

Projet Fil Rouge :

Reconnaissance de mots isolés audio en embarqué

Machine Learning

ENSEEIHT - 1ère année ModIA

Contact :

`remy.dutto@vitesco.com`

`Sandrine.Mouysset@irit.fr`

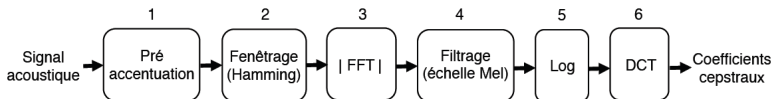


Commander un drone par commande vocale

créer un système de reconnaissance de commandes audio pour contrôler un drone : tourne droite, avance, fais un flip...

Transformation du signal temporel en MFCC (Mel Frequency Cepstral Coefficient)

Le cepstre présente l'avantage de permettre la séparation des contributions respectives de la source et du conduit vocal. Les MFCC s'obtiennent en utilisant, pour le calcul du spectre, une échelle fréquentielle non linéaire tenant compte de la perception auditive de la fréquence.



⇒ Utilisation de la librairie python *Librosa*

Organisation :

Ce projet en 6 séances se décomposent en 5 parties :

- *Partie I* : Sélection de modèles et pénalisation
- *Partie II* : Classification par méthodes à noyau
- *Partie III* : Apprentissage par ensemble : Adaboost et GradientBoosting
- *Partie IV* : Classification par réseaux de neurones
- *Partie V* : Votre étude

Projet :

- **Langage** : Notebook Python
- Travail à réaliser **en binôme**
- **Livrables du projet** : un notebook par binôme et un rapport au format pdf de 10 pages max.
- **Deadline** : le 24 juin !

Réalisez **votre propre étude**, par exemple, en :

- en augmentant les bases de test d'apprentissage avec les corpus fournis et/ou en rajoutant vos propres voix,
- en proposant des variantes des méthodes proposées et/ou en testant d'autres méthodes de classification,
- en utilisant d'autres méthodes de réduction de dimension des MFCC (ACP,t-SNE),
- en rajoutant davantage de classes (jusqu'à 13 ordres vocaux disponibles).

Le tout en testant les approches (parties I à IV) et en interprétant les résultats via les mesures d'évaluation (matrice de confusion et pourcentage de données bien classées) et synthétiser votre étude dans un rapport (10 pages en pdf).

Paramètres pour enregistrements audio de vos voix perso:

16 KHz, mono, 16 bits, format *.wav*