



**ASDLearn**

L'outil d'aide au  
diagnostic  
précoce de  
l'autisme

# 1 - Les étapes de développement de notre Intelligence Artificielle



PRESENTE PAR :

BREUILLER JULIEN - CHEKROUN KATHYNA - DJEMAME YASMINE -  
ELBAZ LEA - LUXIMON DAVINA - VANDERMARLIERE GABIN

A child with curly hair is seen from the back, playing with colorful blocks. The child's head is at the top right, and their hands are visible at the bottom left, interacting with the blocks. The blocks are in various colors: red, green, blue, and white. The background is a solid light blue color.

## TABLE DES MATIÈRES

---

Création de notre base de données p1

Compréhension et extraction des  
caractéristiques des signaux vocaux p2

Tests et choix de l'IA p5

# Création de notre base de données

Après des recherches de bases de données publiques, et après avoir contacté des instituts spécialisés, il s'est avéré qu'il n'y avait pas de bases de données pré-existantes d'enfants autistes en très bas âge. Nous avons donc décidé de créer notre propre base de données.

Nous avons créé notre propre base de données de signaux vocaux d'enfants de moins de 2 ans, en nous servant de YouTube (voir protocole ci-contre). Nous avons alors classé les audios dans deux répertoires : "autistes" et "sains".

## PROTOCOLE DE RECUPERATION

- Installer une application d'enregistrement audio
- Recherche sur Youtube avec mots clés anglais (ex : "babbling")
- Décider des passages pertinents
- Nommer enregistrement :  
ex : **SG3a\_12mois\_joue**
- Convertir en .wav
- Dépôt dans les dossiers correspondants
- Téléchargement BDD complète pour extraction

Résultat : 325 signaux récoltés jusqu'à maintenant.

# Compréhension et extraction des caractéristiques des signaux vocaux

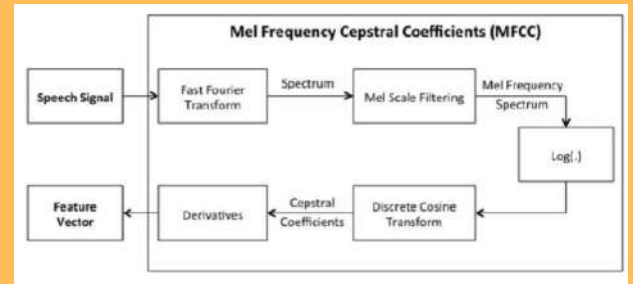
Avant de procéder à l'entraînement d'un réseau de neurones IA, il nous a fallu étudier différentes caractéristiques des signaux vocaux d'enfants.

## Coefficients MFCC ( Mel-Frequency Cepstral Coefficients)

Représentation des spectres de puissance sur des temps courts d'un signal.

C'est le résultat de l'application de 12 à 14 filtres successifs dont l'espacement correspond à l'échelle Mel sur un échantillon de 10 ms d'un signal. Ce processus est similaire à celui appliqué par l'oreille, il permet aux individus de reconnaître une personne grâce à sa voix. Pour obtenir les MFCC d'un signal, on accentue le signal en faisant ressortir les hautes fréquences avec un filtre passe-haut. Ensuite on découpe le signal en fenêtre de 30ms toutes les 10ms et on applique la fenêtre de Hamming sur ces trames. Par la suite, la transformée de Fourier de la trame est réalisée. On applique alors un banc de filtres triangulaires - qui permettent mieux de simuler le fonctionnement de l'oreille humaine - logarithmiquement espacés selon l'échelle de Mel. L'échelle mel est une échelle dont l'espacement entre deux fréquence est assez grand pour que deux sons soient différenciés à l'écoute.

Enfin, on applique une DCT (Transformée en Cosinus Discrète). En reconnaissance de parole, cette méthode est la plus utilisée pour représenter un signal.



## Deltas et Deltas-Deltas

Les Deltas et Deltas-Deltas sont calculés à partir des MFCC. Ils permettent de représenter la variation de la parole dans le temps. Également appelés coefficients différentiels et d'accélération, les deltas représentent les trajectoires des coefficients MFCC dans le temps. On les calcule de la manière suivante :

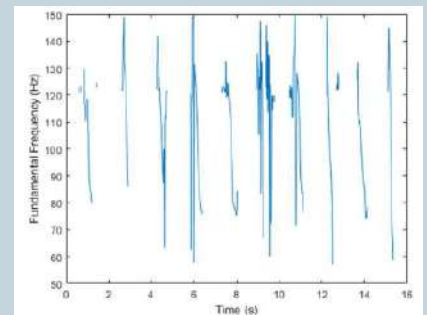
$$d_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^N n^2}$$

Avec  $d_t$  le coefficient delta, calculé à partir des coefficients statiques de MFCC (de  $ct+N$  à  $ct-N$ ) où  $N$  vaut 2 la plupart du temps

Les Deltas-Deltas sont calculés de la même façon mais à partir des Deltas.

## Pitch

Fréquence fondamentale perçue par l'oreille d'un signal vocal. Il caractérise la hauteur de la voix, c'est-à-dire si la voix est aiguë ou grave. Il est obtenu par un algorithme de détection de hauteur aussi appelé ADPs.



Ces caractéristiques vocales numériques vont être extraites pour chaque enfant de moins de 2 ans sous Matlab, et placées dans un type "structure", qui correspond à une forme de tableau pouvant contenir tout type de variables (vecteurs, matrices, caractères, etc.). Les enfants autistes seront labellisés "1" et les non autistes "-1". Cette "structure" sera utilisée pour tester les méthodes de Machine Learning sous Matlab. Elle est également convertie en .csv pour des tests sous Python.



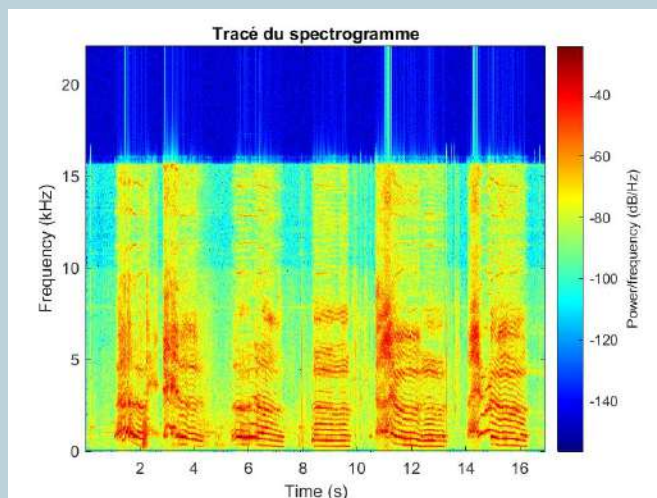
# Compréhension et extraction des caractéristiques des signaux vocaux

*Avant de procéder à l'entraînement d'un réseau de neurones IA, il nous a fallu étudier différentes caractéristiques des signaux vocaux d'enfants.*

## Spectrogrammes

Le spectrogramme est une représentation visuelle du spectre d'un audio qui change à travers le temps. Il est obtenu à l'aide de la transformée de Fourier à court terme (ou transformée de Fourier fenêtrée).

Il possède deux axes : le temps et la fréquence, il nous renseigne donc sur une possible bande de fréquences en particulier pour les autistes, ou bien sur une variation fréquentielle au cours du temps qui pourrait être caractéristique d'un malade. La courbe des formants obtenus, sur certains sons voyelliques par exemple, peut également nous aider à différencier un individu sain d'un malade.

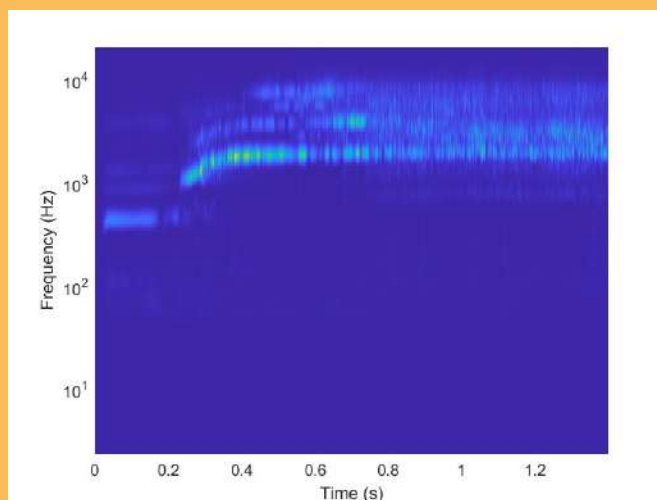


## Scalogrammes

Représentation de la répartition des coefficients d'ondelettes en fonction de l'échelle choisie.

Ces coefficients d'ondelettes sont obtenus grâce à la transformée en ondelettes.

$$g(s, \tau) = \int_{-\infty}^{+\infty} f(t) \psi_{s, \tau}^*(t) dt$$



Ces spectrogrammes et scalogrammes vont être générés sous Matlab et enregistrés séparément sous forme d'images, pour chaque enfant de moins de 2 ans, dans des répertoires labellisés "autistes" et "sains". Ces répertoires vont être utilisés par le Deep Learning, afin de tester différentes méthodes, toujours sous Matlab et sous Python.



**MACHINE  
LEARNING ?**

**QUELLE IA  
CHOISIR ?**



**DEEP LEARNING ?**

# TESTS ET CHOIX DE L'IA

## MACHINE LEARNING

### VS

## DEEP LEARNING

### THEORIE

- **Deep Learning:**  
Entraînement à l'aide d'images de Scalogrammes et de Spectrogrammes
- **Machine Learning:**  
Utilisation d'un Data frame de valeurs numériques représentant les caractéristiques de signaux vocaux (pitch, mfcc, deltas, deltasDeltas)

### TESTS EFFECTUES

- **Deep Learning :**  
AlexNet et GoogleNet avec différentes tailles d'epochs sous Matlab et Python.
- **Machine Learning :**  
CSVM, Logistic regression, KNeighbors, Random forest, GaussianNB, Decision Trees sous Matlab et Python.

### MEILLEURS RESULTATS

- GoogleNet sous Matlab, avec un taux de prédiction de plus de **83%** pour le **Deep Learning**. Ce taux de prédiction a connu une **amélioration progressive** en fonction de l'ajout de signaux dans notre base de données.
- Le CSVM sous Matlab, avec un taux de plus de **70%** pour le **Machine Learning**

### CHOIX FINAL

Le **Deep Learning** avec le réseau de neurones GoogleNet sur les **spectrogrammes**, car on y obtient le meilleur taux de réussite en terme de prédiction.

Nous sommes conscients qu'un taux de réussite à 83% pour du Deep Learning, dans une application médicale n'est pas suffisant. Mais ceci s'explique par le peu de données dont nous disposons en raison de la récupération manuelle de signaux (325 audios au moment de la rédaction de cette brochure), car en effet, nous avons tout de même pu observer une augmentation progressive du taux de réussite, en fonction de l'augmentation du nombre de données. Pour que notre outil soit fiable et robuste, il faut donc largement augmenter la base de données, peut-être avec l'aide d'un partenaire médical car, à notre échelle, nous avons atteint une limite en terme de récupération.