

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	4
20	Introduction générale	5
21	1 Contexte	7
22	1.1 TAL et MIR	7
23	1.2 La transcription automatique de la musique	9
24	1.3 La transcription automatique de la batterie	11
25	1.4 Les représentations de la musique	12
26	2 État de l'art	17
27	2.1 Monophonique et polyphonique	17
28	2.2 Audio vers MIDI	18
29	2.3 MIDI vers partition	19
30	2.4 Approche linéaire et approche hiérarchique	19
31	3 Méthodes	23
32	3.1 La notation de la batterie	23
33	3.2 Modélisation pour la transcription	30
34	3.3 Qparse	32
35	3.4 Les systèmes	33
36	4 Expérimentations	41
37	4.1 Le jeu de données	41
38	4.2 Analyse MIDI-Audio	43
39	4.3 Expérimentation théorique d'un système	47
40	4.4 Résultats et discussion	52
41	Conclusion générale	57
42	Bibliographie	59

LISTE DES FIGURES

44	1.1	Transcription automatique	11
45	1.2	Exemple évènements avec durée	13
46	1.3	Critère pour un évènement	13
47	1.4	Exemple évènements sans durée	13
48	1.5	Exemple de partition de piano	14
49	1.6	MusicXML	15
50	2.1	HMM	20
51	2.2	arbre_jazz	21
52	3.1	Rapport des figures de notes	24
53	3.2	Hauteur et têtes de notes	25
54	3.3	Point et liaison	26
55	3.4	Les silences	26
56	3.5	Silence joué	27
57	3.6	Équivalence	28
58	3.7	Séparation des voix	28
59	3.8	Les accents et les ghost-notes	29
60	3.9	Exemple pour les accentuations et les ghost-notes	29
61	3.10	Présentation de Qparse	32
62	3.11	Métrique	35
63	3.12	Motif 4-4 binaire	36
64	3.13	Motif 4-4 jazz	36
65	3.14	Système 4-4 afro-latin	37
66	3.15	Simplification	38
67	3.16	38
68	4.1	Batterie électronique	42
69	4.2	Partition de référence	46
70	4.3	Motifs et gammes	47
71	4.4	Partition d'un système en 4/4 binaire	48
72	4.5	Arbre de rythme — système	48
73	4.6	Arbre de rythme — voix haute	49
74	4.7	Arbre de rythme — voix basse	49
75	4.8	50
76	4.9	50
77	4.10	50
78	4.11	51

79	4.12	51
----	----------------	----

80

LISTE DES TABLEAUX

81	3.1 Pitches et instruments	30
82	3.2 Systèmes	34

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 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 audio ou symboliques. La musique et les langues naturelles sont deux moyens que nous servent à exprimer nos ressentis sur le monde et les choses : « La musique s'écrit et se lit aussi facilement qu'on lit et écrit les paroles que nous prononçons. » [1]. Cet exercice nécessite la manipulation d'un langage musical codifié par 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 entraînent conditionnent les choix d'écriture. 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.

Voici la proposition de ce mémoire ainsi que les contributions apportées lors du stage : Rédiger entièrement la liste à puce qui suit.

- 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 théorique 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 :

"ayant toutes pour objectif..." → dans le domaine de la transcription musicale automatique objectif d'améliorer **qparse**, un outil de transcription automatique de et plus spécifiquement, dans le cas de la batterie.

transcription musicale, suivant des méthodes communes/issues du TAL (éviter les négations)

"l'écriture" → la génération

il faut revoir la fin, avec une description rapide du problème, de la méthode suivie et des contributions

121	— <code>qparselib/notes/cluster.md</code>
122	— <code>qparselib/src/segment/import/ :</code>
123	<code>DrumCode</code> <code>hpp</code> et <code>cpp</code>
124	
125	Au lieu du paragraphe final : Nous présenterons dans un premier temps
126	les parallèles entre TAL et MIR, puis les spécificités de la notation pour
127	la batterie.
128	Nous présenterons le contexte suivi d'un état de l'art et nous définirons
129	de manière générale le processus de transcription automatique de la mu-
130	sique pour enfin étayer les méthodes utilisées pour la transcription au-
131	tomatique de la batterie, et nous présenterons les principales contribu-
132	tions apportées à l'outil <code>qparse</code> . Nous décrirons ensuite le corpus ainsi
133	que les différentes expérimentations menées. Nous concluerons par une
134	discussion sur les résultats obtenus et les pistes d'améliorations futures
135	à explorer.
136	

ce paragraphe final (plan
est bien. tu peux ajouter les
numéros de sections avec
\ref

CONTEXTE

Sommaire

1.1	TAL et MIR	7
1.2	La transcription automatique de la musique	9
1.3	La transcription automatique de la batterie	11
1.4	Les représentations de la musique	12

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 musicale² est une étude du traitement de la musique [3], en particulier des représentations musicales, de la transformée de Fourier pour la musique³, de l'analyse de la structure de la musique et de la

cet article wikipedia est trop partiel, ne pas citer wikipedia, plutôt M. Müller FMP. je refererais plutôt à "Computer Music" : utilisation de méthodes numériques pour l'analyse et la synthèse d'information musicale, qu'il s'agisse de données audio ou symbolique (aide à l'écriture, transcription, base de partitions...) de musique

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

2. https://en.wikipedia.org/wiki/Music_informatics

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

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?

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

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^{4 5} apparaît vers le début des années 2000 [4]. 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.

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.

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 *et al.* [5], la musique est reconnue comme étant un système complexe spécifique à l'être humain dont une des similitudes avec les langues naturelles est l'émer-

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

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

gence 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.* [6] 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.* [7] 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)

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

exemples / illustration de la proximité thématique ?

Un domaine du TAL qui se rapproche du MIR est...

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

1.2 La transcription automatique de la musique

En musique, la transcription⁶ 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) [8] est de convertir la performance d'un musicien en notation musicale - un peu comme la conversion de la parole en texte dans le traitement du langage naturel. L'AMT a des intérêt multiples, notamment pour la transcription de solos ou encore pour la constitution de corpus musicologiques, ou encore pour l'interprétation de la musique et l'analyse du contenu musical

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

à l'instar de la

applications

préservation du patrimoine

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

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

citer un survey pour les applications (pas [9])

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

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

là on passe aux approches scientifiques

quel rapport?

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

éviter newpage

[9]. 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 [9]. 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 [10]. 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 1977⁷, 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.* [8] 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;
 - détection de l'apparition et du décalage;
 - estimation du tempo et du rythme.
- Apprentissage sur des modèles accoustiques 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).

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

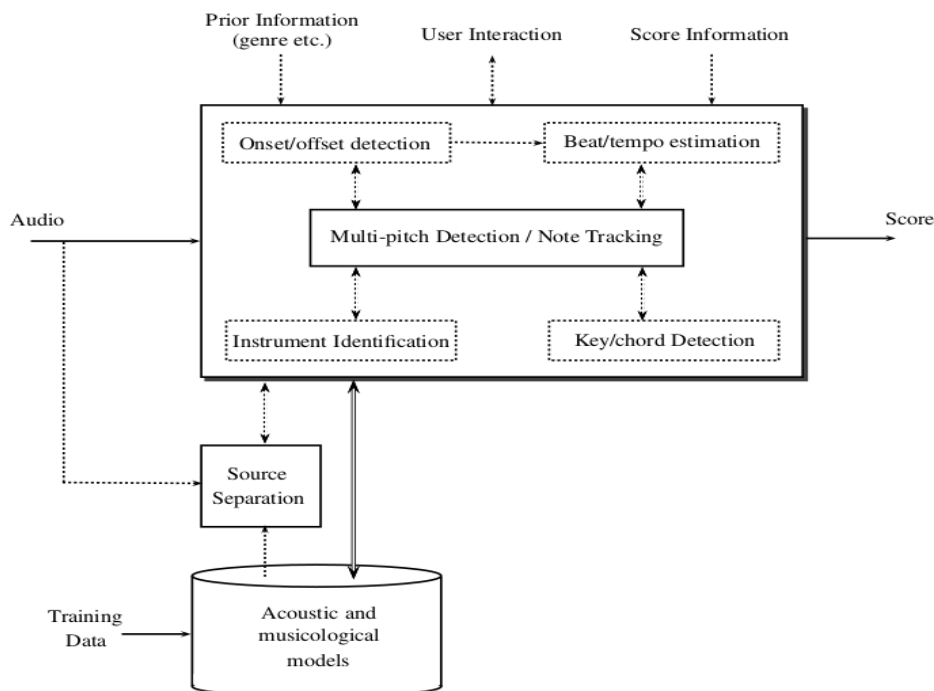


FIGURE 1.1 – Transcription automatique

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.

1.3 La transcription automatique de la batterie

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

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

tres bonne section

cite méthode et école Agostini?

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

citer [11] ici

294 Les applications de l'ADT seraient utiles dans tous les domaines musi-
 295 caux contenant de la batterie dont certains manque de partitions, notam-
 296 ment les musiques d'improvisation (jazz, pop) [8]. Mais aussi de manière
 297 plus générale dans le domaine du MIR. Si les ordinateurs étaient ca-
 298 pables d'analyser la partie de la batterie dans la musique enregistrée, cela
 299 permettrait une variété de tâches de traitement de la musique liées au
 300 rythme. En particulier, la détection et la classification des événements so-
 301 nores de la batterie par des méthodes informatiques est considérée comme
 302 un problème de recherche important et stimulant dans le domaine plus
 303 large de la recherche d'informations musicales [11].

ADT pas défini

"contenant" -> concernés
par

manque verbe

permettrait de faciliter

304 L'ADT est un sujet de recherche crucial pour la compréhension des aspects
 305 rythmiques de la musique, et a un impact potentiel sur des domaines plus
 306 larges tels que l'éducation musicale et la production musicale.

1.4 Les représentations de la musique

307

citer M. Müller FMP pour
cette section?

308

trop technique. ne pas re-
pier wikipédia

309

Les données audio

310 Le fichier WAV⁸ est une instance du Resource Interchange File Format
 311 (RIFF) défini par IBM et Microsoft. Le format RIFF agit comme une "en-
 312 veloppe" pour divers formats de codage audio. Bien qu'un fichier WAV
 313 puisse contenir de l'audio compressé, le format audio WAV le plus cou-
 314 rant est l'audio non compressé au format LPCM (linear pulse-code modu-
 315 lation). Le LPCM est également le format de codage audio standard des
 316 CD audio, qui stockent des données audio LPCM à deux canaux échan-
 317 tillonnées à 44 100 Hz avec 16 bits par échantillon. Comme le LPCM n'est
 318 pas compressé et conserve tous les échantillons d'une piste audio, les uti-
 319 lisateurs professionnels ou les experts en audio peuvent utiliser le format
 320 WAV avec l'audio LPCM pour obtenir une qualité audio maximale.

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

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

321

Les données MIDI

322 Le MIDI⁹ (Musical Instrument Digital Interface) est une norme tech-
 323 nique qui décrit un protocole de communication, une interface numérique
 324 et des connecteurs électriques permettant de connecter une grande va-
 325 riété d'instruments de musique électroniques, d'ordinateurs et d'appareils
 326 audio connexes pour jouer, éditer et enregistrer de la musique.

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

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

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

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

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

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

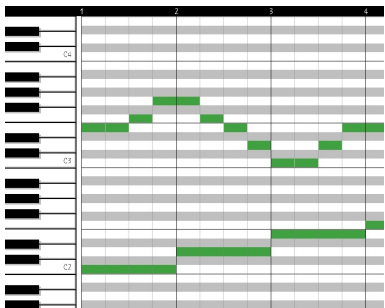


FIGURE 1.2 – Exemple évènements avec durée

331
332 Chaque évènement MIDI rassemble un ensemble d'informations sur la
hauteur, la durée, le volume, etc. . . :

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

FIGURE 1.3 – Critère pour un évènement

333
334 Pour la batterie, les évènements sont considérés sans durée, nous ignore-
335 rons donc les offsets (« Off Event »), les « Off Tick » et les « Duration ». Le
336 *channel* ne nous sera pas utile non plus.
337 Ici, définir *Tick* et *channel*.

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

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

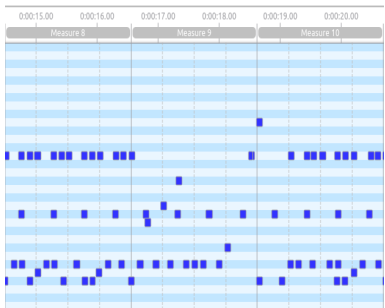


FIGURE 1.4 – Exemple évènements sans durée

338

339 On observe que toutes les durées sont identiques.

340 **Les partitions**

FIGURE 1.5 – Exemple de partition de piano

341 Une partition de musique¹⁰ est un document qui porte la représentation
 342 systématique du langage musical sous forme écrite. Cette représentation
 343 est appelée transcription et elle sert à traduire les quatre caractéristiques
 344 du son musical :

- 345 — la hauteur ;
 346 — la durée ;
 347 — l'intensité ;
 348 — le timbre.

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

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

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

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

356 **Le format MusicXML**

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

357 MusicXML est un format de fichier basé sur XML pour représenter la no-
 358 tation musicale occidentale. Ce format est ouvert, entièrement documenté
 359 et peut être utilisé librement dans le cadre de l'accord de spécification fi-
 360 nale de la communauté du W3C.

standard W3C = MNX (en
 cours)

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

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

inconvénient : format.s v
 beux et ambigus. -> on uti-
 lise pour la transcription
 une représentation inter-
 médiaire abstraite décrite
 plus loin.

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


```
<?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

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 résoudre (autant pour l'AMT que pour l'ADT).

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

ÉTAT DE L'ART

Sommaire

2.1	Monophonique et polyphonique	17
2.2	Audio vers MIDI	18
2.3	MIDI vers partition	19
2.4	Approche linéaire et approche hiérarchique	19

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 monophoniques¹ [8]. 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).

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 [12]. 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 [11], 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 RNN ou des NMF 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 [13] 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 avec un HMM. Les features sont transformés en représentations statistiques indépendantes. L'approche AdaMa [14] est une autre approche de la même catégorie ; elle commence par une estimation initiale des sons de la batterie qui sont itérativement raffinés pour correspondre à (pour matcher) l'enregistrement visé.

2.3 MIDI vers partition

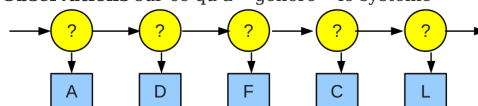
Le plus souvent, lorsque les articles abordent la transcription automatique de bout en bout (de l'audio à la partition), l'appellation « score » (*partition*) désigne un output au format Music XML, ou simplement MIDI. Par exemple, dans [9], la chaîne de traitement va jusqu'à la génération d'une séquence 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 [15]. 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 [15] 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.* [9] ont utilisé le modèle de Markov caché (HMM)² 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://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

476 *Source : Cours de Damien Nouvel*³

477

478 L'évaluation finale des résultats de [9] montre qu'il faut rediriger
 479 l'attention vers les valeurs des notes, la séparation des voix et d'autres
 480 éléments délicats de la partition musicale qui sont significatifs pour
 481 l'exécution de la musique. Or, même si la quantification du rythme se fait
 482 le plus souvent par la manipulation de données linéaires allant notam-
 483 ment des *real time units* (secondes) vers les *musical time units* (temps,
 484 métrique,...), de nombreux travaux suggèrent d'utiliser une approche
 485 hiérarchique puisque le langage musical est lui-même structuré.

486 En effet, l'usage d'arbres syntaxiques est idéale pour représenter le
 487 langage musical. Une méthodologie simple pour la description et l'af-
 488 fichage des structures musicales est présentée dans [16]. Les RT y
 489 sont évoqués comme permettant une cohésion complète de la notation
 490 musicale traditionnelle avec des notations plus complexes. Jacquemard
 491 *et al.* [17] propose aussi une représentation formelle du rythme, inspirée
 492 de modèles théoriques antérieurs et dont l'objectif est la réécriture de
 493 termes. Ils démontrent aussi l'application des arbres de rythmes pour
 494 les équivalences rythmiques dans [18]. La réécriture d'arbres, dans un
 495 contexte de composition assistée par ordinateur, par exemple, pourrait
 496 permettre de suggérer à un utilisateur diverses notations possibles pour
 497 une valeur rythmique, avec des complexités différentes.

498 La nécessité d'une approche hiérarchique pour la production automatique
 499 de partition est évoquée dans [15]. Les modèles de grammaire qui y sont
 500 exposés sont différents de modèles markoviens linéaires de précédents
 501 travaux.

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

Example: Summertime

FIGURE 2.2 – arbre_jazz

*Représentation arborescente d'une grille harmonique [10]***Conclusion**

La plupart des travaux déjà existants sur l'ADT ont été énumérés par Wu *et al.* [11] 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 [19], 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.

522

523

MÉTHODES

524

Sommaire

525	3.1	La notation de la batterie	23
526	3.2	Modélisation pour la transcription	30
527	3.3	Qparse	32
528	3.4	Les systèmes	33
529			
530			
532			

533

Introduction

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

536 Pour commencer, nous exposerons une description de la notation de la
 537 batterie ainsi qu'une modélisation de celle-ci pour la représentation des
 538 données rythmiques en arbres syntaxiques. Nous poursuivrons avec une
 539 présentation de qparse¹, un outil de transcription qui est développé par
 540 Florent Jacquemard (Inria) au sein du laboratoire Cedric au CNAM.

541 Enfin, nous présenterons les systèmes.

542

3.1 La notation de la batterie



543 Une figure de note [1] de musique combine plusieurs critères² :

544 — Une tête de note :

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

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

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

- 547 — Une hampe :
 548 Indicatrice d'appartenance à une voix en fonction de sa direction
 549 et indicatrice d'une durée représentée par sa présence ou non
 550 (blanche \neq ronde)
- 551 — Un crochet : La durée d'une note est divisée par deux à chaque
 552 crochet ajouté à la hampe d'une figure de note.



FIGURE 3.1 – Rapport des figures de notes
[1]

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

560 Les hauteurs et les têtes de notes

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

565 En effet, les hauteurs sur la portée représentent :

- 566 — La hauteur physique des instruments :
 567 La caisse claire est centrale sur la portée et sur la batterie (au
 568 niveau de la ceinture, elle conditionne l'écart entre les pédales et
 569 aussi la position de tous les instruments basiques d'une batterie).
 570 Tout ce qui en-dessous de la caisse-claire sur la portée est en
 571 dessous de la caisse-claire sur la batterie (pédales, tom basse);
 572 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est

573 aussi sur la batterie.

574

575 — La hauteur des instruments en terme de fréquences :

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

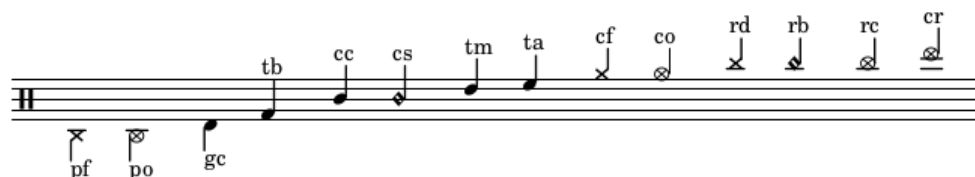


FIGURE 3.2 – Hauteur et têtes de notes

579 Les noms des instruments correspondant aux codes des notes de la figure
580 3.2 sont dans le tableau 3.1.

581 Les durées

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

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

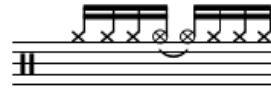
593 — Le point ;

594 — La liaison.

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



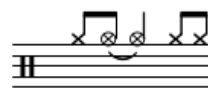
Exemple 1



Exemple 2



Exemple 3



Exemple 4

FIGURE 3.3 – Point et liaison

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

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

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

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

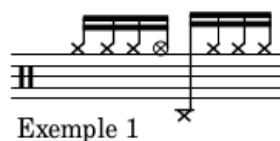
609 En cas de nécessité de rallonger la durée d'une note au-delà de son temps
 610 initial et si cette note correspond à une ouverture de charley, on privilé-
 611 giera la liaison.

612 Les silences

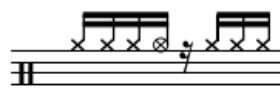
	la pause	la demi-pause	le soupir (2)	le demi-soupir	le quart de soupir	le huitième de soupir	le seizième de soupir
Silences							
Notes							
	la ronde	la blanche	la noire	la croche	la double croche	la triple croche	la quadruple croche

FIGURE 3.4 – Les silences

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



Exemple 1



Exemple 2

FIGURE 3.5 – Silence joué

620 L'exemple 1 de la figure 3.5 montre ce qui est écrit dans les données MIDI
 621 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-
 622 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1
 623 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-
 624 truments jouant simultanément.
 625 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte
 626 la note suivante pour l'écriture :
 627 - Si c'est un charley fermé joué à la main \Rightarrow la note sera cf;
 628 - Si c'est un charley fermé joué au pied \Rightarrow la note sera un silence.
 629

630 Les équivalences rythmiques

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



FIGURE 3.6 – Équivalence

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

646 Les voix

647 Les voix³ désignent les différentes parties mélodiques constituant une
 648 composition musicale et destinées à être interprétées, simultanément ou
 649 successivement, par un ou plusieurs musiciens. En batterie, une voix est
 650 l'ensemble des instruments qui, à eux seuls, constituent une phrase ryth-
 651 mique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant
 652 possibles pour un même rythme, on peut regrouper les instruments de la
 653 batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou
 654 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

655
 656 Ce choix se fera en fonction des instruments joués, de la nature plus ou
 657 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. » [1]

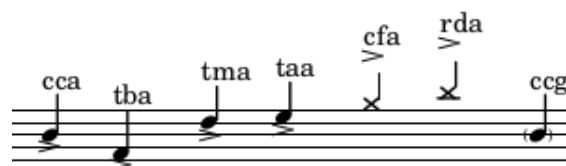


FIGURE 3.8 – Les accents et les ghost-notes

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

680 3.2 Modélisation pour la transcription

681 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

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

687 Malgré le large panel de pitches disponible, il semblerait qu'aucun pitch
 688 ne désigne le charley ouvert joué au pied. Pourtant, dans la batterie mo-
 689 derne, plusieurs rythmes ne peuvent fournir le son du charley ouvert
 690 qu'avec le pied car les mains ne sont pas disponibles pour le jouer. Cela
 691 doit en partie être dû à l'utilisation des boîte à rythmes en MAO qui ne né-
 692 cessitent pas de faire des choix conditionnés par les limitations humaines
 693 (2 pieds, 2 mains, et beaucoup plus d'instruments. . .)

694 La vélocité

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

697 Cette transcription nous a mené aux observations suivantes :

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

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

712 Les arbres de rythmes

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



716
 717 Ci-dessous, le même arbre dont les codes des instruments sont remplacés
 par leurs données MIDI respectives :



718
 719 Chacun des trois exemples de la figure 3.7 est représenté par un des deux
 720 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 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.

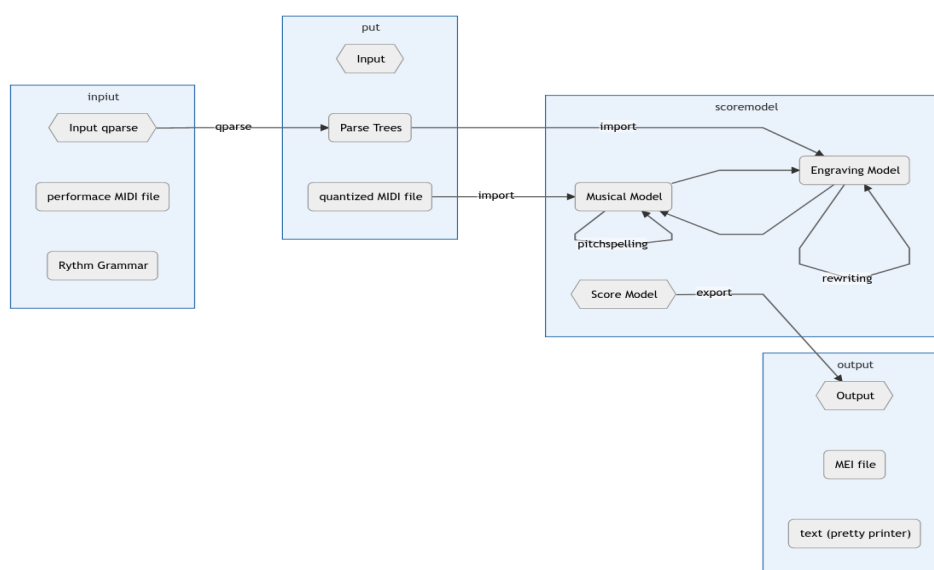


FIGURE 3.10 – Présentation de Qparse

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

Explication des différentes étapes de la figure 3.10⁵ :

- **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>

776 sur 1 ou 2 voix)

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

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

782

783 Un ensemble de systèmes comprenant leur métrique et leurs règles
784 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

785

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

787 — 4/4 binaire

788 — 4/4 jazz

789 — 4/4 afro-cubain

790 Objectif des systèmes

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

792 — définir une métrique ;

793 — choisir une grammaire appropriée ;

794 — fournir les règles de réécriture (séparation des voix et simplifica-
795 tion.

796

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

801 Détection d'indication de mesure

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



Exemple 1



Exemple 2

FIGURE 3.11 – Métrique

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

812 **Choix d'une grammaire**

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

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

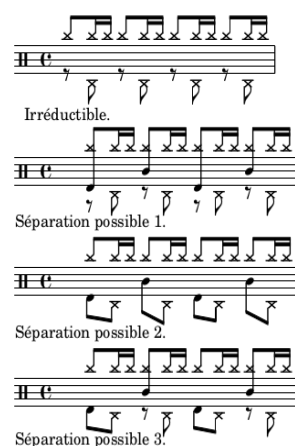
825 **Séparation des voix**

FIGURE 3.12 – Motif 4-4 binaire

826 Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les
 827 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système
 828 est constitué par défaut du motif rd/pf/cc (voir 3.1) et d'un texte joué à la
 829 grosse-caisse. La première ligne de la figure 3.12 est appelée « Irréductible
 830 » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du
 831 charley au pied. La troisième séparation proposée est privilégiée car elle
 832 répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les
 833 pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont
 834 utilisés par voix et plus logique pour le lecteur puisque les mains sont en
 haut et les pieds en bas.

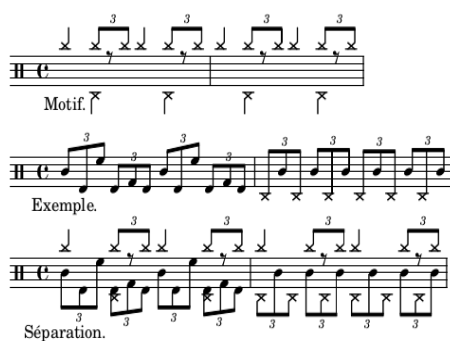


FIGURE 3.13 – Motif 4-4 jazz

835 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 836 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 837 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 838 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 839

840 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 841 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 842 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 843 et de caisse claire et son écriture séparée serait inutilement compliquée
 844 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 845 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

846
 847 La figure 3.14 montre un exemple minimaliste de système afro-latin [22].
 848 Ce système doit être écrit sur trois voix car la voix centrale est souvent
 849 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas
 850 serait surchargé et peu lisible.

851 **Simplification de l'écriture**

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

854
 855 Les gammes qui accompagnent les motifs d'un système étayent toutes les
 856 combinaisons d'un système et elles permettent, combinées avec le motif
 857 d'un système, de définir les règles de simplification propres à celui-ci.

858
 859 Voici les différentes étapes à suivre :

- 860 — Pour chaque gamme du système, faire un arbre de rythme repré-
 861 sentant la gamme combinée avec le motif du système ;
- 862 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
 863 arbre de rythme par voix ;
- 864 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
 865 nœuds qui nécessitent une simplification et écrire la règle.

866
 867 Certaines précisions concernant l'extraction de ces règles sont néces-
 868 saires. Il s'agit de précisions à propos de la durée, des silences et de la
 869 présence ou non d'ouverture de charley dans les instruments joués. Nous
 870 avons discuté de ces problèmes dans le chapitre 3.

871
 872 Voici quelques règles inhérentes à la simplification de l'écriture pour
 873 la batterie :

874 Toutes les continuations (t) qui se trouvent en début de temps (figures 4.9,
 875 4.11 et 4.12) sont transformées en silences (r) sauf si la note précédente
 876 est un charley ouvert ?
 877 Même si on favorise l'usage des silences pour l'écart entre les notes
 878 n'appartenant pas au même temps, on les supprime systématiquement
 879 pour 2 notes au sein d'un même temps et favorise, une liaison si co, un
 880 point si pas co et nécessaire, un simple ajustement de la figure de note si
 suffisant.

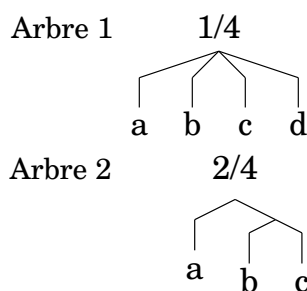


FIGURE 3.15 – Simplification

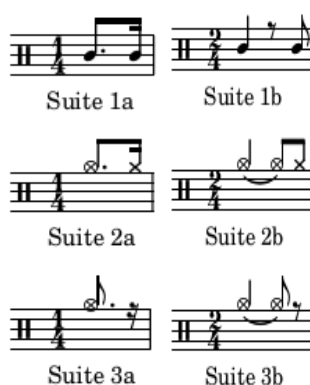


FIGURE 3.16

881
 882 Soit l'arbre 1 de la figure 3.15 dans lequel :
 883 a et d sont des instruments de la batterie (x);
 884 b et c sont des continuations (t); Pour chacune des conditions suivantes,
 885 une suite de la figure 3.16 est attribuée :
 886 — Si a n'est pas un co :
 887 ⇒ Suite 1a.
 888 — Si a est un co :
 889 — Si d est un cf :
 890 ⇒ Suite 2a.

891 — Si d est un pf :
892 ⇒ Suite 3a : d deviens un silence (r).
893
894 Soit l'arbre 2 de la figure 3.15 dans lequel :
895 a et c sont des instruments de la batterie (x);
896 b est une continuation (t); Pour chacune des conditions suivantes, une
897 suite de la figure 3.16 est attribuée :
898 — Si a n'est pas un co :
899 ⇒ Suite 1b, b devient un silence.
900 — Si a est un co :
901 — Si c est un cf :
902 ⇒ Suite 2b, b devient une liaison et c devient un cf.
903 — Si c est un pf :
904 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
905
906 *Rappel :*
907 *cf = charley fermé joué à la main ;*
908 *co = charley ouvert joué à la main ;*
909 *pf = charley fermé joué au pied.*
910
911 **Problème : le cf et le co ne seront jamais sur la même voix que le**
912 **pf... Par conséquent, les règles concernant les charleys ouverts**
913 **doivent-elles être appliquées sur l'arbre de parsing de l'input?...**

914 Conclusion

915 Nous avons formalisé une notation de la batterie, modélisé cette notation
916 pour la transcription de données MIDI en partition, nous avons décrit
917 Qparse.
918 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-
919 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-
920 propriée et énoncer des règles de séparation des voix et de simplification
921 de l'écriture.

922

923

EXPÉRIMENTATIONS

924

Sommaire

925

926

927

928

929

930

932

4.1	Le jeu de données	41
4.2	Analyse MIDI-Audio	43
4.3	Expérimentation théorique d'un système	47
4.4	Résultats et discussion	52

933

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.

941

4.1 Le jeu de données

Nous avons utilisé le Groove MIDI Dataset¹ [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

950 Autres critères spécifiques au GMD :

- 951 — Toutes les performances ont été jouées au métronome et à un tempo
- 952 choisi par le batteur.
- 953 — 80% de la durée du GMD a été joué par des batteurs professionnels
- 954 qui ont pu improviser dans un large éventail de styles. Les don-
- 955 nées sont donc diversifiées en termes de styles et de qualités de jeu
- 956 (professionnel ou amateur).
- 957 — Les batteurs avaient pour instruction de jouer des séquences de
- 958 plusieurs minutes ainsi que des fills²
- 959 — Chaque performance est annotée d'un style (fourni par le batteur),
- 960 d'une métrique et d'un tempo ainsi que d'une identification ano-
- 961 nyme du batteur.
- 962 — Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10
- 963 rythmes dans leurs styles respectifs. Ils sont dans les dossiers eval-
- 964 session du GMD.
- 965 — Les sorties audio synthétisées ont été alignées à 2 ms près sur leur
- 966 fichier MIDI.

967 **Format des données**

968 Le Roland TD-11 divise les données enregistrées en plusieurs pistes dis-
969 tinctes :

- 970 — une pour le tempo et l'indication de mesure ;
- 971 — une pour les changements de contrôle (position de la pédale de
- 972 charley) ;
- 973 — une pour les notes.

974

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

977 Pour simplifier le traitement de ces données, ces trois pistes ont été
978 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.

979
 980 « Control Changes The TD-11 also records control changes speci-
 981 fying the position of the hi-hat pedal on each hit. We have preserved this
 982 information under control 4. »
 983 (<https://magenta.tensorflow.org/datasets/groove>)
 984 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 985 tions dans les fichiers MIDI.
 986 L'utilisation de pretty_midi devient urgente !

987 4.2 Analyse MIDI-Audio

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

990 Comparaisons de transcriptions

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

995 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 996 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 997 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 998 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 999 premières notes (une double-croche et une croche) d'un triolet sur
- 1000 le temps 1 de la mesure 2 de la TA.

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

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

1001 **Exemple d'analyse 2**

Transcription manuelle \Rightarrow Transcription automatique



1002 — Les doubles croches ont été interprétées en quintolet

1003 — La deuxième double-croche est devenue une croche.

1004

1005 **Exemple d'analyse 3**

Transcription manuelle \Rightarrow Transcription automatique



1006 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.

1008 — 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.

1010 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.

1012 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.

1013

1015 **Exemple d'analyse 4**

Transcription manuelle \Rightarrow Transcription automatique



1016

1017 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une manière excessivement complexe!

1018

1019 **Exemple avec des flas**

1020 Transcription manuelle



1021

1022 Transcription automatique

1023



1024

1025

- 1026 — Le premier fla est reconnu comme étant un triolet contenant une
- 1027 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1028 ornementée.
- 1029 — Le deuxième fla est reconnu comme étant un accord.
- 1030 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1031 dans la TA.
- 1032 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1033 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1034 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1035 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1036 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1037 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1038 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1039 temps 3 de la TA.
- 1040 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1041 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1042 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1043 cross-sticks.

1044 **Transcription de partition**

Music engraving by LilyPond 2.22.1—www.lilypond.org

FIGURE 4.2 – Partition de référence

1045 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1046 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1047 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1048 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1049 *Inria*) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1050 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1051 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1052 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1053 et quatre).

1054 La transcription des données audio et MIDI contenues dans ces fichiers
 1055 a permis une analyse plus approfondie des critères à relever pour chaque
 1056 évènement MIDI et de la manière de les considérer dans un objectif de
 1057 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

1081 dont la première ne contient que des croches et la deuxième que des
 1082 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1083 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1084 Systèmes — motifs et gammes combinés

1085 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1086 de la figure 4.3.

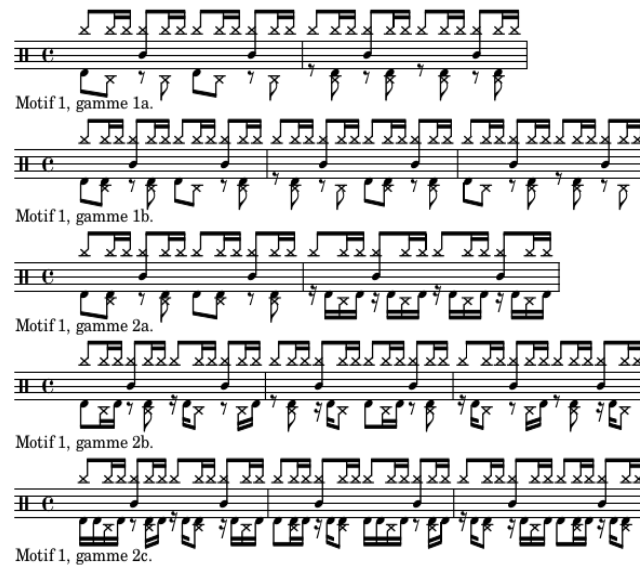


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1087

1088 Représentation du système en arbres de rythmes

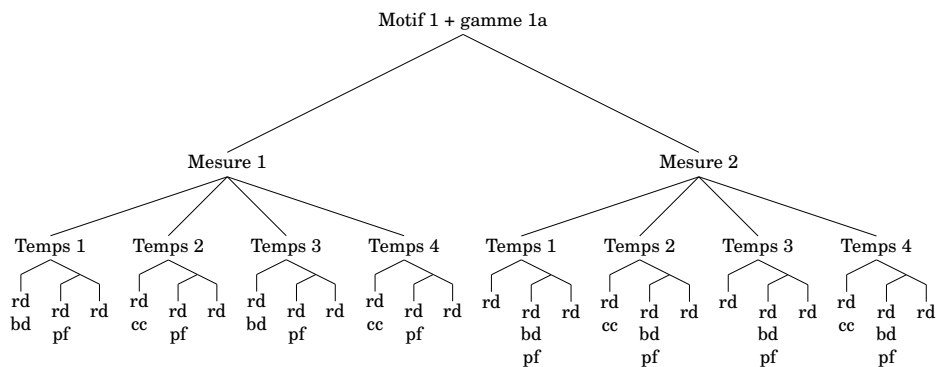


FIGURE 4.5 – Arbre de rythme — système

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.

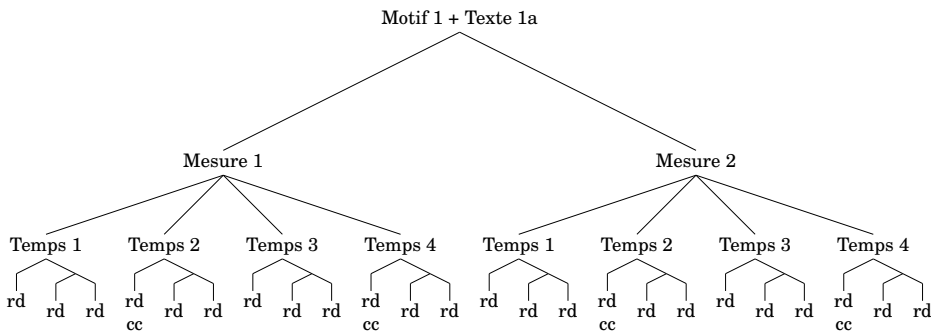


FIGURE 4.6 – Arbres de rythme — voix haute

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

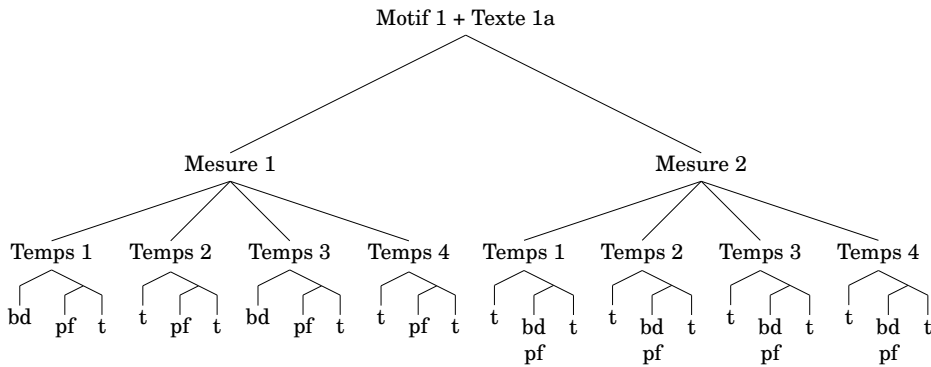


FIGURE 4.7 – Arbres de rythme — voix basse

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

1104 Les règles de simplifications

1105 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1106 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1107 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1108 entre elles (rendre la pulse visuelle).

1109

1110 Pour les figures ci-dessous :

1111 — x = une note ;

1112 — r = un silence ;

1113 — t = une continuation (point ou liaison)

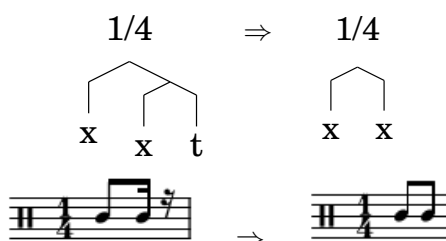


FIGURE 4.8

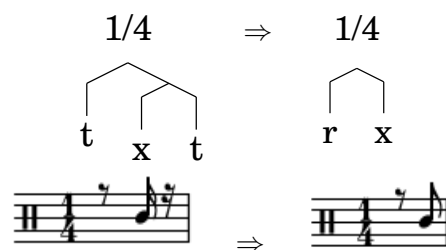


FIGURE 4.9

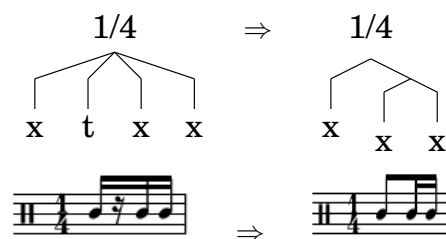


FIGURE 4.10

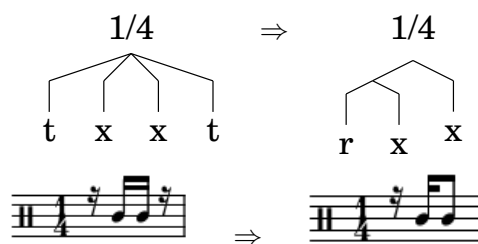


FIGURE 4.11

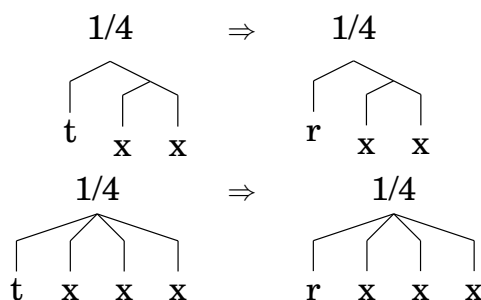


FIGURE 4.12

1114 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
 1115 manquants seront mis en annexe.

1116

1117 Les règles remplacent par un silence les continuations (t) qui sont
 1118 au début d'un temps. Cela est valable pour ce système mais lorsqu'il y a
 1119 des ouvertures de charley, cela n'est pas toujours applicable. Ce problème
 1120 est évoqué de le chapitre 3.

1121

1122 \Rightarrow **Objectif de cette expérimentation théorique :**

1123 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
 1124 expérimentation théorique a pour but d'orienter la recherche d'autres
 1125 systèmes par observation du jeu de données et de montrer comment les
 1126 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

1127 4.4 Résultats et discussion

1128 Cette section regroupe les avancées qui ont été réalisées par rapport aux
 1129 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
 1130 tats de l'ADT avec Qparse. Nous avons amélioré le système de quantifi-
 1131 cation de Qparse pour la batterie, notamment le passage à la polyphonie
 1132 avec les Jams. Nous avons pu obtenir des arbres de parsing correctes en
 1133 améliorant les grammaires avec des fichiers MIDI courts. Puis, une sortie
 1134 MEI a été aussi obtenue (encore à vérifier).

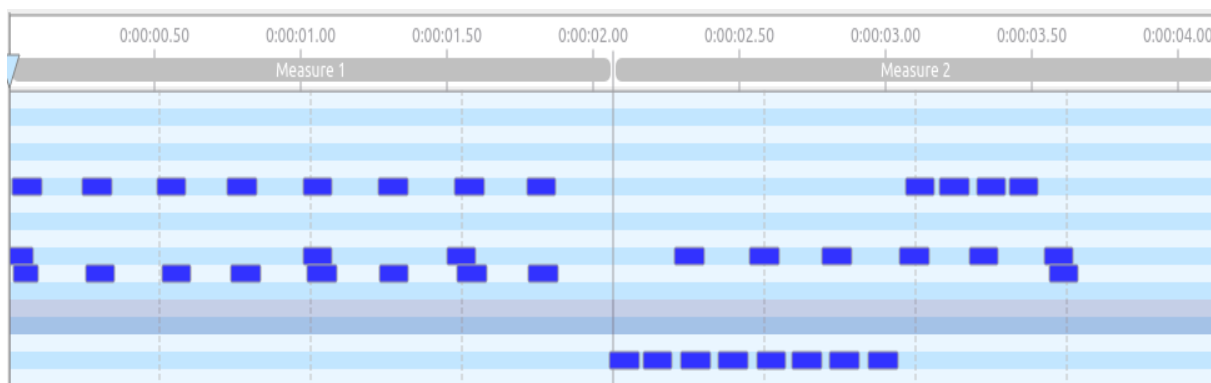
1135 Les Jams

1136 Les Jams permettent de passer du monophonique au polyphonique.

1137 Le parsing

1138 Tests effectués avec le fichier midi suivant :

1139



1140

1141

1142 Un premier test convaincant est effectué avec la grammaire sui-
 1143 vante :

1144

1145 // bar level

1146 0 -> C0 1

1147 0 -> E1 1

1148 0 -> U4(1, 1, 1, 1) 1

1149

1150 // half bar level

1151 9 -> C0 1

1152 9 -> E1 1

1153

1154 // beat level

1155 1 -> C0 1

1156 1 -> E1 1

1157 1 -> T2(2, 2) 1
 1158 1 -> T4(4, 4, 4, 4) 1
 1159

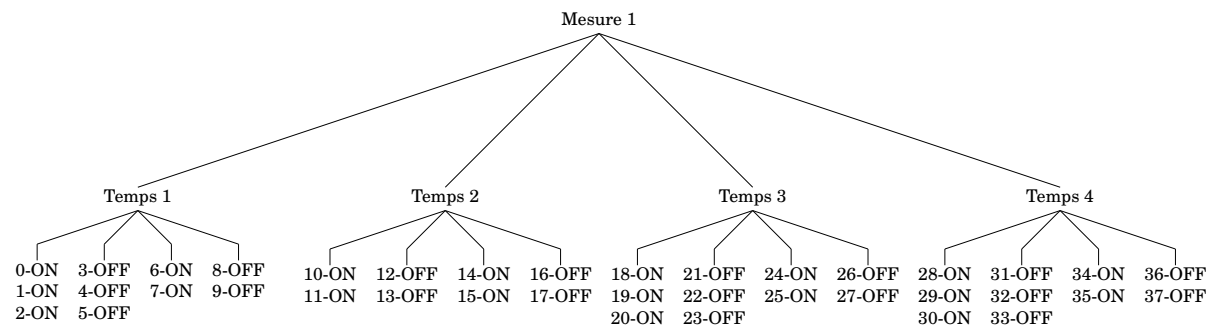
1160 // croche level
 1161 2 -> C0 1
 1162 2 -> E1 1
 1163

1164 // double level
 1165 4 -> C0 1
 1166 4 -> E1 1
 1167 4 -> E2 1
 1168 4 -> T2(6, 6) 1
 1169

1170 // triple level
 1171 6 -> E1 1
 1172

1173 Cette grammaire sépare les ligatures par temps au niveau de la
 1174 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1175 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1176 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1177 il découpe correctement les mesures et les temps.

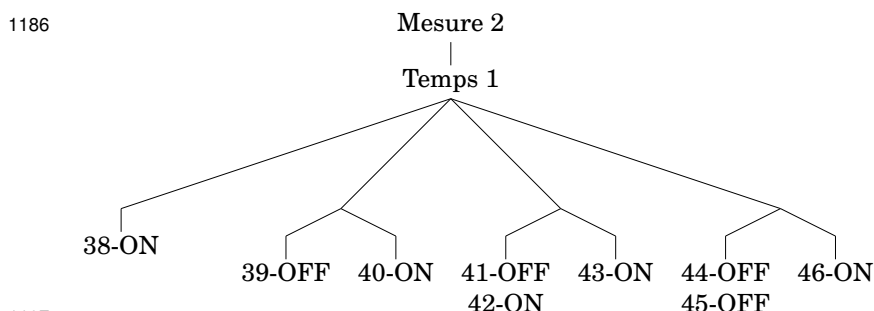
1178
 1179



1180
 1181

1182 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1183 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1184 quantification visibles dès le premier temps.

1185



1189 Les Onsets sont correctement triés au niveau des doubles croches
 1190 mais certaines doubles croches sont inutilement subdivisées en triples
 1191 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1192 ci-dessus).

1193 2ème exemple :

1194 Après une augmentation du poids des triples croches dans la grammaire
 1195 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1196 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1197 sont bien triés et aucuns ne sont subdivisés.
 1198

1199 Évaluation

1200 Pour l'évaluation, il aurait fallu produire un module.
 1201 L'évaluation est-elle automatique ou manuelle?
 1202 Possibilité d'un export lilypond en arbre pour comparer l'output avec la
 1203 transcription manuelle.
 1204 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel
 1205 ou MEI pour les comparer et faire des statistiques. Si transformés en
 1206 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)
 1207 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.
 1208 L'expérimentation peut-être considérer comme une évaluation manuelle?
 1209 (magicien d'Oz)
 1210 Lilypond vers MIDI + output vers MIDI ⇒ Comparaison des MIDI
 1211 dumpés.
 1212

1213 Discussion

1214 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
 1215 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
 1216 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.
 1217 Écrire des règles de réécriture spécifique aux charley avec un système
 1218 approprié. Le jeu de système

- 1219 — implémenter un pattern. . .
- 1220 ⇒ manque de temps ?
- 1221
- 1222 — La partie résultat est manquante car :
- 1223 ⇒ Sujet très difficile ;
- 1224 ⇒ Matcher les motifs peut être fait ultérieurement ;
- 1225 Mais ce travail aurait été indispensable pour obtenir une quan-
- 1226 tité de résultats qui justifieraient une évaluation automatique
- 1227 permettant de faire des graphiques.
- 1228
- 1229 — L'évaluation fut entièrement manuelle car :
- 1230 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
- 1231 VS output)
- 1232 — Le ternaïre jazz (voir expérience 2)
- 1233 — Reconnaissance d'un motif sur le MIDI
- 1234 Reconnaître un motif (système) sur une mesure de l'input (un fi-
- 1235 chier midi représentant des données audios)
- 1236 ⇒ Motif (système) reconnu : true ou false
- 1237 Si true :
- 1238 - Choisir la grammaire correspondante ;
- 1239 - Parser le MIDI ;
- 1240 - Appliquer les règles de réécritures (Séparation des voix et simpli-
- 1241 fication)
- 1242 — Nous travaillerons aussi sur la détection de répétitions sur plu-
- 1243 sieurs mesures afin de pouvoir corriger des erreurs sur une des
- 1244 mesures qui aurait dû être identique aux autres mais qui présente
- 1245 des différences.
- 1246 — dans quelle catégorie mettre le shuffle ?
- 1247 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
- 1248 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
- 1249 à la batterie a été générée. Elle sera un bon point de départ pour les
- 1250 travaux futurs dont plusieurs propositions sont énoncés dans le présent
- 1251 document.

CONCLUSION GÉNÉRALE

1253 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1254 tion automatique de la batterie. Son objectif était de transcrire, à partir
1255 de leur représentation symbolique MIDI, des performances de batteur de
1256 différents niveaux et dans différents styles en partitions écrites.
1257 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1258 cessus de regroupement des événements MIDI qui nous a permis de faire
1259 la transition du monophonique vers le polyphonique. Une des données
1260 importante de ce processus était de différencier les nature des notes d'un
1261 accord, notamment de distinguer lorsque 2 notes constituent un accord
1262 ou un *fla*.
1263 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1264 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1265 tionnables en amont du parsing, soit par indication des noms des fichiers
1266 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1267 naire de patterns prédéfinis⁵ qu'il serait pertinent de mettre en œuvre en
1268 machine learning.
1269 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1270 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1271 d'un système et pour la simplification grâce aux gammes du motif d'un
1272 système. Nous avons aussi montré comment, dans des travaux futurs, un
1273 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1274 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1275 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1276 la chaîne de traitement en terme de réécriture.
1277 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1278 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1279 entièrement automatisée de partitions de batterie et au problème général
1280 de l'AMT de bout en bout.[8]

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

BIBLIOGRAPHIE

- 1282 [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41
1283 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition,
1284 1996. – Cité pages 5, 23, 24 et 29.
- 1285 [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 7
1286 et 10.
- 1287 [3] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité
1288 page 7.
- 1289 [4] Caroline Traube. Quelle place pour la science au sein de la musico-
1290 logie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 8.
- 1291 [5] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions
1292 entre les traitements de la musique et du langage. *La Lettre des*
1293 *Neurosciences*, 58 :24–26, 2018. – Cité page 8.
- 1294 [6] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo.
1295 Techniques de traitement automatique du langage naturel appli-
1296 quées aux représentations symboliques musicales. In *JIM 2021 -*
1297 *Journées d’Informatique Musicale*, Virtual, France, July 2021. –
1298 Cité page 9.
- 1299 [7] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering
1300 music relations with sequential attention. In *NLP4MUSA*, 2020. –
1301 Cité page 9.
- 1302 [8] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger
1303 Kirchhoff, and Anssi Klapuri. Automatic music transcription : Chal-
1304 lenges and future directions. *Journal of Intelligent Information Sys-*
1305 *tems*, 41, 12 2013. – Cité pages 9, 10, 12, 17 et 57.
- 1306 [9] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1307 musical statistics as guides for audio-to-score piano transcription.
1308 *Information Sciences*, 566 :262–280, 2021. – Cité pages 10, 19 et 20.
- 1309 [10] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1310 O’Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1311 Cité pages 10 et 21.
- 1312 [11] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1313 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander

- 1314 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1315 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1316 1483, 2018. – Cité pages 12, 18 et 21.
- 1317 [12] Moshekwa Malatji. Automatic music transcription for two instru-
1318 ments based variable q-transform and deep learning methods, 10
1319 2020. – Cité page 18.
- 1320 [13] Antti J. Eronen. Musical instrument recognition using ica-based
1321 transform of features and discriminatively trained hmms. *Seventh*
1322 *International Symposium on Signal Processing and Its Applications*,
1323 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 18.
- 1324 [14] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1325 drum sound description for real-world music using template adap-
1326 tation and matching methods. *International Conference on Music*
1327 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 18.
- 1328 [15] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1329 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1330 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1331 *Computation in Music*, volume Lecture Notes in Computer Science
1332 of *Proceedings of the Seventh International Conference on Mathema-*
1333 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1334 2019. Springer. – Cité pages 19 et 20.
- 1335 [16] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1336 ring of rhythm structures. In *Proceedings of the First International*
1337 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1338 IEEE Computer Society. – Cité page 20.
- 1339 [17] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1340 Term Rewriting Based Structural Theory of Rhythm Notation. Re-
1341 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1342 Cité page 20.
- 1343 [18] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1344 equivalent rhythmic notations based on rhythm tree languages. In
1345 *Third International Conference on Technologies for Music Notation*
1346 *and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lo-
1347 pez Palma and Mike Solomon. – Cité page 20.
- 1348 [19] R. Marxer and J. Janer. Study of regularizations and constraints in
1349 nmf-based drums monaural separation. In *International Conference*
1350 *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1351 02/09/2013 2013. – Cité page 21.
- 1352 [20] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1353 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 24.

- 1354 [21] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1355 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 24.
- 1356 [22] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1357 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1358 94120 Fontenay-sous-Bois, 2000. – Cité pages 25 et 37.
- 1359 [23] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1360 Bamman. Learning to groove with inverse sequence transforma-
1361 tions. In *International Conference on Machine Learning (ICML)*,
1362 2019. – Cité page 41.

