

## 1 Introduction

Les morceaux de musique sont structurés à différentes échelles de temps (battements, temps, accords, mesure, phrase, section etc.). L'estimation de la structure consiste à décrire l'organisation des éléments qui constituent un morceau de musique à un niveau macro-temporel. Ce problème a de nombreuses applications, dont l'indexation dans de grandes bases de données, la classification des musiques par genre ou la génération de résumés audio, et a attiré l'attention de nombreux chercheurs ces dernières années [1]. L'estimation de la structure d'un morceau se fait souvent en deux étapes. On extrait d'abord une représentation moyen-niveau du signal appelée *matrice d'auto-similarité* (en anglais *self-similarity matrix*, SSM) [2], puis on cherche à en extraire une structure macro-temporelle, soit en se basant sur les répétitions, soit sur les changements dans le signal que cette représentation fait apparaître. Un exemple de matrice d'auto-similarité est donné figure 1.

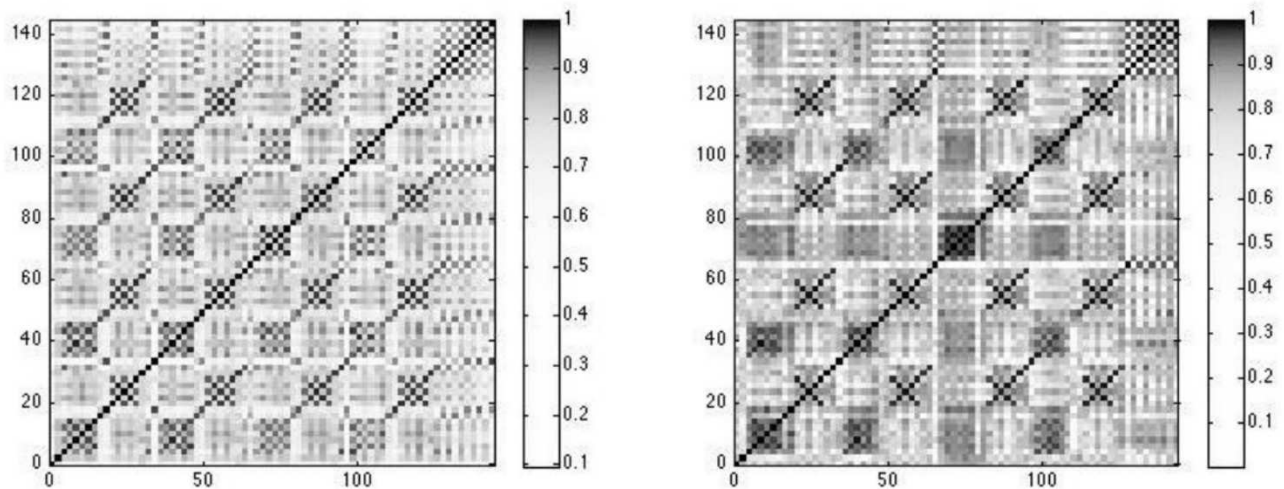


Figure 1: Matrice d'auto-similarité pour le morceau *Drive my car* des Beatles, obtenue à gauche à partir des chromas et à droite des MFCC. Les axes X et Y sont le temps en secondes.

Les étapes principales sont :

- On commence par extraire une représentation du signal "moyen-niveau" qui nous apporte une certaine information d'intérêt. On appelle ça extraire des descripteurs du signal ou des "features". Par exemple, on va regarder comment évolue l'harmonie au cours du morceau. Pour ça on va calculer à des instants successifs du morceau un vecteur qui va indiquer l'importance des 12 demi-tons de la gamme chromatique (on appelle ça un *vecteur de chroma*) et ainsi obtenir une représentation qui décrit de manière grossière l'évolution de l'harmonie et la mélodie au cours du temps (appelée *chromagramme*)
- On rappelle que le but est de trouver quels sont les segments temporels dans le morceau qui sont similaires ou différents. Pour ça on calcule la similarité entre chaque vecteurs du chromagramme (par exemple en prenant le produit scalaire). On obtient un matrice d'auto-similarité qui permet de visualiser la structure (diagonales indiquant des répétitions, blocs indiquant des changements).
- Une fois que l'on a obtenu cette matrice, on cherche à en extraire de manière automatique ces informations sur les répétitions et les blocs pour en déduire la structure du morceau.

*Remarque* : En lisant les articles sur le sujet, on se rendra compte qu'il est difficile de donner une définition précise de la notion de structure. Il y a plusieurs manières de voir les choses pour décomposer le morceau à partir de la matrice d'auto-similarité en différents segments qui ont un sens sémantique : on peut soit chercher des points de rupture (par exemple un changement dans l'instrumentation), soit essayer de trouver des motifs, ou des segments qui se répètent.

## 2 Travail demandé

Le but de ce projet est de réaliser une application sous Android qui permet de visualiser une telle structure sous forme de matrice d'auto-similarité. Cette représentation visuelle permettra de mettre en évidence dans un morceau de musique une décomposition en plusieurs parties qui ont du "sens" musicalement.

Pour extraire les descripteurs du signal on pourra utiliser les toolbox suivantes :

- Pour des descripteurs liés au timbre, on utilisera les MFCC de l'*AuditoryToolbox* <https://engineering.purdue.edu/~malcolm/interval/1998-010/>
- Pour des descripteurs liés à l'harmonie, on utilisera les chroma de la *Chroma Toolbox* <http://www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/>
- On pourra éventuellement ajouter des descripteurs liés au rythme en utilisant la *TemporalToolbox* <http://www.mpi-inf.mpg.de/resources/MIR/tempogramtoolbox/>. Cependant on laissera ce descripteur de côté tant que le projet avec les deux autres descripteurs n'aura pas été réalisé.

*Remarque*: On essaiera de bien comprendre comment sont implémentés les descripteurs, l'influence des différents paramètres (taille de fenêtre utilisée, pas de temps, fréquences considérées etc.).

On construira ensuite une matrice d'auto-similarité  $SSM$  ainsi que proposé dans l'article [2] de la manière suivante :

$$SSM_{i,j} = s(f_i, f_j) \quad (1)$$

Où  $f_k, k \in [1 : N]$  est une séquence de  $N$  descripteurs calculés successivement au cours du temps (le chromagramme par exemple)  $s$  est une fonction de similarité (par exemple produit scalaire entre deux vecteurs de norme unitaire),  $f_i$  and  $f_j$  sont les descripteurs aux instants  $i$  et  $j$ .

On créera une application qui permet de voir un curseur qui parcourt la ou les matrice(s) d'auto-similarité en même temps que le morceau est joué. On pourra imaginer plusieurs manières de visualiser la structure, par exemple par superposition de plusieurs matrices de similarité avec différentes couleurs etc. On pourra aussi donner à l'utilisateur la possibilité de modifier des paramètres d'analyse.

## 3 Morceaux

On donne en pièce jointe de l'énoncé du projet 2 morceaux des Beatles avec la structure annotée à la main.

## References

- [1] Paulus, J. and Müller, M. and Klapuri, A., Audio Based Music Structure Analysis, In ISMIR, 2010.
- [2] Foote, J., Visualizing Music and Audio using Self-Similarity, In ACM Multimedia, pages 77–80, 1999.