

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

4 Département Textes, Informatique, Multilinguisme

5 **Titre du mémoire**

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

8 *Parcours :*
9 *Ingénierie Multilingue*

10 par

11 **Martin DIGARD**

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

14 *Encadrant :*
15 *Florent JACQUEMARD*

16 Année universitaire 2020-2021

TABLE DES MATIÈRES

18	Liste des figures	4
19	Liste des tableaux	5
20	Introduction générale	7
21	1 Contexte	9
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	13
25	1.4 Les représentations de la musique	14
26	2 État de l'art	19
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	3 Méthodes	25
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	4 Expérimentations	43
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	Conclusion générale	59
42	Bibliographie	61

LISTE DES FIGURES

44	1.1	Transcription automatique	13
45	1.2	Exemple évènements avec durée	15
46	1.3	Critère pour un évènement	15
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	25
53	3.2	Rapport des figures de notes	26
54	3.3	Hauteur et têtes de notes	27
55	3.4	Point et liaison	28
56	3.5	Les silences	28
57	3.6	Silence joué	29
58	3.7	Équivalence	30
59	3.8	Séparation des voix	30
60	3.9	Les accents et les ghost-notes	31
61	3.10	Exemple pour les accentuations et les ghost-notes	31
62	3.11	Présentation de Qparse	34
63	3.12	Métrique	37
64	3.13	Motif 4-4 binaire	38
65	3.14	Motif 4-4 jazz	38
66	3.15	Système 4-4 afro-latin	39
67	3.16	Simplification	40
68	3.17	40
69	4.1	Batterie électronique	44
70	4.2	Partition de référence	48
71	4.3	Motifs et gammes	49
72	4.4	Partition d'un système en 4/4 binaire	50
73	4.5	Arbre de rythme — système	50
74	4.6	Arbre de rythme — voix haute	51
75	4.7	Arbre de rythme — voix basse	51
76	4.8	52
77	4.9	52
78	4.10	52
79	4.11	53

80	4.12	53
----	----------------	----

81

LISTE DES TABLEAUX

82	1.1 speechToText vs AMT	11
83	3.1 Pitches et instruments	32
84	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 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

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

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

CONTEXTE

Sommaire

1.1	TAL et MIR	9
1.2	La transcription automatique de la musique	11
1.3	La transcription automatique de la batterie	13
1.4	Les représentations de la musique	14

Introduction

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

cf. ismir.net

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

du problème de l'AMT, de ses applications

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

1.1 TAL et MIR

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

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

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

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

On peut citer : ...

? psycho-acoustique, neurosciences ?

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

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

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

Is Music a Language?



Leonard Bernstein

Norton Lectures at Harvard, 1973

« The Unanswered Question: Six Talks at Harvard »

idea of music as a kind of universal language

notion of a worldwide, « inborn musical grammar »

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

theory of innate grammatical competence

ne pas inclure ce slide, citer Bernstein et Chomsky est suffisant

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

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

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

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

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

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

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

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

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

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

exemples / illustration de la proximité thématique?

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

TABLE 1.1 – speechToText vs AMT

1.2 La transcription automatique de la musique

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

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

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

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

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

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

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

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

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

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

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

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

- 262 Quatres sous-tâches optionnelles accompagnent ces algo-
 263 rithmes :
 264 — identification de l'instrument ;
 265 — estimation de la tonalité et de l'accord ;
 266 — détection de l'apparition et du décalage ;
 267 — estimation du tempo et du rythme.
- 268 — Apprentissage sur des modèles acoustiques et musicologiques.
 269 — *Optionnel* : Informations fournies de manière externe, soit fournie
 270 en amont (genre, instruments,...), soit par interaction avec un uti-
 271 lisateur (infos sur une partition incomplète).

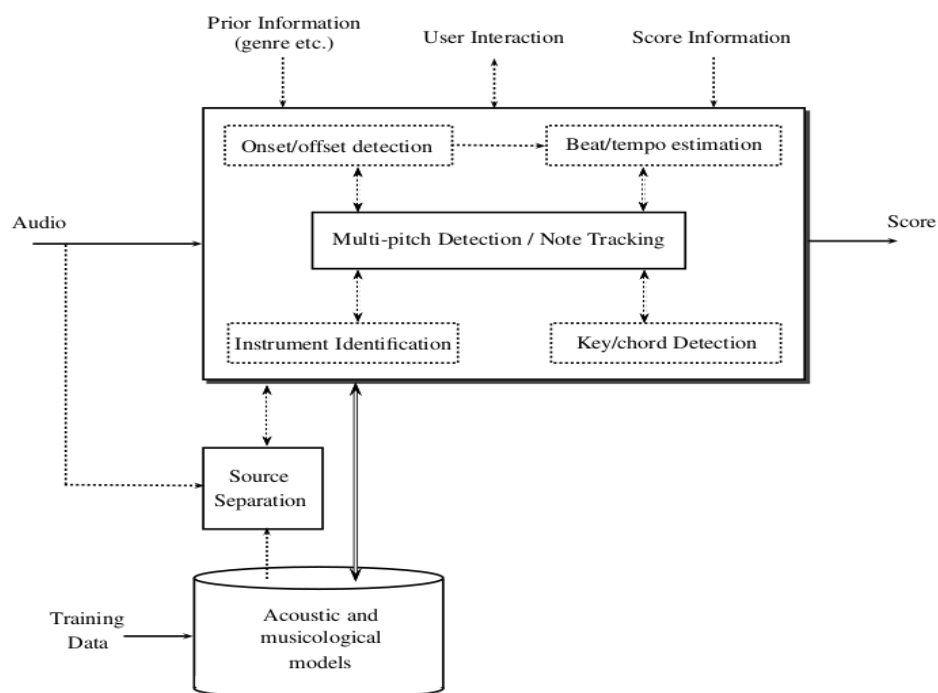


FIGURE 1.1 – Transcription automatique

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

272

éviter newpage

273 1.3 La transcription automatique de la batterie

274

tres bonne section

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

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

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

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).

294 Les applications de l'ADT seraient utiles, non seulement dans tous les
295 domaines musicaux contenant de la batterie dont certains manquent de
296 partitions, notamment les musiques d'improvisation (jazz, pop) [10], mais
297 aussi de manière plus générale dans le domaine du MIR : si les ordina-
298 teurs étaient capables d'analyser la partie de la batterie dans la musique
299 enregistrée, cela permettrait une variété de tâches de traitement de la
300 musique liées au rythme. En particulier, la détection et la classification
301 des événements sonores de la batterie par des méthodes informatiques
302 est considérée comme un problème de recherche important et stimulant
303 dans le domaine plus large de la recherche d'informations musicales [13].
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.

307 1.4 Les représentations de la musique

309 Les données audio

310 Le fichier WAV¹¹ 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

9. Roland TR-808

10. SQ-1

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

cite méthode et école Ago-
tini?

citer [13] ici

ADT pas défini

"contenant" -> concernés
par

permettrait de faciliter

citer M. Müller FMP pou-
cette section?

trop technique. ne pas re-
pier wikipédia

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

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

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

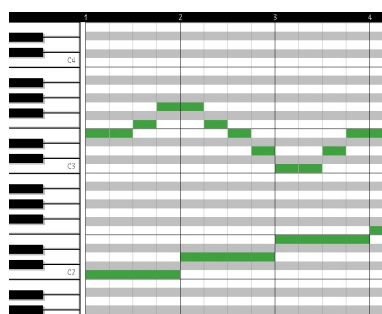
322 Les données MIDI

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

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

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

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



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

FIGURE 1.2 – Exemple événements avec durée

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

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

FIGURE 1.3 – Critère pour un événement

333

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

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

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 :

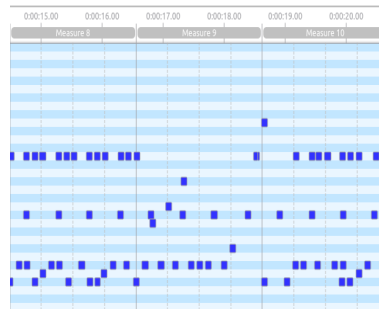


FIGURE 1.4 – Exemple évènements sans durée

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 musique¹³ 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

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

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

Le format MusicXML

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

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

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

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

standard W3C = MNX (en cours)

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



FIGURE 1.6 – MusicXML

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

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

Conclusion

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

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

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

374 soudre (autant pour l'AMT que pour l'ADT).
375 Et enfin, nous avons décrit les représentations de la musique nécessaires
376 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.

présenterons quelques travaux antérieurs

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

2.1 Monophonique et polyphonique

Les premiers travaux en transcription ont été faits sur l'identification des instruments monophoniques¹ [10]. Actuellement, le problème de l'estimation automatique de la hauteur des signaux monophoniques peut être considéré comme résolu, mais dans la plupart des contextes musicaux, les instruments sont polyphoniques². L'estimation des hauteurs multiples

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

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

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

2.2 Audio vers MIDI

Jusqu'à aujourd'hui, les recherches se sont majoritairement concentrées sur le traitement du signal vers la génération du MIDI [14].

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

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

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

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

de signaux audio

MIDI non-quantifié = performance (à expliquer)

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

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

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

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

haute fréquence, aigus ?

classification des événements ? la phrase semble redondante

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

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

448 2.3 MIDI vers partition

449
 450 Le plus souvent, lorsque les articles abordent la transcription automa-
 451 tique de bout en bout (de l'audio à la partition), l'appellation « *score* »
 452 (partition) désigne un ouput au format Music XML, ou simplement MIDI.
 453 Par exemple, dans [11], la chaîne de traitement va jusqu'à la génération
 454 d'une séquence MIDI quantifiée qui est importée dans MuseScore pour en
 455 extraire manuellement un fichier MusicXML contenant plusieurs voix.
 456 Seuls quelques travaux récents s'intéressent de près à la création d'outils
 457 permettant la génération de partition. Le problème de la conversion d'une
 458 séquence d'évènements musicaux symboliques en une partition musicale
 459 structurée est traité notamment dans [17]. Ce travail, qui vise à résoudre
 460 en une fois la quantification rythmique et la production de partition struc-
 461 turée, s'appuie tout au long du processus sur des grammaires génératives
 462 qui fournissent un modèle hiérarchique *a priori* des partitions. Les expé-
 463 riences ont des résultats prometteurs, mais il faut relever qu'elle ont été
 464 menées avec un ensemble de données composé d'extraits monophoniques ;
 465 il reste donc à traiter le passage au polyphonique, en couplant le problème
 466 de la séparation des voix avec la quantification du rythme.
 467 L'approche de [17] est fondée sur la conviction que la complexité de la
 468 structure musicale dépasse les modèles linéaires.

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

de manière conjointe

langage *a priori*

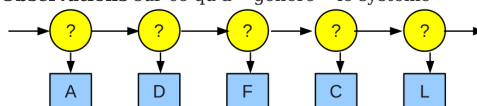
qui nécessite de traiter le
problème supplémentaire
de la séparation de voix. i.e.
pour la batterie on n'vut
quantification + structu-
ration + séparation mais
seules les 2 premières sont
couplées dans l'approche de
tonn stage.

469 2.4 Approche linéaire et approche hiérarchique

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

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

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



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

FIGURE 2.1 – HMM

479 *Source : Cours de Damien Nouvel*⁵

480

481

482 L'évaluation finale des résultats de [11] montre qu'il faut rediriger l'atten-
 483 tion vers les valeurs des notes, la séparation des voix et d'autres éléments
 484 délicats de la partition musicale qui sont significatifs pour l'exécution de
 485 la musique. Or, même si la quantification du rythme se fait le plus souvent
 486 par la manipulation de données linéaires allant notamment des *real time*
 487 *units* (secondes) vers les musical *time units* (temps, métrique, ...), de nom-
 488 breux travaux suggèrent d'utiliser une approche hiérarchique puisque le
 langage musical est lui-même structuré. En effet, l'usage d'arbres syn-
 taxiques est idéale pour représenter le langage musical. Une méthodolo-
 gie simple pour la description et l'affichage des structures musicales est
 présentée dans [18]. Les RT y sont évoqués comme permettant une co-
 hésion complète de la notation musicale traditionnelle avec des notations
 plus complexes. Jacquemard *et al.* [19] propose aussi une représentation
 formelle du rythme, inspirée de modèles théoriques antérieurs issus du
 domaine de la réécriture de termes. Ils démontrent aussi l'application des
 arbres de rythmes pour les équivalences rythmiques dans [20]. La réécri-
 ture d'arbres, dans un contexte de composition assistée par ordinateur,
 par exemple, pourrait permettre de suggérer à un utilisateur diverses
 notations possibles pour une valeur rythmique, avec des complexités dif-
 férentes.

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

je ne comprend pas bien 489
 l'explication. le pb est plu- 490
 tot vue locale (déduction de 491
 la proba d'une durée à par- 492
 tir de la durée précédente, 493
 par ex. dans un HMM) vs
 vue globale, dans une hié-
 rarchie

RT? 494
 495

techniques de réécriture 496
 appliquée à la déduction 497
 automatique, calcul symbo-
 lique 498

le calcul d'équiv. 499
 500

citer thèse de David Rizo 503
 (Valencia) 504

505

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

Example: *Summertime*

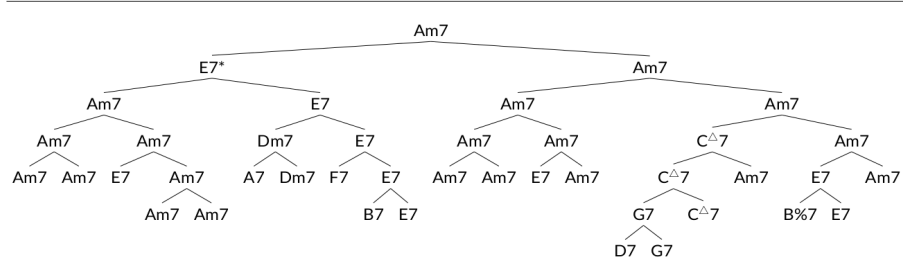


FIGURE 2.2 – arbre_jazz
Représentation arborescente d’une grille harmonique [12]

Conclusion

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

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

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

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

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

latex : enumerate

527

528

MÉTHODES

529

Sommaire

530

531

532

533

534

535

536

537

3.1	La notation de la batterie	25
3.2	Modélisation pour la transcription	32
3.3	Qparse	34
3.4	Les systèmes	35

538

Introduction

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

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

Enfin, nous présenterons les systèmes.

3.1 La notation de la batterie

547



FIGURE 3.1

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

— La ronde, elle vaut 4 ;

550

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

- 551 — La blanche, elle vaut 2 ;
- 552 — La noire, elle vaut 1 ;
- 553 — La croche, elle vaut 1/2.
- 554 Une figure de note [1] de musique combine plusieurs critères² :
- 555 — Une tête de note :
556 Sa position sur la portée indique la hauteur de la note. La tête de
557 note peut aussi indiquer une durée.
- 558 — Une hampe :
559 Indicatrice d'appartenance à une voix en fonction de sa direction
560 et indicatrice d'une durée représentée par sa présence ou non
561 (blanche \neq ronde)
- 562 — Un crochet : La durée d'une note est divisée par deux à chaque
563 crochet ajouté à la hampe d'une figure de note.

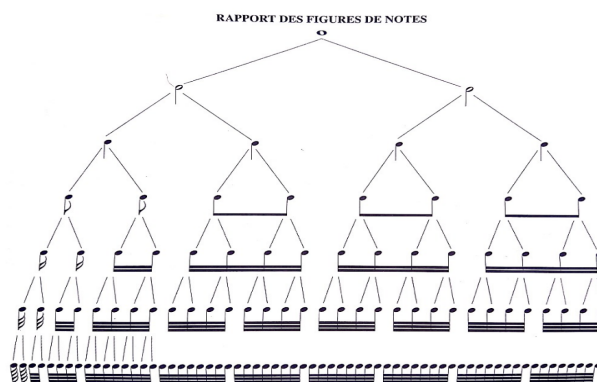


FIGURE 3.2 – Rapport des figures de notes
[1]

564 La figure 3.2 montre les rapports de durée entre les figures de notes. Plus
565 les durées sont longues, plus elles sont marquées par la tête de note (la
566 note carrée fait deux fois la durée d'une ronde) ou la présence ou non de
567 la hampe. À partir de la noire (3ème lignes en partant du haut), on ajoute
568 un crochet à la hampe d'une figure de notes pour diviser sa durée par 2.
569 Les notes à crochet (croche, double-croche, triple...) peuvent être reliées
570 ou non par des ligatures (Voir les 4 dernière lignes de la figure 3.2).

571 Les hauteurs et les têtes de notes

572 Pour la transcription, nous proposons une notation inspirée du recueil de
573 pièces pour batterie de J.-F. Juskowiak [22] et des méthodes de batterie
574 Agostini [23], car nous trouvons la position des éléments cohérente et in-
575 tuitive.
576 En effet, les hauteurs sur la portée représentent :

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

- 577 — La hauteur physique des instruments :
- 578 La caisse claire est centrale sur la portée et sur la batterie (au
- 579 niveau de la ceinture, elle conditionne l'écart entre les pédales et
- 580 aussi la position de tous les instruments basiques d'une batterie).
- 581 Tout ce qui en-dessous de la caisse-claire sur la portée est en
- 582 dessous de la caisse-claire sur la batterie (pédales, tom basse);
- 583 Tout ce qui est au-dessus de la caisse-claire sur la portée, l'est
- 584 aussi sur la batterie.
- 585
- 586 — La hauteur des instruments en terme de fréquences :
- 587 Sauf pour le charley au pied et si l'on sépare en trois groupes
- 588 (grosse-caisse, toms et cymbales), de bas en haut, les instruments
- 589 vont du plus grave au plus aigu.

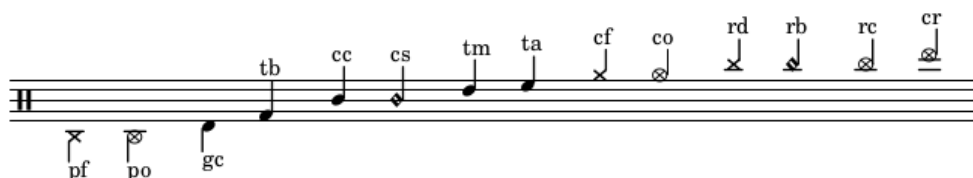


FIGURE 3.3 – Hauteur et têtes de notes

590 Les noms des instruments correspondant aux codes des notes de la figure

591 3.3 sont dans le tableau 3.1.

592 Les durées

593 Comme nous venons de la voir, la majorité des instruments de la batte-

594 rie sont représentés par les têtes des notes. Par conséquent, les symboles

595 rythmiques concernant la tête de note ne pourront pas être utilisés. Cela

596 est valable aussi pour la présence ou non de la hampe puisque ce phé-

597 nomène n'existe qu'avec les têtes de notes de type cercle-vide (opposition

598 blanche-ronde). L'usage des blanches existe dans certaines partitions de

599 batterie [24] mais cela reste dans des cas très rares. Certains logiciels per-

600 mettent de faire des blanches avec des symboles spécifiques à la batterie

601 ou aux percussions mais leur lecture reste peu aisée et leur utilisation

602 pour la batterie est rarissime.

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

- 604 — Le point;
- 605 — La liaison.

606 Ces symboles ne seront utiles que pour l'écriture des ouvertures de char-

607 ley. Le charley est le seul instrument de la batterie dont la durée est quan-

608 titifiée (les cymbales attrapées à la main peuvent l'être aussi mais cela est

609 très rare.)

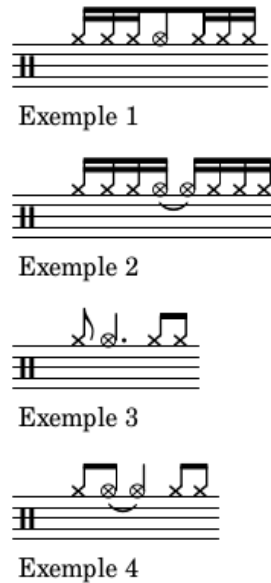


FIGURE 3.4 – Point et liaison

610 L'écriture de la batterie doit faire ressortir la pulsation. La première chose
 611 à prendre en compte pour analyser la figure 3.4 est donc la nécessité de
 612 regrouper les notes par temps à l'aide des ligatures.
 613 Exemple 1 : ouverture de charley quantifiée mais pas notes pas regrou-
 614 pées par temps.
 615 Exemple 2 : Ici, la liaison permet de regrouper les notes par temps en ob-
 616 tenant le même rythme que dans l'exemple 1.
 617 Exemple 3 et exemple 4 : les deux exemples sont valables mais le
 618 deuxième est le plus souvent utilisé car plus intuitif (regroupement par
 619 temps).
 620 En cas de nécessité de rallonger la durée d'une note au-delà de son temps
 621 initial et si cette note correspond à une ouverture de charley, on privilé-
 622 giera la liaison.

623 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 3.5 – Les silences

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

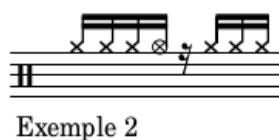
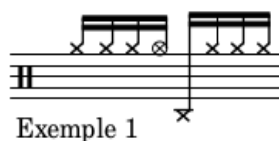


FIGURE 3.6 – Silence joué

631 L'exemple 1 de la figure 3.6 montre ce qui est écrit dans les données MIDI
 632 et l'exemple 2 montre ce que le batteur doit penser en lisant la parti-
 633 tion. Il faut aussi prendre en compte l'écriture surchargée que l'exemple 1
 634 donnerait avec une partition comprenant plusieurs voix et plusieurs ins-
 635 truments jouant simultanément.
 636 Lorsqu'une note est un charley ouvert, il faudra donc prendre en compte
 637 la note suivante pour l'écriture :
 638 - Si c'est un charley fermé joué à la main \Rightarrow la note sera cf;
 639 - Si c'est un charley fermé joué au pied \Rightarrow la note sera un silence.
 640

641 Les équivalences rythmiques

642 Pour les instruments mélodiques, la liaison et le point sont les deux seules
 643 possibilités en cas d'équivalence rythmique pour des notes dont la durée
 644 de l'une à l'autre est ininterrompue. Mais pour la batterie, à part pour
 645 les ouvertures de charley (voir section 3.1), les durées des notes n'ont pas
 646 d'importance. L'usage des silences pour combler la distance rythmique
 647 entre deux notes devient donc possible.
 648 Cela pris en compte, et étant donné que les indications de durée dans les
 649 têtes de notes sont peu recommandées (voir section 3.1), l'écriture à l'aide
 650 de silences sera privilégiée comme indication de durée sauf dans les cas
 651 où cela reste impossible. Ce choix a pour but de n'avoir qu'une manière
 652 d'écrire toutes les notes, que leurs têtes de notes soit modifiées ou non.
 653 Sur la figure 3.7, théoriquement, il faudra choisir la notation de la



FIGURE 3.7 – Équivalence

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

657 Les voix

658 Les voix³ désignent les différentes parties mélodiques constituant une
 659 composition musicale et destinées à être interprétées, simultanément ou
 660 successivement, par un ou plusieurs musiciens. En batterie, une voix est
 661 l'ensemble des instruments qui, à eux seuls, constituent une phrase ryth-
 662 mique et sont regroupés à l'aide des ligatures. Plusieurs écritures étant
 663 possibles pour un même rythme, on peut regrouper les instruments de la
 664 batterie par voix. Sur une portée de batterie, il existe le plus souvent 1 ou
 665 2 voix. Sur la figure 3.8, il faudra faire un choix entre les exemples 1, 2 et
 3 qui sont trois façons d'écrire le même rythme.



FIGURE 3.8 – Séparation des voix

666
 667 Ce choix se fera en fonction des instruments joués, de la nature plus ou
 668 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]

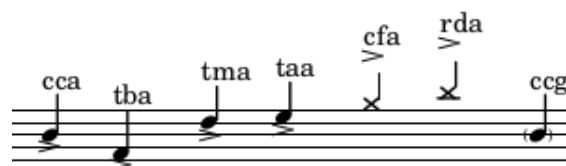


FIGURE 3.9 – Les accents et les ghost-notes

La figure 3.9 ne prend en compte que les accents que nous avons estimés nécessaires (voir la section 3.2). Les accents sont marqués par le symbole « > ». Il est positionné au-dessus des notes représentant des cymbales et en-dessous des notes représentant des toms ou la caisse-claire. Ce choix a été fait pour la partition de la figure 4.2 car elle est plus lisible 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 3.9 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 3.9 sont exposées en situation réelle dans la figure 3.10.



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

691 3.2 Modélisation pour la transcription

692 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

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

698 Malgré le large panel de pitches disponible, il semblerait qu'aucun pitch
 699 ne désigne le charley ouvert joué au pied. Pourtant, dans la batterie mo-
 700 derne, plusieurs rythmes ne peuvent fournir le son du charley ouvert
 701 qu'avec le pied car les mains ne sont pas disponibles pour le jouer. Cela
 702 doit en partie être dû à l'utilisation des boîte à rythmes en MAO qui ne né-
 703 cessitent pas de faire des choix conditionnés par les limitations humaines
 704 (2 pieds, 2 mains, et beaucoup plus d'instruments. . .)

705 La vélocité

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

708 Cette transcription nous a mené aux observations suivantes :

- 709 — Vélocité inférieure à 40 : ghost-note ;
- 710 — Vélocité supérieure à 90 : accent ;
- 711 — Pas d'intention d'accent ni de ghost-note pour une vélocité entre 40
 712 et 89 ;

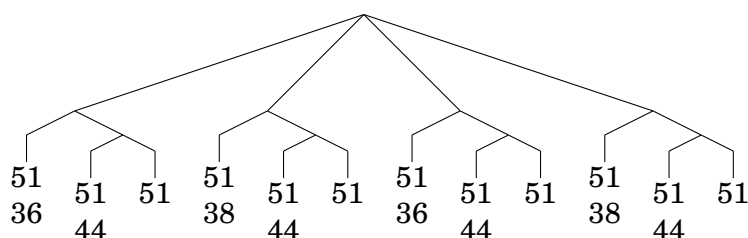
- 713 — Les accents et les ghosts-notes ne sont significatifs ni pour les ins-
 714 truments joués au pied, ni pour les cymbales crash.
 715 En effet, certaines vélocités en dessous de 40 étant détectées et ins-
 716 crites dans les données MIDI sont dues au mouvement du talon du
 717 batteur qui bat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

723 Les arbres de rythmes

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



727
 728 Ci-dessous, le même arbre dont les codes des instruments sont remplacés
 par leurs données MIDI respectives :



729
 730 Chacun des trois exemples de la figure 3.8 est représenté par un des deux
 731 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.

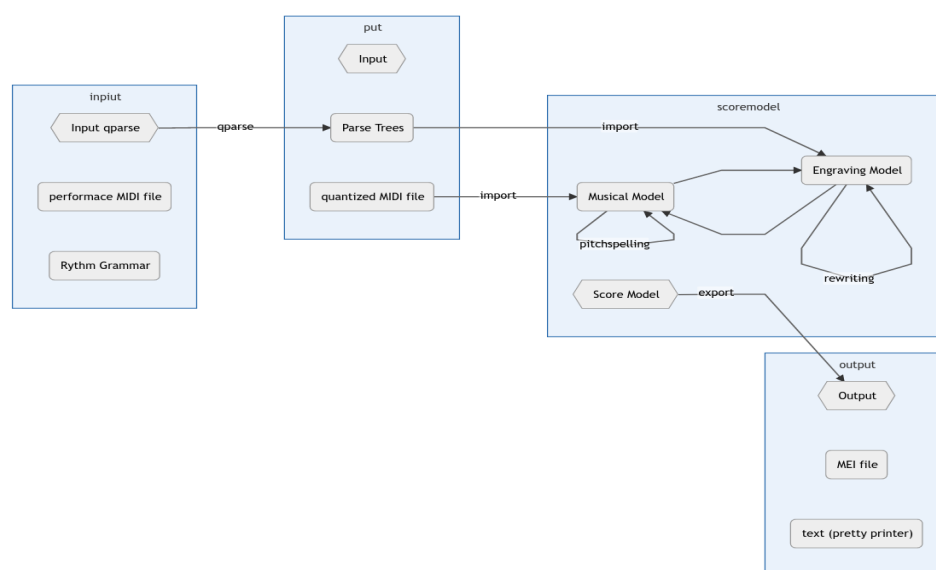


FIGURE 3.11 – Présentation de Qparse

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

Explication des différentes étapes de la figure 3.11⁵ :

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

787 sur 1 ou 2 voix)

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

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

793

794 Un ensemble de systèmes comprenant leur métrique et leurs règles
795 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	duolet, quartelet	2
afros-cubains	simple	croches	-	3

TABLE 3.2 – Systèmes

796

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

798 — 4/4 binaire

799 — 4/4 jazz

800 — 4/4 afro-cubain

801 Objectif des systèmes

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

803 — définir une métrique ;

804 — choisir une grammaire appropriée ;

805 — fournir les règles de réécriture (séparation des voix et simplifica-
806 tion.

807

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

812 Détection d'indication de mesure

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



Exemple 1



Exemple 2

FIGURE 3.12 – Métrique

817 La figure 3.12 montre deux indications de mesure différentes. L'une
818 (exemple 1) est *simple* (2 temps binaires sur lesquels sont joués des trio-
819 lets), l'autre (exemple 2) est *complexe* (2 temps ternaires). Le jazz est tra-
820 ditionnellement écrit en binaire avec ou sans triolet (même si cette mu-
821 sique est dite ternaire alors que le rock ternaire sera plutôt écrit comme
822 dans l'exemple 2).

823 **Choix d'une grammaire**

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

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

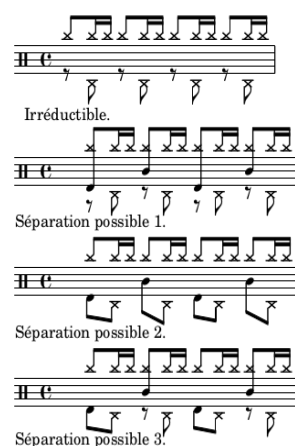
836 **Séparation des voix**

FIGURE 3.13 – Motif 4-4 binaire

837 Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les
 838 2 et 4 avec un choix de répartition des cymbales type fast-jazz. Le système
 839 est constitué par défaut du motif rd/pf/cc (voir 3.1) et d'un texte joué à la
 840 grosse-caisse. La première ligne de la figure 3.13 est appelée « Irréductible
 841 » car il n'y a pas d'autre choix pertinent pour la répartition de la ride et du
 842 charley au pied. La troisième séparation proposée est privilégiée car elle
 843 répartit selon 2 voix, une voix pour les mains (rd + cc) et une voix pour les
 844 pieds (pf + gc). Ce choix paraît plus équilibré car deux instruments sont
 845 utilisés par voix et plus logique pour le lecteur puisque les mains sont en
 haut et les pieds en bas.

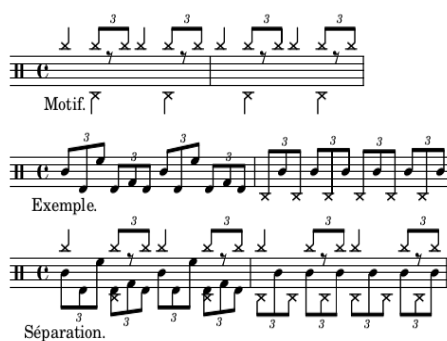


FIGURE 3.14 – Motif 4-4 jazz

846 Dans la plupart des méthodes, le charley n'est pas écrit car il est considéré
 847 comme évident en jazz traditionnel. Ce qui facilite grandement l'écriture :
 848 la ride et les crash sur la voix du haut et le reste sur la voix du bas. Ici,
 849 le parti pris est de tout écrire. Dans l'exemple ci-dessus, les mesures 1 et
 850

851 2 combinées avec le *motif* de la première ligne, sont des cas typiques de
 852 la batterie jazz. Tout mettre sur la voix haute serait surchargé. De plus,
 853 la grosse caisse entre très souvent dans le flot des combinaisons de toms
 854 et de caisse claire et son écriture séparée serait inutilement compliquée
 855 et peu intuitive pour le lecteur. Le choix de séparation sera donc de lais-
 856 ser les cymbales en haut et toms, caisse-claire, grosse-caisse et pédale de
 charley en bas.



FIGURE 3.15 – Système 4-4 afro-latin

857
 858 La figure 3.15 montre un exemple minimaliste de système afro-latin [24].
 859 Ce système doit être écrit sur trois voix car la voix centrale est souvent
 860 plus complexe qu'ici (que des noirs) et la mélanger avec le haut ou le bas
 861 serait surchargé et peu lisible.

862 Simplification de l'écriture

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

865
 866 Les gammes qui accompagnent les motifs d'un système étayent toutes les
 867 combinaisons d'un système et elles permettent, combinées avec le motif
 868 d'un système, de définir les règles de simplification propres à celui-ci.

869
 870 Voici les différentes étapes à suivre :

- 871 — Pour chaque gamme du système, faire un arbre de rythme repré-
 872 sentant la gamme combinée avec le motif du système ;
- 873 — Pour chaque arbre de rythmes obtenus, séparer les voix et faire un
 874 arbre de rythme par voix ;
- 875 — Pour chaque voix (arbre de rythmes) obtenus, extraire tous les
 876 nœuds qui nécessitent une simplification et écrire la règle.

877
 878 Certaines précisions concernant l'extraction de ces règles sont néces-
 879 saires. Il s'agit de précisions à propos de la durée, des silences et de la
 880 présence ou non d'ouverture de charley dans les instruments joués. Nous
 881 avons discuté de ces problèmes dans le chapitre 3.

882
 883 Voici quelques règles inhérentes à la simplification de l'écriture pour
 884 la batterie :

885 Toutes les continuations (t) qui se trouvent en début de temps (figures 4.9,
 886 4.11 et 4.12) sont transformées en silences (r) sauf si la note précédente
 887 est un charley ouvert ?
 888 Même si on favorise l'usage des silences pour l'écart entre les notes
 889 n'appartenant pas au même temps, on les supprime systématiquement
 890 pour 2 notes au sein d'un même temps et favorise, une liaison si co, un
 891 point si pas co et nécessaire, un simple ajustement de la figure de note si
 suffisant.

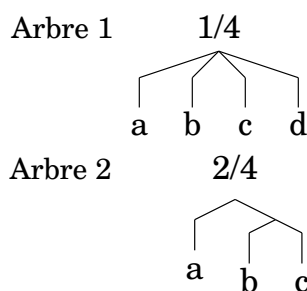


FIGURE 3.16 – Simplification

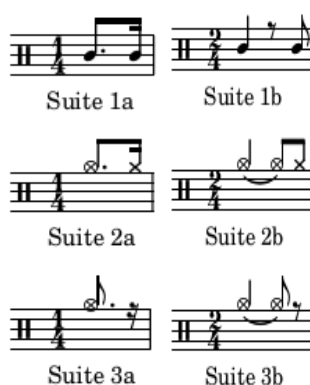


FIGURE 3.17

892
 893 Soit l'arbre 1 de la figure 3.16 dans lequel :
 894 a et d sont des instruments de la batterie (x);
 895 b et c sont des continuations (t); Pour chacune des conditions suivantes,
 896 une suite de la figure 3.17 est attribuée :
 897 — Si a n'est pas un co :
 898 ⇒ Suite 1a.
 899 — Si a est un co :
 900 — Si d est un cf :
 901 ⇒ Suite 2a.

- 902 — Si d est un pf :
903 ⇒ Suite 3a : d deviens un silence (r).
904
- 905 Soit l'arbre 2 de la figure 3.16 dans lequel :
906 a et c sont des instruments de la batterie (x);
907 b est une continuation (t); Pour chacune des conditions suivantes, une
908 suite de la figure 3.17 est attribuée :
- 909 — Si a n'est pas un co :
910 ⇒ Suite 1b, b devient un silence.
911 — Si a est un co :
912 — Si c est un cf :
913 ⇒ Suite 2b, b devient une liaison et c devient un cf.
914 — Si c est un pf :
915 ⇒ Suite 3b : b deviens une liaison et c devient un silence.
916
- 917 *Rappel :*
918 *cf = charley fermé joué à la main ;*
919 *co = charley ouvert joué à la main ;*
920 *pf = charley fermé joué au pied.*
921
- 922 **Problème : le cf et le co ne seront jamais sur la même voix que le**
923 **pf... Par conséquent, les règles concernant les charleys ouverts**
924 **doivent-elles être appliquées sur l'arbre de parsing de l'input ?...**

925 Conclusion

926 Nous avons formalisé une notation de la batterie, modélisé cette notation
927 pour la transcription de données MIDI en partition, nous avons décrit
928 Qparse.
929 Enfin, nous avons exposé une approche de type dictionnaire (les « sys-
930 tèmes ») pour détecter une métrique, choisir une grammaire pondérée ap-
931 propriée et énoncer des règles de séparation des voix et de simplification
932 de l'écriture.

933

934

EXPÉRIMENTATIONS

935

Sommaire

936

937

938

939

940

941

942

943

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

944

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.

952

4.1 Le jeu de données

953

954

955

956

957

958

959

960

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

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

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



FIGURE 4.1 – Batterie électronique

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

Autres critères spécifiques au GMD :

- Toutes les performances ont été jouées au métronome et à un tempo choisi par le batteur.
- 80% de la durée du GMD a été joué par des batteurs professionnels qui ont pu improviser dans un large éventail de styles. Les données sont donc diversifiées en termes de styles et de qualités de jeu (professionnel ou amateur).
- Les batteurs avaient pour instruction de jouer des séquences de plusieurs minutes ainsi que des fills²
- Chaque performance est annotée d'un style (fourni par le batteur), d'une métrique et d'un tempo ainsi que d'une identification anonyme du batteur.
- Il a été demandé à 4 batteurs d'enregistrer le même groupe de 10 rythmes dans leurs styles respectifs. Ils sont dans les dossiers évaluation du GMD.
- Les sorties audio synthétisées ont été alignées à 2 ms près sur leur fichier MIDI.

Format des données

Le Roland TD-11 divise les données enregistrées en plusieurs pistes distinctes :

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

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

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

990
 991 « Control Changes The TD-11 also records control changes speci-
 992 fying the position of the hi-hat pedal on each hit. We have preserved this
 993 information under control 4. »
 994 (<https://magenta.tensorflow.org/datasets/groove>)
 995 ⇒ ??? Je ne comprends pas encore comment trouver ce type d'informa-
 996 tions dans les fichiers MIDI.
 997 L'utilisation de pretty_midi devient urgente !

998 4.2 Analyse MIDI-Audio

999 Ces analyses ont été faites dans le cadre de transcriptions manuelles à
 1000 partir de fichiers MIDI et Audio du GMD.

1001 Comparaisons de transcriptions

1002 Pour les comparaisons de transcriptions, les transcriptions manuelles
 1003 (TM) ont été éditées à l'aide de Lilypond³ ou MuseScore⁴ et les trans-
 1004 criptions automatiques (TA) ont toutes été générées manuellement avec
 1005 MuseScore.

1006 Exemple d'analyse 1

Transcription manuelle ⇒ Transcription automatique



- 1007 — Erreur d'indication de mesure (3/4 au lieu de 4/4);
- 1008 — Les silences de la mesure 1 de la TA sont inutilement surchargés;
- 1009 — La noire du temps 4 de la mesure 1 de la TM est devenue les deux
- 1010 premières notes (une double-croche et une croche) d'un triolet sur
- 1011 le temps 1 de la mesure 2 de la TA.

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

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

1012 **Exemple d'analyse 2**

Transcription manuelle \Rightarrow Transcription automatique



- 1013 — Les doubles croches ont été interprétées en quintolet
 1014 — La deuxième double-croche est devenue une croche.
 1015

1016 **Exemple d'analyse 3**

Transcription manuelle \Rightarrow Transcription automatique



- 1017 — Les grosses-caisses, les charleys et les caisses-claires ont été décalés d'un temps vers la droite.
 1018 — 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.
 1019 — La première caisse-claire de la mesure 1 devient binaire dans la TA alors qu'elle appartenait à un triolet dans la TM.
 1020 — Le triolet de tom-basse du temps 4 de la mesure 2 de la TA n'existe pas la TM.
 1021
 1022
 1023
 1024
 1025

1026 **Exemple d'analyse 4**

Transcription manuelle \Rightarrow Transcription automatique



- 1027
 1028 Sur le temps 4 de la mesure 1, la deuxième croche a été transcrite d'une
 1029 manière excessivement complexe!

1030 **Exemple avec des flas**

1031 Transcription manuelle



1032

1033 Transcription automatique

1034



1035

1036

- 1037 — Le premier fla est reconnu comme étant un triolet contenant une
- 1038 quadruple croche suivie d'une triple croche au lieu d'une seule note
- 1039 ornementée.
- 1040 — Le deuxième fla est reconnu comme étant un accord.
- 1041 — Les deux double en l'air sur le temps 4 de la TM sont mal quantifiée
- 1042 dans la TA.
- 1043 — La TA ne reconnaît qu'une mesure quand la TM en transcrit deux.
- 1044 En effet, la TA a divisé par deux la durée des notes afin de les faire
- 1045 tenir dans une mesure à 4 temps dont les unités de temps sont
- 1046 les noires. Par exemple, le soupir du temps 2 de la TM devient un
- 1047 demi-soupir sur le contre-temps du temps 1 dans la TA. Ou encore,
- 1048 la noire (pf, voir le tableau 3.1) sur le temps 1 de la mesure 2 de
- 1049 la TM suivie d'un demi-soupir devient une croche pointée sur le
- 1050 temps 3 de la TA.
- 1051 — Autre problème : certaines têtes de notes sont mal attribuées. Par
- 1052 exemple, le charley ouvert en l'air sur le temps 2 de la mesure 2
- 1053 de la TM devrait avoir le même symbole sur la TA. Idem pour les
- 1054 cross-sticks.

1055 **Transcription de partition**

FIGURE 4.2 – Partition de référence

1056 La figure 4.2 est la transcription manuelle des fichiers *004_jazz-*
 1057 *funk_116_beat_4-4.mid* et *004_jazz-funk_116_beat_4-4.wav* du GMD.

1058 Cette transcription a été entièrement faite avec Lilypond (voir le code
 1059 lilypond sur le git [https://github.com/MartinDigard/Stage_M2_](https://github.com/MartinDigard/Stage_M2_Inria)
 1060 [Inria](https://github.com/MartinDigard/Stage_M2_Inria)) Il s'agit d'une partition d'un 4/4 binaire dont le fichier MIDI est
 1061 annoncé dans le GMD de style «jazz-funk» probablement en raison de
 1062 la ride de type shabada rapide (le ternaire devient binaire avec la vi-
 1063 tesse) combiné avec l'after-beat de type rock (caisse-claire sur les deux
 1064 et quatre).

1065 La transcription des données audio et MIDI contenues dans ces fichiers
 1066 a permis une analyse plus approfondie des critères à relever pour chaque
 1067 évènement MIDI et de la manière de les considérer dans un objectif de
 1068 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 mesure 23 à la mesure 28 avec des variations. Ces deux motifs sont très classiques et pourront être détectés dans de nombreuses performances.

Gammes

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

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

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

1092 dont la première ne contient que des croches et la deuxième que des
 1093 doubles-croches en l'air. Ces deux possibilités sont combinées de manière
 1094 circulaire dans les lignes 4 et 5 qui contiennent chacune 3 mesures.

1095 Systèmes — motifs et gammes combinés

1096 Pour la suite de l'expérimentation théorique, nous utiliserons le motif 1
 1097 de la figure 4.3.

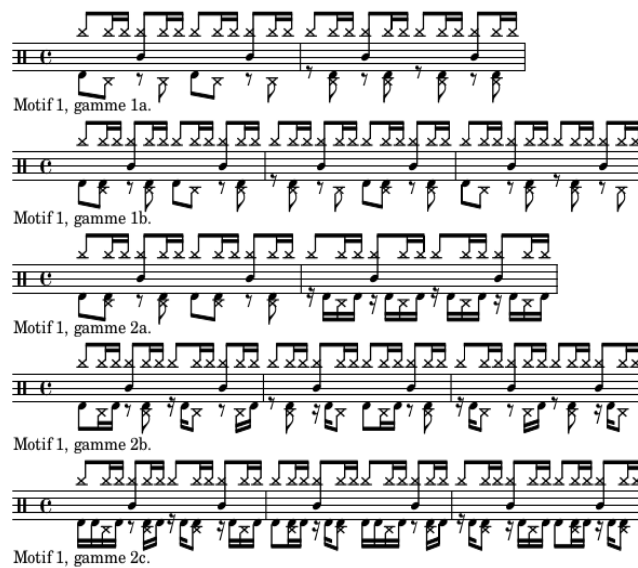


FIGURE 4.4 – Partition d'un système en 4/4 binaire

1098

1099 Représentation du système en arbres de rythmes

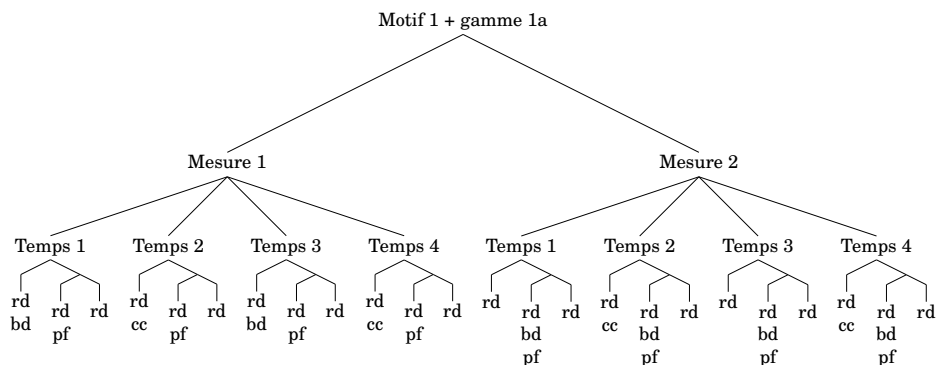


FIGURE 4.5 – Arbre de rythme — système

1100 L’arbre de la figure 4.5 servira de base pour le suite de l’expérimentation.
1101 Comme indiqué à la racine de l’arbre, il représente la première ligne de la
1102 figure 4.4. Même si cet arbre représente parfaitement le rythme concerné,
1103 il manque des indications de notation telles que les voix spécifiques à
1104 chaque partie du rythme ainsi que les choix d’écriture pour les distances
1105 qui séparent les notes de chaque voix entre elles en termes de durée.

1106 **Réécriture — séparation des voix et simplification**

1107 **La séparation des voix**

1108 Ainsi l’arbre syntaxique de départ est divisé en autant d’instruments qui
1109 le constituent et les voix seront regroupées en suivant les règles du sys-
tème.

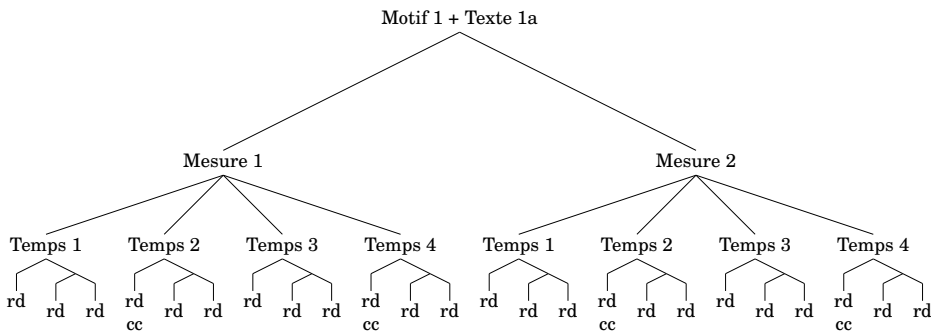


FIGURE 4.6 – Arbres de rythme — voix haute

1110
1111 La voix haute regroupe la ride et la caisse-claire sur les ligatures du haut.

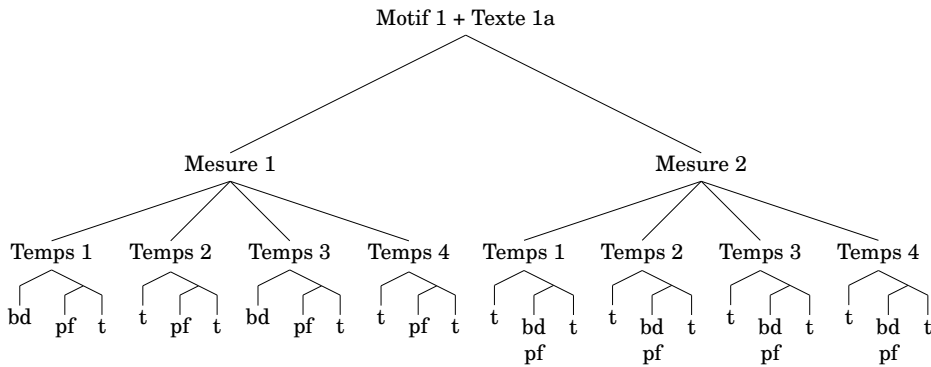


FIGURE 4.7 – Arbres de rythme — voix basse

1112
1113 La voix basse regroupe la grosse-caisse et le charley au pied sur les liga-
1114 tures du bas.

1115 Les règles de simplifications

1116 L'objectif des règles de simplifications est de réécrire les écarts de durées
 1117 qui séparent les notes d'une manière appropriée pour la batterie et qui
 1118 soit la plus simple possible. Les ligatures relient les notes d'un temps
 1119 entre elles (rendre la pulse visuelle).

1120

1121 Pour les figures ci-dessous :

1122 — x = une note ;

1123 — r = un silence ;

1124 — t = une continuation (point ou liaison)

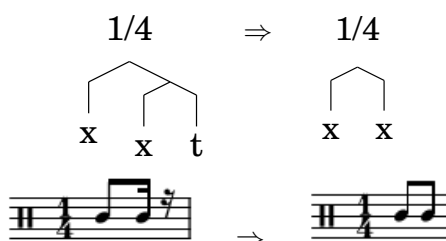


FIGURE 4.8

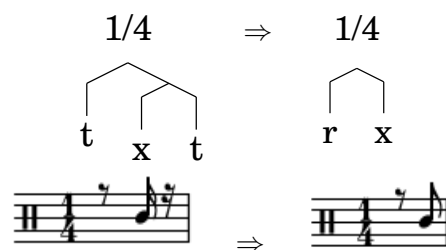


FIGURE 4.9

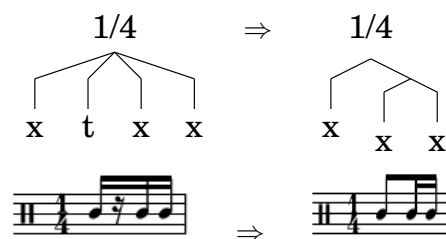


FIGURE 4.10

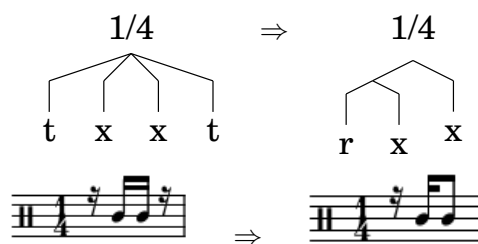


FIGURE 4.11

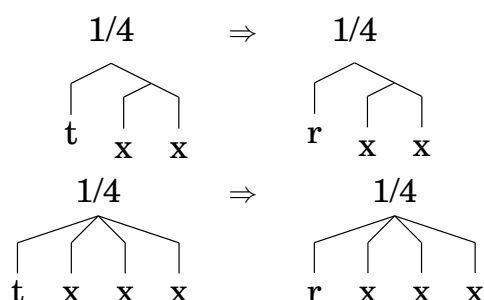


FIGURE 4.12

1125 Ces règles ont été tirées de l'ensemble des arbres du système. Les arbres
 1126 manquants seront mis en annexe.

1127

1128 Les règles remplacent par un silence les continuations (t) qui sont
 1129 au début d'un temps. Cela est valable pour ce système mais lorsqu'il y a
 1130 des ouvertures de charley, cela n'est pas toujours applicable. Ce problème
 1131 est évoqué de le chapitre 3.

1132

1133 \Rightarrow **Objectif de cette expérimentation théorique :**

1134 La méthode des *systèmes* étant basée sur une approche dictionnaire, cette
 1135 expérimentation théorique a pour but d'orienter la recherche d'autres
 1136 systèmes par observation du jeu de données et de montrer comment les
 1137 construire pour agrandir la base de connaissance de Qparse pour l'ADT.

1138 4.4 Résultats et discussion

1139 Cette section regroupe les avancées qui ont été réalisées par rapport aux
 1140 objectifs de départ ainsi qu'une réflexion sur le moyen d'évaluer les résul-
 1141 tats de l'ADT avec Qparse. Nous avons amélioré le système de quantifi-
 1142 cation de Qparse pour la batterie, notamment le passage à la polyphonie
 1143 avec les Jams. Nous avons pu obtenir des arbres de parsing correctes en
 1144 améliorant les grammaires avec des fichiers MIDI courts. Puis, une sortie
 1145 MEI a aussi été obtenue (encore à vérifier).

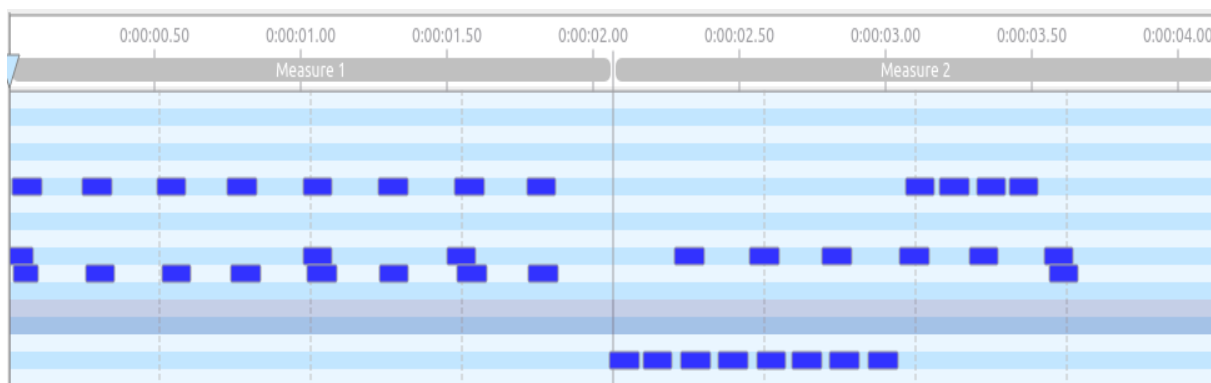
1146 Les Jams

1147 Les Jams permettent de passer du monophonique au polyphonique.

1148 Le parsing

1149 Tests effectués avec le fichier midi suivant :

1150



1151

1152

1153 Un premier test convaincant est effectué avec la grammaire sui-
 1154 vante :

1155

1156 // bar level

1157 0 -> C0 1

1158 0 -> E1 1

1159 0 -> U4(1, 1, 1, 1) 1

1160

1161 // half bar level

1162 9 -> C0 1

1163 9 -> E1 1

1164

1165 // beat level

1166 1 -> C0 1

1167 1 -> E1 1

1168 1 -> T2(2, 2) 1
 1169 1 -> T4(4, 4, 4, 4) 1
 1170

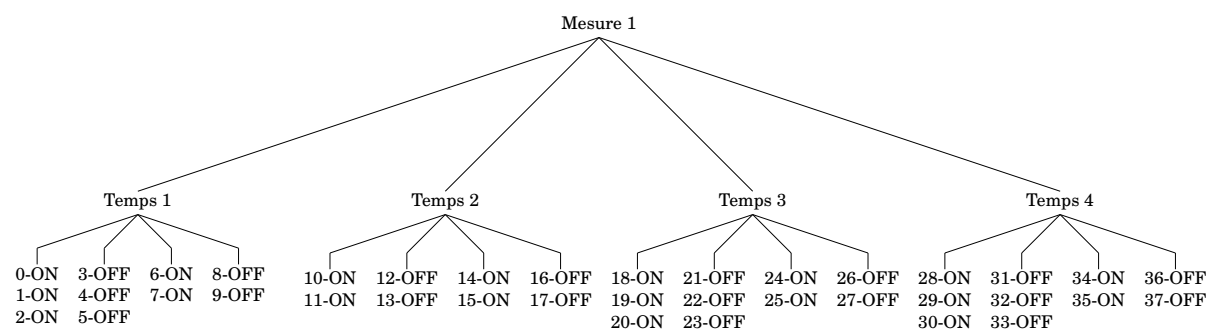
1171 // croche level
 1172 2 -> C0 1
 1173 2 -> E1 1
 1174

1175 // double level
 1176 4 -> C0 1
 1177 4 -> E1 1
 1178 4 -> E2 1
 1179 4 -> T2(6, 6) 1
 1180

1181 // triple level
 1182 6 -> E1 1
 1183

1184 Cette grammaire sépare les ligatures par temps au niveau de la
 1185 mesure. Puis, au niveau du temps, elle autorise les divisions par deux
 1186 (croches) et par quatre (doubles-croches). Tous les poids sont réglés sur 1.
 1187 L'arbre de parsing en résultant est considéré comme « convaincant » car
 1188 il découpe correctement les mesures et les temps.

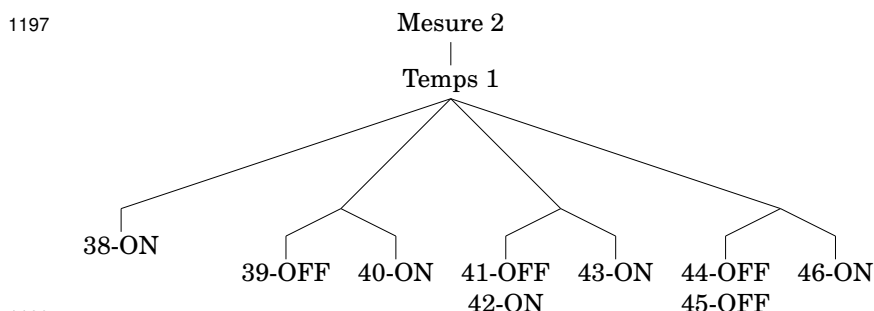
1189
 1190



1191
 1192

1193 Les temps de la première mesure du fichier MIDI sont bien quanti-
 1194 fié mais ceux de la deuxième mesure présentent quelques défauts de
 1195 quantification visibles dès le premier temps.

1196



1200 Les Onsets sont correctement triés au niveau des doubles croches
 1201 mais certaines doubles croches sont inutilement subdivisées en triples
 1202 croches (les 2ème, 3ème et 4ème doubles croches sur le premier temps
 1203 ci-dessus).

1204 2ème exemple :

1205 Après une augmentation du poids des triples croches dans la grammaire
 1206 (monté de 1 à 5) et une baisse de tous les autres poids (descendu de 1 à
 1207 0.5), et mis à part le troisième temps de la 2ème mesure, tous les Onsets
 1208 sont bien triés et aucuns ne sont subdivisés.

1210 Évaluation

1211 Pour l'évaluation, il aurait fallu produire un module.
 1212 L'évaluation est-elle automatique ou manuelle?
 1213 Possibilité d'un export lilypond en arbre pour comparer l'output avec la
 1214 transcription manuelle.
 1215 Possibilité de transformer lilypond(output) et lilypond(ref) en ScoreModel
 1216 ou MEI pour les comparer et faire des statistiques. Si transformés en
 1217 MEI : diffscore de Francesco. Possibilité de transformer lilypond(output)
 1218 et lilypond(ref) en MusicXML pour les comparer ou dans Music21.
 1219 L'expérimentation peut-être considérer comme une évaluation manuelle?
 1220 (magicien d'Oz)
 1221 Lilypond vers MIDI + output vers MIDI ⇒ Comparaison des MIDI
 1222 dumpés.

1224 Discussion

1225 Dans cette section, nous discuterons sur la pertinence de l'ensemble des
 1226 choix qui ont été faits. Nous ferons un bilan des différentes avancées qui
 1227 ont été faites ou non et nous tenterons d'en expliquer la ou les raisons.
 1228 Écrire des règles de réécriture spécifique aux charley avec un système
 1229 approprié. Le jeu de système

- 1230 — implémenter un pattern. . .
- 1231 ⇒ manque de temps ?
- 1232
- 1233 — La partie résultat est manquante car :
- 1234 ⇒ Sujet très difficile ;
- 1235 ⇒ Matcher les motifs peut être fait ultérieurement ;
- 1236 Mais ce travail aurait été indispensable pour obtenir une quan-
- 1237 tité de résultats qui justifieraient une évaluation automatique
- 1238 permettant de faire des graphiques.
- 1239
- 1240 — L'évaluation fut entièrement manuelle car :
- 1241 ⇒ Très dure automatiquement : il faut comparer 2 partitions (réf
- 1242 VS output)
- 1243 — Le ternaïre jazz (voir expérience 2)
- 1244 — Reconnaissance d'un motif sur le MIDI
- 1245 Reconnaître un motif (système) sur une mesure de l'input (un fi-
- 1246 chier midi représentant des données audios)
- 1247 ⇒ Motif (système) reconnu : true ou false
- 1248 Si true :
- 1249 - Choisir la grammaire correspondante ;
- 1250 - Parser le MIDI ;
- 1251 - Appliquer les règles de réécritures (Séparation des voix et simpli-
- 1252 fication)
- 1253 — Nous travaillerons aussi sur la détection de répétitions sur plu-
- 1254 sieurs mesures afin de pouvoir corriger des erreurs sur une des
- 1255 mesures qui aurait dû être identique aux autres mais qui présente
- 1256 des différences.
- 1257 — dans quelle catégorie mettre le shuffle ?
- 1258 Sujet passionnant mais difficile. Obtenir la totalité des critères pour le
- 1259 mémoire n'aurait pas pu être fait sans bâcler. Une base solide spécifique
- 1260 à la batterie a été générée. Elle sera un bon point de départ pour les
- 1261 travaux futurs dont plusieurs propositions sont énoncés dans le présent
- 1262 document.

CONCLUSION GÉNÉRALE

1264 Dans ce mémoire, nous avons traité de la problématique de la transcrip-
1265 tion automatique de la batterie. Son objectif était de transcrire, à partir
1266 de leur représentation symbolique MIDI, des performances de batteur de
1267 différents niveaux et dans différents styles en partitions écrites.
1268 Nous avons avancé sur le parsing des données MIDI établissant un pro-
1269 cessus de regroupement des événements MIDI qui nous a permis de faire
1270 la transition du monophonique vers le polyphonique. Une des données
1271 importante de ce processus était de différencier les nature des notes d'un
1272 accord, notamment de distinguer lorsque 2 notes constituent un accord
1273 ou un *fla*.
1274 Nous avons établis des *grammaires pondérées* pour le parsing qui corres-
1275 pondent respectivement à des métriques spécifiques. Celles-ci étant sélec-
1276 tionnables en amont du parsing, soit par indication des noms des fichiers
1277 MIDI, soit par reconnaissance de la métrique avec une approche diction-
1278 naire de patterns prédéfinis⁵ qu'il serait pertinent de mettre en œuvre en
1279 machine learning.
1280 Nous avons démontré que l'usage des *systèmes* élimine un grand nombre
1281 de calcul lors de la réécriture. Pour la séparation des voix grâce au motif
1282 d'un système et pour la simplification grâce aux gammes du motif d'un
1283 système. Nous avons aussi montré comment, dans des travaux futurs, un
1284 système dont le motif serait reconnu en amont dans un fichier MIDI pour-
1285 rait prédéfinir le choix d'une grammaire par la reconnaissance d'une mé-
1286 trique et ainsi améliorer le parsing et accélérer les choix ultérieurs dans
1287 la chaîne de traitement en terme de réécriture.
1288 Il sera également intéressant d'étudier comment l'utilisation de LM peut
1289 améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la génération
1290 entièrement automatisée de partitions de batterie et au problème général
1291 de l'AMT de bout en bout.[10]

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

BIBLIOGRAPHIE

- [1] A. Danhauser. *Théorie de la musique*. Edition Henry Lemoine, 41 rue Bayen - 75017 Paris, Édition revue et augmentée - 1996 edition, 1996. – Cité pages 7, 26 et 31.
- [2] H. C. Longuet-Higgins. Perception of melodies. 1976. – Cité pages 9 et 12.
- [3] Wikipedia. Music informatics. Available at https://en.wikipedia.org/wiki/Music_informatics (2021/01/06). – Cité page 9.
- [4] Meinard Müller. *Fundamentals of Music Processing*. 01 2015. – Cité page 9.
- [5] Gaël Richard et al. De fourier à la reconnaissance musicale. Available at <https://interstices.info/de-fourier-a-la-reconnaissance-musicale/> (2019/02/15). – Cité page 9.
- [6] Caroline Traube. Quelle place pour la science au sein de la musicologie aujourd’hui? *Circuit*, 24(2) :41–49, 2014. – Cité page 10.
- [7] Bénédicte Poulin-Charronnat and Pierre Perruchet. Les interactions entre les traitements de la musique et du langage. *La Lettre des Neurosciences*, 58 :24–26, 2018. – Cité page 11.
- [8] Mikaela Keller, Kamil Akesbi, Lorenzo Moreira, and Louis Bigo. Techniques de traitement automatique du langage naturel appliquées aux représentations symboliques musicales. In *JIM 2021 - Journées d’Informatique Musicale*, Virtual, France, July 2021. – Cité page 11.
- [9] Junyan Jiang, Gus Xia, and Taylor Berg-Kirkpatrick. Discovering music relations with sequential attention. In *NLP4MUSA*, 2020. – Cité page 11.
- [10] Emmanouil Benetos, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri. Automatic music transcription : Challenges and future directions. *Journal of Intelligent Information Systems*, 41, 12 2013. – Cité pages 12, 14, 19 et 59.

- 1324 [11] Kentaro Shibata, Eita Nakamura, and Kazuyoshi Yoshii. Non-local
1325 musical statistics as guides for audio-to-score piano transcription.
1326 *Information Sciences*, 566 :262–280, 2021. – Cité pages 12, 21 et 22.
- 1327 [12] Daniel Harasim, Christoph Finkensiep, Petter Ericson, Timothy J
1328 O'Donnell, and Martin Rohrmeier. The jazz harmony treebank. –
1329 Cité pages 12 et 23.
- 1330 [13] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Ge-
1331 rhard Widmer, Jason Hockman, Meinard Müller, and Alexander
1332 Lerch. A review of automatic drum transcription. *IEEE/ACM Tran-*
1333 *sactions on Audio, Speech, and Language Processing*, 26(9) :1457–
1334 1483, 2018. – Cité pages 14, 20 et 23.
- 1335 [14] Moshekwa Malatji. Automatic music transcription for two instru-
1336 ments based variable q-transform and deep learning methods, 10
1337 2020. – Cité page 20.
- 1338 [15] Antti J. Eronen. Musical instrument recognition using ica-based
1339 transform of features and discriminatively trained hmms. *Seventh*
1340 *International Symposium on Signal Processing and Its Applications*,
1341 *2003. Proceedings.*, 2 :133–136 vol.2, 2003. – Cité page 20.
- 1342 [16] Hiroshi G. Okuno Kazuyoshi Yoshii, Masataka Goto. Automatic
1343 drum sound description for real-world music using template adap-
1344 tation and matching methods. *International Conference on Music*
1345 *Information Retrieval (ISMIR)*, pages 184–191, 2004. – Cité page 21.
- 1346 [17] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Ma-
1347 sahiko Sakai. A Parse-based Framework for Coupled Rhythm Quan-
1348 tization and Score Structuring. In *MCM 2019 - Mathematics and*
1349 *Computation in Music*, volume Lecture Notes in Computer Science
1350 of *Proceedings of the Seventh International Conference on Mathema-*
1351 *tics and Computation in Music (MCM 2019)*, Madrid, Spain, June
1352 2019. Springer. – Cité pages 21 et 22.
- 1353 [18] C. Agon, K. Haddad, and G. Assayag. Representation and rende-
1354 ring of rhythm structures. In *Proceedings of the First International*
1355 *Symposium on Cyber Worlds (CW'02)*, CW '02, page 109, USA, 2002.
1356 IEEE Computer Society. – Cité page 22.
- 1357 [19] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A
1358 Term Rewriting Based Structural Theory of Rhythm Notation. Re-
1359 search report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. –
1360 Cité page 22.
- 1361 [20] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating
1362 equivalent rhythmic notations based on rhythm tree languages. In
1363 *Third International Conference on Technologies for Music Notation*

- 1364 *and Representation (TENOR)*, Coroña, Spain, May 2017. Helena Lo-
1365 pez Palma and Mike Solomon. – Cité page 22.
- 1366 [21] R. Marxer and J. Janer. Study of regularizations and constraints in
1367 nmf-based drums monaural separation. In *International Conference*
1368 *on Digital Audio Effects Conference (DAFx-13)*, Maynooth, Ireland,
1369 02/09/2013 2013. – Cité page 23.
- 1370 [22] J.-F. Juskowiak. *Rythmiques binaires 2*. Alphonse Leduc, Editions
1371 Musicales, 175, rue Saint-Honoré, 75040 Paris, 1989. – Cité page 26.
- 1372 [23] Dante Agostini. *Méthode de batterie, Vol. 3*. Dante Agostini, 21, rue
1373 Jean Anouilh, 77330 Ozoir-la-Ferrière, 1977. – Cité page 26.
- 1374 [24] O. Lacau J.-F. Juskowiak. *Systèmes drums n. 2*. MusicCom publica-
1375 tions, Editions Joseph BÉHAR, 61, rue du Bois des Joncs Marins -
1376 94120 Fontenay-sous-Bois, 2000. – Cité pages 27 et 39.
- 1377 [25] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David
1378 Bamman. Learning to groove with inverse sequence transforma-
1379 tions. In *International Conference on Machine Learning (ICML)*,
1380 2019. – Cité page 43.

