UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À

L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE

APPLIQUÉES

PAR
Aymen BEN MESSAOUD

UTILISATION DES SIGNAUX DU CERVEAU (EEG) ET VOCAUX POUR LA DÉTECTION ET LE MONITORING DES FACULTÉS D'UNE PERSONNE

AVRIL 2020

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

Sommaire

La conduite avec les facultés affaiblies par l'alcool ou les drogues demeure la principale cause des décès sur les routes. Avec la légalisation du cannabis au Canada, l'état continu toujours de s'attaquer intensément à la conduite avec les facultés affaiblies par l'alcool et aussi par le cannabis, dans le but d'assurer la sécurité des conducteurs automobiles.

Des règlements et des mesures sévères ont été appliqué depuis cette légalisation telles que la tolérance zéro pour les nouveaux conducteurs et la suspension du permis de conduire.

Malgré le risque élevé, la conduite avec facultés affaiblies par l'alcool demeure un comportement répandu au Québec. À la suite d'un sondage publié en 2012 par la fondation de recherche sur les blessures de la route, 17.6% des québécois ont avoué avoir conduit après la consommation d'alcool au cours des 30 derniers jours et 5.7% ont admis avoir conduit en dépassant la limite permise au code de la sécurité routières au cours des 12 derniers mois [1].

Le comportement de la conduite en facultés affaiblies a poussé le gouvernement à mettre en œuvre des tests de dépistage déployés par des policiers formés à passer ces tests sur des personnes soupçonnées de conduire avec des facultés affaiblies.

Le policier peut soumettre une personne soupçonnée à des épreuves de coordination des mouvements (comme le test du nystagmus horizontal effectué sur les yeux, le test de la promenade et tour pour vérifier la démarche en ligne droite et retourner ou un test de l'équilibre pour se tenir sur un pied) ou de passer un test salivaire.

Les tests de dépistage déployés par les policiers permettant de déterminer si la personne est en facultés affaiblies se basent sur le taux d'alcool, de la drogue ou de certains médicaments dans le sang, dans l'urine ou la salive plutôt que la capacité de jugement ou la réflexion engendrés par la consommation.

Cependant les résultats de ces tests sont assujettis à l'interprétation humaine et il y aura certains cas où cette interprétation peut échouer.

De nos jours, les avancées technologiques et les neurosciences travaillent à déterminer avec une grande précision, si une personne sujette à certains tests, est en facultés affaiblies ou non. Et ce en analysant l'habilité de conduite à un instant précis sachant que les effets de l'alcool ou du cannabis varient d'une personne à l'autre.

L'objectif de cette recherche est d'enregistrer et ensuite analyser des signaux EEG et vocaux d'une personne effectuant des tâches demandant un certain niveau de cognition comme la conduite automobile. En premier lieu, une capture de signaux vocaux et EEG est effectuée sur des personnes effectuant des tâches cognitives, avant et après la consommation d'alcool ou de cannabis.

Dans un second temps, une analyse sera effectuée sur les signaux vocaux capturés, permettant d'extraire les caractéristiques requises pour faire la distinction entre un discours normal associé à un état de fatigue ou de facultés normales ou celui dont l'articulation des mots est déficiente, situation pouvant découler de fatigue ou de facultés affaiblies en générale.

En troisième lieu, l'ensemble des caractéristiques déduites des signaux EEG et vocaux seront aussi analysés pour ainsi en extraire les caractéristiques les plus discriminantes et qui permettront d'évaluer optimalement les capacités d'une personne à effectuer ces types de tâches cognitives dans des contextes normaux et de facultés affaiblies.

Afin de pouvoir observer les déformations de la voix et les fonctions du cerveaux influencées par celles-ci, nous avons mené notre recherche en nous basant sur un ensemble de mots à prononcer avec une certaine complexité afin de pouvoir détecter les déformations linguistiques ou de la parole tout en capturant les signaux EEG de certaines zones actives du cerveau lors de la production de la parole.

D'après nos observations et d'après d'autres recherches, nous avons pu constater que la consommation de cannabis, impacte la concentration et la mémoire à court terme et diminue la capacité de réflexion et augmente le temps de réaction. Le test de prononciation des mots, affectué d'une façon aléatoire aux personnes sujettes à cette expérience, nous a prouvé que la diminution de la capacité de réflexion est observée après chaque consommation de cannabis.

Des travaux antérieurs sur le traitement de la parole et du son ont démontré la capacité d'obtenir d'excellentes performances lors de l'apprentissage direct à partir de formes d'ondes audio utilisant des réseaux de neurones convolutionnels (CNN).

Dans ce mémoire, nous analysons et expliquons théoriquement comment les CNN apprennent en profondeur à partir de formes d'onde audio brutes et identifions les limites potentielles des structures de réseaux existants. Sur la base de cette analyse, nous proposons en outre une nouvelle architecture de réseau de neurones, qui offre une structure simple mais concise et une interprétabilité élevée de ce modèle.

Les enregistrements vocaux ont été capturés à l'aide d'un casque Yeti pro, l'analyse de ces signaux vocaux a été faite par les logiciels Praat et Matlab et un modèle d'apprentissage machine basé sur le modèle Inception V3 de Google dont nous avons modifié sa dernière couche pour créer notre propre modèle sur le système Machine Learning Tensorflow.

Néanmoins, la prise de signaux EEG a été faite par un logiciel d'interface machine OpenVibe, l'analyse de ces signaux a été effectuée avec le logiciel MATLAB et un modèle d'apprentissage machine supervisé déployé sur une interface machine Azure Machine Learning pour faire la classification de ces signaux EEG et faire le monitoring des facultés d'une personne.

En conséquence, pour mieux quantifier les facultés de conduire d'une personne, nous pouvons énoncer un autre test de dépistage basé sur l'analyse des signaux vocaux et EEG avec une précision élevée.

Les résultats de notre recherche sont prometteurs ce qui permettra d'ouvrir des perspectives sur un nouveau test proposé aux agents de police pour leurs méthodes de dépistage des conducteurs soupçonnés de conduire après avoir consommé de la drogue ou de l'alcool.

ABSTRACT

Driving while impaired by alcohol or drugs remains the leading cause of traffic fatalities. With the legalization of cannabis in Canada, the state has consistently tackled impaired driving by alcohol and even cannabis, to keep drivers safe.

Strict regulations and measures have been applied since this legalization, such as zero tolerance for new drivers and the suspension of driving licenses.

Despite the high risk, alcohol-impaired driving remains a common behaviour in Quebec. Following a survey published in 2012 by the research foundation on road injuries, 17.6% of Quebecers admitted having driven after drinking alcohol in the past 30 days and 5.7% admitted having driven by exceeding the limit allowed in the road safety code during the last 12 months [1].

The behavior of impaired driving has prompted the government to implement screening tests by police officers trained to apply these tests on people suspected of driving under the influence.

Police officers can subject a suspected person to movement coordination tests (such as the horizontal gaze nystagmus eye test, the walk and turn test for walking on a straight line and turning, or the balance test for standing on one foot) or take a saliva test. Police tests to determine if the person is impaired are based on the level of alcohol, drugs or certain medications in the blood, urine or saliva rather than the ability of judgment or reflection altered by the consumption. However, the results of these tests are subject to human interpretation and there will be few cases where interpretation may fail.

Nowadays, technological advances and neuroscience are working to determine with great precision, whether a person subjected to specific tests, is impaired or not. By analyzing driving ability at a specific time because the effects of alcohol or cannabis can vary from person to person.

The objective of this research is to record and then to analyze the EEG and voice signals of a person performing tasks requiring a certain level of cognition such as driving. First,

voice and EEG signals are captured on people before and after consuming alcohol or cannabis.

Secondly, an analysis will be carried out, making it possible to extract the characteristics required to distinguish between normal speech associated with a state of rest or normal faculties and whose word articulation is deficient, a situation which may arise from fatigue or impairement in general.

Thirdly, the set of characteristics deduced from these EEG and vocal signals will also be analyzed in order to extract their most discriminating characteristics and which will make it possible to optimally assess a person's capacities to perform these types of cognitive tasks in normal and impaired contexts.

In order to be able to observe and analyse the deformation of the voice and the brain functions (EEG signals), we conducted our research based on a set of words to be pronounced according to a certain complexity in order to be able to detect linguistic or speech deformations by capturing EEG signals from certain active areas during speech production.

Based on our observations and other research, we have note that cannabis use affects concentration and short-term memory and shortens thinking time. The word pronunciation test, randomly assigned to people subject to this experience, showed that the reduction of thinking is present after each use of cannabis.

Previous work on speech and sound processing has demonstrated the ability to achieve excellent performance when learning directly from audio waveforms using convolutional neural networks (CNN). However, the exact internal functioning of a CNN network remains unclear, which prevents any further development and any improvement in this direction. In this dissertation, we analyze and theoretically explain how CNNs learn in depth from raw audio waveforms and identify the potential limits of existing network structures. Based on this analysis, we also propose a new network architecture that we used, which offers a very simple but concise structure and high interpretability of the model.

The voice recordings were taken using a Yeti pro headset, the analysis of these voice signals was performed by Praat and Matlab software and a machine learning model based on Google's Inception V3 model where its last layer was modified to create our own model on the Machine Learning Tensorflow system.

However, the EEG signals were taken by OpenVibe machine interface software, the analysis of these signals was carried out with MATLAB software and a supervised machine learning model deployed on an Azure Machine Learning machine interface to make classification of these EEG signals and monitoring of a person's faculties.

Therefore, to better quantify a person's driving ability, we can design another screening test based on the analysis of voice signals and EEG with high precision.

The results of our research are promising, which will open up perspectives on a new test proposed to police officers for their methods of detecting drivers suspected of driving after using drugs or alcohol.

REMERCIEMENT

Au terme de la rédaction de ce mémoire de recherche, je tiens à remercier tous ceux qui ont contribué, aussi bien par leur patience morale ou pratique, à élaborer et au déroulement de cette recherche dans les meilleures conditions.

Je tiens, en particulier, à remercier et à exprimer ma profonde et respectueuse gratitude ainsi que ma reconnaissance à mon directeur de recherche Monsieur François Meunier professeur à l'Université de Québec à Trois Rivières qui, en dépit de ses responsabilités, était toujours présent par ses directives fructueuses et ses encouragements me guidant tout au long de l'élaboration de ce projet avec sa patience et son savoir-faire qui le caractérisent.

Je le remercie infiniment pour son aide matériel et moral et qui s'est toujours montré coopératif et dévoué par son aide continu et ses conseils rigoureux, et qui, à tout moment, m'a soutenu moralement tout au long de ce mémoire de recherche.

Je profite de cette occasion pour exprimer ma reconnaissance à tous les professeurs qui m'ont enseigné durant mon cursus d'études supérieures et m'ont aidé d'une manière directe ou indirecte à trouver ma voie et m'épanouir. Je tiens aussi à remercier mes amis et les personnes qui se sont portées volontaires et qui grâce à eux j'ai pu construire une riche base de connaissances.

Enfin je ne peux pas clore cette rubrique sans exprimer ma gratitude aux membres de jury qui ont pris la peine d'évaluer ce modeste travail en faisant preuve d'attention et de patience.

DÉDICACE

Je dédie ce travail à ma chère Mère.

En témoignage de ma profonde gratitude, mon incontestable reconnaissance, pour la confiance qu'elle m'accorde et l'amour dont elle m'entoure.

À mon cher père.

Grâce à son encouragement, sa confiance et son soutien moral et matériel et pour son amour infini en exprimant mes gratitudes, mon profond amour et ma passion.

À toute ma famille, à mes amis, chaque personne qui a cru en moi et à tous ceux qui me sont chers.

À ceux qui ne sont plus là et à ceux qui l'ont toujours été.

Que Dieu vous Garde!

Table des matières

Sommaire ii

ABSTRACT v

REMERCIEMENT viii

DÉDICACE ix

Table des matières x

Liste des figures xiv

Chapitre 1 : Introduction 1

1.1 Divers faits saillants 1

1.2 Objectifs et hypothèses 2

Chapitre 2 : Revue de littérature 4

2.1 Introduction 4

2.2 Effet du cannabis et de l'alcool sur le cerveau humain, la voix et la conduite automobile 4

2.2.1 Effet du cannabis et de l'alcool sur le cerveau 4

2.2.2 Effet du cannabis et de l'alcool sur la voix humaine 6

2.2.3 Effet du cannabis et de l'alcool sur la conduite routière 6

2.3 L'électroencéphalographie EEG 7

2.3.1 Principe et fonction de signaux EEG 7

2.3.2 Bandes de fréquences-ondes cérébrales 10

2.3.3 Structure et fonction des neurones 12

2.4 La parole humaine 13

2.4.1 Les moyens d'investigation de la parole 17

2.4.2 Techniques expérimentales d'analyse de la voix 18

2.4.2.1 Analyse des gestes respiratoires 19

2.4.2.2 Analyse des gestes phonatoires 20

2.4.2.3 Analyse des gestes articulatoires 20

2.4.2.4 Analyse acoustique de la voix 21

2.5 Aires de langage dans le cerveau humain 23

2.6 Données EEG 25

2.6.1 Traitement de signal 26

2.6.1.1 Signaux EEG bruités 26

2.6.1.2 Analyse en composantes indépendantes des données (ICA) 32

2.6.2 Transformée de Fourier 36

2.6.3 Transformée en ondelettes 37
2.6.4 Densité spectrale de puissance (PSD) 39
2.7 Différentes approches de classification de signaux EEG 41
2.7.1 Interface cerveau-ordinateur - vue d'ensemble 41
2.7.2 Réseau de neurones artificiel et Perceptron multicouche 42
2.7.3 Algorithme de retro propagation 43
2.7.4 SVM (support vector machine) 44
2.8 La communication vocale 46
2.8.1 Mécanisme de production de la parole 46
2.8.2 Mécanisme de perception de la parole 46
2.8.3 Classification des sons de la parole 47
2.8.3.1 Sons Sonores 47
2.8.3.2 Sons non vocaux 47
2.8.3.3 Silence 48
2.9 La parole et les sons du français 48
2.9.1 Les voyelles 49
2.9.2 Les consonnes 50
2.10 Dysarthrie et problème d'élocution 51
2.11 La représentation de la parole dans les domaines temps et fréquence 53
2.11.1 Le spectrogramme 55
2.12 Les différentes approches de classification de la parole 57
2.12.1 PLP Perceptual Linear Prediction 57
2.12.2 LPC Linear Predictive Analysis 58
2.12.3 LPCC Linear Predictive Cepstral Coefficient 59
2.12.4 MFCC Mel Frequency Cepstral Coefficient 60
2.12.5 FFT Power Spectral Analysis 60
2.12.6 MEL Mel Scale Cepstral Analysis 61
2.13 Conclusion 62
Chapitre 3 : Méthodes expérimentales 63
3.1 Introduction 63
3.2 Casque Multi Chanel Emotiv Epoc + ... 64
3.3 Blue Yeti Pro 66
3.4 Audacity 67
3.5 OpenVibe 68
3.5.1 Acquisition de données EEG 69

3.5.2 Enregistrement de signaux EEG 72
3.5.3 EEGLAB 72
3.5.4 Rejet d'artéfacts avec ICA 74
3.6 Extraction des caractéristiques en utilisant BrainStorm 76
3.6.1 La transformation en ondelettes 78
3.7 Classification des signaux EEG 80
3.7.1 Azure Machine Learning studio 80
3.8 Enregistrement de signaux audio 82
3.8.1 Pré-Traitement des signaux vocaux 82
3.8.2 Matlab 83
3.9 Classification des signaux vocaux 86
3.9.1 Réseau de neurones Convolutif Profond 87
3.9.2 Couche convolutionnelle (Convolutional Layer) 88
3.9.3 Pooling Layer 90
3.9.4 Fonction d'activation (Activation Function) 90
3.9.5 Normalisation des lots 91
3.9.6 Dropout 91
3.9.7 Couche entièrement connectée 92
3.9.8 Google Inception Model 92
3.9.9 Classification audio avec Tensorflow 93
3.10 Conclusion 93
Chapitre 4 Résultats et discussions 95
4.1 Introduction 95
4.2 Élimination des bruits du signal audio 95
4.3 Détection des voyelles 96
4.4 Automatisation de la classification des signaux Audio 101
4.4.1 Préparation des données 101
4.4.2 Le réentraînement du bottleneck et l'affinement du modèle 104
4.4.3 Tester le modèle avec des nouvelles entrées audios 109
4.5 Automatisation de la classification de signaux EEG 110
4.5.1 Puissance de signaux EEG 110
4.5.2 Réseau de neurones artificiels 112
4.5.3 Support vector machine 113
4.5.4 Résultat de l'entraînement du modèle 113
4.5.6 Tester le modèle avec des nouvelles entrées EEG 114

4.6 Conclusion **115**
Chapitre 5 Conclusion **117**
Annexe **119**
Références **120**

Liste des figures

Figure 1 Casque traditionnel avec plusieurs électrodes (à gauche), et à droite un nouveau type de casque à 14 électrodes (Emotiv epoc +) [47] 8
Figure 2 Schéma des différentes zones du cerveau et leur rôle [48] 9
Figure 3 Positionnement et identification standards des électrodes EEG selon le système international 10-20 [49] 10
Figure 4 Les ondes associées aux bandes de fréquences des signaux EEG [50] 10
Figure 5 Structure d'un neurone [51] 13
Figure 6 Variation du flux d'air circulant dans la glotte [52] 15
Figure 7 Période des vibrations sonores [52] 16
Figure 8 Organes articulatoires 17
Figure 9 Source de compréhension du langage dans le cerveau humain [61] 24
Figure 10 Estimation du clignement des yeux avec l'algorithme DANSE [62] 26
Figure 11 Exemples de modèles d'artéfacts IC similaires malgré les différentes configurations de canaux EEG. Première rangée: images IC d'une configuration à 25 canaux; deuxième rangée: images IC d'une configuration à 30 canaux 34
Figure 12 Quinze secondes de données EEG sur 9 canaux (sur 100) du cuir chevelu (panneau supérieur) avec activités de 9 (sur 100) composants indépendants (CI, panneau inférieur) 35
Figure 13 Fonction de base de l'ondelette Morlet [63] 39
Figure 14 Densité du spectre de puissance de l'onde thêta d'une personne normale (rouge) et d'une personne alcoolique (bleue), l'axe vertical représente la puissance et l'axe horizontal représente le temps en miliseconde [63] 40
Figure 15 Densité du spectre de puissance de l'onde beta d'une personne normale (rouge) et d'une personne alcoolique (bleue), l'axe vertical représente la puissance et l'axe horizontal représente le temps en milliseconde [63] 40
Figure 16 Neurone dans un réseau de propagation à réaction. f est la fonction d'activation [63] 43
Figure 17 Illustration d'un classificateur de machine à vecteurs de support [63] 45
Figure 18 Mécanisme de production et de perception de la parole [66] 47
Figure 19 Forme d'onde d'amplitude du signal de parole étiqueté avec les régions V (voisé), U (non voisé) et S (silence) [65] 48
Figure 20 Trapèze vocalique représentant un classement des voyelles orales des langues du monde selon des critères articulatoires 49
Figure 21 Valeurs formantiques des voyelles a, i, u (en haut) et coupe sagittale du tractus vocal (en bas) avec les différentes positions de la langue pour la production de ces trois voyelles, les plus distinctes du point de vue articulatoire [70] 50
Figure 22 Signal de parole, s [n] pour l'énoncé « l'oublierons-nous jamais » [78] ... 54
Figure 23 Segments de niveau phonème du signal pour «will» comme «w», «ih» et «l» [77] 55

Figure 24 Spectrogramme à échelle logarithmique du signal de parole utilisant une taille de fenêtre de 30 ms et une taille de saut de 7,5 ms 56
Figure 25 Schéma fonctionnel du processus LPCC [27] 59
Figure 26 Schéma fonctionnel du processus MFCC [27] 60
Figure 27 Schéma de la méthodologie proposée pour la classification de signaux vocaux et de signaux EEG 64
Figure 28 Emplacement des 14 électrodes du Emotiv Epoc+ sur le cuir chevelu 65
Figure 29 Qualité de signal des 14 électrodes sur Emotiv Xavier Control Panel 66
Figure 30 Microphone d'enregistrement Audio Blue Yeti Pro 67
Figure 31 Interface du logiciel Audacity et exemple de signal vocal enregistré 68
Figure 32 Workflow de OpenVibe 69
Figure 33 OpenVibe acquisition Server 70
Figure 34 Scénario d'acquisition de signaux EEG 71
Figure 35 Un fichier CSV contenant des signaux EEG bruts 72
Figure 36 Interface utilisateur de EEGLAB 73
Figure 37 Signal EEG Brut d'un sujet en état de facultés affaiblies 73
Figure 38 Algorithme Runica pour l'analyse en composantes indépendantes de EEGLAB 74
Figure 39 Rejet d'artéfacts en utilisant la décomposition en composantes indépendantes sur un signal EEG en facultés affaiblies 75
Figure 40 Composant lié à un artéfact dû au clignement d'œil 76
Figure 41 La densité de la puissance spectrale d'un sujet en état normal sur les différentes bandes de fréquences 77
Figure 42 densité de la puissance spectrale d'un sujet en état de facultés affaiblies sur les différentes bandes de fréquences 78
Figure 43 Transformée en ondelette du canal F3 d'un sujet en état normal 79
Figure 44 Transformée en ondelette du canal F3 d'un sujet en état facultés affaiblies 79
Figure 45 Données d'entraînement de sujets en état normal 81
Figure 46 Données d'entraînement de sujets en état de facultés affaiblies 81
Figure 47 Élimination de bruit d'un signal vocal audio 82
Figure 48 Interface utilisateur de Praat 83
Figure 49 Représentation graphique du mot Abattu prononcé par un sujet Femme en état normal (signal Vert) et en état de facultés affaiblies (signal Rouge) 84
Figure 50 Représentation graphique du mot Abattu prononcé par un sujet Homme en état normal (signal Vert) et en état de facultés affaiblies (signal Rouge) 85
Figure 51 Spectrogramme du mot Abattu d'un sujet en état Normal réalisé avec Matlab 85
Figure 52 Spectrogramme du mot Abattu d'un sujet en état de facultés affaiblies réalisé avec Matlab 86
Figure 53 Architecture d'un réseau de neurones convolutionnel [27] 88
Figure 54 Enregistrement audio avant et après l'élimiation de bruits sur Audacity 95
Figure 55 Formants de mot Abattu prononcé par le sujet A en état normal 97

Figure 56 Formants de voyelles du mot Abattu prononcé par le sujet A en état de facultés affaiblies 97
Figure 57 Formant de la voyelle A extrait du mot Abattu sur Praat 98
Figure 58 Caractéristiques de formants liées aux voyelles et leurs amplitudes 99
Figure 59 Formants de la deuxième voyelle A du mot Abattu 99
Figure 60 Caractéristiques de formants liées à la seconde voyelle A et leurs amplitudes 99
Figure 61 Répartition des données audios en 2 dossiers 101
Figure 62 Fonction de segmentation de fichier audio en segments de 0,5 seconde et génération de spectrogrammes sur ces segments audios 102
Figure 63 Segmentation de données audios en segments 0.5 seconde 102
Figure 64 Fonction de création de spectrogramme à partir de fichiers contenant des segments audio de 0,5 seconde 103
Figure 65 Génération de spectrogrammes 104
Figure 66 Génération de spectrogrammes et comparaison entre un spectrogramme à l'état normal (à droite) et un autre à l'état facultés affaiblies (à gauche) 104
Figure 67 Fonction de création de liste d'images spectrogrammes pour l'entraînement du modèle 106
Figure 68 Fonction de distorsion de données 107
Figure 69 Entraînement du modèle pour la classification des signaux vocaux 109
Figure 70 Résultat de test sur un nouvel enregistrement audio du sujet F en état de facultés affaiblies 110
Figure 71 Résultat de test sur un nouvel enregistrement audio du Sujet G en facultés affaiblies 110
Figure 72 Résultat de test sur un nouvel enregistrement audio du Sujet H à l'état normal 110
Figure 73 Densité de puissance spectrale en état normal sur différentes bandes de fréquences de Sujet A 111
Figure 74 Densité de puissance spectrale en état facultés affaiblies sur différentes bandes de fréquences de Sujet A 112
Figure 75 Modèle de classification utilisant l'algorithme de réseau de neurones et SVM 113
Figure 76 Courbe ROC des résultats de classification utilisant l'algorithme de réseau de neurones. 114
Figure 77 Résultats de précision et d'exactitude de la classification de signaux EEG 114
Figure 78 Résultat de classification de signaux EEG du sujet F 115

Liste des tableaux

Tableau 1 Amplitude de formants de voyelles du sujet A en état normal. 100
Tableau 2 Amplitude de formants de voyelles du sujet A en état facultés affaiblies 100
Tableau 3 Effets rapportés de la consommation aiguë de cannabis sur les paramètres nécessaires lors de la conduite automobile 119

Chapitre 1 : Introduction

1.1 Divers faits saillants

La consommation récréative du cannabis a été légalisée au Canada le 17 octobre 2018 dans un marché strictement règlementé. Cette légalisation engendre nécessairement la mise en place de nouvelles lois et de nouvelles sanctions concernant la conduite avec les facultés affaiblies par le cannabis et autres drogues.

La conduite routière avec les facultés affaiblies dû à la consommation d'alcool ou des drogues a été l'une des principales causes de plusieurs accidents routiers durant les dernières années.

Selon un rapport de la fondation de recherche sur les blessures de la route au Canada mentionne que le pourcentage des accidents routiers mortels et après le test de dépistage de drogues est passé de 15.9% en 2000 à 20.9% en 2015 [1].

Selon l'enquête nationale sur le cannabis de statistique Canada, un consommateur de cannabis sur sept détenant un permis de conduire valide prend le volant dans les deux heures suivant sa consommation. Cette statistique est vraiment inquiétante parce que les canadiens doutent que les autorités policières disposent des moyens nécessaires pour identifier les conducteurs ayant les facultés affaiblies par le cannabis ou d'autres types de drogues.

La consommation du cannabis ou d'alcool est caractérisée par la perturbation du système nerveux central des consommateurs et altère leurs facultés et foncions. Dans une étude de François Meunier et Souhail Berdaa [63], un système a été développé permettant d'analyser et diagnostiquer les facultés affaiblies d'une personne suite à la consommation d'alcool ou de la drogue à partir de signaux EEG.

Le but principal de la consommation du cannabis à usage non médical appelée aussi non thérapeutique est l'effet d'euphorie appelé en anglais « high » engendré par la présence de THC qui est la substance psychoactive du Cannabis [2].

La consommation du cannabis affecte les fonctions cognitives qui sont les facultés du cerveau qui permettent de communiquer, d'analyser l'environnement et de se concentrer sur une tâche spécifique.

Toutefois, chaque consommateur de cannabis aura une diminution de la concentration et de l'attention, une faible capacité de jugement ainsi qu'un ralentissement du temps de réaction ce qui augmentera la probabilité d'avoir des accidents routiers en conduisant un véhicule [1].

En vertu du code criminel, les tests de dépistage se basent sur le taux d'alcool, de la drogue ou de certains médicaments dans le sang. Si le taux d'alcool est supérieur ou égal à 80 mg par 100 ml de sang (0.08g/100) le conducteur est considéré en facultés affaiblies et des sanctions sont appliquées [1].

D'autre part, si le taux de cannabis (THC) est dans l'intervalle de 2 à 5 nanogrammes de THC par millilitre de sang ou 2.5 nanogramme de THC par millilitre de sang combiné à un taux d'alcool de 50 milligrammes d'alcool par 100 millilitres de sang, le conducteur sera accusé de conduite avec les facultés affaiblies [1].

1.2 Objectifs et hypothèses

L'objectif de ce mémoire de recherche est de capturer des signaux vocaux et EEG sur des personnes ayant été sujettes à des tests expérimentaux avant et après la consommation du cannabis ou d'alcool. Ensuite, ces signaux sont analysés afin de détecter la déformation linguistique causée par la consommation dans le but de déterminer si une personne est en état de facultés affaiblies ou en état normal. Ce processus sera explicité dans les prochains chapitres.

Ce mémoire de recherche sera divisé en plusieurs parties, dans un premier temps, au chapitre 2, nous allons énumérer les effets du cannabis, de l'alcool sur le cerveau humain ainsi que l'effet de la consommation sur le cerveau, ensuite nous allons exposer les

différents tests de dépistages que les autorités utilisent pour détecter les facultés affaiblies d'une personne en contexte de conduite routière.

Après, nous allons faire une étude sur les recherches et les études connexes à notre mémoire dans les domaines de l'analyse de la voix et l'analyse des signaux EEG.

Par la suite, nous expliquerons les tests de production de la parole et les tests effectués pour la détection de la dysarthrie et pourquoi nous avons opté pour cette approche comme base expérimentale de ce mémoire. Postérieurement, au chapitre 3, nous allons décrire la méthodologie de recherche que nous avons suivie tout au long de ce travail expérimental. pour résoudre notre problématique de recherche. Au chapitre 4, nous présentons les résultats de nos expérimentations et commentons ces résultats obtenus, leur précision et fiabilité.

Au chapitre 5, nous allons faire une conclusion afin d'exprimer nos jugements et avis sur la problématique de ce mémoire et nous allons présenter un sommaire des méthodes utilisées pour la capture des signaux vocaux et EEG, les analyser, et les classifier pour pouvoir détecter automatiquement un discours associé à un état de facultés affaiblies, ainsi que la déficience de l'articulation de quelques mots. Les perspectives d'avenir de ce mémoire seront aussi exposées dans ce chapitre.

Chapitre 2 : Revue de littérature

2.1 Introduction

Dans ce chapitre, nous allons dans un premier temps exposer le contexte général de notre mémoire de recherche traitant des effets du cannabis et de l'alcool sur le cerveau humain, la voix humaine ainsi que la conduite automobile. En outre, nous allons définir les concepts de base que nous avons utilisés dans notre recherche.

Par la suite, nous allons donner un aperçu des différentes études qui ont traité de la classification des signaux vocaux et des ondes cérébrales (EEG).

2.2 Effet du cannabis et de l'alcool sur le cerveau humain, la voix et la conduite automobile

Dans cette section, nous allons discuter de l'effet du cannabis et de l'alcool sur les comportements de conduite d'un véhicule automobile.

2.2.1 Effet du cannabis et de l'alcool sur le cerveau

Dans cette section nous allons traiter les effets du cannabis et de l'alcool sur les différentes zones du cerveau humain et expliquer le danger de la conduite automobile en facultés affaiblies et la déficience de la cognition.

L'alcool est considéré comme substance dépressante ayant comme principal effet de ralentir les fonctions du cerveau. Après chaque consommation d'alcool, différentes parties du cerveau humain sont touchées par l'effet de l'alcool telles que :

Le Lobe frontal : L'alcool peut avoir comme effet la perte de la capacité décisionnelle, du raisonnement et de résolution de problème, ainsi que la perte du contrôle émotionnel de la coordination. L'alcool accentue les risques de réactions impulsives et réduit l'action de frein naturel tel que l'arrêt de l'envie de boire.

Le lobe pariétal : L'alcool ralentit la communication entre les neurotransmetteurs du cerveau et empêche l'information de passer de manière fluide comme lorsque nous sommes sobres d'où le ralentissement du temps de réaction.

Le lobe occipital : L'alcool réduit les capacités des pupilles qui ne peuvent pas se dilater ou se contracter selon la lumière de l'environnement ce qui provoque une vision floue ou double.

Le lobe temporal : La consommation de l'alcool provoque des problèmes d'élocution et des propos incohérents.

Le cervelet : La consommation de l'alcool a comme effet une mauvaise coordination des muscles et le déséquilibre.

Par ailleurs, le cannabis est la deuxième substance psychoactive la plus consommée au Canada après l'alcool.

La consommation de cette dernière a comme effet le ralentissement de la communication entre les neurones. Plusieurs parties du cerveau humain sont alors affectées par l'effet de la consommation du cannabis.

L'Amygdale : c'est une partie du cerveau humain qui a hérité de son nom à cause de sa forme semblable à celle d'une amande située dans la partie frontale du lobe temporal. Elle est primordiale pour notre capacité de sentir et percevoir des émotions. À la suite de la consommation du cannabis, la molécule de tétrahydrocannabinol appelée THC se fixe sur les récepteurs CB1 d'où la sensation d'euphorie légère, de relaxation et de perceptions auditives et visuelles amplifiées.

Le cervelet et le Ganglion basal : Les ganglions basaux sont fortement liés, à la fois anatomiquement et fonctionnellement à plusieurs régions corticales et au cervelet.
Ces deux parties du cerveau humain ont comme rôle d'assurer l'équilibre, le mouvement et la coordination. Étant donné que la présence de la molécule THC déstabilise le fonctionnement de ces deux parties, le consommateur aura une sensation de vertige comme effet secondaire indésirable à court terme ou sentir des étourdissements après s'être levé soudainement. Le consommateur devient désorienté avec une attention réduite. Pour cela, il ne faut jamais conduire après avoir consommé du cannabis.
Cortex cérébral et Hippocampes : Les effets cognitifs les plus fréquemment rapportés après la consommation de cannabis sont l'altération de l'apprentissage et de la mémoire

d'où nous pouvons expliquer le disfonctionnement à court terme du cortex cérébral et de l'hippocampe, deux parties du cerveau, responsables de la mémoire.

2.2.2 Effet du cannabis et de l'alcool sur la voix humaine

Les consommateurs de cannabis ont tendance à avoir la gorge sèche, ce phénomène de déshydratation au niveau de la gorge affecte les cordes vocales et pousse l'être humain à mettre plus d'efforts nécessaires pour produire un son.

À court terme, le consommateur exerce une pression supplémentaire sur les cordes vocales enflées ce qui pourra provoquer des problèmes d'élocution ou de mauvaise articulation.

2.2.3 Effet du cannabis et de l'alcool sur la conduite routière

Plusieurs études ont traité la relation entre la consommation de cannabis et l'incidence accrue d'accidents de la route ou de la conduite avec facultés affaiblies. Une étude a démontré un taux élevé de cas d'hospitalisations chez les consommateurs de cannabis pour des blessures de toutes causes comparé aux personnes qui ne consomment pas ou ne consomme plus [3].

Plusieurs études ont montré qu'il n'y a pas nécessairement de relation directe entre la capacité de conduire et la performance au volant et les concentrations plasmatiques de la molécule de THC, qui seraient non représentatives de celles détectées dans le cerveau, soit celles entraînant les effets d'altérations cognitives du THC. Toutefois, Fischer et al. dans une étude publiée en 2011, ont montré qu'une concentration plasmatique de THC de 7 à 10 ng/ml serait équivalente à 0,05 % d'éthanol sanguin en termes d'altération de la capacité à conduire une automobile [4].

Selon le centre canadien sur les dépendances et l'usage de substances ainsi que lors des études portant sur l'habilité de conduire, des tests psychomoteurs et cognitifs après une consommation aigue de cannabis chez certains consommateurs, une diminution de leurs performances a été remarquée. Ces effets étant maximaux, durant l'heure suivant la consommation de cannabis, alors qu'ils diminuent sur une période de temps qui varie entre 3 et 4 heures, pouvait aller jusqu'à 24 heures. Ces effets affectent l'exécution des tâches nécessitant plus de concentration.

Une autre étude réalisée à l'université McGill a montré que même après 5 heures d'inhalation d'une dose de cannabis inférieure à celle que nous trouvons dans un seul joint, le comportement et les facultés associés à la conduite automobile sur un simulateur de conduite s'altèrent une fois que le conducteur sera exposé à des distractions comme celles dans la vraie vie [5].

La conduite en état d'ébriété est responsable d'une grande partie des accidents de la route.
Plusieurs recherches ont étudié les effets de l'alcool sur la conduite automobile et ont déduit que ces derniers affectaient les conducteurs de nombreuses façons, notamment: l'attitude, la capacité de jugement, la vigilance, la perception, la réaction et le contrôle durant la conduite.

L'analyse du taux d'accident a montré qu'une alcoolémie élevée entraînera également un taux d'accident élevé. Les résultats ont montré aussi que la conduite en état de faculté affaiblie dû à la consommation d'alcool avec un taux d'alcoolémie plus élevé était plus facile à distinguer que la conduite normale [6].

Une étude réalisée par l'institut national de santé publique du Québec [109], sur les effets de la consommation aigue de cannabis sur les paramètres de la conduite automobile, a démontré que les paramètres de la conduite sont affectés suite à la consommation du cannabis. Cette étude a été réalisée sur un simulateur de conduite afin de faire des tests psychomoteurs de conduite en état de facultés affaiblies et observer les effets selon le contexte de consommation. Les résultats de cette étude sont présentés en annexe de ce mémoire de recherche.

2.3 L'électroencéphalographie EEG

2.3.1 Principe et fonction de signaux EEG

Le EEG pour électroencéphalographie, est un examen reposant sur la mesure de l'activité électrique du cerveau permettant une exploration cérébrale. Celui-ci est effectué grâce à des électrodes placées sur le cuir chevelu du patient (Voir figure 1).

L'électroencéphalographie est la représentation sous forme d'un tracé des variations dans le temps de l'activité électrique du cerveau appelé électroencéphalogramme. Généralement cette approche est utilisée afin de diagnostiquer certaines maladies telles que l'épilepsie, l'insomnie, différencier les différents types d'ondes cérébrale (alpha, bêta, thêta et delta) et même pour étudier le fonctionnement du cerveau d'une personne saine [7].

Comparable à l'électrocardiogramme (ECG) qui permet d'étudier le fonctionnement du cœur, l'EEG est un examen indolore et non-invasif qui renseigne sur l'activité neurophysiologique du cerveau au cours du temps et en particulier du cortex cérébral soit dans un but diagnostique en neurologie, soit dans la recherche en neurosciences cognitives.
Le signal électrique à la base de l'EEG est la résultante de la sommation des potentiels d'action post-synaptiques synchrones issus d'un grand nombre de neurones [8].

Figure 1 Casque traditionnel avec plusieurs électrodes (à gauche), et à droite un nouveau type de casque à 14 électrodes (Emotiv epoc +) [47]

Généralement, l'application principale de l'EEG est l'étude d'une manière manuelle des enregistrements, à étudier les relations des liaisons ou des troubles cérébraux, tels que l'épilepsie rapporté par Sundaram, Sadler, Young, & Pillay en 1999 [110].

En 1875, Richard Caton détecta la présence de courants électriques chez les lapins et les singes à la surface du cerveau confirmée par l'occurrence d'oscillations de l'aiguille d'un galvanomètre.

Hans Berger appliqua à l'homme cette approche et enregistra le premier EEG en 1929 sous variations permanentes de potentiel, enregistrées avec des électrodes impolarisables sur un crâne intact. Berger cherchait à matérialiser l'énergie physique de l'être humain afin d'identifier le support de la pensée. Cette expérience mena alors à la découverte de l'EEG chez l'être humain (Millet, D. 2001) [9].

Les bases de certains examens diagnostiques en neurologie ont été établies depuis 1945 et sont toujours appliquées de nos jours, avec l'évolution technologique, les amplificateurs EEG peuvent supporter plusieurs canaux allant à 512 canaux sur certains appareils.

Les enregistrements effectués sont généralement basés sur une carte cérébrale comme présentée dans la figure 2 à l'aide des électrodes fabriquées en argent ou en chlorure d'argent placées selon des systèmes de positionnement standards afin de pouvoir visualiser l'activité cérébrale (Voir figure 3).

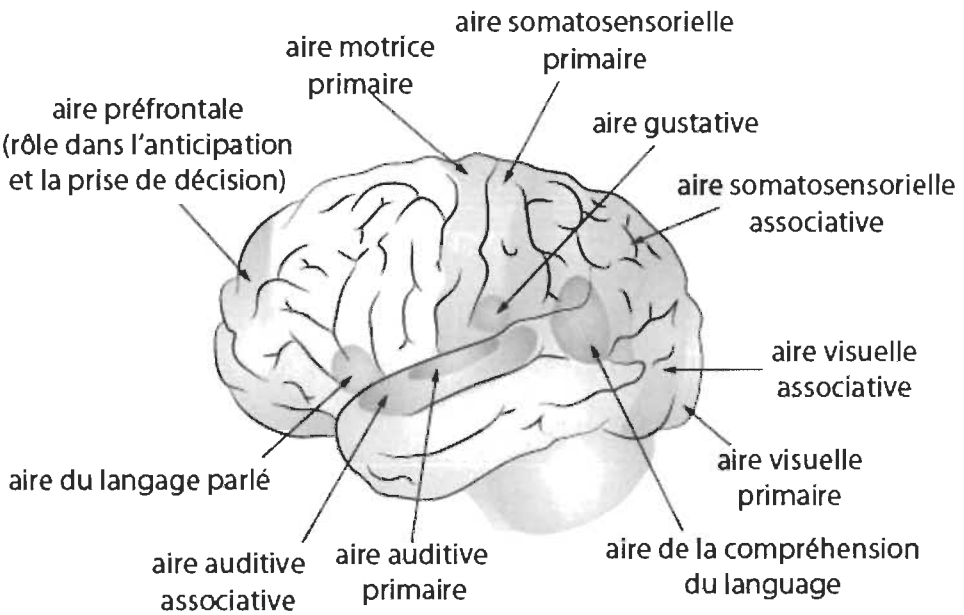

Figure 2 Schéma des différentes zones du cerveau et leur rôle [48]

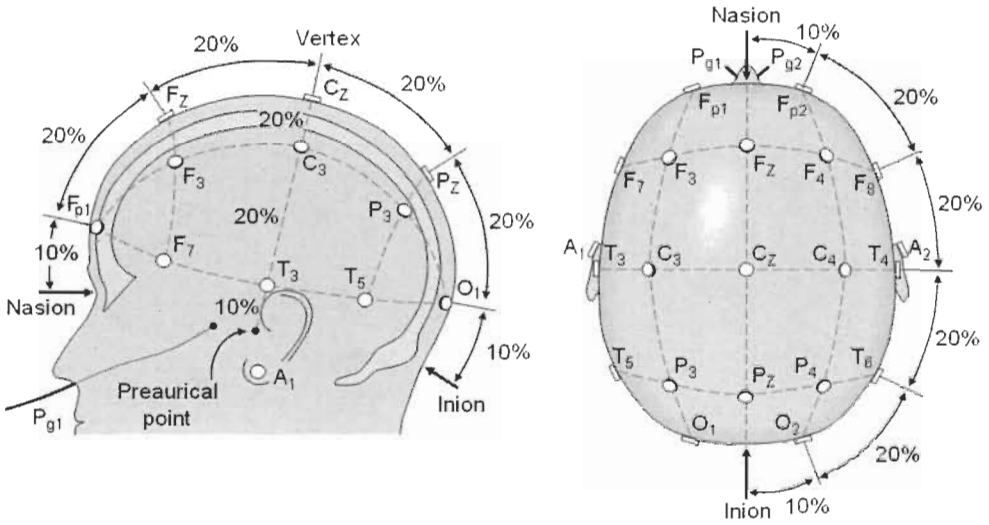

Figure 3 Positionnement et identification standards des électrodes EEG selon le système international 10-20 [49]

2.3.2 Bandes de fréquences-ondes cérébrales

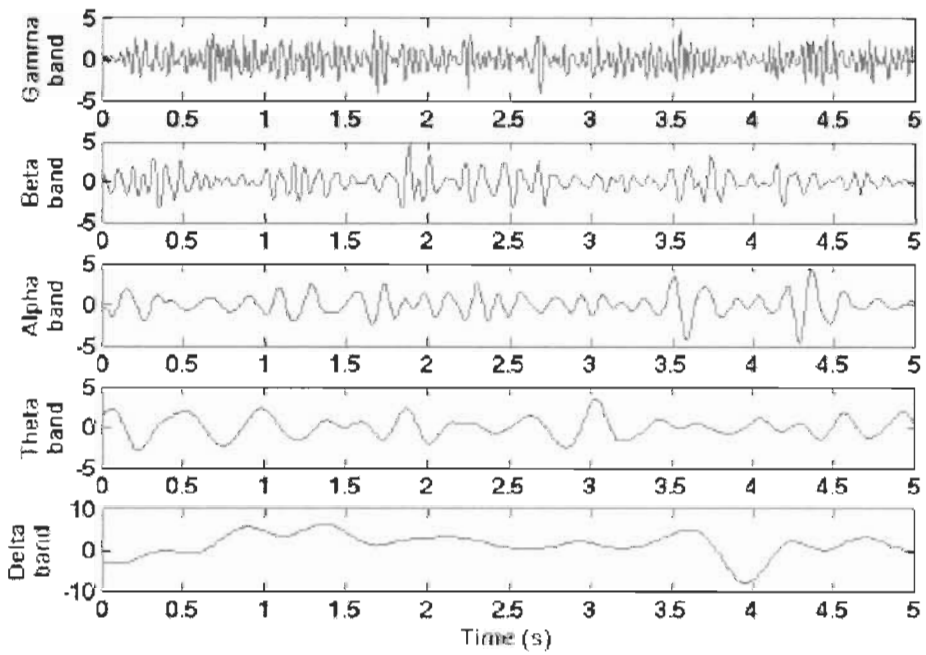

Figure 4 Les ondes associées aux bandes de fréquences des signaux EEG [50]

Nous distinguons plusieurs rythmes cérébraux (Voir figure 4) qui sont en réalité un ensemble de signaux spontanés, autrement dit, ce ne sont pas des signaux générés par une stimulation externe. Ces rythmes, nommés aussi ondes cérébrales, permettent d'identifier

des patrons d'activité neuronale anormale découlant de pathologies telles que les tumeurs cervicales, ou de classifier les états de sommeil ou les états de conscience.

La bande de fréquence Delta: Cette fréquence est comprise dans l'intervalle de 0 à 4 Hz (Voir figure 4). Elle se manifeste lors d'un état de sommeil sans rêve ou un état de méditation profonde. Cette fréquence n'est jamais observée dans les états de veille. Les ondes delta suspendent la conscience externe et sont une source d'apathie. La guérison et la régénération sont stimulées dans cet état, et c'est pourquoi le sommeil réparateur profond est si essentiel au processus de guérison.

La bande de fréquence Thêta : Les ondes cérébrales thêta surviennent le plus souvent pendant le sommeil mais sont également dominantes dans la méditation profonde. Thêta est notre passerelle vers l'apprentissage, la mémoire et l'intuition. Dans cette bande de fréquences, nos sens sont retirés du monde extérieur et concentrés sur les signaux provenant de l'intérieur. C'est cet état crépusculaire que nous n'éprouvons habituellement que furtivement lorsque nous nous réveillons ou passons au sommeil. En état thêta nous sommes dans un rêve; images vives, intuition et information au-delà de notre conscience normale. Ces ondes sont comprises dans l'intervalle de fréquence de 4 à 7 Hz (Voir figure 4).

Lorsqu'une personne glisse dans un état de somnolence après avoir été consciente, des ondes thêta sont observées. Elles jouent un rôle crucial chez les jeunes enfants. La pensée créatrice, la médiation profonde, l'accès au matériel inconscient sont associés aux ondes thêta [10].

La bande de fréquence Alpha : Les ondes cérébrales alpha sont dominantes lorsque les pensées évoluent doucement et dans certains états méditatifs. Ce type d'ondes est comprise dans l'intervalle de 8 à 12 Hz (Voir figure 4).

Cette plage de fréquences correspond à l'activité cérébrale s'insérant entre notre pensée consciente et notre esprit subconscient. En d'autres termes, alpha est la plage de fréquences entre bêta et thêta. Dans cette bande de fréquences nous sommes calmes ce qui favorise des sentiments de relaxation profonde. Si nous devenons stressés, un phénomène appelé blocage alpha peut se produire, impliquant une activité bêta excessive

et très peu d'alpha. Essentiellement, les ondes bêta bloquent la production d'alpha parce que nous sommes trop excités.

Une étude a démontré une augmentation de la puissance alpha après la consommation du cannabis (Lukas, Mendelson et Benedikt, 1995) [11].

La bande de fréquence Bêta Les ondes cérébrales bêta, comprises dans l'intervalle de 12 à 30 Hz (voir figure 4), dominent notre état de conscience éveillé normal lorsque l'attention est dirigée sur les tâches cognitives et le monde extérieur. Celles-ci sont caractérisées par des ondes cérébrales de basse fréquence et d'amplitude élevée que l'on observe couramment lorsque nous sommes éveillés. Elles sont impliquées dans la pensée consciente et la pensée logique.

Avoir la bonne quantité d'ondes bêta nous permet de nous concentrer et de terminer facilement des tâches scolaires ou professionnelles. Tandis qu'avoir un excès d'ondes bêta peut nous amener à faire l'expérience d'un stress excessif et d'une anxiété. Les fréquences bêta les plus élevées sont associées à des niveaux élevés d'excitation. Lorsqu'une personne est en état de panique, une onde bêta de haut niveau est générée (Sterman, M. B., MacDonald, L. R., et Stone, R 1974) [12].

La bande de fréquence Gamma: Les ondes cérébrales gamma sont impliquées dans des **tâches** de traitements supérieurs ainsi que dans le fonctionnement cognitif. Les ondes gamma sont primordiales pour l'apprentissage, la mémoire et le traitement de l'information.
Nous pensons que l'onde gamma à 40 Hz reflète le mécanisme de la conscience. Il a été constaté que les personnes ayant une déficience intellectuelle et des troubles d'apprentissage ont généralement une activité gamma inférieure à la moyenne.

La détection de ces ondes, ayant une fréquence supérieure à 30Hz, peut être utilisée pour confirmer certaines maladies neurologiques et sont considérées comme un bon indicateur de la synchronisation du cerveau lié à un certain évènement [13].

2.3.3 Structure et fonction des neurones

Un neurone est une cellule de notre système nerveux qui joue un rôle principal dans la communication et le traitement de l'information comme le présente la figure 5 [14].

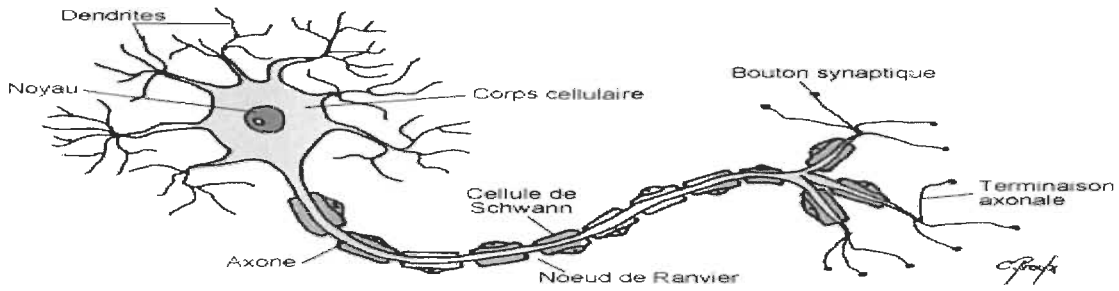

Figure 5 Structure d'un neurone [51]

Un neurone est une cellule permettant la réception et l'envoi d'un message à travers le corps, du cerveau appelé aussi l'encéphale, vers les muscles et les différents organes via la moelle épinière et les nerfs sous forme de messages bioélectriques. Son corps cellulaire est composé d'un noyau et d'un cytoplasme. Les dendrites sont des capteurs de messages des autres neurones émetteurs, ces messages sont analysés par le neurone récepteur.

Ce neurone qui devient un neurone transmetteur passera le message par un influx nerveux dans l'axone jusqu'aux boutons synaptiques où l'information est transmise au neurone suivant. L'axone est enduit d'une gaine de myéline permettant au message de circuler rapidement. Ce dernier est entouré par les cellules de Schwann et séparé par les nœuds de Ranvier.

2.4 La parole humaine

Nous sommes comme nous parlons. Le son de la voix humaine en dit long sur notre personnalité, nos émotions, notre confiance en nous et nos sentiments sur nous-même, ainsi que sur ce que nous pensons vraiment.

D'une part, la parole est une idée cognitive traduite en construction grammaticale. D'autre part le cerveau humain donne un ordre sous forme de commande musculaire afin d'activer le système phonatoire à produire un signal acoustique.

Fondamentalement, la parole est la capacité de communiquer des pensées et des sentiments en articulant des sons.

Toutefois, le mécanisme de la production de la voix humaine est assez complexe. La production de la voix est assistée par trois unités fonctionnelles à savoir, la génération de

vibrations par la modulation de la pression atmosphérique, la régulation des vibrations et le contrôle des résonateurs.

La phonation est la production du son par les organes de la voix. Elle est possible notamment grâce au souffle produit par les poumons à la vibration du larynx et des cordes vocales, à la résonnance des organes résonateurs (cavité buccale et cavité nasale) et à l'articulation (lèvre inférieure, langue, partie avant du gosier, glotte, pharynx, luette, voile de palais, alvéoles dentaires, les dents et lèvre supérieure).

Pour parler il nous faut de l'air, celui-ci est stocké dans les poumons qui se remplissent puis l'air est chassé des poumons et remonte jusqu'au larynx situé dans le cou et constitué de muscles, de tendon et de cartilage. Dans le larynx se trouve les cordes vocales, lorsque l'air remonte des poumons, les cordes vocales se rapprochent pour former un obstacle au passage de l'air. Ce dernier fait alors pression sur les cordes vocales qui finissent par s'entrouvrir pour en laisser passer une petite partie. Les cordes vocales vont ainsi s'entrouvrir et se refermer des centaines de fois par seconde, puis se mettre à vibrer et communiquer ses vibrations dans l'air ambiant.

L'air arrivant ensuite dans le pharynx, c'est à dire la gorge et la bouche, ou il est amplifié par les résonateurs (cavités buccale et nasale) et modulé par le mouvement de la mâchoire, de la langue, des lèvres et de la voile de palais. Ainsi la voix humaine est la plus complexe de tous les instruments, en quelque sorte, un instrument à air, à corde et à percussion au même temps.

Le mécanisme de production de la voix est divisé en deux parties :

1) **La phonation:** Les organes phonatoires sont constitués de poumons et du larynx. La phonation est comparable à un instrument de musique puisque ce dernier comporte un fournisseur d'énergie; la baguette ou l'archet, par exemple, permettant de transmettre à la corde du violon ou à au tambour une énergie sonore, un générateur d'énergie sonore lors de la vibration des cordes du violon sous l'action de l'archet et un résonateur afin d'amplifier le son. La phonation agit aussi comme un système de production vocale, car elle crée les sons de la source vocale.

Ceci est fait en ajustant la pression de l'air dans les poumons et la vibration des cordes vocales au niveau du larynx. Ces deux organes règlent collectivement le volume, la prosodie, la hauteur et la qualité de la voix.

Pour pouvoir prononcer, entre autres, faire vibrer les cordes vocales, il faut que ces dernières puissent s'accoler. Si elles ne sont que légèrement accolées, elles ne laissent passer qu'une partie du souffle d'air, elles donneront lieu au chuchotement.

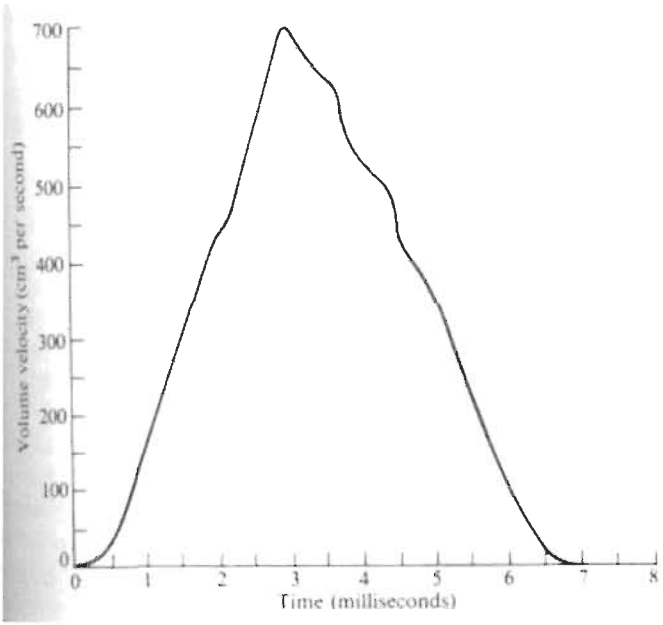

Figure 6 Variation du flux d'air circulant dans la glotte [52]

La figure 6 expose la variation de flux d'air qui circule dans la glotte en fonction du temps sur un cycle complet. Lorsque la parole est émise, la courbe se reproduit d'une façon périodique en moyenne toutes les 8 ms (Voir figure 7) chez l'homme (une fréquence de vibration d'environ 120 Hz) et toutes les 4 ms chez la femme (soit une fréquence de 240 Hz) et toutes les 3 ms chez les enfants (avec une fréquence de 300 Hz) [52].

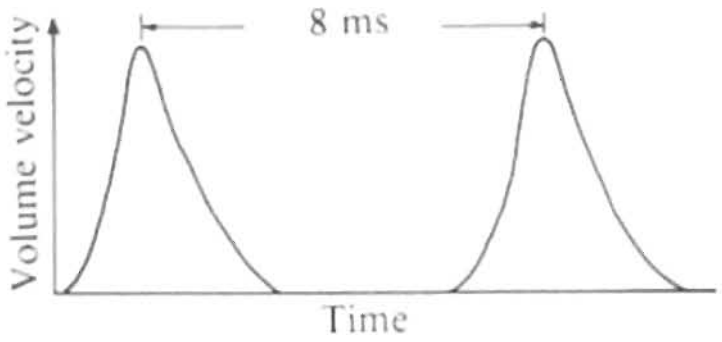

Figure 7 Période des vibrations sonores [52]

L'amplitude du son émis dépend de la fréquence de vibration. Son intensité dépend en principe de l'amplitude des vibrations, déterminée par l'importance du courant d'air qui traverse la glotte, mais aussi du degré de fermeture de la glotte pour chaque vibration. Si la fermeture reste incomplète et si une partie du courant d'air passe sans être transformée en vibrations, le son devient plus faible. L'amplitude et l'intensité sont également fonction de la tension des cordes vocales. Les voix féminines sont plus aiguës que les voix masculines à cause de la longueur des cordes vocales. On admet dans l'ensemble que les cordes vocales longues correspondent plutôt à des voix graves, et les cordes vocales courtes, aux voix aiguës. Les hommes ont des cordes vocales de 17 mm (voix de ténor) à 24 mm (voix de basse) et les femmes de 14 mm (voix de soprano) à 18 mm (voix de contralto). L'onde source produite par la glotte est un signal périodique s(t) [52].

D'après le théorème de Fourier, on peut décomposer ce signal périodique en une somme de composantes sinusoïdales dont les fréquences sont des multiples de la fréquence du signal de base. On parle de fréquence fondamentale et de ses harmoniques.

2) L'articulation:

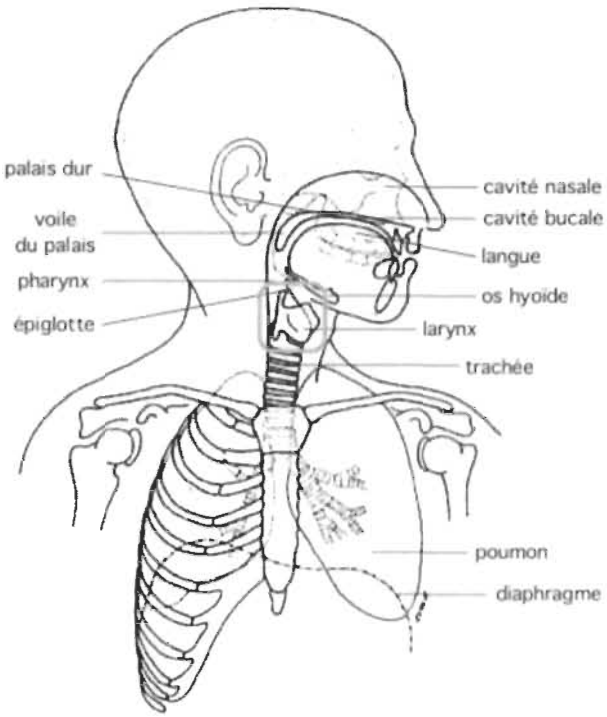

Figure 8 Organes articulatoires

Les organes articulatoires (Voir figure 8) sont constitués des lèvres, de la langue, de la mâchoire inférieure et de la luette appelée aussi uvule. Ces organes donnent des modulations ou des résonances à la source vocale et produisent également des sons supplémentaires pour certaines consonnes. Les propriétés du résonateur acoustique dépendent de la position des organes articulatoires.

2.4.1 Les moyens d'investigation de la parole

La compréhension des principes de la phonation a eu un grand progrès depuis quelques décennies grâce aux progrès des moyens d'observation de l'appareil phonatoire en temps réel.

En 1935, le laryngoscope était considéré comme un grand avancement; il est placé au bout d'une tige formant un angle incliné à 45° avec un miroir semblable à celui des dentistes, permettant d'observer le plan glottique, avec une lumière stroboscopique, le mouvement des cordes vocales. Le laryngoscope fût inventé par Manuel Garcia un chanteur et professeur de chant. Cet outil est encore utilisé de nos jours pour effectuer un examen rapide des cordes vocales, en demandant au patient d'étirer la langue pour

pouvoir dégager l'entrée de pharynx, ce qui empêche de parler librement et ne permet de ce fait qu'une inspection rudimentaire que pour des raisons médicales [16].

En 1975, l'èlectroglottographe de Fabre, qui repose sur une fonctionnalité complétement différente à celle du laryngoscope permettant de visualiser d'une façon indirecte le mouvement des cordes vocales.

Le système est composé de deux électrodes placées sur le cou du sujet entre lesquelles circule un courant électrique d'une haute fréquence (de l'ordre de 20 KHz) et d'une faible intensité. Le mouvement des cordes vocales permet de varier l'impédance transversale de la gorge qui se traduit par l'oscillation d'une courbe sur un oscilloscope auquel sont reliées les électrodes.

Un an après l'apparition de l'èlectroglottographe, le naso-fibroscope est apparu en 1976 en donnant la possibilité de filmer l'intérieur du larynx en introduisant un faisceau de fibres optique à travers le rhino-pharynx et les fosses nasales. Les fibres optiques constituant cette invention, de faible diamètre (3 mm), sont souples, peuvent être courbées de l'extérieur et dirigées à volonté vers le point à observer.

Le naso-fibroscope, comme son nom l'indique, contient aussi un dispositif d'éclairage stroboscopique (500 Watts), une caméra vidéo reliée à un écran cathodique pour l'observation en temps réel, ainsi qu'à un magnétoscope pour l'enregistrement et éventuellement à un ordinateur pour le traitement du signal acoustique.

De nos jours, l'une des tendances les plus importantes en intelligence artificielle est l'analyse du langage naturel humain dans le but d'interagir avec ce dernier et d'apprendre comment comprendre le langage naturel humain dont la voix. Cependant, le discours humain n'est pas toujours précis, il est souvent ambigu et sa structure linguistique peut dépendre d'un grand nombre de variables complexes.

2.4.2 Techniques expérimentales d'analyse de la voix

La production de la voix humaine fait intervenir trois niveaux essentiels pour la parole, le chant et toutes formes d'expressions vocales (Henrich, 2012) [53]. Le premier niveau porte sur le contrôle respiratoire. Il permet de fournir une énergie aérodynamique au système vocal. Le second niveau porte sur le contrôle laryngé de la phonation. L'énergie

de l'air se transforme en son après une interaction avec des parois mobiles, comme les plis vocaux ou tout autre pli au sein du larynx, ou par la rencontre d'obstacles sur son parcours dans le conduit vocal (comme par exemple les dents ou le rapprochement de la langue vers le palais). Une source sonore est ainsi produite. Le troisième et dernier niveau est celui du contrôle articulatoire. La source acoustique est modulée par les variations temporelles lentes de l'aire des cavités qui constituent le conduit vocal, sous l'effet du mouvement des articulateurs que sont le larynx, le palais, la langue, les dents, les lèvres, la mandibule. Nous allons exposer comment analyser la voix selon ces trois niveaux de contrôle [17].

2.4.2.1 Analyse des gestes respiratoires

La technique la plus directe et la plus aboutie reste la ciné radiographie par rayons X, qui ne peut néanmoins plus être appliquée à des fins de recherche en raison de la dangerosité des expositions aux rayons X.

Cette technique d'imagerie permet notamment de visualiser de façon statique ou dynamique les mouvements des organes essentiels au contrôle respiratoire, comme la cage thoracique ou les deux hémi-coupoles diaphragmatiques. Elle permet également l'observation d'organes adjacents, comme la structure de la colonne vertébrale dorso-lombaire, l'ombre du cœur ou la « bulle » gastrique (Trolliet-Cornut, 2002) [18]. Des films illustratifs sont à la disposition du lecteur dans l'ouvrage collectif de Cornut (2002) [18].

Nous pouvons évaluer les actions musculaires qui participent à la gestion respiratoire. Pour ce faire, la technique la plus utilisée est l'électromyographie de surface.

Les activités électriques des muscles pendant leurs phases de détente et de contraction sont détectées à l'aide d'électrodes de contact plates et circulaires positionnées sur la peau au niveau des muscles à étudier. Des couples d'électrodes sont communément utilisées. Cette technique requiert de repérer avec précision la position des muscles chez le sujet, par palpation et observation de l'amplitude du signal mesuré. Les électrodes sont positionnées d'une distance donnée dans le sens des fibres musculaires. Cette technique est souvent complétée par une mesure simultanée des pressions thoracique et abdominale

à l'aide d'une ceinture de contrainte mesurant les variations de circonférence des zones ciblées (e.g. Pettersen et al., 2005) [19].

2.4.2.2 Analyse des gestes phonatoires

L'espace laryngé est logé dans notre cou, l'innovation technologique a rendu disponible plusieurs approches directes ou indirectes pour observer son fonctionnement. D'une façon directe mais invasive, nous pouvons utiliser les approches actuelles d'endoscopie laryngée [54].

Le laryngoscope à miroir inventé par Babington en 1827 et largement utilisé en phonation par Garcia [55], est l'outil de visualisation des plis vocaux par excellence. Il nécessite l'apport de lumière en fond de gorge à l'aide d'une source de lumière blanche. Il peut être complété par un stroboscope qui, réglé sur la fréquence fondamentale de vibration des plis vocaux, permet une reconstitution visuelle du mouvement vibratoire glottique. Les possibilités technologiques actuelles permettent de prendre des images en couleur et d'accroître la vitesse d'acquisition à plusieurs milliers d'images par seconde (Crevier-Buchman et Vincent, 2009) [20].

D'une façon indirecte, l'accolement des plis vocaux peut être détecté par l'électroglottographie, une technique caractérisée par sa simplicité et par son caractère informatif (Nathalie Henrich B, 2001) [21]. Cette technique mesure la résistance au passage d'un courant électrique de faible intensité entre deux électrodes placées sur le cou du patient comme nous avons expliqué précédemment. Cette résistance est fonction du degré de contact des plis vocaux au sein du larynx, étant donné que l'air est moins bon conducteur que les tissus humains et oppose donc une plus grande résistance au passage d'un courant électrique. Le signal électroglottographique résultant est donc modulé par les mouvements d'accolement et décollement des plis vocaux lors de la phonation.

2.4.2.3 Analyse des gestes articulatoires

L'Imagerie par Résonance Magnétique (IRM) est devenue l'approche communément utilisée pour étudier la position et la forme des articulateurs.

Il s'agit d'une technique relativement peu invasive, mais qui implique d'allonger le patient dans une machine IRM dont le niveau sonore peut être perturbant pour la

production vocale. Même si l'acquisition dynamique se développe peu à peu (Echternach et al. Sur la voix chantée lyrique, e.g. 2010) [22], l'acquisition statique par la production d'un son vocal sur plusieurs secondes reste prédominante, limitant donc les possibilités exploratoires de cette technique aux configurations articulatoires quasi-statiques.

Le mouvement dynamique des articulateurs peut se mesurer à partir d'un articulographe basé sur le principe de l'électromagnétographie (EMA). Cette technique consiste à suivre le mouvement de marqueurs positionnés sur des points de chair liés aux articulateurs à étudier. Les marqueurs sont de petites bobines qui génèrent des courants électriques sous l'effet du champ magnétique auquel elles sont soumises. Les modèles actuels d'articulographes permettent de mesurer le déplacement de 12 marqueurs à une fréquence d'échantillonnage de 200 Hz.

La visualisation des mouvements de la langue par échographie est une alternative intéressante à la technique EMA (Hueber, 2009, 2013) [23]. Une sonde ultrasonore est positionnée sous le menton du patient. Elle permet d'observer le mouvement de la langue dans les plans sagittal et coronal avec une résolution temporelle de 60 à 100 images par seconde. Cette technique permet également un retour visuel des mouvements de la langue lors de la génération de la parole, ce qui a potentiellement de belles perspectives d'applications en rééducation orthophonique [24].

2.4.2.4 Analyse acoustique de la voix

Souvent, en cabinet, l'orthophoniste n'a pas à sa disposition les outils technologiques que nous venons de décrire. Pour analyser la voix de son patient, il s'appuie sur la puissance informative de son écoute analytique.

Plusieurs questions se posent, telles, comment garder une trace d'écoute entre les séances?

À coût réduit, il est possible de s'équiper d'un système d'enregistrement audio composé d'un microphone, d'un ordinateur et d'une carte de son, néanmoins, le choix de matériel est important et en particulier le choix du microphone (Ternström et Granqvist, 2010) [25].

Suivant les recommandations publiées par Svec et Granqvist (2010) [56], la réponse en fréquences du microphone doit être uniforme (à 2 dB près) dans la zone de fréquence

d'intérêt (dans l'intervalle de 20 Hz à 20 kHz), la dynamique appropriée pour permettre l'enregistrement sans distorsion des signaux dont le rapport de puissance du signal sur celui du bruit est suffisamment élevé (au moins 15 dB) pour permettre l'enregistrement des signaux sonores de plus faible intensité.

L'analyse la plus simple à effectuer à partir d'un enregistrement du signal audio est de représenter visuellement le son par les fréquences acoustiques qu'il contient et leur évolution au cours du temps.

L'analyse temps-fréquences d'un son met en évidence les fréquences qui le constituent, leurs niveaux d'amplitude et leurs variations temporelles. Cette représentation visuelle est appelée spectrogramme ou sonagramme.

Le premier paramètre d'importance est la fréquence fondamentale, qui renseigne sur l'amplitude de la voix, sa stabilité lors de sa génération, ainsi que sa plage de variabilité. La fréquence fondamentale se mesure sur les parties vocales du signal acoustique, c'est-à-dire pour la génération de sonorités qui met en jeu la vibration des plis vocaux. Sa définition et son calcul requièrent une stabilité de la durée du cycle vibratoire glottique sur plusieurs cycles consécutifs. Quand cette durée est modifiée de façon notable d'un cycle glottique à l'autre lors de la production de voix pathologique, la mesure de fréquence fondamentale perd de son sens. Il peut être alors intéressant de comparer les durées de cycles glottiques successifs.

L'analyse du signal audio mesuré en sortie des lèvres permet aussi d'estimer la fonction de transfert acoustique du conduit vocal et d'en déduire les fréquences et largeurs de bande des formants. Les formants sont des zones spectrales d'énergie renforcées par l'action de résonance des cavités qui constituent le conduit vocal [57].

Avec les avancées technologiques et la médecine, une nouvelle approche d'analyse acoustique de la voix fit son apparition. Des personnes atteintes de la dysarthrie ont été les sujets d'une étude d'analyse du signal de la parole et le diagnostic de cette maladie (Minu George.T et al., 2017) [58]. Cette étude expérimentale a été menée en enregistrant la parole de 70 sujets dont 25 personnes normales et 45 personnes atteintes de la dysarthrie.

Les formes d'onde vocale ont été analysées à l'aide de la boîte à outils Praat et MATHLAB. Le contour de l'amplitude (enveloppe) des signaux vocaux, la variation de la formation et la durée de la parole sont représentés sous forme de graphiques qui sont ensuite analysés.

L'analyse de l'amplitude de la parole de la population étudiée a montré un modèle spécifique pour chaque type (normal vs avec dysarthrie). Une tonalité cassée et une variabilité de F0 (premier formant d'un mot prononcé) a été observées pour la dysarthrie spastique, la rupture de pas avec la dysarthrie ataxique, et l'amplitude la monotonie avec la dysarthrie extrapyramidale. Après cette analyse, Minu et al. ont pu identifier 19 cas pour lesquels un ou plusieurs modèles reconnus ont coexisté et pour lesquels il y avait une corrélation significative entre la sévérité de la dysarthrie et la variation des formants [58].

Dans une autre étude de Abdellah Kachaab et al. [59], basée sur l'analyse de composantes principales du spectrogramme de la parole pour détecter la parole dysarthrique, a montré des relations significatives entre le locuteur et la dysarthrie.

2.5 Aires de langage dans le cerveau humain

L'année 1861 a marqué le début de l'identification des régions du cerveau impliquées dans le langage. Paul Broca, un neurochirurgien français examina le cerveau d'un patient récemment décédé qui ne pouvait prononcer d'autres syllabes que « tan » sachant qu'il comprenait ce qui était dit [60].

Ce patient ne pouvait produire aucune phrase complète ni exprimer ses idées en les écrivant, sans être atteint d'aucun trouble moteur de la bouche et de la langue qui aurait pu entraver ses capacités d'élocution.

Après l'autopsie effectuée sur le cerveau de ce patient, Paul Broca a constaté une lésion importante dans le cortex frontal inférieur gauche, ce qui l'a poussé à étudier huit patients aux déficits identiques qui avaient une lésion dans l'hémisphère frontal gauche. Après ces examens, il a déclaré son célèbre « Nous parlons avec l'hémisphère gauche » et à identifier pour la première fois l'existence d'un centre du langage dans la partie

postérieure du lobe frontal de cet hémisphère et ce fut la première zone du cerveau associée au langage, appelée l'air de Broca [60].

Quelque dix ans après Broca, un psychiatre allemand du nom de Carl Wernicke mit en évidence un autre type d'aphasie associée à des lésions situées dans d'autres régions du cerveau, dans la partie postérieure du lobe temporal gauche. Alors que le patient de Broca comprenait ce qu'on lui disait tout en ayant des difficultés à parler, le patient de Wernicke avait un débit verbal en apparence normal mais de grandes difficultés à comprendre ce qui était dit, avec une tendance à tenir des discours incohérents ou dénués de sens. On en conclut que les processus de la compréhension du langage et ceux de la production de la parole ne mettaient pas en jeu les mêmes aires du cerveau. Il n'y avait donc pas un seul siège du langage, mais plusieurs centres coordonnés, aux fonctions différentes.

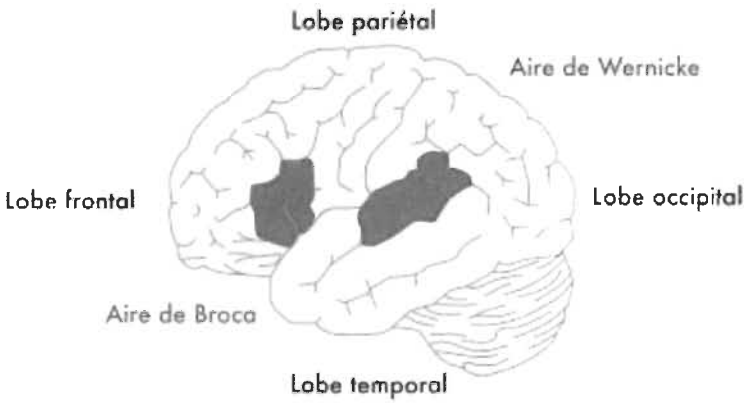

Figure 9 Source de compréhension du langage dans le cerveau humain [61]

Ces observations ont été confirmées et ont s'entend aujourd'hui sur le fait qu'il y a, autour du sillon latéral de l'hémisphère gauche, une sorte de boucle neuronale impliquée dans la compréhension orale du langage et sa production par la parole (Voir figure 9). À l'extrémité frontale de cette boucle, on trouve l'aire de Broca, habituellement associée à la production du langage. À l'autre extrémité, plus précisément dans la partie supérieure et postérieure du lobe temporal, se situe l'aire de Wernicke, associée au traitement des paroles entendues, autrement dit à l'input du langage. L'aire de Broca et l'aire de

Wernicke sont connectées par un important faisceau de fibres nerveuses appelé le faisceau arqué.

Cette boucle présente chez environ 90% des droitiers et 70% des gauchers, le langage étant l'une des fonctions traitées de manière asymétrique dans le cerveau. Étrangement, on la retrouve aussi au même endroit chez les sourds qui s'expriment en language des signes. On peut dire que cette boucle n'est pas spécifique au langage oral ou des signes mais considérée en association à la modalité principale du langage d'un individu.

2.6 Données EEG

Les appareils permettant d'acquérir les données EEG doivent permettre la collecte des données analogiques et leurs conversions en signal numérique. Cette acquisition peut être réalisée par une multitude de façons en assurant les traitements ainsi que leurs séquences.

Les signaux EEG peuvent contenir des données pertinentes et d'autres non pertinentes à la fois. Certains experts en EEG considèrent des données non pertinentes comme signal de bruit alors que d'autres les considèrent comme signal contenant des informations pertinentes.

Lors de l'analyse temps-fréquences des signaux EEG, des personnes seront intéressées par l'approche ERPs (Event Related Potentiel) caractérisée par la manifestation d'un stimulus visuel spécifique, que nous allons exploiter par la suite, en conséquence, si une personne sujette à un enregistrement de ses signaux EEG fait un clignement d'œil, celui-ci, aura un effet sur les signaux EEG enregistrés par le fait que le globe oculaire agit comme un dipôle avec un pôle positif orienté vers l'avant (cornée) et un pôle négatif orienté vers l'arrière (rétine). Lorsque le globe oculaire tourne autour de son axe, il génère un champ de courant alternatif de grande amplitude détectable par l'une des électrodes positionnées près de l'œil. Le clignement amène le pôle positif (la cornée) à se rapprocher des électrodes front polaires FP1, FP2, produisant des déviations symétriques vers le bas.

Les signaux EEG peuvent alors ne pas représenter les processus corticaux associés à la vision d'un stimulus à l'écran, parce que le signal contiendra des signaux bruités par des artefacts de fortes amplitudes liés aux clignements des yeux.

Nous avons jugé opportun d'éliminer les signaux bruités de l'analyse des signaux EEG puisqu'ils ne contiennent pas d'informations pertinentes. Pour comprendre visuellement, la différence entre un signal bruité et un autre non bruité est présentée à la figure 10 [26]. Cette figure représente une estimation du clignement des yeux lors de l'acquisition des signaux EEG capturés d'une sonde située proche des yeux (ex : channel 11). Ce tracé peut nous donner une idée de la position des clignements, en comparant un signal EEG enregistré avec des artefacts de clignement des yeux et un autre sans artefacts. Le tracé du haut (Voir figure 10) montre le signal EEG d'origine avec les artefacts de clignement des yeux estimés en rouge tandis que le tracé inférieur montre le signal EEG nettoyé où l'artefact est soustrait du signal EEG.

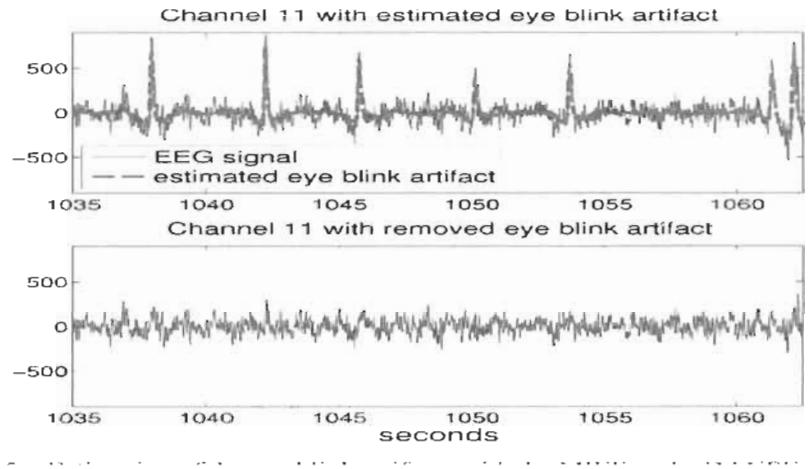

Figure 10 Estimation du clignement des yeux avec l'algorithme DANSE [62]

2.6.1 Traitement de signal

Dans cette partie de ce mémoire, nous allons passer en revue les différentes techniques de traitement des signaux adoptées dans la littérature et dans des recherches connexes à la nôtre tout en commençant par le prétraitement jusqu'à l'extraction des caractéristiques.

2.6.1.1 Signaux EEG bruités

L'électroencéphalographie (EEG) est l'une des technologies principales utilisées de nos jours pour observer l'activité cérébrale dans la recherche. Par rapport à d'autres approches, cette méthode est précise: mesurer les changements d'état cérébral qui se produisent en millisecondes, en assurant une excellente résolution de temps, littéralement à la vitesse de la pensée. Cette approche est également relativement peu coûteuse et facile

à utiliser. Pour ces raisons, l'EEG est mondialement utilisé dans plusieurs domaines dont le domaine de la recherche et le domaine médical. Cependant, l'utilisation de l'EEG n'est pas complètement sans certains obstacles. Le rapport signal / bruit (SNR) est l'un des défis méthodologiques le plus important pour la collecte et l'analyse des données EEG. Il s'agit du rapport de puissance des signaux entre tous les signaux que nous voulons mesurer pour l'analyse et tous les signaux indésirables (bruits) quand même capturés dans le signal EEG. Ce bruit, parfois, est considéré comme problématique car il y a deux sources principales de bruits dans les signaux EEG. La première est le bruit de fond général qui provient de l'extérieur du cerveau. La deuxième est le bruit naturel qui provient de notre cerveau car notre cerveau est toujours occupé à faire beaucoup de tâches à la fois, pas seulement la tâche sur laquelle nous souhaitons nous concentrer dans notre étude. Toutes ces activités contaminent le signal EEG.

L'un des plus grands défis de l'utilisation de l'EEG est le très faible rapport signal / bruit des signaux cérébraux que nous essayons d'observer, couplé avec la grande variété de sources de bruit. Quatre stratégies générales sont utilisées pour traiter la question du bruit dans l'enregistrement et l'analyse des signaux EEG, chacune avec ses propres avantages, défis et limites: l'élimination des sources de bruit, le moyennage du signal, le rejet des données bruitées et l'élimination du bruit [28].

Pour l'élimination des sources de bruit, la meilleure façon de faire face au bruit est de ne pas en avoir en premier lieu. Certaines sources de bruit peuvent être éliminées facilement, d'autres présentent plus de difficultés et peuvent introduire des conséquences indésirables, tandis que certaines sources de bruit sont en principe inévitables. Les sources de bruit les plus faciles à gérer sont les sources de bruit externes et environnementales, telles que les lignes électriques, l'éclairage et un large éventail d'équipements électroniques (ordinateurs, écrans, téléviseurs, routeurs sans fil, ordinateurs portables, téléphones portables). Les étapes les plus élémentaires de la lutte contre le bruit ambiant consistent à éliminer toute source inutile de bruit électromagnétique (EM) de la salle d'enregistrement et de ses environs immédiats et, si possible, à remplacer l'équipement utilisant un courant alternatif par un équipement utilisant du courant continu (tel que le courant continu pour l'éclairage). Une mesure plus

avancée et plus coûteuse consiste à isoler la salle d'enregistrement du bruit électromagnétique en utilisant une cage de Faraday. Bien qu'elle soit très efficace pour éliminer la plupart des bruits électromagnétiques environnementaux, l'isolation électromagnétique nécessite une planification préalable ou des travaux de reconstruction coûteux.

Une autre source de bruit traitable dans l'enregistrement EEG est le bruit physiologique qui peut être provoqué par divers générateurs de bruit. Des exemples courants de ce bruit sont le signal cardiaque (électrocardiogramme, ECG), les artéfacts de mouvement provoqués par la contraction musculaire (électromyogramme, EMG) et le signal oculaire provoqué par le mouvement du globe oculaire (électrooculogramme, EOG). Parmi ceux-ci, le signal ECG n'est pas évitable, mais a également le plus faible effet sur le signal EEG enregistré. Le bruit provoqué par les signaux EMG et EOG peut souvent être évité. Le bruit EMG peut être évité ou réduit en demandant au participant de trouver une position confortable et de se détendre avant le début d'une session d'enregistrement, et en évitant les tâches qui nécessitent des réponses verbales ou de grands mouvements.

Le moyennage du signal est le moyen le plus simple de supprimer la composante de bruit dans les données EEG enregistrées. L'hypothèse clé derrière le moyennage du signal EEG est que le bruit dans le signal est aléatoire, ou au moins se produit avec une phase aléatoire par rapport au signal EEG associé à un événement d'intérêt, tandis que ce signal d'intérêt est stable. Si nous enregistrons le signal EEG à plusieurs reprises, dans une période de temps assez courte, pour faire en sorte que le signal EEG d'intérêt reste stable. Avec une composante de bruit pouvant être considérée comme suivant une distribution normale de moyenne nulle et d'écart-type σ, le calcul de la moyenne pour chaque intervalle de temps des signaux EEG fera en sorte d'annuler le bruit, produisant ainsi la réponse EEG idéale associée à l'événement d'intérêt.

Le moyennage du signal EEG est un moyen simple et puissant d'éliminer le bruit, mais elle présente un certain nombre de limitations et de mises en garde. Premièrement, la moyenne du signal EEG ne peut être utilisée que lorsque nous avons un signal EEG stable pendant une période de temps assez longue, associé à un événement d'intérêt, qui peut être enregistré à un certain nombre de reprises, comme c'est le cas dans les études de

potentiel lié à un événement ERP (Event related potential). Le moyennage du signal EEG ne peut pas être utilisé dans les cas où nous étudions des événements rares qui ne peuvent pas être détectés dans le temps à un moment connu, ou lorsque le signal EEG d'intérêt est lui-même variable. Un exemple de ce genre d'événements rares, pourrait être l'étude des décharges épileptiformes découlant des crises d'épilepsie.

Deuxièmement, seul le bruit aléatoire et symétrique peut être éliminé en utilisant le moyennage du signal EEG. Si pour une raison quelconque le bruit est ponctuel, donc déterministe plutôt qu'aléatoire. Ce signal ne peut donc pas être moyenné puisque la composante de bruit ponctuel ne sera ajoutée au signal EEG d'intérêt que lors de son occurence. Un tel exemple peut être le bruit provenant de la présentation des stimuli. De grands changements de luminosité sur un écran CRT mal isolé peuvent entraîner des pics de bruit ponctuels s'ajoutant au signal EEG généré par l'événement.

De même, si le bruit n'est pas symétrique (introduisant des augmentations équilibrées ainsi que des diminutions de signal), sa moyenne dans le temps ne sera pas nulle mais elle entraînera plutôt des augmentations ou des diminutions globales du signal moyen. Cela peut ne pas être un problème lorsque la quantité de bruit est constante tout au long de la session d'enregistrement ou au moins à chaque essai enregistré, car dans ce cas, le signal qui nous intéresse restera le même par rapport à la ligne de base. Il peut cependant conduire à des artéfacts importants lorsqu'il ne se produit que sur certaines parties de l'essai, où il peut apparaître comme une diminution ou une augmentation systématique du signal EEG et peut donc être confondu avec des composants ERP.

Enfin, s'appuyer sur la moyenne du signal comme principale stratégie d'élimination du bruit peut être assez coûteuse en termes de nombre d'essais nécessaires pour augmenter suffisamment le rapport signal / bruit.

Pour cette raison, il est préférable de s'appuyer sur la moyenne du signal en dernier recours, et uniquement, pour les sources de bruit vraiment inévitables, et d'utiliser d'autres stratégies pour éviter le bruit avant l'enregistrement et/ou le supprimer après l'enregistrement.

La troisième approche pour traiter les bruits est le rejet des données bruitées. Chaque fois que le bruit dans les données enregistrées est facilement reconnaissable, la façon la plus

évidente d'y faire face est d'éliminer les parties des données où le bruit est présent. La procédure la plus simple pour rejeter des données bruitées est l'inspection visuelle. La plupart des mouvements oculaires, des clignements et des artéfacts de mouvement sont relativement facilement reconnaissables et peuvent être marqués pour être rejetés avant la phase de moyennage et l'analyse des données.

Mais il n'est pas toujours possible de s'appuyer sur une inspection visuelle des données. Lorsqu'il s'agit de grands ensembles de données, l'inspection visuelle peut prendre beaucoup de temps. De plus certains types de bruits peuvent être difficiles à reconnaître et à identifier, même pour l'analyste le plus expérimenté.

Pour identifier les mauvais canaux, un certain nombre d'outils d'analyse EEG offrent des options pour visualiser les spectres de fréquence et tester la distribution des données. Les canaux qui possèdent plus de bruit sont généralement caractérisés par une forte puissance dans les hautes fréquences. Les canaux bruités peuvent également montrer une variabilité significativement plus élevée du signal dans le temps par rapport aux autres canaux, ainsi qu'une déviation plus forte de la distribution gaussienne.

Nous allons expliquer maintenant la quatrième approche de traitement de bruit qui consiste à éliminer ce bruit. La façon la plus simple de supprimer le bruit des données brutes est probablement le filtrage. Pour être en mesure de le filtrer, le bruit doit appartenir à l'une des trois catégories de bandes de fréquences: soit inférieure à la fréquence des phénomènes que nous essayons d'observer, soit au-dessus de cette fréquence, ou soit comprise dans une gamme très étroite et bien spécifiée.
Le filtrage passe-haut (filtrage qui ne laisse passer que les harmoniques du signal au-dessus de la fréquence de coupure sélectionnée) est couramment utilisé pendant l'acquisition elle-même afin de réduire la contamination du signal EEG par un bruit et pour atténuer les basses fréquences. Un certain nombre de facteurs tels que la transpiration et les dérives de l'impédance des électrodes peuvent entraîner des changements lents dans la tension mesurée, ce qui peut à son tour conduire à la saturation de l'amplificateur et à la perte de données pendant l'enregistrement, ainsi qu'à des distorsions importantes des signaux EEG capturés n'ayant pas de lien avec l'événement

mesuré. Pour ces raisons, il est souvent recommandé de filtrer (éliminer) les fréquences inférieures à 0,01Hz.

Le deuxième filtrage que nous pouvons appliquer est le filtrage passe-bas (filtrage qui ne laisse passer que les harmoniques du signal en dessous de la fréquence de coupure sélectionnée) est utilisé pour éliminer le bruit à l'autre extrémité du spectre de fréquences qui nous intéresse. La contraction des muscles conduit généralement à un signal fort avec des fréquences supérieures à 100 Hz, la suppression des fréquences supérieures à 100 Hz supprimerait les artéfacts de mouvement dans le signal EEG acquis ayant quant à lui un contenu fréquentiel dans l'intervalle [1 Hz, 20 Hz]. De plus, l'échantillonnage des données lui-même peut conduire à un phénomène de repliement spectral dans lequel des fréquences supérieures à la fréquence d'échantillonnage peuvent être induites comme des basses fréquences artificielles dans les données échantillonnées. Pour cela, les données doivent être filtrées passe-bas à des fréquences au moins un tier inférieur à la fréquence d'échantillonnage [29].

Lors du filtrage des données, il faut être conscient que le type de filtre utilisé peut également affecter de manière significative les données dans les plages de fréquences non nécessairement filtrées, affectant ainsi les estimations des débuts et / ou de l'amplitude des ondes ERP observées. Introduisant parfois des oscillations artéfactuelles. Par conséquent, il est conseillé de limiter le filtrage uniquement à ce qui est nécessaire et inévitable.

Une autre approche est suggérée par Croft RJ, Barry RJ [30] dans une étude prometteuse. Dans cette étude les sources de bruit prévisible sont enregistrées indépendamment sur un canal séparé. Il devient alors possible de supprimer le bruit des données en estimant la composante de bruit ajoutée aux données à l'aide d'une régression linéaire, puis ensuite en faire la soustraction du signal EEG capturé. Un exemple typique est le bruit produit par les clignements et les mouvements oculaires. Pour éliminer ce bruit, une régression linéaire est effectuée entre chaque canal de données et les canaux nuisibles utilisés pour enregistrer les signaux de contractions musculaires, horizontalement (différence HEOG entre les tensions enregistrées au-dessus et en dessous des yeux), verticalement (VEOG - différence entre les tensions enregistrées aux canthi externes gauche et droit des yeux) et radialement (REOG - différence entre la tension moyenne aux yeux et la référence EEG).

Les coefficients ß estimés sont ensuite utilisés pour soustraire les valeurs de chaque canal nuisible multipliées par le coefficient ß correspondant du canal de données mesurées.

2.6.1.2 Analyse en composantes indépendantes des données (ICA)

L'analyse en composantes indépendantes est une technique statistique et informatique permettant de révéler des facteurs cachés qui sous-tendent des ensembles de variables, de mesures ou de signaux aléatoires.

L'ICA définit un modèle génératif pour les données multivariées observées, qui est généralement donné comme une grande base de données d'échantillons. Dans le modèle, les variables de données sont supposées être des mélanges linéaires de certaines variables latentes inconnues, et le système de mélange est également inconnu. Les variables latentes sont supposées non gaussiennes et mutuellement indépendantes, et elles sont appelées les composantes indépendantes des données observées. Ces composantes indépendantes, également appelées sources ou facteurs, peuvent être trouvées par l'ICA. L'ICA est superficiellement liée à l'analyse en composantes principales et à l'analyse factorielle. L'ICA est cependant une technique beaucoup plus puissante, capable de trouver les facteurs ou sources sous-jacents lorsque ces méthodes classiques échouent complètement.

Les données analysées par l'ICA pourraient provenir de nombreux types de domaines d'application, notamment des images numériques, des bases de données documentaires, des indicateurs économiques et des mesures psychométriques. Dans de nombreux cas, les mesures sont données sous la forme d'un ensemble de signaux parallèles ou de séries chronologiques. Le terme séparation aveugle des sources est aussi utilisé pour caractériser ce problème. Des exemples typiques sont des mélanges de signaux vocaux simultanés captés par plusieurs microphones, des ondes cérébrales enregistrées par plusieurs capteurs, des signaux radio brouilleurs arrivant sur un téléphone mobile ou des séries chronologiques parallèles obtenues à partir d'un processus industriel.

L'illustration classique de la séparation de sources est le problème de la soirée cocktail (cocktail party problem). Lors d'une telle soirée, on dispose de P microphones dans une salle dense, où N personnes discutent par groupes de tailles diverses. Chaque microphone

enregistre la superposition des discours des personnes dans ses alentours et le problème consiste à retrouver la voix de chaque personne (« débarrassée » des autres voix considérées comme parasites). L'ICA permet de résoudre ce problème en considérant simplement que les personnes qui parlent à un instant donné ont des discours « indépendants » [31].

Alors que pour les signaux EEG, éliminer des parties entières de ces signaux dues au bruit est une approche amplement utilisée dans les contextes de recherche et engendre la perte de données expérimentales. Cela posera un problème dans le cas de la présence fréquente des artéfacts tels que les clignements des yeux et seulement quand des époques (epoch) sont disponibles. L'époque EEG est une procédure dans laquelle des délais spécifiques sont extraits du signal continu de l'EEG. Ces fenêtres temporelles sont appelées « époques », et sont généralement verrouillées dans le temps dans le respect d'un événement, par exemple un stimulus visuel. En outre, cette méthode est considérée inadaptée quand les signaux EEG mesurés des sujets représentent une activité qui n'est pas couplée avec un évènement (Vanhatalo S et al., 2004) [32]. Comme par exemples, dans les applications d'interface cerveau-ordinateur (BCI ou IND) en temps réel et surveillance en ligne de l'état mentale (Jung et al., 2000) [33], ou de corrélation temporelle à longue portée (Linkenkaer-Hansen et al., 2001) [34].

Certaines approches proposées pour l'exclusion d'artéfacts se basent sur la régression dans le domaine temporel ou fréquentiel (Kenemans et al., 1991) [35]. Elles se concentrent particulièrement sur l'exclusion des artéfacts oculaires (Jung et al., 2000) [33]. Ces types d'artéfacts dans les enregistrements EEG (Theodore C. Weerts et Lang, 1973) [36] sont rejetés dans certaines applications temps réel (Jung et al., 2000) [33]. Une approche importante pour l'analyse des signaux EEG, est la séparation de l'activité neuronale et les artéfacts développée par Sean P.Fitzgibbon et al. en 2007 [37] et l'application de l'analyse par composantes indépendantes (ACI) pour décomposer les données (Jung et al., 2000) [33] (Julie Onton et al., 2006) [38]. Le but général de cette approche est de prendre en considération que le signal EEG est un ensemble de composantes de sources (IC) linéairement indépendantes qui peuvent être éliminées et de composantes de signal EEG restantes pouvant être projetées dans le domaine temporel

original. Cette approche permettra la reconstruction d'un signal EEG sans artéfacts. L'idée sous-jacente d'une des approches d'élimination d'artéfacts présentée dans la littérature est inspirée de l'observation d'expert lors de la classification visuelle. Pour ce processus, ils inspectent visuellement les projections cartographiques du cuir chevelu 2D des circuits intégrés, appelées topoplots. Sur la base du modèle d'image des topoplots et du savoir-faire des experts, ils décident si le circuit intégré est un artefact ou un composant de signal EEG.

La vision par ordinateur vise à imiter les capacités de la vision humaine en percevant et en comprenant électroniquement une image pour une prise de décision ultérieure. Par conséquent, nous introduisons dans cet article une nouvelle méthode de rejet d'artéfacts basée sur la classification par machine des caractéristiques dérivées d'images topoplot en appliquant des algorithmes de traitement d'image pour améliorer les performances. Cette méthode révolutionnaire ne dépend pas de l'enregistrement direct des signaux d'artéfacts, elle n'est pas non plus limitée à un nombre ou à un type d'artéfacts spécifique. Un avantage supplémentaire de cette nouvelle méthode est son adaptabilité à un nombre arbitraire de configurations de canaux EEG. Dans la figure 11, nous pouvons voir une illustration de la similitude des images IC par artéfact malgré leurs différentes configurations de canaux EEG [39].

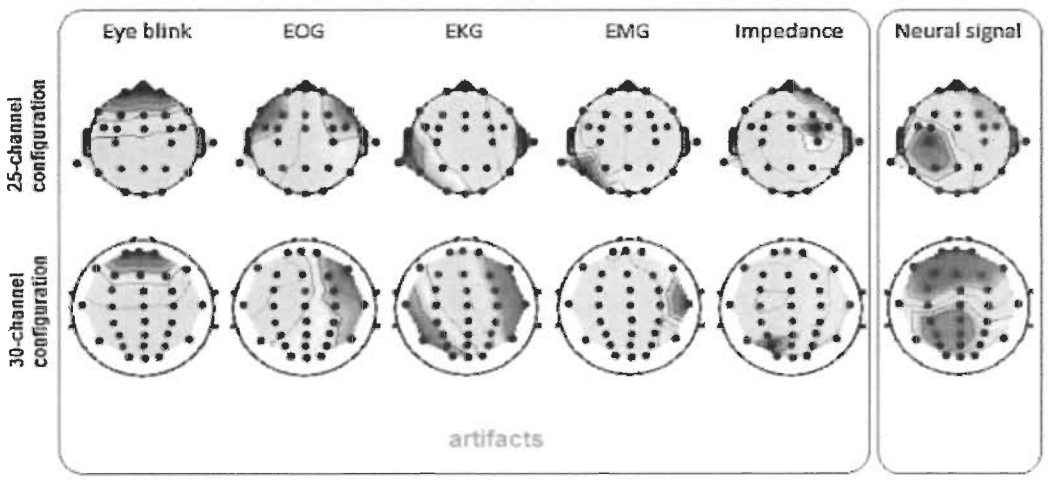

Figure 11 Exemples de modèles d'artéfacts IC similaires malgré les différentes configurations de canaux EEG. Première rangée: images IC d'une configuration à 25 canaux; deuxième rangée: images IC d'une configuration à 30 canaux

L'exemple présenté dans la figure 12, tiré d'Onton et al,. (2006) [38], met l'accent sur la variété des informations sources généralement contenues dans les données EEG, et la capacité frappante de l'ICA à séparer ces activités distinctes (signaux distincts) des mélanges de canaux enregistrés.

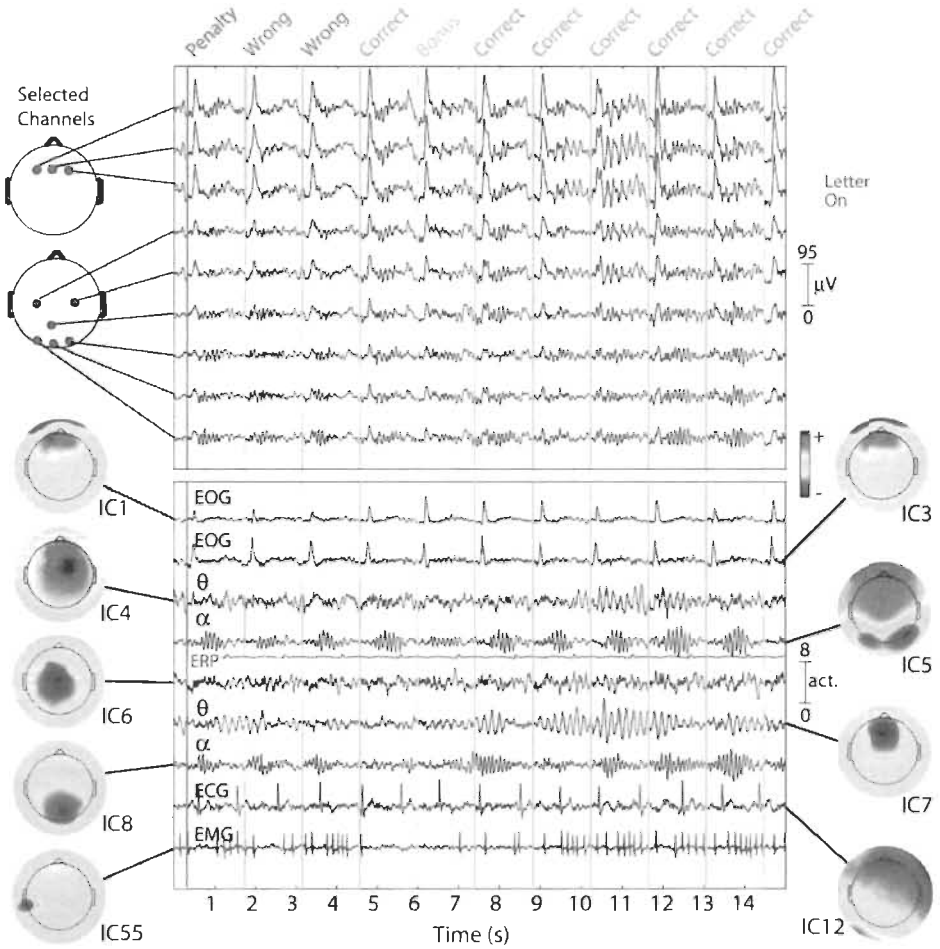

Figure 12 Quinze secondes de données EEG sur 9 canaux (sur 100) du cuir chevelu (panneau supérieur) avec activités de 9 (sur 100) composants indépendants (CI, panneau inférieur)

Lorsque les électrodes à proximité (panneau supérieur de la figure 12) enregistrent des mélanges très similaires d'activités cérébrales et non cérébrales, les activités des composantes ICA (panneau inférieur de la figure 12) sont temporellement distinctes (c'est-à-dire elles sont fortement indépendantes dans le temps), même lorsque leurs cartes du cuir chevelu se chevauchent. Comparez, par exemple, IC1 et IC3, en tenant compte des différentes phases des artéfacts de clignement des yeux produits par ce sujet après

chaque présentation visuelle de lettres (fond gris) et du signal de rétroaction de performance auditive qui en résulte (lignes colorées). Comparez également IC4 et IC7, qui expliquent le chevauchement des activités de la bande thêta frontale (4-8 Hz) apparaissant pendant une séquence de performances correctes (secondes 7 postérieure (8-12 Hz) qui se chevauchent et qui apparaissent lorsque le sujet attend la présentation de la lettre suivante (fond blanc). En guise de comparaison, la réponse évoquée visuelle moyenne répétée d'un processus IC occipital bilatéral (IC5) est montré (en rouge) sur la même échelle (relative). Des IC d'artéfacts ECG et EMG typiques sont également présentés, ainsi que des rafales superposées dans la bande alpha.

2.6.2 Transformée de Fourier

La transformée de Fourier est une fonction mathématique qui permet de représenter le contenu fréquenciel (développement sur une base d'exponentielles complexes) des signaux qui ne sont pas périodiques. Il s'agit de l'analogue des séries de Fourier pour les fonctions périodiques (développement sur la base de fonctions sinusoïdales). Une opération non périodique pouvant être considérée comme opération dont la période est infinie [40].

La transformée de Fourier d'une fonction x est déduite par la formule (Voir eq. 1) suivante :

$$\mathrm{TF}[x(t)] = X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft}\, dt \qquad (1)$$

Où X(f) représente une mesure de l'importance d'une harmonique f dans tout le domaine de la fonction x(t).

La transformée de Fourier dans son application permet de caractériser un signal par son spectre fréquentiel qui, notamment, permet de mettre en évidence l'importance de l'harmonique fondamental ainsi que la décroissance plus ou moins rapide de l'amplitude des harmoniques de rang plus élevé. Elle peut aussi nous aider à déterminer le nombre d'harmoniques nécessaires pour transmettre la quasi-totalité de l'énergie du signal (bande passante, coupe-bande, etc.) [40].

Selon Fabien Lotte et al., (2007) [41], Lyons RG (2011) [42] et Proakis JG, Manolakis DG (2007) [43], la transformée de Fourier ainsi que l'extraction des puissances des bandes est considérée comme l'approche la plus utilisée pour l'analyse et le traitement des signaux. L'algorithme basé sur l'équation de transformée de Fourier discrète (DFT), telle qu'exprimée par l'équation 2, pourra être appliqué sur les signaux EEG, ce qui nous permettra de séparer les différents rythmes EEG [44].

La transformée de Fourier Discrète d'une fonction x est formulée par :

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2k\pi\frac{n}{N}} \quad k = 0, \dots, N-1 \qquad (2)$$

Où X_k représente une mesure de l'importance d'une harmonique k dans tout le domaine de la fonction x_n avec n allant de *0 à N-1* et N le nombre d'observations d'un signal.

La complexité de l'algorithme générique DFT est de O (N^2), est considérée comme la solution la plus simple au problème DFT et est capable de résoudre toutes les tailles de DFT [45], cependant un autre algorithme est considéré plus fiable et efficace appelé transformation de Fourier Rapide (FFT), il est l'un des quelques algorithmes dont la publication a provoqué une véritable révolution dans le champ technique. Cet algorithme développé par J.W Tuckey et J.W Cooley (1965) permet de calculer la transformée de Fourier discrète qui avait été maintes fois redécouverte depuis Gauss, notamment par Danielson et Lanczos (1942). La FFT permet de ramener le calcul de la transformée de Fourier discrète de N^2 à ($NlogN$) opérations. Cette réduction de complexité suffit à faire passer d'impossible à facilement résolvables nombre de problèmes [46].

2.6.3 Transformée en ondelettes

Depuis les années 1930, les activités électriques du cerveau ont été mesurées en utilisant des électrodes de surface connectées au cuir chevelu. Mais de nos jours, divers outils mathématiques tels que la transformation de Fourier (FT), la transformation de Fourier rapide (FFT), la transformation de Fourier à court terme (STFT) et la transformation en ondelettes (WT) ont été introduits pour l'extraction de caractéristiques de signaux EEG. Cependant, dans la transformée de Fourier rapide, vu la possible non-stationnarité des signaux, il survient une perte d'information par rapport au domaine temporel qui ne donne

que des informations spectrales dans le domaine fréquentiel. Pour surmonter les problèmes liés à la FFT, la technologie STFT a été introduite. Elle représente le signal aussi bien dans le temps que dans le domaine fréquentiel en utilisant une fonction de fenêtrage mobile. Le principal problème associé à la STFT est qu'elle ne donne pas d'informations à résolution multiple des signaux car elle a toujours une taille constante. Afin de surmonter les problèmes liés à la transformation de Fourier, à la transformation de Fourier rapide et à la transformation de Fourier à court terme, une méthode puissante a été proposée à la fin des années 1980, connue sous le nom de transformée en ondelettes. La transformée en ondelettes peut être considérée comme une extension de la transformée de Fourier qui au lieu de travailler sur une seule échelle (temps ou fréquence), fonctionne plutôt sur plusieurs échelles et aborde également les problèmes liés aux signaux non stationnaires. Ces dernières années, l'analyse par ondelettes a joué un rôle important dans l'analyse des signaux du domaine temporel. L'ondelette est un type d'analyse temps-fréquence qui fournit des informations sur la fréquence et le temps sur les signaux. La transformation en ondelettes s'est avérée plus efficace pour l'analyse du signal que d'autres méthodes de transformation telles que la transformation de Fourier, la transformation de Fourier à court terme. Le principal avantage de la transformation en ondelettes est qu'elle a une taille de fenêtrage variable qui est large aux basses fréquences et étroite aux hautes fréquences, menant à une résolution de fréquence temporelle optimale dans toutes les plages de fréquences, c'est-à-dire qu'elle conserve les propriétés multi résolution. En mathématiques, une série d'ondelettes est une représentation de carré intégrable qui peut être une fonction à valeur réelle ou complexe. Une ondelette est une onde qui oscille avec une amplitude qui commence à zéro, augmente puis décroît à zéro. Les ondelettes sont généralement conçues pour avoir des propriétés spécifiques qui les rendent utiles pour le traitement du signal. Les ondelettes peuvent être combinées en utilisant une technique inversée, décalée, multipliée et intégrée appelée convolution. Les transformées en ondelettes sont généralement divisées en trois classes : continue, discrète et multi résolution [63].

La méthode utilisée dans notre approche est l'ondelette Morlet (ondelette de Gabor) qui est une transformée en ondelettes de classe continue (Voir figure 13).

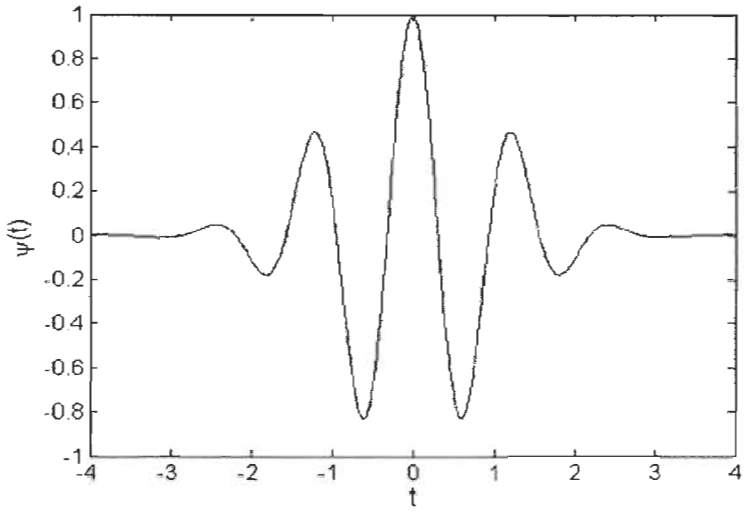

Figure 13 Fonction de base de l'ondelette Morlet [63]

Dans la transformée en ondelettes continu, un signal donné d'énergie finie est projeté sur une famille continue de bandes de fréquences [63].

$$\gamma(s,\tau) = \int f(t)\,\psi_{s,\tau} * (t)dt \qquad (3)$$

Où * désigne la conjuguée complexe, $\psi_{s,\tau}$ la fonction d'ondelette pour une échelle s et une translation τ.

L'équation (3) montre qu'une fonction *f(t)* peut être décomposée en un ensemble de fonctions $\gamma(s, \tau)$ correspondant à la réponse de l'application d'une ondelette comme celle de Morlet sur un signal *f(t)*. Sachant que l'ondelette de Morlet est en fait un filtre passe-bande, il devient possible de déterminer la puissance d'un signal *f(t)* pour une bande de fréquences (échelle s) et temps (translation τ) donnés [63].

2.6.4 Densité spectrale de puissance (PSD)

L'analyse de la densité spectrale de puissance est effectuée pour déterminer la puissance du signal sur une bande de fréquences particulière [63].

Les études dans la littérature révèlent que la puissance du signal EEG dans la région frontale diminue avec l'augmentation de la quantité d'alcool ingérée et que la puissance du signal EEG dans les régions centrales, occipitales augmente. Cela indique que la stimulation nerveuse de l'alcool a une forte influence sur la région centrale, de sorte que les gens sont excités après avoir bu [53].

Les ondes thêta (θ) commencent à apparaître et augmentent progressivement après avoir consommé de l'alcool, ou au fur et à mesure que les sujets sont conduits dans l'état de sommeil ou quand le système nerveux des sujets est inhibé. (Voir figure 14)

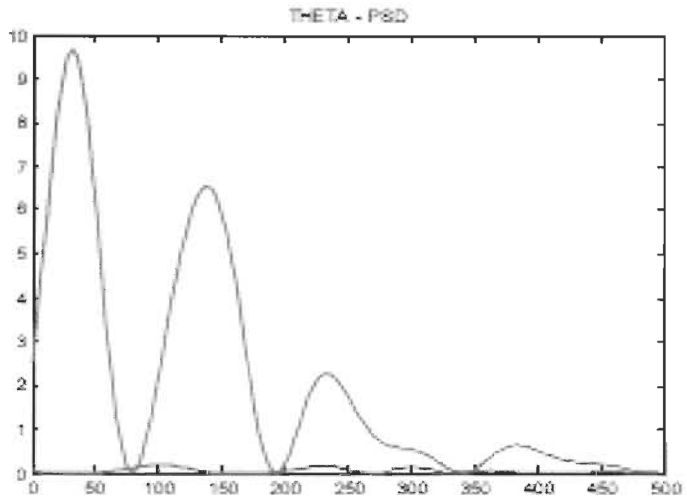

Figure 14 Densité du spectre de puissance de l'onde thêta d'une personne normale (rouge) et d'une personne alcoolique (bleue), l'axe vertical représente la puissance et l'axe horizontal représente le temps en miliseconde [63]

Les ondes bêta (β) : Le cortex cérébral reste dans un état excitable après avoir consommé de l'alcool, ce qui conduit à une augmentation progressive des ondes bêta (β) et à une extension de la zone des ondes bêta (β). (Voir figure 15)

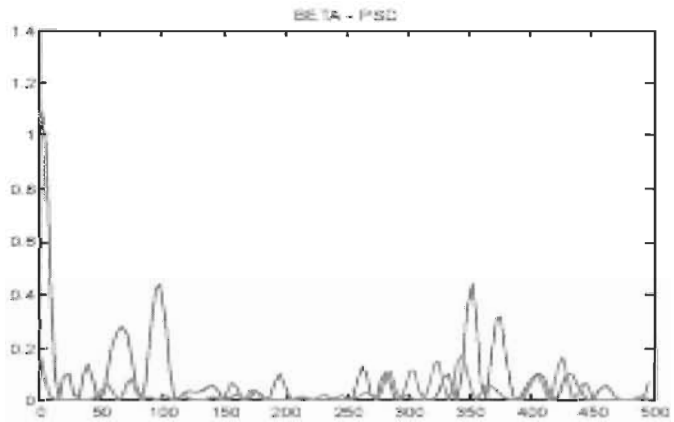

Figure 15 Densité du spectre de puissance de l'onde beta d'une personne normale (rouge) et d'une personne alcoolique (bleue), l'axe vertical représente la puissance et l'axe horizontal représente le temps en milliseconde [63]

Il ressort clairement des figures 13, 14 et 15 que l'activité alpha diminue chez une personne alcoolique par rapport à une personne normale. La même activité thêta est accrue chez une personne alcoolique par rapport à une personne normale. En se basant sur ces deux faits, une personne normale peut être différenciée de l'alcoolique par son EEG [63].

D'autre part, une étude prometteuse de Berdaa S. et Meunier F. [63] a justement basée la différentiation de l'état des facultés d'une personne sur les variations de l'activité alpha et bêta.

Dans cette dernière recherche, la puissance spectrale a été calculée, à partir des signaux EEG des personnes en état normal et en état de facultés affaiblies, à l'aide de la transformée de Fourier et appliquée sur les bandes de fréquences alpha (8–12 Hz) et bêta (13–30 Hz).

Nous nous sommes basés sur cette étude comme hypothèse pour le classement des signaux EEG que nous allons explorer dans le chapitre 3 portant sur la méthodologie de notre recherche.

2.7 Différentes approches de classification de signaux EEG

Les mesures EEG sont des séries temporelles. Il existe deux manières de classifier ce genre de mesures. Un classificateur statique qui utilise chaque mesure dans les séries temporelles comme caractéristique mais qui n'utilise pas l'information globale pouvant être extraite de ces séries. En revanche un classificateur dynamique utilise chaque série temporelle comme caractéristique.

2.7.1 Interface cerveau-ordinateur - vue d'ensemble

Une interface cerveau-ordinateur (BCI) est un système qui prend des entrées du cerveau, par exemple les EEG, et les traduit en commandes sur un ordinateur. Ceci est généralement fait par classification, où plusieurs commandes sont entraînées. L'entraînement consiste en l'exécution par l'utilisateur des différentes commandes. L'objectif est alors de construire un système qui utilise cette information pour classer les commandes futures [63].

Cependant, la classification des signaux EEG pour la détection de facultés affaiblies n'est pas un système d'interface cerveau-ordinateur vu qu'il n'y a pas de commandes à exécuter. De ce fait, tout le processus de classification est identique jusqu'à l'exécution des commandes qui n'est généralement pas la partie la plus difficile pour construire un système BCI [63].

Il existe cinq catégories d'algorithmes utilisés dans les systèmes de classification BCI : les classificateurs linéaires, les classificateurs bayésiens non linéaires, les classificateurs des plus proches voisins (Nearest Neighbors), les réseaux de neurones et une combinaison de classificateurs (Lotte et al., 2007). De bonnes performances BCI ont été obtenues dans toutes ces catégories, à l'exception des classificateurs de plus proches voisins, qui ne semblent pas très bien gérer la dimensionnalité [63].

Cependant, les réseaux de neurones et les classificateurs linéaires (type SVM) sont les plus populaires dans la recherche sur le BCI et constituent les méthodes également choisies dans ce présent projet de recherche pour faire la classification de signaux EEG [63].

2.7.2 Réseau de neurones artificiel et Perceptron multicouche

Perceptron multicouche signifie que le réseau de neurones comprend une couche d'entrée, possiblement une couche cachée au minimum, et une couche de sortie. La sortie de chaque nœud d'une couche se connecte à l'entrée de la couche suivante, mais pas à l'intérieur de la même couche.

L'inspiration vient du réseau de neurones du cerveau humain, mais en est une version très simplifiée. La grande force des réseaux de neurones réside dans le fait qu'ils ont un mécanisme d'approximation flexible et presque universel, ce qui leur permet d'avoir la configuration appropriée : les réseaux de neurones doivent être formés et entraînés avant de pouvoir être utilisés. En cas d'apprentissage supervisé, des ensembles d'apprentissage avec une entrée composée d'échantillons et la sortie attendue (étiquette) correspondante sont fournis au réseau, puis un algorithme ajuste les poids des synapses afin qu'il puisse mapper une entrée valide avec une sortie correcte. Un algorithme très courant pour cette phase d'apprentissage est l'algorithme de rétro propagation [63].

À l'exception des nœuds d'entrée, tous les autres nœuds verront leur entrée déterminée par la somme de toutes les synapses entrantes. Cela devient une entrée de la fonction d'activation dans les nœuds, qui détermine la sortie des nœuds : si elle "déclenche" ou non (Voir figure 16).

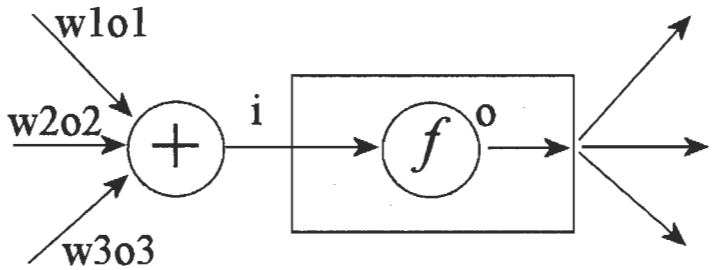

Figure 16 Neurone dans un réseau de propagation à réaction. f est la fonction d'activation [63]

2.7.3 Algorithme de retro propagation

Pour mieux comprendre les réseaux de neurones et le fonctionnement de la rétro propagation, l'algorithme d'apprentissage est présenté ci-dessous en pseudocode. Pour plus de détails, veuillez consulter Dario et Mattiussi (2008) [43] et NeuronDotNet (n.d.-a) [63].

Initialize the weights in the network

while stopping criterion has reached **do**

for all example e 2 in the training set **do**

O = actual, output (network, e); propagate forward

T = wanted output for e

Calculate error (T - O) at each neuron in the output layer

Compute Mean Squared Error value; propagate the error backward

Compute $delta_{weightupdate}$ for all weights

Update all the weights in the network such that the sum-squared value of error is minimized.

end for

end while

La valeur de l'erreur quadratique moyenne (MSE) est calculée à l'aide de l'équation (4). Cette valeur reflète l'efficacité de la formation effectuée jusqu'à présent. Le critère d'arrêt peut être soit lorsque la MSE a atteint une limite acceptable, soit lorsque le nombre de cycles de formation est atteint.

$$\hat{\theta} = E[(\hat{\theta}(X) - \theta)^2] \qquad (4)$$

Il existe plusieurs types de fonction d'activation couramment utilisées, celles-ci sont :

Fonction d'activation sigmoïde : $y = 1/(1 + Exp(-x))$

Fonction d'activation linéaire : $y = x$

Fonction d'activation logarithmique : $y = Log(1 + |x|)$

Fonction d'activation sinusoïdale : $y = Sin(x)$

Fonction d'activation Tanh : $y = Tanh(x)$

2.7.4 SVM (support vector machine)

SVM est un nouveau type de classificateur qui est basé sur deux concepts. Le premier concept comprend la transformation des données en un espace de grande dimension. Ce concept peut transformer des problèmes complexes (avec des surfaces de décision complexes) en des problèmes plus simples utilisant des fonctions discriminantes linéaires (voir figure 17). Et le second concept soujacent des SVM est motivé par la formation et l'utilisation uniquement des entrées proches de la surface de décision, car ils fournissent les informations les plus importantes sur la classification (voir figure 17). L'algorithme SVM est basé sur la théorie de l'apprentissage statistique. La SVM est utilisée pour la classification des données, la reconnaissance de formes, la bio-informatique et les analyses de régression en raison de leur précision et de leur capacité à traiter un grand nombre de prédicteurs. Le classificateur de vecteur de support présente de nombreux avantages. Avec SVM, les limites non linéaires peuvent être utilisées sans effort de calcul supplémentaire. De plus, les performances de SVM sont très compétitives par rapport aux autres méthodes. Un inconvénient des SVM est la complexité du problème qui n'est pas

de l'ordre de la dimension des échantillons, mais de l'ordre du nombre d'échantillons [63].

Un nouvel échantillon $\overrightarrow{x_i}$ peut-être classé selon :

$$\text{Si } \vec{w}.\overrightarrow{x_i} + b \geq 0 \text{ alors positive, sinon négative.}$$

Donc, l'orientation de la normale $\vec{w}$ et le biais b doivent être trouvés. Pour les échantillons de formation il faut que :

$$\begin{cases} \vec{w}.\overrightarrow{x_+} + b \geq 1 \\ \vec{w}.\overrightarrow{x_-} + b \leq -1 \end{cases} \quad (5)$$

Où $\overrightarrow{x_+}$ sont des échantillons positifs et $\overrightarrow{x_-}$ des échantillons négatifs. L'égalité correspondra aux points avec la plus petite marge. Il y aura au moins un vecteur de chaque classe donnant l'égalité, sinon la marge pourrait être augmentée davantage. Ces vecteurs sont appelés vecteurs de support. Le but est de trouver les $\vec{w}$ et b qui maximisent la marge.

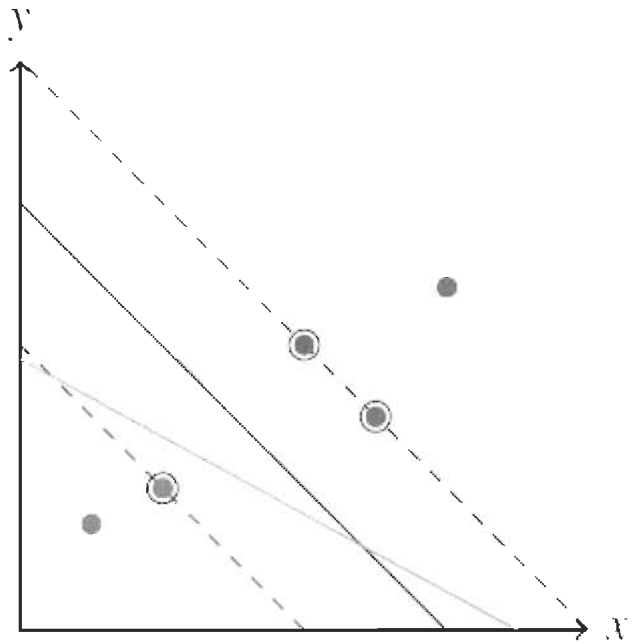

Figure 17 Illustration d'un classificateur de machine à vecteurs de support [63]

Dans la figure 17, la ligne noire continue représente la limite de classification qui maximise la marge des échantillons d'apprentissage. Les points entourés sur la ligne pointillée sont les vecteurs de support. La ligne noire continue représente la limite de classification qui maximise la marge des échantillons d'apprentissage. Les points entourés sur la ligne pointillée sont les vecteurs de support. Ils ont la plus petite marge à la limite de décision. La ligne pleine orange sépare également les données, mais sa marge est moins optimale [63].

2.8 La communication vocale

2.8.1 Mécanisme de production de la parole

Le locuteur initie le processus de génération de la parole en articulant un message dans son cerveau qu'il veut communiquer à l'auditeur par la parole. À l'étape suivante, le message est converti en un ensemble de séquence de phonèmes correspondant aux sons qui composent les mots, ainsi qu'une convention de prosodie représentant des caractéristiques telles que l'intensité, la hauteur et la durée du son. Par la suite le locuteur exécute une série de commandes neuromusculaires qui font vibrer les cordes vocales et contrôlent simultanément la position des organes articulatoires. Puis la séquence de sons de parole appropriée est créée et prononcée par le locuteur, qui fournit un signal acoustique comme sortie finale.

2.8.2 Mécanisme de perception de la parole

Le mécanisme de la perception de la parole commence lorsque le signal de parole est généré par le locuteur et transmis à l'auditeur. L'acoustique se propage le long de la membrane de base dans l'oreille interne, cela permet une analyse du spectre acoustique de l'onde acoustique d'entrée.

Le signal spectral est converti en signaux d'activité sur le nerf auditif par transduction neuronale. Cette dernière est un processus d'extraction de caractéristiques sonores. Les signaux d'activité sont en outre convertis en codes de langue comme indiqué sur la figure 18. Le cerveau transforme le code du langage en messages et la compréhension des messages est atteinte [65].

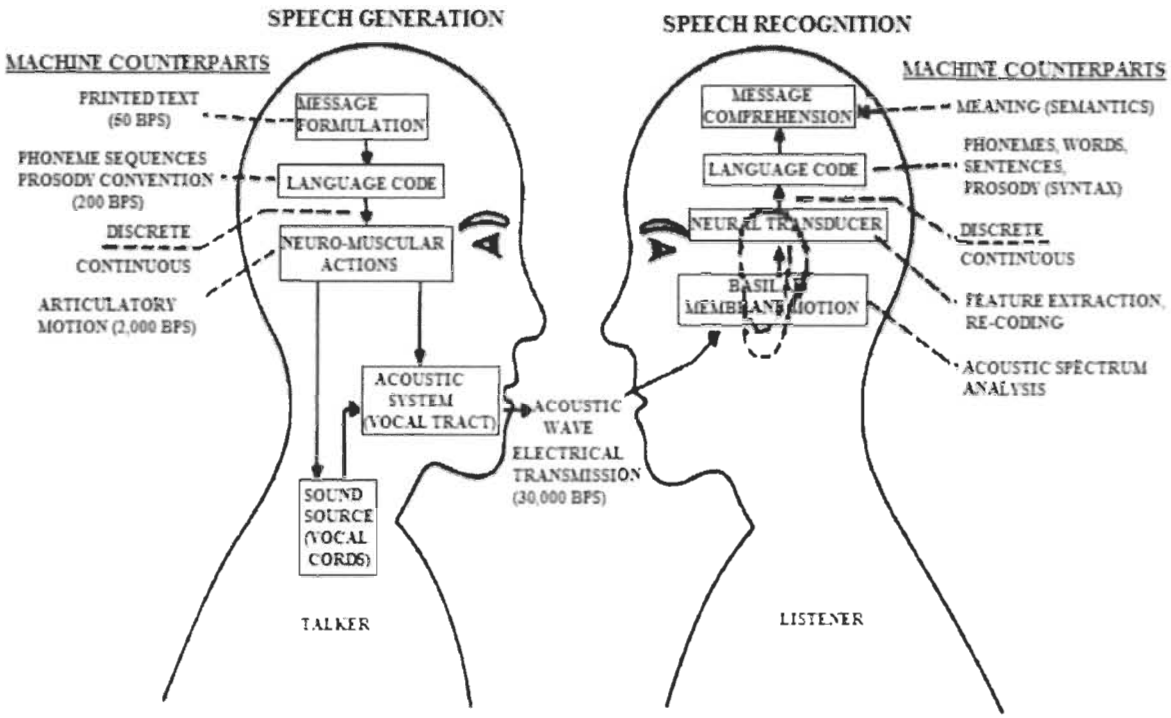

Figure 18 Mécanisme de production et de perception de la parole [66]

2.8.3 Classification des sons de la parole

2.8.3.1 Sons Sonores

La région voisée dans la parole est de nature presque périodique. L'air sortant des poumons est périodiquement interrompu par les cordes vocales vibrantes, ce qui se traduit par une onde glottale. L'onde glottale (flux d'air) est modulée par les organes articulatoires, ce qui entraîne une parole vocale. La région voisée a un caractère périodique dans le domaine temporel et une structure harmonique dans le domaine fréquentiel. Le spectre a plus d'énergie dans la région des basses fréquences.

2.8.3.2 Sons non vocaux

La région non voisée dans la parole a un bruit aléatoire comme ceux générés dans la nature (c'est-à-dire non périodique). L'air sortant des poumons n'est pas interrompu par les cordes vocales vibrantes. La fermeture partielle ou totale se produit le long de la longueur des voies vocales et entraîne une obstruction étroite ou complète du flux d'air. Le flux d'air entraîne l'arrêt ou l'excitation de la frication et est modulé par les organes articulatoires, ce qui entraîne une parole non vocalisée. La parole non voisée n'a pas de

périodicité et de forme d'onde semblable à du bruit dans le domaine temporel et n'a pas la structure harmonique dans le domaine fréquentiel. Le spectre a plus d'énergie dans les hautes fréquences.

2.8.3.3 Silence

La parole voisée et non voisée est séparée par des périodes de silence. Il n'y a pas d'air fourni aux voies vocales et donc pas de sortie de parole dans la région de silence comme le montre la figure 19. Une période de silence n'a pas de signal dans le domaine temporel et aussi aucune information spectrale dans le domaine fréquentiel [67].

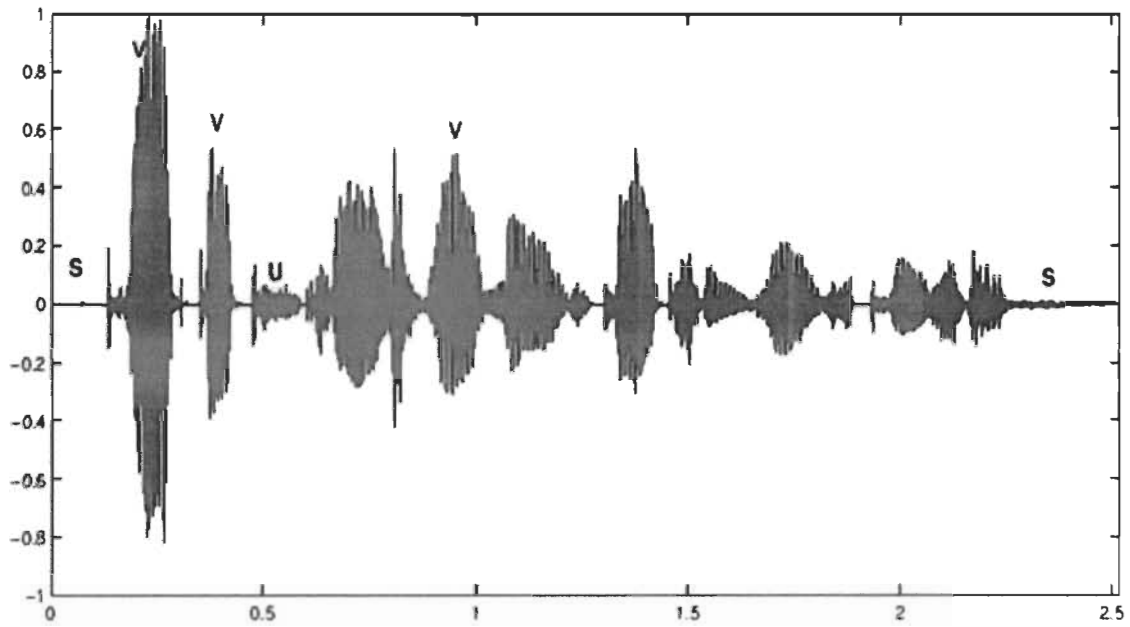

Figure 19 Forme d'onde d'amplitude du signal de parole étiqueté avec les régions V (voisé), U (non voisé) et S (silence) [65]

2.9 La parole et les sons du français

Comme nous avons expliqué précédemment, les sons résultants des mécanismes articulatoires font opérer l'ensemble de composants du système vocal.

Le mouvement de ces organes responsables de la parole peut conduire à la création d'environ 150 sons distincts qui forme le centre de toutes les langues de monde (Teston, 2007) [54].

En français, les phonèmes sont répartis en deux catégories, les voyelles et les consonnes.

D'une perspective articulatoire, la différence entre ces deux dernières est connexe à la circulation de l'air dans le conduit vocal. En effet, à l'encontre des voyelles dont la circulation d'air est proportionnellement libre à travers le conduit vocal, la production des consonnes est caractérisée par la contraction de conduit vocal qui engendre une suspension ou un désordre dans la circulation de l'air (Meunier, 2007) [68]. Néanmoins, la différence flagrante entre ces deux catégories est plus d'ordre linguistique qu'articulatoire.

Les voyelles sont primordiales pour une syllabe et y sont habituellement le centre, ce qui n'est habituellement pas le cas des consonnes. Si le passage de l'air est complétement ou partiellement engorgé en un ou plusieurs endroits, une consonne sera produite.

2.9.1 Les voyelles

Les voyelles, étant par essence portées par la voix, elles sont naturellement voisées et ce même s'il leur arrive de subir un processus de dévoisement, avec une atténuation de la vibration laryngale, dans le cas de la parole chuchotée, par exemple.

On distingue trois catégories de voyelles : les voyelles orales, les voyelles nasales et celles semi-voyelles qui sont représentées dans le triangle vocalique dans la figure 20 à partir de la relation entre le degré d'ouverture de la bouche et la position de la langue. Les voyelles en français sont entourées dans la figure 20.

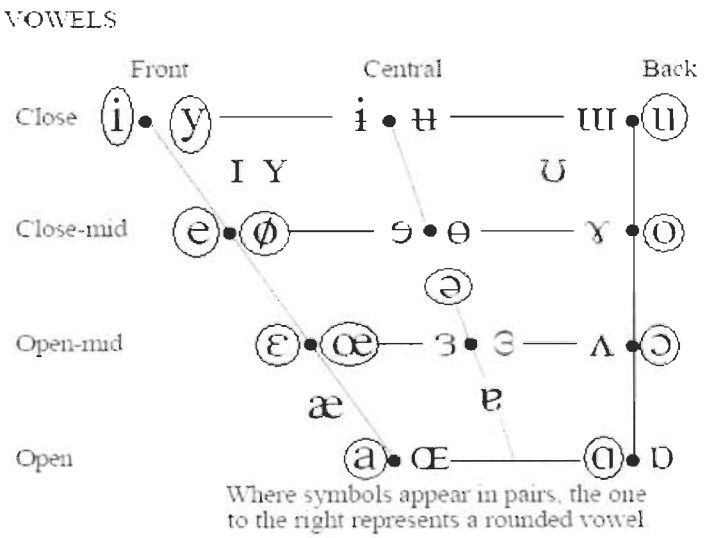

Figure 20 Trapèze vocalique représentant un classement des voyelles orales des langues du monde selon des critères articulatoires

Dans la plupart des cas, les voyelles orales sont exposées sous la forme d'un trapèze, appelé vocalique, qui joue le rôle d'une représentation graphique de la bouche et qui prend en compte les dimensions de l'ouverture de cette dernière et la hauteur de la langue (F1) et de la position de la langue et des lèvres (F2) comme présentée dans la figure 21.

Le trapèze vocalique permet de classer les voyelles selon deux axes comme dans la figure 20 : un axe horizontal représentant la profondeur du point d'articulation avec la position de la langue, et un axe vertical qui représente le degré de l'ouverture de la bouche lors de l'émission d'un son appelé aperture.

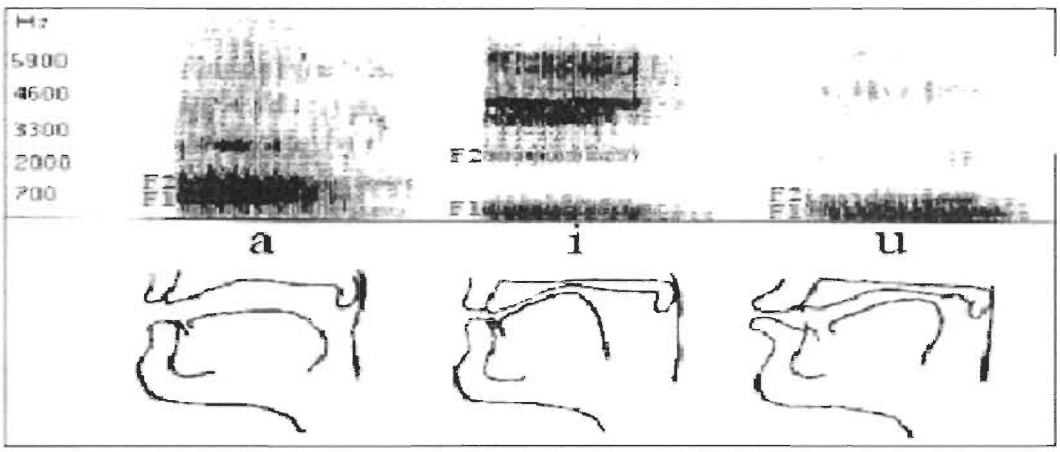

Figure 21 Valeurs formantiques des voyelles a, i, u (en haut) et coupe sagittale du tractus vocal (en bas) avec les différentes positions de la langue pour la production de ces trois voyelles, les plus distinctes du point de vue articulatoire [70]

2.9.2 Les consonnes

Les consonnes, contrairement aux voyelles, sont considérées comme un ensemble de sons hétérogènes, qui peuvent êtres sonores ou sourdes. Elles sont décomposées en deux classes : les occlusives et les fricatives.

Les consonnes occlusives sont produites suite à une tenue pendant laquelle le flux d'air est engorgé dans le conduit vocal suivie d'une explosion et du relâchement de l'occlusion permettant ainsi la libération de flux d'air ce qui produit à l'écoute une impression de frottement ou de friction. Cependant, les consonnes fricatives se distinguent par la présence d'un bruit durant leurs tenues [71].

2.10 Dysarthrie et problème d'élocution

La dysarthrie est un trouble de la parole, non pas en raison d'un problème de phonation, mais d'un dysfonctionnement des organes qui peut être causé par des maladies neurologiques telles que la maladie de Parkinson, la paralysie cérébrale, etc. L'analyse acoustique de la parole dysarthrique vise à comprendre la production de la parole dysarthrique et à développer des méthodes pour détecter la maladie à un stade précoce ou pour quantifier le degré de gravité et de suivre ses progrès. La dysarthrie chez les patients atteints de la maladie de Parkinson est appelée hypokinétique. Des études antérieures de locuteurs souffrant de parkinson ont montré des déficits dans l'articulation des voyelles [72].

Les analyses de la dysarthrie peuvent impliquer des marqueurs acoustiques qui documentent les propriétés phonatoires, prosodiques et articulatoires de la parole [73].

L'analyse est basée principalement sur des voyelles soutenues, des mots ou des phrases isolés. L'obtention de marqueurs acoustiques pour ces derniers implique une segmentation manuelle. Les études des articles lus ou de la parole spontanée sont peu nombreuses [74] et peuvent demander un grand nombre d'indices acoustiques (J. R. Orozco-Arroyave et al., 2016) provoquant des difficultés d'interprétation. De plus le faible pouvoir discriminant de ces indices rendent difficile la distinction entre les locuteurs dysarthriques et ceux témoins. Le spectre moyen à long terme (LTAS) de la parole continue a été utilisé dans plusieurs études pour rendre compte de la qualité de la voix ainsi que d'autres propriétés ou états du locuteur. La plupart des évaluations cliniques basées sur le LTAS de la qualité de la voix se sont concentrées sur les patients atteints de dysphonie causée par des maladies organiques ou fonctionnelles du larynx. Seules quelques études ont ciblé le discours dysarthrique [74].

D'autre part, une recherche basée sur une analyse de la parole dysarthrique au moyen des composantes principales du spectrogramme a été publiée par Grenez F. et al. [75].

L'analyse en composantes principales (ACP) implique une réduction de la dimensionnalité qui permet la projection d'un spectrogramme dans un sous-espace de plus faible dimension tout en conservant la majeure partie de la variation des données

originales. Chaque composante principale (PC) du spectrogramme est une somme pondérée de tous les spectres.

Les marqueurs acoustiques utilisés pour documenter les déficits d'articulation chez les locuteurs souffrants de Parkinson sont des moyennes spectrales de la première composante principale et les deux sommes pondérées des spectres de trame associées aux coefficients positifs et négatifs de la deuxième composante principale.

L'analyse est appliquée à un corpus comprenant des monologues produits par 50 témoins et 50 locuteurs souffrants de Parkinson et les résultats montrent que les marqueurs acoustiques diffèrent statistiquement de manière significative entre ces deux groupes de locuteurs [75].

Cependant d'autres recherches se sont intéressées au diagnostic de la parole afin de pouvoir détecter le problème de dysarthrie, une hypothèse appliquée dans une étude de détection de la maladie de Parkinson à partir de la neuromécanique de l'articulation de la parole (Pedro Gómez-Vilda et al., 2017) [76] a permis d'explorer une autre piste, se basant sur la fonction de densité de la probabilité de la vitesse articulaire absolue comprend des informations sur la stabilité de la phonation lorsqu'elle est appliquée à des voyelles soutenues, ainsi que sa fluidité si elle est appliquée à la parole connectée.

Cette étude a été réalisée sur un ensemble de données de voyelles prolongées enregistrées chez des patients atteints de la maladie de Parkinson contraste avec des enregistrements similaires de sujets normaux. Un réseau de transmission directe des moindres carrés aléatoires (RLSFN) a été utilisé comme classificateur binaire travaillant sur les jeux de données pathologiques et normaux. Deux ensembles de données pour chaque sexe ont été testés (Hommes et femmes), dont 26 sujets normaux et 53 pathologiques dans l'ensemble masculin, et 25 normaux et 38 pathologiques dans l'ensemble féminin.

Les sous-ensembles de données masculins et féminins ont été testés en une seule fois, donnant des taux d'erreur inférieurs à 0.6% avec une précision supérieure à 99.4%. Les résultats sont pertinents suite à l'utilisation d'une méthodologie selon laquelle l'utilisation des descripteurs hautement normalisés comme distribution de probabilité des variables cinématiques de la stabilité de l'articulation des voyelles, qui a des propriétés

intéressantes en termes de théories de l'information, augmente le potentiel des classificateurs simples mais puissants pour produire des résultats de détection tout à fait acceptables pour la détection de la maladie de Parkinson.

2.11 La représentation de la parole dans les domaines temps et fréquence

Les variabilités complexes de la parole rendent l'analyse plus compliquée mais fournissent des informations supplémentaires en utilisant la variation de tonalité et d'amplitude.

L'analyse de la parole et du texte a de vastes applications dans le monde actuel. Ils ont des représentations différentes, et de nombreuses similitudes. Plusieurs défis sont rencontrés dans leur analyse.

Examinons un signal vocal, comme schématisé à la figure 22, extrait d'un énoncé de la base de données vocale CMU US RMS ARCTIC [77]. Chacun des énoncés est enregistré en tant que signal 16 bits à une fréquence d'échantillonnage de 16 kHz, ce qui signifie qu'il y a 16000 échantillons pour chaque seconde du signal et chaque échantillon a une résolution de 16 bits par échantillon. La fréquence d'échantillonnage d'un signal audio détermine la résolution des échantillons audio, plus la fréquence d'échantillonnage est élevée, plus la résolution du signal est élevée. Le signal de parole est lu à partir du fichier 'arctic_a0005.wav' dans la base de données de parole qui a une durée d'environ 1,4 seconde, équivalent à une séquence de 22640 échantillons, chaque échantillon étant un nombre de 16 bits. La représentation vocale exposée dans la figure 22 est un tracé du signal vocal de «arctic_a0005.wav» dont le texte équivalent est «l'oublierons-nous jamais».

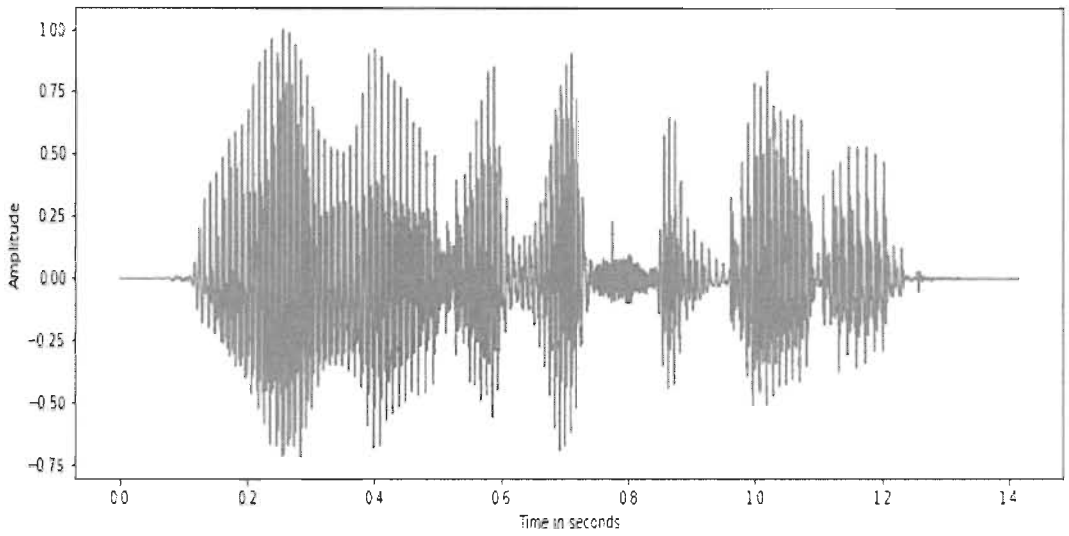

Figure 22 Signal de parole, s [n] pour l'énoncé « l'oublierons-nous jamais » [78]

On peut voir sur la figure 22 que la parole peut être représentée comme une variation d'amplitude en fonction du temps. L'amplitude est normalisée de telle sorte que la valeur maximale soit 1. La parole est essentiellement une séquence de phonèmes. Le signal vocal peut être segmenté en une séquence de phonèmes et de segments silence / non vocal. Dans la figure 23, nous exposons la représentation d'une partie du signal vocal montrant les phonèmes et leur intervalle de temps correspondant.

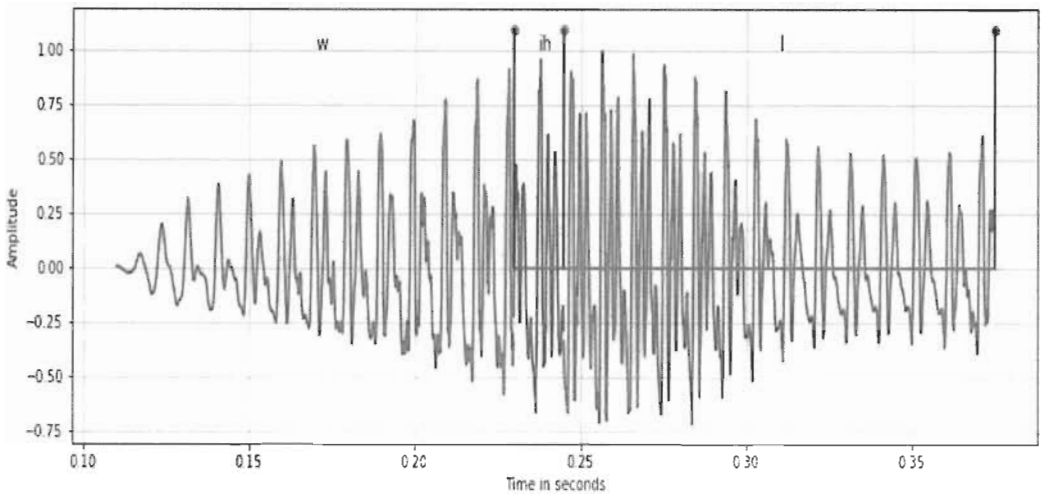

Figure 23 Segments de niveau phonème du signal pour «will» comme «w», «ih» et «l» [77]

On voit sur la figure 23 que les phonèmes «w», «ih» et «l» sont de nature quasi-périodique et sont classés comme phonèmes vocaux car ils sont produits par la vibration périodique des cordes vocales. De plus, «ih» est une voyelle tandis que «w» et «l» sont des semi-voyelles. Les classes sonores et non vocales sont la large catégorisation des sons de la parole basée sur la vibration des cordes vocales. L'étude et la classification des différents phonèmes est appelées phonétiques.

De plus, nous examinerons la représentation du domaine de fréquences au niveau de la trame de la parole, également connue sous le nom de transformée de Fourier à court terme (STFT) dans la communauté de la recherche vocale.

2.11.1 Le spectrogramme

La représentation visuelle des fréquences d'un signal donné dans le temps est appelée spectrogramme. Dans un tracé de représentation de spectrogramme, un axe représente le temps, le deuxième axe représente les fréquences et les couleurs représentent la magnitude (amplitude) de la fréquence observée à un moment particulier.

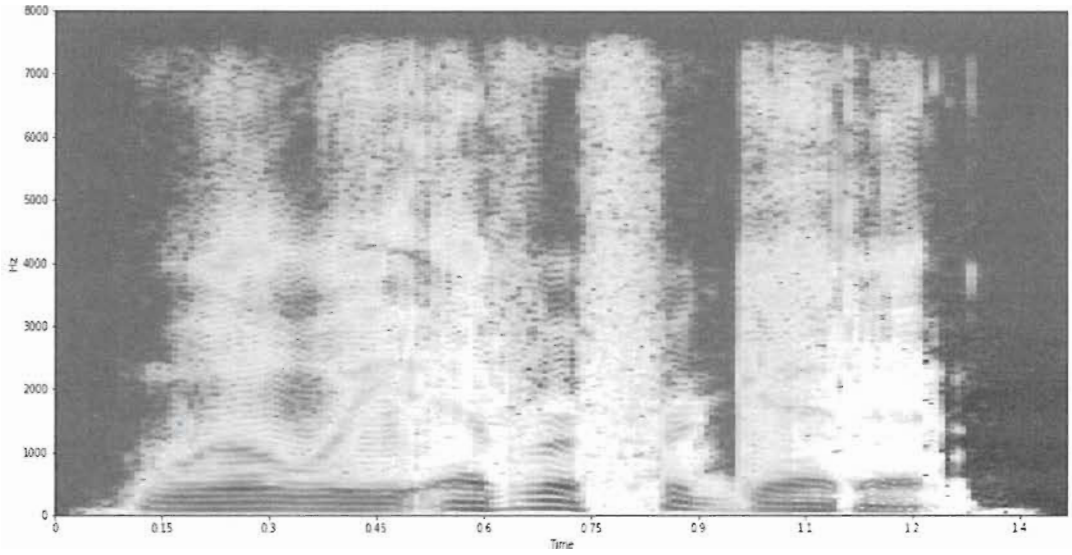

Figure 24 Spectrogramme à échelle logarithmique du signal de parole utilisant une taille de fenêtre de 30 ms et une taille de saut de 7,5 ms

Le spectrogramme à l'échelle logarithmique tracé dans la figure 24 permet de visualiser l'amplitude exprimée dans l'échelle logarithmique des harmoniques déduites par l'application de la STFT d'un signal vocal. Une taille de trame de 30 ms est choisie, ce qui équivaut à 30 ms x 16 kHz = 480 échantillons du signal de parole, tandis qu'un décalage de trame de 7,5 ms équivaut à 120 échantillons. La raison derrière la division du signal de parole en trames de petite durée est que le signal de parole n'est pas stationnaire et ses caractéristiques temporelles changent très rapidement. Ainsi, en prenant une petite taille de trame, nous faisons l'hypothèse que le signal de parole sera stationnaire et que ses caractéristiques ne varieront pas beaucoup à l'intérieur de la trame. En outre, un décalage de trame plus court est choisi pour suivre la continuité du signal de parole et ne pas manquer de changements brusques aux bords des trames. On peut voir sur le graphique de la figure 24 que la représentation du domaine fréquentiel de chaque trame nous aide à mieux analyser le signal vocal; car nous pouvons plus facilement voir les harmoniques comme des segments rouges parallèles dans les régions voisées et comment l'amplitude varie pour chaque fréquence et chaque index de trame. Par conséquent, la plupart des analyses du signal de parole se font dans le domaine fréquentiel. Mais

l'extraction d'informations temporelles telles que les changements brusques du signal (début de salve comme «t») sont mieux capturés dans le domaine temporel car la division du signal de parole en trames rejette les changements instantanés du signal.

Nous pouvons dire que nous pouvons obtenir une meilleure résolution temporelle dans le domaine fréquentiel en prenant une taille de trame plus petite. Mais il y a un compromis entre la résolution dans le domaine temporel et fréquentiel. Prendre une très petite taille de trame donnera une résolution plus élevée dans le temps mais donnera peu d'échantillons dans une seule trame et les composantes de Fourier correspondantes auront peu de composantes de fréquence. Et prendre une taille d'image plus grande donnera une résolution temporelle plus faible mais une résolution en fréquence plus élevée en raison du nombre plus élevé d'échantillons. Il n'est donc pas possible d'obtenir simultanément une haute résolution en temps et en fréquence.

Nous observons que l'axe des y dans le tracé du spectrogramme logarithmique a une fréquence allant jusqu'à 8 kHz pour une fréquence d'échantillonnage de 16kHz. En effet, selon le théorème de Nyquist-Shannon [79], la fréquence maximale qui peut être observée dans un signal discret est au plus la moitié de la fréquence d'échantillonnage qui est de 8 kHz

2.12 Les différentes approches de classification de la parole

Il existe un certain nombre de techniques disponibles pour l'extraction des caractéristiques vocales comme : PLP (Perceptual Linear Predictive Coefficient), LPC (Linear Predictive Coder Analysis), LPCC (Linear Predictive Cepstral Coefficient), MFCC (Mel Frequency Cepstral Coefficient), FFT (Power Spectral Analysis), MEL (Mel Scale Cepstral Analysis), Dans la prochaine section de ce mémoire, nous allons donner une idée générale de ces approches de classification de la parole.

2.12.1 PLP Perceptual Linear Prediction

Le modèle PLP a été développé par Hermensky en 1990, dont le but était d'élaborer la compréhension de la psychophysique de l'audition humaine et entendre avec précision lorsque les caractéristiques sont extraites par un processus d'extraction.

Le calcul des coefficients cepstraux PLP, utilise des fonctions PLP qui sont déjà définies dans une bibliothèque d'analyse. La méthode mentionnée est vulnérable lorsque les valeurs spectrales à court terme [80] ont été modifiées par la réponse en fréquence des canaux de communication.

Afin d'effectuer les calculs basés sur l'analyse PLP, nous définissons des directives qui régiront ce processus de calcul. Le PLP fonctionne de manière similaire à celle de l'analyse LPC basée sur le spectre fréquenciel à court terme des signaux de la parole par certaines transformations basées sur la psychophysique.

PLP a un vecteur caractéristique résultant de faible dimension. Les pics de la PLP sont indépendants de la longueur du tractus vocal.

L'inconvénient de ce modèle est qu'en raison du canal de communication, le bruit et l'équipement utilisé provoquent une altération du canal spectral.

2.12.2 LPC Linear Predictive Analysis

Le LPC est la méthode la plus profonde et la plus puissante utilisées pour coder la parole de qualité à faible débit binaire. Le LPC est une approche utilisée principalement dans le traitement du signal audio et le traitement de la parole. Elle permet à des échantillons de paroles spécifiques d'être approximés en temps réel sous la forme d'une combinaison linéaire d'échantillons de la parole précédents.

Le modèle LPC est basé sur la production de la parole humaine. En utilisant un modèle de source de filtre conventionnel, où les voies vocales glottales et la fonction de transfert de rayonnement labial sont interprétées en un seul filtre multipolaire [81]. Son principe consiste juste à minimiser la somme des différences quadratiques entre la parole (originale) et le signal de parole estimé sur une durée finie.

Le LPC est une approche simple à mettre en œuvre et mathématiquement précise et nous pouvons la considérer comme une puissante technique d'analyse de la parole. En revanche, elle a l'inconvénient d'avoir des composants caractéristiques hautement corrélés.

2.12.3 LPCC Linear Predictive Cepstral Coefficient

Le codage par prédiction linéaire est une approche alternative pour l'estimation de l'enveloppe spectrale. Cette méthode est également connue sous le nom de modèle tout-pôle ou modèle autorégressif.

Les coefficients cepstraux de la prédiction linéaire LPCC sont des coefficients cepstraux dérivés de l'enveloppe spectrale calculée par la méthode LPC.

Dans l'approche LPCC, la seule différence avec l'approche MFCC, une approche que nous allons introduire dans la suite, possède aussi un processus de fenêtrage mais l'étape de déformation est différée avec l'approche LPCC, les composants de caractéristique sont décorrélés en raison de l'analyse spectrale. L'approche LPCC a une meilleure robustesse par rapport à LPC.

Dans LPCC, les échelles linéaires ne sont pas adéquates pour représenter la production ou la perception de la parole. Le processus LPCC est illustré dans la figure 25.

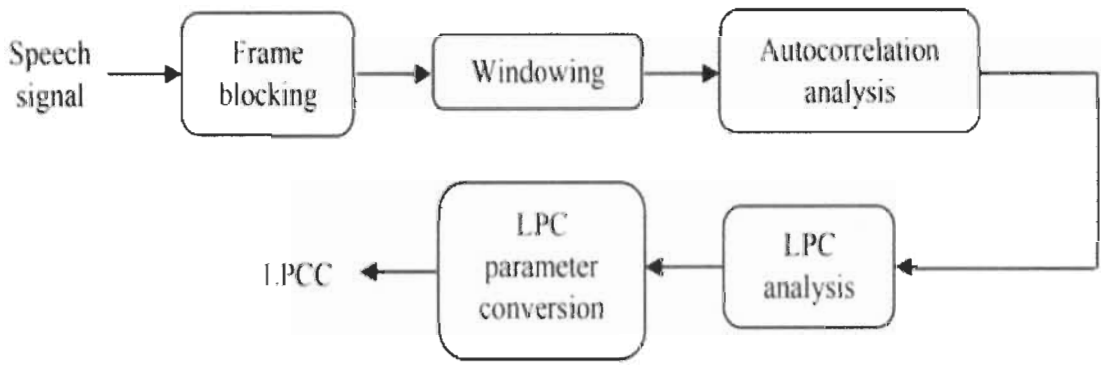

Figure 25 Schéma fonctionnel du processus LPCC [27]

Les LPCC ont une faible vulnérabilité au bruit. Les fonctionnalités LPCC produisent un taux d'erreur inférieur à celui des fonctionnalités LPC [82]. Les coefficients cepstraux d'ordre supérieur sont mathématiquement limités, ce qui entraîne une gamme extrêmement étendue de variances lors du passage des coefficients cepstraux d'ordre inférieur aux coefficients cepstraux d'ordre supérieur [83]. De même, les estimations LPCC sont réputées pour leur grande sensibilité au bruit de quantification. L'analyse

cepstrale sur un signal de parole aigu donne une petite séparabilité source-filtre dans le domaine fréquentiel [84]. Les coefficients cepstraux d'ordre inférieur sont sensibles à la pente spectrale, tandis que les coefficients cepstraux d'ordre supérieur sont sensibles au bruit.

2.12.4 MFCC Mel Frequency Cepstral Coefficient

MFCC est une approche considérée comme faisant partie des méthodes standards pour l'extraction d'entités vocales. Grâce à de nombreuses années de recherche et d'analyse sur le logiciel de reconnaissance, une grande variété de représentations de signaux vocaux ont été expérimentées. La méthode MFCC est le plus populaire et la plus précise de toutes les approches connues. La méthode MFCC fonctionne en réduisant les informations de fréquence des signaux vocaux donnant un plus petit nombre de coefficients. C'est un modèle simplifié du traitement auditif des signaux qui est relativement rapide et facile à calculer. Dans la figure 26, le procesus d'extraction des coefficients cepstraux de la méthode MFCC est schématisé.

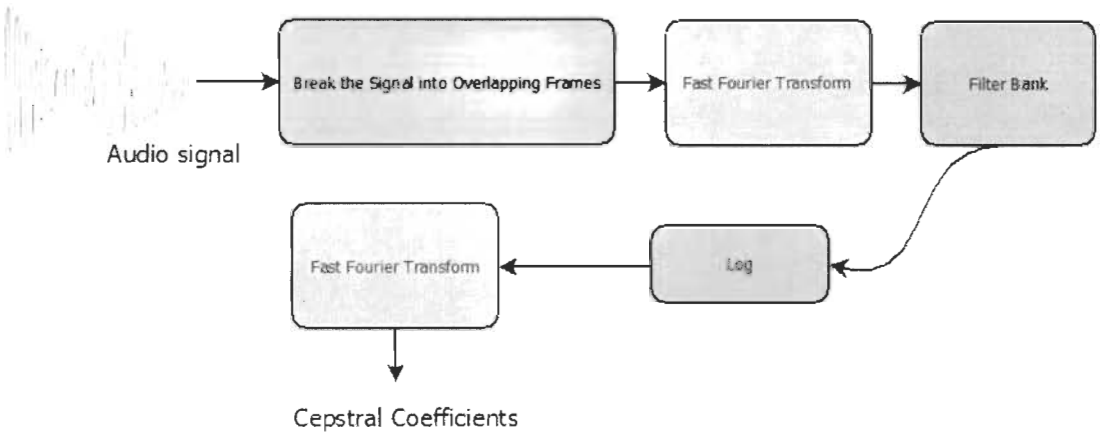

Figure 26 Schéma fonctionnel du processus MFCC [27]

2.12.5 FFT Power Spectral Analysis

L'une des techniques courantes de réalisation d'études sur un signal spectral est le spectre de puissance. Les composantes fréquentielles du signal dans le temps sont décrites dans le spectre de puissance du signal vocal. Durant l'analyse, une question majeure est généralement posée : la puissance d'un signal est contenue dans quelles fréquences spécifiques ? Cette réponse est : Le spectre de puissance. Il s'agit généralement d'une

forme de distribution de la valeur de puissance en fonction de la fréquence, où la moyenne du signal de la parole est caractérisée pas son amplitude. Dans le domaine fréquentiel, c'est plutôt la magnitude de la FFT au carré.

Le spectre de puissance peut être déduit pour le signal entier en appliquant une méthode simple qui calcule la moyenne des segments de périodogrammes du signal temporel donnant la sortie pour former la densité spectrale de puissance PSD.

En effet, cette moyenne des segments de périodogrammes de signaux de longue durée attribue en outre la puissance aux fréquences appropriées et pertinentes avec une précision absolue, réduisant ainsi les fluctuations causées par le bruit dans le spectre de puissance [85].

Cette réduction de la résolution en fréquence est due à un sous-échantillonnage des points de données qui sont maintenant disponibles pour le calcul de la FFT. Le fenêtrage spectral, qui permet d'attribuer une pondération à chaque segment, est la méthode la plus efficace pour améliorer la densité spectrale de puissance. Cependant, la plupart des fonctions de fenêtrage de par leurs pondérations non uniformes offrent une plus grande importance aux données au centre par rapport aux données sur les bords de chaque segment, ce qui représente une perte d'information. Pour diminuer cette perte, les segments de données adjacents sont alors chevauchés dans le temps.

Le calcul du spectre de puissance consiste alors à effectuer la Transformée de Fourier Discrète (DFT) qui met en évidence les informations de fréquence d'un signal capturé dans le domaine temporel. Les informations concernant l'amplitude et la phase d'un signal audio original dans le domaine fréquenciel sont contenues dans la sortie résultante.

2.12.6 MEL Mel Scale Cepstral Analysis

L'analyse MEL est très similaire à la méthode PLP. Dans l'approche MEL, le spectre de fréquences est toujours adapté (mis à l'échelle) conformément à l'échelle MEL, tandis que dans l'approche PLP, cette mise à l'échelle est effectuée conformément aux coefficients d'échelle BARK.

Les analyses d'échelle MEL offrent également une option similaire à celle de l'approche PLP d'utiliser un filtre RASTA pour la compensation du canal linéaire de distorsions.

Tous les modèles de pôles sont utilisés dans le PLP, cela aide à lisser le spectre de puissance modifié tandis que le lissage cepstral est utilisé dans la méthode MEL pour lisser le spectre de puissance modifié.

2.13 Conclusion

Pour résoudre la problématique de détection des facultés affaiblies, la présente recherche se base sur une approche qui permet la détection des déformations vocales ainsi que la variation d'activité cérébrale (EEG) découlant de la consommation du cannabis ou de l'alcool.

À partir des différentes approches décrites dans ce chapitre, nous allons dans le prochain chapitre (chapitre 3) nous inspirer des méthodes répertoriées pour déduire notre propre méthode expérimentale.

Chapitre 3 : Méthodes expérimentales

3.1 Introduction

Dans ce chapitre, nous allons exposer la méthodologie expérimentale de cette recherche résumée dans le diagramme présenté dans la figure 27. Tout d'abord nous allons énumérer les dispositifs matériels que nous avons utilisés pour la capture des signaux EEG et vocaux. Ensuite, nous allons énoncer les méthodologies et les approches déployées pour la classification des signaux EEG et vocaux capturés lors d'expérimentations faites avec des volontaires testés d'abord dans leur état normal et ensuite avec les facultés affaiblies. Les expérimentations comportaient la prononciation de mots prédéfinis de complexité diverses. Les modèles des classificateurs développés pour la détection des facultés affaiblies sont aussi exposés.

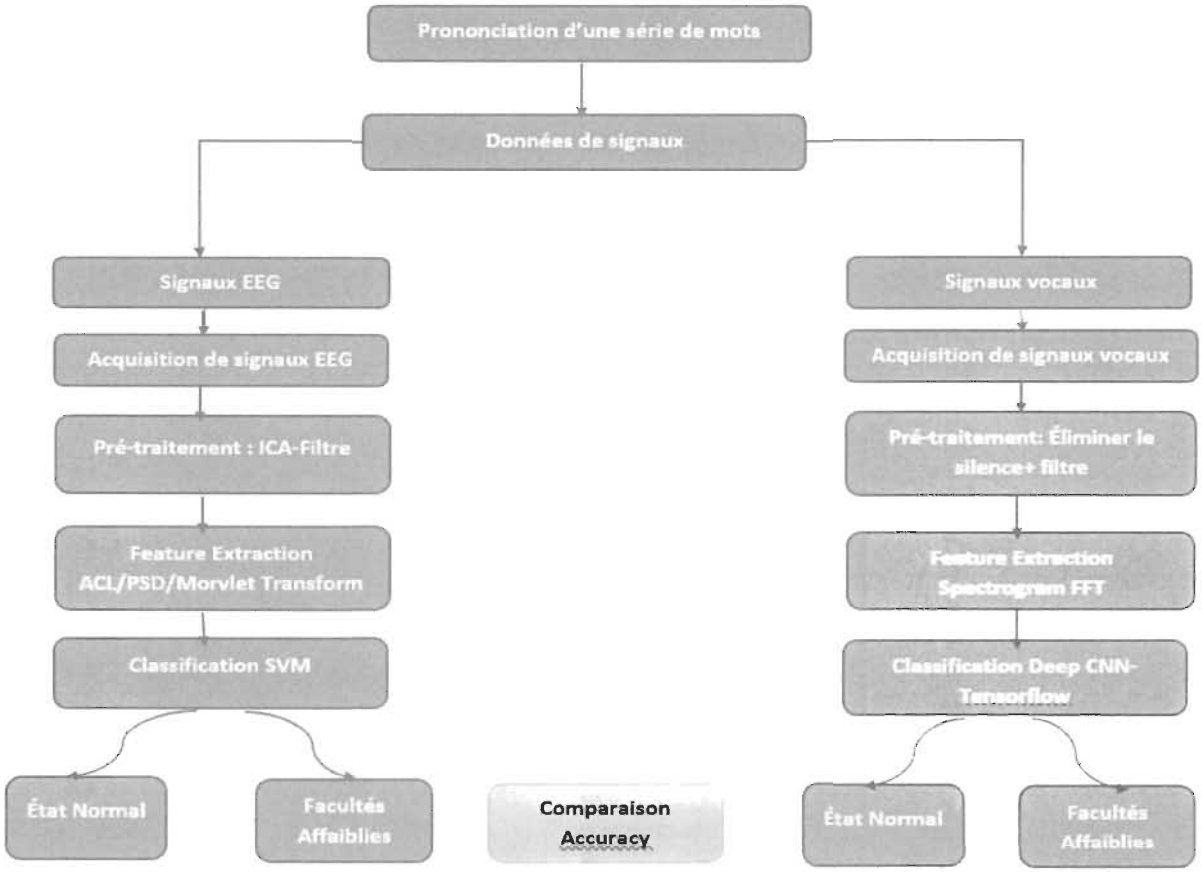

Figure 27 Schéma de la méthodologie proposée pour la classification de signaux vocaux et de signaux EEG

3.2 Casque Multi Chanel Emotiv Epoc +

Pour la capture des signaux EEG, nous avons utilisé le casque Emotiv Epoc + à 14 électrodes. Ce casque est un système portable de haute résolution, multicanal et conçu pour des applications de recherche pratique. Il sert à recevoir des données EEG brutes à l'aide d'une boîte à outils logicielle exclusive qui comporte les API et les bibliothèques, des commandes mentales et des mesures de performance. Les 14 électrodes représentent les 14 canaux EEG basé sur les emplacements du système international 10-20 et ayant les identifiants suivants : AF3, F3, F7, FC5, AF4, F4, F8, FC6, T7, CMS, P7, O1, T8, DRL, P8 et O2 (présentés dans la figure 28).

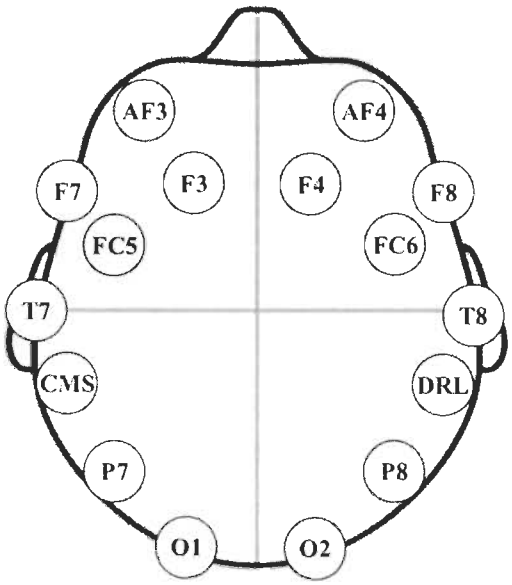

Figure 28 Emplacement des 14 électrodes du Emotiv Epoc+ sur le cuir chevelu

Le casque Emotiv Epoc + que nous avons utilisé dispose d'une connexion sans fil Bluetooth qui permet de faire le couplage avec le logiciel Emotiv Xavier Control Panel (voir figure 29).

Une fois la connexion Bluetooth est établie, le logiciel nous donne la possibilité de visualiser la qualité de signal des 14 électrodes (voir figure 29).

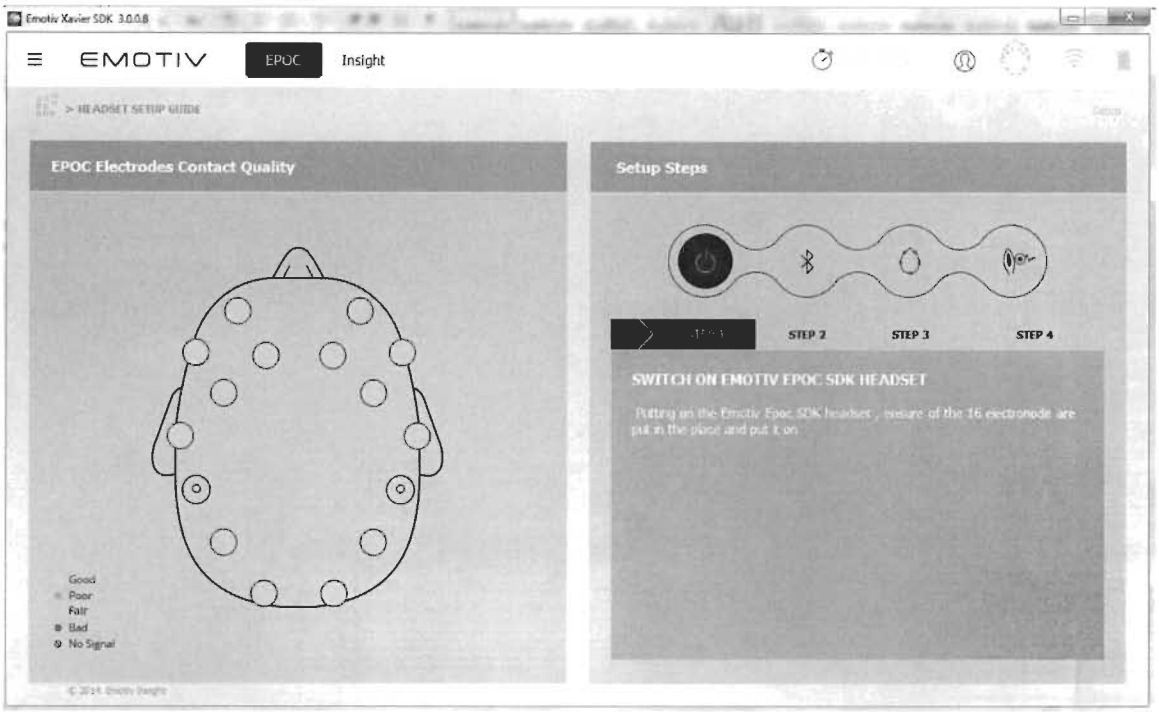

Figure 29 Qualité de signal des 14 électrodes sur Emotiv Xavier Control Panel

La qualité de signal est traduite par un code couleur, la couleur verte traduit une excellente qualité de signal, la couleur orange pour une qualité moyenne et celle rouge pour une qualité insuffisante. Si toutes les électrodes sont vertes comme sur la figure 29, nous pouvons commencer à acquérir les signaux EEG.

3.3 Blue Yeti Pro

Le microphone d'enregistrement audio Yeti Pro (voir figure 30) que nous avons utilisé pour faire des enregistrements vocaux lors de l'expérience est caractérisé par une précision exceptionnelle d'un enregistrement audio avec une fréquence d'échantillonnage de 129 Khz et une quantification sur 24 bits ce qui lui permet d'être le microphone idéal pour capturer des signaux audios en stéréo car ce mode d'enregistrement est recommandé pour la capture de la voix.

Le microphone Yeti pro peut prélever des fréquences allant de 22 Khz à 192 Khz, offrant une résolution plus de quatre fois supérieure à celle d'un enregistrement CD typique (44 KHz) mais créant proportionnellement davantage de données à capturer par ordinateur.

Figure 30 Microphone d'enregistrement Audio Blue Yeti Pro

Après avoir connecté le microphone à l'ordinateur par câble USB, nous faisons en sorte que les sujets soient assis droit devant le microphone afin de commencer la capture des enregistrements audio des mots prononcés par ces sujets. Les signaux vocaux sont enregistrés directement sur le disque à l'aide du logiciel d'enregistrement de son et d'édition de source audionumérique Audacity. Ces signaux vocaux seront ultérieurement traités pour supprimer tout bruit ou articulations étrangères avant ou après les énoncés des mots utilisés durant les expérimentations.

3.4 Audacity

Audacity est un logiciel de traitement et d'édition audio gratuit. Il permet notamment de modifier des fichiers audios pour leur appliquer des effets, d'en changer les paramètres ou encore de les raccourcir. Il comporte deux fenêtres principales : le menu complètement en haut et le plan de travail.

Grâce à Audacity, nous avons effectué les enregistrements audios mono, c'est-à-dire sur un seul canal puisque tous les sons que l'on enregistre sont monophoniques. Ces enregistrements vocaux ont été configurés comme sur la figure 31, avec un format d'échantillonnage (appelé aussi profondeur de bits sur 16 Bits) et une fréquence d'échantillonnage de 44 100 Hz.

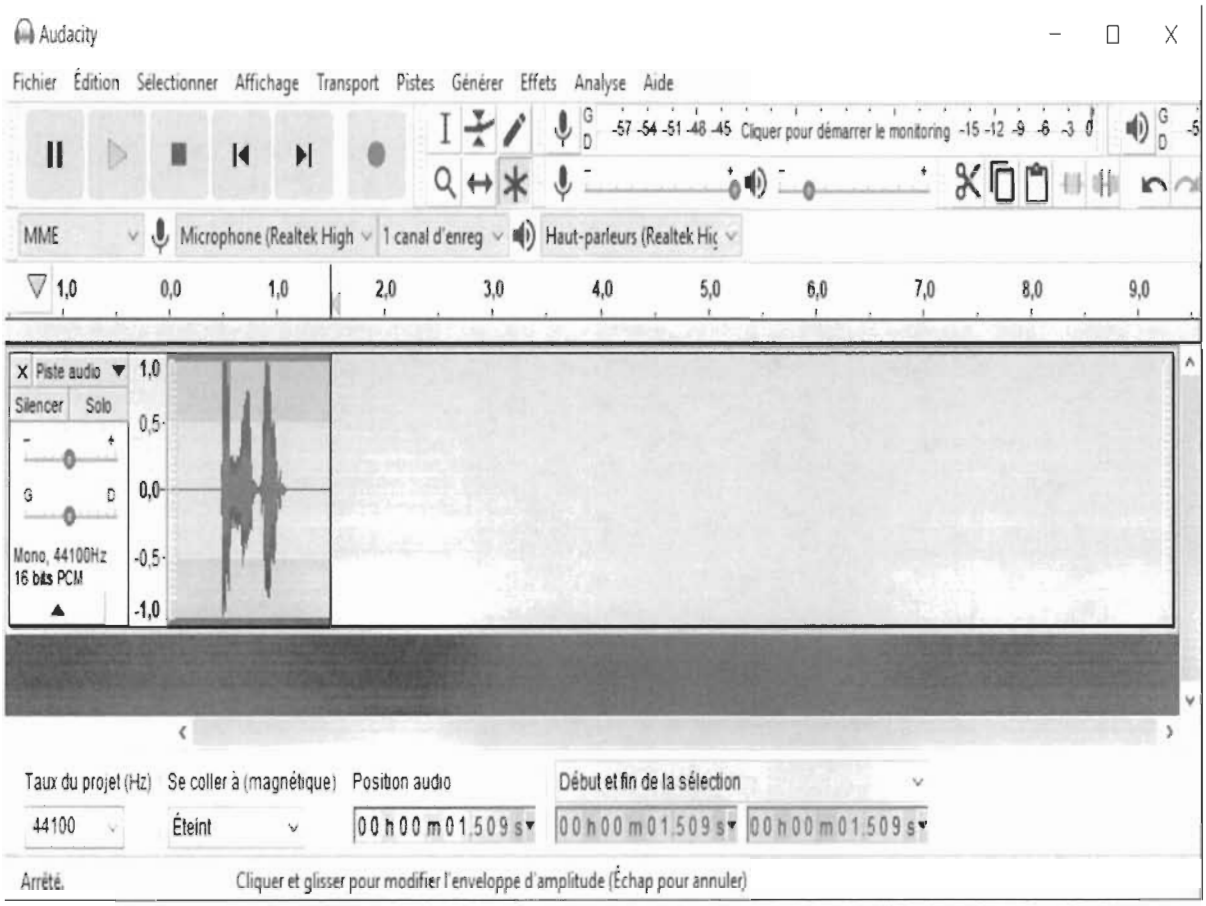

Figure 31 Interface du logiciel Audacity et exemple de signal vocal enregistré

3.5 OpenVibe

OpenVibe est un logiciel pour les neurosciences en temps réel (c'est-à-dire pour le traitement en temps réel des signaux cérébraux). Cette plateforme est consacrée à la conception, au test et à l'utilisation d'interfaces cerveau-ordinateur livrée sous forme de package comprenant un outil de conception pour créer et exécuter des applications personnalisées, ainsi que plusieurs programmes de démonstrations prêts à l'emploi. La figure 32 représente un diagramme des éléments de OpenVibe.

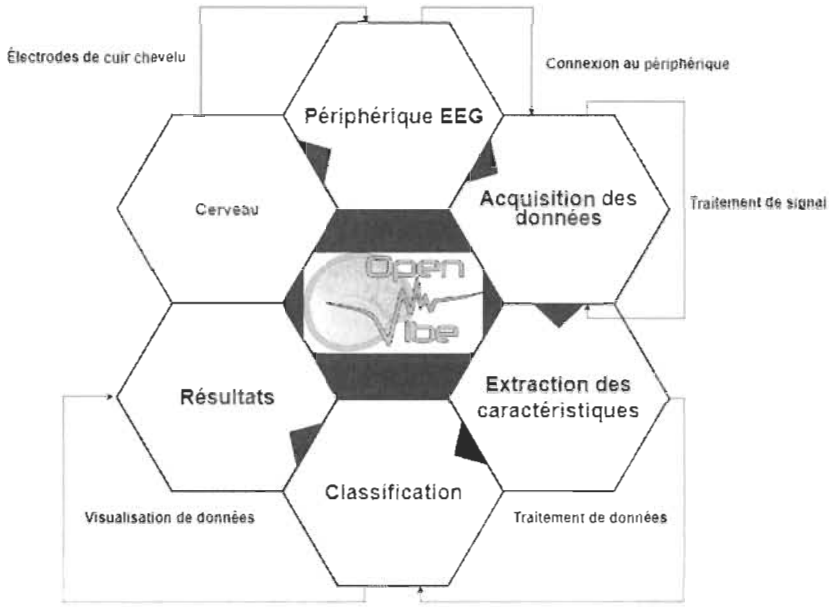

Figure 32 Workflow de OpenVibe

Cette plateforme nous permet d'acquérir, filtrer, traiter, classer et visualiser les signaux du cerveau en temps réel. Son but principal est d'obtenir des données du périphérique d'acquisition via le serveur d'acquisition, puis de les envoyer à un ou plusieurs clients. Ce client dans notre cas et généralement est le concepteur OpenVibe. Le serveur d'acquisition et le concepteur peuvent se trouver sur la même machine ou sur des machines distantes sur le même réseau.

OpenVibe fonctionne sur un système de scénarios qui peuvent être créés à partir de différentes boîtes à outils spécifiques liées entre elles pour répondre à nos besoins.

Dans notre recherche, comme sur la figure 33, on peut trouver des composantes élémentaires du scénario que nous avons utilisé afin de créer un prototype.

D'autre part, OpenVibe nous a donné la possibilité d'acquérir des données EEG à travers des modules en offrant plus d'options pour la capture des signaux EEG comparées à d'autres logiciels.

3.5.1 Acquisition de données EEG

Comme nous avons expliqué précédemment, l'acquisition de signaux EEG se base sur une architecture client-serveur. Le serveur d'acquisition de OpenVibe est un outil conçu pour communiquer avec divers périphériques d'acquisition de signaux matériels, comme dans notre choix matériel de recherche, le casque Emotiv Epoc +. Il joue le rôle d'un

transmetteur de signaux et d'autres informations de test acquis aux applications OpenVibe dans un format générique standardisé. Pour pouvoir communiquer avec les appareils d'acquisition, le serveur d'OpenVibe recours à des modules appelés pilote. Il nous fournit un ensemble de pilotes parmi lesquels choisir, chacun est dédié à un ou plusieurs périphériques.

Notre première étape d'acquisition de signaux consiste à connecter le casque au logiciel OpenVibe Acquisition Server comme nous pouvons constater sur la figure 33.

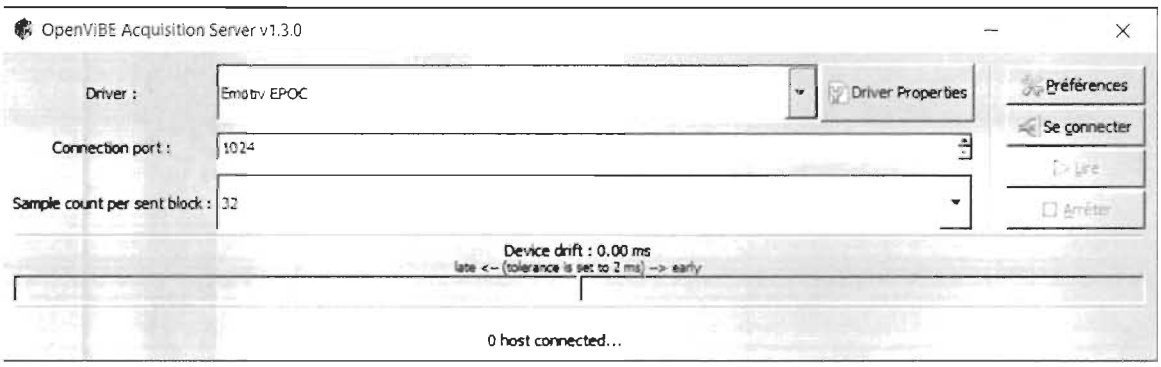

Figure 33 OpenVibe acquisition Server

Une liste de pilotes disponibles est accessible sous l'option pilote « driver ». Selon la distribution d'OpenVibe, une variété de pilotes sont disponibles.

Afin d'utiliser un pilote, nous devons définir des détails de connexion de sorte que le serveur communique avec les applications OpenVibe connectés comme nous souhaitons. Donc dans notre cas, nous avons adopté le pilote destiné à notre casque Emotiv Epoc comme nous pouvons voir sur la configuration dans la figure 33.

Ensuite, le port auquel sont envoyées les données peut être modifié dans la section « connection port » (nous avons gardé le paramètre standard par défaut).

Enfin, nous avons défini le nombre d'échantillons qui doit être envoyé par canal acquis dans un seul tampon. Nous avons conservé la valeur par défaut, soit 32.

Après avoir configurer le serveur d'acquisition, il est temps de faire la configuration de notre client d'acquisition sur le logiciel OpenVibe Designer en suivant le scénario que nous avons choisi tel que décrit sur la figure 34.

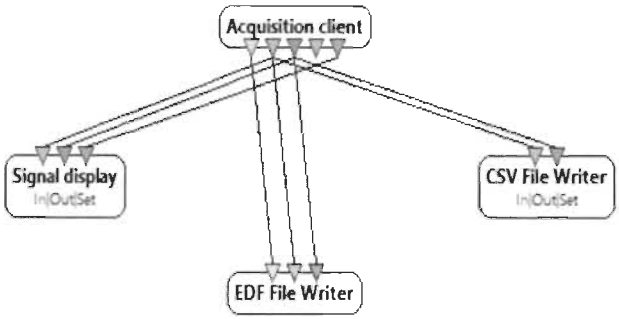

Figure 34 Scénario d'acquisition de signaux EEG

Notre expérience a été réalisé sur 4 sujets âgés entre 25 et 30 ans et s'étend sur deux phases.

La première phase : Nous avons demandé au sujet en état normal de prononcer un ensemble de mots que nous avons défini sur une période allant entre 1.5 secondes et 3 secondes en s'assurant de capturer les signaux EEG avec un taux d'échantillonnage de 128 Hz.

La deuxième phase : Nous avons demandé au même sujet de prononcer des mots après avoir consommé du cannabis, et la prise des signaux est aussi faite sur une période allant entre 1.5 secondes et 3 secondes dépendamment de la longueur du mot prononcé avec aussi un taux d'échantillonnage de 128 Hz.

Les mots utilisés selon la complexité de prononciation et l'occurrence des voyelles sont énumérés comme suit:

Complexité faible de prononciation : Abouti, abattu, accent, avouer, absolu, airbag, amande, anneau, annuel, auteur, aviser, bague, billet, blâmer, blonde, caisse, camion, Canada, classe et client.

Complexité élevée de prononciation :

Anticonstitutionnellement, inéligibilité, phénylalanine et prestidigitateur.

3.5.2 Enregistrement de signaux EEG

Les fonctionnalités de EDF file writer et de CSV file writer sur le scénario exposé à la figure 34 nous permettent d'enregistrer les signaux EEG acquis dans des fichiers de types EDF et CSV (Voir figure 35).

	Time:128	Epoch	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4	
2	0	0	4138.077	4150.897	4032.179	4137.308	4156.154	4151.539	4148.205	4152.82	4202.82	4177.949	4150.128	4136.667	4160.897	4145	
3	0.007813	0	4136.026	4150	4031.41	4142.436	4160.513	4136.539	4146.41	4150.641	4195.256	4163.846	4149.744	4128.077	4148.59	4079.359	
4	0.015625	0	4140.513	4151.026	4032.308	4140.897	4167.051	4128.974	4134.359	4143.59	4193.718	4156.282	4152.308	4132.949	4151.539	4074.231	
5	0.023438	0	4142.82	4152.82	4032.308	4146.154	4166.539	4131.923	4135.641	4132.051	4186.41	4153.846	4153.59	4138.59	4153.974	4112.692	
6	0.03125	0	4139.872	4158.846	4030.513	4158.461	4168.59	4137.82	4142.18	4123.718	4167.82	4150.769	4152.692	4136.667	4152.18	4116.41	
7	0.039063	0	4138.846	4164.744	4030	4159.359	4172.692	4143.59	4142.564	4121.154	4154.872	4147.692	4150.513	4136.282	4153.974	4098.333	
8	0.046875	0	4139.487	4159.615	4023.59	4153.461	4167.18	4145.769	4136.282	4123.59	4166.154	4151.539	4146.282	4131.667	4150.769	4130.513	
9	0.054688	0	4138.205	4152.692	4020.385	4151.667	4162.949	4146.667	4140.256	4134.359	4190.641	4165.385	4149.231	4126.923	4148.974	4168.461	
10	0.0625	0	4137.18	4155.769	4020	4152.436	4164.103	4152.82	4147.82	4141.41	4205.256	4176.282	4153.718	4128.461	4154.615	4136.41	
11	0.070313	0	4139.103	4156.667	4017.051	4151.41	4165.513	4155.897	4143.974	4130.385	4203.59	4170	4154.359	4129.231	4156.539	4111.539	
12	0.078125	0	4136.667	4153.205	4014.103	4150.128	4165.256	4153.59	4143.59	4122.949	4194.359	4153.461	4153.333	4124.615	4150.128	4146.154	
13	0.085938	0	4122.18	4142.949	4011.026	4141.539	4160.128	4155.641	4157.18	4138.718	4191.282	4149.872	4152.308	4120.385	4142.308	4166.41	
14	0.09375	0	4107.436	4127.82	4001.667	4125.256	4152.308	4155.256	4151.154	4147.82	4190.128	4156.282	4150.385	4121.539	4140.128	4150.513	
15	0.101563	0	4109.615	4126.667	4009.615	4130.128	4154.103	4150.513	4134.231	4134.487	4182.82	4153.077	4149.231	4126.923	4141.539	4160.769	
16	0.109375	0	4120.256	4137.18	4046.795	4130.897	4159.872	4154.103	4142.18	4137.436	4191.026	4161.41	4150.897	4135.641	4147.18	4180.897	
17	0.117188	0	4127.436	4142.564	4073.077	4141.282	4160.641	4153.846	4150.769	4157.18	4206.923	4188.205	4154.872	4141.923	4156.923	4152.051	
18	0.125	0	4122.18	4139.872	4072.821	4141.026	4160.641	4145.513	4148.205	4154.615	4200	4188.59	4154.231	4138.59	4152.82	4135.897	
19	0.132813	0	4104.487	4136.923	4071.41	4136.41	4163.461	4150.128	4151.41	4139.872	4183.077	4160.385	4145	4123.077	4133.846	4152.564	

Figure 35 Un fichier CSV contenant des signaux EEG bruts

Après avoir enregistré les signaux EEG acquis sous format CSV et PDF nous passons maintenant à l'étape de pré-traitement de ces signaux à l'aide du module EEGLAB qui nécessite Matlab pour être exécuté.

3.5.3 EEGLAB

Afin d'arriver à nos objectifs, nous avons utilisé l'application EEGLAB qui a été conçue pour être exécutée sur Matlab. C'est une boîte à outils Matlab interactive pour le traitement des données EEG intégrant l'analyse indépendante des composants (ICA), l'analyse temps/fréquence, le rejet d'artéfacts, les statistiques liées aux événements et plusieurs modes utiles de visualisation des données.

Cette boîte à outils nous a fourni une interface utilisateur graphique interactive comme sur la figure 36 nous permettant de traiter de manière flexible et interactive nos données EEG capturées à l'aide de l'approche ICA, que nous allons expliquer par la suite, ainsi qu'une analyse temps/ fréquence.

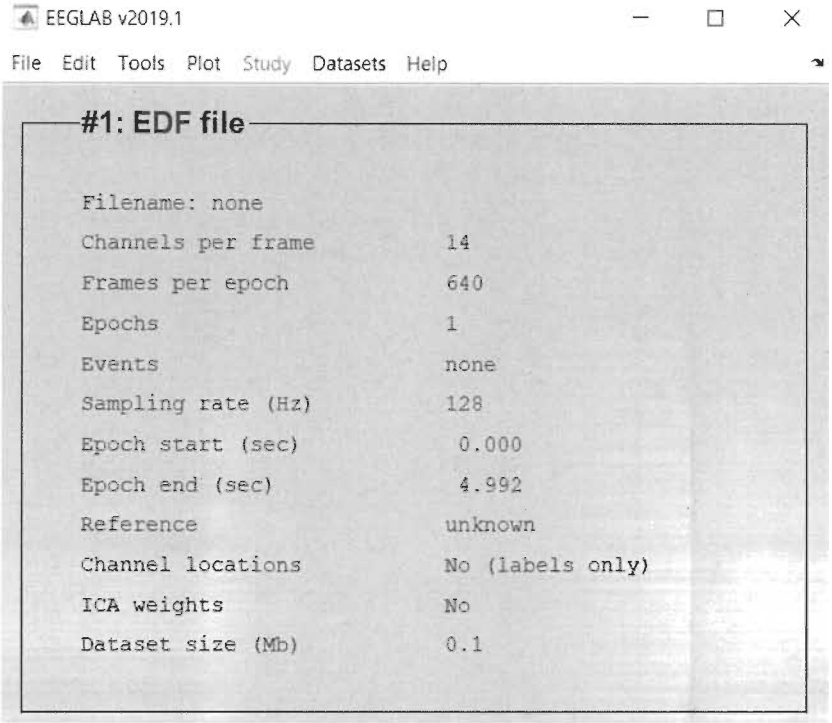

Figure 36 Interface utilisateur de EEGLAB

Comme nous pouvons voir sur la figure 36, nous avons choisi un signal EEG à l'état brut (voir figure 37) afin d'éliminer les artéfacts en appliquant l'ICA.

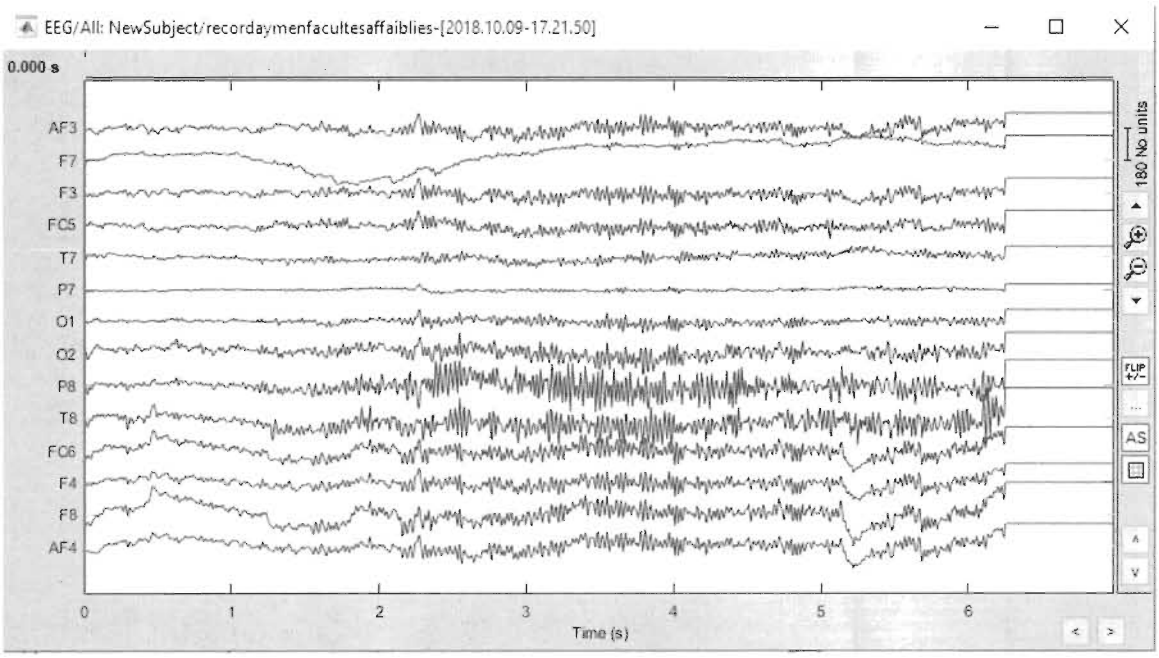

Figure 37 Signal EEG Brut d'un sujet en état de facultés affaiblies

Une fois que nous avons sélectionné un fichier EDF des enregistrements des signaux EEG à l'état brut, nous devons appliquer un filtre FIR entre 1 et 40 Hz afin de séparer les bruits électriques de nos signaux EEG correspondants à l'activité cérébrale que nous voulons étudier, sachant que ce genre de bruit a une fréquence plus élevée, généralement plus que 40 Hz.

3.5.4 Rejet d'artéfacts avec ICA

Par la suite, nous devons procéder au rejet des artéfacts inclus dans nos signaux EEG, qui sont dû aux clignements des yeux et à certaines contractions musculaires.

Pour effectuer cette opération, EEGLAB propose un algorithme d'analyse en composantes indépendantes nommé Runica dont l'interface de configuration est présentée à la figure 38, que nous avons appliqué sur le fichier EDF importé.

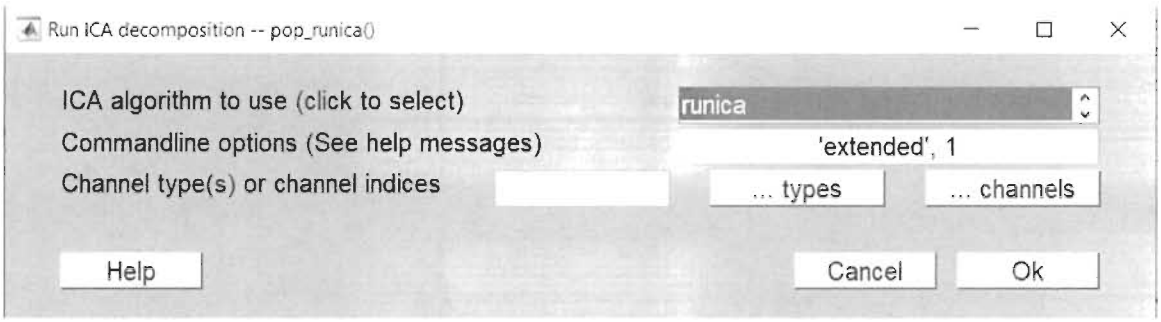

Figure 38 Algorithme Runica pour l'analyse en composantes indépendantes de EEGLAB

Tout d'abord, nous commençons par effectuer l'importation du fichier EDF contenant les enregistrements des signaux EEG dans EEGLAB, et ensuite le filtre FIR est appliqué sur les signaux EEG importés. L'algorithme implémenté est basé sur la maximisation de l'entropie (InfoMax) et permet de trouver la matrice de transformation W tel que $s^i(t) = W x^i(t)$, $i=1..n$, où n est le nombre de signaux mixés $x(t)$ et $s(t)$ les signaux séparés (composantes indépendantes). Cet algorithme est implémenté sous une forme de gradient naturel pour le calcul des composantes indépendantes. Il peut être considéré comme une méthode d'apprentissage neuronal basée sur une approche de descente de gradient maximal, telle que suggérée par la formulation mathématique suivante :

$$W(t+1) = W(t) + \eta(t)(I - f(s)(s)^T)W(t) \qquad (6)$$

où η(t) est une fonction exprimant le taux d'apprentissage en fonction du temps (itérations) et f(s) une fonction d'activation non linéaire choisie selon la nature de la distribution. I est la matrice identité de dimensions m*m. Il est important de noter que la valeur initiale des poids W est généralement une matrice aléatoire.

Donc une fois la décomposition ICA effectuée, nous pouvons sélectionner visuellement les composantes qui représentent des artéfacts comme sur la figure 39.

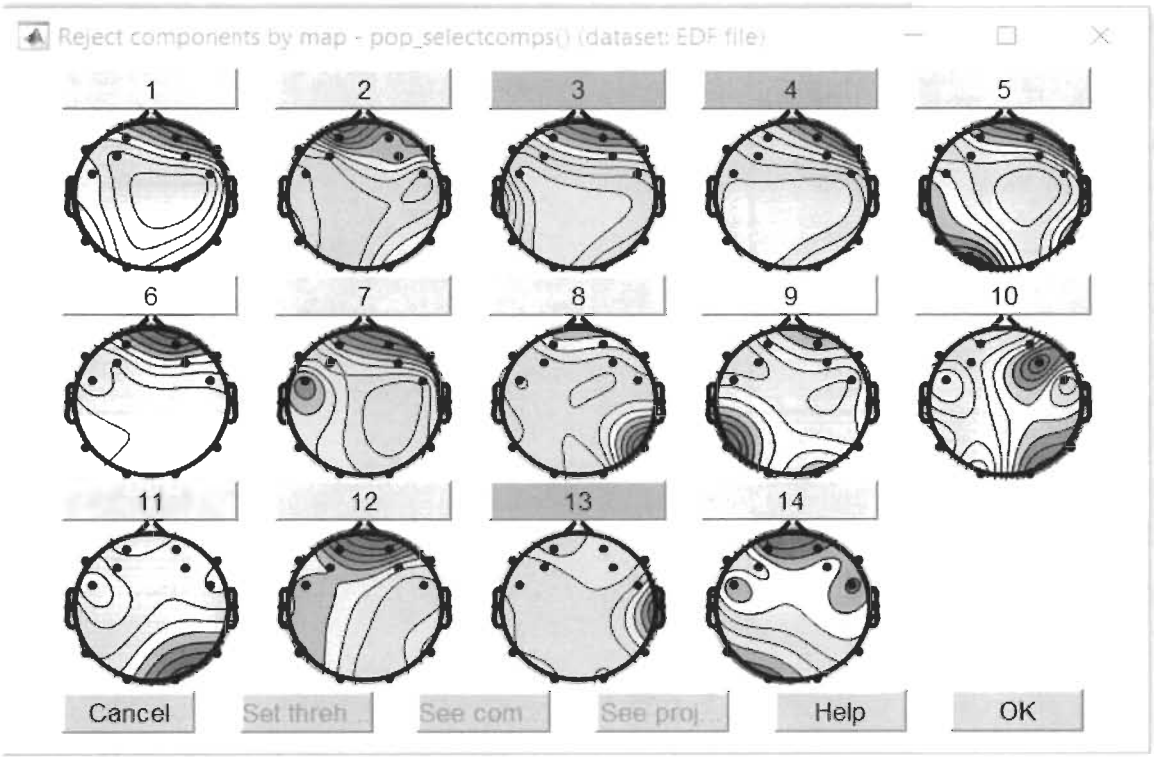

Figure 39 Rejet d'artéfacts en utilisant la décomposition en composantes indépendantes sur un signal EEG en facultés affaiblies

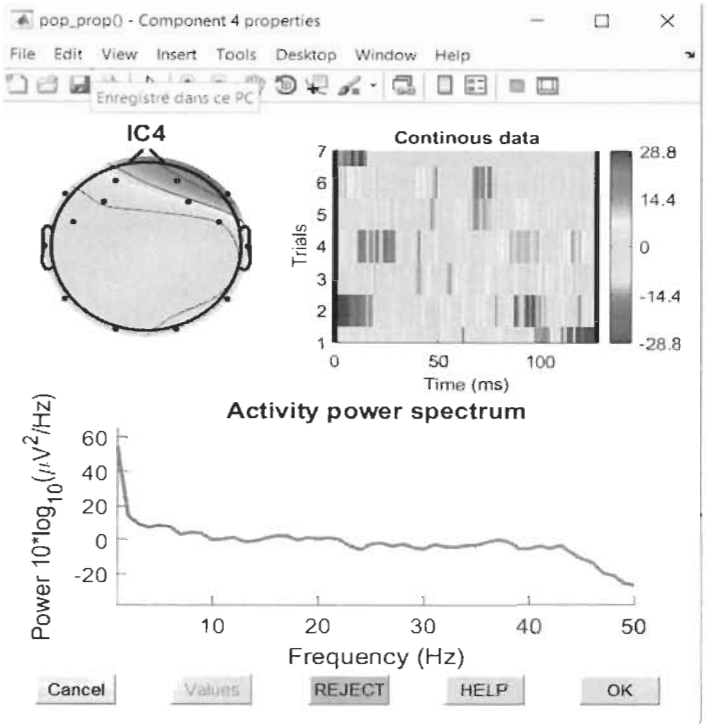

Figure 40 Composant lié à un artéfact dû au clignement d'œil

Nous avons identifié la composante IC4 sur la figure 40 comme étant un artéfact de clignement d'œil, nous l'avons donc sélectionné afin de la rejeter ainsi que les composantes 3 et 13. Une fois les composantes associées aux artéfacts indésirables rejetées, nous pouvons enregistrer le signal EEG, sans les artéfacts, et seulement lié à la cognition. L'étape suivante consiste à extraire un ensemble de caractéristiques des signaux EEG enregistrés en suivant la même approche de Berdaa S. et Meunier F. [63].

3.6 Extraction des caractéristiques en utilisant BrainStorm

Afin de sélectionner les caractéristiques valides du signal EEG liées à l'activité cérébrale qui seront ensuite utilisées pour déterminer l'état des facultés d'une personne, nous avons proposé l'utilisation de méthodes non paramétriques pour l'estimation de la densité spectrale de puissance. Les méthodes non paramétriques pour calculer la densité spectrale de puissance du signal, peuvent soit être basée sur l'estimation de la covariance ou du spectre du périodogramme sans supposer à priori que ce dernier possède une structure particulière. Dans cette recherche, la méthode de Welch est utilisée.

La méthode de Welch (également appelée méthode de périodogramme) pour estimer le spectre de puissance est réalisée en divisant d'abord le signal temporel en blocs (segments) successifs, en formant ensuite le périodogramme de chaque segment exprimé par le calcul des carrés des modules des coefficients de la transformée de Fourier discrète, puis le résultat final est la moyenne arithmétique des périodogrammes individuels [63].

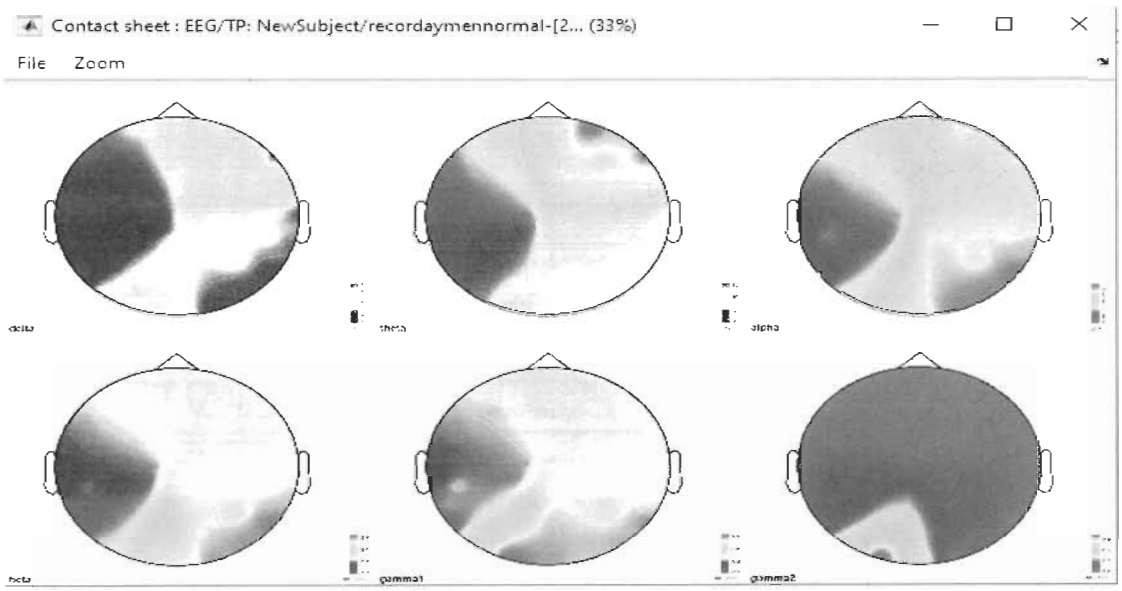

Figure 41 La densité de la puissance spectrale d'un sujet en état normal sur les différentes bandes de fréquences

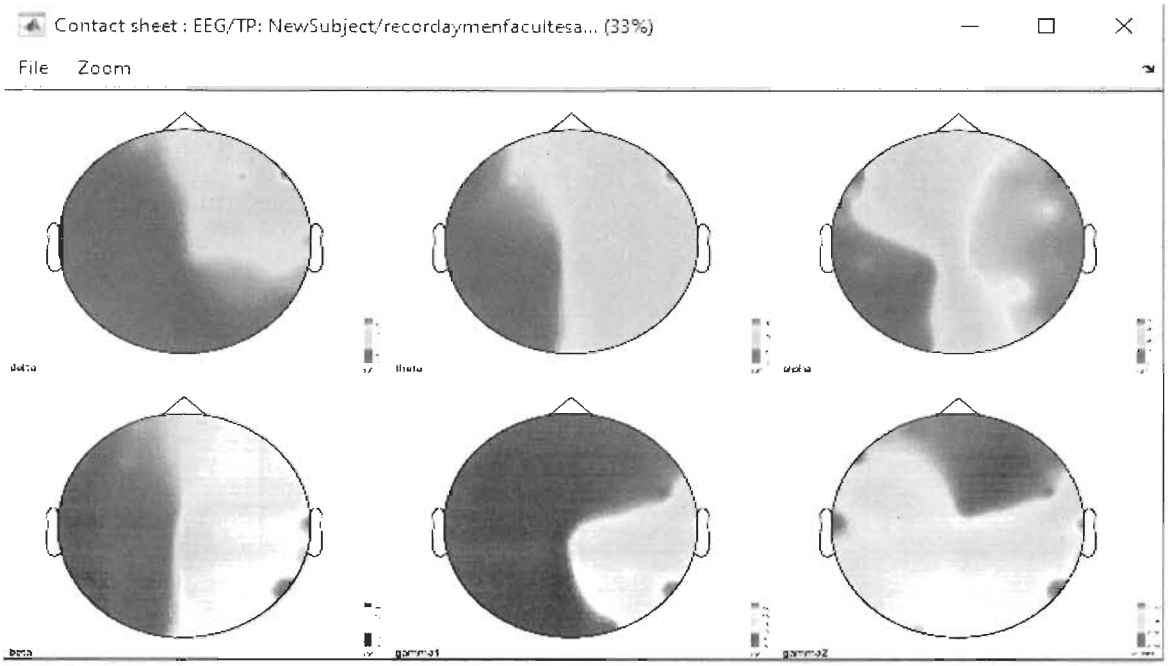

Figure 42 La densité de la puissance spectrale d'un sujet en état de facultés affaiblies sur les différentes bandes de fréquences

Cette approche nous a permis de déterminer et de voir les changements de puissance des signaux sur différentes parties du cerveau et sur les différentes bandes de fréquences en comparant l'état normal avec l'état facultés affaiblies comme sur les figures 41 et 42, ce qui nous a permis par la suite de procéder à l'identification et l'extraction des caractéristiques dans les bandes de fréquences significatives découlant de notre analyse.

3.6.1 La transformation en ondelettes

Dans cette approche nous faisons l'extraction des caractéristiques dans le domaine temps/fréquence ce qui nous a permis par la suite de déceler quels sont les changements les plus indicatifs en fonction du temps et pour quelles bandes de fréquences ces changements surviennent.

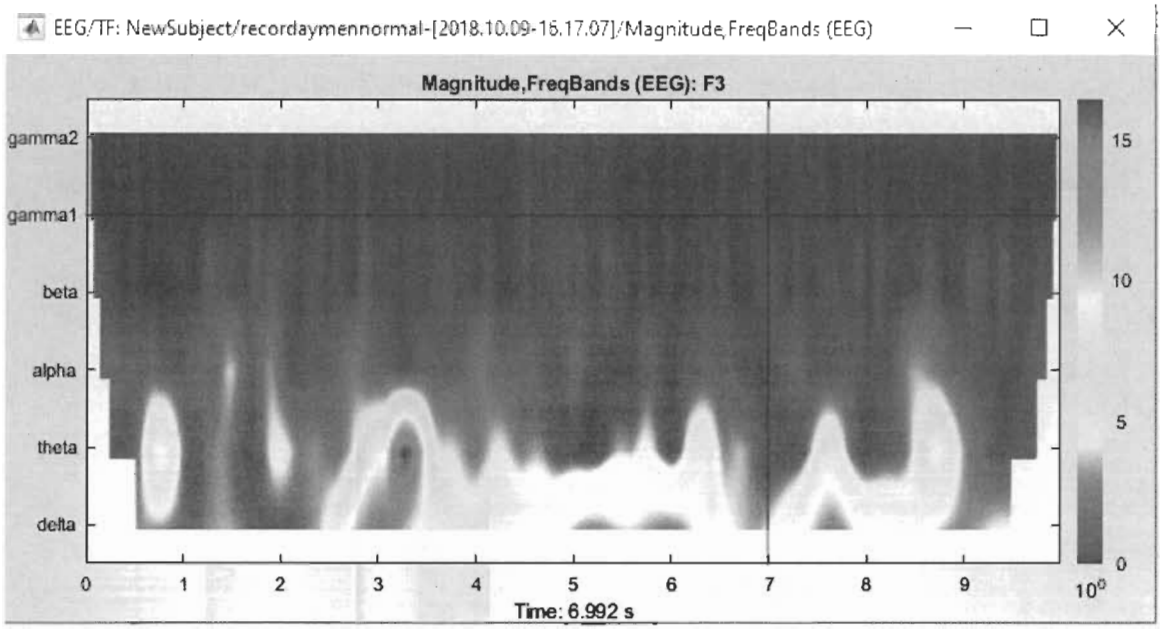

Figure 43 Transformée en ondelette du canal F3 d'un sujet en état normal

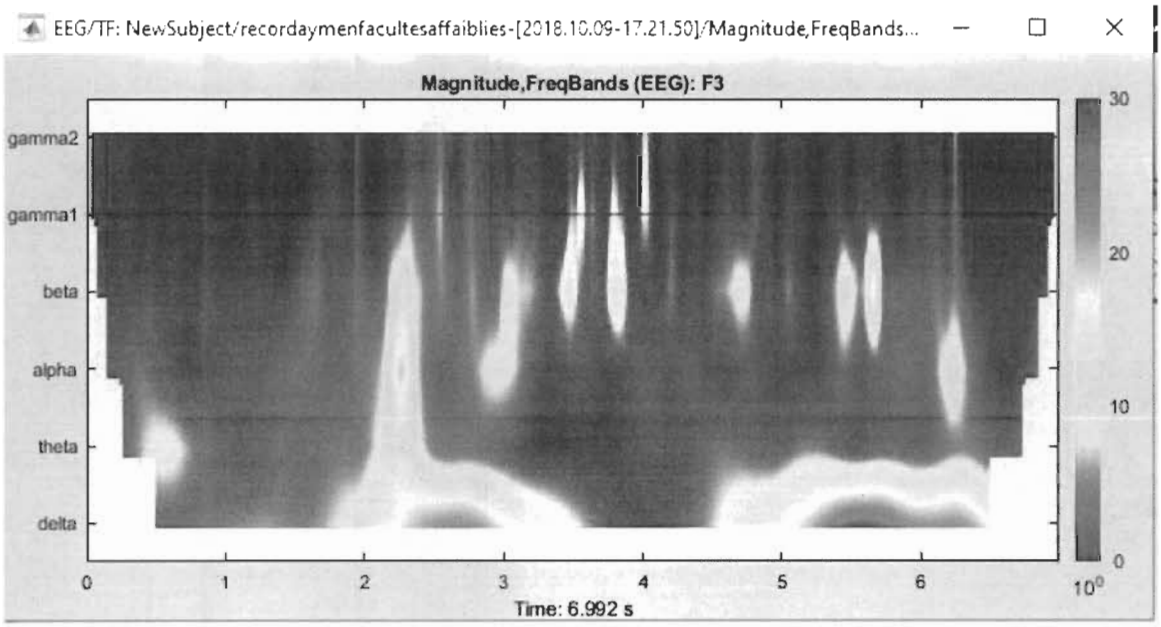

Figure 44 Transformée en ondelette du canal F3 d'un sujet en état facultés affaiblies

Ces transformées en ondelette, sont effectuées sur nos signaux EEG après avoir rejeté les artéfacts. Les coefficients d'ondelette sont ensuite extraits et sont utilisés comme entrées pour entraîner notre réseau de neurones qui sera décrit avec plus de détails dans la prochaine section.

Dans notre étude, nous nous sommes intéressés aux bandes de fréquences Alpha et Beta que nous avons choisi en ajoutant une autre caractéristique qui est la moyenne arithmétique des signaux de la transformée en ondelette comme nous pouvons voir sur l'équation 7 et la déviation standard de ces signaux comme décrit par l'équation 8 :

La moyenne arithmétique de la puissance des signaux peut être présentée comme dans l'équation 7:

$$A = \frac{1}{n} * \sum_{i=1}^{n} x_i \qquad (7)$$

Où A est la moyenne arithmétique, n le nombre de termes moyennés, et x_i la valeur de chaque article individuel dans la liste des nombres qui sont moyennés.

Cependant, l'équation de la déviation standard des signaux est définie comme décrit dans l'équation 8 :

$$SD = \sqrt{\frac{\sum |x-\mu|^2}{N}} \qquad (8)$$

Où x est une valeur dans l'ensemble de données, μ est la moyenne de l'ensemble de données, et N est le nombre de points de données dans l'échantillon.

Les coefficients d'extraction de caractéristiques sont ensuite enregistrés dans un fichier CSV afin de procéder à la classification.

3.7 Classification des signaux EEG

3.7.1 Azure Machine Learning studio

Avant de commencer la classification de nos signaux EEG, nous avons ajouté une colonne intitulée « Event Id » dans les fichiers CSV de caractéristiques extraites de la transformée en ondelette. Ce nouvel attribut permet d'identifier la classe associée au type de facultés d'une personne. Nous avons attribué le chiffre 1 aux caractéristiques des sujets en état de facultés affaiblies et le chiffre 0 pour les caractéristiques des sujets en état normal comme dans l'étude de Berdaa S. et Meunier F. [63].

À partir de ces jeux de données, nous avons entraîné notre modèle en utilisant deux algorithmes : un réseau de neurones artificiels et une machine à support de vecteurs (SVM) pour la facilité de leur mise en œuvre et leur grande vitesse. Ces algorithmes

d'apprentissage et de classification ont été déployés sur Azure Machine Learning, une plateforme collaborative que nous pouvons utiliser pour créer, tester et déployer des solutions d'analyse prédictive de nos données que nous avons regroupées dans deux fichiers CSV comme présentés aux figures 45 et 46. Ces fichiers sont utilisés pour l'entraînement de notre modèle de classification des signaux EEG.

Ces deux fichiers sont ensuite utilisés comme ensemble de données d'entraînement pour la classification des signaux EEG.

Time:128	Epoch	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4	Event Id	Event Dat	Event Duration
0	0	4138.077	4150.897	4032.179	4137.308	4156.154	4151.539	4148.205	4152.82	4202.82	4177.949	4150.128	4136.667	4160.897	4145	0		
0.007813	0	4136.026	4150	4031.41	4142.436	4160.513	4136.539	4146.41	4150.641	4195.256	4163.846	4149.744	4128.077	4148.59	4079.359	0		
0.015625	0	4140.513	4151.026	4032.308	4140.897	4167.051	4128.974	4134.359	4143.59	4193.718	4156.282	4152.308	4132.949	4151.539	4074.231	0		
0.023438	0	4142.82	4152.82	4032.308	4146.154	4166.539	4131.923	4135.641	4132.051	4186.41	4153.846	4153.59	4138.59	4153.974	4112.692	0		
0.03125	0	4139.872	4158.846	4030.513	4158.461	4168.59	4137.82	4142.18	4123.718	4167.82	4150.769	4152.692	4136.667	4152.18	4116.41	0		
0.039063	0	4138.846	4164.744	4030	4159.359	4172.692	4143.59	4142.564	4121.154	4154.872	4147.692	4150.513	4136.282	4153.974	4098.333	0		
0.046875	0	4139.487	4159.615	4023.59	4153.461	4167.18	4145.769	4136.282	4123.59	4166.154	4151.539	4146.282	4131.667	4150.769	4130.513	0		
0.054688	0	4138.205	4152.692	4020.385	4151.667	4162.949	4146.667	4140.256	4134.359	4190.641	4165.385	4149.231	4126.923	4148.974	4168.461	0		
0.0625	0	4137.18	4155.769	4020	4152.436	4164.103	4152.82	4147.82	4141.41	4205.256	4176.282	4153.718	4128.461	4154.615	4136.41	0		
0.070313	0	4139.103	4156.667	4017.051	4151.41	4165.513	4155.897	4143.974	4130.385	4203.59	4170	4154.359	4129.231	4156.539	4111.539	0		
0.078125	0	4136.667	4153.205	4014.103	4150.128	4165.256	4153.59	4143.59	4122.949	4194.359	4153.461	4153.333	4124.615	4150.128	4146.154	0		
0.085938	0	4122.18	4142.949	4011.026	4141.539	4160.128	4155.641	4157.18	4138.718	4191.282	4149.872	4152.308	4120.385	4142.308	4166.41	0		
0.09375	0	4107.436	4127.82	4001.667	4125.256	4152.308	4155.256	4151.154	4147.82	4190.128	4156.282	4150.385	4121.539	4140.128	4150.513	0		
0.101563	0	4109.615	4126.667	4009.615	4120.128	4154.103	4150.513	4134.231	4134.487	4182.82	4153.077	4149.231	4126.923	4141.539	4160.769	0		
0.109375	0	4120.256	4137.18	4046.795	4130.897	4159.872	4154.103	4142.18	4137.436	4191.026	4161.41	4150.897	4135.641	4147.18	4180.897	0		
0.117188	0	4127.436	4142.564	4073.077	4141.282	4160.641	4153.846	4150.769	4157.18	4206.923	4188.205	4154.872	4141.923	4156.923	4152.051	0		
0.125	0	4122.18	4139.872	4072.821	4141.026	4160.641	4145.513	4146.205	4154.615	4200	4188.59	4154.231	4138.59	4152.82	4135.897	0		
0.132813	0	4104.487	4136.923	4071.41	4136.41	4163.461	4150.128	4151.41	4139.872	4183.077	4160.385	4145	4123.077	4133.846	4152.564	0		
0.140625	0	4094.487	4137.18	4081.539	4137.692	4161.923	4162.692	4159.487	4143.333	4182.18	4155.256	4144.231	4121.026	4128.59	4142.949	0		

Figure 45 Données d'entraînement de sujets en état normal

Time:128	Epoch	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4	Event Id	Event Dat	Event Duration
0	0	4123.461	4149.103	4134.231	4155	4104.744	4138.59	4193.333	4114.872	4107.82	4133.077	4145.769	4153.718	4134.103	4150.256	1		
0.007813	0	4126.154	4149.744	4137.82	4161.41	4105.667	4112.18	4187.692	4108.846	4104.103	4120.769	4145	4151.795	4130.513	4146.795	1		
0.015625	0	4120.769	4151.154	4134.744	4157.308	4107.564	4122.051	4193.59	4110.256	4100.128	4134.744	4142.82	4144.231	4134.487	4151.026	1		
0.023438	0	4114.744	4148.59	4136.41	4152.051	4106.282	4146.539	4199.231	4122.436	4108.077	4144.615	4143.974	4143.333	4143.974	4150.128	1		
0.03125	0	4109.615	4144.359	4130.897	4159.615	4103.077	4138.846	4191.667	4123.077	4114.872	4131.923	4144.231	4141.795	4135.128	4138.205	1		
0.039063	0	4101.282	4147.564	4131.667	4163.718	4101.795	4124.744	4191.923	4118.333	4110.128	4128.077	4140	4135	4122.82	4132.308	1		
0.046875	0	4099.231	4148.205	4127.82	4155.385	4094.615	4141.282	4205.513	4132.18	4115.641	4141.282	4139.487	4132.692	4133.59	4134.103	1		
0.054688	0	4104.744	4142.82	4127.564	4145.769	4083.461	4152.949	4207.436	4145.513	4131.923	4146.795	4142.308	4136.41	4145.256	4130.513	1		
0.0625	0	4102.82	4139.744	4128.718	4142.949	4081.026	4124.744	4188.718	4136.41	4131.667	4137.949	4141.923	4133.333	4136.667	4124.487	1		
0.070313	0	4105.256	4143.718	4137.564	4154.744	4091.41	4111.667	4181.539	4127.18	4122.564	4131.539	4143.461	4136.539	4135.641	4132.051	1		
0.078125	0	4124.359	4149.359	4143.846	4172.949	4097.692	4149.103	4192.051	4137.564	4132.051	4146.026	4145.641	4153.59	4156.923	4154.231	1		
0.085938	0	4128.974	4150.513	4142.949	4165.641	4092.692	4160.641	4187.949	4144.231	4140.769	4158.333	4136.41	4155.256	4156.539	4156.026	1		
0.09375	0	4115.513	4146.461	4134.744	4149.744	4092.436	4129.872	4171.795	4133.718	4126.539	4151.923	4125.385	4140.513	4135.897	4135.385	1		
0.101563	0	4119.872	4148.846	4134.359	4159.872	4100.256	4128.974	4172.564	4134.359	4118.59	4150.769	4130.769	4141.923	4141.667	4139.615	1		
0.109375	0	4134.872	4149.103	4140.128	4168.333	4104.359	4141.795	4180.513	4142.436	4132.692	4152.82	4137.051	4154.615	4159.615	4158.461	1		
0.117188	0	4140.513	4150.513	4138.974	4167.949	4098.59	4130.641	4175.897	4141.154	4140.513	4154.231	4135	4157.308	4158.974	4155.897	1		
0.125	0	4136.589	4149.744	4137.949	4169.359	4095.128	4123.461	4171.282	4147.436	4137.051	4161.539	4135	4155.128	4156.539	4146.539	1		
0.132813	0	4126.282	4144.744	4136.795	4162.82	4101.154	4128.59	4167.82	4151.667	4137.051	4159.359	4135.385	4147.82	4159.359	4143.59	1		

Figure 46 Données d'entraînement de sujets en état de facultés affaiblies

Nous allons voir au chapitre 4 la réalisation de ce processus sur Azure Machine Learning Studio.

3.8 Enregistrement de signaux audio

Lors de la prononciation des mots par les personnes sujettes de nos expérimentations, nous avons capturé les signaux vocaux avec Audacity et enregistrer ces signaux dans des fichiers de type wav dont on se servira ultérieurement comme nous allons l'expliquer par la suite.

3.8.1 Pré-Traitement des signaux vocaux

Durant cette étape nous avons effectué une élimination de bruit qui est une étape très importante dans le traitement d'un signal audio.

Puisque les enregistrements audios ont été réalisés dans une chambre, nous avons jugé que les clicks de la souris ou la nature de la chambre, considérée comme un endroit non confiné (avec écho), peuvent présenter un bruit sur notre signal audio.

Une fois que nous avons importé le fichier audio, nous devons sélectionner les parties qui comportent du bruit ou un silence comme nous pouvons constater sur la figure 45.

À ce niveau, nous devons définir l'échelle de réduction de bruit (en dB), ajuster la sensibilité et définir une bande de lissage de fréquences pour commencer à réduire le bruit. En outre, nous effectuons un pré-traitement permettant de filtrer le signal tout en maintenant l'intégrité du signal associé à la parole.

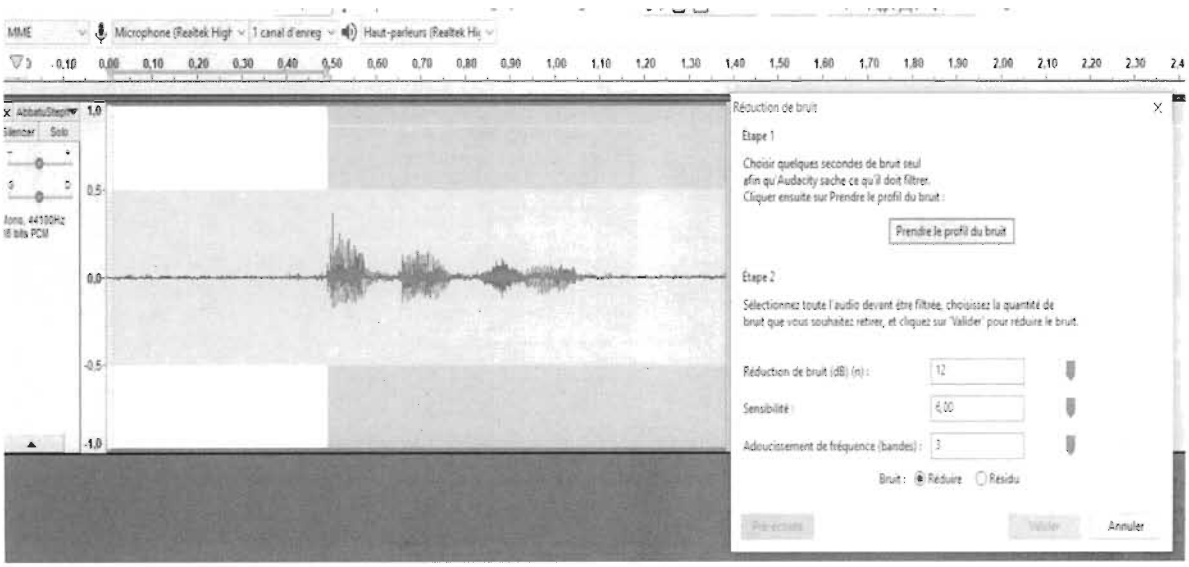

Figure 47 Élimination de bruit d'un signal vocal audio

Logiciel Praat

Praat est un logiciel qui nous a aidé à analyser, synthétiser et manipuler les enregistrements de la parole à l'état pure. Ce logiciel nous a aussi permis de créer un signal source à partir de la concaténation des enregistrements vocaux originaux et d'extraire les segments vocaux qui nous intéressaient.

Praat est considéré comme outil d'analyse spectrale et d'analyse de « Formant » dont nous avons besoin pour détecter les voyelles dans les mots enregistrés. Ces analyses seront aussi expliquées au chapitre 4.

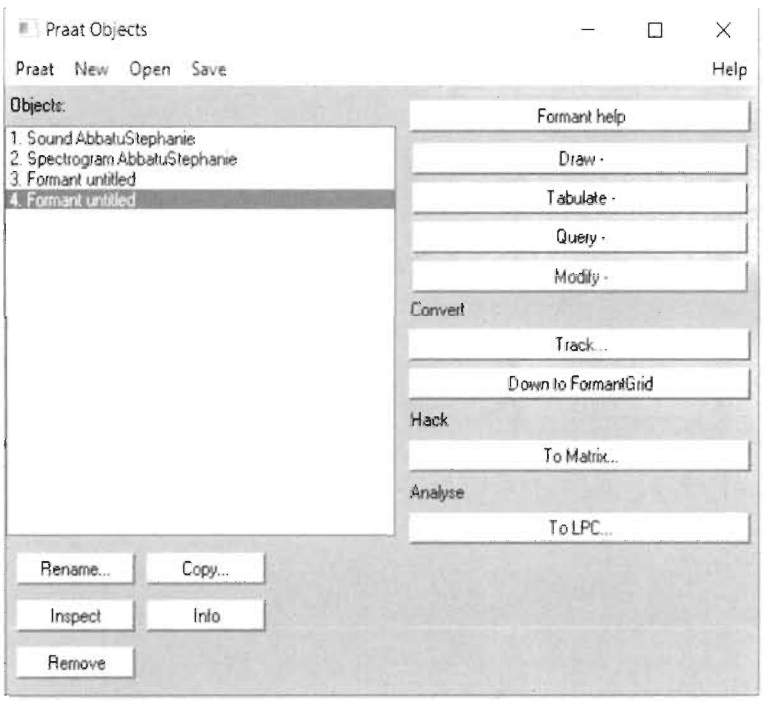

Figure 48 Interface utilisateur de Praat

3.8.2 Matlab

Matlab est un environnement informatique numérique à paradigmes multiples et considéré aussi comme un langage de programmation. Il a été développé par MathWorks permettant les manipulations matricielles, le traçage des fonctions et des données, la mise en œuvre d'algorithmes et la création d'interfaces utilisateurs.

C'est un langage de hautes performances pour l'informatique technique puisqu'il permet de faire le calcul, la visualisation et la programmation dans un environnement facile à utiliser où les problèmes et les solutions sont exprimés en notation mathématique familière. Les utilisations typiques incluent les mathématiques et le calcul, le développement d'algorithmes, la modélisation, la simulation et le prototypage, l'analyser et la visualisation des données, les graphiques scientifiques et techniques ainsi que le développement d'applications y compris la création d'une interface utilisateur graphique [86].

Dans les figures 49 et 50, et à l'aide de Matlab, nous avons présenté graphiquement les signaux vocaux d'un mot prononcé par deux sujets, l'un étant une femme et l'autre un homme, en état normal et en état de facultés affaiblies. La visualisation de ces signaux permet de constater la différence entre les enregistrements dans les deux états.

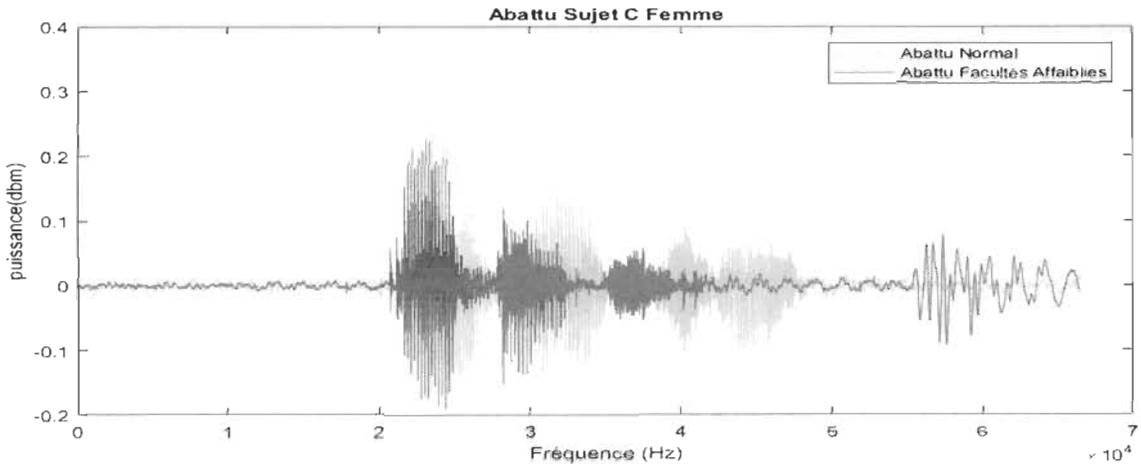

Figure 49 Représentation graphique du mot Abattu prononcé par un sujet Femme en état normal (signal Vert) et en état de facultés affaiblies (signal Rouge)

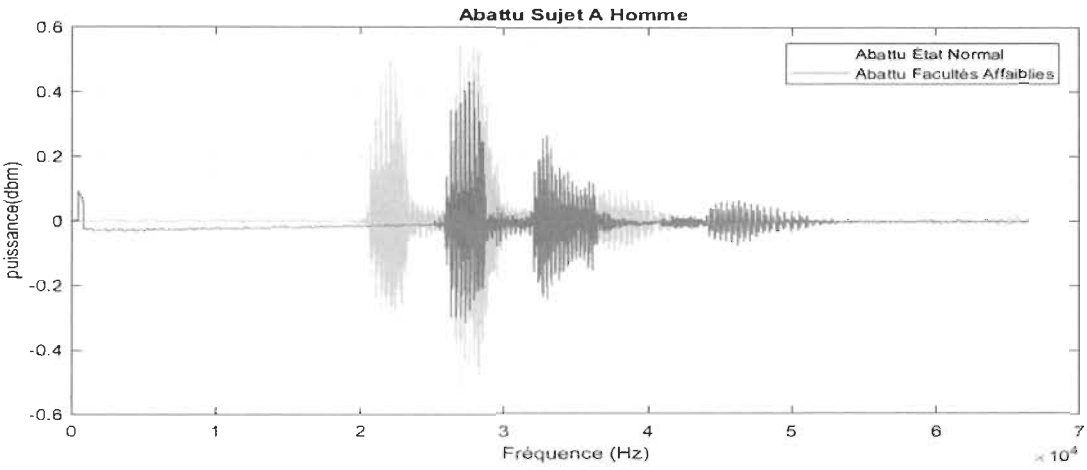

Figure 50 Représentation graphique du mot Abattu prononcé par un sujet Homme en état normal (signal Vert) et en état de facultés affaiblies (signal Rouge)

Afin de confirmer notre hypothèse, nous avons présenté l'enregistrement du mot prononcé dans les deux états sous forme de spectrogrammes pour pouvoir détecter visuellement quelques différences au niveau du contenu spectral comme nous pouvons le constater sur les figures 51 et 52.

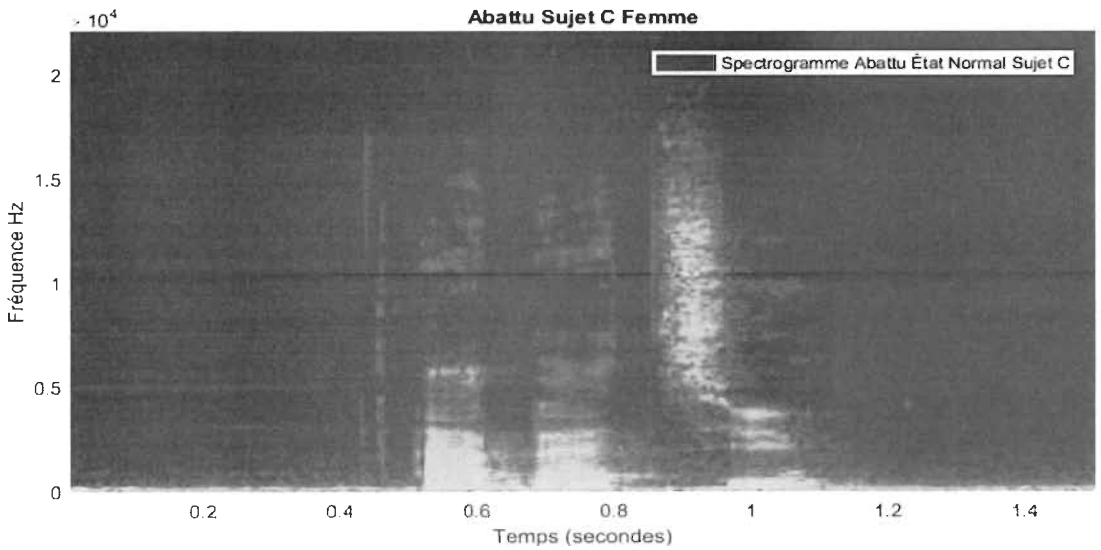

Figure 51 Spectrogramme du mot Abattu d'un sujet en état Normal réalisé avec Matlab

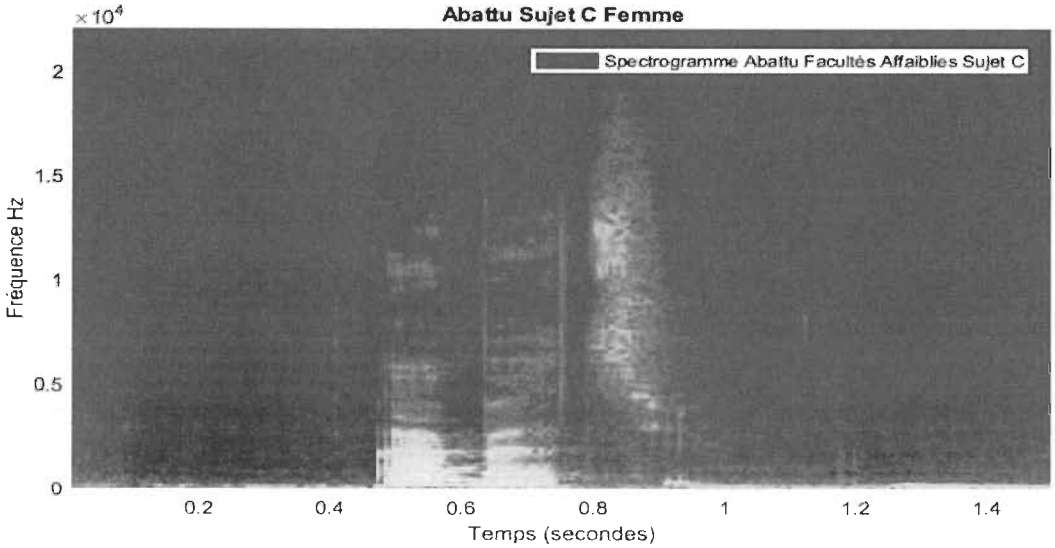

Figure 52 Spectrogramme du mot Abattu d'un sujet en état de facultés affaiblies réalisé avec Matlab

3.9 Classification des signaux vocaux

Il existe de nombreux projets et services différents pour la reconnaissance de la parole humaine. Nous avons été confrontés à cette tâche et avons décidé de construire un modèle qui sera en mesure de classer différents sons à l'aide d'algorithmes d'apprentissage automatique.

Dans un premier temps, nous avons choisi une approche nous permettant de faire une classification de sons représentés sous forme de spectrogrammes. Cette représentation des sons sous forme d'images permet l'entraînement d'un réseau de neurones convolutionnel dédié ensuite à la classification automatique de ces sons.

Les réseaux de neurones convolutionnels (CNN) sont conçus pour capturer certaines caractéristiques à partir des images de spectrogrammes en commençant par des caractéristiques de base au niveau initial (ex : contours) et des caractéristiques de très haut niveau aux niveaux plus profonds (ex : formes plus complexes).

Par niveau élevé, nous voulons dire les caractéristiques qui sont plus généralisées pour une classe particulière.

3.9.1 Réseau de neurones Convolutif Profond

Les CNN sont parmi les meilleurs algorithmes d'apprentissage pour comprendre le contenu d'une image et ont montré d'excellentes performances de segmentation, de classification, de détection d'images [87, 88].

Les réseaux de neurones convolutionels sont des réseaux neuronaux utilisés principalement pour classer les images (c'est-à-dire nommer ce qu'ils voient), les images de cluster par similitude (recherche de photos), et effectuer la reconnaissance d'objets dans les scènes. Par exemple, les réseaux neuronaux convolutionnels (ex : ConvNets) sont utilisés pour identifier les visages, les individus, les panneaux de rue, les tumeurs, les ornithorynques et de nombreux autres aspects des données visuelles.

La principale spécificité des CNN est sa capacité à exploiter la corrélation spatiale ou temporelle dans les données. La topologie des CNN est divisée en plusieurs étapes d'apprentissage composées d'une combinaison de couches convolutionnelles, des unités de traitement non linéaires et des couches de sous-échantillonnage [87]. Les CNN sont des réseaux hiérarchiques à couches multiples feedforward, où chaque couche, à l'aide d'une banque de filtres convolutionels, effectuent de multiples transformations [88]. L'opération de convolution aide à l'extraction des fonctionnalités utiles à partir de points de données localement corrélés. La sortie des filtres convolutionnels est ensuite dirigée vers l'unité de traitement non linéaire (fonction d'activation), qui aide non seulement à apprendre les abstractions, mais intègre également la non-linéarité dans l'espace des fonctionnalités. Cette non-linéarité génère différents modèles d'activations pour différentes réponses et facilite ainsi l'apprentissage des différences sémantiques dans les images. La sortie de la fonction d'activation non linéaire est généralement suivie d'un sous-échantillonnage, ce qui permet de réduire la dimension des résultats et rend également l'entrée invariante aux distorsions géométriques [89], [90]. Le CNN, avec sa capacité d'extraction automatique des caractéristiques, réduit le besoin d'un extracteur de caractéristiques distinct [91].

Les classificateurs d'images CNN prennent une image d'entrée, la traitent et la classe dans certaines catégories prédéterminées (ex : Car, Truck, Van, Bicycle). Les ordinateurs

voient une image d'entrée comme un tableau de pixels et cela dépend de la résolution de l'image. Sur la base de la résolution de l'image, la couche d'entrée du CNN recevra h X w X d (h : Hauteur, w : Largeur, d : Dimension) pixels.

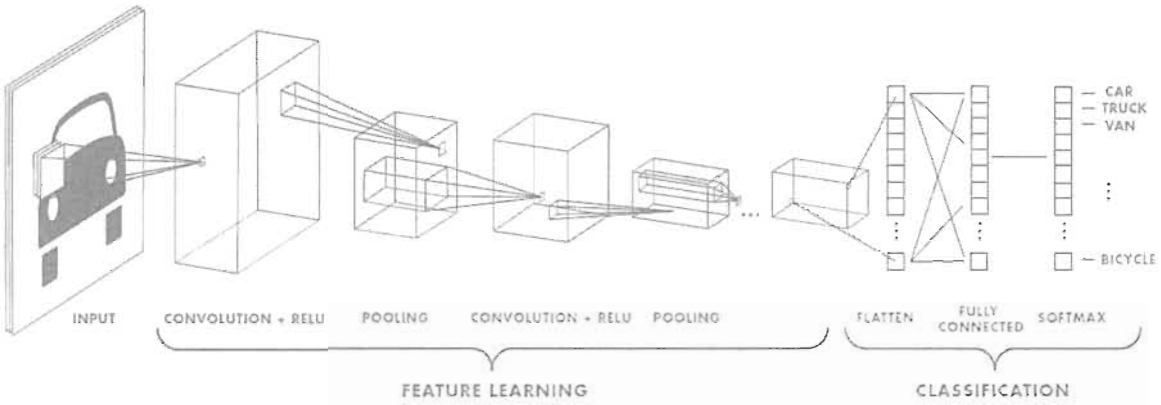

Figure 53 Architecture d'un réseau de neurones convolutionnel [27]

La figure 53 permet d'observer l'architecture d'un CNN typique qui est constituée de couches convolutionnelles permettant l'extraction de caractéristiques représentatives des images et de couches entièrement connectées qui quant à elles permettent la classification des images. Chaque couche convolutionnelle (premières couches du CNN) sert donc à extraire des entités d'une image d'entrée. L'opération de convolution favorise l'extraction des relations spatiales entre les pixels adjacents d'une image en appliquant des opérateurs (filtres) de formes carrées sur les données d'entrées de chaque couche convolutionnelle. Il s'agit d'une opération mathématique qui prend deux entrées telles que la matrice d'images et une banque de filtres.

3.9.2 Couche convolutionnelle (Convolutional Layer)

La couche convolutionnelle est composée d'un ensemble de filtres convolutionnels (chaque neurone agit donc comme un noyau). Toutefois, quand le noyau est symétrique, l'opération de convolution devient une opération de corrélation [92]. Le noyau convolutionnel fonctionne en divisant l'image en petites morceaux communément appelées champs réceptifs afin d'extraire les caractéristiques nécessaires pour le traitement ultérieur de classification. La division d'une image en petits blocs aide à extraire ces caractéristiques. Le noyau est ensuite appliqué aux images à l'aide d'un

ensemble spécifique de poids, en multipliant ces poids avec les éléments correspondants du champ réceptif à l'image [93]. Nous pouvons exprimer l'opération de convolution comme dans la formule mathématique qui suit :

$$f_l^k(p,q) = \sum_c \sum_{x,y} i_c(x,y) \,.\, e_l^k(u,v) \qquad (9)$$

Où $i_c(x,y)$ est un élément de l'entrée de l'image tensor I_c qui est un élément multiplié par $e_l^k(u,v)$ l'index du k^{th} ième du noyau convolutionnelle k^l de la couche l^{th}.

Considérons que la carte de caractéristiques de sortie de l'opération convolutionnelle peut être exprimée comme dans la formule $F_l^k = \left[f_l^k(1,1), \ldots, f_l^k(p,q), \ldots, f_l^k(P,Q)\right]$ (10)

Sachant que:

- K étant le numéro du noyau
- l étant le numéro de la ligne
- x étant la coordonnée x de l'image.
- y étant la coordonnée y de l'image.
- u étant la u^{th} ligne concernée.
- v étant la v^{th} colonne concernée
- $i_c(x,y)$ étant les éléments (x,y) du canal c^{th} de l'image.
- $f_l^k(p,q)$ étant les éléments de la matrice concernée
- (p,q) étant les éléments (p,q) de matrice concernée
- F_l^k étant l'entrée de la matrice de caractéristiques pour la l^{th} couche et la k^{th} neurone.

En raison de la capacité de partage des poids de l'opération convolutionnelle, différents ensembles de caractéristiques dans l'image peuvent être extraits par le noyau coulissant avec le même ensemble de poids sur l'image et rend ainsi le CNN plus efficace par rapport aux réseaux entièrement connectés. L'opération de convolution peut en outre être classée en fonction du type de filtre et de la taille des filtres, du type de remplissage et de la direction de la convolution [94]. Par souci de simplicité, nous pouvons laisser tomber l'index de canal.

3.9.3 Pooling Layer

Les motifs des caractéristiques d'images (ex : contours, arêtes), qui résultent d'une sortie d'opération de convolution peuvent être produits à différents endroits dans l'image. Une fois ces entités extraites, son emplacement exact devient moins important tant que sa position approximative par rapport aux autres est préservée. La mise en commun ou l'échantillonnage vers le bas est une opération locale intéressante. Il résume des informations similaires dans le voisinage du champ réceptif et produit la réponse dominante au sein de cette région [97].

$$Z_l^k = g_p(F_l^k) \qquad (11)$$

L'équation 11 nous montre l'opération de l'opération Pooling, dans laquelle Z_l^k représente la carte de caractéristiques de la couche l^{th} pour la k^{th} entrée de la carte de caractéristiques F_l^k.

L'utilisation de la fonction de Pooling nous permet d'extraire une combinaison de caractéristiques qui sont invariantes aux petites distorsions et aux décalages [95], [96].

La réduction de la taille de la carte de caractéristiques à l'ensemble de caractéristiques invariantes régule non seulement la complexité du réseau, mais contribue également à augmenter la généralisation en réduisant le chevauchement. Différents types de formulations de mise en commun sont possibles, telles que les opérateurs max, moyenne, ou L2. Le surapprentissage, le pooling des pyramides spatiales, etc. sont aussi utilisés dans les CNN [97].

3.9.4 Fonction d'activation (Activation Function)

La fonction d'activation sert de fonction de décision et aide à l'apprentissage de modèles complexes. La sélection d'une fonction d'activation appropriée peut accélérer le processus d'apprentissage. La fonction d'activation d'une carte de caractéristiques adjacentes est définie par l'équation 12.

$$T_l^k = g_a(F_l^k) \qquad (12)$$

Dans l'équation 12, F_l^k est la sortie de la convolution, qui est associée à la fonction d'activation $g_a(.)$ qui sert à rajouter la non-linéarité et renvoie une sortie transformée T_l^k de la couche l^{th}.

Dans la littérature, différentes fonctions d'activation telles que Sigmoid, Tanh, Maxout, SWISH, ReLU et des variantes ReLU telles que leaky ReLU, PreLU, TReLU, RReLU, ELU, et SeLU sont utilisées pour introduire une combinaison non linéaire de caractéristiques [98], [99], [52], [54].

3.9.5 Normalisation des lots

La normalisation des lots est utilisée pour résoudre les problèmes liés au changement de covariance interne dans les cartes de fonctionnalités. Le changement de covariance interne est un changement dans la distribution des valeurs des unités cachées, qui ralentit la convergence (en forçant le taux d'apprentissage à une faible valeur) et nécessite une initialisation minutieuse des paramètres. La normalisation par lots d'une carte est affichée dans l'équation 13)

$$N_l^k = \frac{F_l^k - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{13}$$

Dans l'équation 13, N_l^k représente une carte de caractéristiques normalisées, F_l^k est l'entrée de la carte de caractéristiques, μ_B et σ_B^2 représentent la moyenne et la variance d'une carte de caractéristiques pour un mini-lot respectivement.

Afin d'éviter la division par zéro, ε est ajouté pour la stabilité numérique. La normalisation par lots unifie la distribution des valeurs de la carte de caractéristiques en les fixant à zéro la moyenne et l'unité de variance [102].

En outre, la normalisation des lots lisse le flux de gradient et agit comme un facteur de régulation, ce qui contribue ainsi à améliorer la généralisation du réseau.

3.9.6 Dropout

Le dropout introduit la régularisation au sein du réseau, ce qui améliore en fin de compte la généralisation en sautant au hasard certaines unités ou connexions avec une certaine probabilité. Dans les réseaux de neurones, les connexions multiples qui apprennent une relation non linéaire sont parfois coadaptées, ce qui provoque un surapprentissage [103].

Cette chute aléatoire de certaines connexions ou unités produit plusieurs architectures de réseau éclaircies, et enfin un réseau représentatif est sélectionné avec de petits poids. Cette architecture sélectionnée est alors considérée comme une approximation de tous les réseaux proposés [104].

3.9.7 Couche entièrement connectée

La couche entièrement connectée (Fully Connected Layer) est principalement utilisée à la fin du réseau à des fins de classification.

Contrairement à la mise en commun et à la convolution, il s'agit d'une opération globale prenant les entrées des étapes d'extraction des entités et analyse globalement la production de toutes les couches précédentes [105].

Cela présente une combinaison non linéaire de caractéristiques sélectionnées, qui sont utilisées pour la classification des données [106].

3.9.8 Google Inception Model

Google Inception Model est un modèle de classification d'image open source qui est développé et maintenu par Google. Il est classé comme l'un des meilleurs modèles de classification d'image et entraîné sur 1000 classes d'objets.

Inception-V3 est une version améliorée de Inception-V1 et V2. L'idée d'Inception-V3 était de réduire le coût de calcul des réseaux profonds sans affecter la généralisation. À cette fin, Szegedy et al. ont remplacé les filtres de grande taille (5x5 et 7x7) par des petits filtres asymétriques (1x7 et 1x5) et ont utilisé la convolution 1x1 comme Bottelneck avant les grands filtres [107]. Cela transforme l'opération de convolution traditionnelle en une corrélation cross-canal. Dans l'une des publications précédentes, Lin et coll. ont exploité le potentiel des filtres 1x1 dans l'architecture NIN [108]. Szegedy et al. [111] ont utilisé le même concept de façon à construire un micro-réseau de neurones avec des structures plus complexes pour extraire les données dans le champ récepteur. Cela est réalisé en créant une instance de micro-réseau avec un perceptron multicouche afin de laisser les informations circuler de la couche d'entrée vers la couche de sortie uniquement. Dans Inception-V3, l'opération convolutionnelle 1x1 a été utilisée pour cartographier les données d'entrée dans 3 ou 4 espaces distincts qui sont plus petits que

l'espace d'entrée d'origine, puis cartographie toutes les corrélations dans des espaces 3D plus petits, via des convolutions standards (3x3 ou 5x5).

Ce modèle nous permet de recycler sa dernière couche et de la modifier avec notre propre ensemble de données audios sous forme de spectrogrammes. Le modèle Inception de Google est le seul modèle que nous avons trouvé donnant des prédictions précises en moins de temps et facile à utiliser tout en utilisant des réseaux neuronaux de convolution profonds (Deep CNN).

3.9.9 Classification audio avec Tensorflow

Créer un nouveau réseau de neurones convolutif est coûteux en termes d'expertise, de matériel et de quantité de données annotées nécessaires. Cette complexité de création peut être évitée en adoptant des réseaux pré-entraînés publiques. Ces approches sont appelées transfert learning parce que nous pouvons utiliser la connaissance acquise sur un problème de classification afin de l'appliquer à une nouvelle problématique particulière.

Afin de faire la classification de nos données audios en utilisant le réseau CNN Inception V3 comme modèle spécialisé pour la classification d'images, nous avons en premier lieu, converti nos données audios en spectrogrammes. De ce dataSet d'images sont extraits les vecteurs de caractéristiques utilisés pour entraîner notre classificateur.

Avec les spectrogrammes générés à partir des signaux vocaux capturés lors de nos expérimentations, subdivisées en segments de 0.5 seconde, nous pouvons utiliser Tensorflow pour entraîner un modèle de classification des facultés affaiblies. Ces processus seront détaillés dans le chapitre suivant du présent mémoire.

3.10 Conclusion

Dans ce chapitre, nous avons décrit les techniques utilisées dans notre analyse de signaux vocaux et signaux EEG pour la classification de ces signaux en deux groupes distincts : état des facultés normal ou état des facultés affaiblies.

Nous avons d'abord décrit le processus de capture de signaux vocaux. Le pré-traitement de ces signaux qui consiste à l'élimination du bruit dans nos enregistrements qui peut nuire à nos résultats de détections de formants liés aux voyelles et de classification.

Ensuite, l'approche de classification des signaux vocaux (dont nous allons voir l'automatisation dans le chapitre suivant de réalisation) en utilisant un Deep convolutional Neural Networks dont nous avons décrit l'architecture et les différentes couches et leurs rôles. Le processus d'extraction de caractéristiques automatisée par la couche convolutionnelle (contrairement aux signaux EEG dont l'extraction de caractéristiques n'est pas automatisée) ainsi que le modèle Google Inception V3 que nous avons modifié sa dernière couche pour créer notre propre modèle a aussi été exposé.

Nous avons notamment décrit le pré-traitement de signaux EEG en appliquant le pré-filtrage avant de passer au rejet des artéfacts en utilisant un algorithme ICA. L'extraction des coefficients de la transformée en ondelette (ondelette de Morlet) génère les caractéristiques utilisées pour entraîner notre modèle en ajoutant une colonne « Event Id » permettant de distinguer les données en facultés affaiblie ayant un attribut Event Id= 1 et les données en état normal ayant un Event Id = 0.

Dans le chapitre 4 nous allons décrire l'automatisation de la classification de signaux vocaux et EEG d'après notre méthodologie expérimentale ainsi que les résultats obtenus. Nous allons aussi discuter de la différence et de la fiabilité des deux méthodes de classification.

Chapitre 4 Résultats et discussions

4.1 Introduction

Après avoir introduit la méthodologie que nous avons adoptée durant notre recherche afin de pouvoir classifier les signaux vocaux et les signaux EEG pour détecter les facultés affaiblies d'une personne, dans ce chapitre nous allons exposer les résultats obtenus, présenter et valider nos observations ainsi que les avantages des signaux vocaux et les signaux EEG pour la classification de l'état des facultés d'une personne. Enfin nous allons en venir aux différents résultats obtenus lors de la classification en spécifiant le degré de précision de chaque classification des deux types de signaux (vocaux et EEG) et faire une comparaison entre la précision des deux approches de classifications.

4.2 Élimination des bruits du signal audio

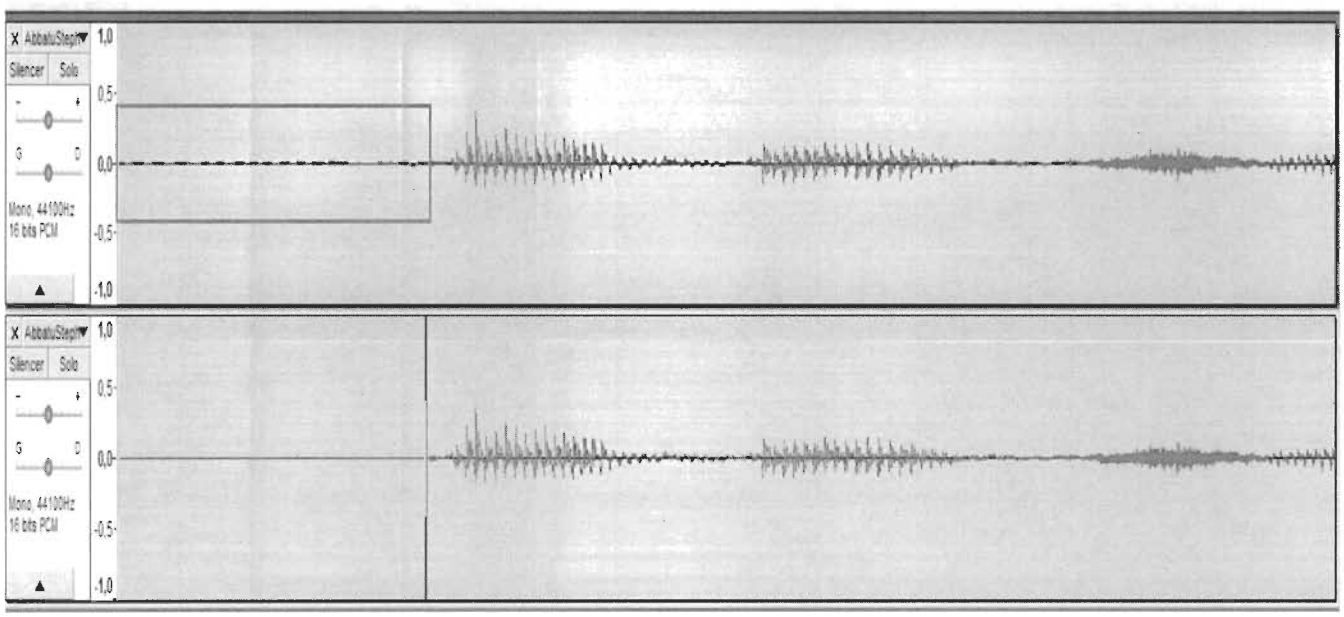

Figure 54 Enregistrement audio avant et après l'élimiation de bruits sur Audacity

Dans la figure 54, nous pouvons constater que le signal audio tracé dans la partie supérieure de cette figure représente le signal à l'état brut ayant une présence de bruit. La

partie en rouge représente le bruit détecté dans notre signal et nous pouvons remarquer que cette partie de signal est bosselée.

Après avoir procéder à une élimination de bruit comme nous pouvons voir dans la partie inférieure de la figure 54, le premier fragment de notre signal audio est désormais complètement plat ce qui signifie qu'on s'est débarrassé de ce type de bruit dans notre signal audio.

4.3 Détection des voyelles

Afin de distinguer les voyelles les unes des autres, nous nous intéressons davantage à une autre représentation qui elle est dans le domaine fréquentiel. Ce mode de représentation présente des courbes de réponse en fréquence (indiquant les fréquences de résonance préférées de l'appareil vocal) correspondant au spectre brut de l'onde sonore. En outre il est reconnu que l'articulation de l'appareil phonatoire génère des sonorités desquelles il est possible d'extraire les fréquences principales pour chaque son produit par les sujets de notre recherche.

Chacune des fréquences de résonance (fréquences principales) préférées de l'appareil phonatoire (chaque maximum dans la courbe de réponse de fréquence) est connue sous le nom de formant. Ils sont généralement annotés F1, F2, F3, etc, selon l'ordre croissant des fréquences.

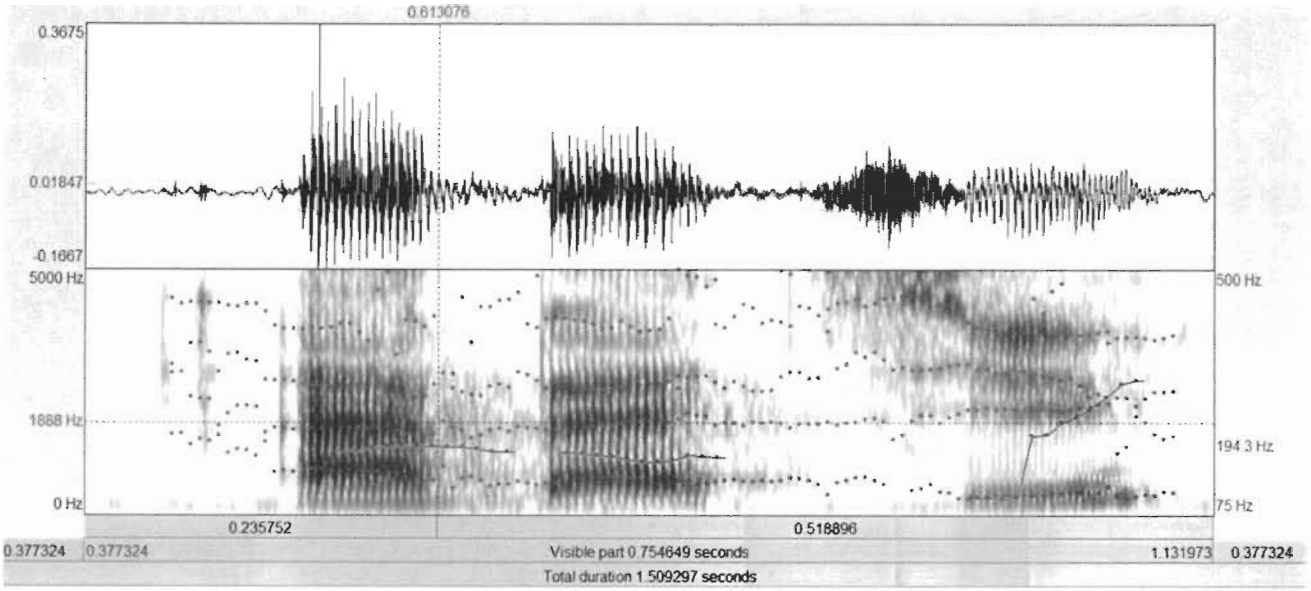

Figure 55 Formants de mot Abattu prononcé par le sujet A en état normal

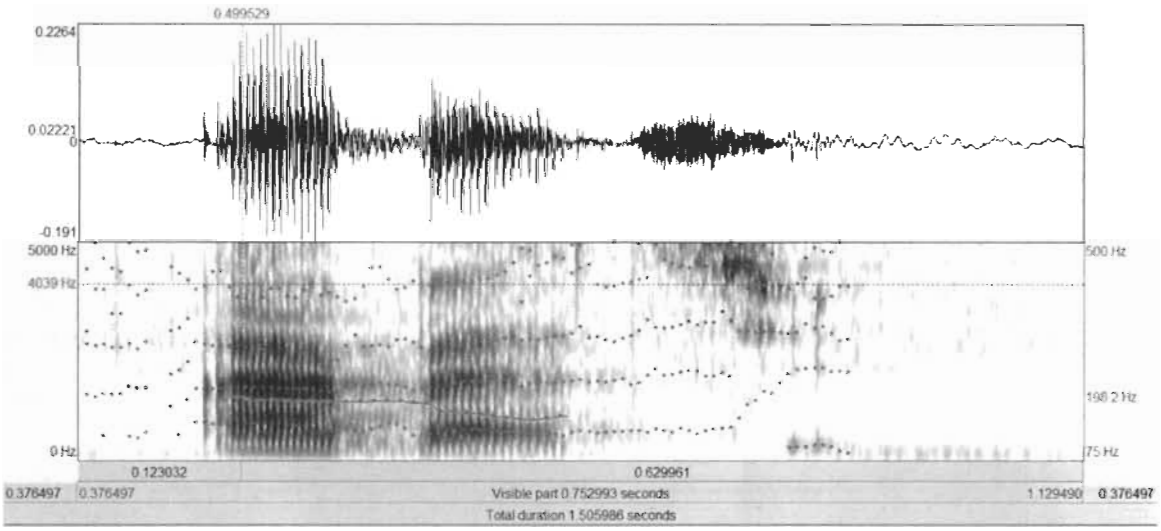

Figure 56 Formants de voyelles du mot Abattu prononcé par le sujet A en état de facultés affaiblies

Nous avons expérimenté la prononciation du mot « Abattu », mot qui comporte l'occurrence de voyelles. La détection des voyelles basée sur l'analyse des formants permettra ensuite de comparer la prononciation du mot faite par un sujet à l'état normal et un à l'état facultés affaiblies.

À ce niveau nous pouvons observer les formants extraits à partir du menu formant du logiciel Praat comme présenté dans les sections inférieures des figures 55 et 56.

Il est alors possible de visualiser la sonorité produite par la prononciation de la voyelle A comme présenté dans la partie supérieure de la figure 55. Nous pouvons constater que la voyelle A est prononcé deux fois, et correspond au premier et troisième segment sonore (voir partie supérieure, figure 55). Dans la partie inférieure de la figure 55, le spectre de fréquences des signaux vocaux générés par la prononciation du mot « Abattu » est affiché. Les lignes bleus représentent l'amplitude du son appelée hauteur tonale (Pitch), nous pouvons aussi observer les formants correspondant aux points rouges et identifiés explicitement par les étiquettes F1, F2, F3 et F4 pour la première voyelle A prononcée (voir partie inférieure, figure 55) schématisés dans la figure 57.

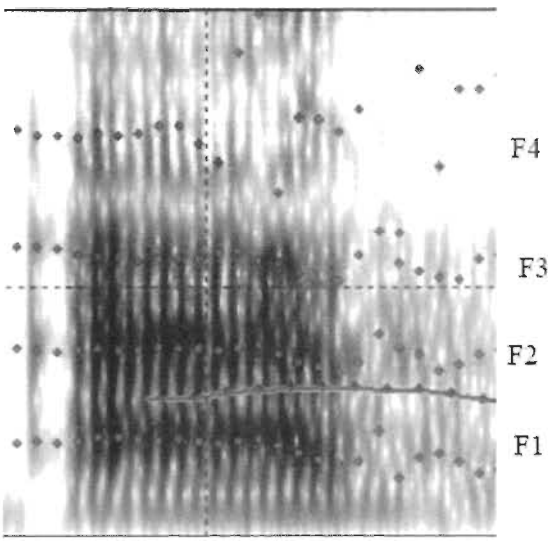

Figure 57 Formant de la voyelle A extrait du mot Abattu sur Praat

Dans l'outil Praat, en cliquant sur une voyelle il devient possible d'obtenir quelques informations sur les formants. Ces informations nous montrent que pour le temps 0.563976 seconde (voir ligne rouge verticale pointillée, figure 57), F1 correspond à une fréquence de 887.522516 Hz, F2 à 1756.293623 Hz, F3 à 2648.152433 Hz et F4 à 3628.969519 Hz comme observer sur la figure 58.

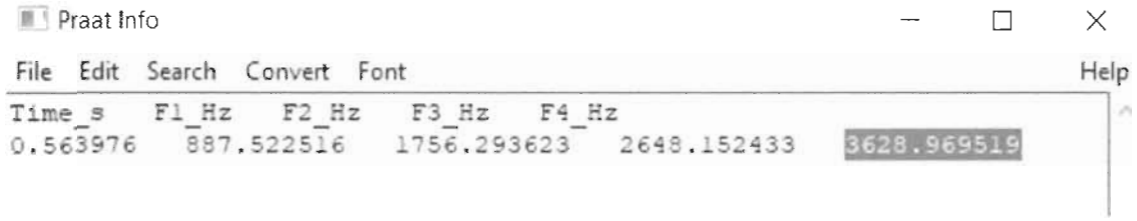

Figure 58 Caractéristiques de formants liées aux voyelles et leurs amplitudes

Si nous considérons que pour la parole, la prononciation des voyelles génère des sonorités dont les deux formants les plus importants sont F1 et F2. Il est alors possible sur la base de ces deux formants F1 et F2, de faire la distinction entre les prononciations des voyelles A prononcées dans un mot.

La deuxième occurrence de la voyelle A est apparue lors de la prononciation du mot « Abattu » durant la période du temps de 0.739921 seconde (Voir les figures 59 et 60).

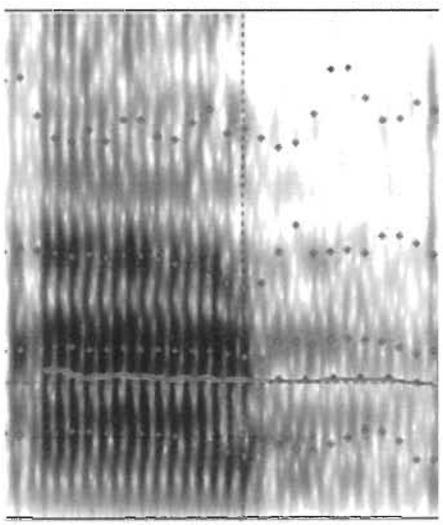

Figure 59 Formants de la deuxième voyelle A du mot Abattu

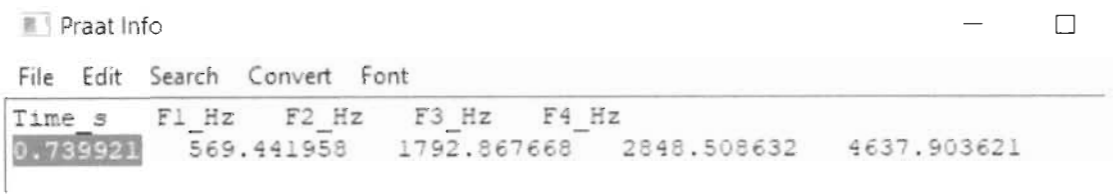

Figure 60 Caractéristiques de formants liées à la seconde voyelle A et leurs amplitudes

Nous avons ensuite effectué cette même analyse des formants sur les enregistrements vocaux des sujets en état normal et en état de facultés affaiblies afin de comparer les fréquences de formants générés des voyelles A, E, I, O et U. Dans le tableau 1, nous pouvons voir les formants de ces voyelles enregistrées en état normal, tandis que le tableau 2 contient les formants de ces voyelles prononcées en état de facultés affaiblies.

L'observation des deux spectrogrammes des figures 55 et 56, illustre bien l'évolution de la composition spectrale et le changement de position des formants avec le temps selon l'état des facultés d'une personne.

Nous pouvons donc déduire que la position des formants varie beaucoup entre l'enregistrement du même mot à l'état normal et celui à l'état faculté affaiblie.

Voyelle	F1	F2
A	890 Hz	1760 Hz
E	927 Hz	1890 Hz
I	426 Hz	1860 Hz
O	216 Hz	1535 Hz
U	395 Hz	2140 Hz

Tableau 1 Amplitude de formants de voyelles du sujet A en état normal.

Voyelle	F1	F2
A	596 Hz	1580 Hz
E	502 Hz	1716 Hz
I	854 Hz	2093 Hz
O	341 Hz	1338 Hz
U	1719 Hz	2930 Hz

Tableau 2 Amplitude de formants de voyelles du sujet A en état facultés affaiblies

En comparant les résultats présentés dans les tableaux 1 et 2, nous pouvons confirmer notre hypothèse de comparaison de la prononciation des voyelles. Nous pouvons observer une différence dans les fréquences des formants de voyelles entre un discours associé à un état normal et un autre associé à un état de facultés affaiblies.

Après avoir effectué une analyse de voyelles, nous passons à l'automatisation de cette analyse et à la classification de nos données audios avec Tensorflow.

4.4 Automatisation de la classification des signaux Audio

4.4.1 Préparation des données

Dans cette étape, nous avons réparti nos données audios en deux dossiers distincts, un pour chaque état des facultés, c'est-à-dire un dossier où nous avons stocké les enregistrements de la parole des sujets à l'état normal et un autre dossier contenant les enregistrements audios de la parole à l'état facultés affaiblies.

En deuxième lieu, nous avons regroupé les deux dossiers créés sous un dossier que nous avons nommé « data_audio » dans notre espace de travail comme nous pouvons voir dans la figure 61.

Figure 61 Répartition des données audios en 2 dossiers

Après avoir préparé nos dossiers de données, nous avons créé une classe en python pour pouvoir décomposer chaque enregistrement audio en morceaux de taille de 0.5 seconde et générer les spectrogrammes de ces derniers.

```
def makechunks(path):
    folders=glob.glob(path+'*')
    for folder in folders:
      waves = glob.glob(folder+'/'+ '*.wav')
      print 'w',waves
      if len(waves) == 0:
          return 10
      for i in waves:
          w = i
          myaudio = AudioSegment.from_file(i, 'wav')
          chunk_length_ms = 500
          chunks = make_chunks(myaudio, chunk_length_ms)
          print chunks
          for i, chunk in enumerate(chunks):
              chunk_name = w.split('.')[0] + "chunk{0}.wav".format(i)
              print chunk_name
              print "exporting", chunk_name
              chunk.export(folder+'/'+chunk_name, format="wav")
```

Figure 62 Fonction de segmentation de fichier audio en segments de 0,5 seconde et génération de spectrogrammes sur ces segments audios

Comme nous pouvons observer dans la figure 62, la fonction *makechunks()* est celle responsable d'accéder à nos dossiers d'enregistrements audios et de segmenter chaque fichier wav en morceaux de 500 ms (0.5 secondes) qui seront ensuite utilisés pour extraire les caractéristiques servant à entraîner notre modèle de classification.

```
les/data_audio/Affaiblie/ClientAymen.wav', '../tf_files/data_audio/Affaiblie/AbbatuAhmed.wav']
[<pydub.audio_segment.AudioSegment object at 0x7f000b533910>, <pydub.audio_segment.AudioSegment object at 0x7f000b533950>]
chunk0.wav
exporting chunk0.wav
chunk1.wav
exporting chunk1.wav
[<pydub.audio_segment.AudioSegment object at 0x7f000b533990>, <pydub.audio_segment.AudioSegment object at 0x7f000b5339d0>]
chunk0.wav
exporting chunk0.wav
chunk1.wav
exporting chunk1.wav
[<pydub.audio_segment.AudioSegment object at 0x7f000b533950>, <pydub.audio_segment.AudioSegment object at 0x7f000b533890>]
chunk0.wav
exporting chunk0.wav
chunk1.wav
exporting chunk1.wav
[<pydub.audio_segment.AudioSegment object at 0x7f000b5339d0>, <pydub.audio_segment.AudioSegment object at 0x7f000b5338d0>]
chunk0.wav
exporting chunk0.wav
chunk1.wav
exporting chunk1.wav
[<pydub.audio_segment.AudioSegment object at 0x7f000b533890>, <pydub.audio_segment.AudioSegment object at 0x7f000b533910>]
```

Figure 63 Segmentation de données audios en segments 0.5 seconde

La figure 63 représente une segmentation des fichiers audios en morceaux de 0.5 seconde. Durant ce traitement, une extraction des spectres de fréquences de chaque

morceau audio est réalisé et ensuite sauvegarder dans des fichiers textes pour un traitement ultérieur.

Une fois que les enregistrements audios bruts sont segmentés, nous faisons appelle à une fonction permettant de créer les spectrogrammes de ces derniers.

```
def graph_spectrogram(wav_file):
    rate, data = get_wav_info(wav_file)
    print type(data), len(data)
    nfft = 256  # Longueur des segments de fenêtrage
    fs = 256  # Fréquence d'échantillonnage
    pxx, freqs, bins, im = plt.specgram(data, nfft, fs)
    print "pxx : ", len(pxx)
    print "freqs : ", len(freqs)
    print "bins : ", len(bins)
    # plt.axis('on')
    # plt.show()
    plt.axis('off')
    print wav_file.split('.wav')[0]
    plt.savefig(wav_file.split('.wav')[0] + '.png',
                dpi=100,  # Points par pouce
                frameon='false',
                aspect='normal',
                bbox_inches='tight',
                pad_inches=0)  # Spectrogramme enregistré au format .png
    try:
      im = Image.open(wav_file.split('.wav')[0] + '.png')
      rgb_im = im.convert('RGB')
      rgb_im.save(wav_file.split('.png')[0] + '.jpg')
    except Exception as e:
      print e
    if os.path.exists(wav_file.split('.wav')[0] + '.png'):
        os.system('convert '+(wav_file.split('.wav')[0] + '.png') + ' '+(wav_file.split('.wav')[0] + '.jpg'))
        os.remove(wav_file.split('.wav')[0] + '.png')
```

Figure 64 Fonction de création de spectrogramme à partir de fichiers contenant des segments audio de 0,5 seconde

Dans la figure 64 nous observons la fonction permettant de générer les spectrogrammes de morceaux audios de 0,5 seconde des enregistrements audios en parcourant les deux racines de dossier de notre espace de travail et en générant les spectrogrammes à partir desquels il sera possible d'extraire les caractéristiques des signaux audio.

Cette fonction de génération de spectrogrammes retourne quatre objets :

- pxx : un périodogramme qui est l'estimation de la densité spectrale du signal audio en prenant le carré de la transformée de Fourier calculé dans une fenêtre comme suit :

$$\hat{S}_{per}(\omega) = \frac{1}{N}\left(\left|\sum_{n=1}^{N} x(n).e^{(-j\omega n)}\right|\right)^2.\sigma_B^2 \qquad (14)$$

Où N représente le nombre d'échantillons fixés et ω représente la fréquence.

- freeqs : le vecteur de fréquences.

- bins : intervalles de temps

- im : instance de matlplotlib.AxesImage représentant les données dans le graphe.

```
<type 'numpy.ndarray'> 66609
pxx :  129
freqs :  129
bins :  519
../tf_files/data_audio/Affaiblie/AuteurStephanie
Generating spectrograms..
<type 'numpy.ndarray'> 66560
pxx :  129
freqs :  129
bins :  519
../tf_files/data_audio/Affaiblie/AccentStephanie
Generating spectrograms..
<type 'numpy.ndarray'> 66560
pxx :  129
freqs :  129
bins :  519
../tf_files/data_audio/Affaiblie/AbboutiStephanie
Generating spectrograms..
<type 'numpy.ndarray'> 68608
pxx :  129
```

Figure 65 Génération de spectrogrammes

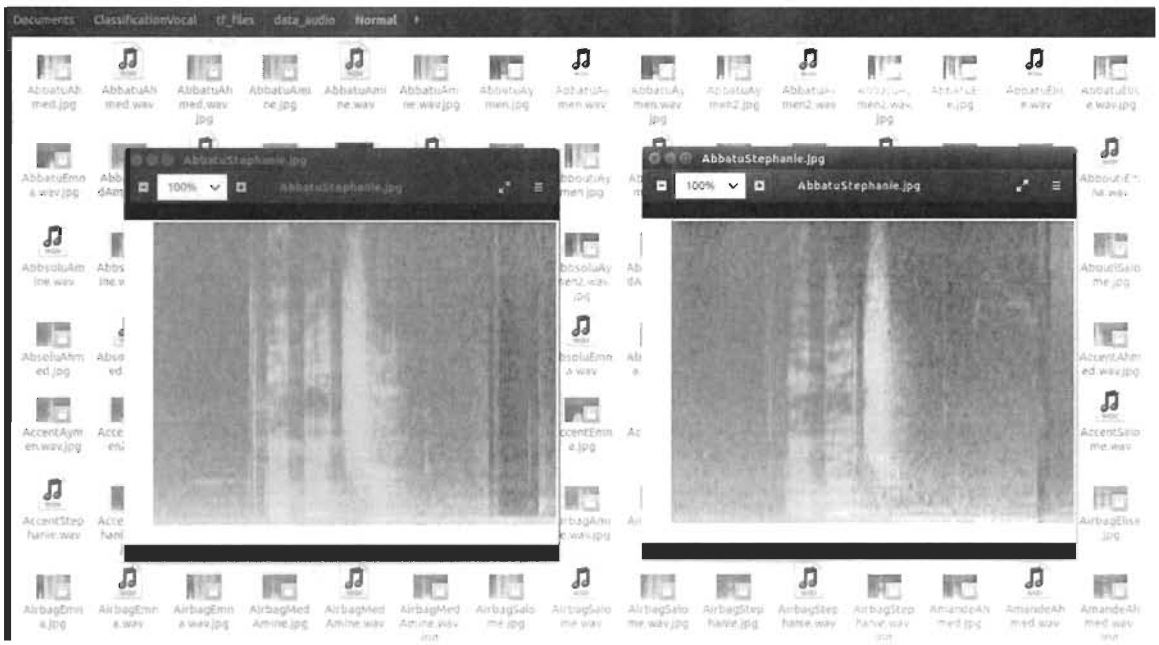

Figure 66 Génération de spectrogrammes et comparaison entre un spectrogramme à l'état normal (à droite) et un autre à l'état facultés affaiblies (à gauche)

4.4.2 Le réentraînement du bottleneck et l'affinement du modèle

La prochaine étape consistera à reconstruire le modèle d'Inception et ainsi réentraîner sa dernière couche avec les ensembles de données d'entraînement capturées dans nos

expérimentations et à tester son efficacité de classification des données du jeu données de tests. La couche de sortie du modèle de réseau implémentée dans notre recherche aura alors deux nœuds correspondant aux types de classifications possibles : soit état normal ou état facultés affaiblies.

Dans notre étude, nous utilisons un réseau neuronal convolutif profond (Deep CNN). Ce choix découle du fait que les CNN font généralement de bons classificateurs et exécutent particulièrement bien les tâches de classification d'images en raison de leurs couches d'extraction et de classification des caractéristiques représentatives contenues dans ces images. Ces CNN devraient donc être très efficaces pour trouver des motifs dans les spectrogrammes tout comme ils sont aussi efficaces pour trouver des structures dans des images.

Pour commencer, nous avons téléchargé le modèle pré-entraîné de Inception V3 et modifier le script de la classe de réentraînement qui est la composante principale de notre algorithme et de toutes tâches de classification d'images. Ce script nous permet d'entraîner une nouvelle couche supérieure (appelé bottleneck) qui reconnaîtra la classe spécifique de chaque spectrogramme. Cette couche supérieure reçoit comme entrées 2048 vecteurs dimensionnels pour chaque image. Une couche softmax est ensuite entraînée au-dessus de cette représentation.

Supposant que la couche softmax contient N étiquettes, cela correspond à l'apprentissage des paramètres du modèle N + 2048 * N correspondant aux biais et poids appris.

Pour cela, nous avons utilisé une fonction qui nous a permis de construire une liste d'images d'entraînement à partir des spectrogrammes dont nous pouvons voir le code python dans la figure 67.

```
def create_image_lists(image_dir, testing_percentage, validation_percentage):
  if not gfile.Exists(image_dir):
    tf.logging.error(" dossier d images '" + image_dir + "' non trouvé.")
    return None
  result = collections.OrderedDict()
  sub_dirs = [
    os.path.join(image_dir,item)
    for item in gfile.ListDirectory(image_dir)]
  sub_dirs = sorted(item for item in sub_dirs
                    if gfile.IsDirectory(item))
  for sub_dir in sub_dirs:
    extensions = ['jpg', 'jpeg', 'JPG', 'JPEG']
    file_list = []
    dir_name = os.path.basename(sub_dir)
    if dir_name == image_dir:
      continue
    tf.logging.info("Recherche images dans '" + dir_name + "'")
    for extension in extensions:
      file_glob = os.path.join(image_dir, dir_name, '*.' + extension)
      file_list.extend(gfile.Glob(file_glob))
    if not file_list:
      tf.logging.warning('aucun fichier trouvé')
      continue
    if len(file_list) < 20:
      tf.logging.warning(
          'Attention: le dossier contient moins de 20 images, ce qui peut engendrer des problemes.')
    elif len(file_list) > MAX_NUM_IMAGES_PER_CLASS:
      tf.logging.warning(
          'Attention: le dossier {}contient plus d {} images.certaines images'
          'ne seront pas selectionnées.'.format(dir_name, MAX_NUM_IMAGES_PER_CLASS))
    label_name = re.sub(r'[^a-z0-9]+', ' ', dir_name.lower())
    training_images = []
    testing_images = []
    validation_images = []
    for file_name in file_list:
      base_name = os.path.basename(file_name)
```

Figure 67 Fonction de création de liste d'images spectrogrammes pour l'entraînement du modèle

Cette fonction nous permet de faire l'analyse des sous-dossiers du répertoire d'images, les diviser en ensembles stables d'entraînement, de test et de validation, et renvoie une structure de données décrivant les listes d'images de spectrogrammes pour chaque étiquette et leurs chemins.

La fonction de création de liste de spectrogrammes retourne une liste contenant une entrée pour chaque sous-dossier d'étiquette, avec des spectrogrammes divisés en ensembles d'entraînement, de test et de validation pour chaque étiquette et qui possède comme arguments :

Image_dir : étant le chemin vers le dossier qui contient les sous dossiers contenant nos spectrogrammes.

Testing_percentage : étant le pourcentage de spectrogrammes à réserver pour les tests.
Validation_percentage : étant le pourcentage de spectrogrammes réservés pour la validation.
À ce niveau, nous décidons dans quel ensemble mettre une image, le créateur de l'ensemble de données à un moyen de regrouper les photos qui sont des variations proches les unes des autres.

Nous devons décider si ce fichier doit entrer dans les ensembles d'entraînement, de test ou de validation, et nous voulons conserver les fichiers existants dans le même ensemble même si d'autres fichiers sont ajoutés par la suite.

Pour y arriver, nous exécutons une méthode qui renvoie un chemin d'accès à une image pour une étiquette à l'index donné ayant comme argument :

- Une liste d'images d'entraînement pour chaque étiquette, une Étiquette pour laquelle nous voulons obtenir une image.
- Un décalage de l'index de l'image concernée cela sera modulé par le nombre d'images disponibles pour l'étiquette, il peut donc être arbitrairement grand.
- Le chemin d'accès du dossier racine des sous dossiers contenant les images d'apprentissage.
- Le nom de l'ensemble à partir duquel extraire les images d'entraînement, de test et de validation.

Cependant, nous avons pensé aux distorsions que notre algorithme peut rencontrer après avoir été déployé à la suite d'une mauvaise qualité de spectrogramme (spectrogramme flou).

Pour ce faire, nous avons utilisé une fonction comme celle décrite dans la figure 68.

```
def should_distort_images(flip_left_right, random_crop, random_scale,
                          random_brightness):
  jpeg_data = tf.placeholder(tf.string, name='DistortJPGInput')
  decoded_image = tf.image.decode_jpeg(jpeg_data, channels=input_depth)
  decoded_image_as_float = tf.cast(decoded_image, dtype=tf.float32)
  decoded_image_4d = tf.expand_dims(decoded_image_as_float, 0)
  margin_scale = 1.0 + (random_crop / 100.0)
  resize_scale = 1.0 + (random_scale / 100.0)
  margin_scale_value = tf.constant(margin_scale)
  resize_scale_value = tf.random_uniform(tensor_shape.scalar(),
                                         minval=1.0,
                                         maxval=resize_scale)
  scale_value = tf.multiply(margin_scale_value, resize_scale_value)
  precrop_width = tf.multiply(scale_value, input_width)
  precrop_height = tf.multiply(scale_value, input_height)
  precrop_shape = tf.stack([precrop_height, precrop_width])
  precrop_shape_as_int = tf.cast(precrop_shape, dtype=tf.int32)
  precropped_image = tf.image.resize_bilinear(decoded_image_4d,
                                              precrop_shape_as_int)
  precropped_image_3d = tf.squeeze(precropped_image, squeeze_dims=[0])
  cropped_image = tf.random_crop(precropped_image_3d,
                                 [input_height, input_width, input_depth])
```

Figure 68 Fonction de distorsion de données

Pour remédier aux distorsions et éviter de perdre l'exactitude de la prédiction dans telles circonstances, nous avons opté pour deux approches :

- Le recadrage de distorsions qui se fait en plaçant un cadre de sélection à une position aléatoire dans le spectrogramme complet. Le paramètre de recadrage contrôle la taille de ce cadre par rapport au spectrogramme d'entrée. Si la valeur est de 50%, la zone de recadrage correspondra à la moitié de la largeur et de la hauteur de l'entrée.
- La mesure de distorsions qui ressemble beaucoup au recadrage, sauf que le cadre de sélection est toujours centré et que sa taille varie de façon aléatoire dans la plage de donnée. Si le pourcentage de la mesure est nul, le cadre de sélection aura la même taille que l'entrée et aucune mesure n'est appliquée. Sinon si le pourcentage est de 50%, le cadre englobant sera dans une plage aléatoire entre la moitié de la largeur et de la hauteur de l'entrée.

Étant donné que les paramètres de la fonction de la figure 66 sont les suivants :

- flip_left_right: booléen s'il faut inverser aléatoirement les images horizontalement.
- random_crop: pourcentage entier définissant la marge totale utilisée autour de la zone de recadrage.
- random_scale: pourcentage entier de la variation de la mesure
  random_brightness: plage entière pour multiplier au hasard les valeurs des pixels
  input_height: taille verticale de l'image d'entrée attendue à modéliser.
- input_depth: combien de canaux l'image d'entrée attendue devrait avoir.
- input_mean: valeur de pixel qui doit être nulle dans l'image du graphique.
- input_std: de combien diviser les valeurs des pixels.

Une fois que nous avons remédié au problème de distorsion, il est temps d'entraîner notre système dont la progression est affichée dans la figure 69. Nous avons commencé l'entraînement par un faible nombre d'époches soit 500 et si le modèle converge à la sortie, nous augmenterons le nombre d'époches.

```
I0220 23:42:22.419667 139926896920384 retrain.py:1071] 2020-02-20 23:42:22.419640: Step 460: Cross entropy = 0.285909
INFO:tensorflow:2020-02-20 23:42:22.719806: Step 460: Validation accuracy = 93.0% (N=100)
I0220 23:42:22.720051 139926896920384 retrain.py:1087] 2020-02-20 23:42:22.719806: Step 460: Validation accuracy = 93.0% (N=100)
INFO:tensorflow:2020-02-20 23:42:24.580894: Step 470: Train accuracy = 89.0%
I0220 23:42:24.580970 139926896920384 retrain.py:1069] 2020-02-20 23:42:24.580894: Step 470: Train accuracy = 89.0%
INFO:tensorflow:2020-02-20 23:42:24.581263: Step 470: Cross entropy = 0.314284
I0220 23:42:24.581285 139926896920384 retrain.py:1071] 2020-02-20 23:42:24.581263: Step 470: Cross entropy = 0.314284
INFO:tensorflow:2020-02-20 23:42:24.864003: Step 470: Validation accuracy = 91.0% (N=100)
I0220 23:42:24.864262 139926896920384 retrain.py:1087] 2020-02-20 23:42:24.864003: Step 470: Validation accuracy = 91.0% (N=100)
INFO:tensorflow:2020-02-20 23:42:26.647327: Step 480: Train accuracy = 88.0%
I0220 23:42:26.647403 139926896920384 retrain.py:1069] 2020-02-20 23:42:26.647327: Step 480: Train accuracy = 88.0%
INFO:tensorflow:2020-02-20 23:42:26.647785: Step 480: Cross entropy = 0.352311
I0220 23:42:26.647809 139926896920384 retrain.py:1071] 2020-02-20 23:42:26.647785: Step 480: Cross entropy = 0.352311
INFO:tensorflow:2020-02-20 23:42:26.934335: Step 480: Validation accuracy = 91.0% (N=100)
I0220 23:42:26.934672 139926896920384 retrain.py:1087] 2020-02-20 23:42:26.934335: Step 480: Validation accuracy = 91.0% (N=100)
INFO:tensorflow:2020-02-20 23:42:28.626181: Step 490: Train accuracy = 89.0%
I0220 23:42:28.626262 139926896920384 retrain.py:1069] 2020-02-20 23:42:28.626181: Step 490: Train accuracy = 89.0%
INFO:tensorflow:2020-02-20 23:42:28.626607: Step 490: Cross entropy = 0.312587
I0220 23:42:28.626631 139926896920384 retrain.py:1071] 2020-02-20 23:42:28.626607: Step 490: Cross entropy = 0.312587
INFO:tensorflow:2020-02-20 23:42:28.919122: Step 490: Validation accuracy = 96.0% (N=100)
I0220 23:42:28.919369 139926896920384 retrain.py:1087] 2020-02-20 23:42:28.919122: Step 490: Validation accuracy = 96.0% (N=100)
INFO:tensorflow:2020-02-20 23:42:30.628775: Step 499: Train accuracy = 90.0%
I0220 23:42:30.628853 139926896920384 retrain.py:1069] 2020-02-20 23:42:30.628775: Step 499: Train accuracy = 90.0%
INFO:tensorflow:2020-02-20 23:42:30.629147: Step 499: Cross entropy = 0.296715
I0220 23:42:30.629169 139926896920384 retrain.py:1071] 2020-02-20 23:42:30.629147: Step 499: Cross entropy = 0.296715
INFO:tensorflow:2020-02-20 23:42:30.915653: Step 499: Validation accuracy = 90.0% (N=100)
I0220 23:42:30.916121 139926896920384 retrain.py:1087] 2020-02-20 23:42:30.915653: Step 499: Validation accuracy = 90.0% (N=100)
INFO:tensorflow:Final test accuracy = 88.9% (N=45)
I0220 23:42:31.838120 139926896920384 retrain.py:1113] Final test accuracy = 88.9% (N=45)
```

Figure 69 Entraînement du modèle pour la classification des signaux vocaux

Après l'entraînement de notre modèle, nous avons eu une précision d'entraînement de 88.9%. Les performances sont bonnes et le modèle semble bien se généraliser, et avoir une bonne efficacité de prédiction. D'autres tests seront effectués sur des nouvelles données audios et présentés dans la prochaine section du mémoire.

4.4.3 Tester le modèle avec des nouvelles entrées audios

Nous avons testé le modèle sur trois nouveaux sujets, deux sujets étant en état de facultés affaiblies et l'autre étant à l'état normal. Nous avons enregistré leur prononciation de mots lors d'une séance d'acquisition de signaux vocaux. Leurs enregistrements ont été sélectionnés afin de prédire leurs états.

Les figures 70 et 71 représentent les résultats de classifications des 2 sujets en état de facultés affaiblies. Nous avons fixé un seuil de 60 % pour estimer qu'une personne est en état de facultés affaiblies en supposant qu'il pourrait être déterminant pour la détection.

Figure 70 Résultat de test sur un nouvel enregistrement audio du sujet F en état de facultés affaiblies

Figure 71 Résultat de test sur un nouvel enregistrement audio du Sujet G en facultés affaiblies

La figure 72 représente le résultat de la prédiction de sujet en état normal

Figure 72 Résultat de test sur un nouvel enregistrement audio du Sujet H à l'état normal

Notre classificateur était capable de classifier nos données audios et d'avoir une précision satisfaisante avec un pourcentage d'exactitude de 88.9%.

4.5 Automatisation de la classification de signaux EEG

4.5.1 Puissance de signaux EEG

Comme nous pouvons constater dans les figures 73 et 74, chez le sujet A la puissance des signaux est plus importante en état normal par rapport à l'état facultés affaiblies, cela est dû à une diminution de la puissance des signaux en état de facultés affaiblies à cause de la consommation du cannabis. Cette diminution est significative surtout pour les bandes de fréquences alpha, beta et gamma.

Comme nous pouvons le constater en comparant les figures 73 et 74, la puissance du signal augmente dans les bandes de fréquences Alpha, Beta et Gamma 1 chez le sujet A. Cela peut être expliqué par l'effort exercé par ce sujet pour prononcer le mot que nous

avons proposé. Cette augmentation significative peut être aussi expliquée par l'hyperexcitation dû à la consommation de cannabis puisque ce dernier stimule surtout ces deux bandes de fréquences.

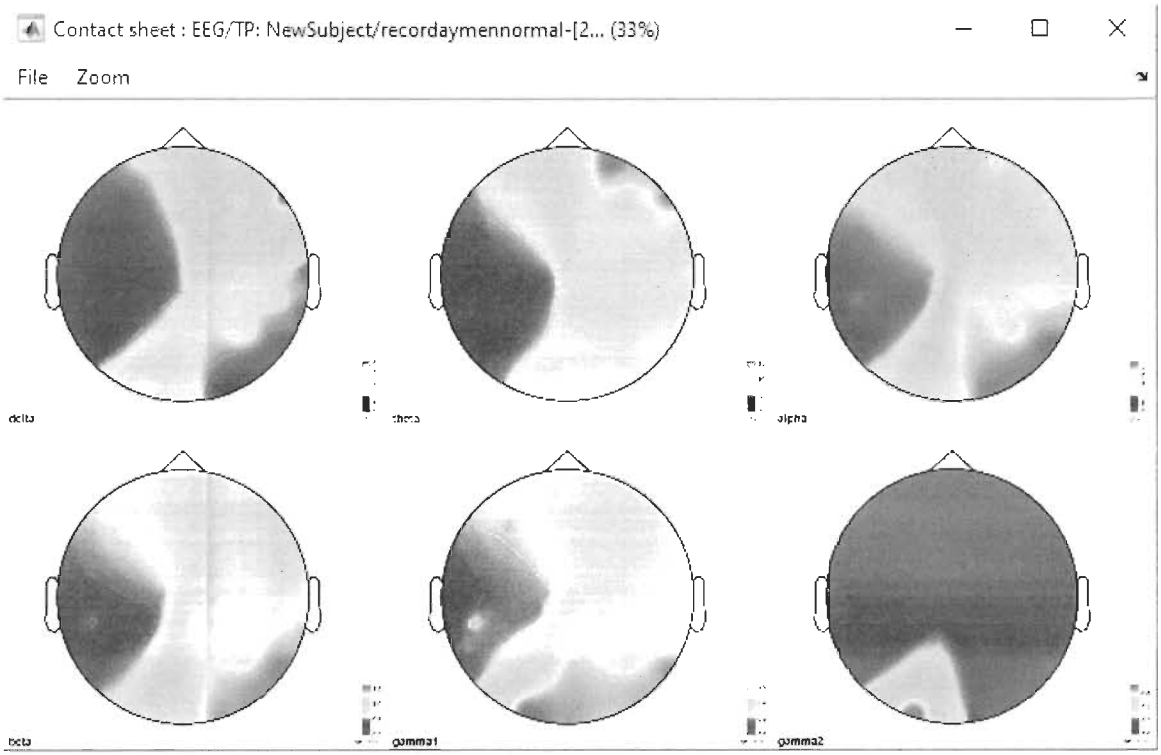

Figure 73 Densité de puissance spectrale en état normal sur différentes bandes de fréquences du sujet A

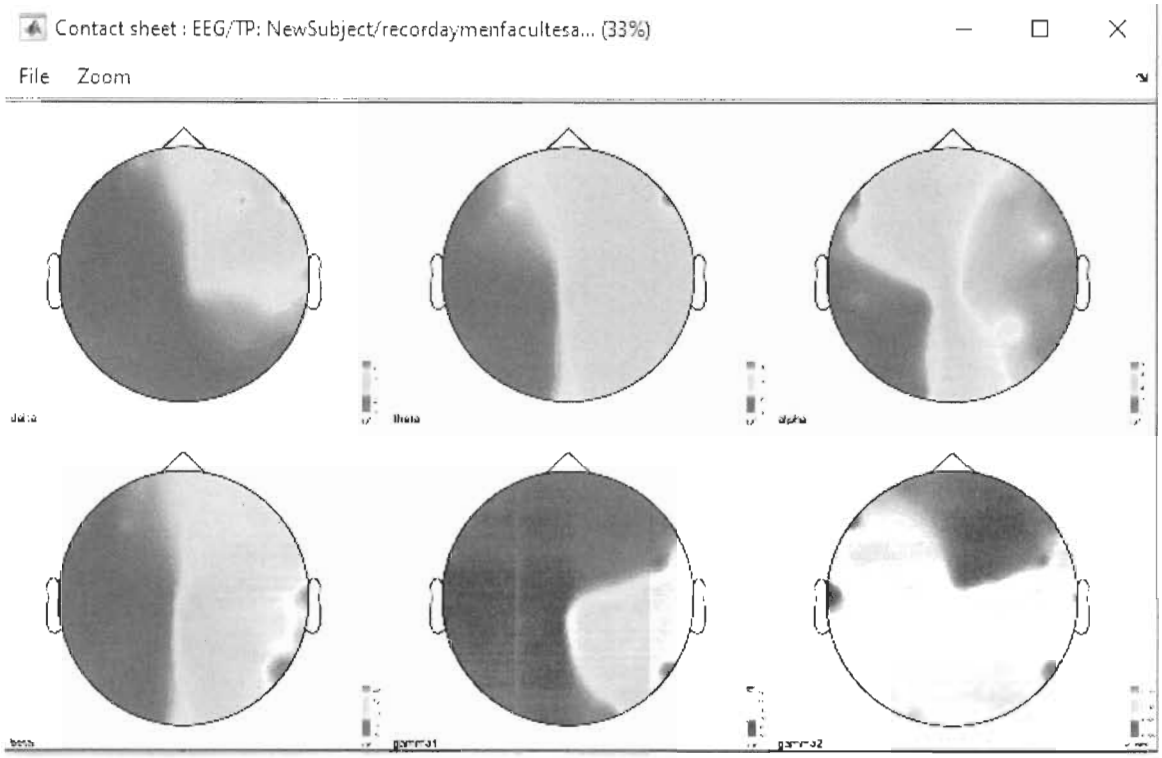

Figure 74 Densité de puissance spectrale en état facultés affaiblies sur différentes bandes de fréquences du sujet A

4.5.2 Réseau de neurones artificiels

Afin de faire la classification de signaux EEG, nous avons créé notre modèle en utilisant un module « two-class neural network » qui permet de créer un classificateur binaire en utilisant un algorithme de réseau de neurones comme celui schématisé sur la figure 75.

Le module « Train model » entraîne la classification créée en utilisant un ensemble de données fournis. Le réseau de neurones pleinement connecté a été entraîné avec un nombre de nœuds cachés de 1000 et un taux d'apprentissage configuré à 0.001 ainsi qu'un nombre d'itérations de 1000 et un poids initial de 0.1.

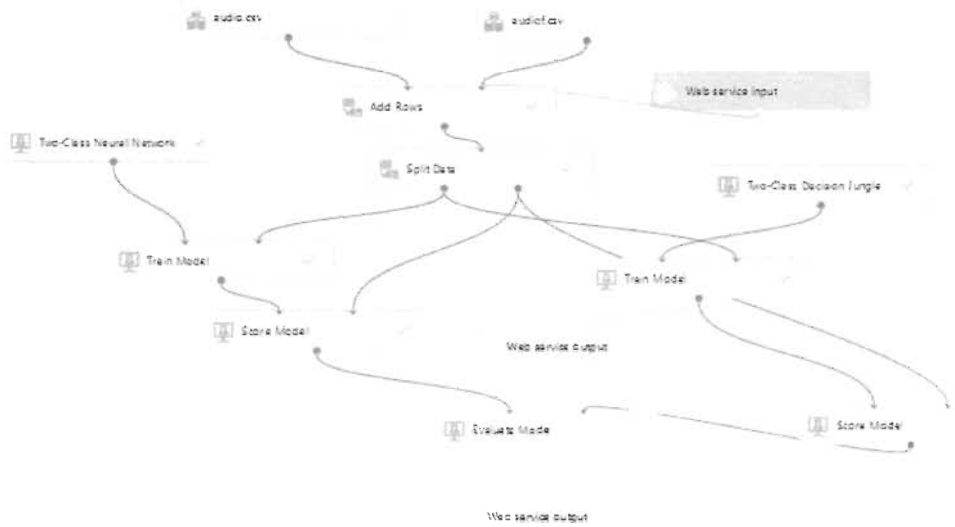

Figure 75 Modèle de classification utilisant l'algorithme de réseau de neurones et SVM

4.5.3 Support vector machine

Dans cette approche nous avons utilisé le module « two-class support vector machine » qui crée un modèle de classification binaire utilisant un algorithme SVM (support vector machine).

Le module train model entraîne la classification créée précédemment dans le modèle de réseau de neurones en utilisant l'ensemble des données fournies.

Ce modèle a été entraîné avec un seul paramètre et le nombre d'itérations a été fixé à 1500 itérations.

4.5.4 Résultat de l'entraînement du modèle

Le pourcentage de précision et d'exactitude de la classification comparé au même ensemble de données avec un threshold de 0.5, est de 0.96 ce qui est excellent et très satisfaisant comme résultat (Voir les figures 76 et 77). Le seuil de détection des facultés affaiblies a été établit à 40%. Nous supposons que ce seuil pourrait être déterminant pour la détection de facultés affaiblies.

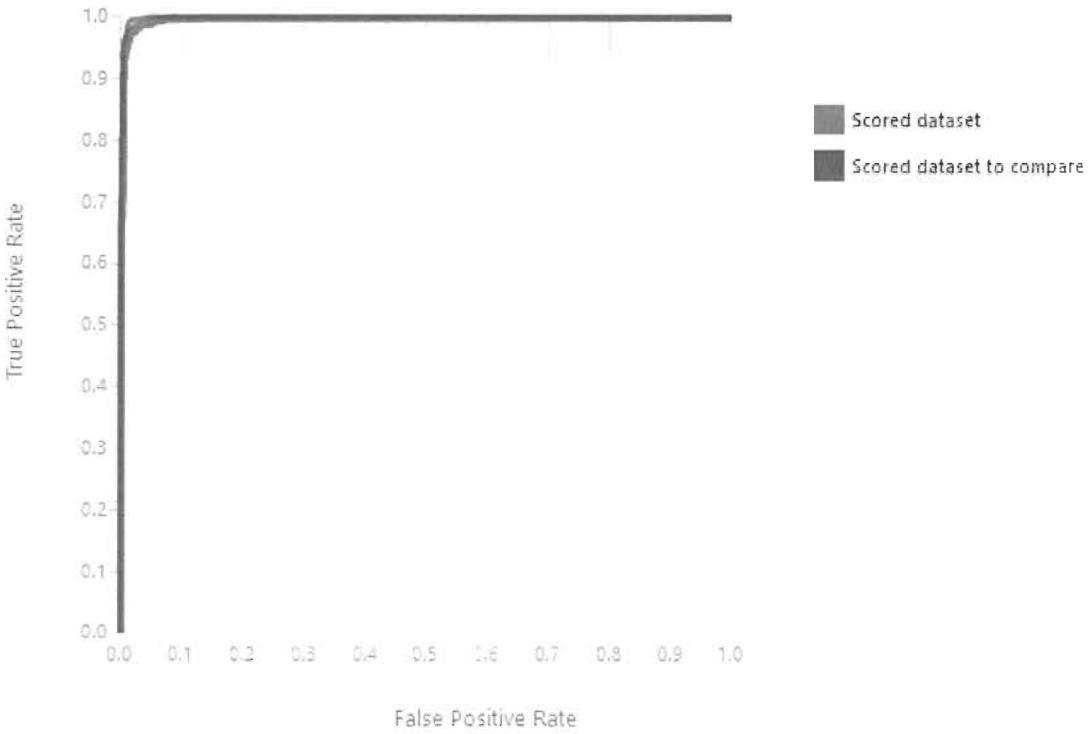

Figure 76 Courbe ROC des résultats de classification utilisant l'algorithme de réseau de neurones.

True Positive	False Negative	Accuracy	Precision
999	50	0.967	0.994
False Positive	**True Negative**	**Recall**	**F1 Score**
6	657	0.952	0.973
Positive Label	**Negative Label**		
1	0		

Figure 77 Résultats de précision et d'exactitude de la classification de signaux EEG

4.5.6 Tester le modèle avec des nouvelles entrées EEG

Afin de tester notre modèle, nous avons utilisé une application de classification de signaux EEG réalisé par Berdaa S. et Meunier F. [63]. Qui nous permet d'introduire de nouveaux signaux EEG à tester.

Nous avons sélectionné un nouveau fichier Edf de signaux EEG de sujet F (celui qui a donné une meilleure prédiction pour la classification audio) à l'état facultés affaiblies

après les avoir acquis pour réaliser ce test et comparer la précision des deux classificateurs.

Le classificateur était capable de prédire que l'état d'un sujet, soit en facultés affaiblies, avec une précision significative par rapport au classificateur audio à 95% facultés affaiblies comme observée dans la figure 78.

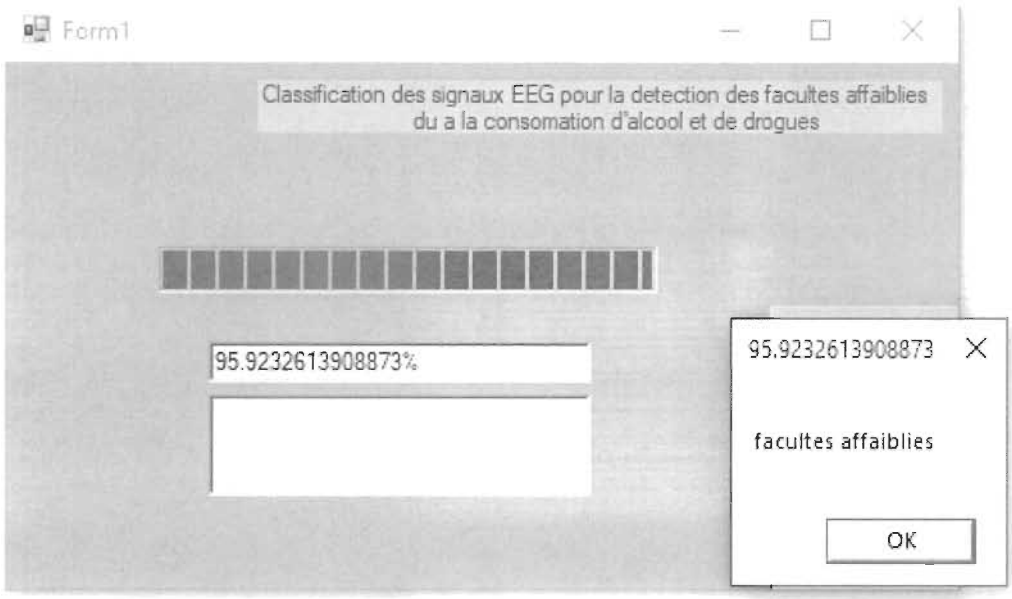

Figure 78 Résultat de classification de signaux EEG du sujet F

4.6 Conclusion

Les résultats que nous avons obtenus sont encourageants pour les deux classificateurs et surtout pour les données EEG avec leurs précisions élevées.

Cette expérimentation nous a permis de détecter tous les sujets en état de facultés affaiblies, de réaliser que les signes de facultés affaiblies sont présents dans les signaux vocaux et signaux EEG et de détecter aussi les états normaux des sujets.

Cependant, les algorithmes de classification nous ont permis de classifier les états de sujets avec succès, donc nous pouvons distinguer entre un discours dont la prononciation de mots est affectée par l'état de la personne entre état normal ou état de facultés affaiblies.

La classification de signaux EEG était plus efficace en termes de précision, cependant des améliorations peuvent être apportées au classificateur de signaux vocaux en exploitant plus de données en augmentant la base de connaissances de l'algorithme proposé par d'autres données ou par des données d'un DataSet publique.

L'amélioration du classificateur de signaux audios pourra valider la détection du discours dysarthrique

Il reste encore du travail pour avoir des résultats plus optimaux, mais sans l'ombre d'un doute nous sommes capables d'automatiser la classification des signaux vocaux et EEG pour la détection de facultés affaiblies dû à la consommation d'alcool ou de cannabis.

De plus, l'EEG enregistré pendant une tâche cognitive est relativement facile à séparer d'un signal EEG de base que les signaux EEG enregistrés dans deux tâches cognitives différentes (Lecture du mot à prononcer et sa prononciation).

Chapitre 5 Conclusion

Dans ce travail de mémoire de recherche, nous avons décrit plusieurs sujets. D'abord nous avons décrit les effets de l'alcool et du cannabis sur la conduite automobile, sur la voix et sur les différentes parties du cerveau ainsi que les ondes cérébrales.

Ensuite, nous avons fait une revue de littérature en parcourant ce qui a été publié dans ce le domaine de traitement et classification des signaux vocaux et EEG et ainsi que des études dont le sujet portait sur des analyses de la voix pour détecter le discours dysarthrique.

En premier lieu, nous avons vu les techniques d'acquisition de signaux vocaux et EEG et puis les approches d'élimination de bruits ainsi que les artéfacts dans ces signaux.

Différentes méthodes de traitements et de classification de ces signaux ont été décrites et analysées. La revue de littérature nous a permis de bien nous orienter et nous a donné des repères à suivre pour réaliser notre recherche.

Nous avons aussi parcouru quelques études et recherches connexes à notre recherche ce qui a été la base de cette dernière et qui nous ont permis d'atteindre nos objectifs et avoir des résultats encourageants.

Ensuite, nous avons décrit une approche de classification de signaux EEG en utilisant la méthode SVM et le réseau de neurone, une approche pour confirmer l'exactitude de l'approche de classification de signaux vocaux dont nous avons utilisé l'analyse de spectrogrammes représentants les données audios en utilisant un réseau de neurone convolutionnels profonds et un modèle de classification d'image entraîné Inception V3 de Google que nous avons réentraîné avec nos données.

Les deux approches ont permis de détecter les facultés affaiblies avec succès, cependant, le classificateur de signaux EEG était plus efficace en termes de précision de prédiction.

Des améliorations pourront être apportées dans les prochaines recherches pour valider une augmentation de précision des deux approches de classifications, en exploitant plus de données afin d'enrichir la base de connaissance des modèles.

Enfin, avec les résultats obtenus, nous pouvons confirmer que notre étude est novatrice dans le domaine et permettrons d'avoir des perspectives pour la détection des facultés affaiblies.

Annexe

Effets rapportés de la consommation aiguë de cannabis sur les paramètres nécessaires à la conduite automobile.

Types de tests	Paramètre	Effet observé selon le contexte de consommation	Références
Tests psychomoteurs	Attention	Diminuée chez consommateurs occasionnels, mais tolérance possible chez les consommateurs chroniques Effet peut durer plusieurs heures après la consommation	(9,38,60)
	Capacité de jugement	Diminuée significativement à court terme et possiblement à long terme si consommation chronique	(27,38,60)
	Coordination psychomotrice	Diminuée selon tâches de conduite critique ou de contrôle moteur[1, 20]	(1,17,20)
	Fonctions exécutives	Diminution significative	(20,61)
	Fonctions visuelles	Peu ou pas d'effets	(9,27)
	Impulsivité motrice	Diminution significative et contrôle des pulsions diminué	(20,59)
	Mémoire à court terme	Altérée après une consommation aiguë, mais pas d'effets résiduels	(9,60)
	Prise de risques	Altérée, diminution de la vigilance	(38)
	Suivi de trajectoire	Diminuée de manière significative	(1,9,20,27)
	Temps de réaction	Augmenté, toutefois de faibles doses auraient un impact beaucoup moindre	(9,20,38,51)
Simulateurs de conduite	Attention et concentration	Diminuées	(17)
	Capacité à éviter des obstacles	Diminuée	(27,38)
	Capacité d'apprentissage	Diminuée	(27)
	Prise de risque	Diminuée	(17,27)
	Omission de panneaux	Augmentée	(38)
	Réponse à des tâches secondaires	Diminuée, oubli de faire tâches préassignées	(27,38)
	Sorties de route	Augmentées	(52)
	Suivi de trajectoire et contrôle de la voiture	Déviation de la position latérale, distance augmentée avec les autres véhicules	(17,27,51)
	Temps de réaction	Délai de réaction et temps de décision augmentés.	(17,38,59)
	Temps de freinage	Augmenté	(27,38)
	Vitesse	Diminuée, par compensation du conducteur	(27)
	Vigilance	Diminuée	(17)
Conduite dans un contexte réel	Omission de panneaux	Augmentée	(38)
	Performance lors de situations imprévues	Diminuée	(14)
	Prise de risques	Diminuée, par compensation du conducteur	(14, 38)
	Suivi de trajectoire, contrôle de la voiture	Déviation de la position latérale	(27,38)
	Temps de décision	Augmenté	(27)
	Vitesse	Diminuée par compensation du conducteur	(14,38)
	Relation avec autres voitures	Vitesse entre véhicules augmentée, mais dépassements non sécuritaires	(38)

Tableau 3 Effets rapportés de la consommation aiguë de cannabis sur les paramètres nécessaires lors de la conduite automobile

Références

[1] Erin E.Beasley, Douglas J.Beirness, Amy J.Porath-Waller, centre canadien de lutte contre l'alcoolisme et les toxicomanies, Étude visant à comparer les décès sur les routes impliquant l'alcool et la drogue, 2011

[2] https://saaq.gouv.qc.ca/securite-routiere/comportements/alcool/ce-que-dit-la-loi/ consulté janvier 2018.

[3] Hall W, Degenhardt L. Adverse health effects of non-medical cannabis use. The Lancet. 2009; 374(9698): 1383-91.

[4] Fischer B, Jeffries V, Hall W, Room R, Goldner E, Rehm J. Lower Risk cannabis use Guidelines for Canada (LRCUG): A narrative review of evidence and recommendations. Canadian Journal of Public Health. 2011.

[5] Publication de l'université McGill, les jeunes canadiens sont plus à risque d'être impliqués dans un accident de la route après avoir consommé du cannabis, indiquent les résultats d'une nouvelle étude. https://publications.mcgill.ca/lebulletel/2018/10/15/les-jeunes-canadiens-sont-plus-a-risque-detre-impliques-dans-un-accident-de-la-route-apres-avoir-consomme-du-cannabis-indiquent-les-resultats-dune-nouvelle-etude/ , consulté Janvier 2018.

[6] Xiaohua Zhao, Xingjian Zhang, and Kian Rong, Study of the Effects of Alcohol on Drivers and Driving Performance on Straight Road, 2014.

[7] http://www.cinq.ulaval.ca/EEG consulté janvier 2018.

[8] https://lnc.univ-amu.fr/fr/techniques-utilisees consulté janvier 2018.

[9] Millet, D. 2001 « Hans Berger: from psychic energy to the EEG ». Perspectives in biology and medicine 44 (4): 522-542

[10] Ashwal, S., and Rust, R., „Child neurology in the 20th century", Pedia. Res., 53, 2003, 345–361

[11] Lukas SE, Mendelson JH, Benedikt R, Electroencephalographic correlates of marihuana-induced euphoria.1995, Alcohol and Drug Abuse Research Center, McLean Hospital/Harvard Medical School: Belmont

[12] Sterman, M. B., MacDonald, L. R., and Stone, R. K.Biofeedback training of sensor motor EEG in man and its effect on epilepsy", Epilepsies, 15, 1974, 395–416

[14] https://thanatofrance.wordpress.com/2015/08/15/le-systeme-nerveux-les-neurones/ consulté Mars 2018.

[15]http://www.claudegabriel.be/Cine%20acoustique%209.pdf Consulté Avril 2018

[16] Gardé, la voix, Paris, Presse Universitaire de France, 1970

[17] Physiologie de la voix chantée: vibrations laryngées et adaptations phono-résonantielles, Entretiens de Médecine Physique et de réadaptation, Montpellier Mars 2012.

[18] Trolliet-Cornut, A. (2002). « La dynamique diaphragmatique », Moyens d'investigation et pédagogie de la voix chantée, coordination de Guy Cornut, Symétrie, p. 1-7

[19] Pettersen, V., Bjørkøy, K., Torp, H. et al. (2005). Neck and shoulder muscle activity and thorax movement in singing and speaking tasks with variation in vocal loudness and pitch. Journal of Voice.

[20] Crevier-Buchman, L., Vincent, C. (2009). "Visualisation du larynx : endoscopie rigide et fibroscopie, stroboscopie et caméra haute vitesse", in A. Marchal & C. Cavé (Eds.), Imagerie médicale pour l'étude de la parole. Paris : Hermès-Lavoisier

[21] Nathalie Henrich Bernardoni, Étude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception. Thèse de doctorat de l'Université Paris 6, 2001.

[22] Echternach, M., Sundberg, J., Arndt, S. et al. (2010). Vocal tract in female registers—A dynamic real-time MRI study. Journal of Voice

[23] Hueber T. (2013). Ultraspeech-player: Intuitive visualization of ultrasound articulatory data for speech therapy and pronunciation training, Proceedings of Interspeech (show&tell), Lyon, France, 2013, 752-753.

[24] Nathalie Henrich Bernardoni, Audrey Acher Comment analyser la voix humaine dans la parole et dans le chant ? Les outils scientifiques et méthodes de la recherche fondamentale à disposition de la recherche clinique sur la voix et leurs implications en orthophonie

[25] Ternström, S., Granqvist, S. (2010). Personal computers in the voice laboratory: Part two-audio devices. Logopedics Phoniatrics Vocology, 2010.

[26] Alexander Bertrand, Marc Moonen, Distributed eye blink artifact removal in a wireless EEG sensor network, 2016.

[27] A Survey of the recent Architectures of Deep Convolutional Neural Networks:

Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi.

[28] Grega Repovš Dealing with Noise in EEG Recording and Data Analysis Research Review Paper.2010, University of Ljubljana: Ljubljana. 2010.

[29] Luck S: An Introduction to the Event-Related Potential Technique. Cambridge, MA 2005: MIT Press.

[30] Removal of ocular artifact from the EEG: a review. Neurophysiologie clinique 2000; 30(1): 5-19

[31] https://fr.wikipedia.org/wiki/Analyse_en_composantes_ind%C3%A9pendantes. Consulté Juin 2018.

[32] Vanhatalo S, Palva JM, Holmes MD, Miller JW, Voipo J, Kaila K, Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep.2004, University of Helsinki: Helsinki

[33] Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ, Removing electroencephalographic artifacts by blind source separation.2000, Howard Hughes Medical Institute, Salk Institute: San Diego, California.

[34] Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ, Long-range temporal correlations and scaling behavior in human brain oscillations.2001, Helsinki University Central Hospital: Helsinki.

[35] J. Leon Kenemans ,Peter C.Molenaar, Marinus N. Verbaten, Jef L. Slangen, Removal of the Ocular Artifact from the EEG: A Comparison of Time and Frequency Domain Methods with Simulated and Real. 1991, Department of Psychonomics, University of Amsterdam: Amsterdam.

[36] Theodore C. Weerts, Peter J.Lang, The effects of eye fixation and stimulus and response location on the Contingent Negative Variation (CNV).1972, University of Wisconsin: Madison.

[37] Sean P.Fitgibbon et al., Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG.2007, Flinders University : Adelaide.

[38] Julie Onton, Marissa Westerfield, Jeanne Townsend, Scott Makeig , Imaging human EEG dynamics using independent component analysis.2006, University Of California : San Diego.

[39] (T.Radüntz J.Scouten O.Hochmuth B.Meffertb: EEG artifact elimination by extraction of ICA-component features using image processing algorithms)

[40] https://www.futura-sciences.com/sciences/definitions/mathematiques-transformee-fourier-11880/ Consulté Juin 2018.

[41] Fabien Lotte, Marco Congedo, Anatole L´ecuyer, Fabrice Lamarche, Bruno Arnaldi: A review of classification algorithms for EEG-based brain–computer interfaces

[42] Lyons RG. Understanding digital signal processing, 3/E. Upper Saddle River: Prentice Hall; 2011.

[43] Proakis JG, Manolakis DG. Digital signal processing: principles algorithms and applications.

[44] Erik Andreas Larsen, Classification of EEG Signals in a Brain-Computer Interface System.2011, Norwegian University of Science and Technology.

[45] Guang R.Gao et al, Languages and Compilers for Parallel Computing, 2009

[46] http://cermics.enpc.fr/polys/oap/node86.html Consulté juin 2018.

[47]https://server.nitrado.net/fre/news2/view/realite-virtuelle-emotiv/?f=index/mode:0/page:1

[48] https://e.maxicours.com/img/4/4/3/1/443187.jpg consulté Juillet 2018.

[49] https://www.diytdcs.com/media/010_EEG_standard.gif consulté juillet 2018.

[50] M. Abo-Zahhad, Sabah M. Ahmed, and Sherif N. Abbas, A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals. 2015, Department of Electrical and Electronic Engineering, Assiut University: Assiut.

[51] http://ressources.unisciel.fr/DAEU-biologie/P2/co/P2_chap4_c03.html consulté juillet 2018.

[52] http://www.claudegabriel.be/Cine%20acoustique%209.pdf consulté Juillet 2018.

[53] Nathalie Henrich Bernardoni: Physiologie de la voix chantée: vibrations laryngées et adaptations phono-résonantielles

[54] Système d'endoscopie basé sur la tomographie par cohérence optique dédié à l'étude des cordes vocales pédiatriques et des pathologies associées, Fouzi Benboujia, Université de Montréal, 2016.

[55] The Vocal Principles of Garcia as Represented by His Pupils: Bataille, Marchesi, and Stockhausen, James Richard Joiner

[56] Guidelines for Selecting Microphones for Human Voice Production Research: Jan G.; Granqvist, Svante, 2010.

[57] http://newt.phys.unsw.edu.au/jw/voice.html Consulté Septembre 2018.

[58] Minu George Thoppil, C. Santhosh Kumar, Anand Kumar, and John Amose : Speech Signal Analysis and Pattern Recognition in Diagnosis of Dysarthria, 2017

[59] Abdellah Kachaab, Francis Grenezb, Juan Rafael Orozco Arroyavecd, Jean Schoentgenb : Principal component analysis of the spectrogram of the speech signal: Interpretation and application to dysarthric speech

[60] https://lecerveau.mcgill.ca/flash/d/d_10/d_10_cr/d_10_cr_lan/d_10_cr_lan.html Consulté Septembre 2018.

[61] https://www.frcneurodon.org/comprendre-le-cerveau/a-la-decouverte-du-cerveau/le-langage/ Consulté Septembre 2018.

[62] Alexander Bertrand*, Marc Moonen: Distributed eye blink artifact removal in wireless EEG sensor network, mai 2014.

[63] Berdaa, S., Meunier, F., Classification des signaux EEG pour la détection des facutlés Affaiblies du à la consommation d'alcool et de drogues, 2018.

[64] E.Malar, M.Gauthaam, D.Chakravarthy, A Novel Approach for the Detection of Drunken Driving using the Power Spectral Density Analysis of EEG.2011.

[65] IIT GUWAHATI Virtual Lab, "identification of Voice/Unvoiced/Silent regions of speech", 2011

[66] Rabiner L. R. and Schafer R. W., Digital processing of speech signals, 1979

[67] Atal B.S. and Rabiner L.R, "A Pattern Recognition Approach To Voiced-Unvoiced-Silence Classification With Applications To Speech Recognition," IEEE Trans. Acoust., Speech, and signal processing. vol. 24(3), pp. 201-211, 1976.

[68] http://www.resodys.org/IMG/pdf/Meunier_Production_2010.pdf

[69] https://fr.wikipedia.org/wiki/Voisement

[70] PHONÉTIQUE ACOUSTIQUE Christine Meunier Laboratoire Parole et Langage, CNRS UMR 6057 - Université de Provence

[71] Évaluation de la parole dysarthrique : Apport du traitement automatique de la parole face à l'expertise humaine Imed Laaridh- Université d'Avignon, 2017

[72] S Skodda, W Gronheit, and U Schlegel, "Intonation and speech rate in Parkinson's disease: General and dynamic aspects and reponsiveness ti levodopa admission,"J. voice, vol. 25, pp. 199-205, 2011

[73] R. D. Kent, "Acoustic studies of dysarthric speech methods, progress, and potential" J. commun. Disord, vol. 32 pp. 141-186, 1999

[74] J. R. Orozco-Arroyave et al., "Automatic detection of Parkinson's disease in running speech spoken in three different languages," J. Acoust. Soc. Am., 139(1), 148-500,2016.

[75] Dysarthric speech analysis by means of the principal components of the spectrogramA. Kacha, F. Grenez, Juan R. Orozco-Arroyave, J. Schoentgen

[76] Parkinson Disease Detection from Speech Articulation Neuromechanics Pedro Gómez-Vilda et al., 2017 (Pedro Gómez-Vilda, Jiri Mekyska, José M. Ferrández, Daniel Palacios-Alonso, Andrés Gómez-Rodellar, Victoria Rodellar-Biarge, Zoltan Galaz, Zdenek Smekal, Ilona Eliasova, Milena Kostalova, and Irena Rektorova)

[77] http://festvox.org/cmu_arctic/dbs_rms.html

[78] https://towardsdatascience.com/beginners-guide-to-speech-analysis-4690ca7a7c05

[79] https://www.f-legrand.fr/scidoc/docimg/numerique/tfd/echantillonnage/echantillonnage.html

[80] Nikolaos Mitianoudis, A Graphical Framework For The Evaluation Of Speaker Verification Systems,

[81] Urmila Shrawankar, Dr. Vikas Thakrey: Techniques for feature extraction in speech recognition system: A comparative Study

[82] Wu QZ, Jou IC, Lee SY. On-line signature verification using LPC cepstrum and neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 1997, 27(1):148-153

[83] Young S, Evermann G, Gales M, Hain T, Kershaw D, Liu X, et al. The HTK Book, Version 3.4. Cambridge, United Kingdom: Cambridge University; 2006

[84] Wang TT, Quatieri TF. High-pitch formant estimation by exploiting temporal change of pitch. IEEE Transactions on Audio, Speech, and Language Processing. 2010

[85] https://www.wavemetrics.com/products/igorpro/dataanalysis/signalprocessing/powerspectra.

[86] https://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab.htm

[87] X. Liu, Z. Deng, and Y. Yang, —Recent progress in semantic image segmentation,‖ Artif. Intell. Rev., vol. 52, no. 2, pp. 1089–1106, 2019

[88] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, —Deep neural networks segment neuronal membranes in electron microscopy images, in Advances in neural information processing systems, 2012.

[89] Y. LeCun, K. Kavukcuoglu, C. C. Farabet, and others, —Convolutional networks and applications in vision, in ISCAS, 2010.

[90] D. Scherer, A. Müller, and S. Behnke, —Evaluation of pooling operations in convolutional architectures for object recognition, in Artificial Neural Networks--ICANN 2010.

[91] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E.Muharemagic, —Deep learning applications and challenges in big data analytics, J. Big Data.

[92] Ian Goodfellow, Y. Bengio, and A. Courville, Deep learning, Nat. Methods

[93] J. Bouvrie, 1 Introduction Notes on Convolutional Neural Networks,2006

[94] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature, 2015.

[95] K. He, X. Zhang, S. Ren, and J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell.,

[96] Y. Boureau, Icml2010B.Pdf, 2009

[97] C.-Y. Lee, P. W. Gallagher, and Z. Tu, Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree, in Artificial Intelligence and Statistics, 2016.

[98] J. Gu et al., Recent advances in convolutional neural networks, 2018

[99] T. Wang, D. J. D. J. Wu, A. Coates, and A. Y. Ng, —End-to-end text recognition with convolutional neural networks, 2012.

[100] B. Xu, N. Wang, T. Chen, and M. Li, —Empirical Evaluation of Rectified Activations in Convolutional Network, May 2015.

[101] P. Ramachandran, B. Zoph, and Q. V. Le, —Swish: a Self-Gated Activation Function, 2017.

[102] S. Ioffe and C. Szegedy, —Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015.

[103] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, 2012

[104] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfittin, 2014

[105] M. Lin, Q. Chen, and S. Yan, Network In Network, 2013

[106] W. Rawat and Z. Wang, Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review, 2016

[107] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016

[108] M. Lin, Q. Chen, and S. Yan, Network In Network, 2013

[109] https://www.inspq.qc.ca/pdf/publications/2044_effets_cannabis_conduite_automobile.pdf

[110] EEG in Epilepsy: Current Perspectives, M. Sundaram, R.M. Sadler, G.B. Young and N. Pillay

[111] C. Szegedy et al.: Going deeper with convolutions,IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, vol. 57, no. 3, pp. 1–9.