
Institut National des Langues et Civilisations Orientales

Département Textes, Informatique, Multilinguisme

Titre du mémoire

MASTER

TRAITEMENT AUTOMATIQUE DES LANGUES

Parcours :

Ingénierie Multilingue

par

Martin DIGARD

Directeur de mémoire :

Damien NOUVEL

Encadrant :

Florent JACQUEMARD

Année universitaire 2020/2021

TABLE DES MATIÈRES

Liste des figures	4
Liste des tableaux	4
Introduction	7
I Contexte général	9
1 État de l'art	11
1.1 Introduction	11
1.2 Contenu	11
1.3 Conclusion	12
2 Méthodes	13
2.1 Introduction	13
2.2 Contenu	15
2.3 Conclusion	20
II Expérimentations	21
3 Corpus	23
3.1 Introduction	23
3.2 Contenu	26
3.3 Conclusion	26
4 Résultats	27
4.1 Introduction	27
4.2 Contenu	27
4.3 Conclusion	27
5 Discussion	29
5.1 Introduction	29
5.2 Contenu	29
5.3 Conclusion	29
Conclusion générale	31
Bibliographie	33

LISTE DES FIGURES

LISTE DES TABLEAUX

2.1	Pitchs et instruments	15
-----	---------------------------------	----

En règle générale, l'introduction et la conclusion sont les deux sections de contenu à ne pas être numérotées. Idéalement, chaque chapitre commence par une introduction rapide et se termine par une conclusion rapide pour aider le lecteur à mémoriser et comprendre ce qui a été fait.

INTRODUCTION

Introduction : présentation générale du contexte et de la problématique traitée, plan suivi dans le mémoire.

Présentation générale

Sujet du stage

L'objectif de la transcription automatique de la musique (AMT) [1] 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. Elle est considérée comme l'un des problèmes de recherche les plus anciens et les plus difficiles dans le domaine de la recherche d'information musicale (MIR).

Le cas de la transcription de la batterie (DT) est très particulier puisqu'il s'agit d'instruments sans hauteur, d'événements avec (presque) aucune durée et de notations spécifiques. Il a été la source de nombreuses études MIR, voir [2] pour un aperçu. La plupart de ces travaux se concentrent sur des méthodes de calcul pour la détection d'événements sonores de batterie à partir de signaux acoustiques, et sur l'extraction de caractéristiques de bas niveau telles que la classe d'instrument et le moment de l'apparition du son (peak picking). Cependant, très peu d'entre eux ont abordé la tâche de générer une notation musicale (rythmique) lisible à partir des caractéristiques ci-dessus, une étape cruciale dans un contexte musical et loin d'être triviale.

La présente proposition s'intéresse à ce dernier problème, et plus précisément à :

1. l'étude de modèles de langage (LM) incorporant certaines informations musicales de haut niveau nécessaires à la génération de partitions de qualité. On devrait en particulier considérer des hiérarchies d'événements de batterie induisant des placements temporels cohérents et se prêtant à des notations rythmiques faciles à lire pour un batteur entraîné; voir [3] pour des modèles structurés en arbre basés sur la théorie formelle du langage, que nous développons dans le contexte d'outils AMT plus généraux.

2. l'intégration de ces LM avec l'état de l'art des modèles acoustiques (AM) et des méthodes pour les tâches de traitement du signal ci-dessus. Cela nécessite la prise en compte du contexte musical et des informations musicales de haut niveau des LM en plus des caractéristiques acoustiques de bas niveau ci-dessus.

En outre, certaines expériences seront menées sur la base d'ensembles de données publiques, afin d'évaluer l'approche intégrée. Elles devraient couvrir certains cas de motifs rythmiques complexes se chevauchant.

Au-delà de l'intégration de modèles, il sera également intéressant d'étudier comment l'utilisation de LM peut améliorer les résultats de l'AM, voir [2], et ouvrir la voie à la

génération entièrement automatisée de partitions de batterie et au problème général de l'AMT de bout en bout. [2]

Problématique traitée

En entrée : midi (séquence d'événements datés (piano roll) accompagné d'une grammaire pondérée)

⇒ parsing

⇒ global parsing tree

⇒ RI (Représentation Intermédiaire) arbres locaux par instruments

⇒ Sortie (xml, mei, lilypond, . . .)

Minimiser la distance entre le midi et la représentation en arbre.

Le but du stage est d'améliorer qparse, un outil de transcription et d'écriture automatique de la batterie (entre autre)

Le sujet de ce mémoire est de proposer une tâche de reconnaissance du regroupement des notes par les ligatures dans l'écriture de la batterie.

Pour cela, nous utiliserons la logique des systèmes (selon la définition agostinienne).

⇒ Motif répétitif de plusieurs instruments coordonnés accompagnés d'un texte varié joué par un autre instrument de la batterie.

Nous partirons de propositions génériques de systèmes (environs trois systèmes dans différents style de batterie) que nous tenterons de détecter dans le jeu de données groove.

Nous travaillerons aussi sur la détection de répétitions sur plusieurs mesures afin de pouvoir corriger des erreurs sur une des mesures qui aurait dû être identique au autres mais qui présente des différences.

Plan suivi dans le mémoire

- écouter le dataset groove
- observer la structure midi
- décrire la notation de la batterie et transcrire manuellement pour comparer l'input et l'output idéal d'un point de vu théorique.
- Détecter les systèmes dans le dataset.
- Faire des règles de réécriture (rewriting/simplification) et de séparation des voix (reconnaissance de systèmes)

Première partie

Contexte général

ÉTAT DE L'ART

Sommaire

1.1	Introduction	11
1.2	Contenu	11
1.3	Conclusion	12

L'état de l'art (chapitre 1) : les articles qui traitent du même sujet que vous, présentés en un tout cohérent (extraire de chaque article les points essentiels et présenter dans ce chapitre le résultat de ces lectures en regroupant les articles par point essentiel)

1.1 Introduction

Dans ce chapitre, nous présentons les différentes avancées qui ont déjà eues lieu dans le domaine de la transcription.

1.2 Contenu

Automatic music transcription : Challenges and future directions [1]
(introduction[1])

Les applications de l'AMT ont aussi de la valeur dans les domaines où il manque de partition (jazz, pop, (et donc batterie, note perso))
(abstract [1])

Les différents travaux existants se préoccupent plus de la transcription à partir de l'audio en passant par le traitement du signal.

Les humains sont encore meilleurs que les machines et la précision à l'air d'avoir atteint sa limite.

Analyse des limites des méthodes courantes et identification des directions prometteuses.

Les modèles généraux utilisés ne traitent pas correctement la riche diversité des signaux musicaux.

2 moyens pour surmonter cela :

- adapter l'algorithme pour des cas d'utilisations spécifiques.
- utiliser les approches semi-automatiques.

La richesse des partitions musicales et des données audio correspondantes, désormais disponibles, constitue une source potentielle de données d'apprentissage, grâce à l'alignement forcé des données audio sur les partitions, mais l'utilisation à grande échelle de ces données n'a pas encore été tentée.

D'autres approches prometteuses incluent l'intégration d'informations provenant de plusieurs algorithmes et de différents aspects musicaux.

A Review of Automatic Drum Transcription[\[2\]](#)

1.3 Conclusion

Conclusion de ce chapitre.

MÉTHODES

Sommaire

2.1	Introduction	13
2.1.1	Chaîne de traitement	13
2.2	Contenu	15
2.2.1	Caractéristiques pour la représentation et la notation	15
2.2.2	Les systèmes	16
2.3	Conclusion	20

Méthodes (chapitre 2) : les méthodes appliquées, avec le détail des expériences réalisées (différentes configurations);

Corpus (chapitre 3) : le corpus utilisé (caractéristiques, pré-traitements appliqués)

2.1 Introduction

Dans ce chapitre...

2.1.1 Chaîne de traitement

- Reconnaître un motif (système) dans une partition (un fichier midi)
Sur une mesure de l'input \Rightarrow Motif (système) reconnu : true ou false
- Si true :
 - Séparer les voix
 - Simplifier l'écriture de chaque voix

Chaîne de traitement :

```
if match(system_parse_tree(system + pitch), input_midi_parse_tree)
  then voice_split;
  for voice in voice_split :
    simplification voice
```

Les systèmes en batterie

SYSTÈME \Rightarrow MOTIF (2-3 instruments sur 1-2 voix, joué en boucle) + TEXTE (1 instrument sur 1 voix, irrégulier)

ex : système afro-cubain, trois voix. Proposition pour la détection de la direction des hampes et pour les ligatures (regroupement des notes et séparation des voix.)

— **Les systèmes :**

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

Définition d'un système :

En cas de système, les ligatures forment deux voies :

- Le texte ;
- Le système.

Mettre des exemples de différents systèmes.

— **Les moulins :**

Lorsqu'il y a plus d'une voie, ils sont prioritaires pour les ligatures.

Mettre des exemples.

Gestion des silences et des têtes de notes

Rythme tree (RT)

- Le symbole "-" est une continuation (liaison) mais pour une partition de batterie, ça serait un silence par défaut sauf peut-être pour les ouvertures de charley et éventuellement les cymbales ou un tom basse qui résonne.

⇒ Question de la (notation noir + silence) vs blanche.

⇒ On privilégierait (noir + silence) puisque les symboles « x » des cymbales ne peuvent pas porter d'indication de durée dans la tête de notes.

Les 3 parties d'une note :

- durée
- hampe
- tête de note (peut aussi indiquer la durée mais en batterie on évitera les blanches, etc.)

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

Définition des têtes de notes et des hauteurs

Proposition de définition d'un standard de départ

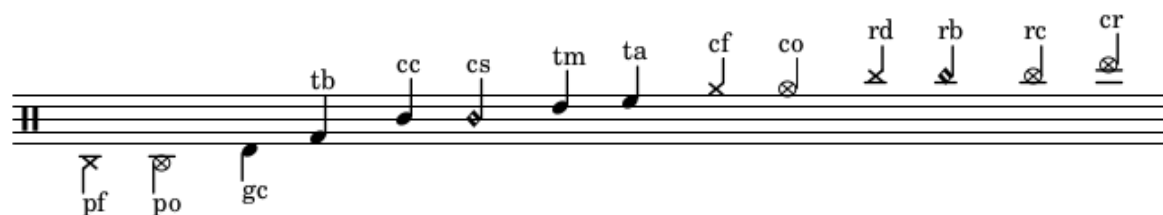
Pour la transcriptions, nous proposons de choisir la base Agostini. La caisse claire centrale sur la portée est aussi centrale sur la batterie est elle est un élément qui conditionne la position des jambes (écart entre les pédales, etc.) ainsi que l'organisation des éléments en hauteur (toms, cymbales, etc.). On pensera en terme de symétrie la répartition des éléments par rapport au point central que constitue la caisse claire.

Cette symétrie s'opère en trois dimensions :

- Les hauteurs en terme de fréquences ;
- La hauteur physique des éléments :
Du bas vers le haut : pédales, toms et caisse, cymbales
- L'ergonomie, qui hiérarchise l'importance des éléments sur la portée (caisse claire au centre, hh-pied et ride sont aux deux extrémités).

2.2 Contenu

2.2.1 Caractéristiques pour la représentation et la notation



Codes	Instruments	Pitches
pf	charley-pied-fermé	44
po	charley-pied-ouvert	?
gc	grosse-caisse	36
tb	tom-basse	43, 58
cc	caisse-claire	38, 40
cs	cross-stick	37
tm	tom-medium	45, 47
ta	tom-alto	48, 50
cf	charley-main-fermé	00
co	charley-main-ouvert	00
rd	ride	51
rb	ride-cloche (bell)	53
rc	ride-crash	59
cr	crash	55

TABLE 2.1 – Pitches et instruments



Si la vélocité est en dessous de 40, il s'agit de ghost-notes : la tête de note devra être entourée de parenthèses et le suffixe *p* (*piano*) devra être ajouté au codes de l'instrument. (Voir ccp ci-dessus.)

Si la vélocité est au dessus de 90, il s'agit de notes accentuées : le symbole « > » et le suffixe *f* (*forte*) devra être ajouté au codes de l'instrument. (Voir ccf ci-dessus.)

Lorsque la vélocité va de 40 à 89, on considèrera le volume comme normal et aucun symbole supplémentaire ne sera ajouté à la note.

L'instrument qui sera difficile à placer sera la caisse claire car elle ne sera pas toujours affiliée aux mêmes instruments.

2.2.2 Les systèmes

Créer un ensemble de systèmes :

- Trouver les systèmes intéressants dans groove ;
- Compléter avec des systèmes ago personnalisés ;
- Tout transcrire avec Lilypond et en arbres d'analyse syntaxique.
- Créer les arbres de voix séparées.
- Créer les arbres de voix séparées simplifiés (rewriting).

Choix d'écriture

Irréductible.

Séparation possible 1.

Séparation possible 2.

Ici, le système est construit sur un modèle rock en 4/4 : after-beat sur les 2 et 4 avec un choix de répartition des cymbales type fast-jazz.

Le système est constitué par défaut du motif ride/ch-pf/cc et d'un texte joué à la grosse-caisse.

Le motif 1 est privilégié car il réparti selon 2 voix, une voix pour les mains (ride + cc) et une voix pour les pieds (ch-pf + gc). Ce choix paraît plus équilibré car deux instruments sont utilisés par voix et plus logique pour le lecteur puisque les mains sont en haut et les pieds en bas.

D'autres choix d'écriture auraient été possibles :

- Toutes les hampes en haut ;
- Combinaison motif 1 et 2 en donnant 2 directions aux hampes de la cc).

À partir de ce choix d'écriture, le système suivant est défini :

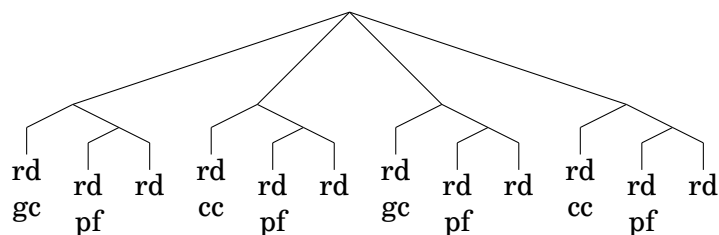
The image shows three staves of musical notation. The top staff, labeled 'Texte.', is in common time (C) and contains a sequence of eighth notes. The middle staff, labeled 'Motif.', is also in common time and contains a sequence of eighth notes, with some notes marked with 'x' to indicate specific rhythmic values. The bottom staff, labeled 'Système.', is in common time and contains a sequence of eighth notes, with some notes marked with 'x' to indicate specific rhythmic values.

Dans le motif, la grosse caisse est notée car elle fait partie de l'équilibre musical de ce rythme mais rigoureusement, elle ne devrait pas y apparaître puisque qu'elle sera indiquée par le texte.

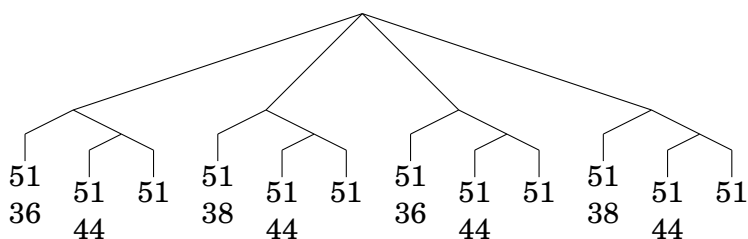
Dans ce système, les noires ne doivent jamais tomber en même temps que les caisses-claires, donc, quand le texte donnera des noires sur les deuxième et quatrième temps, elles ne seront jamais jouées par le batteur (exemple sur la première mesure du système).

Représentations en arbres

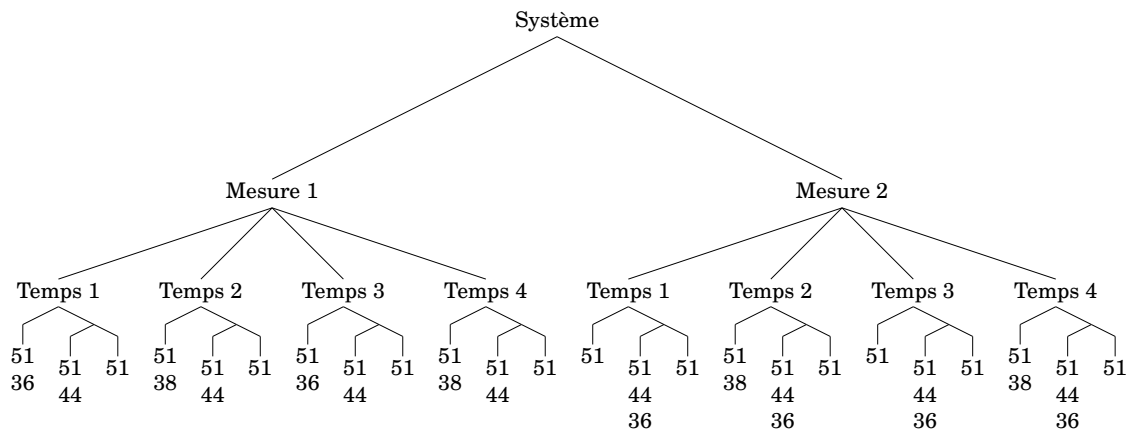
Quelque soit le choix d'écriture, la représentation arborescente du motif ci-dessus donnera le résultat suivant :



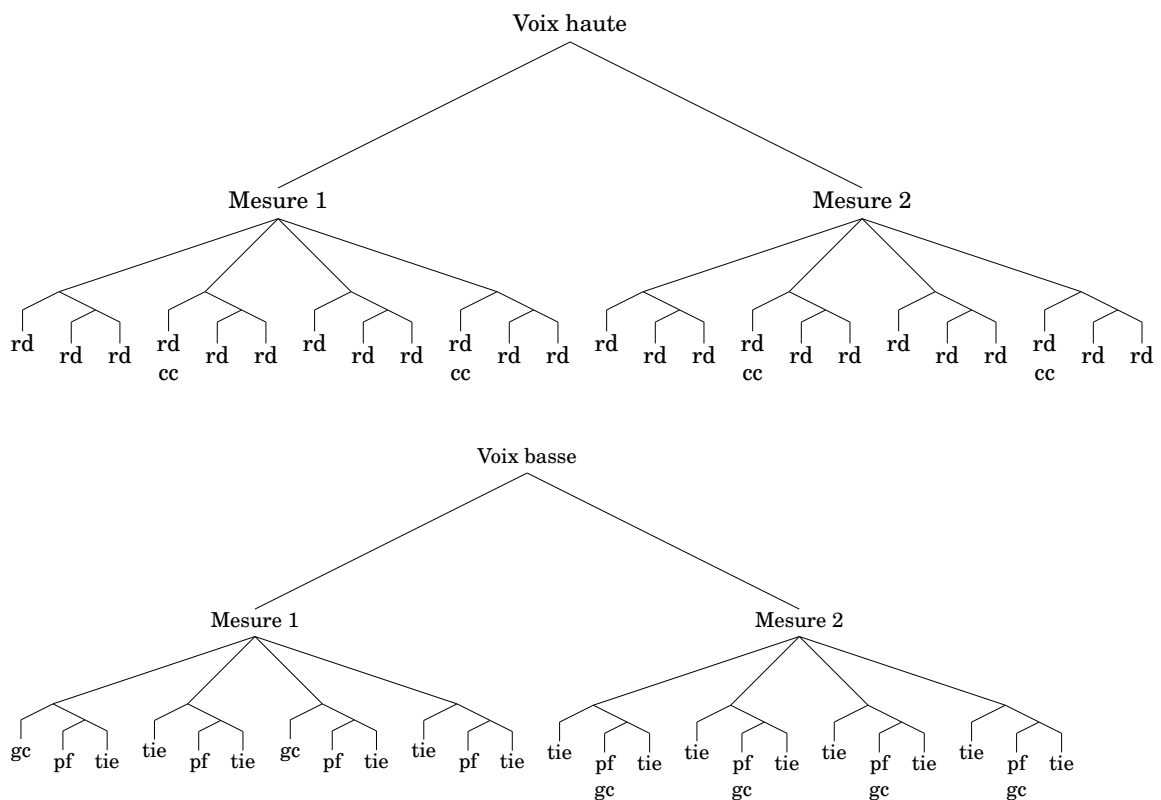
Représentation du même motif avec les pitches midi :



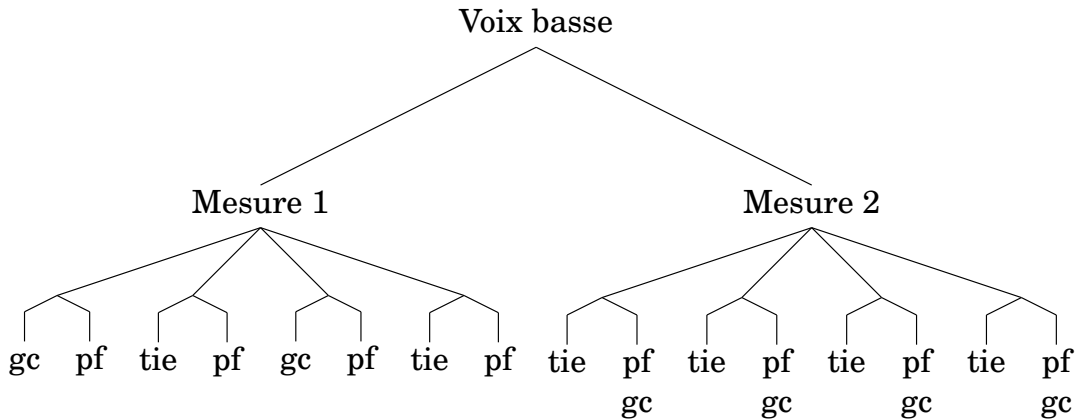
Représentation du système (motif + texte) avec les pitches midi :



Séparation des voix



Réécriture (Simplification)



La voix haute reste inchangée.

La réécriture des événements MIDI pour la batterie

Basé sur [4] et sur [5]

Pour la plupart des instruments mélodiques, la liaison et le point sont les deux seules possibilités en cas d'équivalence rythmique pour des notes dont la durée de l'une à l'autre est ininterrompue. Mais puisque les durées des notes n'ont pas d'importance en batterie, l'usage des silences pour combler la distance rythmique entre deux notes devient possible.

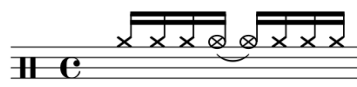
Les cymbales-crash et les ouvertures de charley constituent le seul cas qui exclut cette option. Le charley car ses ouvertures/fermetures sont presque toujours quantifiées et les cymbales-crash car elles peuvent être arrêtées à la main de manière quantifié aussi mais ce cas est très rare, nous allons donc nous concentrer sur les ouvertures de charley et considérer les crashes comme des événements sans durée.

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

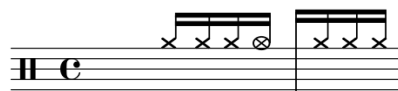
Exemples :



A1 — Évènement MIDI.



A2 — Réécriture.



B1 — Évènement MIDI.

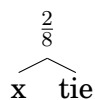


B2 — Réécriture.

Exemples à écrire en arbre :

- SI (pas pf) ET (note sur un temps suivie de note en l'air) :
⇒ (Temps1 : Note pertinente) + (Temps2 : Silence pertinent + Note pertinente.)
- Si (po ou co) déborde sur le temps suivant :
⇒ Liaison car marchera dans tous les cas même la où le point ne marchera pas (voir A2).
- Une blanche sera écrite noir + soupir.

Les règles de réécriture



2.3 Conclusion

Conclusion de ce chapitre.

Deuxième partie

Expérimentations

CORPUS

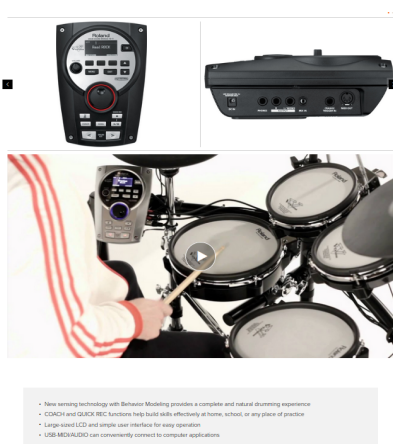
Sommaire

3.1	Introduction	23
3.2	Contenu	26
3.3	Conclusion	26

3.1 Introduction

Dans ce chapitre... **groove MIDI dataset**

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



Des batteurs pro ont été engagés pour jouer sur un roland td-11

The Groove MIDI Dataset (GMD), has several attributes that distinguish it from existing ones:

- The dataset contains about 13.6 hours, 1,150 MIDI files, and over 22,000 measures of drumming.
- Each performance was played along with a metronome set at a specific tempo by the drummer.
- The data includes performances by a total of 10 drummers, with more than 80% of duration coming from hired professionals. The professionals were able to improvise in a wide range of styles, resulting in a diverse dataset.
- The drummers were instructed to play a mix of long sequences (several minutes of continuous playing) and short beats and fills.
- Each performance is annotated with a genre (provided by the drummer), tempo, and anonymized drummer ID.
- Most of the performances are in 4/4 time, with a few examples from other time signatures.
- Four drummers were asked to record the same set of 10 beats in their own style. These are included in the test set split, labeled `eval-session/groove1-10`.
- In addition to the MIDI recordings that are the primary source of data for the experiments in this work, we captured the synthesized audio outputs of the drum set and aligned them to within 2ms of the corresponding MIDI files.

Les métadatas :

The metadata file (`info.csv`) has the following fields for every MIDI/WAV pair:

Field	Description
drummer	An anonymous string ID for the drummer of the performance.
session	A string ID for the recording session (unique per drummer).
id	A unique string ID for the performance.
style	A string style for the performance formatted as "<primary>/<secondary>". The primary style comes from the Genre List below.
bpm	An integer tempo in beats per minute for the performance.
beat_type	Either "beat" or "fill"
time_signature	The time signature for the performance formatted as "<numerator>-<denominator>".
midi_filename	Relative path to the MIDI file.
audio_filename	Relative path to the WAV file (if present).
duration	The float duration in seconds (of the MIDI).
split	The predefined split the performance is a part of. One of "train", "validation", or "test".

Genre List: afrobeat, afrocuban, blues, country, dance, funk, gospel, highlife, hiphop, jazz, latin, middleeastern, neworleans, pop, punk, reggae, rock, soul

A train/validation/test split configuration is provided for easier comparison of model accuracy on various tasks.

Split	Beats	Fills	Measures (approx.)	Hits	Duration (minutes)
Train	378	519	17752	357618	648.5
Validation	48	76	2269	44044	82.2
Test	77	52	2193	43832	84.3
Total	503	647	22214	445494	815.0

Détails (entre autres tensorflow avec le dataset) à : <https://magenta.tensorflow.org/datasets/groove#license>

3.2 Contenu

Une section dans ce chapitre...

3.3 Conclusion

Conclusion de ce chapitre.

RÉSULTATS

Sommaire

4.1	Introduction	27
4.2	Contenu	27
4.3	Conclusion	27

Résultats (chapitre 4) : les résultats obtenus sur chacune des expériences ;

4.1 Introduction

Dans ce chapitre...

4.2 Contenu

Une section dans ce chapitre...

4.3 Conclusion

Conclusion de ce chapitre.

DISCUSSION

Sommaire

5.1	Introduction	29
5.2	Contenu	29
5.3	Conclusion	29

Discussion (chapitre 5) : la discussion des résultats obtenus (quelle expérience a produit les meilleurs résultats, de manière globale, dans le détail des catégories) avec, si possible, une analyse des erreurs pour comprendre les possibilités d'amélioration ;

5.1 Introduction

Dans ce chapitre...

5.2 Contenu

Une section dans ce chapitre...

5.3 Conclusion

Conclusion de ce chapitre.

CONCLUSION GÉNÉRALE

Conclusion : la conclusion globale du mémoire.

Dans ce mémoire, nous avons traité de la problématique...

BIBLIOGRAPHIE

- [1] 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 7 et 11.
- [2] Chih-Wei Wu, Christian Dittmar, Carl Southall, Richard Vogl, Gerhard Widmer, Jason Hockman, Meinard Müller, and Alexander Lerch. A review of automatic drum transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(9) :1457–1483, 2018. – Cité pages 7, 8 et 12.
- [3] Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, and Masahiko Sakai. A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring. In *MCM 2019 - Mathematics and Computation in Music*, volume Lecture Notes in Computer Science of *Proceedings of the Seventh International Conference on Mathematics and Computation in Music (MCM 2019)*, Madrid, Spain, June 2019. Springer. – Cité page 7.
- [4] Florent Jacquemard, Pierre Donat-Bouillud, and Jean Bresson. A Term Rewriting Based Structural Theory of Rhythm Notation. Research report, ANR-13-JS02-0004-01 - EFFICACe, March 2015. – Cité page 19.
- [5] Florent Jacquemard, Adrien Ycart, and Masahiko Sakai. Generating equivalent rhythmic notations based on rhythm tree languages. In *Third International Conference on Technologies for Music Notation and Representation (TENOR)*, Coruña, Spain, May 2017. Helena Lopez Palma and Mike Solomon. – Cité page 19.

