modélisation de la transcription de la batterie (3.2)
- Analyse MIDI-Audio (4.2)
- Transcription manuelle de partition 4.2
- Expérimentation théosique d'un système 4.3
- Théorie et tests unitaires pour le passage au polyphonique (4.4)
- Création de grammaires pondérées pour la batterie (4.4)
- Contributions sur la branche « distance » dans :  
— `qparselib/notes/cluster.md`

119

120 — `qparselib/src/segment/import/` :

121 `DrumCode` hpp et cpp

122

123 Nous présenterons le contexte suivi d'un état de l'art et nous définirons  
124 de manière générale le processus de transcription automatique de la mu-  
125 sique pour enfin étayer les méthodes utilisées pour la transcription au-  
126 tomatique de la batterie, et nous présenterons les principales contribu-  
127 tions apportées à l'outil `qparse`. Nous décrirons ensuite le corpus ainsi  
128 que les différentes expérimentations menées. Nous concluerons par une  
129 discussion sur les résultats obtenus et les pistes d'améliorations futures  
130 à explorer.



# CONTEXTE

## Sommaire

134	1.1	TAL et MIR . . . . .	9
135	1.2	La transcription automatique de la musique . . . . .	11
136	1.3	La transcription automatique de la batterie . . . . .	14
137	1.4	Les représentations de la musique . . . . .	14

## Introduction

La transcription automatique de la musique (AMT) est un défi ancien [1] 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)<sup>1</sup>.

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).

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<sup>2</sup> est une étude du traitement de la musique [2], en particulier des représentations musicales, de la transformée de Fourier pour la musique<sup>3</sup>, 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.

2. [https://en.wikipedia.org/wiki/Music\\_informatics](https://en.wikipedia.org/wiki/Music_informatics)

3. <https://interstices.info/de-fourier-a-la-reconnaissance-musicale/>

et de la reconnaissance des accords. 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 MIR<sup>4 5</sup> apparaît vers le début des années 2000 [3]. 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 qui veulent développer des systèmes de recommandation de musique ou des moteurs de recherche dédiés au son et à la musique.

173

#### 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

174

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 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

180

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

5. <https://ismir2021.ismir.net/>

et al. [4], 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. [5] 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.

*Ici, Digression sur la musicologie calculatoire vs linguistique computationnelle ?*

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. [6] 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.

Il semblerait que le domaine du TAL qui se rapproche le plus du MIR soit la reconnaissance de la parole. 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.

Similitudes :

Reconnaissance automatique de la parole :

signal  $\Rightarrow$  phonèmes  $\Rightarrow$  texte Transcription automatique de la musique :

signal  $\Rightarrow$  MIDI  $\Rightarrow$  partition Différence :

Texte (données linéaires)  $\neq$  partition (données structurées hiérarchiques)

## 1.2 La transcription automatique de la musique

En musique, la transcription<sup>6</sup> est la pratique consistant à noter un morceau ou un son qui n'était auparavant pas noté et/ou pas populaire en tant que musique écrite, par exemple, une improvisation de jazz ou une bande sonore de jeu vidéo. Lorsqu'un musicien est chargé de créer une partition à partir d'un enregistrement et qu'il écrit les notes qui composent le morceau en notation musicale, on dit qu'il a créé une transcription musicale de cet enregistrement.

L'objectif de la transcription automatique de la musique (AMT) [7] est de convertir la performance d'un musicien en notation musicale - un peu

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

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 [8]. 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 [8]. 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 [9]. 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 [1] évoquait déjà la représentation musicale en arbre syntaxique dans le but d'écrire automatiquement des partitions à partir de données audio en se basant sur un mimétisme psychologique de l'approche humaine. De même pour les chercheurs en audio James A. Moorer, Martin Piszczalski et Bernard Galler qui, en 1977<sup>7</sup>, ont utilisé leurs connaissances en ingénierie de l'audio et du numérique pour programmer un ordinateur afin de lui faire analyser un enregistrement musical numérique de manière à détecter les lignes mélodiques, les accords et les accents rythmiques des instruments à percussion.

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

La figure 1.1 est une proposition de Benetos *et al.* [7] qui représente l'architecture générale d'un système de transcription musicale. On y observe plusieurs sous-tâches de l'AMT :

- La séparation des sources à partir de l'audio.
- Le système de transcription :
  - Cœur du système :
    - ⇒ Algorithmes de détection des multi-pitches et de suivi des notes.
  - Quatres sous-tâches optionnelles accompagnent ces algorithmes :
    - identification de l'instrument ;
    - estimation de la tonalité et de l'accord ;

---

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

- détection de l'apparition et du décalage ;
- estimation du tempo et du rythme.
- Apprentissage sur des modèles acoustiques et musicologiques.
- *Optionnel* : Informations fournies de manière externe, soit fournie en amont (genre, instruments,...), soit par interaction avec un utilisateur (infos sur une partition incomplète).

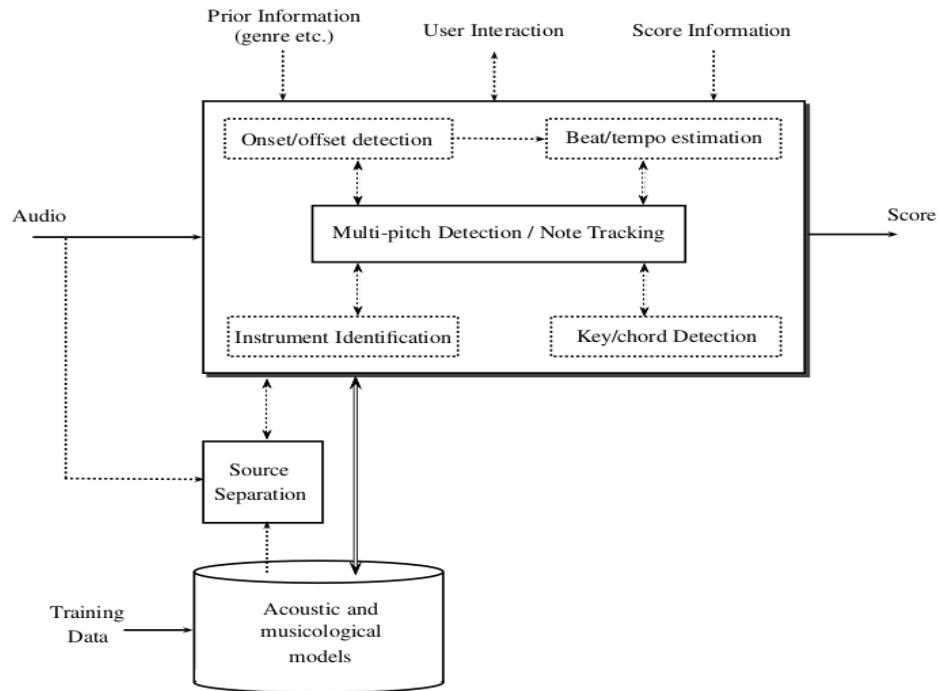


FIGURE 1.1 – Transcription automatique

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

### 269 1.3 La transcription automatique de la batterie

270 La batterie est un instrument récent qui s'est longtemps passé de parti-  
271 tion. En effet pour un batteur, la qualité de lecteur lorsqu'elle était néces-  
272 saire, résidait essentiellement dans sa capacité à lire les partitions des  
273 autres instrumentistes (par exemple, les grilles d'accords et la mélodie du  
274 thème en jazz) afin d'improviser un accompagnement approprié que per-  
275 sonne ne pouvait écrire pour lui à sa place.

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

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

289 Les applications de l'ADT seraient utiles dans tous les domaines musi-  
290 caux contenant de la batterie dont certains manque de partitions, notam-  
291 ment les musiques d'improvisation (jazz, pop) [7].

292 Mais aussi de manière plus générale dans le domaine du MIR. Si les or-  
293 dinateurs étaient capables d'analyser la partie de la batterie dans la mu-  
294 sique enregistrée, cela permettrait une variété de tâches de traitement de  
295 la musique liées au rythme. En particulier, la détection et la classification  
296 des événements sonores de la batterie par des méthodes informatiques  
297 est considérée comme un problème de recherche important et stimulant  
298 dans le domaine plus large de la recherche d'informations musicales [10].  
299 L'ADT est un sujet de recherche crucial pour la compréhension des aspects  
300 rythmiques de la musique, et a un impact potentiel sur des domaines plus  
301 larges tels que l'éducation musicale et la production musicale.

### 302 1.4 Les représentations de la musique

#### 303 Les données audio

304 Le fichier WAV<sup>8</sup> est une instance du Resource Interchange File Format  
305 (RIFF) défini par IBM et Microsoft. Le format RIFF agit comme une "en-  
306 veloppe" pour divers formats de codage audio. Bien qu'un fichier WAV

---

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

puisse contenir de l'audio compressé, le format audio WAV le plus courant est l'audio non compressé au format LPCM (linear pulse-code modulation). Le LPCM est également le format de codage audio standard des CD audio, qui stockent des données audio LPCM à deux canaux échantillonnées à 44 100 Hz avec 16 bits par échantillon. Comme le LPCM n'est pas compressé et conserve tous les échantillons d'une piste audio, les utilisateurs professionnels ou les experts en audio peuvent utiliser le format WAV avec l'audio LPCM pour obtenir une qualité audio maximale.

### Les données MIDI

Le MIDI<sup>9</sup> (Musical Instrument Digital Interface) est une norme technique qui décrit un protocole de communication, une interface numérique et des connecteurs électriques permettant de connecter une grande variété d'instruments de musique électroniques, d'ordinateurs et d'appareils audio connexes pour jouer, éditer et enregistrer de la musique.

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



FIGURE 1.2 – Exemple évènements avec durée

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

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

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

FIGURE 1.3 – Critère pour un évènement

326 Pour la batterie, les évènements sont considérés sans durée, nous ignore-  
 327 rons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le  
 328 *channel* ne nous sera pas utile non plus.  
 329 Ici, définir *Tick* et *channel*.

330

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



FIGURE 1.4 – Exemple évènements sans durée

331

332 On observe que toutes les durées sont identiques.

### 333 Les partitions



FIGURE 1.5 – Exemple de partition de piano



Une partition de musique<sup>10</sup> est un document qui porte la représentation systématique du langage musical sous forme écrite. Cette représentation est appelée transcription et elle sert à traduire les quatre caractéristiques du son musical :

- la hauteur;
- la durée;
- l'intensité;
- le timbre.

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

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

### 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.

```
<?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

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

354

355 Le figure 1.6 représente un do en clef de sol de la durée d'une ronde sur  
356 une mesure en 4/4.

## 357 **Conclusion**

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

362 Nous avons également établi que le MIR est né de l'AMT qui est un pro-  
363 blème ancien et très difficile et qu'il serait toujours très utile de le ré-  
364 soudre (autant pour l'AMT que pour l'ADT).

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

## ÉTAT DE L'ART

## Sommaire

2.1	Monophonique et polyphonique . . . . .	19
2.2	Audio vers MIDI . . . . .	20
2.3	MIDI vers partition . . . . .	21
2.4	Approche linéaire et approche hiérarchique . . . . .	21

## 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.

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 ont été faits sur l'identification des instruments monophoniques <sup>1</sup> [7]. 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 polyphoniques. L'estimation des hauteurs multiples (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

1. Instruments produisant une note à la fois, ou plusieurs notes de même durée (monophonie par accord).

396 plusieurs instruments différents. Ce défi est donc majeur pour la batte-  
397 rie puisque c'est un instrument qui est lui-même constitué de plusieurs  
398 instruments (caisse-claire, grosse-caisse, cymbales, toms, etc...). Le fort  
399 degré de chevauchement entre les durées ainsi qu'entre les fréquences  
400 complique l'identification des instruments polyphoniques. Cette tâche est  
401 étroitement liée à la séparation des sources et concerne aussi la sépara-  
402 tion des voix. Les performances des systèmes actuels ne sont pas encore  
403 suffisantes pour permettre la création d'un système automatisé capable  
404 de transcrire de la musique polyphonique sans restrictions sur le degré  
405 de polyphonie ou le type d'instrument. Cette question reste donc encore  
406 ouverte.

## 407 2.2 Audio vers MIDI

408 Jusqu'à aujourd'hui, les recherches se sont majoritairement concentrées  
409 sur le traitement du signal vers la génération du MIDI [11]. Cette partie  
410 englobe plusieurs sous-tâches dont la détection multi-pitches, la détection  
411 des onset et des offset, l'estimation du tempo, la quantification du rythme,  
412 la classification des genres musicaux, etc...

413 En ADT [10], plusieurs stratégies de répartition pré/post-processing  
414 sont possibles pour la détection multi-pitches. Entamer la détection dès  
415 le pré-processing, en supprimant les features non-pertinentes pendant  
416 la séparation des sources afin d'obtenir une meilleure détection des  
417 instruments de la batterie, est une démarche intuitive : supprimer  
418 la structure harmonique pour atténuer l'influence des instruments à  
419 hauteurs sur la détection grosse-caisse et caisse-claire en est un exemple.  
420 Mais certaines études montrent que des expériences similaires ont donné  
421 des résultats non-concluants et que la suppression des instruments à  
422 hauteurs peut avoir des effets néfastes sur les performances de l'ADT.  
423 En outre, les systèmes d'ADT basés sur des RNN ou des NMF font la  
424 séparation des sources pendant l'optimisation, ce qui réduit la nécessité  
425 de la faire pendant le pré-processing.

426 Pour la reconnaissance des instruments, une approche possible [12]  
427 est de mettre un modèle probabiliste dans l'étape de la classification  
428 des événements afin de classer les différents sons de la batterie. Cette  
429 méthode permet de se passer de samples audio isolés en modélisant la  
430 progression temporelle des features avec un HMM. Les features sont  
431 transformés en représentations statistiques indépendantes. L'approche  
432 AdaMa [13] est une autre approche de la même catégorie ; elle commence  
433 par une estimation initiale des sons de la batterie qui sont itérativement  
434 raffinés pour correspondre à (pour matcher) l'enregistrement visé.

435

## 2.3 MIDI vers partition

Le plus souvent, lorsque les articles abordent la transcription automatique de bout en bout (de l'audio à la partition), l'appellation « score » (*partition*) désigne un output au format Music XML, ou simplement MIDI. Par exemple, dans [8], la chaîne de traitement va jusqu'à la génération d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en extraire manuellement un fichier MusicXML contenant plusieurs voix. Seuls quelques travaux récents s'intéressent de près à la création d'outils permettant la génération de partition. Le problème de la conversion d'une séquence d'événements musicaux symboliques en une partition musicale structurée est traité notamment dans [14]. Ce travail, qui vise à résoudre en une fois la quantification du rythme et la production de partition, s'appuie tout au long du processus sur des grammaires génératives qui fournissent un modèle hiérarchique *a priori* des partitions. Les expériences ont des résultats prometteurs, mais il faut relever qu'elle ont été menées avec un ensemble de données composé d'extraits monophoniques ; il reste donc à traiter le passage au polyphonique en couplant le problème de la séparation des voix avec la quantification du rythme. L'approche de [14] est fondée sur la conviction que la complexité de la structure musicale dépasse les modèles linéaires.

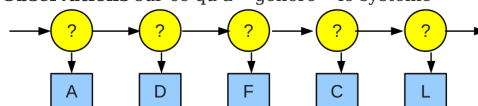
## 2.4 Approche linéaire et approche hiérarchique

Plusieurs travaux ont d'abord privilégié l'approche stochastique. Par exemple, Shibata *et al.* [8] ont utilisé le modèle de Markov caché (HMM)<sup>2</sup> pour la reconnaissance de la métrique. Les auteurs utilisent d'abord deux réseaux de neurones profonds, l'un pour la reconnaissance des pitches et l'autre pour la reconnaissance de la vélocité. Pour la dernière couche, la probabilité est obtenue par une fonction sigmoïde. Ils construisent ensuite plusieurs HMM métriques étendus pour la musique polyphonique correspondant à des métriques possibles, puis ils calculent la probabilité maximale pour chaque modèle afin d'obtenir la métrique la plus probable.

---

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

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



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

FIGURE 2.1 – HMM

466 *Source : Cours de Damien Nouvel*<sup>3</sup>

467

468 L'évaluation finale des résultats de [8] montre qu'il faut rediriger  
 469 l'attention vers les valeurs des notes, la séparation des voix et d'autres  
 470 éléments délicats de la partition musicale qui sont significatifs pour  
 471 l'exécution de la musique. Or, même si la quantification du rythme se fait  
 472 le plus souvent par la manipulation de données linéaires allant notam-  
 473 ment des *real time units* (secondes) vers les *musical time units* (temps,  
 474 métrique,...), de nombreux travaux suggèrent d'utiliser une approche  
 475 hiérarchique puisque le langage musical est lui-même structuré.

476 En effet, l'usage d'arbres syntaxiques est idéale pour représenter le  
 477 langage musical. Une méthodologie simple pour la description et l'af-  
 478 fichage des structures musicales est présentée dans [15]. Les RT y  
 479 sont évoqués comme permettant une cohésion complète de la notation  
 480 musicale traditionnelle avec des notations plus complexes. Jacquemard  
 481 *et al.* [16] propose aussi une représentation formelle du rythme, inspirée  
 482 de modèles théoriques antérieurs et dont l'objectif est la réécriture de  
 483 termes. Ils démontrent aussi l'application des arbres de rythmes pour  
 484 les équivalences rythmiques dans [17]. La réécriture d'arbres, dans un  
 485 contexte de composition assistée par ordinateur, par exemple, pourrait  
 486 permettre de suggérer à un utilisateur diverses notations possibles pour  
 487 une valeur rythmique, avec des complexités différentes.

488 La nécessité d'une approche hiérarchique pour la production automatique  
 489 de partition est évoquée dans [14]. Les modèles de grammaire qui y sont  
 490 exposés sont différents de modèles markoviens linéaires de précédents  
 491 travaux.

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

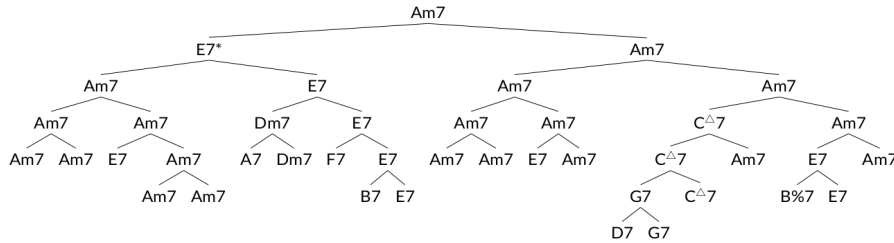
**Example: Summertime**

FIGURE 2.2 – arbre\_jazz

*Représentation arborescente d'une grille harmonique [9]***Conclusion**

La plupart des travaux déjà existants sur l'ADT ont été énumérés par Wu *et al.* [10] 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 [18], 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 flaps ou autres ornements ;
- 2) les choix d'écritures spécifiques à la batterie concernant la séparation des voix et les continuations.





512

513

## MÉTHODES

514

## Sommaire

515	3.1	La notation de la batterie . . . . .	25
516	3.2	Modélisation pour la transcription . . . . .	32
517	3.3	Qparse . . . . .	34
518	3.4	Les systèmes . . . . .	35
519			
520			
522			

523

## Introduction

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

526 Pour commencer, nous exposerons une description de la notation de la  
 527 batterie ainsi qu'une modélisation de celle-ci pour la représentation des  
 528 données rythmiques en arbres syntaxiques. Nous poursuivrons avec une  
 529 présentation de qparse<sup>1</sup>, un outil de transcription qui est développé par  
 530 Florent Jacquemard (Inria) au sein du laboratoire Cedric au CNAM.

531 Enfin, nous présenterons les systèmes.

## 532 3.1 La notation de la batterie



533 Une figure de note [19] de musique combine plusieurs critères<sup>2</sup> :

534 — Une tête de note :

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

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

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

- 537 — Une hampe :  
 538 Indicatrice d'appartenance à une voix en fonction de sa direction  
 539 et indicatrice d'une durée représentée par sa présence ou non  
 540 (blanche  $\neq$  ronde)
- 541 — Un crochet : La durée d'une note est divisée par deux à chaque  
 542 crochet ajouté à la hampe d'une figure de note.



FIGURE 3.1 – Rapport des figures de notes  
[19]

543 La figure 3.1 montre les rapports de durée entre les figures de notes. Plus  
 544 les durées sont longues, plus elles sont marquées par la tête de note (la  
 545 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de  
 546 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute  
 547 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.  
 548 Les notes à crochet (croche, double-croche, triple...) peuvent être reliées  
 549 ou non par des ligatures (Voir les 4 dernière lignes de la figure 3.1).

## 550 Les hauteurs et les têtes de notes

551 Pour la transcription, nous proposons une notation inspirée du recueil de  
 552 pièces pour batterie de J.-F. Juskowiak [20] et des méthodes de batterie  
 553 Agostini [21], car nous trouvons la position des éléments cohérente et in-  
 554 tuitive.

555 En effet, les hauteurs sur la portée représentent :

- 556 — La hauteur physique des instruments :  
 557 La caisse claire est centrale sur la portée et sur la batterie (au  
 558 niveau de la ceinture, elle conditionne l'écart entre les pédales et  
 559 aussi la position de tous les instruments basiques d'une batterie).  
 560 Tout ce qui en-dessous de la caisse-claire sur la portée est en  
 561 dessous de la caisse-claire sur la batterie (pédales, tom basse);  
 562 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est

563 aussi sur la batterie.

564

565 — La hauteur des instruments en terme de fréquences :

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

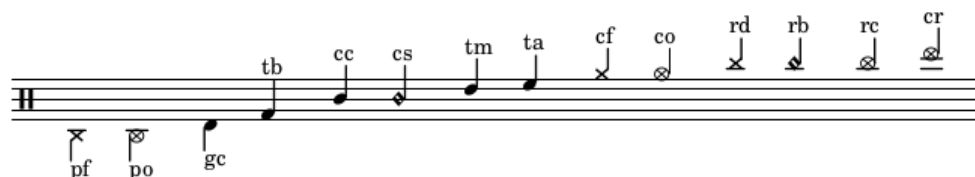


FIGURE 3.2 – Hauteur et têtes de notes

569 Les noms des instruments correspondant aux codes des notes de la figure  
570 3.2 sont dans le tableau 3.1.

### 571 Les durées

572 Comme nous venons de la voir, la majorité des instruments de la batte-  
573 rie sont représentés par les têtes des notes. Par conséquent, les symboles  
574 rythmiques concernant la tête de note ne pourront pas être utilisés. Cela  
575 est valable aussi pour la présence ou non de la hampe puisque ce phé-  
576 nomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition  
577 blanche-ronde). L'usage des blanches existe dans certaines partitions de  
578 batterie [22] mais cela reste dans des cas très rares. Certains logiciels per-  
579 mettent de faire des blanches avec des symboles spécifiques à la batterie  
580 ou aux percussions mais leur lecture reste peu aisée et leur utilisation  
581 pour la batterie est rarissime.

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

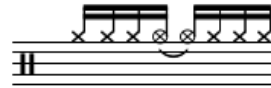
583 — Le point ;

584 — La liaison.

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



Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

589 L'écriture de la batterie doit faire ressortir la pulsation. La première chose  
 590 à prendre en compte pour analyser la figure 3.3 est donc la nécessité de  
 591 regrouper les notes par temps à l'aide des ligatures.

592 Exemple 1 : ouverture de charley quantifiée mais pas notes pas regrou-  
 593 pées par temps.

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

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

599 En cas de nécessité de rallonger la durée d'une note au-delà de son temps  
 600 initial et si cette note correspond à une ouverture de charley, on privilé-  
 601 giera la liaison.

## 602 Les silences

	la pause	la demi-pause	le soupir (2)	le demi-soupir	le quart de soupir	le huitième de soupir	le seizième de soupir
Silences							
Notes							
	la ronde	la blanche	la noire	la croche	la double croche	la triple croche	la quadruple croche

FIGURE 3.4 – Les silences

603 Les silences sont parfois utilisés pour quantifier les ouvertures de charley.  
 604 Les fermetures du charley sont notées soit par un silence (correspondant  
 605 à une fermeture de la pédale), soit par un écrasement de l'ouverture par  
 606 un autre coup de charley fermé, au pied ou à la main. Physiquement, le  
 607 charley est fermé par une pression du pied sur la pédale de charley. Dans  
 608 les fichiers MIDI, cette pression est traduite par un charley joué au pied.  
 609 Mais dans une vraie partition, cette écriture ne traduirait pas ce que le  
 batteur doit penser.

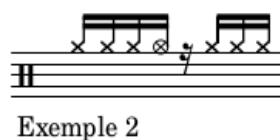
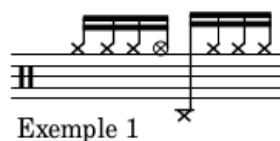


FIGURE 3.5 – Silence joué

610 L'exemple 1 de la figure 3.5 montre ce qui est écrit dans les données MIDI  
 611 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-  
 612 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1  
 613 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-  
 614 truments jouant simultanément.  
 615 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte  
 616 la note suivante pour l'écriture :  
 617 - Si c'est un charley fermé joué à la main  $\Rightarrow$  la note sera cf;  
 618 - Si c'est un charley fermé joué au pied  $\Rightarrow$  la note sera un silence.  
 619

## 620 Les équivalences rythmiques

621 Pour les instruments mélodiques, la liaison et le point sont les deux seules  
 622 possibilités en cas d'équivalence rythmique pour des notes dont la durée  
 623 de l'une à l'autre est ininterrompue. Mais pour la batterie, à part pour  
 624 les ouvertures de charley (voir section 3.1), les durées des notes n'ont pas  
 625 d'importance. L'usage des silences pour combler la distance rythmique  
 626 entre deux notes devient donc possible.  
 627 Cela pris en compte, et étant donné que les indications de durée dans les  
 628 têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide  
 629 de silences sera privilégiée comme indication de durée sauf dans les cas  
 630 où cela reste impossible. Ce choix a pour but de n'avoir qu'une manière  
 631 d'écrire toutes les notes, que leurs têtes de notes soit modifiées ou non.  
 632 Sur la figure 3.6, théoriquement, il faudra choisir la notation de la



FIGURE 3.6 – Équivalence

633 deuxième mesure mais dans certains contextes, pour des raisons de lisi-  
 634 bilité ou de surcharge, la version sans les silences de la troisième mesure  
 635 pourra être choisie.

### 636 Les voix

637 Les voix<sup>3</sup> désignent les différentes parties mélodiques constituant une  
 638 composition musicale et destinées à être interprétées, simultanément ou  
 639 successivement, par un ou plusieurs musiciens. En batterie, une voix est  
 640 l'ensemble des instruments qui, à eux seuls, constituent une phrase ryth-  
 641 mique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant  
 642 possibles pour un même rythme, on peut regrouper les instruments de la  
 643 batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou  
 644 2 voix. Sur la figure 3.7, il faudra faire un choix entre les exemples 1, 2 et  
 3 qui sont trois façons d'écrire le même rythme.



FIGURE 3.7 – Séparation des voix

645  
 646 Ce choix se fera en fonction des instruments joués, de la nature plus ou  
 647 moins systématique de leurs phrasés, et des associations logiques entre

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

les instruments dans la distribution des rythmes sur la batterie (voir la section 3.4).

### 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. » [19]

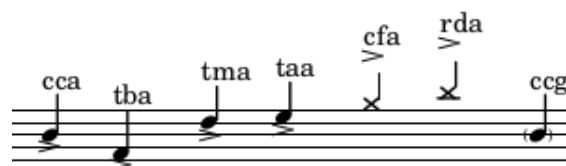


FIGURE 3.8 – Les accents et les ghost-notes

La figure 3.8 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 ainsi, mais ces choix devront être adaptés en fonction des différents systèmes reconnus (voir la section 3.4). Par exemple, pour les systèmes jazz, les ligatures pour les toms et la caisse-claire seront dirigés vers le bas, il faudra donc mettre les symboles d'accentuation correspondants au-dessus des têtes de notes.

La dernière note de la figure 3.8 montre un exemple de ghost-notes. Le parenthésage a été choisi car il peut être utilisé sur n'importe quelle note sans changer la tête de note.

Pour les codes, on prend le code de la note et on ajoute un « a » pour un accent et un « g » pour une ghost-note. Toutes les notes de la figure 3.8 sont exposées en situation réelle dans la figure 3.9.



FIGURE 3.9 – Exemple pour les accentuations et les ghost-notes

## 670 3.2 Modélisation pour la transcription

### 671 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

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

677 Malgré le large panel de pitches disponible, il semblerait qu'aucun pitch  
 678 ne désigne le charley ouvert joué au pied. Pourtant, dans la batterie mo-  
 679 derne, plusieurs rythmes ne peuvent fournir le son du charley ouvert  
 680 qu'avec le pied car les mains ne sont pas disponibles pour le jouer. Cela  
 681 doit en partie être dû à l'utilisation des boîte à rythmes en MAO qui ne né-  
 682 cessitent pas de faire des choix conditionnés par les limitations humaines  
 683 (2 pieds, 2 mains, et beaucoup plus d'instruments. . .)

### 684 La vélocité

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

687 Cette transcription nous a mené aux observations suivantes :

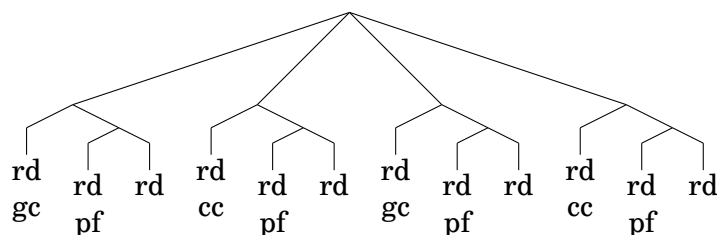
- 688 — Vélocité inférieure à 40 : ghost-note ;
- 689 — Vélocité supérieure à 90 : accent ;
- 690 — Pas d'intention d'accent ni de ghost-note pour une vélocité entre 40  
 691 et 89 ;



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

## 702 Les arbres de rythmes

703 Les arbres de rythmes représentent un rythme unique dont les possibi-  
 704 lités de notation sur une partition sont théoriquement multiples. Voici  
 705 une représentation de la figure 3.7 en arbre de rythmes avec les codes de  
 chaque instrument :



706  
 707 Ci-dessous, le même arbre dont les codes des instruments sont remplacés  
 par leurs données MIDI respectives :



708  
 709 Chacun des trois exemples de la figure 3.7 est représenté par un des deux  
 710 arbres syntaxiques ci-dessus.

### 3.3 Qparse

La librairie Qparse<sup>4</sup> implémente la quantification des rythmes basée sur des algorithmes d'analyse syntaxique pour les automates arborescents pondérés. En prenant en entrée une performance musicale symbolique (séquence de notes avec dates et durées en temps réel, typiquement un fichier MIDI), et une grammaire hors-contexte pondérée décrivant un langage de rythmes préférés, il produit une partition musicale. Plusieurs formats de sortie sont possibles, dont XML MEI. Les principaux contributeurs sont :

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

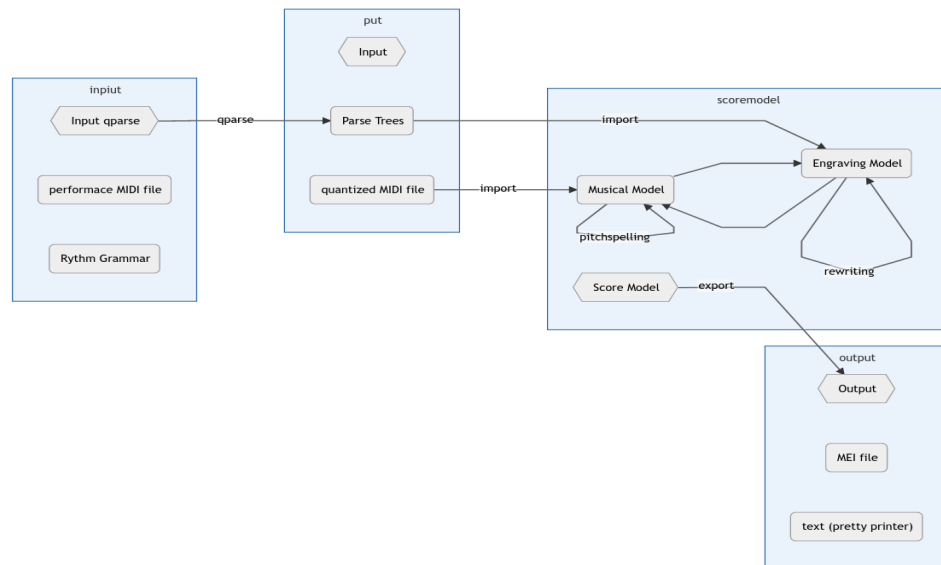


FIGURE 3.10 – Présentation de Qparse

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

Explication des différentes étapes de la figure 3.10<sup>5</sup> :

— **Input Qparse :**

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

— **Arbre de parsing :**

Les données MIDI sont quantifiées, les notes de dates proches sont 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

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

766 sur 1 ou 2 voix)

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

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

772

773 Un ensemble de systèmes comprenant leur métrique et leurs règles  
774 spécifiques de réécriture sera nécessaire. Les systèmes devront être  
distribués dans 4 grandes catégories :

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

TABLE 3.2 – Systèmes

775

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

777 — 4/4 binaire

778 — 4/4 jazz

779 — 4/4 afro-cubain

## 780 Objectif des systèmes

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

782 — définir une métrique ;

783 — choisir une grammaire appropriée ;

784 — fournir les règles de réécriture (séparation des voix et simplifica-  
785 tion.

786

787 La partie *motif* des systèmes sera utilisée pour la **définition des mé-**  
788 **triques**. Le *motif* et la gammes des systèmes seront utilisés pour la **sé-**  
789 **paration des voix**. Les règles de **simplification** (les combinaisons de  
790 réécritures) seront extraites des voix séparées des systèmes.

## 791 Détection d'indication de mesure

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



Exemple 1



Exemple 2

FIGURE 3.11 – Métrique

796 La figure 3.11 montre deux indications de mesure différentes. L'une  
797 (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des trio-  
798 lets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est tra-  
799 ditionnellement écrit en binaire avec ou sans triolet (même si cette mu-  
800 sique est dite ternaire alors que le rock ternaire sera plutôt écrit comme  
801 dans l'exemple 2).

#### 802 **Choix d'une grammaire**

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

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

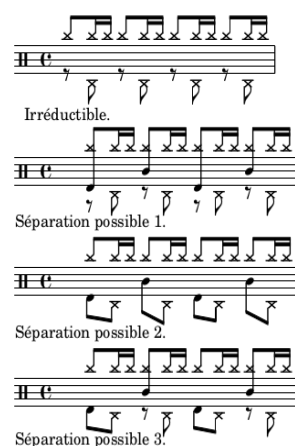
815 **Séparation des voix**

FIGURE 3.12 – Motif 4-4 binaire

816 Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les  
 817 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système  
 818 est constitué par défaut du motif rd/pf/cc (voir 3.1) et d'un texte joué à la  
 819 grosse-caisse. La première ligne de la figure 3.12 est appelée « Irréductible  
 820 » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du  
 821 charley au pied. La troisième séparation proposée est privilégiée car elle  
 822 répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les  
 823 pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont  
 824 utilisés par voix et plus logique pour le lecteur puisque les mains sont en  
 haut et les pieds en bas.

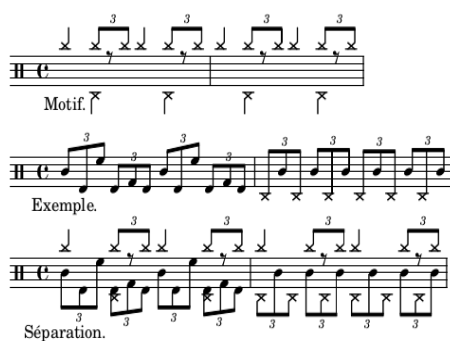


FIGURE 3.13 – Motif 4-4 jazz

825 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré  
 826 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :  
 827 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,  
 828 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et  
 829

830 2 combinées avec le *motif* de la première ligne, sont des cas typiques de  
 831 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,  
 832 la grosse caisse entre très souvent dans le flot des combinaisons de toms  
 833 et de caisse claire et son écriture séparée serait inutilement compliquée  
 834 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-  
 835 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de  
 charley en bas.



FIGURE 3.14 – Système 4-4 afro-latin

836  
 837 La figure 3.14 montre un exemple minimaliste de système afro-latin [22].  
 838 Ce système doit être écrit sur trois voix car la voix centrale est souvent  
 839 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas  
 840 serait surchargé et peu lisible.

#### 841 **Simplification de l'écriture**

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

844  
 845 Les gammes qui accompagnent les motifs d'un système étayent toutes les  
 846 combinaisons d'un système et elles permettent, combinées avec le motif  
 847 d'un système, de définir les règles de simplification propres à celui-ci.

848  
 849 Voici les différentes étapes à suivre :

- 850 — Pour chaque gamme du système, faire un arbre de rythme repré-  
 851 sentant la gamme combinée avec le motif du système ;
- 852 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un  
 853 arbre de rythme par voix ;
- 854 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les  
 855 nœuds qui nécessitent une simplification et écrire la règle.

856  
 857 Certaines précisions concernant l'extraction de ces règles sont néces-  
 858 saires. Il s'agit de précisions à propos de la durée, des silences et de la  
 859 présence ou non d'ouverture de charley dans les instruments joués. Nous  
 860 avons discuté de ces problèmes dans le chapitre 3.

861  
 862 Voici quelques règles inhérentes à la simplification de l'écriture pour  
 863 la batterie :

864 Toutes les continuations (t) qui se trouvent en début de temps (figures 4.9,  
 865 4.11 et 4.12) sont transformées en silences (r) sauf si la note précédente  
 866 est un charley ouvert ?  
 867 Même si on favorise l'usage des silences pour l'écart entre les notes  
 868 n'appartenant pas au même temps, on les supprime systématiquement  
 869 pour 2 notes au sein d'un même temps et favorise, une liaison si co, un  
 870 point si pas co et nécessaire, un simple ajustement de la figure de note si  
 suffisant.

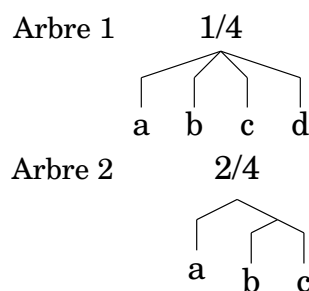


FIGURE 3.15 – Simplification

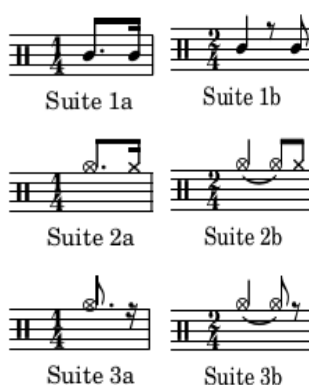


FIGURE 3.16

871  
 872 Soit l'arbre 1 de la figure 3.15 dans lequel :  
 873 a et d sont des instruments de la batterie (x);  
 874 b et c sont des continuations (t); Pour chacune des conditions suivantes,  
 875 une suite de la figure 3.16 est attribuée :  
 876 — Si a n'est pas un co :  
 877       ⇒ Suite 1a.  
 878 — Si a est un co :  
 879       — Si d est un cf :  
 880       ⇒ Suite 2a.



881 — Si d est un pf :  
882     ⇒ Suite 3a : d deviens un silence (r).  
883  
884 Soit l'arbre 2 de la figure 3.15 dans lequel :  
885 a et c sont des instruments de la batterie (x);  
886 b est une continuation (t); Pour chacune des conditions suivantes, une  
887 suite de la figure 3.16 est attribuée :  
888 — Si a n'est pas un co :  
889     ⇒ Suite 1b, b devient un silence.  
890 — Si a est un co :  
891 — Si c est un cf :  
892     ⇒ Suite 2b, b devient une liaison et c devient un cf.  
893 — Si c est un pf :  
894     ⇒ Suite 3b : b deviens une liaison et c devient un silence.  
895  
896 *Rappel :*  
897 *cf = charley fermé joué à la main ;*  
898 *co = charley ouvert joué à la main ;*  
899 *pf = charley fermé joué au pied.*  
900  
901 **Problème : le cf et le co ne seront jamais sur la même voix que le**  
902 **pf... Par conséquent, les règles concernant les charleys ouverts**  
903 **doivent-elles être appliquées sur l'arbre de parsing de l'input ?...**

## 904 Conclusion

905 Nous avons formalisé une notation de la batterie, modélisé cette notation  
906 pour la transcription de données MIDI en partition, nous avons décrit  
907 Qparse.  
908 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-  
909 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-  
910 propriée et énoncer des règles de séparation des voix et de simplification  
911 de l'écriture.



# EXPÉRIMENTATIONS

## Sommaire

4.1	Le jeu de données . . . . .	43
4.2	Analyse MIDI-Audio . . . . .	45
4.3	Expérimentation théorique d'un système . . . . .	49
4.4	Résultats et discussion . . . . .	54

## 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 Dataset<sup>1</sup> [23] (GMD) qui est un jeu de données mis à disposition par Google sous la licence Creative Commons Attribution 4.0 International (CC BY 4.0).

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

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



FIGURE 4.1 – Batterie électronique

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

940 Autres critères spécifiques au GMD :

- 941 — Toutes les performances ont été jouées au métronome et à un tempo
- 942 choisi par le batteur.
- 943 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 944 qui ont pu improviser dans un large éventail de styles. Les don-
- 945 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 946 (professionnel ou amateur).
- 947 — Les batteurs avaient pour instruction de jouer des séquences de
- 948 plusieurs minutes ainsi que des fills<sup>2</sup>
- 949 — Chaque performance est annotée d'un style (fourni par le batteur),
- 950 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 951 nyme du batteur.
- 952 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 953 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 954 session du GMD.
- 955 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 956 fichier MIDI.

## 957 **Format des données**

958 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-  
959 tinctes :

- 960 — une pour le tempo et l'indication de mesure ;
- 961 — une pour les changements de contrôle (position de la pédale de
- 962 charley) ;
- 963 — une pour les notes.

964

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

967 Pour simplifier le traitement de ces données, ces trois pistes ont été  
968 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.

969  
 970 « Control Changes The TD-11 also records control changes speci-  
 971 fying the position of the hi-hat pedal on each hit. We have preserved this  
 972 information under control 4. »  
 973 (<https://magenta.tensorflow.org/datasets/groove>)  
 974 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-  
 975 tions dans les fichiers MIDI.  
 976 L'utilisation de pretty\_midi devient urgente !

## 977 4.2 Analyse MIDI-Audio

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

### 980 Comparaisons de transcriptions

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

#### 985 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



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

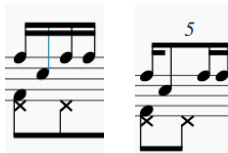
---

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

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

991 **Exemple d'analyse 2**

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 992 — Les doubles croches ont été interprétées en quintolet  
 993 — La deuxième double-croche est devenue une croche.  
 994

995 **Exemple d'analyse 3**

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 996 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.  
 997 — Les toms basses des temps 1 et 2 de la mesure 2 de la TM ont été décalés d'une double croche vers la droite dans la TA.  
 998 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.  
 999 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.  
 1000  
 1001  
 1002  
 1003  
 1004

1005 **Exemple d'analyse 4**

Transcription manuelle  $\Rightarrow$  Transcription automatique



- 1006  
 1007 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une  
 1008 manière excessivement complexe!

1009 **Exemple avec des flas**

1010 Transcription manuelle



1011

1012 Transcription automatique

1013



1014

1015

- 1016 — Le premier fla est reconnu comme étant un triolet contenant une
- 1017 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1018 ornementée.
- 1019 — Le deuxième fla est reconnu comme étant un accord.
- 1020 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1021 dans la TA.
- 1022 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1023 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1024 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1025 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1026 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1027 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1028 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1029 temps 3 de la TA.
- 1030 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1031 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1032 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1033 cross-sticks.

## 1034 Transcription de partition

FIGURE 4.2 – Partition de référence

1035 La figure 4.2 est la transcription manuelle des fichiers *004\_jazz-*  
 1036 *funk\_116\_beat\_4-4.mid* et *004\_jazz-funk\_116\_beat\_4-4.wav* du GMD.

1037 Cette transcription a été entièrement faite avec Lilypond (voir le code  
 1038 lilypond sur le git [https://github.com/MartinDigard/Stage\\_M2\\_](https://github.com/MartinDigard/Stage_M2_Inria)  
 1039 *Inria*) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est  
 1040 annoncé dans le GMD de style «jazz-funk» probablement en raison de  
 1041 la ride de type shabada rapide (le ternaire devient binaire avec la vi-  
 1042 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux  
 1043 et quatre).

1044 La transcription des données audio et MIDI contenues dans ces fichiers  
 1045 a permis une analyse plus approfondie des critères à relever pour chaque  
 1046 évènement MIDI et de la manière de les considérer dans un objectif de  
 1047 transcription en partition lisible pour un musicien (Voir la section 3.2).



### 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.

#### 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

1071 dont la première ne contient que des croches et la deuxième que des  
 1072 doubles-croches en l'air. Ces deux possibilités sont combinées de manière  
 1073 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

#### 1074 **Systèmes — motifs et gammes combinés**

1075 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1  
 1076 de la figure 4.3.

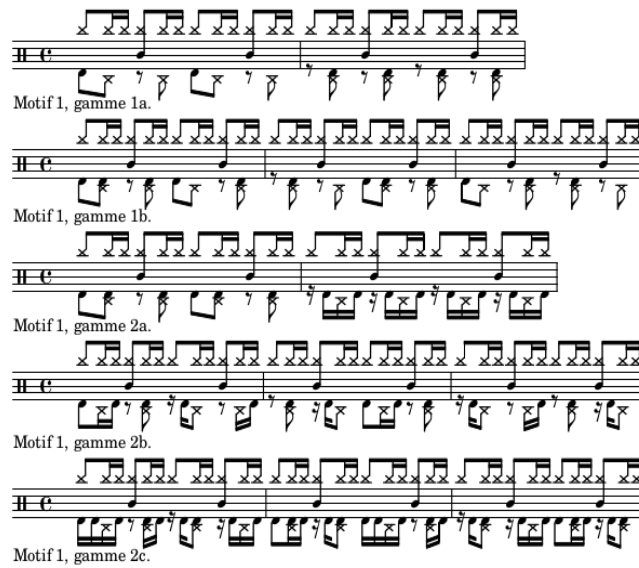


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1077

#### 1078 **Représentation du système en arbres de rythmes**

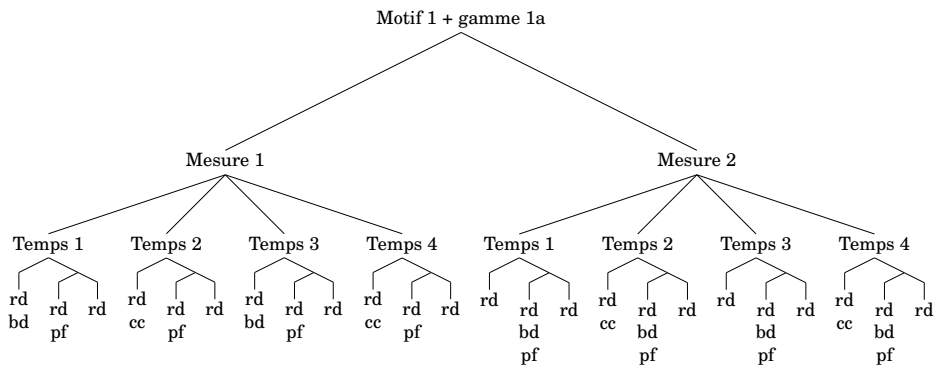


FIGURE 4.5 – Arbre de rythme — système

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.

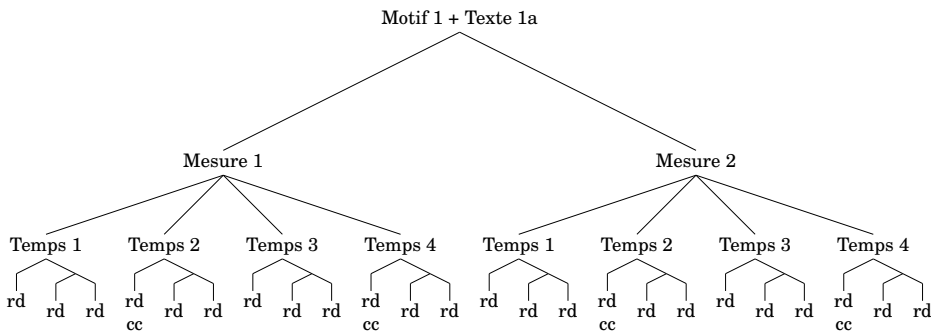


FIGURE 4.6 – Arbres de rythme — voix haute

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

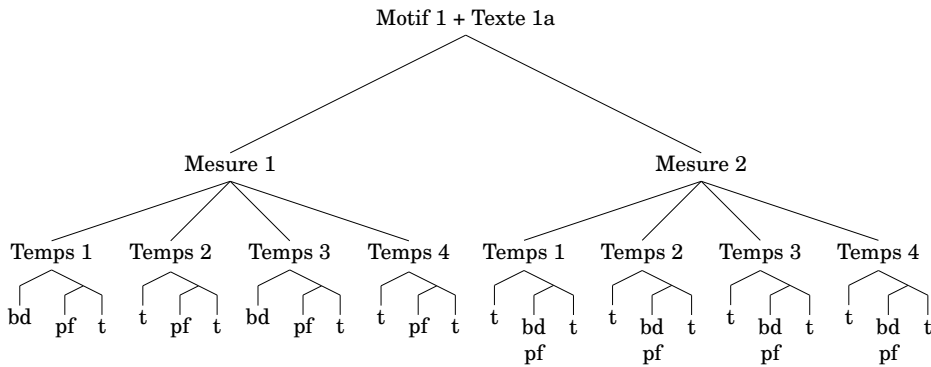


FIGURE 4.7 – Arbres de rythme — voix basse

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

### 1094 Les règles de simplifications

1095 L'objectif des règles de simplifications est de réécrire les écarts de durées  
 1096 qui séparent les notes d'une manière appropriée pour la batterie et qui  
 1097 soit la plus simple possible. Les ligatures relient les notes d'un temps  
 1098 entre elles (rendre la pulse visuelle).

1099

1100 Pour les figures ci-dessous :

1101 — x = une note ;

1102 — r = un silence ;

1103 — t = une continuation (point ou liaison)

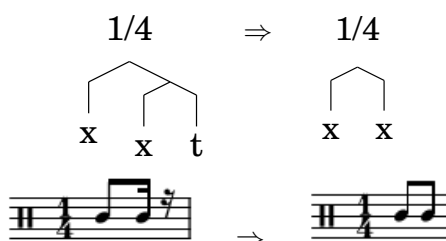


FIGURE 4.8

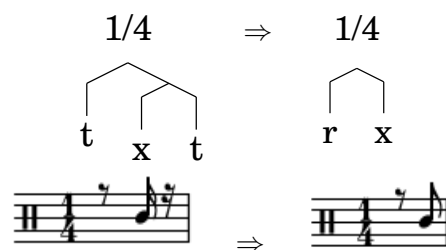


FIGURE 4.9

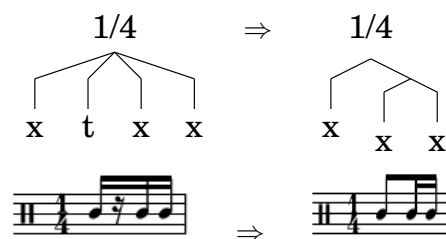


FIGURE 4.10

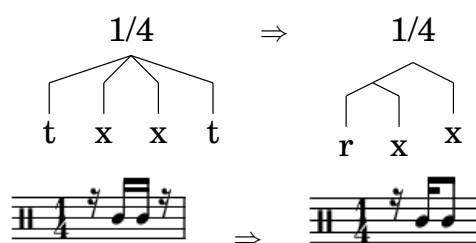


FIGURE 4.11

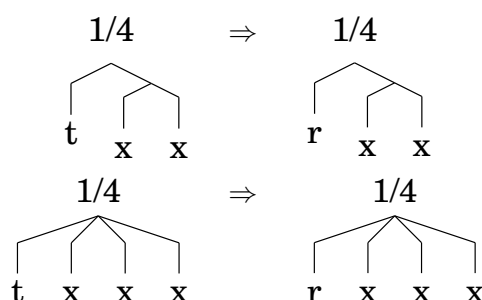


FIGURE 4.12

1104 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres  
 1105 manquants seront mis en annexe.

1106

1107 Les règles remplacent par un silence les continuations (t) qui sont  
 1108 au début d'un temps. Cela est valable pour ce système mais lorsqu'il y a  
 1109 des ouvertures de charley, cela n'est pas toujours applicable. Ce problème  
 1110 est évoqué de le chapitre 3.

1111

1112  $\Rightarrow$  **Objectif de cette expérimentation théorique :**

1113 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette  
 1114 expérimentation théorique a pour but d'orienter la recherche d'autres  
 1115 systèmes par observation du jeu de données et de montrer comment les  
 1116 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

## 1117 4.4 Résultats et discussion

1118 Cette section regroupe les avancées qui ont été réalisées par rapport aux  
 1119 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-  
 1120 tats de l'ADT avec Qparse. Nous avons amélioré le système de quantifi-  
 1121 cation de Qparse pour la batterie, notamment le passage à la polyphonie  
 1122 avec les Jams. Nous avons pu obtenir des arbres de parsing correctes en  
 1123 améliorant les grammaires avec des fichiers MIDI courts. Puis, une sortie  
 1124 MEI a aussi été obtenue (encore à vérifier).

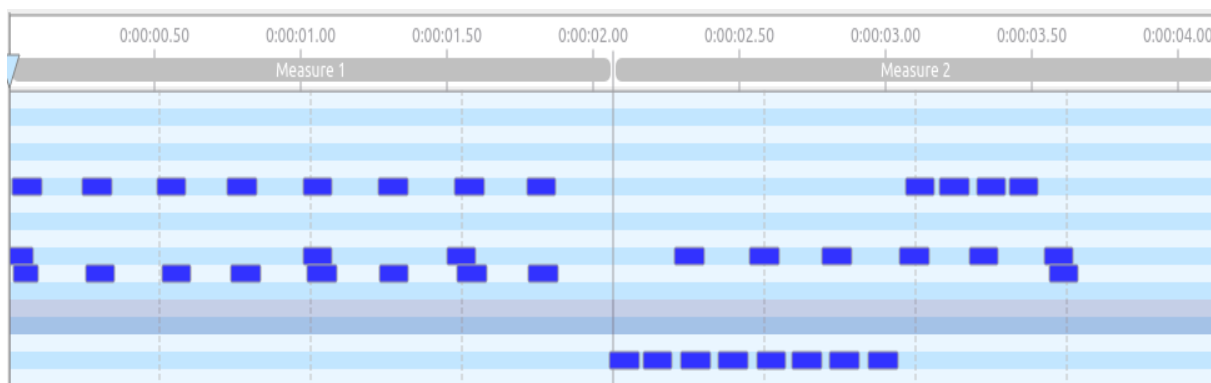
### 1125 Les Jams

1126 Les Jams permettent de passer du monophonique au polyphonique.

### 1127 Le parsing

1128 Tests effectués avec le fichier midi suivant :

1129



1130

1131

1132 Un premier test convaincant est effectué avec la grammaire sui-  
 1133 vante :

1134

1135 // bar level

1136 0 -> C0 1

1137 0 -> E1 1

1138 0 -> U4(1, 1, 1, 1) 1

1139

1140 // half bar level

1141 9 -> C0 1

1142 9 -> E1 1

1143

1144 // beat level

1145 1 -> C0 1

1146 1 -> E1 1

```

1147 1 -> T2(2, 2) 1
1148 1 -> T4(4, 4, 4, 4) 1
1149

```

```

1150 // croche level
1151 2 -> C0 1
1152 2 -> E1 1
1153

```

```

1154 // double level
1155 4 -> C0 1
1156 4 -> E1 1
1157 4 -> E2 1
1158 4 -> T2(6, 6) 1
1159

```

```

1160 // triple level
1161 6 -> E1 1
1162

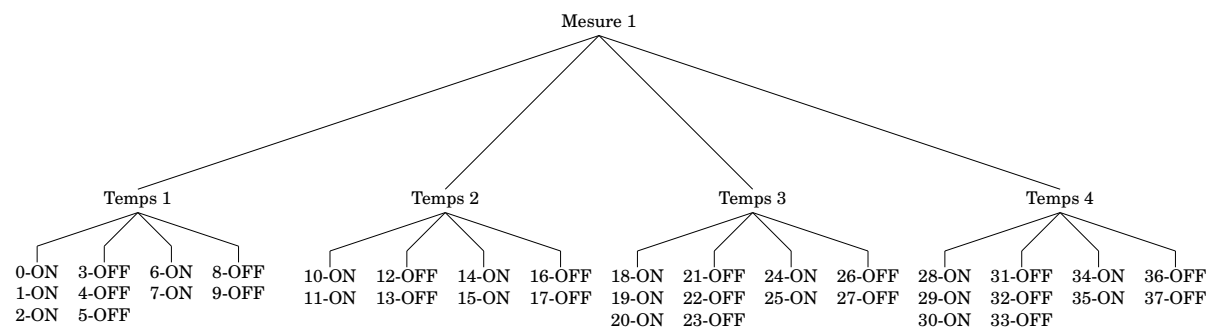
```

1163 Cette grammaire sépare les ligatures par temps au niveau de la  
 1164 mesure. Puis, au niveau du temps, elle autorise les divisions par deux  
 1165 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.  
 1166 L'arbre de parsing en résultant est considéré comme « convaincant » car  
 1167 il découpe correctement les mesures et les temps.

```

1168
1169

```



```

1170
1171

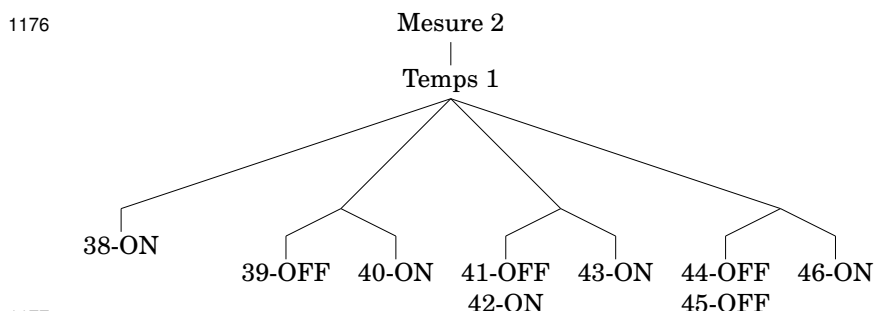
```

1172 Les temps de la première mesure du fichier MIDI sont bien quanti-  
 1173 fié mais ceux de la deuxième mesure présentent quelques défauts de  
 1174 quantification visibles dès le premier temps.

```

1175

```



1179 Les Onsets sont correctement triés au niveau des doubles croches  
 1180 mais certaines doubles croches sont inutilement subdivisées en triples  
 1181 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps  
 1182 ci-dessus).

### 1183 2ème exemple :

1184 Après une augmentation du poids des triples croches dans la grammaire  
 1185 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à  
 1186 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets  
 1187 sont bien triés et aucuns ne sont subdivisés.  
 1188

## 1189 Évaluation

1190 Pour l'évaluation, il aurait fallu produire un module.  
 1191 L'évaluation est-elle automatique ou manuelle?  
 1192 Possibilité d'un export lilypond en arbre pour comparer l'output avec la  
 1193 transcription manuelle.  
 1194 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel  
 1195 ou MEI pour les comparer et faire des statistiques. Si transformés en  
 1196 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)  
 1197 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.  
 1198 L'expérimentation peut-être considérer comme une évaluation manuelle?  
 1199 (magicien d'Oz)  
 1200 Lilypond vers MIDI + output vers MIDI  $\Rightarrow$  Comparaison des MIDI  
 1201 dumpés.  
 1202

## 1203 Discussion

1204 Dans cette section, nous discuterons sur la pertinence de l'ensemble des  
 1205 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui  
 1206 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.  
 1207 Écrire des règles de réécriture spécifique aux charley avec un système  
 1208 approprié. Le jeu de système



- 1209 — implémenter un pattern. . .
- 1210     ⇒ manque de temps ?
- 1211
- 1212 — La partie résultat est manquante car :
- 1213     ⇒ Sujet très difficile ;
- 1214     ⇒ Matcher les motifs peut être fait ultérieurement ;
- 1215         Mais ce travail aurait été indispensable pour obtenir une quan-
- 1216         tité de résultats qui justifieraient une évaluation automatique
- 1217         permettant de faire des graphiques.
- 1218
- 1219 — L'évaluation fut entièrement manuelle car :
- 1220     ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
- 1221         VS output)
- 1222 — Le ternaïre jazz (voir expérience 2)
- 1223 — Reconnaissance d'un motif sur le MIDI
- 1224     Reconnaître un motif (système) sur une mesure de l'input (un fi-
- 1225     chier midi représentant des données audios)
- 1226     ⇒ Motif (système) reconnu : true ou false
- 1227     Si true :
- 1228         - Choisir la grammaire correspondante ;
- 1229         - Parser le MIDI ;
- 1230         - Appliquer les règles de réécritures (Séparation des voix et simpli-
- 1231         fication)
- 1232 — Nous travaillerons aussi sur la détection de répétitions sur plu-
- 1233     sieurs mesures afin de pouvoir corriger des erreurs sur une des
- 1234     mesures qui aurait dû être identique aux autres mais qui présente
- 1235     des différences.
- 1236 — dans quelle catégorie mettre le shuffle ?
- 1237 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
- 1238 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
- 1239 à la batterie a été générée. Elle sera un bon point de départ pour les
- 1240 travaux futurs dont plusieurs propositions sont énoncés dans le présent
- 1241 document.



## CONCLUSION GÉNÉRALE

1243 Dans ce mémoire, nous avons traité de la problématique de la transcrip-  
1244 tion automatique de la batterie. Son objectif était de transcrire, à partir  
1245 de leur représentation symbolique MIDI, des performances de batteur de  
1246 différents niveaux et dans différents styles en partitions écrites.  
1247 Nous avons avancé sur le parsing des données MIDI établissant un pro-  
1248 cessus de regroupement des événements MIDI qui nous a permis de faire  
1249 la transition du monophonique vers le polyphonique. Une des données  
1250 importante de ce processus était de différencier les nature des notes d'un  
1251 accord, notamment de distinguer lorsque 2 notes constituent un accord  
1252 ou un *fla*.  
1253 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-  
1254 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-  
1255 tionnables en amont du parsing, soit par indication des noms des fichiers  
1256 MIDI, soit par reconnaissance de la métrique avec une approche diction-  
1257 naire de patterns prédéfinis<sup>5</sup> qu'il serait pertinent de mettre en œuvre en  
1258 machine learning.  
1259 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre  
1260 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif  
1261 d'un système et pour la simplification grâce aux gammes du motif d'un  
1262 système. Nous avons aussi montré comment, dans des travaux futurs, un  
1263 système dont le motif serait reconnu en amont dans un fichier MIDI pour-  
1264 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-  
1265 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans  
1266 la chaîne de traitement en terme de réécriture.  
1267 Il sera également intéressant d'étudier comment l'utilisation de LM peut  
1268 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération  
1269 entièrement automatisée de partitions de batterie et au problème général  
1270 de l'AMT de bout en bout.[7]

---

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



## BIBLIOGRAPHIE

- 1272 [1] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 9  
1273 et 12.
- 1274 [2] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité  
1275 page 9.
- 1276 [3] Caroline Traube. Quelle place pour la science au sein de la musico-  
1277 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 10.
- 1278 [4] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions  
1279 entre les traitements de la musique et du langage. *La Lettre des*  
1280 *Neurosciences*, 58 :24–26, 2018. – Cité page 11.
- 1281 [5] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.  
1282 Techniques de traitement automatique du langage naturel appli-  
1283 quées aux représentations symboliques musicales. In *JIM 2021 -*  
1284 *Journées d’Informatique Musicale*, Virtual, France, July 2021. –  
1285 Cité page 11.
- 1286 [6] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering  
1287 music relations with sequential attention. In *NLP4MUSA*, 2020. –  
1288 Cité page 11.
- 1289 [7] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger  
1290 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-  
1291 lenges and future directions. *Journal of Intelligent Information Sys-*  
1292 *tems*, 41, 12 2013. – Cité pages 11, 12, 14, 19 et 59.
- 1293 [8] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local  
1294 musical statistics as guides for audio-to-score piano transcription.  
1295 *Information Sciences*, 566 :262–280, 2021. – Cité pages 12, 21 et 22.
- 1296 [9] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J  
1297 O’Donnell, and Martin Rohrmeier. The jazz harmony treebank. –  
1298 Cité pages 12 et 23.
- 1299 [10] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-  
1300 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander  
1301 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*  
1302 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–  
1303 1483, 2018. – Cité pages 14, 20 et 23.

- 1304 [11] Moshekwa Malatji. Automatic music transcription for two instru-  
1305        ments based variable q-transform and deep learning methods, 10  
1306        2020. – Cité page 20.
- 1307 [12] Antti J. Eronen. Musical instrument recognition using ica-based  
1308        transform of features and discriminatively trained hmms. *Seventh*  
1309        *International Symposium on Signal Processing and Its Applications,*  
1310        2003. *Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 20.
- 1311 [13] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic  
1312        drum sound description for real-world music using template adap-  
1313        tation and matching methods. *International Conference on Music*  
1314        *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 20.
- 1315 [14] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-  
1316        sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-  
1317        tization and Score Structuring. In *MCM 2019 - Mathematics and*  
1318        *Computation in Music*, volume Lecture Notes in Computer Science  
1319        of *Proceedings of the Seventh International Conference on Mathema-*  
1320        *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June  
1321        2019. Springer. – Cité pages 21 et 22.
- 1322 [15] C. Agon, K. Haddad, and G. Assayag. Representation and rende-  
1323        ring of rhythm structures. In *Proceedings of the First International*  
1324        *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.  
1325        IEEE Computer Society. – Cité page 22.
- 1326 [16] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A  
1327        Term Rewriting Based Structural Theory of Rhythm Notation. Re-  
1328        search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –  
1329        Cité page 22.
- 1330 [17] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating  
1331        equivalent rhythmic notations based on rhythm tree languages. In  
1332        *Third International Conference on Technologies for Music Notation*  
1333        *and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-  
1334        pez Palma and Mike Solomon. – Cité page 22.
- 1335 [18] R. Marxer and J. Janer. Study of regularizations and constraints in  
1336        nmf-based drums monaural separation. In *International Conference*  
1337        *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,  
1338        02/09/2013 2013. – Cité page 23.
- 1339 [19] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41  
1340        rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,  
1341        1996. – Cité pages 25, 26 et 31.
- 1342 [20] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions  
1343        Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 26.

- 1344 [21] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue  
1345 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 26.
- 1346 [22] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-  
1347 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -  
1348 94120 Fontenay-sous-Bois, 2000. – Cité pages 27 et 39.
- 1349 [23] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David  
1350 Bamman. Learning to groove with inverse sequence transforma-  
1351 tions. In *International Conference on Machine Learning (ICML)*,  
1352 2019. – Cité page 43.

