Dans le cadre de ce projet, nous avons effectué une analyse audio et une analyse d’image.

# Analyse audio

L’objectif de l’analyse audio est d’identifier la séquence musicale de l’introduction et du générique de fin d’une émission télévisée. En effet, l’émission est divisée en quatre sections :

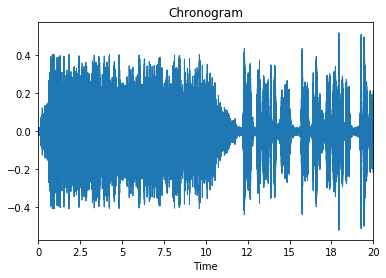
* Introduction : musique
* Reportage : voix
* Débat : voix
* Générique de fin : musique

Nous avons choisi de définir un classifieur entre deux catégories : voix et musique.

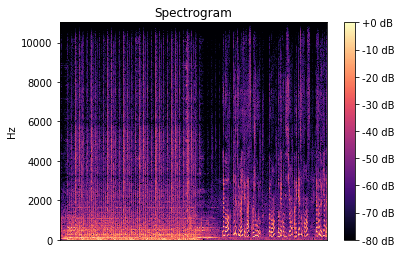
Pour cela nous avons récupéré des morceaux de musique sur internet[[1]](#footnote-1) et un jeu de données de voix basé sur l’émission puis avons construit des observations de 5 secondes en Python en utilisant la librairie *MoviePy*.

Par la suite, nous avons extrait des *features* en utilisant la libraire *Librosa*.

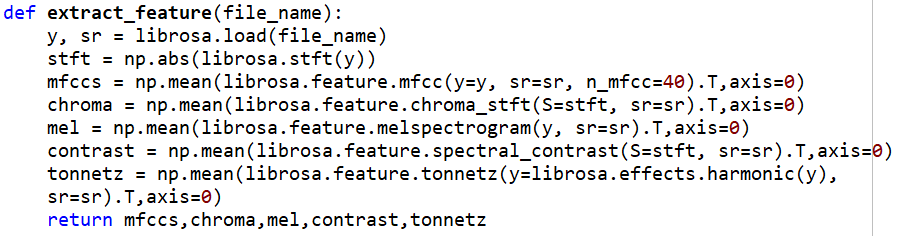
En affichant, le chronogramme des 20 premières secondes de l’émission (10 secondes de musique puis 10 secondes de voix), nous observons des différences entre ces 2 types de classes audio.



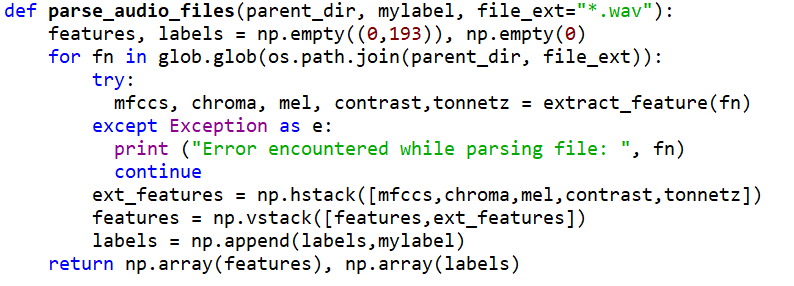
De même avec le spectrogramme.



La librairie open source *Librosa* fournit plusieurs méthodes d’extraction de *features* de piste audio : *melspectrogram, mfcc et chroma-stft.* Ces méthodes ont d’ailleurs été présentées dans le résumé de cours sur l’analyse audio.



Pour chaque piste audio, nous définissons ainsi les *features* et le label associé (voix ou musique).

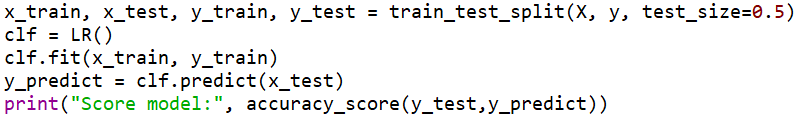


On obtient alors un *dataset* de 200 observations ayant chacune 193 *features*.

Dans le domaine de la classification supervisée, il existe de nombreuses méthodes : *KNN, Naive Bayes, LDA/QDA, Logistic, SVM, Neural Network*…

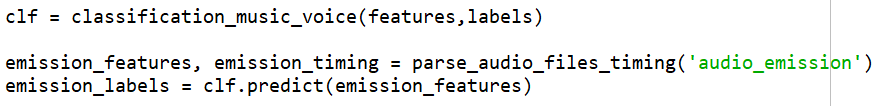
Pour notre étude, nous aurions pu tester l’implémentation d’un neural network en *Tensorflow*. Cette méthode est en effet devenue très populaires pour le traitement de signal et nous avons trouvé beaucoup d’articles sur ce sujet.

Le problème de classification étant assez simple, nous avons opté pour un modèle de régression logistique.



Avec ce modèle, nous obtenons un score de précision de 100%.

Il ne nous reste maintenant plus qu’à appliquer notre classifieur sur la piste audio de l’émission pour identifier les séquences musicales d’introduction et de générique de fin.



Par soucis d’affichage, nous n’avons représenté que les 20 premières et 20 dernières secondes de l’émission. Le classifieur donne l’introduction musicale sur les 10 premières secondes et le générique de fin à partir de la 2342 ème seconde.

1. Source : http://freemusicarchive.org [↑](#footnote-ref-1)