

Apprentissage des structures temporelles hiérarchiques pour l'improvisation guidée

Tissot Priscilla

Index Terms—Musique, Improvisation, Grammaire hors-contexte, Structure multi-niveaux ...

1 OBJECTIF

L'objectif de cette recherche est de générer des improvisations de musique de jazz. Afin d'atteindre cet objectif, l'étude s'est déroulée en trois grandes phases consécutives.

- 1) La modélisation de la structure temporelle hiérarchique d'une musique de Jazz.
- 2) La mise en place d'une méthode d'apprentissage automatique d'une grammaire hors-contexte.
- 3) La génération d'improvisations à partir de cette grammaire.

2 DONNÉES UTILISÉES

Les données qui sont étudiées pour générer une grammaire sont issues du corpus *Omnibook*[1]. Peu d'informations ont été données concernant l'origine de ce corpus et le traitement de ces données. Concernant l'origine, ce corpus est disponible sur le lien suivant[2].

Ces fichiers sont des partitions de musique de Jazz qui ont été mises au format musicXML. Ce sont des fichiers xml qui ont des balises spécifiques au format musicXML. De nombreuses informations sont donc présentes. Celles qui ont été utilisées pour reproduire l'étude sont les balises `<harmony>` et `<note>` qui sont contenues dans des `<measure>` tel que

`<measure>`

```
...
<harmony print-frame="no">
  <root>
    <root-step>B</root-step>
    <root-alter>-1</root-alter>
  </root>
  <kind>major</kind>
</harmony>
<note default-x="73.12" default-y="-30.00">
  <pitch>
    <step>G</step>
    <octave>4</octave>
  </pitch>
  <duration>120</duration>
  <voice>1</voice>
  <type>quarter</type>
  <stem>up</stem>
</note>
...
```

```
<harmony ... >
<note ... >
<note ... >
</measure>
```

Dans la balise `harmony` nous recensons la balise `root-step` qui nous permet de connaître les accords utilisés. De plus, les balises `note` en dessous des balises `harmony` contiennent chacune une balise `duration` qui nous permet d'obtenir la durée d'une note. La durée des notes cumulées permet de nous donner la durée de l'accord.

Dans cette étude, les accords ne sont pas utilisés directement. Ce sont des degrés qui sont traités. En effet, les accords en eux-mêmes ne sont pas importants. Ce qui est le plus important c'est la notion de changement de rythme qui est perceptible à l'aide des degrés. Les degrés correspondent à un degré de variation à partir d'une note donnée. Par exemple, si la note "de référence" est un DO alors son degré est de 1, RE de 2, MI de 3 ... comme représenté sur la figure 1. Ces degrés permettent de définir ce qu'on appelle *la tonalité*. Dans cet exemple, nous pouvons donc dire que la tonalité est en DO majeur.

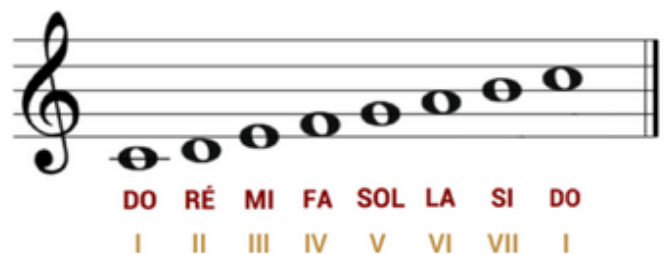


Fig. 1. Exemple des degrés

Cette variation sous forme de degré est intéressante pour notre apprentissage puisque certains accords peuvent effectuer les mêmes variations dans différentes musiques, mais en se basant sur des tonalités différentes (en ayant donc une note "de référence" différente).

Pour convertir les accords en degrés, j'ai considéré le premier accord de la partition comme étant de degré 1 puis j'ai calculé les degrés des autres accords selon ce premier accord, tout au long de la musique.

De plus, seulement les 32 premières mesures sont prises en compte pour l'apprentissage.

3 METHODES

3.1 Modélisation d'une structure temporelle hiérarchique

Afin de modéliser une structure temporelle hiérarchique, une analyse a été réalisée avec *Pascal Mabit*, un musicien professionnel. Cette phase est une analyse théorique de la musique. En pratique, d'un point de vue informatique, cette phase n'est donc pas reproductible. Mais par la suite, nous pourrions nous baser sur les informations fournies dans cette partie pour obtenir une grammaire hors-contexte. En effet, la grammaire représentée sur la Figure 2 a été mise en place dans cette partie.

Grammar 2 *Rhythm changes* phrase structure grammar

- 1: $RhythmChanges \rightarrow A_1 A B A$
- 2: $A_1 \rightarrow \tau_I \tau \sigma \tau$
- 3: $A_2 \rightarrow \tau_I \tau \sigma \omega$
- 4: $A \rightarrow A_1 \mid A_2$
- 5: $B \rightarrow \delta_{III} \delta_{VI} \delta_{II} \delta_V$
- 6: $\tau_I, \tau, \sigma, \omega, \delta_{III}, \delta_{VI}, \delta_{II}, \delta_V$ are sets of harmonic fun

Fig. 2. Grammaire hors-contexte pour les changements de rythme

Dans un premier temps, AABA est donné. Ces structure sont des sections de huit mesures. Les sections A ont des accords changeant rapidement avec généralement deux accords par mesure et restant proches de la tonalité initiale. Alors que la section B a une progression plus lente où chaque accord est généralement joué sur deux mesures, et avec une sensation de changement tonal.

De ces sections, en découle des fonctions tonales de deux mesures chacune. Ce sont les degrés de chaque fonctions tonales qui seront appris à partir du corpus.

3.2 Apprentissage de grammaire

Afin d'associer les degrés aux fonctions tonales, une méthode de sélection de séquences de mots partageant une information mutuelle a été utilisée. Cette méthode est issue de travaux de Zitouni et al. [3]. De cette étude, la fonction J est mise en exergue.

$$J = \log(count(a, b) * T / (count(a) * count(b)))$$

Avec a et b deux symboles consécutifs, $count$ une fonction de comptage et T la taille du corpus.

Or les degrés peu fréquents génèrent un agglomérat d'association avec cette fonction. Pour palier à ce problème, une normalisation est effectuée, ce qui donne la fonction

$$J = \log(count(a, b) * T / (count(a) * count(b)) / (l(a) + l(b)))$$

où 1 correspond à la durée du symbole.

Ensuite, un alphabet trop grand est généré. certains symboles ayant la même fonction tonale peuvent être regroupés. Pour cela, une nouvelle fonction est mise en place.

Avec :

- 1) K la taille du vocabulaire.
- 2) u la liste des symboles à gauche de a et b
- 3) v la liste des symboles à droite de a et b

$$\Psi(a, b) = \frac{1}{K} \sum_{u,v} (J(u, a) - J(u, b))^2 + (J(a, v) - J(b, v))^2 \leq \xi$$

Fig. 3. Fonction de regroupement de symboles

- 4) ϵ un seuil à déterminer expérimentalement en fonction du corpus.

Avec ces fonctions, l'algorithme Figure 4 est réalisé.

Algorithme 5 Induction de grammaire sur un corpus de scenario

- Entrée :** Corpus de scénarios.
Sortie : Liste de règles de réécriture.
- 1: **Répéter**
 - 2: Trouver a et b tels que $\tilde{J}(a, b) = \max_{x,y} \tilde{J}(x, y)$.
 - 3: Créer la règle de réécriture $X_{ab} \rightarrow a + b$.
 - 4: $l(X_{ab}) \leftarrow l(a) + l(b)$.
 - 5: Remplacer toutes les occurrences de $a + b$ par X_{ab} dans le corpus.
 - 6: **Si** \exists un symbole Y tel que $l(Y) = l(X_{ab})$ **et** $\Psi(Y, X_{ab}) < \xi$ **alors**
 - 7: Créer la règle de réécriture $Y \rightarrow X_{ab}$.
 - 8: Remplacer toutes les occurrences de X_{ab} par Y dans le corpus.
 - 9: **Fin Si**

Fig. 4. Algorithme de génération de grammaire

Une fois que la grammaire est entièrement générée, les improvisations peuvent être réalisées.

3.3 Improvisation sur un scénario temporel

L'improvisation se réalise en 2 étapes. L'anticipation et la navigation.

3.3.1 Anticipation

Premièrement nous regardons dans la grammaire si la position courante correspond à un préfixe d'une règle. Puis, on

$$Futur(T) = \{j \in [0...m] \mid \exists c_{futur} \in \mathbb{N}, \Lambda_j \dots \Lambda_{j+c_{futur}-1} \in Pref(S_T \dots S_s)\}.$$

construit l'ensemble des états de la grammaire partageant un passé commun par rapport à la position courante. Ces

$$Passé(i) = \{j \in [0...m] \mid \exists c_{passé} \in [1, j], \Lambda_{j-c_{passé}+1} \dots \Lambda_j \in Suff(0...i)\}.$$

étapes sont nécessaires donc l'anticipation se traduit par la recherche de positions j tel que la mise en place de futur et de passé est possible.

$$Ant(T, i) = \{j \in [0...m] \mid j \in Futur(T) \wedge j - 1 \in Passé(i)\}.$$

3.3.2 Navigation

L'étape de navigation consiste à chercher des événements dans la grammaire partageant un contexte commun avec la position courante tout en respectant l'architecture de l'arbre de dérivation de l'improvisation. Cette étape permet de générer des variations locales.

$$\text{Nav}(T, i) = \{j \in [0...m] \mid \Lambda_j = S_T \wedge j - 1 \in \text{Passé}(i)\} .$$

4 RÉSULTATS

Mon programme est disponible sur Github[4]. Je n'ai pas eu le temps de générer des improvisations. En revanche, j'ai pu réaliser des arbres de dérivation. Nous pouvons comparer les résultats que j'ai obtenu (Fig 5) avec ceux de l'article (Fig 6). Les arbres de dérivations correspondent à la musique Celerity.

Regles de dérivations obtenues

A \rightarrow 3 6

D \rightarrow 2 5

E \rightarrow 1 4

G \rightarrow 1 6

B \rightarrow A 6

F \rightarrow 5 E

K \rightarrow A D

J \rightarrow 1 D

L \rightarrow 2 D

M \rightarrow G D

C \rightarrow 3 B

H \rightarrow F 4

I \rightarrow C 2

N \rightarrow H 1

O \rightarrow N 1

Nous pouvons observer des différences dès la première feuille de l'arbre. En effet, J correspond à τ_1 . Nous pouvons voir dans l'arbre de dérivation du groupe de recherche qu'il y a I I dans la première mesure puis II IV alors que j'obtiens I | II IV. De plus, en observant le début de la partition sur la figure 7, nous pouvons observer qu'il y a bien un seul accord dans la première mesure et donc un seul I et non deux.

Ensuite,

- K correspond bien à τ
- Pour $F=\sigma$, V, I et IV sont mal découpés, et là encore il y a une répétition du IV dans les résultats des chercheurs.
- $J=\tau$

Pour A2 :

- $J=\tau_1$
- $K=\tau$
- $F=\sigma$ qui est de nouveau mal découpé.
- $O=\omega$ qui est aussi mal découpé.

Dans la partie B, tous les degrés se doublent donc l'arbre dérivé de mon côté ne correspond pas pour les noeuds I et L.

Pour A \rightarrow A1 :

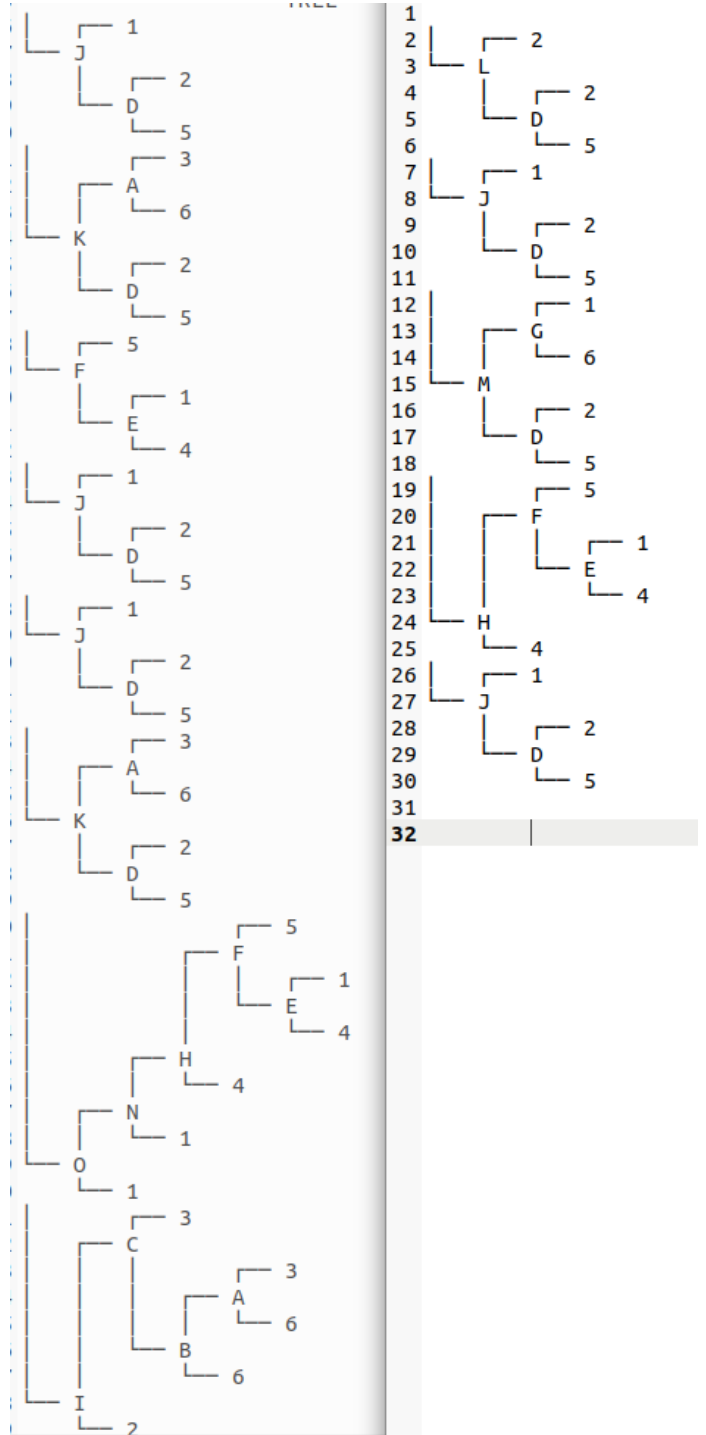
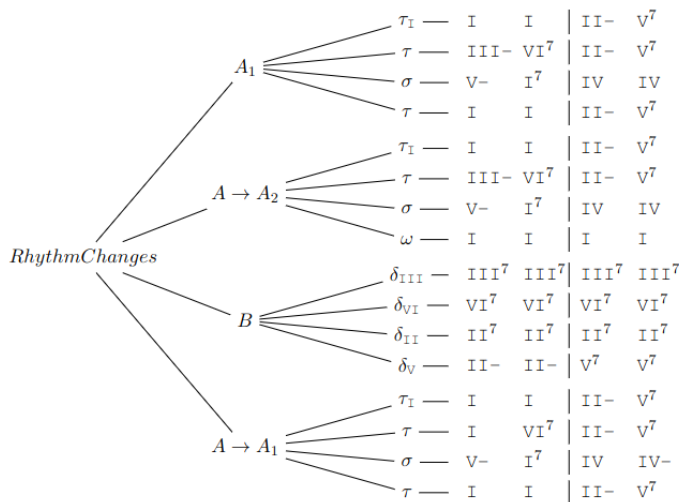


Fig. 5. Arbre de dérivation pour la musique Celerity

- $J=\tau_1$
- $M=\tau$
- $H=\sigma$ qui est mal découpé
- $J=\tau$

5 VALIDATION

Afin d'évaluer leurs recherches, des résultats ont été présentés à trois musiciens. En effet, les données qualitatives sont plus compliquées à évaluer que des données



quantitatives. En revanche, il existe tout de même des méthodes d'évaluations qui permettent d'éviter différents biais. Dans ce cas, nous pouvons dire que trois personnes ne représentent pas un échantillon assez important pour représenter l'ensemble des musiciens de Jazz et encore moins la population en général.

Pour obtenir des variables quantitatives il est possible de réaliser des questionnaires dans lesquels des personnes peuvent noter une caractéristique précise selon la question posée. Le questionnaire peut être présenté avec les résultats d'évaluations pour connaître les caractéristiques prises en compte lors de l'évaluation.

De plus, l'échantillon de personnes sondées doit être randomisé afin d'éviter les biais. Et cet échantillon doit être assez grand. Pour effectuer un intervalle de confiance il faut au moins 30 personnes avec des variables quantitatives. Et pour des variables qualitatives il faut que $NP \geq 5$ et $N(1-P) \geq 5$ tel que N est le nombre de personnes et P une probabilité.

Les questionnaires peuvent être posés sur des improvisations humaines et générées par ordinateur afin d'avoir des cas témoins et des cas expérimentaux. Pour déterminer si les résultats sont les mêmes, avec des variables quantitatives nous pouvons faire des tests d'indépendance linéaire et avec des variables qualitatives des tests de Chi 2.

6 CONCLUSION

Cette étude comporte trois grandes étapes. Mes résultats s'arrêtent à l'étape 2 qui comporte des différences avec les résultats des recherches. En effet, lorsqu'une mesure a un nombre impair d'accords ils ajoutent un accord pour

équilibrer l'arbre. Or je n'ai pas vu d'informations concernant l'ajout d'accord. Des questions peuvent se poser concernant cet ajout d'accord. En effet, pour chaque accord un temps doit être donné. Il est nécessaire pour la deuxième fonction J et donc selon le cas, ce temps utilisé pour la normalisation peut faire J_{max} dans l'algorithme de génération de grammaire. Ainsi, son ordre de traitement peut différer.

Concernant l'évaluation des résultats, les études sont insuffisantes. Une démarche plus poussée est nécessaire pour apprécier la qualité des résultats.

REFERENCES

- [1] Ken Déguernel, Emmanuel Vincent, and Gérard Assayag. *Using Multidimensional Sequences for Improvisation in the OMax Paradigm*, in Proceedings of the 13th Sound and Music Computing Conference, 2016.
- [2] <https://homepages.loria.fr/evincent/omnibook/>
- [3] Zitouni, I., K. Smaili, and J.-P. Haton. 2000. "Beyond the conventional statistical language models: the variable-length sequences approach." In Proceedings of Interspeech. pp. 562–565.
- [4] Dépot git : <https://github.com/Tpris/Apprentissage-de-structures-musicales>