Sorbonne Nouvelle

Réseaux de neurones appliqués à l'oral

# Rapport Classification Parole Pathologique

*Auteures :*
Léna GAUBERT
Qinliang QI
Natacha MINICONI

14 janvier 2024

# Table des matières

# 1 Introduction

La maladie ataxie spinocérébelleuse, en raison de sa nature génétique, constitue une pathologie rare touchant un nombre restreint d'individus. Dans ce contexte, l'importance d'une classification automatique réside dans la possibilité d'offrir un diagnostic précoce, favorisant ainsi une intervention thérapeutique plus prompte et efficace. La recherche d'un modèle de classification optimal ainsi que d'un input adéquat est donc d'une importance cruciale

Les patients atteints de cette maladie présentent des perturbations rythmiques dans leur production de parole, caractérisées par des transitions prolongées entre les phonèmes, avec une irrégularité persistante tout au long de l'énoncé. Ici, l'utilisation de la modulation cepstrale permet de visualiser les variations dans le spectre mais aussi dans les changements articulatoires. Cette approche offre ainsi une représentation visuelle de l'espace temporel des locutions, permettant la compréhension des caractéristiques distinctives des discours de ces patients.

Les spectrogrammes, en capturant les déformations potentielles des voyelles et des consonnes induites par la maladie, offrent une autre perspective pour les modèles de classification. Nous avons constitué deux modèles qui vont effectuer la même tâche : la classfication de la parole c'est à dire si elle est pathologique ou non. Nous avons anticipé que ces modèles pourraient éprouver des difficultés particulières dans la catégorisation des patients présentant une sévérité modérée. De plus le logatome *wi* pourrait poser problème en raison de sa complexité, tant chez les contrôles que chez les malades.

# 2 Corpus

## 2.1 Nature des données

*Au vu du caractère pathologique du corpus, aucun audio n'a été mis sur internet. Tout a été fait en local.*
Le corpus utilisé au sein de ce projet est un corpus audio comprenant trois logatomes : *aj*, *uj*, *wi* que nous a prêté Mme Fougeron. Les locuteurs produisent ces logatomes sous la forme de trois répétitions : *ajajaj*, *ujujuj*, *wiwiwi* sans pause. Il provient du projet *MonPage* et est composé de 33 locuteurs atteints de la pathologie avec trois types de sévérités ( peu atteints, moyennement et très sévère) et 33 contrôles. Chaque locuteur a produit ces logatomes 3 à 4 fois. Le corpus total a 768 données audio. 376 pour les malades et 392 pour les personnes sans maladies.

# 3 Préparation des données

## 3.1 Annotation Orthographique et Phonémique

Pour la réalisation de la classification, une annotation manuelle orthographique a été réalisée sur l'intégralité des données. Par la suite, les données avec leurs Textgrids ont été passées au système d'alignement forcé MFA pour obtenir les transcriptions phonémiques. Cependant, en raison du caractère pathologique des données, une révision manuelle de l'annotation phonémique a été nécessaire. L'aligneur avait tendance à couper le dernier phonème avant la fin de la production.

## 3.2 Transformation données pour la Classification

Pour nos classifications, deux types de données ont été choisis comme entrée. Il a été jugé important de maintenir la même durée d'extraction pour les deux types d'entrées afin de les rendre comparables :

1. La modulation cepstrale de l'entièreté du logatome.

2. Le spectrogramme de l'entièreté du logatome.

### 3.3 Modulation Cepstrale

Après l'obtention de l'annotation phonémique et orthographique, nous avons utilisé la fonction 'get mfcc' présente dans le notebook prétraitement. Elle a été utilisée pour obtenir la modulation cepstrale. Ce script n'utilise pas l'annotation mais génère un fichier texte synchronisé avec les fichiers .wav. Ce fichier texte est visible sur praat à l'aide de la fonction lecture, elle-même présente sur le notebook de prétraitement.

### 3.4 Extraction de bande pour la modulation cepstrale

L'utilisation de la modulation cepstrale nous a placé au coeur d'une reflexion importante. Mettre uniquement la courbe avec le fond blanc risquerait de produire trop de bruits pour le modèle. Il pourrait être biaisé par cet amas de blanc en fond. La modulation cepstrale a donc été reproduite sous forme d'une bande de pixele montrant l'intensité de la courbe, l'échelle de couleur pour son intensité a été normalisée sur la totalité du corpus. L'avantage en plus de cette bande de pixel est que la courbe sera plus facile à lire pour les modèles. Cette bande a été extraite à l'aide de python. Nous avons demandé à ce qu'il représente cette bande sur la totalité du logatome , étant donnée que ce sont des logatomes relativement courts nous dépassons rarement les 3-4 secondes.

### 3.5 Extractions des spectrogrammes

L'extraction des spectrogrammes a été faites à l'aide du module python parselmouth. Celui-ci nous a permis d'obtenir un spectrogramme très ressemblant à celui de praat.

### 3.6 Répartitions corpus train et test

Nous avons donc un total de 176 données en test et de 592 données en entraînement. Le choix de la répartition dans le corpus a été fait de manière manuelle et réflechie en fonction des métadonnées. Le nombre de locuteurs malades et contrôles a été réparti de manière équitable, de même pour le sexe. Il a été important d'avoir un corpus équilibré et symétrique.

## 4 Modèles

Nous avons choisi de comparer deux modèles : *Support Vector Machine* et *Réseau de neurones convolutif*. Cela va nous permettre de voir lequel classe le mieux et avec quel input. Nous avons été curieuses de voir si un modèle comme le SVM pouvait concurrencer un CNN dans le milieu de l'image. Afin de garantir une comparaison équitable, les deux modèles ont été entraînés sur le même corpus d'entraînement et test.

### 4.1 Réseau de Neurones Convolutifs

Vous retrouverez dans les notebooks 'cnn_modulation.ipynb' et 'cnn_spectro.ipynb' les choix d'hyperparamètres, leur justifications ainsi que de multiples essais.

#### 4.1.1 Avec modulation cepstrales

Nous obtenons un score de 91% en accuracy. Les erreurs du modèle sont ciblées sur les locuteurs malades ayant un score de sévérité très bas. Ils sont donc confondus avec les contrôles. De plus, le

logatome *wi* se révèle particulièrement difficile à discerner pour les locuteurs contrôles , ce qui peut expliquer cette moins bonne classification par rapport aux logatomes *uj* et *aj*.

#### 4.1.2 Avec spectrogrammes

Les spectrogrammes se sont montrés moins performants : le modèle est à 75 % d'accuracy avec les même paramètres que celui avec la modulation cepstrale. Les locuteurs mal classés sont une nouvelle fois ceux avec une sévérité de la maladie basse. Les logatomes mal classés sont tout d'abord *wi* une nouvelle fois puis *aj*. Le logatome *aj* était l'un des mieux classé sur la modulation cepstrale. Cela veut donc dire que les deux modèles n'ont pas la même préférence de logatomes. Cette différence est très probablement explicable au vu de la différence des deux entrées.

### 4.2 Support Vector Machine

Vous retrouverez dans les notebooks 'cnn_svm_modulation.ipynb' et 'cnn_svm_spectro.ipynb' les choix d'hyperparamètres, leur justifications ainsi que de multiples essais.

#### 4.2.1 Avec modulations cepstrales

Nous obtenons avec le modèle ayant la modulation cepstrale comme entrée 80% d'accuracy. Néanmoins il a beaucoup plus de difficulté pour classer les locuteurs malades par rapport aux locuteurs contrôles. Les logatomes ayant le plus d'erreur ne sont pas les mêmes de même pour les locuteurs malades qui moyennement sévère qui sont mal classés. Le logatome le moins bien classé est *aj* suivi de *uj*.

#### 4.2.2 Avec spectrogrammes

Le modèle s'en sort beaucoup moins avec les spectrogrammes comme entrée. Nous sommes à 65% d'accuracy. Nous pouvons observer sur la matrice de confusion une moins bonne discrimination pour les personnes malades. Néanmoins, nous avons une mauvaise classification pour les deux classes tout de même. Les logatomes mal classés sont tous mal classé de manière équitable, et *aj* sort légèrement du lot.

## 5 Conclusion

En conclusion, il apparaît que le réseau de neurones surpasse le SVM avec les deux types d'entrée. Cependant, le SVM démontre une performance honorable dans la classification des locutions des patients lorsqu'il prend en entrée la modulation.

Les résultats révèlent des similarités dans les difficultés de classification rencontrées par les deux modèles : nous avons pu en effet relever cette difficulté récurrente à classer *wi*. Cette difficulté peut amemner vers une discussion intéréressante : quel logatome donner à ces modèles pour les perfectionner par exemple. Neanmoins, il est important de souligner que pour l'intégralité des modèles présentés ici, nous avons une classifation correcte pour tout les patients très atteints. Par la suite, nous pourrions également étudier certains phonèmes ou syllabes à l'aide de la transcription phonémique : cela enrichirait les données. Une une procédure de type leave one out afin de tester tous les locuteurs pourrait également être envisageable pour tester la robustesse du modèle.

Par ailleurs, les modulations cepstrales se démarquent comme des données plus efficaces que les spectrogrammes, indépendamment du modèle utilisé. Cette constatation met en évidence que les informations contenues dans les modulations cepstrales offrent des indices plus discriminants pour la classification

des discours ataxiques. En contraste, les spectrogrammes, bien que informatifs, présentent davantage de bruit et montrent une performance légèrement inférieure dans cette tâche spécifique. Néanmoins il serait intéressant de tester des spectrogrammes ciblés sur un phonème ou bien une catégorie de phonèmes. Prendre la totalité du spectrogramme lui a peut être fait défaut.

Vous trouverez sur *Github* les notebooks reprenant la totalité du travail fait ici avec davantage de précision. `https://github.com/NataTCHA/Projet-CNN`