

Classification du genre musical par apprentissage machine

Noé Aubin-Cadot

20 février 2020

But

But : Déterminer ce qu'est du *v a p o r w a v e*.

Idée : Utiliser l'apprentissage machine pour classifier le genre musical d'un morceau de musique puis évaluer le classificateur sur des morceaux de *v a p o r w a v e*.

Il existe plusieurs méthodes pour établir le genre musical.

Approche naïve : Faire de l'apprentissage machine sur l'amplitude spectrale moyenne de la quatrième seconde d'un morceau de musique.

Plan

Plan :

1. Trouver des données.
2. Préparer les données.
3. Visualiser les données.
4. Apprentissage machine sur les données.
5. Déterminer ce qu'est du *vaporwave*.

Trouver des données

On veut des données qui contiennent :

- *source* \mathbf{X} = spectre moyen de la quatrième seconde.
- *but* \mathbf{y} = genre musical.

Le spectre moyen de la quatrième seconde peut être extrait des morceaux d'une librairie *iTunes* via un script Python.

On considère 1069 morceaux dans les trois genres suivants :

- *classique* (348 morceaux)
- *électronique* (363 morceaux)
- *metal* (358 morceaux)

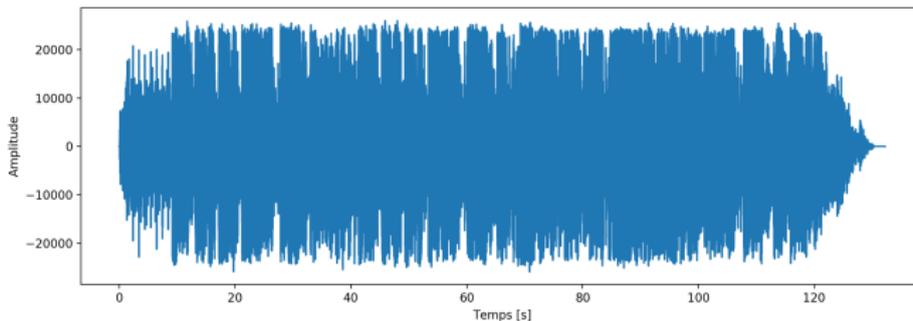
Préparer les données

En étapes :

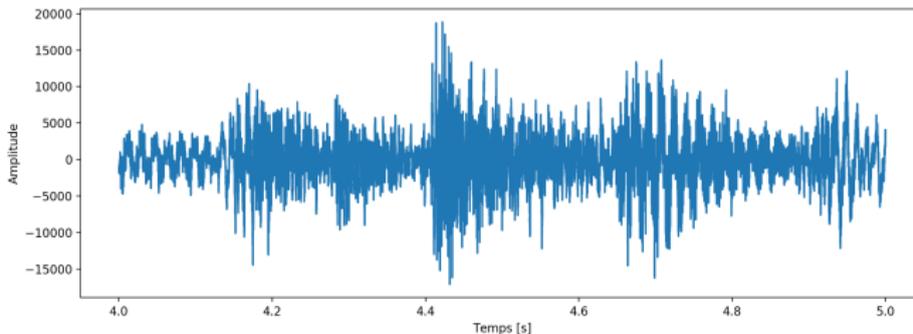
1. Extraction de l'onde sonore à la quatrième seconde.
2. Transformée de Fourier avec fenêtres de 1024 points.
3. Couper le spectre f à $< 20\text{kHz}$.
4. Passer aux décibels via $10 \log_{10}(f)$.
5. Moyenne de l'amplitude spectrale sur une seconde.
6. Normaliser les amplitudes spectrales moyennes.
7. Ajuster les amplitudes normalisées (optionnel).

Préparer les données

Onde sonore d'un morceau de musique :

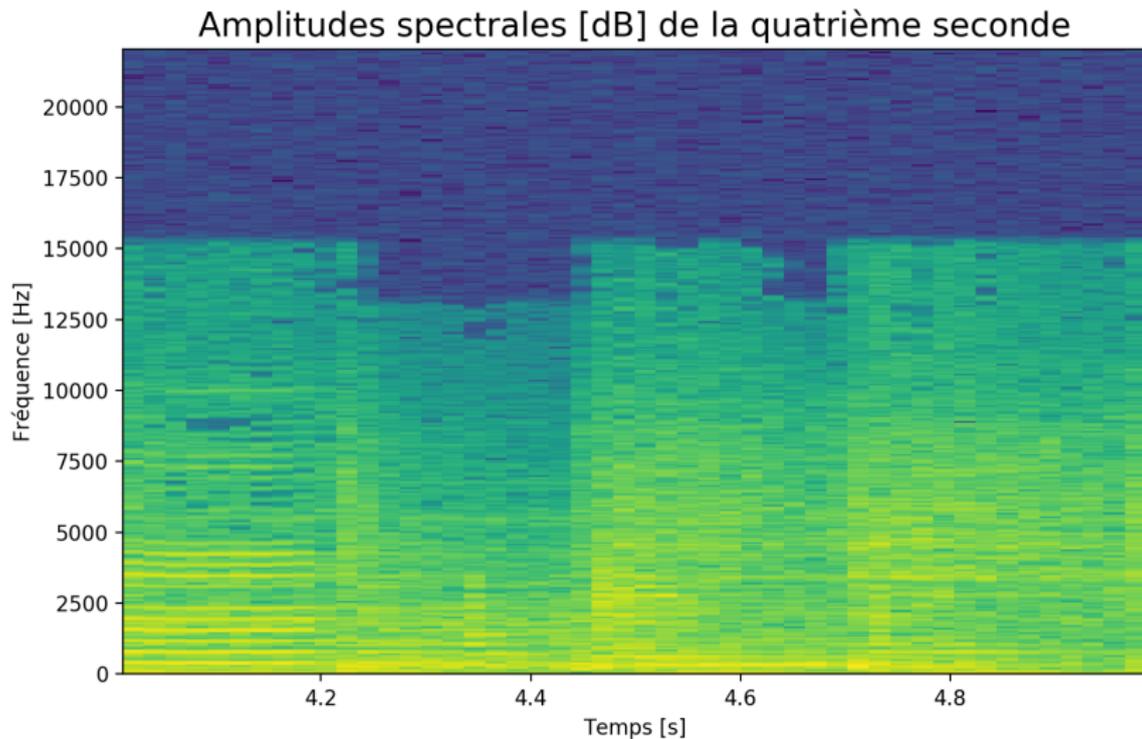


Extraction de la quatrième seconde :



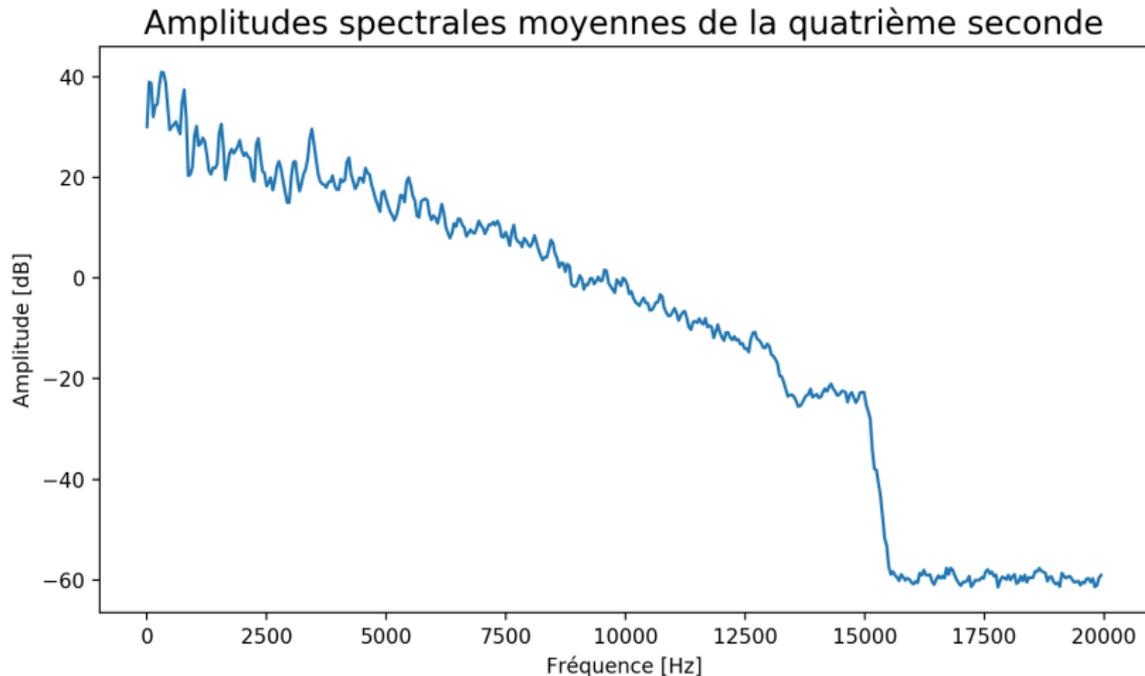
Préparer les données

Transformée de Fourier avec fenêtres de 1024 points :



Préparer les données

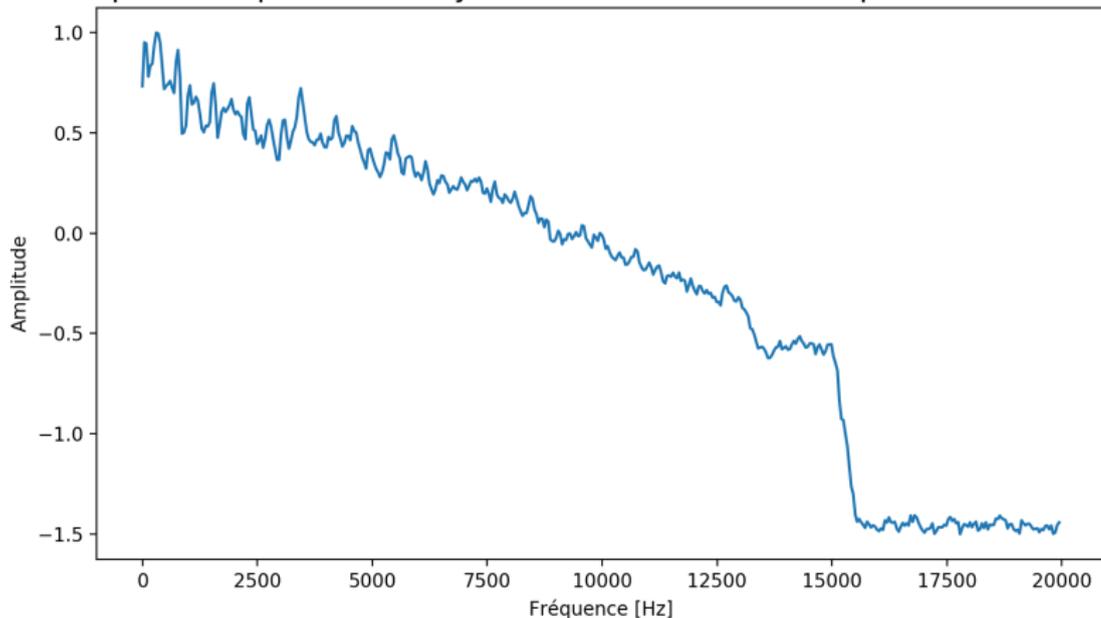
Amplitudes spectrales moyennes de la quatrième seconde :



Préparer les données

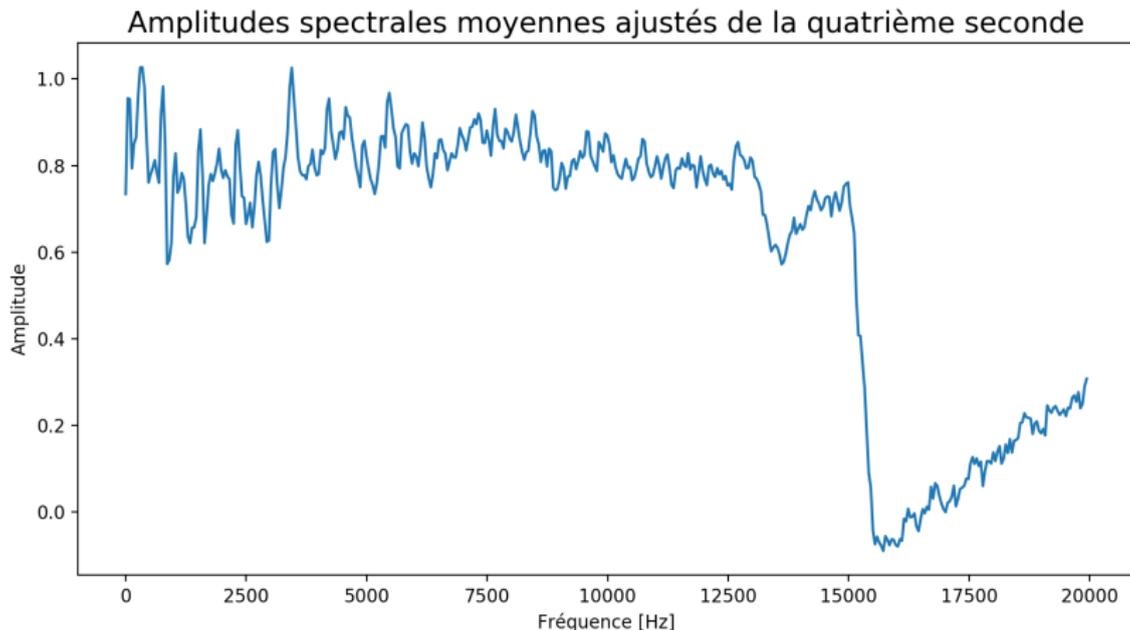
Amplitudes spectrales normalisées de la quatrième seconde :

Amplitudes spectrales moyennes normalisées de la quatrième seconde



Préparer les données

Amplitudes spectrales ajustées de la quatrième seconde :



Préparer les données

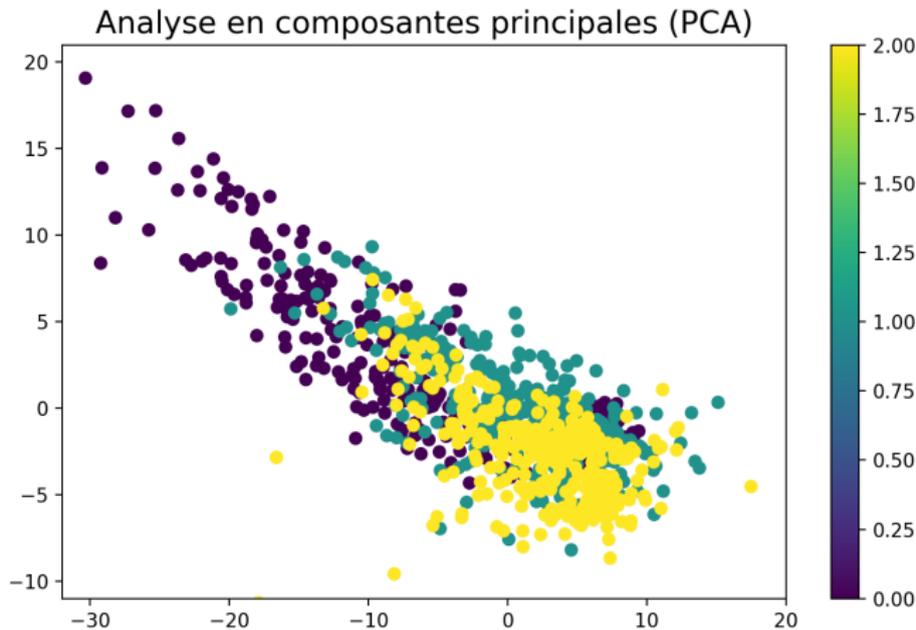
Il existe plusieurs manières plus ou moins *ad hoc* de choisir et de préparer les amplitudes en vue d'y faire de l'apprentissage machine.

Gardons en tête deux manières différentes :

1. Amplitudes spectrales de la quatrième seconde, normalisées mais non ajustées.
2. Amplitudes spectrales de la douzième seconde, normalisées et ajustées.

Visualiser les données

Analyse en composantes principales (PCA) des amplitudes spectrales de la douzième seconde, normalisées et ajustées :



classique, électronique, métal.

Apprentissage machine sur les données

On scinde les données (\mathbf{X}, \mathbf{y}) en deux sous-ensembles :

- 75% : *entraînement* $(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}})$, 801 lignes.
- 25% : *test* $(\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}})$, 268 lignes.

On entraîne un classificateur sur les données d'entraînement et on évalue ses résultats sur les données de test.

On peut essayer divers classificateurs `scikit-learn` e.g. KNN, BNG, BNB, SVM, lbfgs, liblinear, RFC, Perceptron, SGDC, DTC, etc.

Apprentissage machine sur les données

Scores d'apprentissage pour la 4^{ème} sec. (non ajustée) vs la 12^{ème} seconde (ajustée) :

Nom	Train	Test
KNN	100.0%	79.8%
BNG	35.6%	30.7%
BNB	44.8%	41.2%
SVM	86.8%	80.5%
lbf	77.8%	70.8%
lib	82.6%	73.0%
RFC	98.6%	74.9%
Per	43.2%	47.9%
SGD	68.3%	61.0%
DTC	100.0%	64.8%

Scores pour la 4^{ème} sec.

Nom	Train	Test
KNN	100.0%	75.7%
BNG	40.6%	42.5%
BNB	45.7%	50.4%
SVM	86.8%	81.0%
lbf	82.4%	76.9%
lib	85.8%	75.0%
RFC	98.8%	78.4%
Per	65.7%	58.2%
SGD	66.9%	59.3%
DTC	100.0%	69.8%

Scores pour la 12^{ème} sec.

Apprentissage machine sur les données

Matrice de confusion $(i, j) = (\text{réel}, \text{prédit})$ pour classificateur SVM sur la 12ième seconde ajustée :

	<i>classique</i>	électronique	metal
<i>classique</i>	92	6	3
électronique	11	54	11
metal	11	9	71

Il y a **51** mauvaises classifications sur 268 prédictions.

Déterminer ce qu'est du *v a p o r w a v e*.

On peut enfin déterminer ce qu'est du *v a p o r w a v e*.

On entraîne le classificateur SVM sur les 1069 morceaux de départ et on l'évalue sur 47 morceaux de *v a p o r w a v e*.

Déterminer ce qu'est du *v a p o r w a v e*.

Selon la 4ième seconde (non ajustée), le classificateur classe les 47 morceaux de *v a p o r w a v e* en :

- 7 morceaux de *classique*
- 26 morceaux d'électronique
- 14 morceaux de metal

Selon la 12ième seconde (ajustée), ils sont classifiés en :

- 9 morceaux de *classique*
- 16 morceaux d'électronique
- 22 morceaux de metal

Déterminer ce qu'est du *vaporwave*.

Donc, en moyenne, le *vaporwave* c'est :

- 17.0% de *classique*
- 44.7% d'électronique
- 38.3% de metal

Merci de votre attention 😊