

Département Textes, Informatique, Multilinguisme

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 Représentations, systèmes et réécriture	15
2.3 Conclusion	21
II Expérimentations	23
3 Corpus	25
3.1 Introduction	25
3.2 Contenu	26
3.3 Conclusion	34
4 Résultats	35
4.1 Introduction	35
4.2 Contenu	35
4.3 Conclusion	35
5 Discussion	37
5.1 Introduction	37
5.2 Contenu	37
5.3 Conclusion	37
Conclusion générale	39
Bibliographie	41

LISTE DES FIGURES

LISTE DES TABLEAUX

2.1	Pitches et instruments	16
2.2	Vélocité et nuances	16

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 les précisions 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	Représentations, systèmes et réécriture	15
2.2.1	La notation de la batterie	15
2.2.2	La notation midi de la batterie	15
2.2.3	Les systèmes	18
2.3	Conclusion	21

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.

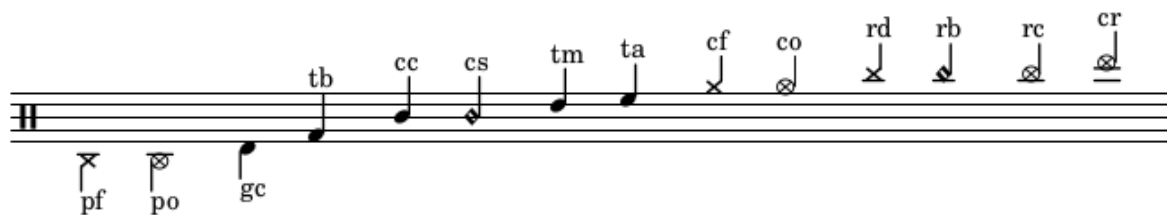
Cette symétrie s'opère en trois dimensions :

- 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 Représentations, systèmes et réécriture

2.2.1 La notation de la batterie

La hauteur des notes et les têtes de notes



partition 1

Les nuances



partition 2

La séparation des voix



partition 3

2.2.2 La notation midi de la batterie

Les pitches et la vitesse

Nous ne prendrons en compte la vitesse que pour la cc, les toms et les cymbales jouées aux mains. Les nuances de grosse caisse et charley aux pieds sont le plus souvent insignifiants, elles ne sont marquées sur le figure qu'à titre indicatif.

Si la vitesse 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.)

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	?
co	charley-main-ouvert	26
rd	ride	51
rb	ride-cloche (bell)	53
rc	ride-crash	59
cr	crash	55

TABLE 2.1 – Pitches et instruments

Codes	Instruments	Pitches	Vélocité
cop	charley-main-ouvert	46	?

TABLE 2.2 – Vélocité et nuances

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.

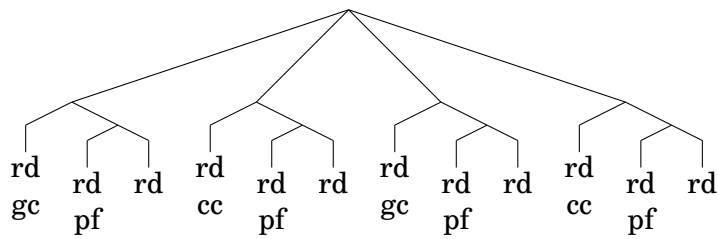
L'instrument qui sera difficile à placer sera la caisse claire car elle ne sera pas toujours affiliée aux mêmes instruments.

Les dilemmes

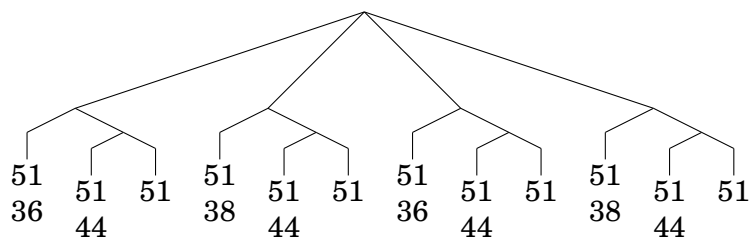
Le charley de pitch 46 est considéré comme le charley ouvert joué à la main sur le haut de la cymbale mais souvent, ça correspond au geste « tranche-olive » de la baguette lorsque le batteur accentue avec la tranche et joue moins fort avec l'olive sur le plat de la cymbale. Je vais dans un premier temps considérer le pitch comme **charley-main-ouvert-piano** (ghost-note)

Représentations en arbres

Voici une représentation de la *partition 3* en arbre de rythme avec les codes de chaque instrument :



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



Cet arbre représente un rythme unique dont les possibilités de notation sur une partition sont théoriquement multiples. (Voir *partition 3*).

2.2.3 Les systèmes

Définition

Système = motif + texte

motif = rythmes coordonnés joués avec 2 ou 3 membres en boucle

texte = rythme irrégulier joué avec un seul membre sur le motif.

Quatre systèmes standards :

- binaire
- ternaire (shuffle, afro, rock)
- jazz
- afro-cubain

Utilité

- Séparation des voix
- Définir une métrique
- Conditionner des règles spécifiques de réécriture

Créer un ensemble de systèmes :

- 4/4 binaire FAIT
- jazz vs ternaire(12/8) EN COURS...
- afro-cubain
- Tout transcrire avec lilypond et en arbres d'analyse syntaxique.
- Créer les arbres de voix séparées.
- Écrire les règles de réécriture.
- Créer les arbres de voix séparées simplifiés (rewriting).

Pour la **séparation des voix** et la **définition des métriques**, nous nous intéresserons principalement à la partie *motif* des systèmes qui seront présentés. La partie *texte* nous intéressera plus pour les **combinaisons de réécritures**.

Pour la séparation des voix

Motif 4-4 binaire

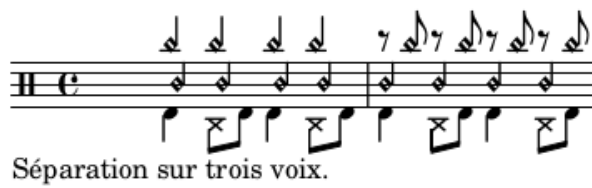
The image shows three staves of musical notation for a 4/4 time signature. Each staff starts with a treble clef and a common time signature 'C'. The first staff is labeled 'Irréductible.' and shows a sequence of eighth notes with 'x' marks above them. The second staff is labeled 'Séparation possible 1.' and shows the same sequence with some notes grouped. The third staff is labeled 'Séparation possible 2.' and shows a different grouping. The fourth staff is labeled 'Séparation possible 3.' and shows a third grouping. The notation uses 'x' marks to indicate specific notes or groups of notes.

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. La troisième séparation proposée est privilégiée car elle répartit 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.

Motif 4-4 jazz

The image shows three staves of musical notation for a 4/4 time signature. Each staff starts with a treble clef and a common time signature 'C'. The first staff is labeled 'Motif.' and shows a sequence of eighth notes with 'x' marks above them. The second staff is labeled 'Exemple.' and shows a sequence of eighth notes with 'x' marks above them. The third staff is labeled 'Séparation.' and shows a sequence of eighth notes with 'x' marks above them. The notation uses 'x' marks to indicate specific notes or groups of notes.

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

Système 4-4 afro-cubain**Pour la reconnaissance de la métrique*****12/8 vs 4/4 ternaire*****Motif 12/8****Pour les règles de réécriture**

Les textes qui accompagnent les motifs étayent toutes les combinaisons d'un système.

Construction des systèmes pour les expérimentations**2.3 Conclusion**

Conclusion de ce chapitre.

Deuxième partie

Expérimentations

CORPUS

Sommaire

3.1	Introduction	25
3.2	Contenu	26
3.2.1	Expérience 1 - 4/4 binaire	26
3.2.2	Expérience 2	34
3.3	Conclusion	34

3.1 Introduction

Dans ce chapitre...

3.2 Contenu

3.2.1 Expérience 1 - 4/4 binaire

Partition de référence pour l'ouput

Music engraving by LilyPond 2.22.1—www.lilypond.org

Systèmes recherchés

Textes :


Débit croches.




Débit doubles-croches.



Motifs :




Motif 1.




Motif 2.


Systèmes résultants :




Motif 1, texte 1a.




Motif 1, texte 1b.




Motif 1, texte 2a.




Motif 1, texte 2b.




Motif 1, texte 2c.




Motif 2, texte 1a.




Motif 2, texte 1b.



Motif 2, texte 2a.

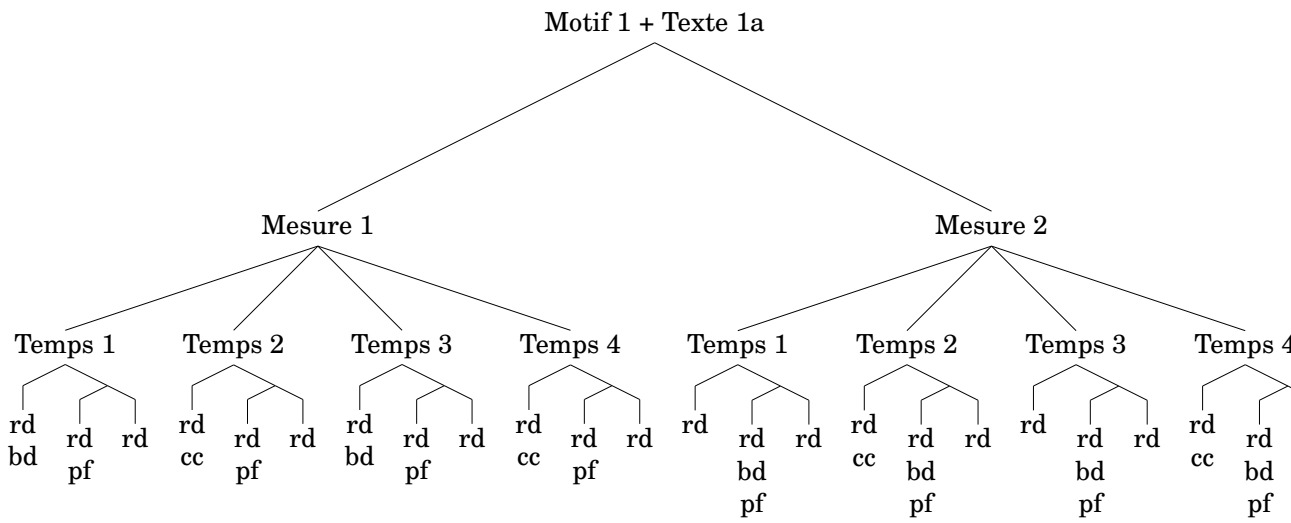


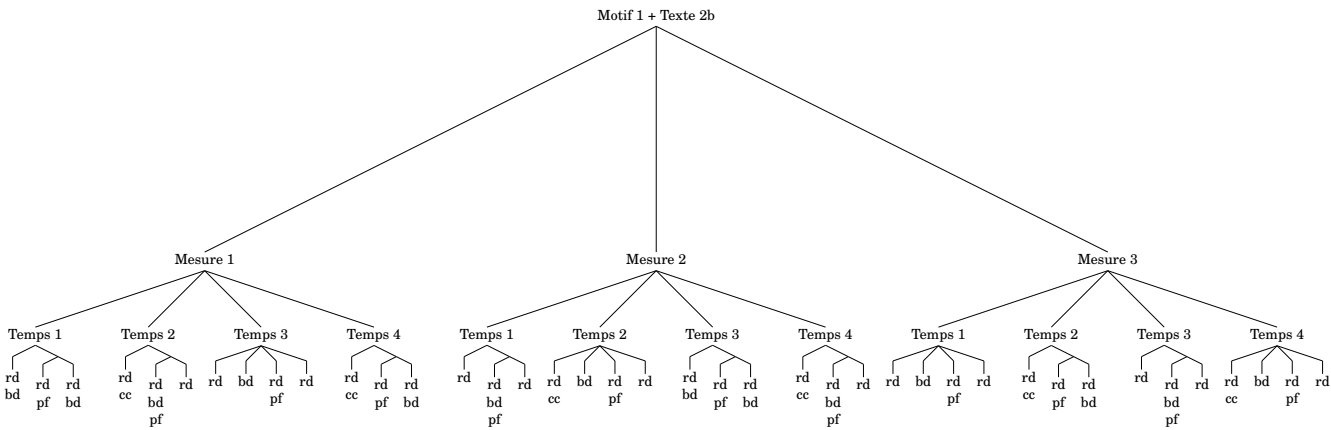
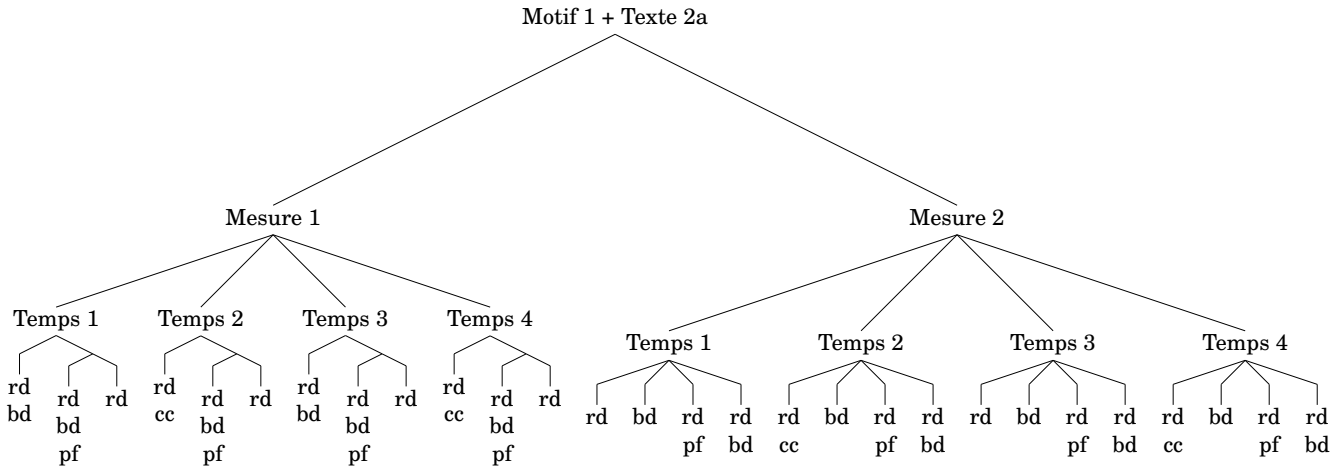
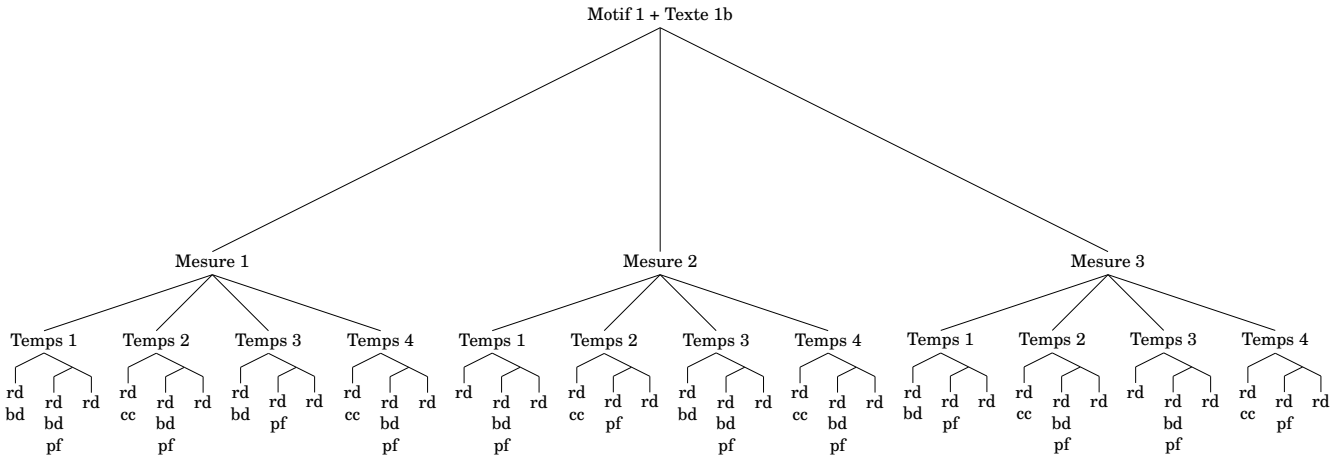
Motif 2, texte 2b.

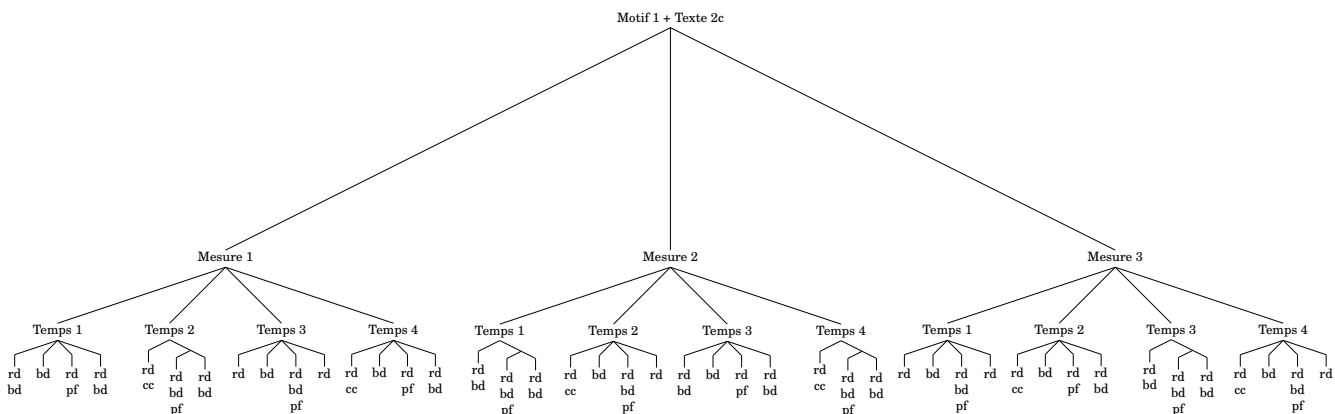


Motif 2, texte 2c.

Représentation des systèmes en arbres de rythmes



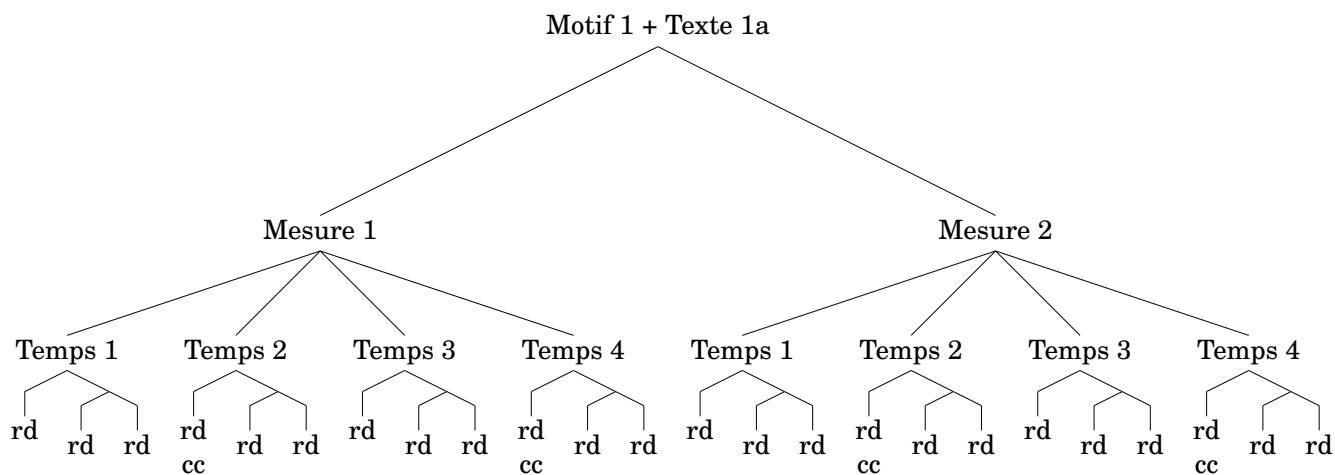




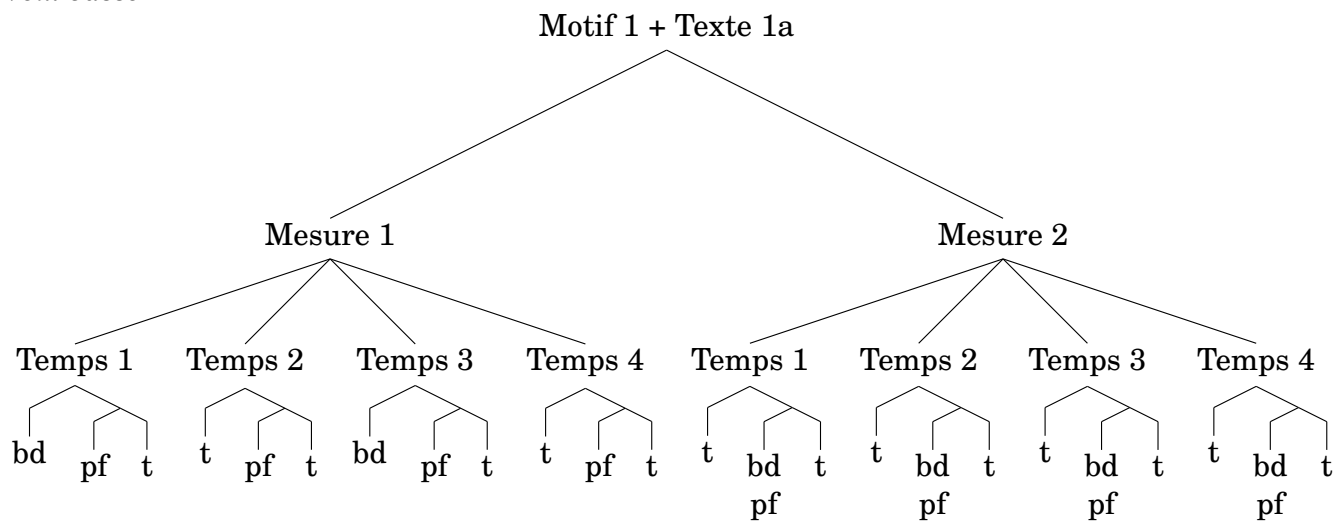
Séparation des voix

Motif 1 + Texte 1a

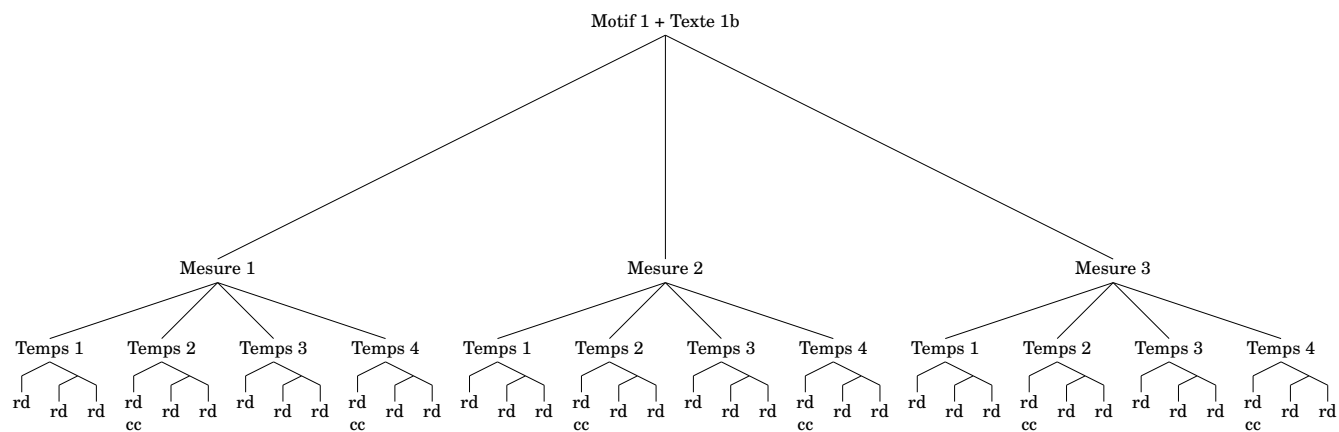
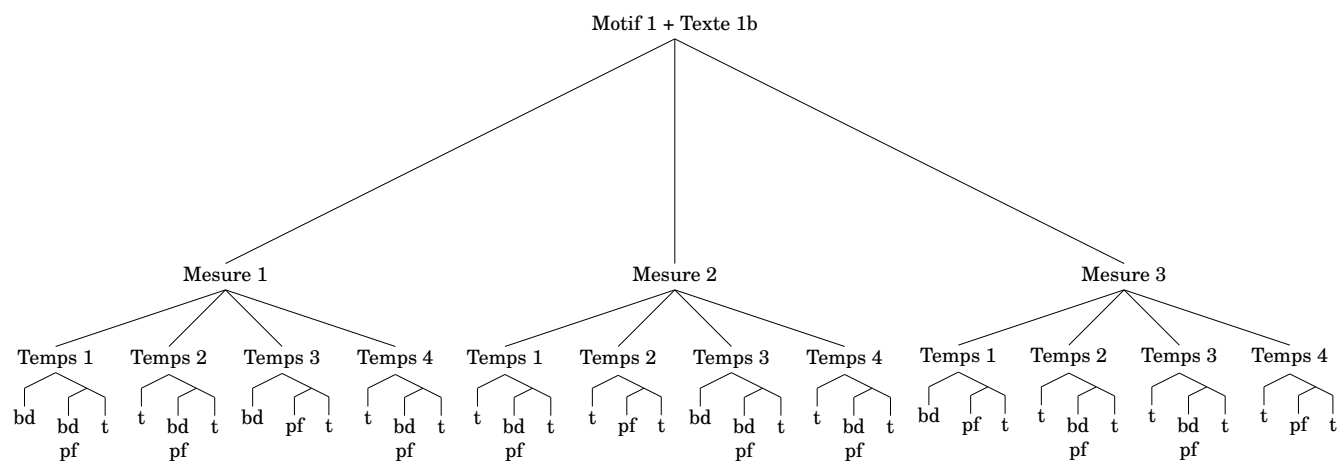
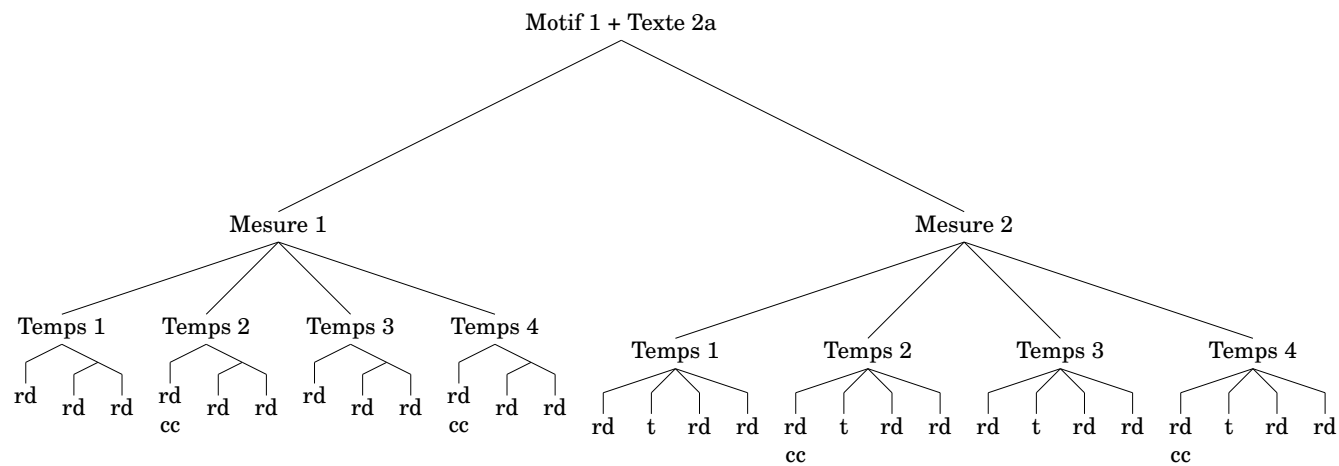
Voix haute

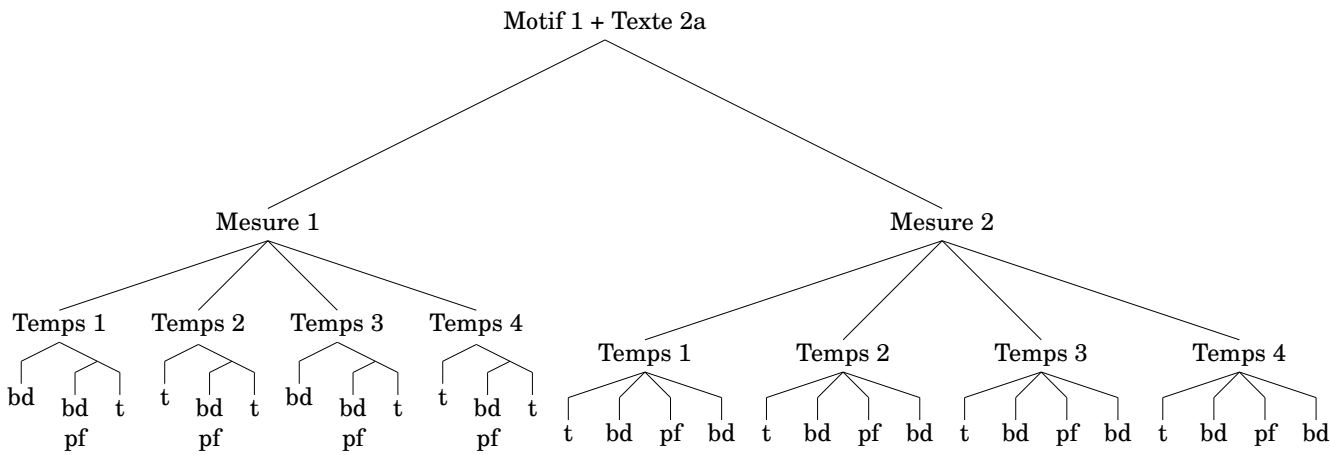


Voix basse



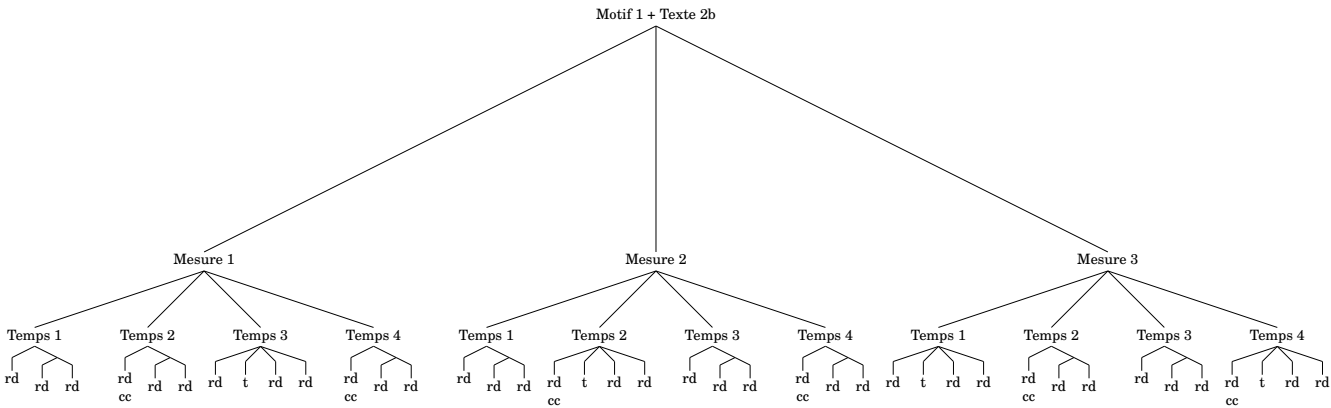
Motif 1 + Texte 1b

Voix haute*Voix basse***Motif 1 + Texte 2a***Voix haute**Voix basse*

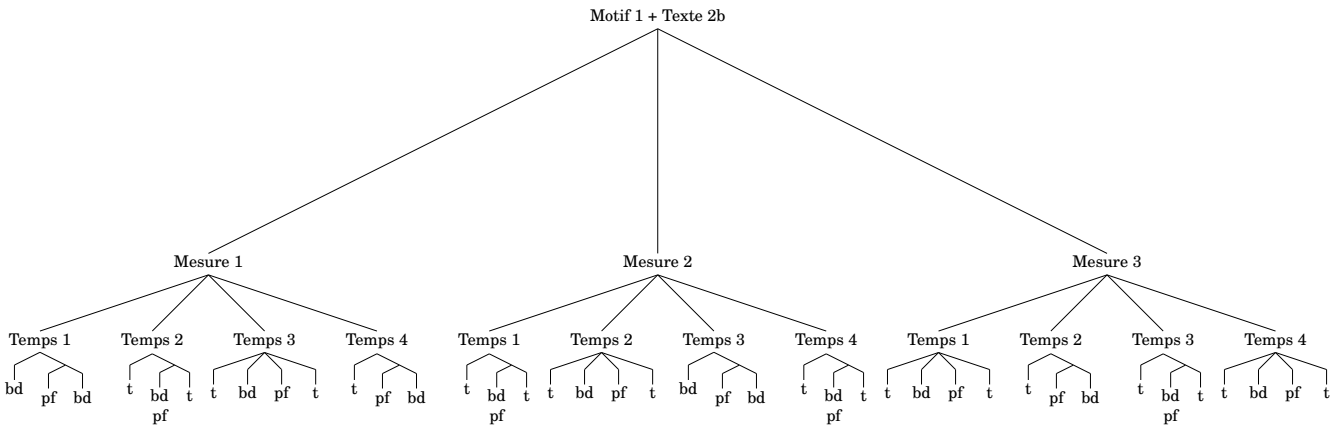


Motif 1 + Texte 2b

Voix haute

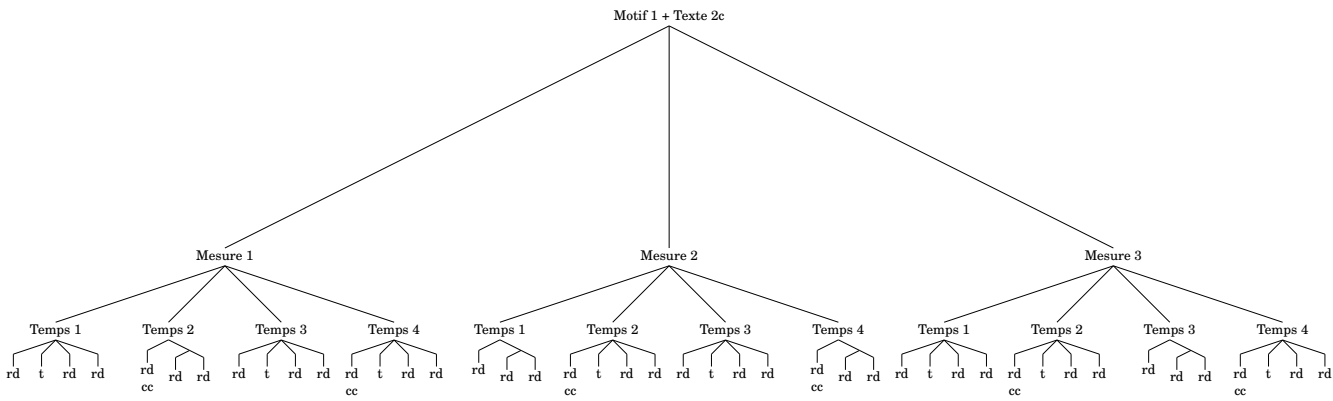


Voix basse

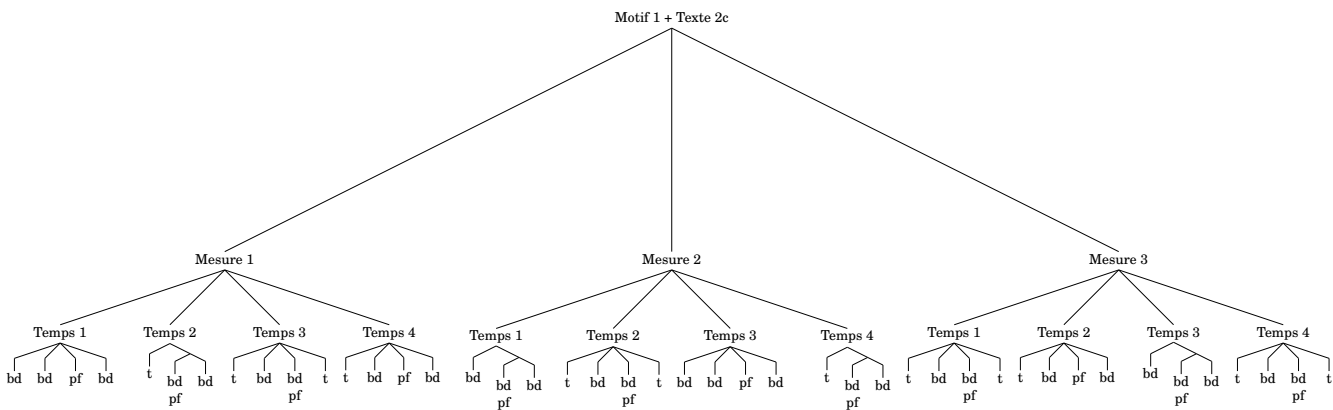


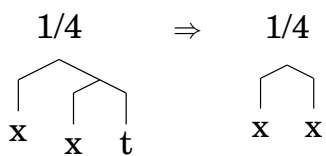
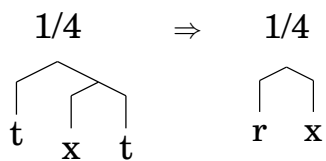
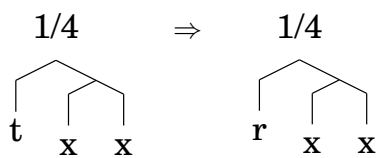
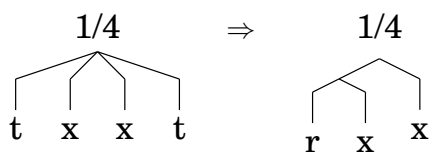
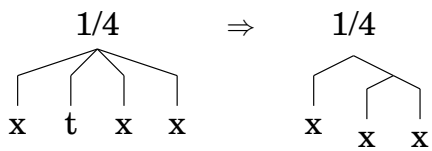
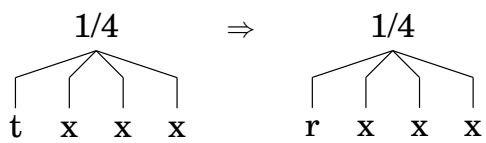
Motif 1 + Texte 2c

Voix haute



Voix basse



Règles de réécriture pour le 4/4 binaire

3.2.2 Expérience 2

Partition de référence pour l'output

The image displays a musical score for three staves, enclosed in a black rectangular frame. The top staff is in C major, indicated by a C-clef and a key signature of one sharp (F#). It contains a sequence of notes with various articulations, including slurs and accents, and is marked with a '3' above several groups of notes. The middle staff is in C major, indicated by a C-clef and a key signature of one sharp (F#). It contains a sequence of notes with various articulations, including slurs and accents, and is marked with a '3' above several groups of notes. The bottom staff is in C major, indicated by a C-clef and a key signature of one sharp (F#). It contains a sequence of notes with various articulations, including slurs and accents, and is marked with a '3' above several groups of notes. The score is a reference partition for the output.

En cours...

3.3 Conclusion

Conclusion de ce chapitre.

RÉSULTATS

Sommaire

4.1	Introduction	35
4.2	Contenu	35
4.3	Conclusion	35

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	37
5.2	Contenu	37
5.3	Conclusion	37

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.

