



# Yohann Perez

Ingénieur en informatique  
Orientation logiciels et systèmes complexes

# Yohann Perez

Ingénieur en informatique  
Orientation logiciels et systemes complexes

## Mon parcours

---

- CFC Informatique accéléré (15-18ans)
- Bachelor I.T.I. (18-21ans)
- Service civil (21-22ans)
  
- B2 : Anglais (FCE) & Espagnol
- B1 : Allemand
  
- Loisirs : VTT, ...

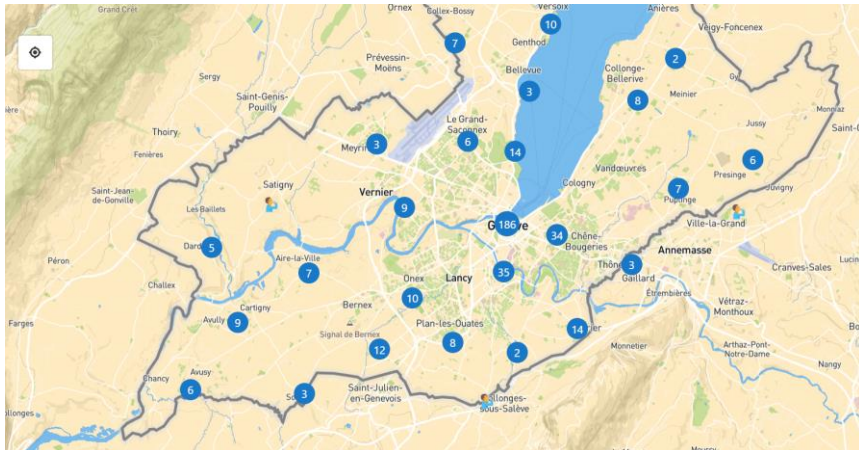
## CFC, ce que j'en retiens

---

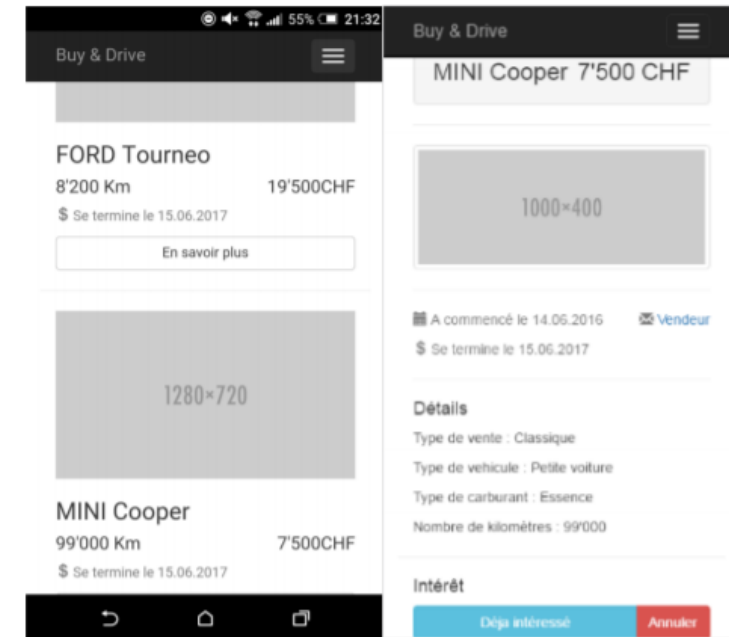
- Développement WEB
  - HTML, JS - ES6, PHP, SQL & CSS
- Développement d'application Windows
  - C#
- Configuration de poste et de réseaux
  - Windows 7 & Cisco
- J'ai appris le métier d'informaticien

# Quelques travaux de mon CFC

## GE-SOIF, AGENDA21



## BUY & DRIVE, CFC



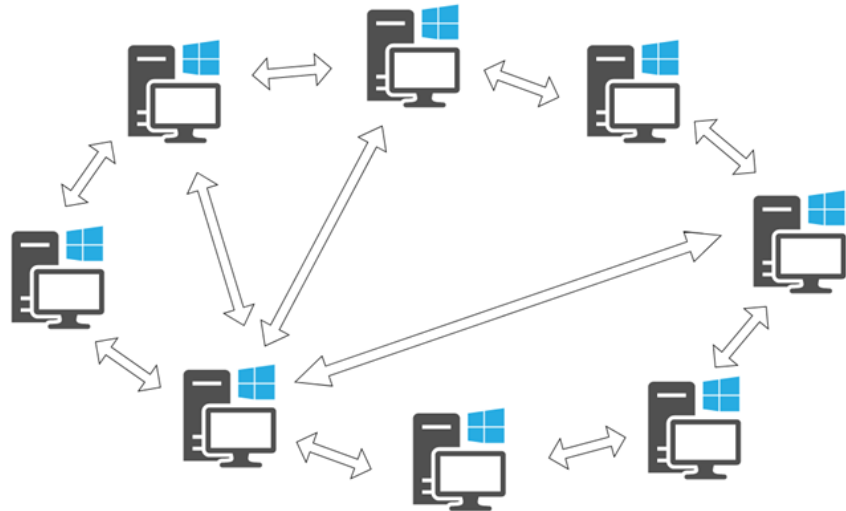
# Bachelor ITI, ce que j'en retiens

---

- Développement WEB & apps mobiles avec interactions (senseurs & API)
  - Angular, NodeJS, Express & MongoDB
- Algorithmiques, mathématiques avancées & machine learning
  - Octave, Python & Java
- Programmation bas niveau (système)
  - C & C++
- Meilleure compréhension du fonctionnement d'un ordinateur
- Développement personnel (capacité de réflexion & capacité d'analyse)

# Quelques travaux de mon bachelor

## SIMPLIFIED BITTORENT



## FUNSPEECH



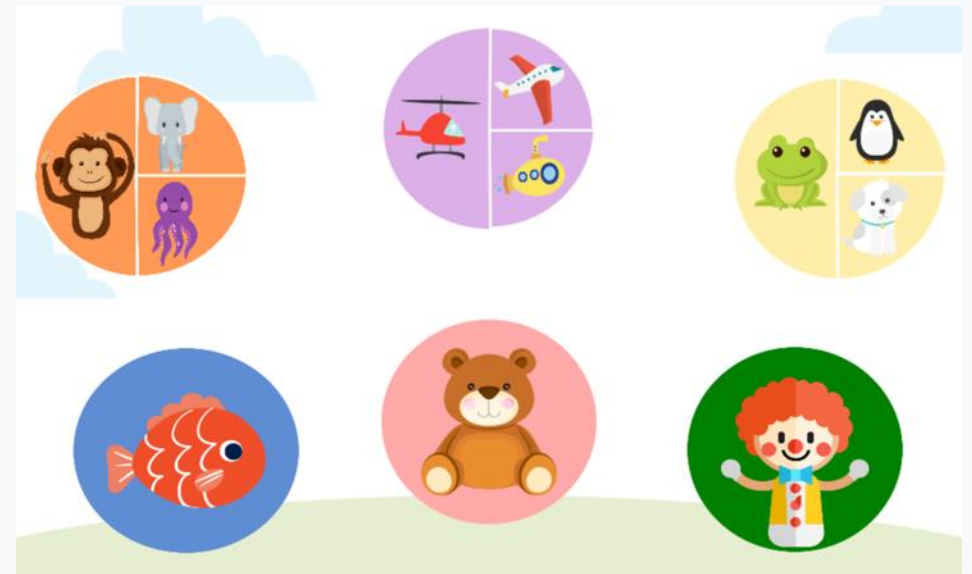
# FunSpeech

Travail final de bachelor



# Description de FunSpeech

- Application de rééducation orthophonique
- Destinée aux enfants sourds implantés
- A été développée par HEPIA (+2ans)
- But : Encourager les enfants à la production de sons



## FunSpeech – la problématique

- Algorithme de reconnaissance de sons de base de la parole (phones) ne fonctionne pas
- Investiguer les algorithmes de reconnaissance de sons, d'extraction de caractéristiques et de classification.
- Trouver une meilleure combinaison d'algorithmes

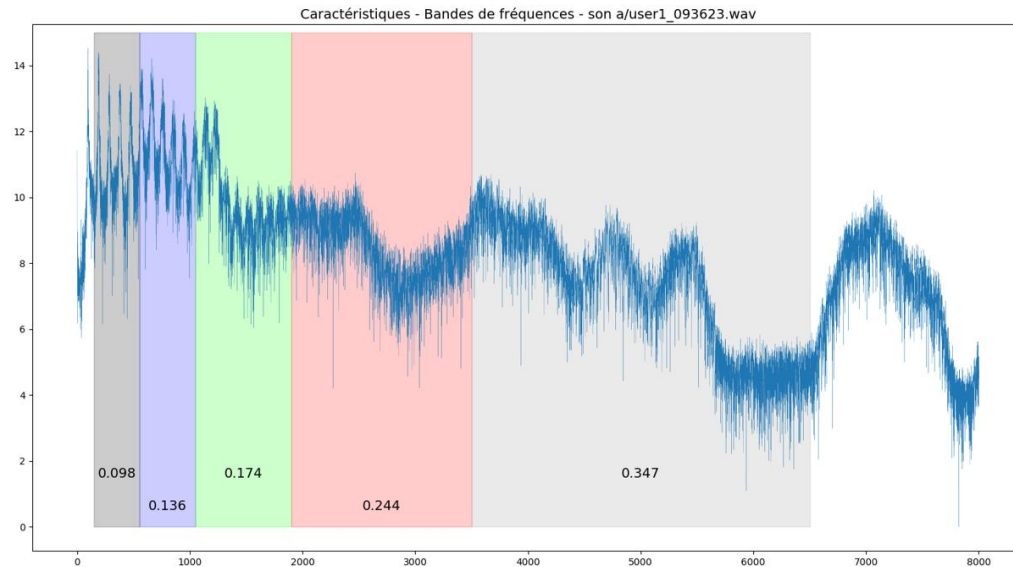


# Algorithme actuel

- Energies des bandes

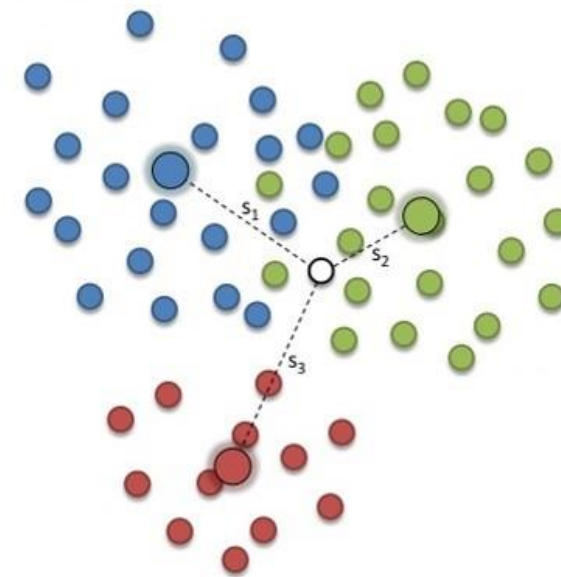
- $E_{\text{bande}} = \frac{\sum x_i^2}{\text{énergie tot.}}$

- [0.098, 0.136, 0.174, 0.244, 0.347]



- Plus proche barycentre

- $Distance(p, q) = \sqrt{\sum_{i=1}^{n=5} (q_i - p_i)^2}$



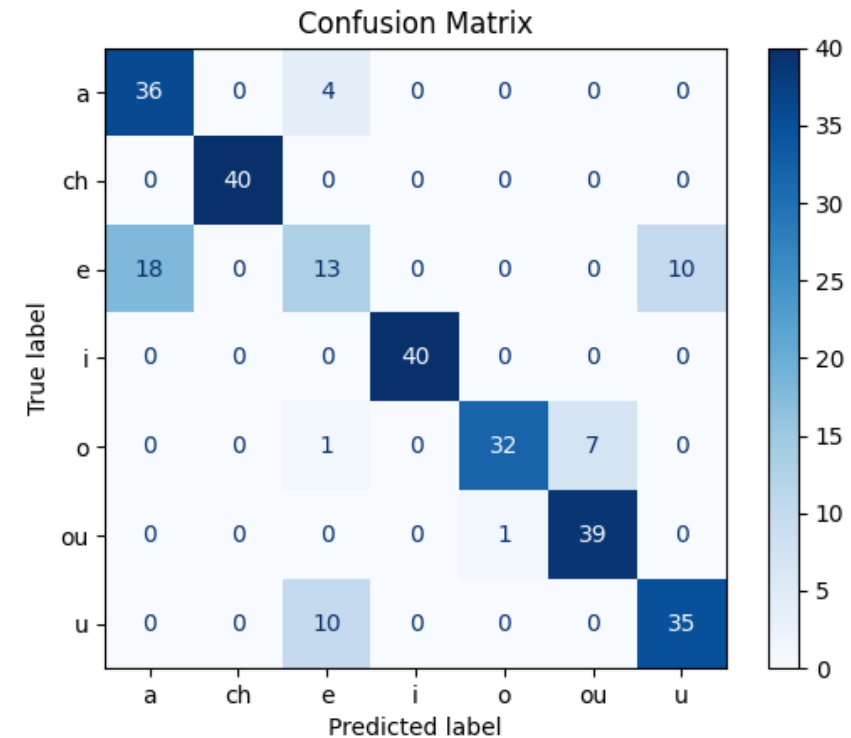
# Démarche pour résoudre la problématique

---

- Acquérir des jeux de données
- Comprendre le mécanisme de production d'un son
- Investiguer les caractéristiques acoustiques
- Investiguer les algorithmes de classification
- Définir et implémenter des protocoles de tests et de validations

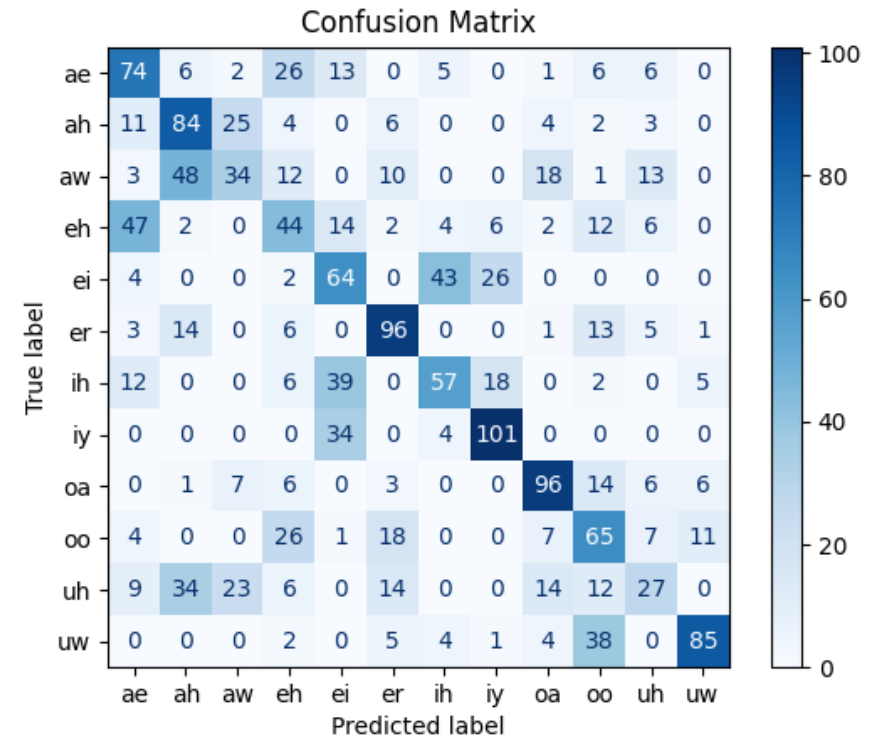
# FunSpeech 4-Voices

- 287 données
- 4 locuteurs
- 7 phones suisse/français
- Classes équitablement réparties
- Inconvénients:
  - Phones adultes → Jeunes enfants
  - Peu de diversité
  - Nécessite un prétraitement (silence), + des vérifications



# Western Michigan Uni.

- 1667 données
- 12 phones américains
- 139 locuteurs
  - 48 femmes + 45 hommes
  - 19 filles + 27 garçons
- Classes équitablement réparties



# Mécanisme de production d'un son

---



- Expiration
- Intensité du son  
(faible ou fort)



- Impulsion glottale
- Hauteur du son  
(aigu ou grave)

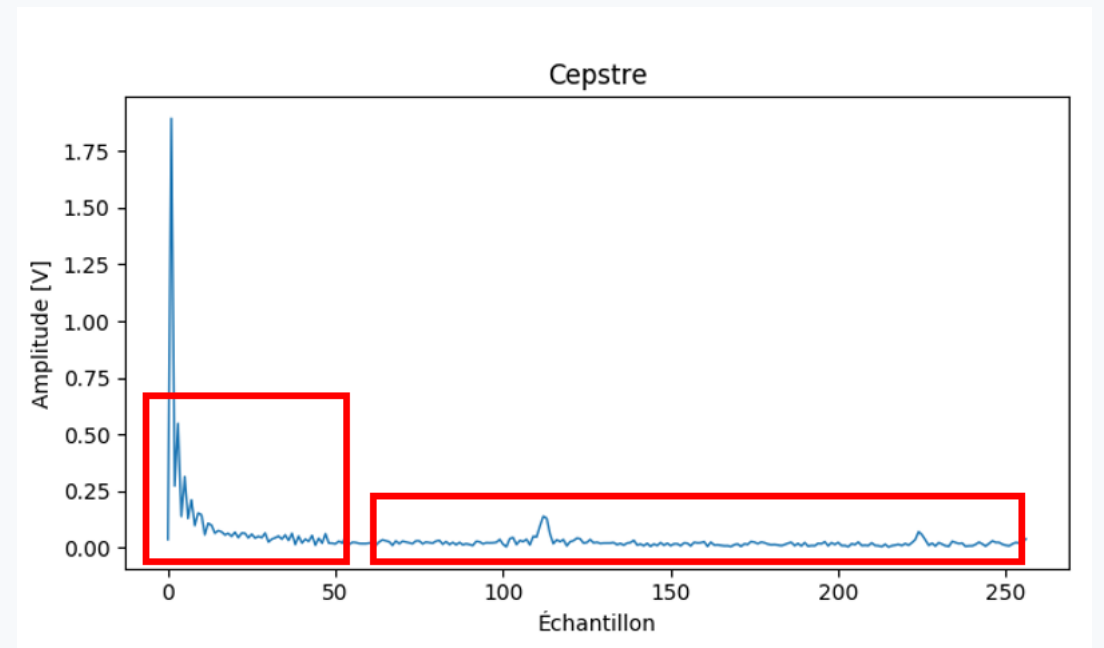


- Forme du canal vocal
- Phone qui va être  
produit

- Dans un signal vocal, l'impulsion glottale et la forme du canal sont convoluées

# Domaine du cepstre

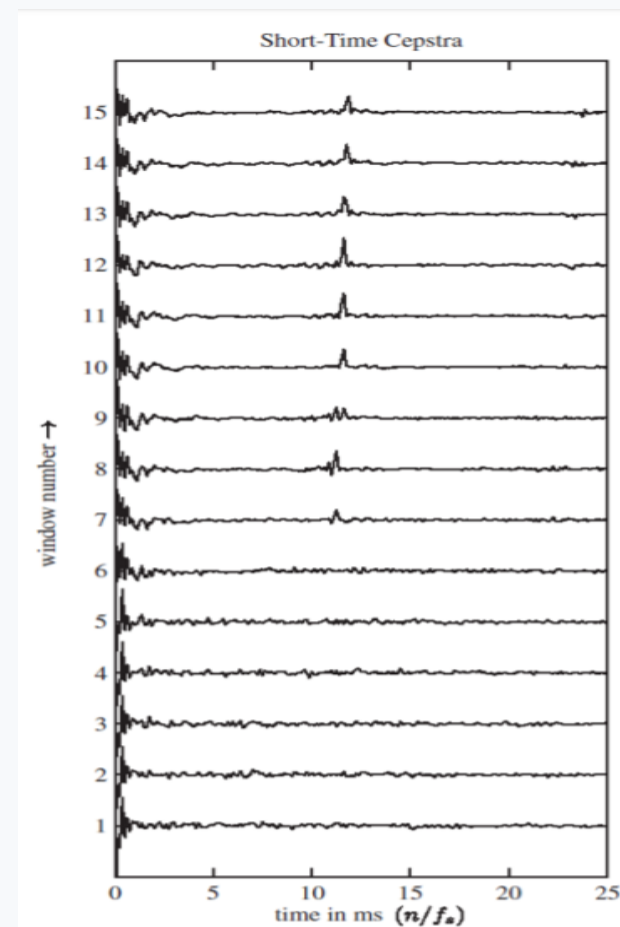
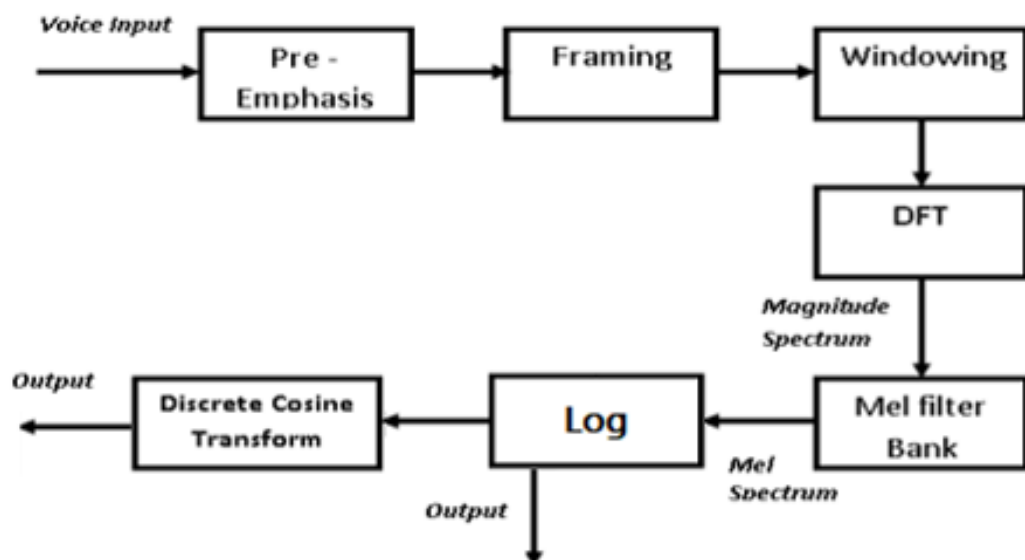
- Distinguer l'impulsion glottale et la forme du canal vocal = Analyse cepstrale
- $Cepstre = FFT^{-1}(\ln(|FFT(x)|))$



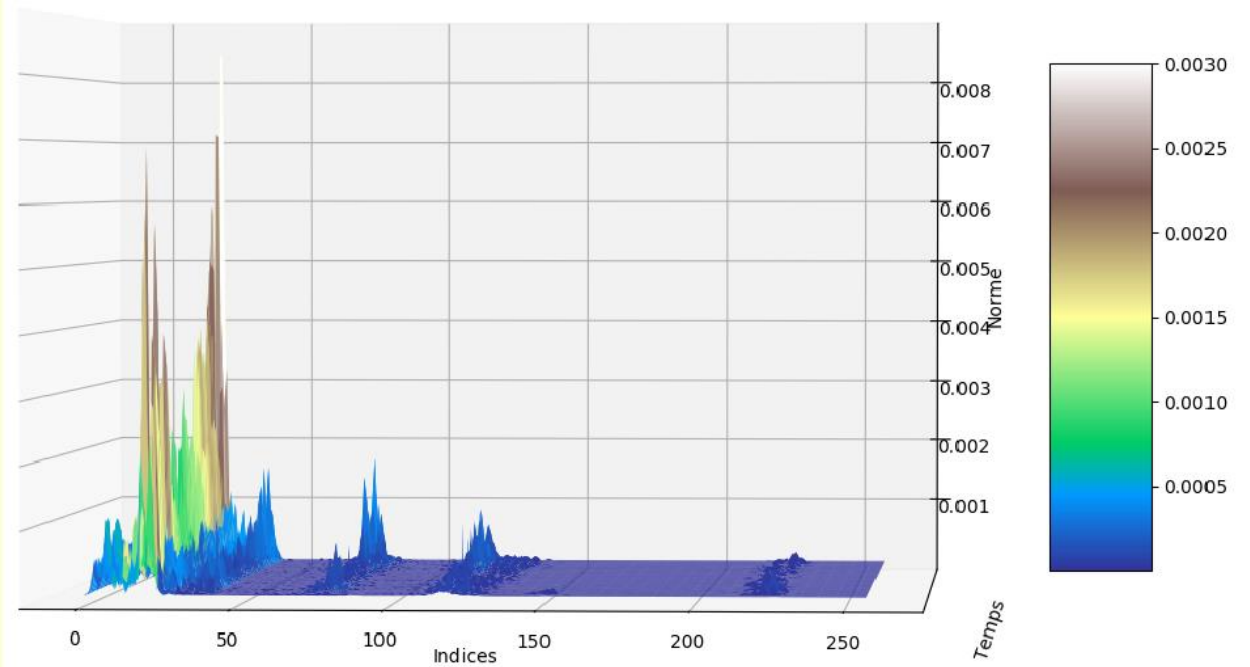
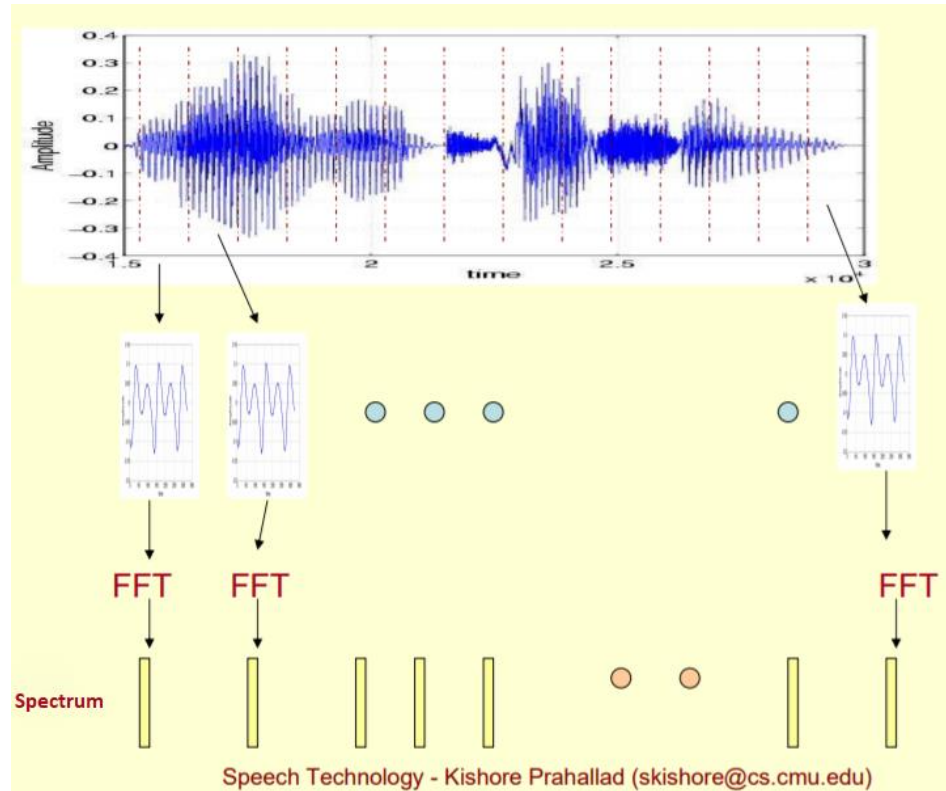


# Mel Frequency Cepstral Coefficients

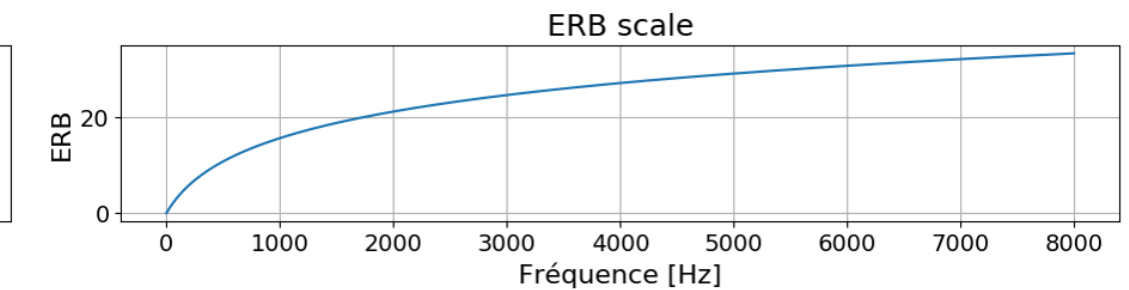
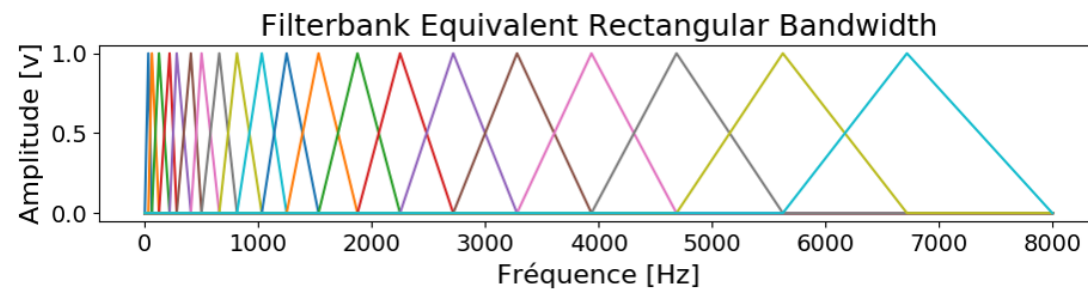
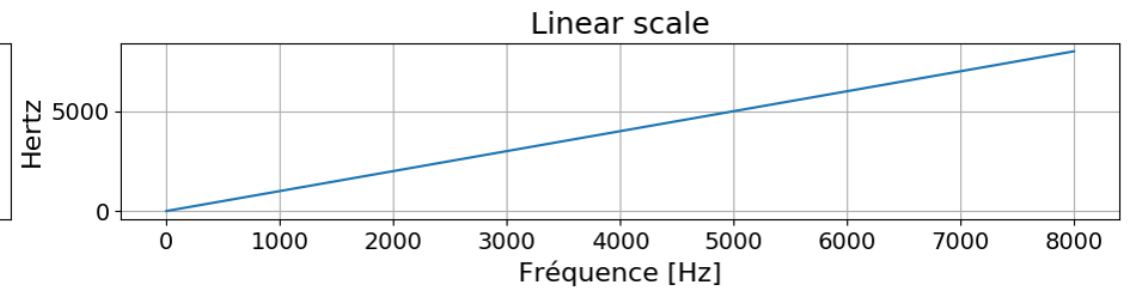
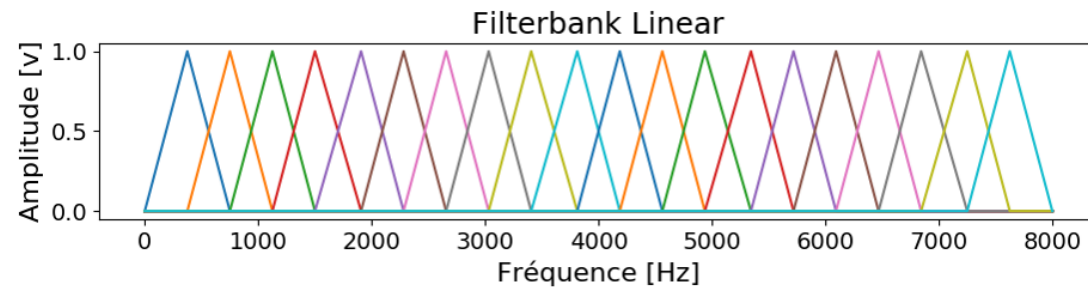
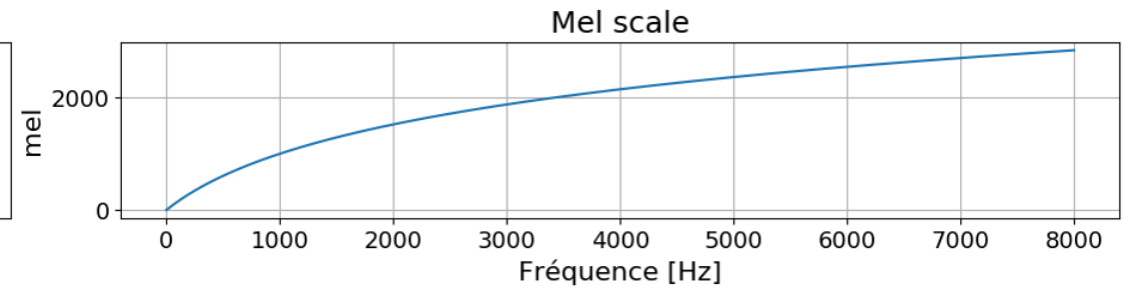
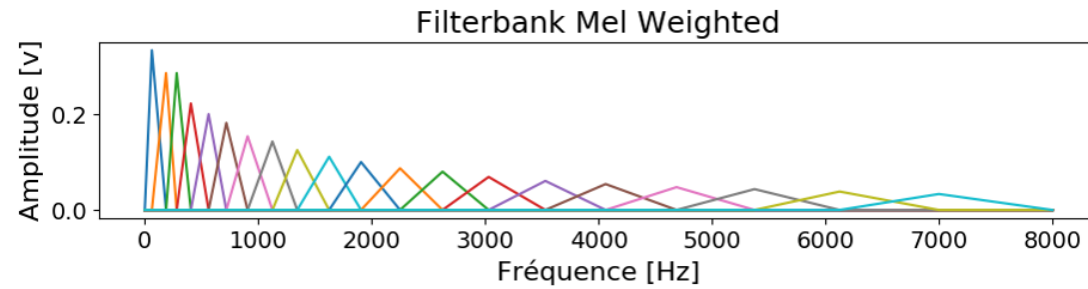
- Evolution de la forme du canal vocal et de l'impulsion glottale au cours du temps
- Banc de filtres ou analyse du cepstre



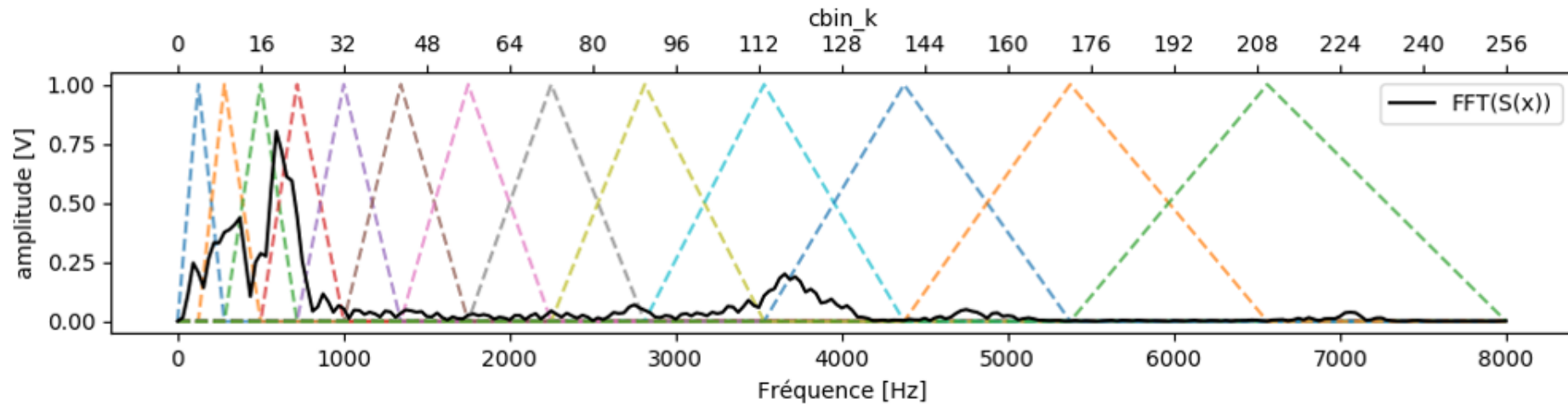
# Mel Frequency Cepstral Coefficients



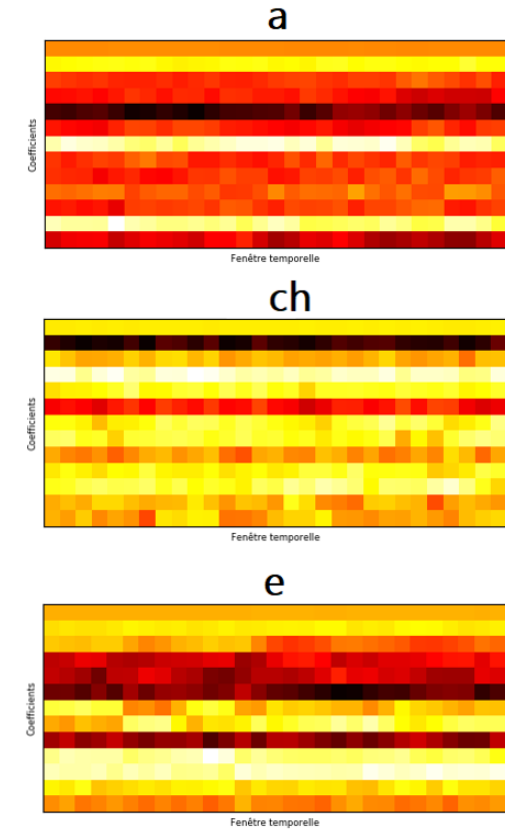
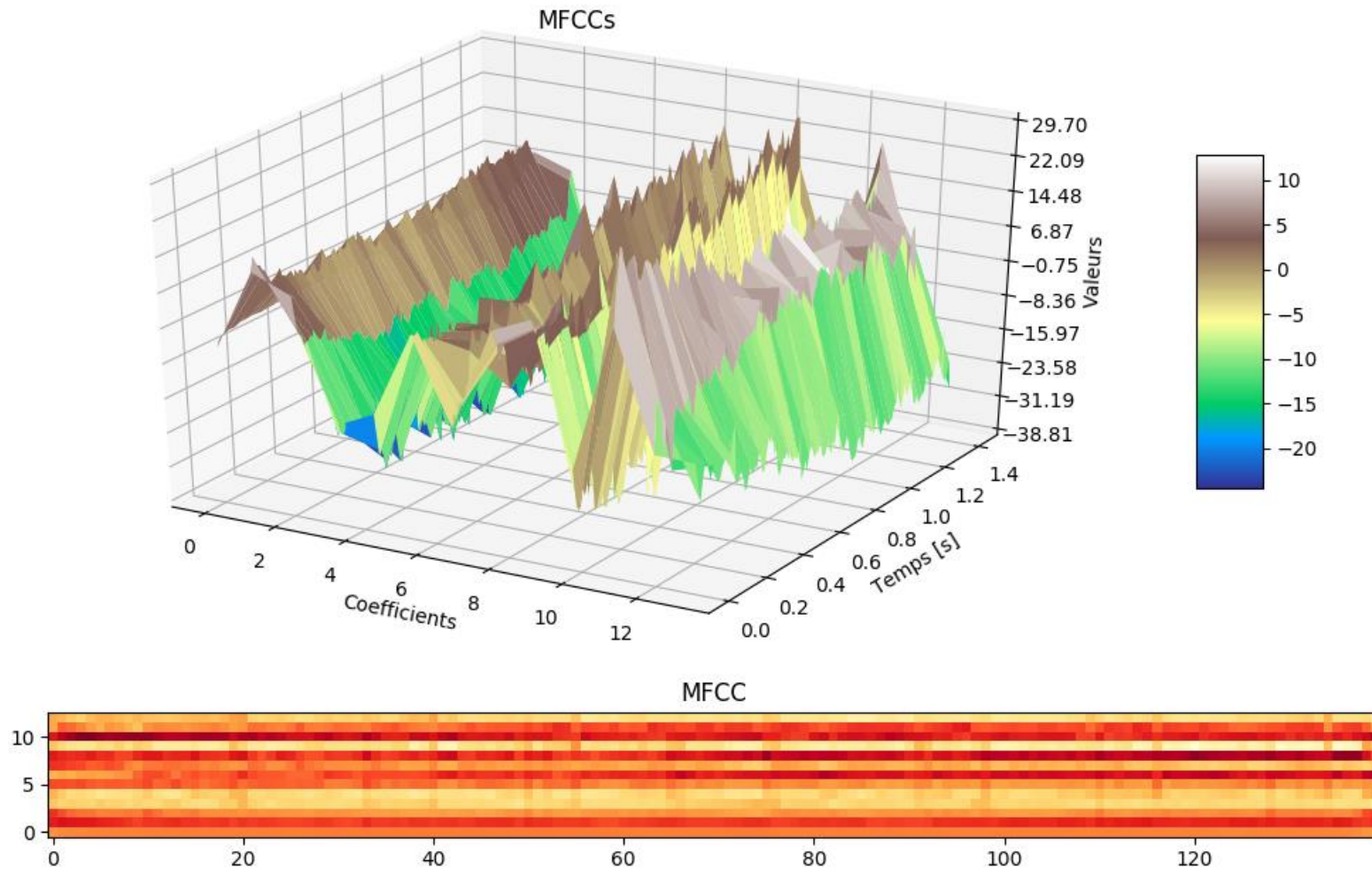
# Mel Frequency Cepstral Coefficients



# Mel Frequency Cepstral Coefficients

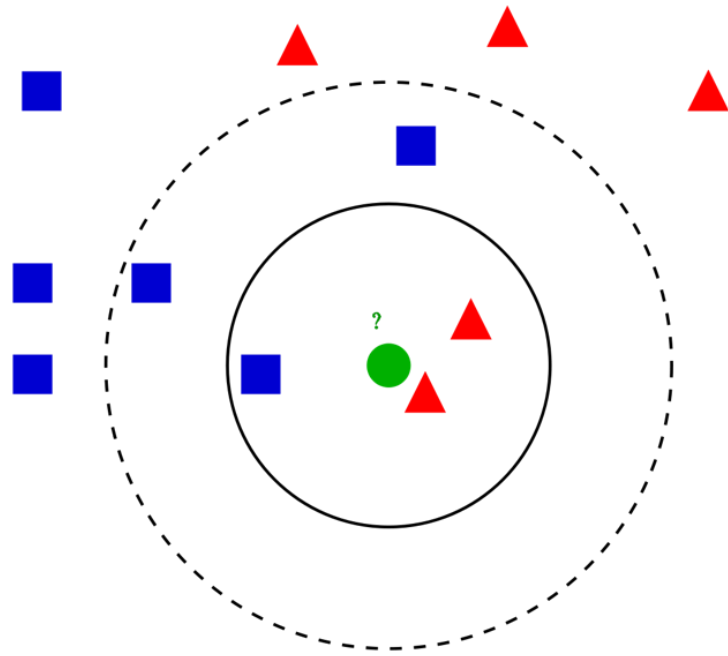


# Mel Frequency Cepstral Coefficients

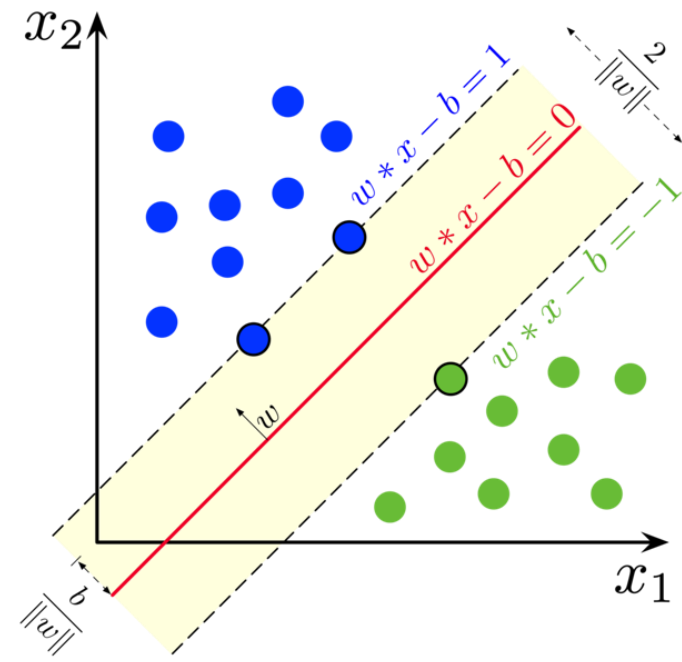


# Algorithmes de classification

## N PLUS PROCHES VOISINS

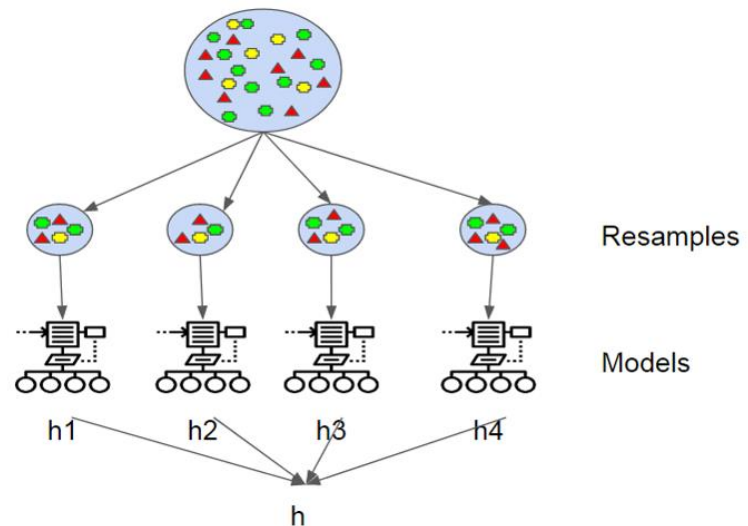


## SÉPARATEURS À VASTES MARGES

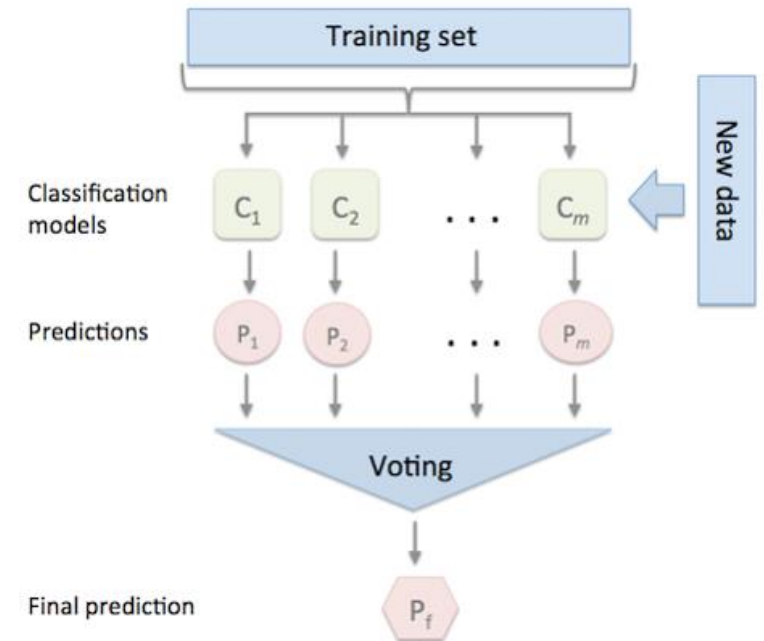


# Algorithmes de classification

## BAGGING



## COMBINAISON DE CLASSIFIEURS



## Que dois faire mon protocole de tests ?

---

- Déterminer si l'utilisation de phones uniquement produits par des adultes pour l'entraînement d'un modèle suffit pour correctement classifier des phones produits par des d'enfants
- Déterminer combien le fait d'ajouter des données d'entraînement est bénéfique à la résolution de la problématique
- Comparer les différents combinaisons d'algorithme d'extraction et de classification pour déterminer laquelle est la plus performante et la plus adaptée pour résoudre la problématique



# Protocole de tests

---

## QUALITÉ DE PRÉDICTION DU MODÈLE

- Métriques
  - Précision
  - Incertitude
  - Temps

## PROCÉDURE DE VALIDATION

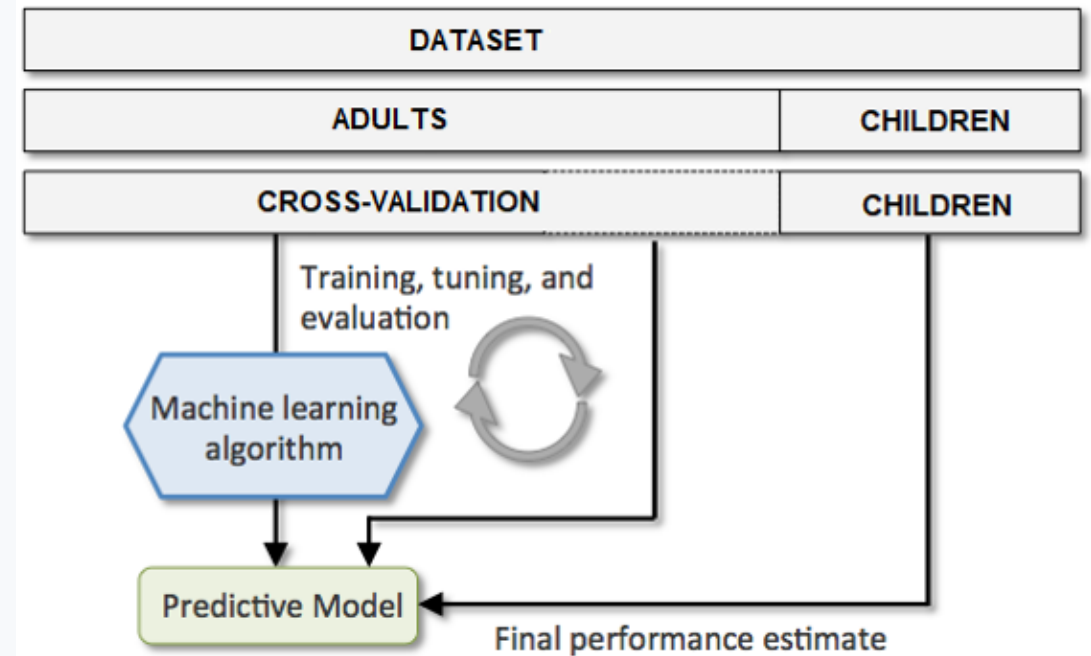
- Validation croisée
- Train, Validation & Test sets
- Courbes d'apprentissage

## OPTIMISATION DES CLASSIFIEURS

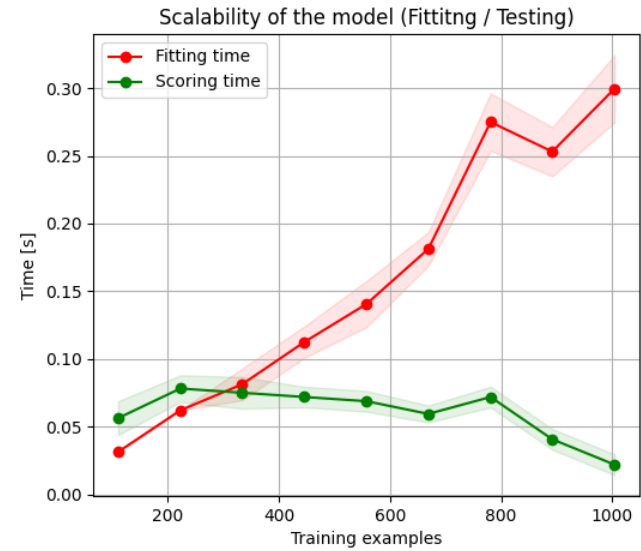
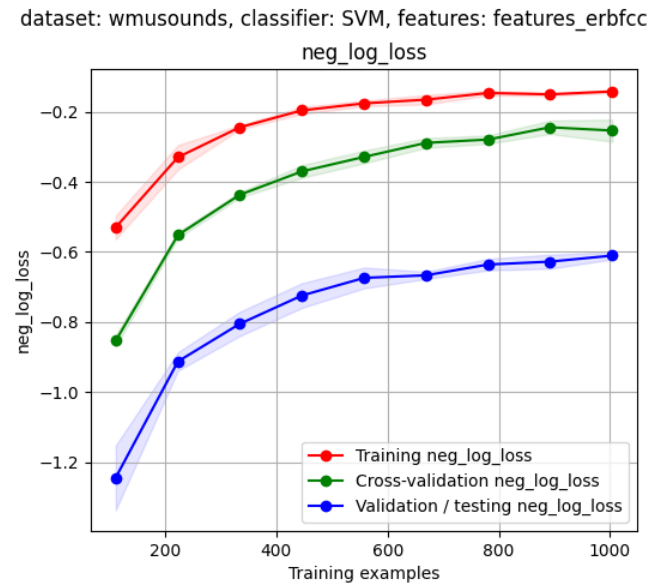
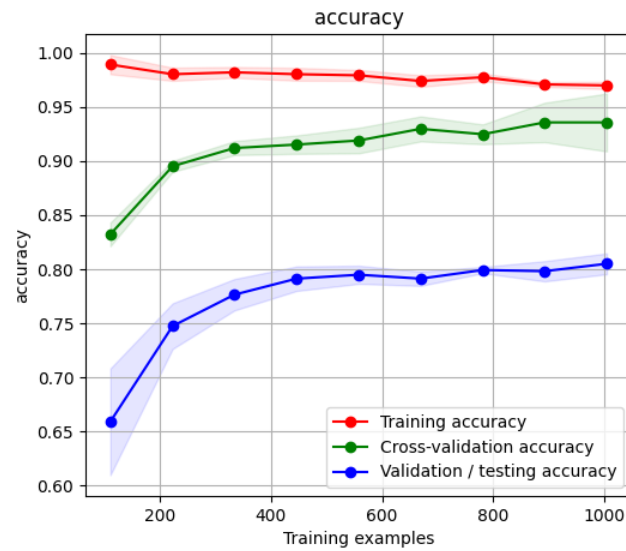
- Recherche exhaustive des paramètres

# Train, validation & test sets

- Train → Pouvoir prédire
- Validation → Ajustement des paramètres
- Test → Estimation non biaisée



# Courbes d'apprentissage

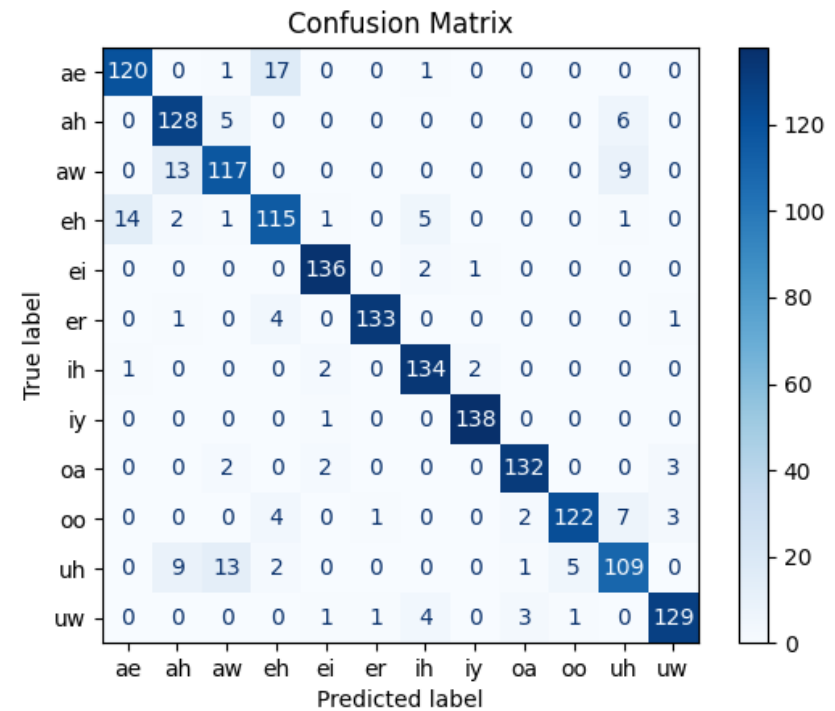
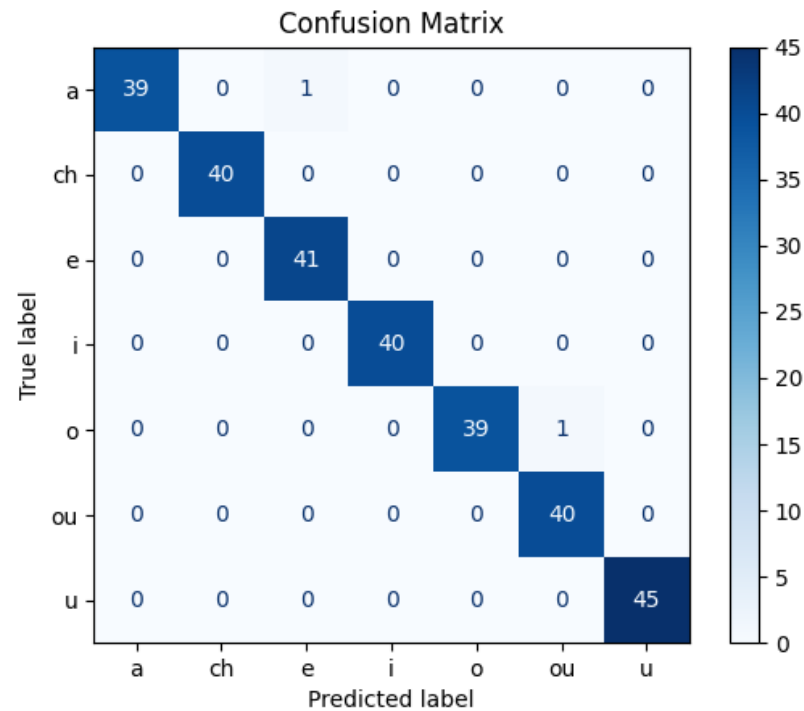


## Résultats ( + de 120 fichiers )

---

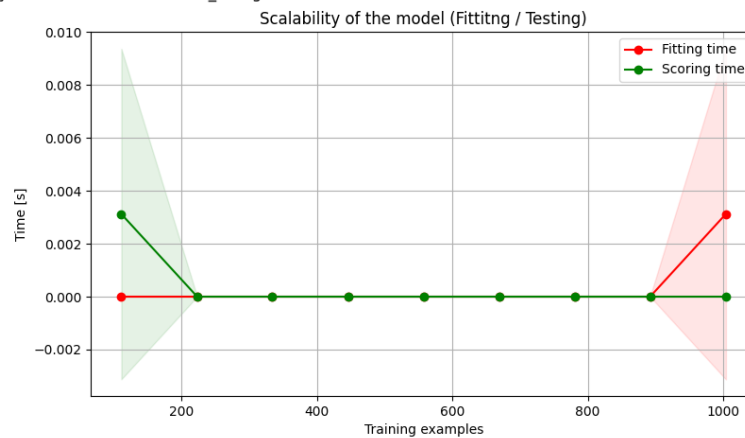
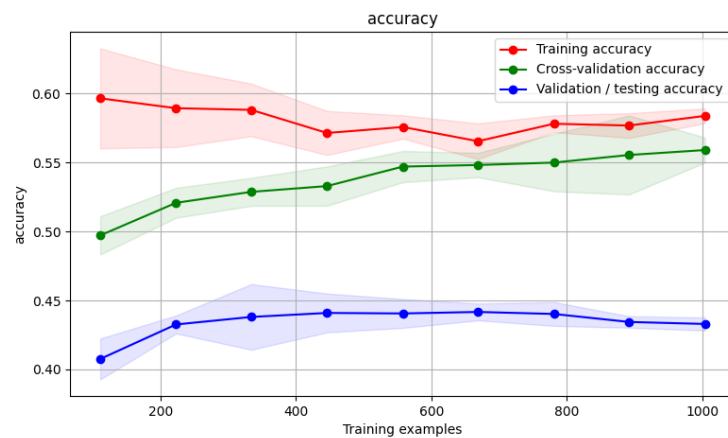
- Caractéristiques
  - Energies
  - MFCC, MW-FCC, L-FCC & ERB-FCC
  - Log(banc de filtres)
- Echantillon et signal complet
- Deux jeux de données
- Algorithmes de classification
  - Plus proche barycentre
  - K-plus proche voisins
  - Séparateurs à vastes marges
  - Bagging
  - Combinaison de classifieur

# Résultat: SVM + ERB-FCC (matrice de confusion)

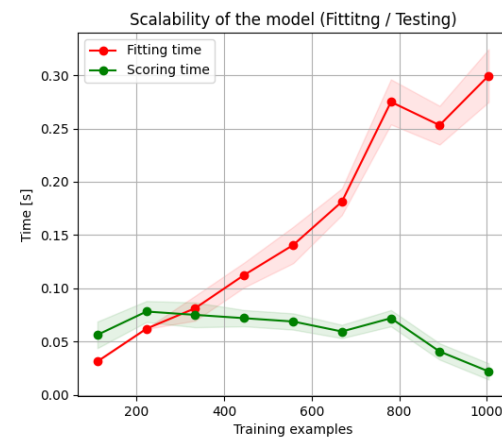
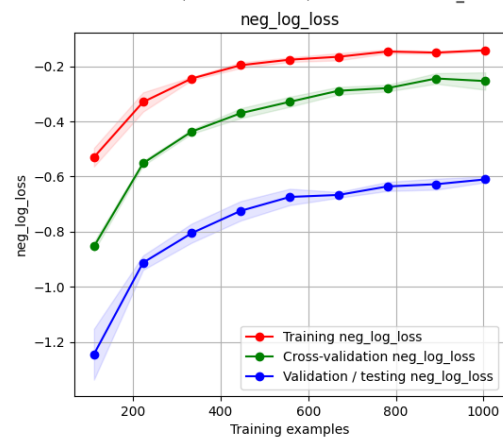
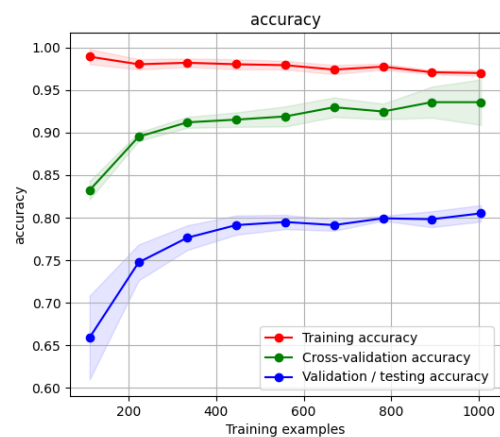


# Résultats (Train, Validation & Test) (avant / après)

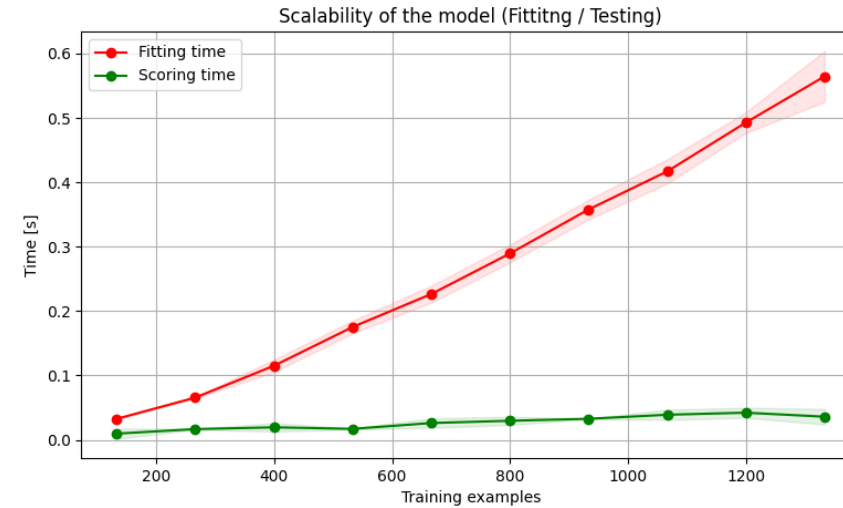
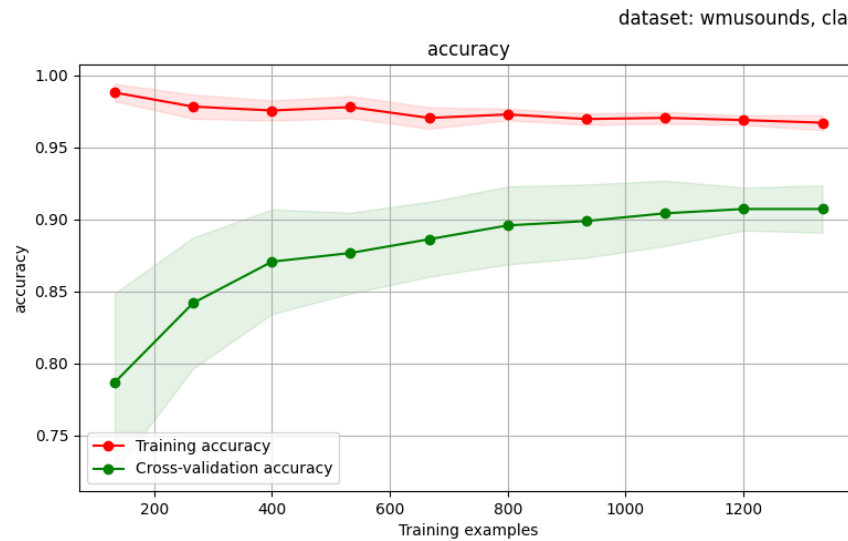
dataset: wmusounds, classifier: Barycentre, features: features\_energies



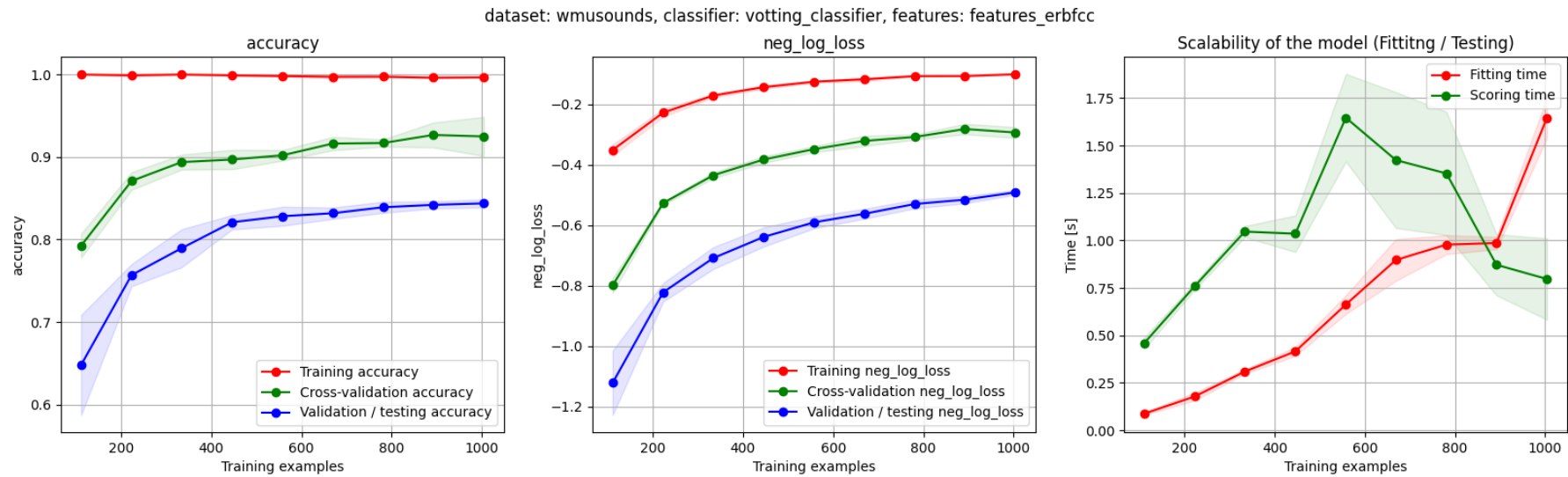
dataset: wmusounds, classifier: SVM, features: features\_erbfc



# Un des meilleurs résultats (Train & Test)



# Meilleur résultat combinaison de classifieurs





## Conclusion

---

- Classifieur : SVM
- Caractéristiques: ERB-FCC
- Des sons adultes suffisent à classifier des sons enfants
- La limitation de la taille d'entraînement est possible
- Grande amélioration, de 44% à 82%
- stagnance à 82% de reconnaissance