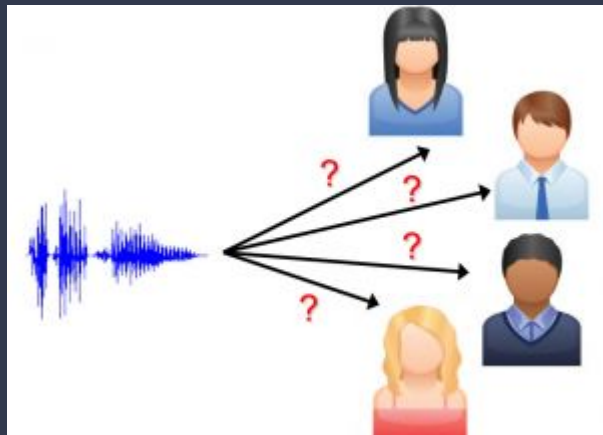


Speaker identification



Sommaire

1. Constitution du dataset
2. Pré-traitement des données
3. Entraînement et évaluation du modèle
4. Limites
5. Retour d'expérience

Dataset

3 classes :

- Aissa : 5 audios = 2800 frames
- Marouan : 5 audios = 2800 frames
- Autre : 19 audios = 3664 frames (bruits de fond, silence, autres speakers, common-voices)

Total : 9264 frames, 128 features

Enregistrement :

- Bibliothèque sounddevice
- Pas avec le même micro...
- Mêmes phrases pour Aissa et Marouan

Prétraitement des données

Filtrage bandpass

iirfilter de scipy

Fenêtre de filtrage : [128, 2048]

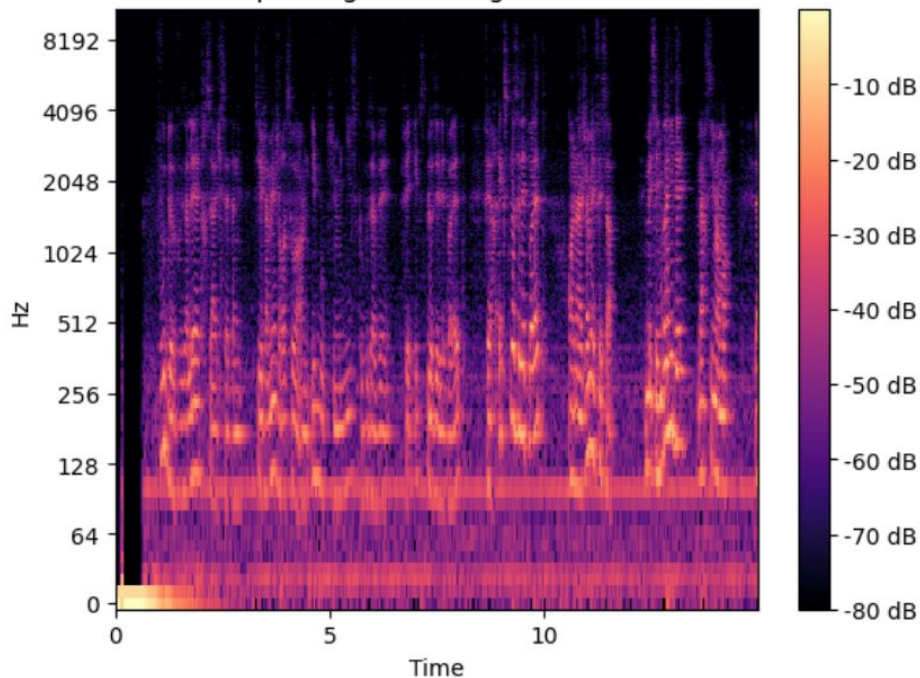
Suppression des silences

Librosa.effects.split

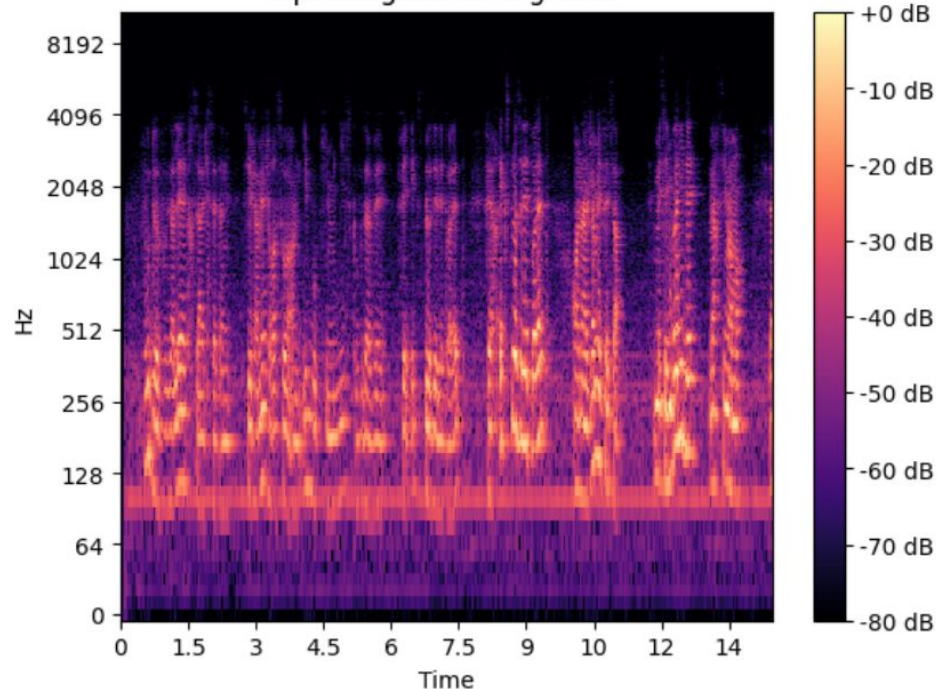
top_db = 35

Filtrage bandpass

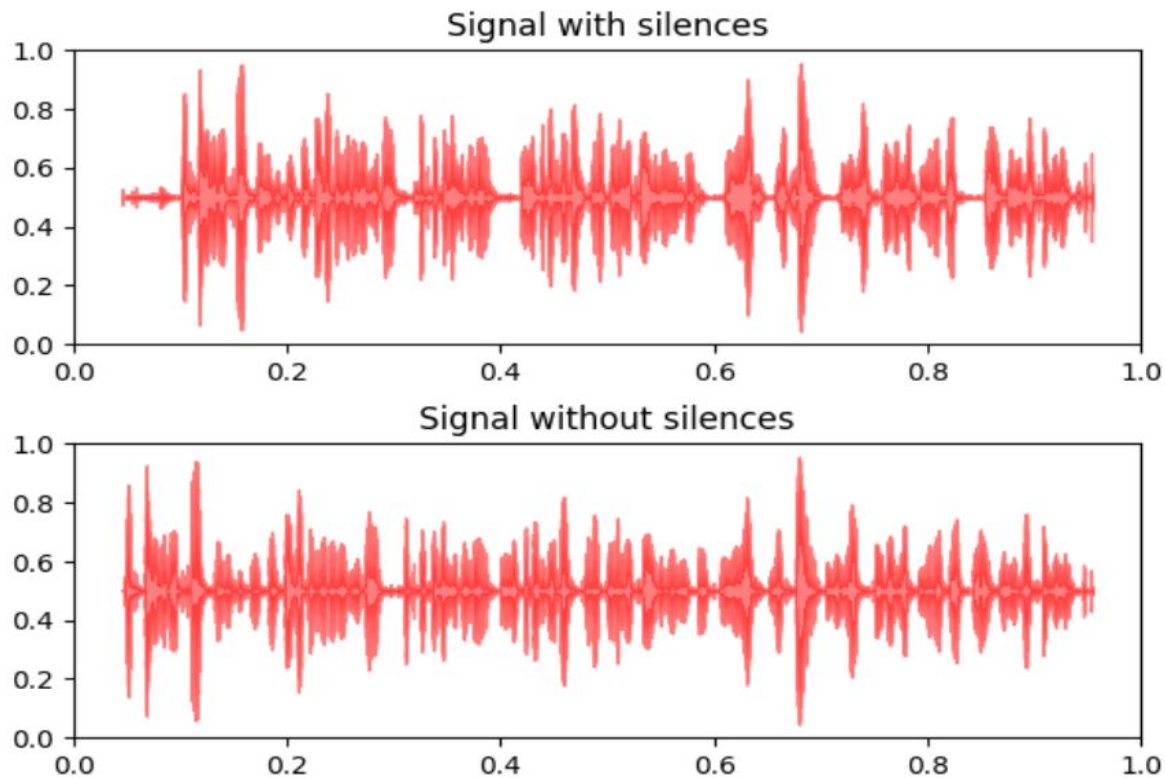
Spectrogramme log non filtré



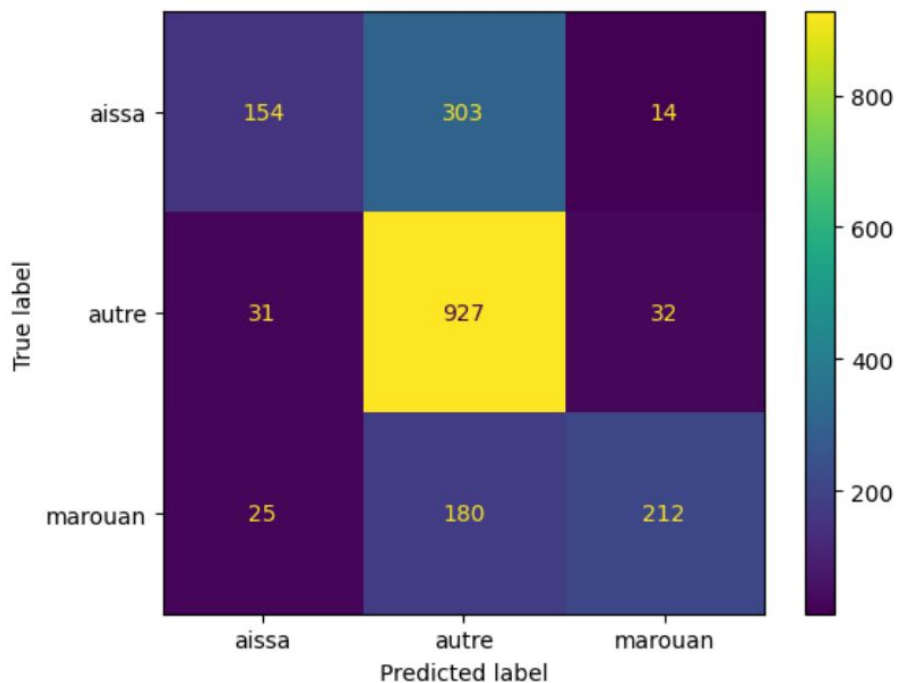
Spectrogramme log filtré



Suppression des silences

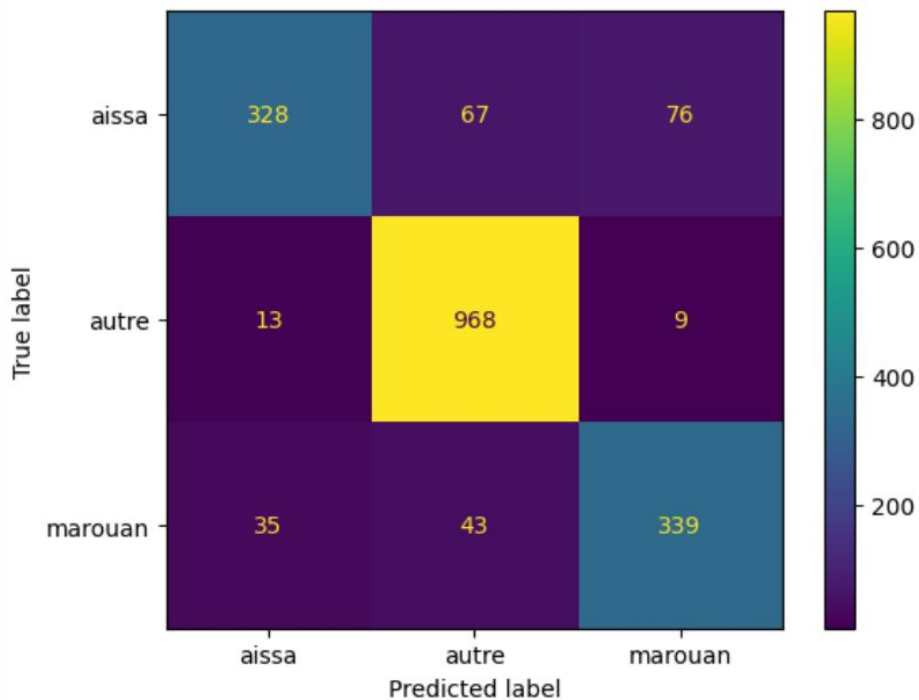


Régression Logistique



	precision	recall	f1-score	support
aissa	0.73	0.33	0.45	471
autre	0.66	0.94	0.77	990
marouan	0.82	0.51	0.63	417
accuracy			0.69	1878
macro avg	0.74	0.59	0.62	1878
weighted avg	0.71	0.69	0.66	1878

Random Forest



`n_estimators = 10`

`max_depth = 10`

`max_leaf_nodes = 70`

	precision	recall	f1-score	support
aissa	0.87	0.76	0.81	552
autre	0.91	0.97	0.94	1020
marouan	0.83	0.84	0.84	512
accuracy			0.88	2084
macro avg	0.87	0.86	0.86	2084
weighted avg	0.88	0.88	0.88	2084

Limites

Enregistrements des audios sur différents micros

Diversité des audios dans la classe “autre”

Filtrage non applicables sur toutes les classes (classe “autre”)

Biclasse > Multiclasse

→ Deep learning ? Convolution pour accentuer les features discriminantes ?

→ Utiliser algo de vision sur images spectrogrammes ? Perte d'infos ?

Retour d'expérience

Compréhension plus en profondeur ++

Preprocessing ++

Domaine complexe, nécessite approfondissement