Ministère de l'Enseignement Supérieur et de la Recherche scientifique
وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR-ANNABA

جامعة باجي مختار - عنابة

Faculté des Sciences de l'Ingéniorat **Année : 2014/2015**
Département d'Électronique

THESE

Présentée en vue de l'obtention du diplôme de ***DOCTORAT***

Système d'Assistance à l'Élocution Altérée en Langue Arabe

Option

Signaux et Images

Par

DAHMANI Habiba

Rapporteur : M. Noureddine DOGHMANE Professeur Université Badji Mokhtar, Annaba
Co-rapporteur : M. Sid-Ahmed SELOUANI Professeur Université Moncton, Canada

Devant le Jury

Président :	M.DJEGHABA Messaoud	Professeur	Université Badji Mokhtar, Annaba
Examinateur :	Mme.BAHI Halima	Professeur	Université Badji Mokhtar, Annaba
Examinateur :	M.BOUKROUCHE Abdelhani	Professeur	Université de Guelma
Examinateur :	M.AMARDJIA Nourredine	MCA	Université de Sétif
Invité	M.FEZARI Mohamed	MCA	Université Badji Mokhtar, Annaba

Année Universitaire : 2014/2015

À ma mère **: Rebeh**

À mes enfants **:** Myra et Ibrahim

Et à la mémoire de **:**

Mon Père Salah

Mon frère Ibrahim

Mon bien aimé : Yakine

Que le Bon Dieu les miséricordes

Remerciements

Je tiens en premier lieu à remercier mes directeurs de thèse, Monsieur Sid-Ahmed SELOUANI (Université de Moncton, Canada), et Monsieur Noureddine DOGHMANE (université de Annaba) auprès duquel j'ai énormément appris. Je les remercie pour toute l'attention et pour le soutien qu'ils m'ont porté durant ces années. Ils m'ont communiqué leur volonté d'aller toujours plus loin et j'ai pris un réel plaisir à effectuer cette thèse sous leurs directions.

Je voudrais également remercier M. Mohamed CHETOUANI (Institut des systèmes intelligents et robotiques, Paris, France : ISIR) et Monsieur Douglas O'SHAUGHNESSY (Institut INRS, Montréal, Canada) de m'avoir bien accueilli dans leur laboratoires.

Je voudrai aussi remercier M.DJEGHABA Messaoud d'avoir voulu présider le jury de thèse, Mme Halima BAHI (Université Badji Mokhtar, Annaba), M.Boukrouche Abdelhani (Université de Guelma) et M. Nourredine AMARDJIA (Université de Sétif) d'avoir accepté d'examiner cette thèse. Également, je tiens à remercier M. Mohamed FEZARI (Université Badji Mokhtar, Annaba) d'avoir bien voulu participer à ce jury à titre d'invité. Leurs remarques et leurs questions pertinentes contribueront à mes futurs travaux de recherche.

Je remercie également mes parents qui m'ont soutenu durant ces années. Une attention toute particulière se tourne vers mon compagnon Wassef Alsekhaneh, qui a réussi à me supporter tout au long de ce chemin et qui a toujours su se montrer positif vis-à-vis de mon travail, même dans les moments les plus difficiles.

Un dernier remerciement tout spécial pour OumAissa le baby-sitter de mes enfants ; son aide était très précieux.

D'autres personnes ont aussi contribué à cette thèse : merci à ma famille surtout mes sœurs, à mes amis et à mes collègues.

ملخص

الهدف الرئيسي لهذه الأطروحة يتمثل في إقامة نظام لتحليل إشارة الكلام لأشخاص مصابين بحبسة بروكا(Aphasie de Broca) والمصابين بمرض الديزرتريا (la dysarthrie) .هذا النظام يقوم على التحليل النغمي أو الإيقاعي (Rythmique) للكلام و الصوت خاصة بالإضافة إلى التحاليل الأخرى كتلك الخاصة بالتحليل الطيفي لإشارة الكلام (L'analyse spectrale). الدعامات النغمية أو المميزات النغمية هي تلك التي أستعملت للتمييز بين اللغات مؤخرا حيث حققت نجاحا كبيرا. ننوه كذلك بنجاح هاته الأخيرة في التمييز بين اللهجات العربية .

على كل حال ، يعتقد أن أهم مما يميز كلام المصابين بحبسة بروكا و الديزرتريا هو تأثرنغمته و هذا كان دافعا قويا للقيام بهذه الدراسة. إن مثل هذه الدراسات الخاصة بكلام المرضى خاصة التي تخص كلام مرضى حبسة بروكا باللهجة الجزائرية يظل قليل جدا. في هذا العمل، نقترح مقاربة للتعرف على كلام المرضى وتصنيفه بالنسبة لكلام الأصحاء و أيضا تصنيف مختلف مستويات شدة المرض والمرجع كلام االأشخاص المعافين.

لتحقيق و تفعيل مختلف التجارب على اللغة الإنجليزية الأمريكية والتي هو ركيزة قاعدة البيان نيمورز (Nemours) للدزرتريا وعلى اللغة العربية بلهجة شرق الجزائر من خلال قاعدة البيانات التي جمعناها من مرضى حبسة بروكا من كلا المستشفيين المتخصصيين (EHS) رأس الما بمدينة سطيف و سرايدي بمدينة عنابة.

لا بد أن ننوه بالنتائج التي تحصلنا عليها بعد ما قمنا بتقسيم إشارة الكلام إلى أجزاء صوتية و غير صوتية و كذلك إلى وحدات ساكنة و وحدات متحركة كانت مشجعة للغاية. فإن النتائج التجريبية للتعرف وتصنيف كلام المرضى باستخدام دعامات النغم أو الإيقاع وبالمقارنة مع استعمال مميزات أخرى مرضية جدا خاصة كذلك فيما يخص تصنيف مستويات خطورة المرض.

أخيرا،نخلص للقول أن التجارب التي أجريناها أظهرت أن الإيقاع هو مميز صوتي و كلامي كافي لتحديد و تمييز خصائص كلام المرضى على نحو فعال.

ABSTRACT

The main objective of this thesis is the development of an acoustic-prosodic analyzer of pathological speech. Few studies have been devoted to the rhythmic aspect of pathological speech especially the aphasic speech. In general pathological speech and particularly the Arab dialects remains one of the research areas rarely explored. In this work, we propose an approach for the recognition and classification of pathological speech also to effectively characterize the different level of severity of dysarthria and aphasia, taking as reference the healthy speech.

Preliminary experiments have shown the ability of rhythmic parameters (proposed recently in the field of identification of languages) to well discriminate pathological from healthy speech and also to identify different levels of severity of pathological speech.

For the validation experiments are carried out on American English which is the substrate of the dysarthric database Nemours and Arabic dialect of eastern Algeria through the database ADAD (Algerian Dialectal Aphasic Database) that we have collected from aphasic patients from both institutions specialized hospital (EHS) Ras Elma in Sétif and Seraidi in Annaba.

Note that the results of the segmentation voiced / unvoiced and consonant / vowel by that we have used to analyze the corpus of Algerian dialect are very encouraging. Finally, the experimental results of the recognition and classification using rhythmic parameters and other prosodic parameters as inputs of various classifiers machines were also very satisfactory. The experiments we have conducted showed that the rhythm is an acoustic parameter sufficient to characterize pathological speech effectively.

Résumé

L'objectif de la thèse est le développement d'un analyseur acoustique prosodique de la parole pathologique. Peu d'études ont été consacrées à l'aspect rythmique de la parole pathologique spécialement l'aphasique. De façon générale, la parole pathologique arabe et particulièrement celle portant sur le dialectal reste l'un des domaines de recherche rarement exploré. Dans le cadre de ce travail, nous proposons une approche pour la reconnaissance et la classification de la parole pathologique permettant également de caractériser efficacement les différentes catégories de sévérité de la dysarthrie et de l'aphasie en prenant comme référence la parole normale (saine).

Des expériences préliminaires ont montré la capacité des paramètres rythmiques (proposés récemment dans le domaine d'identification des langues) de bien discriminer la parole pathologique de la parole normale et aussi d'identifier les différents niveaux de sévérité de la parole pathologique.

Pour la validation, les expériences sont menées sur la langue anglaise américaine qui est le substrat de la base de donnée dysarthrique Nemours et le dialecte arabe de l'est algérien à travers la base de données ADAD (Algerian Dialectal Aphasic Database) que nous avons collecté auprès des patients aphasiques des deux établissements hospitaliers spécialisés (EHS) de Ras-Elma de Sétif et de Seraidi d'Annaba.

Notons que les résultats de la segmentation voisée / non voisée et voyelle / consonne que nous avons utilisés pour analyser le corpus du dialecte algérien sont très encourageants. Enfin, les résultats des expériences de la reconnaissance et de la classification utilisant les paramètres rythmiques et les autres paramètres prosodiques comme entrées des divers classifieurs automatiques étaient également très satisfaisants. Les expériences que nous avons menées ont démontré que le rythme est un paramètre acoustique pertinent pour caractériser efficacement la parole pathologique.

LISTE DES TABLEAUX

LISTE DES FIGURES

Liste Des Acronymes

ACP	**A**nalyse en **C**omposante **P**rincipale
ADAD	**A**lgerian **D**ialectal **A**phasic **D**ata**B**ase
ANOVA	**AN**alysis **O**f **VA**riance
API	**A**lphabet **P**honétique **I**nternational
AVC	**A**ccidents **V**asculaires **C**érébraux
BD	**B**ase de **D**onnées
CP	**C**erebral **P**alsy
CWT	**C**ontinuous **W**avelet **T**ransform
DCT	**D**iscrete **C**osine **T**ransform
DFA	**D**iscriminant **F**unction **A**nalysis
DFB	**D**ivergence **F**orward **B**ackward
DFT	**D**iscrete **F**ourier **T**ransform
DHMM	Discret Hidden Markov Model
DWT	**D**iscrete **W**avelet **T**ransform
EHS	**É**tablissement **H**ospitalier **S**pécialisé
EPD	**E**nd **P**oint **D**etection
EZW	**E**mbedded coding with **Z**ero tree of **W**avelet coefficients
FDA	**F**renchay **D**ysarthria **A**ssessment
FFT	**F**ast **F**ourier **T**ransform
FR	**F**ollow up **R**ate
GMM	**G**aussian **M**ixture **M**odel
HC	**H**ealthy **C**ontrol
HTK	**H**idden **M**arkov **M**odel **T**oolkit
IMC	**I**nfirmité **m**otrice **C**érébrale
ISIR	**I**nstitut des **S**ystèmes **I**ntelligents et de **R**obotique
MAP	**Ma**ximum **A P**osteriori
MLLR	**M**aximum **L**ikelihood **L**inear **R**egression
MLP	Multi **l**ayer **P**erceptron

MMC	**M**odèles de **M**arkov **C**achés
MMG	**M**odèle De **M**élange **G**aussien
MSA	**M**odern **S**tandard **A**rabic
NLA	**N**on-**L**inear **A**pproximation
PA	**P**arole **A**phasique
PF	**P**atient **F**éminin
PM	**P**atient **M**asculin
PM	**P**atient presque guéri (**M**ILD Patient).
PMO	**P**atient **MO**déré
PN	**P**arole **N**ormale
PS	**P**atient **S**évère
QV	**Q**uantification **V**ectorielle
RAP	**R**econnaissance Automatique de la **P**arole
RMSE	**R**oot-**M**ean-**S**quare **E**rror
SAMPA	**S**peech **A**ssessment **M**ethods **P**honetic **A**lphabet
SNR	**S**ignal-to-**N**oise **R**atio
SR	**S**peaking **R**ate
TO	**T**ransformée en **O**ndelettes
TPZ	**T**aux de **P**assage de **Z**éro
TS	**T**ied **S**tates
WAcc	**W**ord **A**ccuracy
WER	**W**ord **E**rror **R**ate
WT	**W**avelet **T**ransform
ZCR	**Z**ero **C**rossing **R**ate

TABLES DES MATIÈRES

Chapitre 1 : Introduction Générale

« Une bonne parole c'est comme un bon arbre dont la racine est solide et dont les branches vont jusqu'au ciel. Il donne ses fruits en toute saison avec la permission du Seigneur ».

Le Saint Coran

Notre travail s'intéresse au traitement automatique de la parole aphasique de Broca et la parole dysarthrique : étude prosodique, classification et reconnaissance. Ces deux pathologies qui sont également le lot de la paralysie cérébrale et d'autres affections neurodégénératives telles que le parkinson, la sclérose en plaque, le syndrome cérébelleux...etc.

Depuis plus d'une vingtaine d'années, l'étude de la parole pathologique résultante des dysfonctionnements de la voix intéresse les laboratoires de recherche issus des sciences du langage. Les chercheurs confrontent les résultats de leurs recherches établies sur des corpus de parole normale à des situations d'élocution pathologique. En effet, le dysfonctionnement aide à comprendre le fonctionnement.

La parole pathologique désigne les troubles de la parole et du langage et aussi la parole produite par des locuteurs atteints de dysfonctionnement de la voix. La parole pathologique revête différentes formes plus ou moins sévères qui peuvent aller jusqu'à altérer complètement une situation de communication.

La question qui se pose généralement est la suivante : sommes-nous capables de trouver un mode ou un outil de communication de substitution ou d'assistance qui soit compatible entre les locuteurs : enfants et adultes ayant ou non des troubles de langage ou de la parole ? Une solution peut-elle répondre en globalité à ces difficultés ? En fait, le problème de l'intelligence artificielle est un sujet d'actualité et pour l'instant, seules les solutions partielles sont aptes à répondre aux différentes tâches que la machine doit effectuer. Cependant, l'évolution fulgurante des technologies et de systèmes de synthèse et de reconnaissance de la parole a facilité et

multiplié l'utilisation de ces systèmes dans différents domaines. Dans la plupart des cas, ces technologies sont utilisées pour les locutrices et locuteurs normaux. Mais cela n'a pas empêché l'existence d'un certain nombre de systèmes pour aider les personnes handicapées en générale. En effet, il existe des systèmes qui ont été développés pour pallier aux problèmes de reconnaissance et corriger la sortie vocalisée par synthèse de la parole afin d'aider les personnes ayant des difficultés d'articulation ou des locuteurs souffrant de dysfonctionnement de la voix (en général, il s'agit de la dysarthrie ou de la dysphonie) [1-9].

En fait, la difficulté réside dans le fait de trouver un mode de communication de substitution ou d'assistance qui soit compatible entre les locuteurs : enfants et adultes ayant ou non des troubles de langage ou de parole. Cette difficulté est loin d'être simple à surmonter car la sévérité du trouble du langage est extrêmement variable d'un locuteur à un autre et d'une pathologie à une autre. La parole produite par des patients atteints de troubles du langage ou de la parole telle que l'aphasie ou la dysarthrie, est difficilement compréhensible.

1.1. État de l'art de la parole pathologique arabe et algérienne.

D'après la recherche bibliographique menée jusqu'à la date de la rédaction de cette thèse, nous avons trouvé peu de travaux sur la parole pathologique arabe. En fait, nous avons recensé surtout des travaux sur les distorsions de la voix et spécialement la dysphonie qui a été traitée surtout par des chercheurs tunisiens [10, 11] et saoudiens [12-17].

En 2003, Chérif et al. sans donner de détails sur leur base de données, précisent que les données concernent des pathologies reliées à la sphère ORL, à des problèmes neurologiques et ceux relatifs au larynx. Ils ont utilisé les segments voisés pour extraire le pitch, les trois premières valeurs des formants, le Jitter et le Shimmer. Ils ont conclu que les valeurs faibles du pitch peuvent indiquer la présence de dyslexie. Les fluctuations du pitch et des formants peuvent constituer une indication de la pathologie du larynx alors que des valeurs élevées du Jitter et Shimmer peuvent refléter des anomalies du conduit vocal, de la glotte ou du neuro-mécanisme [10]. Salhi et Shérif en 2006 [18], présentent une interface pour l'analyse des voix pathologiques sous Matlab en utilisant toujours la même base de données sans y donner aucun détails et les même paramètres acoustiques utilisés dans [10]. En 2008, Salhi et al. [11] ont proposé l'identification des voix pathologiques par les réseaux de neurones a multicouches MLP. La base de données et les paramètres acoustiques sont presque les même utilisés par les deux travaux que nous venons de citer. Pour évaluer le système de classification, ils ont utilisé

80 mots pour la phase d'apprentissage (40 mots normaux et 40 mots pathologiques) et pour la phase de test, 20 mots (10 normaux et 10 pathologiques). Les résultats étaient respectivement 90% et 80% d'identifications corrects pour la parole saine ou normale et pour la parole pathologique mais nous n'avons aucune indication sur le nombre de locuteurs, ni sur le matériau linguistique utilisé.

Une base de données plus importante et clairement décrite a été présentée en [12-14, 16, 17, 19, 20]. Les échantillons de parole ont été enregistrés sur des patients ayant des troubles de la voix qui ont fréquenté la clinique de la voix de l'hôpital de l'université du roi Abdulaziz (Arabie Saoudite) entre 2009 et 2010. Pour la parole normale, les 71 locuteurs étaient sans histoire antérieure ou actuelle avec des troubles de la voix. Tous les locuteurs étaient des Arabes natifs (53 hommes et 18 femmes) avec une tranche d'âge de 18 à 50 ans. Il a été demandé aux locuteurs de compter des chiffres arabes de 1 à 10 avec 10 répétitions. Pour la voix pathologique, un total de 71 locuteurs de six types différents de troubles de la voix, ont participé avec des échantillons de chiffres arabes. Pour chaque type de trouble de la voix, on trouve au moins 10 locuteurs. Les sept types de troubles des cordes vocales considérés étaient : les kystes, le RGO (Reflux Gastro-Oesophagien), la paralysie, polypes, dysphonie spasmodique (SD), Vergetures ou sillon vocales et nodules [21, 22].

Ghulam et al. en 2011 [16, 17, 19] proposent une reconnaissance et classification automatique de la parole pathologique suivant les types de pathologie en utilisant les formants des voyelles extraites des chiffres arabes. Aussi, ils proposent l'analyse des quatre premières valeurs des formants des voyelles extraites des chiffres arabes. Les méthodes de classification étaient la quantification vectorielle et les réseaux de neurones artificiels. Le meilleur résultat était 67.86% de correctes classifications pour la parole pathologique avec les réseaux de neurones. La reconnaissance automatique des chiffres arabes pathologique était de loin difficile avec les paramètres MFCC et donc ils insistent sur le besoin de trouver de nouveaux paramètres. Tandis que l'analyse des formants était prometteuse. Ghulam et al. (2012) [20] développent une méthode d'extraction de nouveaux paramètres pour la reconnaissance automatique. Les paramètres proposés étaient les distributions des segments voisés et non voisés, ainsi que les temps du Onset et Offset (Onset : temps d'attaque et Offset : le temps de terminaison) dans le domaine temps-fréquence pour détecter la pathologie de la voix. Ils étaient satisfaits des résultats (98 % de réussite). Ghulam et al., (2014) [23, 24] utilisent la base de

données pour l'anglais MEEI [25] pour proposer deux nouvelles méthodes pour la détection de la parole pathologique.

Alsulaiman et al. (2013) [14] présentent les paramètres RASTA-PLP (Spectral Transform relative-Perceptual Linear Predictive) pour la classification des différents types de troubles des cordes vocales. Ils précisent que les résultats étaient encourageants. Alsulaiman (2011 et 2014) [12, 13] ont exploité la totalité de la base de données saoudienne et insistent encore sur la fiabilité des paramètres RASTA-PLP et PLP pour la détection, la classification et la reconnaissance de la parole dysphonique.

Saudi et al. (2011) [26], ont mené le même travail que Alsulaiman; la reconnaissance de la parole dysphonique en utilisant les paramètres RASTA-PLP dans le cadre des HMM mais avec une base de données différente. La collection des données a été réalisée dans une salle acoustiquement isolée du département phoniatrie de l'hôpital Kobri Elkobba. Les échantillons acoustiques correspondent à la phonation tenue de la voyelle / ah / (1-3 s) de 35 patients (hommes et femmes) ayant des troubles des cordes vocales tels que kyste, polypes, nodules, paralysie, œdèmes et carcinome et aussi de la parole normale comme référence.

Une autre contribution sur la parole pathologique en langue arabe sur l'aphasie de Broca a été effectuée par Adam (2013) [27]. Cette étude représente la première tentative de comparer les délais d'établissement du voisement (ou Voice Onset Times : VOT) pour les consonnes /t/ et /d/ des aphasiques de Broca arabophones palestiniens et des personnes saines. Il faut noter que le Onset des voyelles /a/, /u/ et /i/ ont été extraits à partir de simples mots. Quatre locuteurs natifs arabes palestiniens masculins vivant en Cisjordanie ont participé à cette étude. Ils étaient âgés entre 45 et 70 (moyenne = 52 ans). Quatre sujets témoins, qui avaient une certaine correspondance de l'âge et du sexe avec les patients aphasiques de Broca, ont été considérés comme sujets contrôles. Les participants de contrôle n'avaient aucun historique de pathologie de la parole ou de l'audition.

En Algérie, nous pouvons citer les travaux de Maalam (2007) [28] relatifs à l'étude du signal parole normal et pathologique produit par des patients à audition déficiente (malentendants et sourds profonds) et ceux ayant subi une trachéotomie. Une étude qui porte sur l'analyse cepstrale, spectrale et poly-spectrale utilisant une approche récente qui se base sur les statistiques d'ordre supérieur (cumulant), pour extraire les deux principaux paramètres acoustiques qui sont : les formants et la fréquence fondamentale (pitch). Cependant, il n'y avait

pas suffisamment d'informations sur sa base de données collectée à Guelma ni sur le corpus utilisé même si il est mentionné que l'étude a été faite sur des voyelles tenues extraites à partir d'enregistrements de la parole pathologique et normale.

Une étude menée par Benselama et al. [29] concerne un programme de thérapie de la parole pathologique assistée par ordinateur, sur la base de modèles de Markov cachés et les réseaux de l'intelligence artificielle appliqué au stigmatisme en arabe. Malheureusement, aussi dans ce travail, nous ne trouvons pas beaucoup de détails sur les données utilisées comme le nombre de locuteurs, le corpus linguistique, la taille des données, ni sur la pathologie elle-même.

Ghelis et al. (2010) [30], proposent une modélisation statistique pour une classification dysphoniques. L'objectif de leur travail était de développer un système automatique pour évaluer la classification dysphonique arabe/française par la modélisation statistique des signaux de parole basée sur un modèle de mélange gaussien (GMM). Les patients ont été choisis de l'Université d'Annaba : centre hospitalier (CHU) dans le service ORL à la présence d'un groupe composé de 8 spécialistes médicaux. Chaque locuteur a enregistré une session variant de 55 secondes à trois minutes et quelques secondes. Tous les intervenants étaient âgés entre 35 et 60 ans. Les sujets ont été assignés à lire un paragraphe en langue arabe d'un texte sur "l'équitation" et un paragraphe en français du texte "la chèvre de Monsieur Seguin» (d'Alphonse Daudet). Ces informations sur la base de données restent toujours insuffisantes.

En 2012, Daoud [31] et Hemri [32], présentent deux thèses de magister sous la direction de Guerti Mhania. Daoud traite la pathologie du langage oral chez un locuteur arabophone : cas de la substitution des phonèmes. Il a utilisé la technique LPC (Linear Predictive Coding) comme technique d'analyse de la parole et les réseaux de neurones multi-couches MLP (Multi Layer Perceptrons) comme technique de reconnaissance. Le système donne un taux de reconnaissance du phonème cible et ce taux aussi représente le degré de la réussite de la prononciation du patient pour le phonème concerné. Tandis que Hemri présente une étude sur les troubles d'articulé dentaire chez un locuteur arabophone. Le but de ce travail était la correction des problèmes causés par le port de la prothèse dentaire lors de la prononciation des 10 phonèmes dentaires de la langue arabe. Pour atteindre cet objectif, ils ont élaboré un corpus de mots contenant tous ces phonèmes dans les différentes positions qui existent en Arabe (initiale, médiale et finale). Il a calculé le taux de reconnaissance des phonèmes dentaires avec l'application des réseaux de neurones multicouches (Multi Layer Perceptrons : MLP). Là aussi, nous ne trouvons pas beaucoup de détails sur les bases de données collectées.

En 2013, nous avons recensé quatre études. Ferrat et Guerti [33], proposent une étude des sons produits par des locuteurs de l'œsophage algériens. Un corpus de parole a été recueilli d'Octobre 2008 à Septembre 2009. Les participants étaient des patients du sexe masculin qui ont subi une laryngectomie totale. Âge minimum des patients était de 47 ans et l'âge maximum était de 59 ans. Les enregistrements ont été effectués avant le commencement de la rééducation et après trois, six et onze mois en utilisant la voix œsophagienne. L'analyse acoustique comprend l'emplacement F (Hz), formants, l'intensité, la gigue (%), Shimmer (dB), harmonique sur bruit HNR (dB), et le degré de trames non voisées DUF (%). A.Mansour [34] effectue une analyse acoustique multiparamétrique des dysphonies qui résultent d'une lésion organique liée aux cordes vocales non pas aux troubles de langage. Dans cette étude aussi, nous n'avons pas d'informations sur la base de données construite. Amara et Fezari effectuent deux études : la classification de la parole pathologique en utilisant les GMM et les SVM et la détection de la parole pathologique par la technique de l'identification automatique de locuteurs. En fait, Ils ont travaillé sur une base de donné allemande qui contient de nombreuses pathologies [35, 36]. Fezari et al. [37] encore dans une étude récente, il a exploité la même base de données allemande [38] pour une analyse acoustique en utilisant des paramètres adaptatifs et classifieurs pour la détection des troubles de la voix. Fezari et al., construisent un système d'analyse acoustique de la voix, avec des techniques de reconnaissance automatique des locuteurs, en se basant sur les MFCC et la variation en fréquences et amplitudes des paramètres (Jitter et Shimmer) et les GMM comme classifieur.

Après ce bref aperçu sur le traitement automatique de la parole pathologique arabe, nous remarquons évidemment, en premier lieu que les bases de données sont très limitées (de stimuli : chiffres ou phonation tenues des voyelles ou de consonnes, nombre réduit de participants), dispersée, hétérogènes et construites pour des objectifs ponctuels. Nous avons relevé également que la dysphonie est la seule pathologie traitée automatiquement.

1.2. Objectifs et contributions

Compte tenu de l'ampleur du domaine, ce travail considère un objectif global est celui du diagnostic de la pathologie par le biais d'une analyse acoustique-prosodique suivi par un module de classification ou de reconnaissance mais dans des conditions limitées. Effectivement, le diagnostic de différentes pathologies de la parole à partir de la seule analyse perceptive de la parole est difficile. Il reste subjectif. Chercheurs et cliniciens ont besoin d'un instrument simple et fiable pour l'évaluation objective des troubles de la parole.

Ainsi, le développement de mesures automatiques des événements prosodiques représente un objectif important dans notre travail. En fait, nous nous intéressons surtout aux paramètres rythmiques proposés récemment à l'identification des langues par Ramus et Grabe [39-41].

En effet, Nos contributions sont :

- Une contribution à l'étude de la variation rythmique dans la parole pathologique. Notre hypothèse repose sur le fait que les éléments prosodiques tels que le rythme sont des indices pertinents permettant la discrimination de la parole pathologique de la parole normale et aussi la discrimination des différents niveaux de sévérité de la pathologie[42-47].
- Une importante contribution consiste également en la construction d'une base de données pour les aphasiques algériens désignée par l'abréviation ADAD (Algerian Dialectal Aphasic DataBase). Effectivement, La réalité pratique montre qu'en l'absence de corpus linguistique adéquat, les méthodes les plus performantes de traitement du signal parole restent insuffisantes pour caractériser efficacement du point de vue acoustique et phonétique la parole pathologique. C'est dans l'objectif de constituer des ressources linguistiques qui pourraient servir à de nombreuses applications (suivi de l'évolution des patients, système d'assistance vocale, etc.) que nous avons effectué une campagne d'enregistrement auprès des patients aphasiques des régions d'Annaba et de Sétif situées à l'est de l'Algérie. Le corpus linguistique utilisé est issu du bilan de langage structuré, élaboré par Zellal [48-51]. Le protocole d'examen est applicable à des aphasiques algériens et respecte rigoureusement les critères d'élaboration et de sélection adoptés à l'origine par Ribaucourt [25]. Le corpus était encore adapté pour les aphasiques de l'est algérien à l'aide de l'orthophoniste d'Annaba qui a collaboré pour l'enregistrement de la base de données[52].
- Nous avons proposé un nouveau système de classification de la parole pathologique et les différents niveaux de sévérité de la dysarthrie de l'anglais américaine de la base de données Nemours. Nous avons présenté un système hybride utilisant une approche connexionniste et bayésienne. L'objectif est de procéder à une évaluation automatique des niveaux de sévérités de la dysarthrie. Les entrées sont composées des MFCC classiques et les paramètres rythmiques [45, 46]. Enfin, nous avons proposé un système de reconnaissance de la parole dysarthrique par les HMM qui bénéficie des résultats de la

classification des niveaux de sévérités issus du système hybrides des réseaux de neurones et la théorie bayésienne.

- La réalisation de l'aphasiogramme automatisé était aussi un objectif et une contribution prioritaire. L'aphasiogramme est un outil à l'usage des orthophonistes pour le diagnostic et l'évaluation de la progression de l'état et la thérapie du patient aphasique. L'aphasiogramme est une courbe qui illustre les différents pourcentages de prononciations correctes des différentes taches de répétitions de mots simples ou de phrases complexes. Donc, afin d'automatiser l'aphasiogramme, nous avons réalisé un système de reconnaissance de la parole continue en utilisant une sorte de normalisation phonétique de la parole pathologique et la parole normale. En utilisant les techniques des HMM et des techniques de l'adaptation des locuteurs MAP (**Ma**ximum **A** **P**osteriori) et MLLR (**M**aximum **L**ikelihood **L**inear **R**egression) pour avoir de meilleurs résultats pour la reconnaissance de la pathologique aphasique

1.3. Organisation de la thèse

Enfin, Cette thèse se compose de six chapitres : une introduction et une conclusion générale et quatre chapitres. Ces derniers présentent le cadre théorique et méthodologique de notre travail et aussi présentent l'analyse expérimentale de la variation rythmique de la parole pathologique. Étant donné que notre travail porte sur deux bases de données pathologiques avec deux pathologies : l'aphasie et la dysarthrie, il nous a paru utile de présenter brièvement dans le deuxième chapitre les deux pathologies tout en décrivant les deux bases de données.

Cette première étape de notre travail est suivie par le chapitre 3 consacré à l'étude du rythme et à l'extraction de différents paramètres acoustiques et prosodiques. Dans cette partie nous décrivons spécialement les paramètres de rythme en plus les paramètres prosodiques (le pitch, les formants, la durée, etc.) et les paramètres spectraux : (MFCC). Cette étude se fonde sur le fait que le rythme est considéré comme paramètre pertinent pour l'établissement d'une discrimination rythmique de la parole pathologique et de la parole normale. Une analyse statistique relative aux paramètres rythmiques appliquée sur la parole dysarthrique. L'approche statistique est présente aussi avec les méthodes paramétriques et non paramétriques : d'ANOVA ou de Kruskal-Wallis pour déterminer la signifiance des paramètres rythmiques. L'analyse DFA ou la régression logistique comme méthodes de classification et analyse statistique déterminant les paramètres de rythme les plus pertinents. Leur résultats sont exposés et discutés.

Les chapitres 4 et 5 sont consacrés à l'évaluation objective de la dysarthrie et l'aphasie respectivement.

Dans le chapitre 4, nous avons passés en revue la théorie des SVM, les réseaux de neurones et les mélange des lois gaussiennes. Nous nous sommes intéressés surtout à la classification parole pathologique/parole normale et aussi à l'intelligibilité de la parole dysarthrique en classifiant les différents niveaux de sévérité tantôt par les SVM et tantôt par les réseaux de neurones. À la fin nous proposons un système de reconnaissance de la parole dysarthrique par les HMM qui bénéficie des résultats de la classification des niveaux de sévérités issus du système hybrides des réseaux de neurones et de la théorie bayésienne.

Le cinquième chapitre se concentre sur la reconnaissance de la parole continue pathologique aphasique. Tout d'abord un système de référence était construit sous la plateforme HTK. Ce dernier a été transformé par les techniques d'adaptation aux locuteurs les plus connues : la régression linéaire du maximum de vraisemblance (Maximum Likelihood Linear Regression : MLLR) et Maximum A Posteriori (Maximum A-Posteriori : MAP) pour avoir une meilleure performance. Notons que sur le plan pratique, nous automatisons l'aphasiogramme pour les tâches de répétition des syllabes, logatomes, mots simples et complexes et à la fin les phrases.

Enfin, ce travail s'achève par une conclusion générale et une discussion de nos résultats obtenus par le biais d'expériences d'identification automatique de la parole pathologique et la classification de niveaux de sévérité. Quelques perspectives pour des travaux futurs sont énoncées pour une éventuelle continuation de la recherche dans le domaine de la parole pathologique. Une annexe importante contient les résultats non présents à travers les chapitres de la thèse. Aussi, cette annexe expose les différents matériaux linguistiques utilisés dans nos expérimentations.

Chapitre 2 :

Revue de la parole pathologique

Chapitre 2 : Revue de La Parole Pathologique

2.1. Introduction

La réussite d'un traitement automatique de la parole normale ou pathologique dépend principalement d'une base de données (BD) fiable. En effet, la BD tient un rôle primordial dans l'élaboration d'une stratégie du traitement automatique de la parole. Cependant, une BD est difficile à élaborer surtout quand il s'agit de la parole pathologique.

C'est dans l'objectif de constituer des ressources linguistiques qui pourraient servir à de nombreuses applications (suivi de l'évolution des patients, système d'assistance vocale, etc.) que nous avons effectué des enregistrements auprès des patients aphasiques des deux villes d'Annaba et de Sétif situées à l'est de l'Algérie.

Ce chapitre propose une brève présentation des deux pathologies considérées : Aphasie de Broca et dysarthrie. Avec un exposé plus ou moins détaillé des deux matériaux linguistiques des patients algériens et américains des deux bases de données utilisées dans les différentes expérimentations à des fins de classification, reconnaissance et discrimination de la parole pathologique de la parole saine

2.2. L'Aphasie (The Aphasia)

2.2.1. Définition de l'aphasie

L'aphasie est une déficience du langage, affectant la production ou la compréhension de la parole et la capacité de lire ou d'écrire. Souvent l'aphasique n'arrive plus à nommer des objets, ne retrouve plus les noms des personnes qu'il connaît ; il se peut même qu'il ne puisse répondre clairement par oui ou non. L'aphasie résultant de lésions cérébrales peut également provenir de traumatisme crânien, des tumeurs cérébrales, ou d'infections [53].

L'aphasie est d'abord un trouble du langage auquel s'ajoutent souvent des difficultés de parole. Les spécialistes du langage font une différence entre la parole et le langage ; si un individu éprouve des difficultés d'articulation, de prononciation nous dirons qu'il a un trouble de la parole ; et si, il éprouve des difficultés à choisir ses mots, à les combiner pour faire des phrases ou encore à comprendre, nous dirons plutôt qu'il a un problème de langage. Les **troubles du langage** comportent une atteinte du langage spontané et une fluence verbale

effondrée. La fluence verbale est le nombre de mots émis par minute en parlant spontanément ou en décrivant une scène imagée. Ce nombre est d'environ 90 mots par minute pour un individu normal [54].

Le cerveau se compose de deux parties, appelées hémisphères. Chaque hémisphère contrôle diverses activités. Pour certaines d'entre elles, la participation des deux hémisphères est importante. Le contrôle du mouvement par le cerveau se fait de manière croisée, c'est-à-dire que la partie gauche du cerveau contrôle le bras droit et la jambe droite tandis que la partie droite du cerveau contrôle le côté gauche du corps. Donc, l'aphasie est une pathologie du système nerveux central, due à une lésion d'une aire cérébrale [53]. Le mot « aphasie » vient du grec et signifie « perte de la parole ». Ce terme a été créé pour la première fois par Armand Trousseau en 1864 [55]. Depuis cette époque, le mot a pris du sens, en désignant un trouble du langage affectant l'expression ou la compréhension du langage parlé ou écrit survenant en dehors de tout déficit sensoriel ou de dysfonctionnement de l'appareil phonatoire [53, 56].

2.2.2. Les causes d'une aphasie

En générale sont :

- Un accident vasculaire cérébral.
- Une hémorragie cérébrale.
- Un traumatisme cranio-cérébral (lors d'accident de la route, d'une chute).

2.2.3. Principaux types d'aphasie

Les différents types d'aphasie résultent d'atteintes localisées dans des régions particulières du cerveau. Des lésions situées dans la région frontale antérieure du cerveau, qu'on appelle aire de Broca, interfèrent sur la production de la parole. Des dommages dans une aire du cortex temporo-pariétal, dite aire de Wernicke, affectent la compréhension du langage. Des lésions dans la circonvolution supra-marginale perturbent la répétition des paroles entendues. Pour la plupart des individus, ces régions fonctionnelles ne se trouvent que dans l'hémisphère gauche (voir figure 2.1).

2.2.3.1. L'Aphasie de Broca et son aspect cérébrale

Cette aphasie a été mise en évidence par Paul Broca, un chirurgien français, en 1861 [57]. Cette pathologie fait suite à des lésions dans l'aire de Broca, l'aire cérébrale chargée de la mise en place des schémas corporels moteurs du langage. Cette aire est située sur la partie

gauche du cerveau, dans le lobe frontal gauche chez les droitiers et sur l'hémisphère droit chez le gaucher.

Ce type d'aphasie est caractérisé par des troubles oraux et écrits alors que la compréhension est à peu près bonne. L'aphasique de Broca présente généralement des problèmes d'articulation (prononcer les mots et les sons) à des degrés divers, et utilise des phrases qui ne sont pas structurées (l'**agrammatisme**). L'aphasique a du mal à trouver les mots exacts pour s'exprimer, et les mots utilisés ne sont pas adaptés. Par conséquent, le sujet est alors incapable de formuler oralement ses idées alors que celles-ci sont intactes dans son esprit. Donc, il y a un manque de mots, des stéréotypies (le patient utilise un élément linguistique, répétitif, prononcé à tout propos), des troubles arthritiques et une **agraphie apraxique** (des difficultés concernant l'écriture) et une hémiplégie ou une hémiparasite droite associée.

En générale, l'aphasie de Broca se caractérise en particulier par une élocution toujours laborieuse et souvent **dysprosodique (prosodique :** prononciation correcte et régulière des mots selon l'accent et la quantité des syllabes), qui désigne les difficultés de contrôler son intonation tels que la hauteur et l'intensité du son. Ceci aboutit à un langage peu mélodique faisant place à des accentuations dans la phrase. Autrement dit la tonalité de la phrase est presque complètement perturbée [58].

2.2.3.2. L'Aphasie de Wernicke et son aspect cérébrale

Cette aphasie est la conséquence de lésions qui affectent la **zone de Wernicke de l'hémisphère dominant** sur l'hémisphère gauche chez le droitier et sur l'hémisphère droit chez le gaucher [59] (voir Figure 2.1).

L'aphasie de Wernicke est caractérisée par une expression logorrhéique (Jargon abondant). Le débit ou la fluence est normale ou élevé et la compréhension est perturbée, ce qui entraîne une communication difficile. Le **jargon** est qualifié de **sémantique** ou de **phonémique** selon le type de paraphasie qui le constitue ou qui prédomine. Les transformations de type phonémique comportent des omissions, des ajouts, des déplacements, et des persévérations. Les **troubles de langage (paraphasies) de type sémantique** comportent une substitution de mots souvent proches par sa signification du mot substitué. **Paraphasies phonémiques** c'est-à-dire la modification des phonèmes qui constituent les

mots avec au maximum la création de **néologismes** (termes inventés par le patient) ou des **paraphasies sémantiques** autrement dit le remplacement d'un mot par un autre et dont la signification peut-être proche [58].

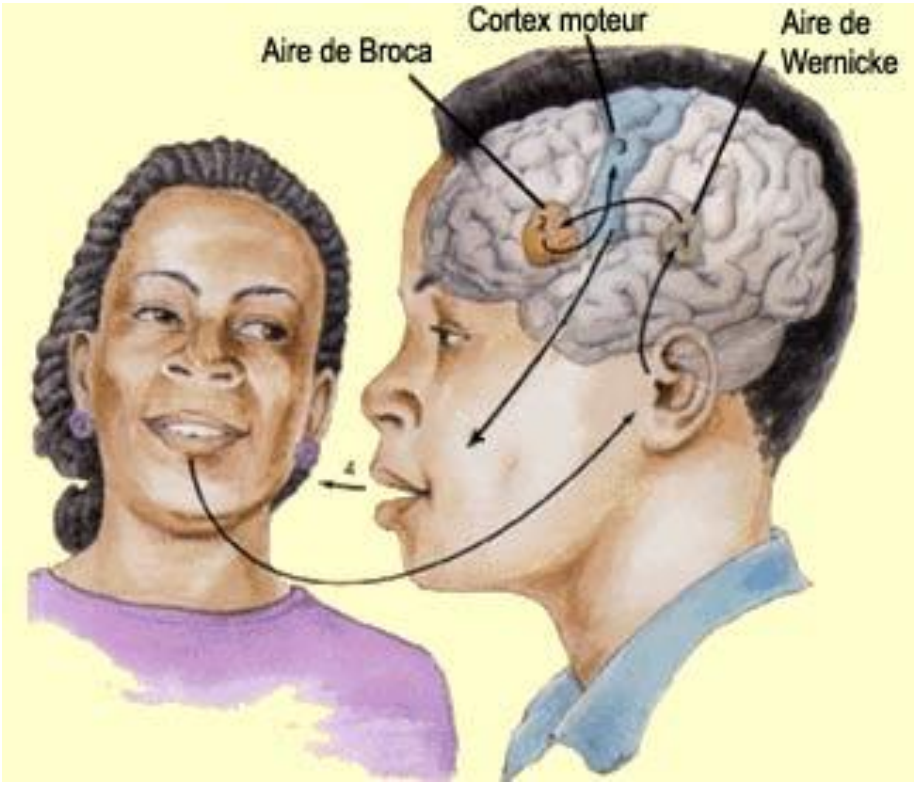

Figure2. 1: L'aire de Broca et de Wernicke dans le cerveau [53].

2.2.4. Pathologies visées

Comme nous nous intéressons qu'à l'aphasie de Broca, nous citons les différentes pathologies rencontrées chez les patients aphasiques participants à la base de données que nous avons construit (ADAD : Algerian Dialectal Aphasic DataBase) [54, 60, 61].

2.2.4.1. Le manque du mot

L'aphasique éprouve de la difficulté à trouver ses mots. Souvent, il n'arrive pas à trouver le mot désiré, il laisse sa phrase inachevée. Il ne s'agit pas d'un trouble de la mémoire mais d'une difficulté à trouver le mot au bon moment. D'ailleurs, le mot cherché peut parfois être produit sans problème, quelque temps après, dans une autre situation.

2.2.4.2. La réduction de l'expression

L'aphasique ne parle pas beaucoup. Il répond souvent aux questions par un « oui » ou un « non », il cherche ses mots et n'arrive pas toujours à faire des phrases. De même qu'il éprouve des difficultés à s'exprimer, il peut aussi éprouver des difficultés à écrire. Parfois, il lui est tout juste possible d'écrire son nom ou de recopier quelques lettres.

2.2.4.3. Le trouble arthrique

La prononciation des sons est anormale ; elle peut être floue ou trop tonique. Habituellement, la personne qui a un trouble arthrique parle plus lentement. Il peut être parfois difficile de la comprendre parce que les sons ne sont pas articulés clairement ou encore ils sont déformés.

2.2.4.4. Le jargon

On parle de jargon ou " jargonaphasie " quand l'aphasique déforme et confond les mots ou même en invente de nouveaux à tel point qu'il devient impossible de le comprendre.

2.2.4.5. Déficience de mots justes

Ce trouble est indépendant des difficultés articulatoires, mais celles-ci peuvent toutefois rendre compte de délais dans la production des mots. La facilitation par ébauche orale permet généralement la production du mot qui fait défaut. Ce trouble est mis en évidence lors de la dénomination d'objets. Le sujet ne peut pas non plus exprimer ce qu'il a sur « le bout de la langue », et utilise des périphrases ou fait des paraphasies verbales.

2.2.4.6. Stéréotypie

Ce trouble se manifeste par une répétition ininterrompue et inappropriée de sons, de mots, des expressions ;

2.2.4.7. Persévérations

Il s’agit d’un trouble qui se manifeste par des mots, des phrases ou des séries qui viennent d'être utilisés, et qui reviennent hors de propos perturber le discours présent.

2.2.4.8. Désintégration phonétique

Ce trouble est souvent associé à un trouble plus général de l’articulation bucco-linguo-faciale appelé apraxie bucco-facial, qui se manifeste par des difficultés voire une incapacité à

exécuter des mouvements articulatoires complexes comme gonfler les joues, siffler, mettre les lèvres en avant, montrer les dents, ou claquer la langue.

2.2.5. Traitement de L'aphasie

Le traitement de l'aphasie dépend de plusieurs facteurs tels que :

- Age de la personne aphasique
- Cause de la lésion cérébrale
- Type de l'aphasie
- La position et la sévérité de la lésion cérébrale

La chirurgie peut parfois améliorer l'état de l'aphasique. Mais, dans la plupart des cas, cependant, la thérapie de la langue devrait commencer dès que possible et être adaptés aux besoins individuels du patient. Réhabilitation d'un orthophoniste implique des exercices vastes dans lesquels les patients lisent, écrivent, suivent les directions, et répètent ce qu'ils entendent. La thérapie assistée par ordinateur peut compléter la thérapie de langue standard. En général, l'orthophoniste est conscient du fait que le cerveau possède une certaine plasticité. Évidemment, un patient atteint d'une aphasie de Broca causée par une lésion de l'hémisphère gauche peut recouvrir des capacités d'expression orale à travers le chant ; la mélodie et la musique qui sont traitées par l'hémisphère droit. À travers une méthode de rééducation appelée la thérapie mélodique et rythmée (ou melodic intonation therapy), le patient va essayer d'utiliser son hémisphère droit pour dire des phrases avec une intonation musicale.

2.2.6. La base de données aphasique ADAD : Algerian Dialectal Aphasic DataBase

Les ressources linguistiques dialectales arabes sont très limitées. Il existe quelque uns pour le dialecte égyptien et le dialecte levantin (Liban, Jordanie, Palestine, Syrie). Ces bases de données sont en général des enregistrements à travers la ligne téléphonique [62, 63]. Concernant la parole pathologique, jusqu'à ce jour, à notre connaissance, ses ressources linguistiques n`existent pas du tout ou si elles existent, elles sont sous formes d'enregistrements de quelques chiffres ou de quelques mots servant à valider des besoins d'une étude. En générale, les ressources linguistiques pathologiques soit pour le MSA ou dialectales sont très rares et ambiguës. Elles manquent également de clarté et de précision comme c'est précisé dans l'état de l'art au premier chapitre [11, 17, 18, 20, 24, 26, 32, 33, 64].

2.2.6.1. Le contexte linguistique arabe et algérien

La langue arabe est parlée dans vingt-quatre pays. Elle constitue la quatrième langue des Nations Unies. Sous ses formes dialectales, l'arabe est la langue maternelle de plus de deux cent cinquante millions de locuteurs. Parlée aussi par quelques minorités dans Presque 33 pays du monde ainsi que dans les pays à forte population musulmane [65].

La langue arabe représente un arabe classique, un arabe standard moderne (ou, en anglais : Modern Standard Arabic : MSA) et un arabe médian (autrement appelé neo-arabe) se situe entre le MSA et l`arabe dialectal. L'arabe médian s'est développé parmi la communauté intellectuelle arabophone. Cette variété, décrite à la fois comme une variante simplifiée de l'arabe littéral moderne et une forme élevée de l'arabe dialectal, atteste la syntaxe et la morphologie du dialecte et un lexique mixte constitué de mots empruntés au dialecte et à l'arabe standard [66].

L`arabe dialectal est une collection de dialectes arabes connexes au MSA, qui sont principalement parlés et qui présentent une importante diversité phonologique, morphologiques, syntaxiques et lexicales entre eux et par rapport à l'arabe standard moderne écrite qui est une version simplifiée de l'arabe classique.

L'arabe classique ou l'arabe coranique n'est plus que la langue du patrimoine culturel passé avec ses œuvres classiques et son livre sacré : le Coran. L'arabe classique est appris dans les établissements d'enseignement à travers la littérature arabe classique et les cours de théologie [65].

L'arabe « fousha » ou communément identifié comme «arabe standard moderne» MSA. Comme c'est entendu, cette langue est moins formelle et ne couvre pas toutes les fonctionnalités de l'arabe classique. C'est la langue de la presse, des médias, de la littérature moderne, des conférences et des discours politiques [67].

Concernant le contexte algérien, Zellal écrit dans [68] « Il est à considérer que la population d'aphasique en Algérie est majoritairement jeune : depuis 20 ans ; nous avons observé 80% de cas âgés entre 18 et 45 ans, 15% de cas âgés entre 45 et 60 ans, 5% de cas âgés entre 60 et 73 ans. Ceci pose évidemment le problème de la nature du bagage linguistique du locuteur algérien en tant que déterminé, depuis l'indépendance, par des

facteurs d'ordre historique : degré d'alphabétisation, scolarisation, provenance par région, etc… ».

La complexité linguistique en Algérie émane du fait que plusieurs langues se partagent le marché linguistique en créant des conflits parfois latents et souvent larvés depuis la plus haute histoire du pays. Cette situation qualifie le marché linguistique d'un champ porteur de tensions le plus souvent idéologiques [69] .

La situation linguistique en Algérie est marquée par la diversité et la coexistence de langues différentes. Ces langues sont les langues nationales, la langue berbère avec ses différentes variétés, l'arabe dialectal et de l'autre côté, les langues étrangères représentées essentiellement par le français et l'espagnol.

En Algérie, La langue arabe se présente sous deux formes : une langue à tradition religieuse, ce qui lui fait acquérir un statut de prestige (l'arabe classique). D'autre part, l'arabe littéraire reste essentiellement écrite et ne constitue aucunement la langue à usage quotidien et spontané (MSA) ; et l'arabe dialectal appelé aussi arabe algérien servant la communication orale et constitue la langue maternelle de la quasi-totalité des algériens.

La langue berbère ou amazighe est une langue parlée par des communautés importantes. En effet, on trouve le Chaoui dans les Aurès ; cela veut dire dans les villes du constantinois comme Oum El Bouaghi, Batna, Sétif. Aussi dans les villes de Tizi Ouzou, Bejaia Borj Bouaririj connus sous le nom de Kabylie. Le mozabite et targuie sont utilisées dans le Mzab et le massif du Hoggar dans le sud algérien. Le chleuh parlé dans l'Oranie.

À côté des langues nationales, nous trouvons le français et un peu moins l'espagnol. Le français est considéré officiellement comme première langue étrangère. Le français reste en Algérie une langue de prestige qui permet l'accès aux nouvelles technologies. Elle est donc la langue de la modernité et est un instrument de recherche qui sert de moyen privilégié partageant avec l'arabe différents domaines [69].

Cette situation présente de grands problèmes difficiles à surmonter pour les technologies langagières arabes et y compris la création des corpus et des bases de données pour les systèmes de synthèse ou pour les systèmes de reconnaissance de la parole arabe. Une

part importante de la variation linguistique se produit et produit de nombreuses variantes de formes qui sont difficiles à identifier et à regrouper.

La plupart de données acoustiques arabes disponibles employées par la communauté traitement automatique de la parole est celle du MSA ; la parole lecture bien planifiée enregistrées dans un environnement sans bruit de fond extraite des émissions d'informations (new broadcasts).

2.2.6.2. Le système de transcription du MSA

L'arabe standard moderne est une variété intermédiaire entre l'arabe classique et l'arabe dialectale, écrite et parlée. Il existe en arabe 3 syllabes sous-jacentes : CV, CVC et CVV et deux syllabes : CVVC et CVCC. Il n'est donc pas permis d'avoir en arabe deux consonnes identiques, ou un cluster consonantique à l'initiale : CCV.

L'alphabet arabe comprend vingt-neuf lettres fondamentales (y compris elhamza). Les consonnes sont sourdes, sonores, emphatiques ou nasales. Voyelles et consonnes peuvent se présenter sous forme de géminées.

Dans les éditions du Coran où les ouvrages didactiques, on utilise une notation vocalique plus ou moins précise sous forme de diacritiques. Il existe, de plus, dans de tels textes, dits « vocalisés », une série d'autres diacritiques de syllabation dont les plus courants sont l'indication de l'absence de voyelle *(sukūn)* et la gémination des consonnes : *chadda* ; Les consonnes doubles se répartissent entre 2 syllabes distinctes et doivent être prononcées séparément comme par exemple dans le nom Muhammad qui est prononcé comme Mu/ham/mad.

Les consonnes sont classées selon leur mode d'articulation (occlusif, fricatif, nasal, glissant ou liquide), leur lieu d'articulation (labial, dental ou velo-palatal) et leur voisement (sonore, ou sourd). Le tableau 2.1 récapitule ces modes d'articulation. Évidemment, nous avons trois colonnes pour le codage phonétique : translittération, le code API et le code SAMPA. le code Sampa est un acronyme de "Speech Assessment Methods Phonetic Alphabet", est un jeu de caractères phonétiques utilisable sur ordinateur utilisant les caractères ASCII 7-bits imprimables basé sur l'Alphabet phonétique international (API) [70]. L'alphabet phonétique international (**API**) est un ensemble de symboles basé sur la traduction d'une langue au moyen des sons qu'elle émet à l'oral. Aujourd'hui, l'alphabet phonétique

international est composé de 107 lettres, 52 signes diacritiques et 4 caractères de prosodie. La **translittération** est l'opération qui consiste à substituer à chaque graphème d'un système d'écriture un graphème ou un groupe de graphèmes d'un autre système, indépendamment de la prononciation. Elle dépend donc du système d'écriture cible, mais pas de la langue [53] .

Tableau 2. 1: Le système consonantique de l'arabe standard moderne

Lettre en arabe	Nom Arabe	Translittération	Code API	Code Sampa	Description
ء	همزة	2	ʔ	?	Occlusive glottale sourde
I	ألف	A/â		a :	Voyelle ouverte antérieure non arrondie
ب	باء	b	b	b	Bilabiale occlusive voisée
ت	تاء	t	t	t	Alvéolaire occlusive sourde
ث	ثاء	th	θ	T	Dentale fricative sourde
ج	جيم	j	ʤ	g	post-alvéolaire affriquée voisée
ح	حاء	7	ħ	X/	Pharyngale fricative sourde
خ	خاء	5	x	x	fricative vélaire sourde
د	دال	d	d	d	occlusive alvéolaire voisée
ذ	ذال	dh	dh	D	fricative dentale voisée
ر	راء	r	r	r	roulée fricative post-alvéolaire sourde
ز	زاي	z	z	z	fricative alvéolaire voisée
س	سين	s	s	s	fricative alvéolaire sourde
ش	شين	ch	ʃ	S	fricative post-alvéolaire sourde
ص	صاد	S	sˤ	s`	**emphatique** Fricative alvéolaire sourde
ض	ضاد	D	dˤ, ðˤ	d`	**emphatique** Plosive Intermentale sonore
ط	طاء	6/T	tˤ	tˤ	**emphatique** occlusive sourde apico-
ظ	ظاء	Z	zˤ, ðˤ	D`	**Emphatique** fricative dentale voisée
ع	عين	3/’	ʔˤ	?’	occlusive glottale emphatique
غ	غين	r/gh	ɣ	G	Consonne fricative vélaire voisée
ف	فاء	f	f	f	fricative labiodentale sourde
ق	قاف	9/q	q	q	occlusive uvulaire sourde
ك	كاف	k	k	k	occlusive vélaire sourde
ل	لام	l	l	l	spirante latérale alvéolaire voisée
م	ميم	m	m	m	occlusive nasale bilabiale voisée
ن	نون	n	n	n	occlusive nasale alvéolaire voisée
ه	هاء	h	h	h	fricative glottale sourde

Les voyelles sont fondamentalement au nombre de 6. Trois **brèves** (Signes diacritique) et trois **longues** (valeur de 2 brèves) : les voyelles sont caractérisées par la vibration des cordes vocales. Elles sont représentées dans le tableau suivant :

Tableau 2. 2 : le système de voyelles pour l'arabe standard

Lettre en arabe	Nom en arabe	Code Sampa	Description
ي	Elyaa	i:	longue : ou y
ا ou ى	*Alif ou alif maqṣūra*	a :	Longue : Respectivement au milieu et en fin du mot
و	Waw	u :	Longue : u : ou w
ـِ	Fatha	i	Brève : est un trait au-dessus de la lettre pour remplacer le 'a' comme dans ba = بَ
ـَ	Kassra	a	Brève : est un trait en-dessous de la lettre pour remplacer le 'i' comme dans bi = بِ
ـُ	Damma	u	Brève : est un petit و (Waw) au-dessus de la lettre pour remplacer le 'ou' comme dans bou = بُ

Il y a aussi les diphtongues qui sont deux : ay et aw dites aussi semi voyelles [71-74].

2.2.6.3. Le système de transcription phonétique du dialecte algérien.

Le dialecte arabe algérien ou l'arabe algérien est utilisé par la quasi-totalité des algériens. Il comporte une importante base lexicale issue de tamazight et du français (issue même du turc, espagnol...). Chaque variété régionale de l'arabe algérien s'identifie par référence à une grande ville.

Le dialecte arabe en général est une langue parlée non écrite. Il n'y a pas de règles d'écriture pour les dialectes arabes ni aucune orthographe. En conséquence, une transcription adéquate pour la parole dialectale pour une modélisation acoustique reste très couteuse.

La structure syllabique en dialecte magrébin et en Algérie spécialement est différente de celle du MSA. La réduction du système vocalique due à la disparition des voyelles brèves en syllabes ouvertes (qui se termine par une voyelle) est responsable de la formation de clusters consonantiques dans la chaîne phonique. Celle de l'arabe maghrébin permet plus d'une consonne dans les positions adjacentes au noyau syllabique. Certains phonèmes existent dans l'arabe algérien, mais non pas dans MSA. Par conséquent, l'ensemble de phonèmes se compose de 41 phonèmes, dont 29 consonnes et 12 voyelles.

Nous avons utilisé le code phonétique utilisé par N. Zellal dans [75]. Nous avons rajouté quelques phonèmes existants dans l'est algérien ; la gémination (chadda) comme signe diacritique et les deux consonnes emphatique th (**ذ**) et dẖ avec le point dessus comme signe diacritique d'emphase de la consonne th. Enfaite, dẖ représente les deux consonnes

(ظ/ض) mais qui sont vraiment confondues dans notre parler surtout dans la ville de Sétif et les villes qui l'entourent.

Tableau 2. 3: le système de transcription pour les consonnes d'après Zellal [75]

Sym	Sym arabe	Mot-clé dialectal	Sig en français	Description phonétique
[P]		[Plaas]	Place	Bilabiale sourde
[b]	ب	[bè:ba]	Papa	Bilabiale sonore
[v]		[vi:lo]	Vélo	Labiodentale sonore
[m]	م	[mu:s]	Couteau	Bilabiale nasale
[f]	ف	[fi:l]	Éléphant	Labiodentale sourde
[t]	ت	[ta3], [tamr]	De, dattes	Apico-dentale sourde non emphatique
[ṭ]	ط	[ṭạ:qa]	Fenêtre	Apico-dentale sourde
[d]	د	[denya]	Vie	Apico-dentale sourde non emphatique
[ḍ]		[ḍo:r]	Tourne-toi	Apico-dentale sonore emphatique
[r]	ر	[ri:x]	Vent	Vibrante non emphatique
[ṛ]		[ṛạ:ħ]	Il est parti	Vibrante emphatique
[s]	س	[senna]	Dent	Sifflante sourde non emphatique
[ṣ]	ص	[ṣghéér]	Petit	Sifflante emphatique
[z]	ز	[ziin]	Beauté	Sifflante sonore
[l]	ل	[li:l]	Nuit	liquide
[n]	ن	[ni:f]	Nez	Apico-dentale nasale
[Ɵ]	ث	[Ɵe3leb]	Chacal	Inter dentale sourde
[ch]	ش	[chu:f]	Vois	Chuintante
[dj]		[djbel]	Montagne	Pré-palatale affriquée sonore
[j]	ج	[jbel][1]	Montagne	Fricative
[tch]		[tchu:f]	Tu vois	Pré-palatale affriquée sourde
[k]	ك	[kursi]	Chaise	Post dorsal post palatale sourde
[g]	-	[gãtra]	Pont	Post dorsal post palatale sonore
[x]	خ	xé:t_]	Fil	Post dorsal post vélaire sourde
[R]	غ	[Rạ:r]	Trou	Postdorso post vélaire sonore
[ħ]	ح	[ħè:l]	Temps	Pharyngale sourde
[3]	ع	[3ạjn]	œil	Pharyngale sonore
[h]	ه	[hu:wwa]	Il	constrictive
[µ]	ء	[sµel]	Il a demandé	Laryngale occlusive
[dh][1]	**ذ**	**[dhékr]**	**Rappel**	**Fricative**
[dḥ][1]	**ظ/ض**	**[dḥrab]**	**Il a tapé**	**Plosive**
[q]	ق	[qạlb]	cœur	uvulaire
[w]	و	[wè:lu]	Rien	Bilabial semi-voyelle
[j]	ي	[ju:m]	Jour	Médiodorsal post-palatale semi-voyelle

[1] : Le phonème ne se trouve pas dans le système de transcription proposé par N.ZELLAL

Tableau 2. 4 : le système de transcription pour les voyelles d'après Zellal[75]

Symbol	Mot-clé dialectal	Signification en français	Description phonétique
[a]	[kla]	Il a mangé	Aperture maxima antérieure
[ạ]	[tạb]	Il est cuit	Aperture maxima postérieure
[è]	[chè:f]	Il a vu	Aperture semi-ouverte centrale
[é]	[xé:r]	Bien	Aperture semi-fermée centrale
[e]	[rajel]	Homme	Aperture moyenne centrale
[i]	[benti]	Ma fille	Aperture minima antérieure
[o]	[fo:q]	Dessus	Aperture minima labialisée postérieure
[u]	3ụmro]	Son âge	Aperture minima non labialisée
[ø]	3øchch]	Nid	Aperture moyenne labialisée

Les signes diacritiques sont comme suit : Emphase (soit pour consonne : c ou voyelle : v) : ç ẹ ; voyelle nasalisée : ṽ ; voyelle longue : v: , gémination : vv, cc. Il faut noter que la gémination ne se trouve pas dans le système de transcription de N.ZELLAL.

2.2.6.4. Matériau linguistique et conditions d'enregistrement

Les enregistrements ont été effectués avec l'aide de l'orthophoniste pendant les séances de la rééducation dans une salle de consultation dans deux **É**tablissements **H**ospitaliers **S**pécialisés : EHS. Il s'agit des EHS Ras Elma et Séraidi. Ras Elma est un village de la ville de Sétif et Séraidi est un village de la ville d'Annaba. Le matériel utilisé pour l'enregistrement est le suivant

- Nous avons utilisé un casque stéréo Andrea NC-65 avec un microphone anti-bruit qui minimise le bruit ambiant.
- Un ordinateur portable ;
- Une carte audio échantillonnant à une fréquence de 16kHz, avec une résolution de16 bits.

Les enregistrements contiennent généralement des conversations entre l'orthophoniste et le patient. Séquentiellement, on retrouve la parole de l'orthophoniste, suivi d'un temps de réponse avant d'avoir la parole du patient. Parfois, il n'y a pas de temps de réponse et par conséquent, il existe un chevauchement entre la parole de l'orthophoniste et du patient. La parole pathologique enregistrée est donc numérisée avec une fréquence d'échantillonnage de de 16 kHz avec une résolution de 16 bits

Le corpus utilisé est issu du bilan de langage structuré, élaboré par Zellal [76]. Le protocole d'examen est applicable à des aphasiques algériens et respecte rigoureusement les critères d'élaboration et de sélection adoptés à l'origine par Ribaucourt [77]. Ce protocole se caractérise par :

Les contextes et référents situationnels ou textuels sont diversifiés afin de susciter réussites, échecs ou perturbations qui ne s'objectivent que dans certaines formes d'actualisation du langage. Les épreuves de parole spontanée ou de récit libre, font suite des questions contraignantes jaugeant, en langue usuelle, le pouvoir de communication et le degré d'information du malade. Les épreuves de répétition de syllabes et de mots, font intervenir les sons propres à l'arabe, et qui ne figurent pas dans le système français, à savoir : les pharyngales ħ (الحرف ح), les emphatiques ṭ ; ḍ ; ṛ ; ṣ ; l'uvulaire q (ق الحرف و ص, ر, ض, ط). En outre, il est tenu compte de l'absence, de l'arabe, de certains groupes consonantiques, en toutes positions : ħ + q ou dans certains contextes : par exemple ch+k n'existe pas en finale implosive. Les phrases à répéter ne sont pas une simple traduction des phrases extraites du corpus original de Ribaucourt. Celles-ci sont constituées en gardant le même nombre de syllabes le même de degré de complexité. Ces phrases sont sériées en deux catégories : les unes, dans un axe pragmatique, variant par rapport à la sélection lexicale, les autres, dans un axe syntagmatique, mettant en jeu les formes combinatoires de la langue.

Rappelons que les épreuves de répétitions servent à observer la progression des difficultés articulatoires classiques. Les épreuves de dénomination, de lecture à voix haute et de désignation d'images, exigeant l'encodage ou le décodage d'unités isolées, ont été sélectionnées selon plusieurs critères. Ils sont : la longueur de pattern, la complexité phonémique ou graphématique de leur signifiant, le degré de fréquence dans la langue, l'indice de similarité phonémique et graphématique, la distance paradigmatique, de la proximité des champs sémantiques, et enfin le principe de disponibilité qui les régit.

Pour la lecture à voix haute, l'arabe dialectal ne s'écrivant pas, c'est l'arabe standard qui a été utilisé. Par conséquent, la caractérisation acoustico-phonétique ne s'appliquera qu'à ceux qui avant la maladie, savaient lire la langue arabe.

2.2.6.5. Choix des patients

Le nombre restreint de patients présents pendant la période de collecte nous a incités à prendre en considération tous les sujets traités dans les deux Établissements Hospitaliers Spécialisés (EHS) cités précédemment. Le tableau 2.5 donne des informations sur les sujets avec lesquels nous avons constitué le corpus de parole pathologique (La lettre P désigne patient alors que la lettre M désigne le sexe Masculin et la lettre F désigne le sexe féminin).

Tableau 2. 5: Les patients participants à la base de données aphasique ADAD[52].

sexe	Réf	Age	Niveau scolaire	Diagnostique	Enregistrement
PM1	419.06	43ans	Secondaire	Une stéréotypie de elhamdoulilah ameen qui a empêché le spontané + manque du mot : Après 26 séances de rééducation ; récupération sans l'absence totale de la stéréotypie mais due à la non compréhension du patient de l'importance de se taire	- Le corpus complet (20min45) - Dénomination et phrases (15min36) - phrases (9min46)
PM2	1103.06	35ans	Secondaire	Manque de mots (stade initial de la maladie)	- Le corpus complet (15min 29) - Lecture de mots isoler pour éviter la persévération (6,5 min)
PM3	210.06	61ans	Instituteur	Trouble de Persévération de l'expression écrite, manque de mots et difficultés articulatoires (stade final de la maladie : récupération)	- Le corpus avec un paragraphe pour la lecture à haute voix en arabe classique (25min 59)
PM4	13.05	31ans	Universitaire	AVC ischémique au territoire profond de l'artère sylvienne gauche a provoqué un mutisme avec préservation de la compréhension orale et écrite puis stéréotypie : 46 séances de PEC : Le patient a récupéré	- Le corpus complet (27min 30)
PM5	1023.06	47ans	Secondaire	Manque de mots + désintégration phonétique (stade initial de la maladie)	- Corpus complet + le spontané (38min75) - Dénomination des images (12min14)
PM6	723.06	69ans	Universitaire	AVC ischémique au territoire sylvien gauche aphasie de conduction + trouble de discrimination phonétique (consultation externe : pas encore récupéré)	- Corpus complet + dénomination des images (26min75)

PM7	444.06	47ans	Moyen	Anarthrie sévère avec une stéréotypie au début (effacée totalement) : Après 26 séances de rééducation : amélioration puis son cas devient stationnaire (une PEC incomplète)	- Bilan orthophonique (30,83 min) -Dénomination des images (23,54 min) - Spontanée (4,19 min) - Répétions de syllabes (18,13 min)
PM8	095.38	56ans	Illettré	Manque du mot (le malade a récupéré)	-Le corpus complet (24min 39) - Dénomination et phrases (31min13) - Le corpus complet (18min27)
PM9	1103.06	50ans	Secondaire	Manque du mot (dans le stade initial de la maladie	- Le corpus complet (15min 29) - Le corpus complet (20 min)
PM10		35	Académique	Compréhension altérée + mutisme (était pris en charge)	-Corpus complet (33 min)
PM11		65	Illettré	Hémiplégie droite accompagne de troubles de compréhension + manque de mots	-Bilan orthophonique incomplet (12.45 min).
PM12	960.06	60 ans	Illettré	Hémiplégie post AVC (n'est pas un cas aphasique) (pour le contrôle	-Spontané +Bilan orthophonique incomplet (le patient à jeun) (15.46 min)
PF1	776.06	62 ans	Illettrée	Désintégration phonétique (stade initial de la maladie mais une PEC incomplète	-Trois enregistrements de : spontané et séries automatiques et répétition des syllabes incomplet + sourate du coran : El Fatiha (33min 87)
PF2	769.06	66 ans	Illettrée	Désintégration phonétique	- Corpus incomplet avec dénomination (18min 59)
PF3	08/00015	66	Illettrée	Hémiplégie droite poste AVC I	-Spontané + Le bilan orthophonique complet (27min6)

Pour les locuteurs de contrôle ou de référence qui ne sont pas aphasiques et qui n'ont aucun historique de maladie neurologiques ou autres maladies affectant la voix ou la parole ou le langage. Ils étaient en majorité, des étudiants entre 17 ans et 30 ans de l'université de

Msila. Bien sûr, il y a avait quelques enregistrements pour des adultes qui dépassent la quarantaine en plus des enregistrements des orthophonistes qui ont aidé à faire les enregistrements des patients des deux EHS. Plus de détails sur l'âge et le niveau scolaire ainsi que la durée des enregistrements sont donnés dans le tableau 2.6.

Tableau 2. 6 : Les locuteurs références sains participants à la base de données aphasique : ADAD.

Sexe	Age (ans)	Niveau scolaire	Enregistrement
RF1	17	secondaire	✓ Bilan orthophonique : 14 min45
RF2	21	universitaire	✓ Bilan orthophonique : 15 min75
RF3	22	universitaire	✓ Bilan orthophonique : 13 min45
RF4	21	universitaire	✓ Bilan orthophonique : 17 min15
RF5	19	universitaire	✓ Bilan orthophonique : 14 min65
RF6	20	universitaire	✓ Bilan orthophonique : 13 min45
RF7	20	universitaire	✓ Bilan orthophonique : 13 min35
RF8	30	fonctionnaire	✓ Bilan orthophonique : 13 min47
RF9	31	fonctionnaire	✓ Bilan orthophonique : 20 min54
RF10	28	fonctionnaire	✓ Bilan orthophonique : 20 min00
RF11	27	fonctionnaire	✓ Bilan orthophonique : 19 min2
RF12	30	fonctionnaire	✓ Bilan orthophonique : 20 min2
RF13	42	fonctionnaire	✓ Bilan orthophonique : 18 min2
RM1	23	universitaire	✓ Bilan orthophonique : 15min65
RM2	23	universitaire	✓ Bilan orthophonique : 16 min89
RM3	23	universitaire	✓ Bilan orthophonique : 14 min00
RM4	23	universitaire	✓ Bilan orthophonique : 13min6
RM5	35	secondaire	✓ Bilan orthophonique : 12 min6
RM6	23	Étudiant universitaire	✓ Bilan orthophonique : 11 min75

RM7	24	Étudiant universitaire	✓ Bilan orthophonique : 9 min65
RM8	24	Étudiant universitaire	✓ Bilan orthophonique : 10 min45
RM9	23	Étudiant universitaire	✓ Bilan orthophonique : 11min35
RM10	22	Étudiant universitaire	✓ Bilan orthophonique : 17 min74
RM11	27	Étudiant universitaire	✓ Bilan orthophonique : 16 min35
RM12	26	Étudiant universitaire	✓ Bilan orthophonique : 17 min45
RM13	29	Étudiant universitaire	✓ Bilan orthophonique : 19 min00
RM14	22	Étudiant universitaire	✓ Bilan orthophonique : 16 min25
RM15	23	Étudiant universitaire	✓ Bilan orthophonique : 14 min69
RM16	23	Étudiant universitaire	✓ Bilan orthophonique : 10 min75
RM17	24	Étudiant universitaire	✓ Bilan orthophonique : 15 min98
RM18	23	Étudiant universitaire	✓ Bilan orthophonique : 18 min45
RM19	22	Étudiant universitaire	✓ Bilan orthophonique : 18min85
RM20	22	Étudiant universitaire	✓ Bilan orthophonique : 24 min47
RM21	22	Étudiant universitaire	✓ Bilan orthophonique : 23 min12
RM22	22	Étudiant universitaire	✓ Bilan orthophonique : 20 min52
RM23	25	Étudiant universitaire	✓ Bilan orthophonique : 18min65
RM24	24	Étudiant universitaire	✓ Bilan orthophonique : 16 min45
RM25	23	Étudiant universitaire	✓ Bilan orthophonique : 17 min73
RM26	22	Étudiant universitaire	✓ Bilan orthophonique : 18 min25
RM27	22	Étudiant universitaire	✓ Bilan orthophonique : 15min60
RM28	23	Étudiant universitaire	✓ Bilan orthophonique : 20min15
RM29	52	Professeur à l'université	✓ Bilan orthophonique : 18min15

Le tableau 2.7 donne un résumé général des caractéristiques de la base de données ADAD. La taille du lexique de la base de données ne peut pas être calculée à cause de la

parole spontanée que nous ne pouvons pas contrôler son lexique évidemment. Donc, nous n'avons recensé que le lexique de la tache de répétitions.

Tableau 2. 7 : Résumé des caractéristiques de la base de données ADAD.

Langue	Dialecte algérien	**Taille de lexique de répétitions**	285
Type de tache de parole	Répétition, spontané, lecture	**Le nombre de phonèmes**	43
Échantillonnage	16 kHz	**Temps total pour la PA***	9 heures et 12 min
Quantification	16 bits	**Temps total pour la PN ****	11heures et 43 min
Nbre de participants	57 (15 patients et 42 normaux)		

*PA : Parole Aphasique, ** PN : Parole Normale ou Saine

2.2.6.6. Évaluation de l'aphasie des patients d'ADAD : Aphasiogramme

L'évaluation de l'état du patient aphasique commence tout d'abord par un examen neurologique et par la suite une synthèse d'un bilan psycholinguistique comprenant plusieurs épreuves : de langage oral (interview dirigée, narration orale, répétitions, dénominations, compréhension orale) et de langage écrit (dictée, écriture libre).

L'évaluation de l'aphasie est basée sur l'outil de test empirique supplémentaire l'aphasiogramme qui comprend l'élaboration de l'examen et le contrôle de la parole et langage du patient. L'aphasiogramme c'est un tracé de courbe des résultats de l'évaluation quantitative en % de réussite des réponses du patient. Il correspond, respectivement, au tracé d'une courbe pour le langage oral et une courbe pour le langage écrit.

Toutefois, le premier aphasiogramme permet le diagnostic du syndrome et il établit une estimation pour la rééducation. Mais il sera plus facile de répéter le même examen après une période pour estimer l'évolution de l'aphasie et le nouvel état du patient, car une comparaison entre au moins deux ensembles de données permet la mise en place d'un programme de réhabilitation.

La procédure d'évaluation de l'aphasie consiste à l'examen de :

- L'expression orale,
- La compréhension de la langue orale,
- L'expression écrite,
- De la compréhension de la langue écrite.

Dans notre étude, nous avons essayé d'obtenir une partie de l'aphasiogramme et se concentrer uniquement sur une partie de l'examen de l'expression orale. Cette partie consiste aux taches de répétitions à haute voix (voir la figure 2.2.).

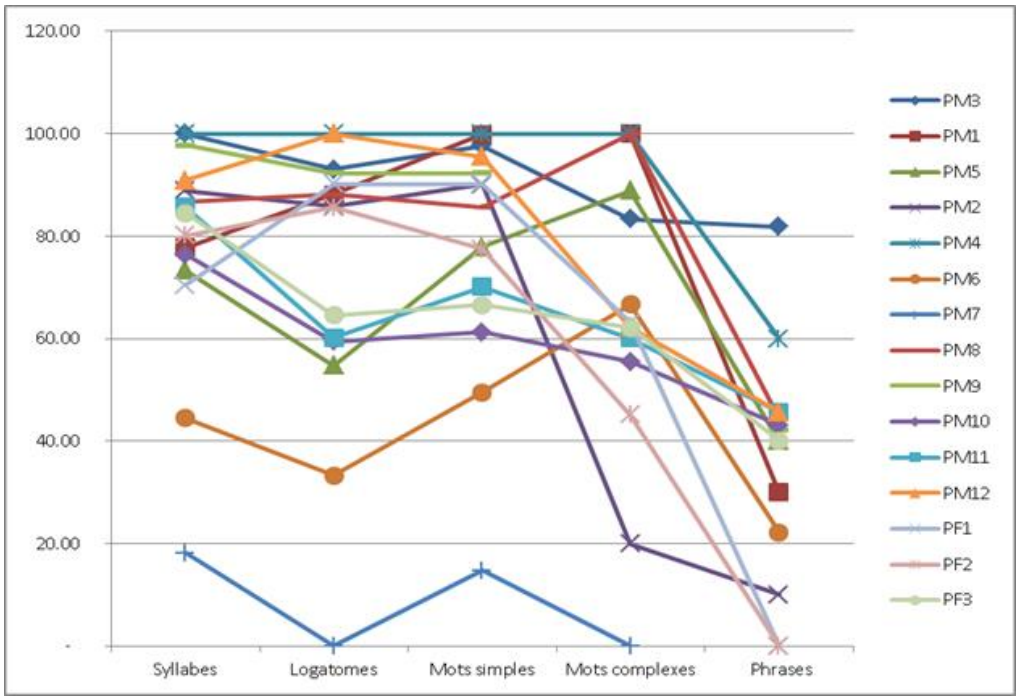

Figure2. 2: Aphasiogramme de tous les patients aphasiques

2.3. La Dysarthrie (The dysarthria)

2.3.1. Définition de la dysarthrie

Le terme dysarthrie désigne un trouble de l'exécution motrice de la parole. La dysarthrie représente un syndrome très complexe, rencontré dans de nombreuses pathologies neurologiques qui peuvent être aigues (accident cérébral, traumatisme crânien) ou évolutives (maladie de Parkinson, sclérose latérale amyotrophique, sclérose en Plaque) et aussi les paralysies faciales, etc. Les troubles résultants concernent la respiration, la résonance, l'articulation, le débit et la prosodie. Voilà donc pourquoi il s'agit d'une production anormale de la parole [78-82].

2.3.2. Types de dysarthrie.

La classification des types de dysarthrie distingue la dysarthrie: spastique, flasque, ataxique, hypokinétique, hyperkinétique et mixte [83, 84]. L'atteinte du faisceau pyramidal sera généralement associée à une dysarthrie spastique, tandis que le syndrome cérébelleux sera associé à une dysarthrie ataxique. Les patients parkinsoniens présentent par définition une dysarthrie hypokinétique. Les syndromes parkinsoniens atypiques, quant à eux, seront

associés à la dysarthrie de type mixte (traumatismes crâniens, sclérose en plaques, sclérose latérale amyotrophique). Les patients atteints de la dysarthrie hyperkinétique sont les patients de la maladie Huntington. Cette dernière, une affection dégénérative héréditaire du système nerveux central, d'évolution inexorable, à révélation tardive, décrite en 1872 par George Huntington.

2.3.3. Symptômes de dysarthrie

Les sujets atteints de dysarthrie présentent les symptômes suivants :

- Difficultés d'articulation.
- Lenteur de la parole ; le débit est altéré, le rythme est saccadé.
- Rythme rapide de la parole qui est difficile à comprendre.
- Altération fréquente de la voix : grinçante, assourdie.
- Le volume inégal de la parole.
- Discours monotone.
- Constance des troubles, qui sont relativement réguliers (à la différence de l'anarthrie, marquée par une plus grande diversité des productions)[84].

2.3.4. La base de données de la parole dysarthrique : Nemours.

La base de données Nemours est une collection de 814 phrases non sensées courtes construites à partir d'un lexique de taille de 111 mots. Chaque patient répète 74 phrases. Les participants sont 11 locuteurs adultes masculins avec des degrés variables de dysarthrie. En addition, la base de données contient deux paragraphes produits par chacun des 11 locuteurs de la parole continue. Elle peut également être utilisée pour étudier les caractéristiques générales du discours dysarthrique tels que les schémas d'erreur de production [85, 86].

Les patients dysarthriques participants au corpus Nemours sont affectés principalement par la dysarthrie spastique. Les dysarthries spastiques sont en relation avec une atteinte supra-bulbaire. Les dysarthries spastiques sont causées par des lésions d'origine dégénérative, vasculaire ou traumatique siégeant à l'étage cortico-sous-cortical.

La dysarthrie spastique se caractérise surtout par une parole laborieuse avec des distorsions faciales, l'émission de phrases courtes, une voix rauque, étranglée, faible, avec des cassures tonales ; une accélération du rythme respiratoire, des inspirations profondes difficiles, des mouvements antagonistes des musculatures diaphragmatiques, abdominales et

thoraciques, des mouvements involontaires de la musculature thoracique. Toutes ces modifications respiratoires se manifestant par : une incapacité à produire plus de deux syllabes dans une expiration, des difficultés à maintenir la pression de l'air pour la vocalisation qui est faible et discontinue. En général, la dysarthrie est associée à une faiblesse de la musculature palato pharyngée et linguale se manifestant par un ralentissement global des contractions participant à la résonance et à l'articulation : imprécision des consonnes (disparition des fricatives, les sourdes deviennent sonores.), monotonie, diminution de l'accentuation, voix rauque, débit lent, hyper nasalité, voix forcée, phrase courtes, distorsion des voyelles, voix soufflée, accentuation excessive [53].

2.3.4.1. Choix des locuteurs et stimulus

Pour examiner l'intelligibilité, une liste de mots est dressée d'une manière semblable à celle décrite par Kent [87]. Chaque mot est associé à un nombre de mots proches en prononciation à ce dernier (par exemple le mot : boat et les mots proches : vote », moat, goat).

Comme c'est indiqué dessus, l'ensemble complet de stimulus se compose de 74 noms monosyllabiques, et de 37 verbes bi-syllabiques inclus dans des phrases insensées (le total de stimulus est 111 mots). Les 37 premières phrases sont de la forme suivante :

The nom1 verbe+ing the nom2

Dont les noms et le verbe sont tirés aléatoirement de l'ensemble de 111 mots. Les 37 dernières phrases sont obtenues par permutation des positions des noms des 37 premières phrases (la phrase devient donc de la forme suivante : The nom2 verbe+ing The nom1) comme c'est donnée par l'exemple suivant : une phrase dans les 37 premières (pour un patient donné) : The thin is sining the badge. Une phrase des 37 dernières : The badge is sining the thin.

Tandis que Kent, qui utilise exclusivement des mots monosyllabiques, des verbes sont utilisés en format « ing » dans lesquels la consonne finale de la première syllabe de l'infinitif pourrait être le phonème d'intérêt. Par exemple le /p/ dans « reaping » peut être testé avec les proches prononciations telles que « reading » et « reeking ».

Les patients sont onze jeunes adultes de sexe masculin atteints de dysarthrie résultant de dysarthrie cérébrale (CP : Cerebral Palsy) ou trauma de cane (Head trauma). Sept locuteurs ont une paralysie cérébrale : trois ont une CP spastique avec quadriplégie. Les quatre locuteurs restants sont victimes de trauma de crane (une quadriplégie et un avec le quadriparesis spastique).

2.3.4.2. La transcription phonétique de Nemours (voir l'annexe B).

La base de données entière a été étiquetée automatiquement au niveau des mots et des phrases à l'exception des données du patient KS dont la parole est sévèrement intelligible (voir tableau 2.14). Les modèles de Markov Cachés discrètes (Discret Harden Markov Model : DHMM) ont été utilisés pour cette fin.

Chaque phonème est codé de un à trois caractères comme utilisé dans la base de données de la parole TIMIT[88]. L'ensemble de phonèmes comprend 25 consonnes, sept semi-voyelles et glides et vingt voyelles. Les intervalles de fermeture des arrêts qui se distinguent de la réalisation de l'arrêt en lui-même. Leurs codes pour les consonnes b, d, g, p, t, k sont BCL, DCL, GCL, PCL, TCK, KCl, respectivement. Concernant les deux affriquées : jh et ch, sont DCL et TCL. Le système phonétique illustré par le tableau 2.8 d'après la base de données TIMIT.

Tableau 2. 8 : le système de transcription de la base données Nemours [88]

Classe phonologique	Phonèmes	Exemple de mot	Transcription phonétique
Les arrêts	b	Bee	BCL B iy
	d	day	DCL D ey
	g	Gay	GCL G ey
	p	pea	PCL P iy
	t	Tea	TCL T iy
	k	Key	KCL K iy
	dx	muddy, dirty	m ah DX iy, dcl d er DX iy
	q	baq	bcl b ae Q
Les Affriquées	jh	joke	DCL JH ow kcl k
	ch	choke	TCL CH ow kcl k
Fricatives	s	Sea	S iy
	sh	She	SH iy
	z	Zone	Z ow n
	zh	Azure	ae ZH er
	f	fin	F ih n

	v	Van	V ae n
	dh	then	DH e n
	th	thin	TH ih n
Nasales	m	Mom	M aa M
	N	Noon	N uw N
	ng	Sing	s ih NG
	em	bottom	b aa tcl t EM
	en	button	b ah q EN
	eng	Shington	w aa sh ENG tcl t ax n
	nxr	Winne	w ih NX axr
Semi-voyelles et Glides :	l	lay	L ey
	r	ray	R ey
	w	Way	W ey
	y	Yacht	Y aa tcl t
	hh	hay	HH ey
	hv	ahead	ax HV eh dcl d
	el	bottle	bcl b aa tcl t EL
Voyelles	iy	Beet	bcl b IY tcl t
	ih	Bit	bcl b IH tcl t
	eh	bet	bcl b EH tcl
	ae	bat	bcl b AE tcl tt
	aa	bott	bcl b AA tcl t
	aw	bout	bcl b AW tcl t
	ay	bite	bcl b AY tcl t
	ah	but	bcl b AH tcl t
	ao	Bought	bcl b AO tcl t
	oy	boy	bcl b OY
	ow	Boat	bcl b OW tcl t
	uh	Book	bcl b UH kcl k
	uw	boot	bcl b UW tcl t
	ux	toot	tcl t UX tcl t
	er	bird	bcl b ER dcl d
	ax	about	AX bcl b aw tcl t
	ix	debit	dcl d eh bcl b IX tcl t
	axr	butter	bcl b ah dx AXR
	ax-h	suspect	s AX-H s pcl p eh kcl k tcl t

2.3.4.3. Les conditions d'enregistrements

Les sessions d'enregistrement ont été menées dans une cabine acoustiquement isolée en chaise roulante. À l'aide d'un microphone omnidirectionnel dynamique Electrovoice RE55 monté sur table reliée à un enregistreur audio numérique Sony, modèle PCM-2500 situé à l'extérieur de la cabine d'enregistrement. Le patient était assis, généralement dans un fauteuil roulant, à côté de l'expérimentateur ou orthophoniste, et environ 12 pouces loin du microphone.

Les sessions d'enregistrement commençaient par un bref examen par l'orthophoniste en considérant l'évaluation Frenchay de la dysarthrie [80]. Suite à l'évaluation et après une courte pause, l'expérimentateur entre dans la cabine d'enregistrement pour conduire le patient à travers une série d'enregistrements qui comprenait l'ensemble des 74 phrases sémantiquement anormales décrites ci-dessus, suivie par deux paragraphes de la parole connectée « My grand Father» et «The Rainbow ». Le matériel linguistique a été tapé en gros caractères sur le papier placé devant le patient. Un peu de temps a été prévu pour familiariser le patient avant le commencement de l'enregistrement. Pour le matériel de phrase, chaque phrase a été lue en premier lieu par l'orthophoniste et ensuite répétée par le patient. L'orthophoniste assistait tous les patients dans la prononciation des mots surtout certains sujets avec un faible niveau de vue ou d'alphabétisation. Enfin, les patients enregistraient deux paragraphes de la parole connectée. En moyenne, chaque session d'enregistrement a été achevée en deux heures et demie à trois heures, y compris le temps pour les pauses.

La parole enregistrée à la fois du patient dysarthrique et l'orthophoniste était plus tard numérisée. La fréquence d'échantillonnage est de 16 kHz avec une résolution de 16 bits après filtrage passe-bas à une valeur nominale de 7500 Hz fréquence de coupure d'un filtre 90 dB / Octave (12 dB d'atténuation à la fréquence de Nyquist.).

2.3.4.4. L'évaluation perceptive de Frenchay de la dysarthrie (Frenchay Dysarthria Assessment : FDA)

Au minimum cinq auditeurs normaux ont été recrutés parmi les étudiants de l'université du Delaware pour les essais d'écoute avec chaque patient dysarthrique. Les phrases ont été présentées aux auditeurs sous la forme normale par un locuteur normal et ensuite par un patient dysarthrique. Au début de chaque essai, une liste de phrases possibles apparait à l'écran et les auditeurs doivent sélectionner la phrase qu'ils pensent que le patient voulait prononcer. Par exemple, une phrase pourrait apparaitre comme suit :

	Fin		Sipping		Bath
The	Thin	Is	Sinning	The	badge
	sin		Sitting		batch
	sin		Sipping		bash

Ainsi le mot cible est associé à plusieurs mots : une liste de 4 à 6 alternatives possibles, qui peuvent avoir la même ou la proche prononciation. L'auditeur doit choisir l'alternative la plus correcte de la liste.

En fait, La FDA est un critère bien établi pour le diagnostic de la dysarthrie. Le test est divisé en 11 sections, à savoir, réflexe, le palais, les lèvres, la mâchoire, la langue, l'intelligibilité… etc. Chaque locuteur dysarthrique est évalué sur un certain nombre de tâches simples. FDA est une échelle de 9 points avec des scores allant de 0 à 8. Dans la FDA, un score de « 8 » représente une fonction normale et « 0 » ne représente aucune fonction. les résultats de l'évaluation Frenchay de la dysarthrie (FDA) pour 9 locuteurs dysarthriques sont donnés par le tableau 2.9 [80].

Tableau 2. 9: L'Évaluation Frenchay des Patients de Nemours (Frenchay Dysarthria Assessment : FDA)

Patients	NS[3] (%)	CI[1]	IP[2]	Classe	Résumé du FDA
KS	-	1	1	Patient Grave (PG ou PS)	fissure palatine sous-muqueuse est possible
SC	49.5	3	1	Patient Grave (PG ou PS)	Un grand nombre de tâches confondues par le manque de soutien de l'air. SC compense nombreuses déficiences en utilisant des moyens alternatifs de production (par exemple, la lèvre inférieure et sous lèvre supérieure pour bilabial, donc le contact est vraiment avec la lèvre supérieure et le menton).
BV	42.5	3	2	Patient Grave (PG ou PS)	----------
BK	41.8	3	0	Patient Grave (PG ou PS)	Produit / p / avec le contact labiodentale. Il pourrait construire une légère pression en utilisant cette tendance. Il a Beaucoup d'articulations compensatoires.
RK	32.4	2	1	Patient Modéré (MOP)	Intelligibilité des mots et des phrases dont il a réussi de lire.
RL	26.7	4	3	Patient Modéré (MOP)	----------
JF	21.5	4	3	Patient Modéré (MOP)	Lèvres : symétrie est bonne ; tremblement perceptible.

LL	15.6	6	4	Les patients à pathologie légère (Mild Patient : MP)	LL compense ses difficultés à la parole spontanée en parlant de courtes phrases. Il brise plus longuement ses énoncés en segments pour maintenir un soutien adéquat pour la phonation. Latéralisation de la langue est rapide mais imprécise, accompagné par des grimaces et un mouvement inapproprié de la mâchoire.
BB	10.3	8	8	Les patients à pathologie légère (Mild Patient : MP)	Lae score d'intelligibilité omet les mots mal lus. Il avait un certain nombre de ce type d'erreur.
MH	7.9	6	4	Les patients à pathologie légère (Mild Patient : MP)	Difficultés avec la distinction de voyelles et de certaines consonnes voisées.
FB	7.1	-	-	Les patients à pathologie légère (Mild Patient : MP)	

1 : Intelligibilité de Conversation, 2 : Intelligibilité de Phrase, 3 : Niveau de Sévérité

Tableau 2. 10 : Résumé des caractéristiques de la base de données Nemours

Langue	Dialecte a**méricai**n	**Nombre d'itérations**	824
Type de tache de parole	Répétition, lecture	**Taille de lexique de répétitions**	111
Échantillonnage	16 kHz	**Le nombre de phonèmes**	52
Quantification	16 bits	**Temps total**	~30 heures
Nbre de participant	12		

2.4. Conclusion

Dans ce chapitre, nous avons vu que la plupart des personnes aphasiques de Broca ont des difficultés à trouver leurs mots. Elles ont de la peine à donner le nom exact d'un objet, bien que, dans leurs têtes, elles savent exactement ce qu'elles veulent dire. Les connaissances et la pensée sont là, mais pas l'expression adéquate. Par ailleurs, de nombreuses personnes aphasiques peuvent avoir du mal pour prononcer les mots et /ou pour les articuler. Tandis que la dysarthrie peut être le fruit de la lésion d'un seul des neurones par lesquels transitent les informations de parole. Cette lésion entraîne une altération des mouvements de la langue, de la gorge, des lèvres. Les symptômes de la dysarthrie sont : parole marmonnée, débit de parole altéré, rythme saccadé, parole plus nasale, dysphonie, voix rauque, voix monotone.

Pour étudier la parole aphasique, nous avons suivi l'examen oral du professeur Zellal constitué de la parole spontanée et les différents types de répétitions. Nous avons donc, décrit surtout notre base de données aphasique. Une base de données que nous considérons importante étant donné la difficulté de la collection des données pathologique et surtout

l'aphasique : chaque cas aphasique est un cas unique et le manque de cas disponibles. Aussi, une autre qualité de la base est le fait qu'elle soit une base dialectale (rareté de telles ressources linguistiques). La base de données Nemours était aussi décrite en se concentrant surtout sur l'évaluation Frenchay de la dysarthrie.

Chapitre 3 :

Paramétrisation De La Parole Pathologique

CHAPITRE 3 : PARAMÉTRISATION DE LA PAROLE PATHOLOGIQUE

3.1.Introduction

Il existe une grande variété de paramétrisations possibles afin de discriminer la parole pathologique de la parole saine. Notons que cette première étape, commune à tout système d'extraction de la parole, se limite trop souvent aux calculs des coefficients cepstraux (MFCC). Alors qu'actuellement les efforts se portent plus sur les méthodes de classification (modélisation). L'étape de la paramétrisation est essentielle puisqu'il s'agit d'extraire d'un signal fortement redondant, l'information pertinente relative à la tâche de classification donnée. Elle est donc pertinente et mérite de s'y attarder un peu plus. En effet, la paramétrisation influe notablement sur les résultats obtenus. Il faut cependant noter qu'il est difficile de paramétriser la parole pathologique compte tenu de son extrême variabilité.

Des travaux récents ont montré qu'en prenant en compte les connaissances prosodiques nous pouvons améliorer considérablement les performances de la reconnaissance automatique de la parole et surtout la parole spontanée. L'information prosodique doit être intégrée de façon efficace dans les systèmes de classification ou reconnaissance automatique de la parole. Cette dernière peut également constituer un outil important de diagnostic et de rééducation pour l'orthophoniste.

Dans ce contexte, nous essayons de construire plusieurs outils afin d'avoir le maximum d'informations prosodiques sur nos données. Notre objectif est également de déterminer quels paramètres sont les plus pertinents et en corrélation avec la qualité de la parole dysarthrique ou aphasique. L'objectif est donc d'exploiter les ressources linguistiques disponibles à des fins de suivi thérapeutique et / ou d'effectuer un traitement automatique de la parole pathologique. Pour ce faire, nous avons mené trois analyses : prosodique, spectrale et rythmique. Nous accorderons une attention particulière aux paramètres rythmiques [39, 89-96]. En effet, leur utilisation dans le domaine de la parole pathologique reste limitée mais semble prometteuse [44, 97].

3.2.Prétraitement de la parole aphasique (ADAD : Aphasic Dialectal Algerian Database).

La parole de la base de données **ADAD** contient de nombreux types de bruits comme bruit de claquement de portes, fond de conversations dans le couloir, bruit de papier…etc. Ceci est dû au fait que la parole a été enregistrée dans un milieu hospitalier et non pas dans une chambre acoustiquement isolée pour une utilisation spécifique de la tâche de reconnaissance automatique de la parole. Nous avons donc exclu des enregistrements entiers jugés de mauvaise qualité pour une tache de reconnaissance de la parole. Aussi, nous avons écarté les segments de parole entachés par les types de bruits suivants : musique, chevauchements de parole, les appels téléphoniques, et tout bruit indéterminé. Nous avons ensuite essayé et testé plusieurs algorithmes de réduction de bruit sur certains exemples d'enregistrements. Une nette amélioration a été obtenue essentiellement en appliquant la réduction de bruit par soustraction spectrale.

Une détection du début et fin de parole / silence était nécessaire. S'en suit une préaccentuation par passage du signal dans un filtre de transmission passe-haut. Cette opération vise à accentuer la partie haute fréquence utile avec un coefficient de 0.97.

3.3. Analyse temporelle prosodique

La prosodie c'est l'ensemble des phénomènes d'intonation d'une langue. La prosodie regroupe, en fait, les caractéristiques de la parole continue qui contribuent à la perception de celle-ci comme constituant un flux auditif cohérent, rythmé et intoné de façon naturelle. Dans une conversation, le ton, le rythme et le volume de la voix changent pour souligner les mots importants ou traduire l'émotion qui s'y rattache. L'intonation, dans la parole "spontanée", "naturelle", est également, et peut être essentiellement, pertinente au niveau des autres fonctions du langage, surtout au niveau de la manifestation des attitudes et des émotions [96].

La prosodie est interprétée par les paramètres acoustiques suivants :

- **La fréquence Fondamentale ou le pitch** : variations subies d'une voix grave à une voix aiguë / tonalités atypiques et niveaux de fréquences fondamentales élevées ;
- **L'intensité :** un chuchotement se transforme en cri/excès ou défaut du volume de la voix ;
- **Le débit :** trop rapide ou trop lent ;

- **L'intonation :** une élocution monotone pourrait relever de difficultés dans l'expression des émotions ou discordante avec la situation d'interlocution. D'autres ont une façon de parler plus chantante et plus mélodieuse mais dépourvue d'émotion et d'intention communicative. Ils présentent rarement l'accent régional de leur lieu d'habitation [96, 97]. Dans ce qui suit, nous allons définir en détails les caractéristiques acoustiques prises en considération dans notre étude de la parole aphasique algérienne[42].

3.3.1. Paramètres acoustiques de La prosodie

3.3.1.1. La fréquence fondamentale et / ou le pitch : Fo

Un signal voisé est le résultat de l'ouverture et la fermeture périodique des cordes vocales. L'inverse de la période est la fréquence fondamentale de la parole. Pitch est la sensation de la fréquence fondamentale des impulsions du flux d'air provenant des plis glottiques. Les termes : pitch et fréquence fondamentale de la parole sont utilisés de manière interchangeable.

Lors de l'évolution temporelle d'un signal vocal, on constate une alternance de zones assez périodiques et de zones bruitées. On les nomes respectivement zones voisées et non voisées. Dans les zones non voisées, on ne trouve aucune fréquence fondamentale, contrairement aux zones voisées où elle évolue lentement au cours du temps. La fréquence fondamentale détermine la hauteur du son. Pour un son de basse fréquence est un son grave, et donc pour un son de haute fréquence est un son aigu. Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants.

Le calcul de la fréquence fondamentale est essentiel pour le traitement automatique de la parole : codage, synthèse acoustique et l'étude prosodique de la parole. Il existe de nombreuses méthodes pour déterminer la fréquence fondamentale. La plupart sont basées sur un calcul d'autocorrélation. La corrélation entre les deux formes d'onde est une mesure de leur similitude. Cette méthode est basée sur la détection des maximas de la fonction d'autocorrélation d'un signal. Les positions de ces maximas nous informe sur l'existence du fondamental d'un signal. La définition mathématique de la fonction d'autocorrélation est montrée dans l'équation 3.1, pour une fonction discrète infinie $x[n]$ et l'équation 3.2 montre

la définition mathématique de l'autocorrélation d'une fonction discrète finie $x'[n]$ de taille N. [41, 44, 95, 97]

$$R_x(v) = \sum_{n=-\infty}^{\infty} x[n]x[n+v] \quad \textbf{(3. 1)}$$

$$R_{x'}(v) = \sum_{n=0}^{N-1-v} x'[n]x'[n+v] \quad \textbf{(3. 2)}$$

Donc si aussi simple que $R_{x'}$ est périodique x(n) est périodique.

3.3.1.2. L'intensité ou l'énergie

L'énergie d'un signal est une caractéristique liée à la quantité de l'information représentée. Résultant de la pression sous glottique, l'intensité est le paramètre prosodique le plus simple à évaluer. Elle est mesurée sur des portions du signal allant de 5 à 10 ms (énergie à court terme) et exprimée en décibel pour respecter l'échelle perceptive. Sa formule est la suivante : [41, 47]

$$Edb = 10\,\mathrm{Log}_{10}\left(\frac{1}{T}\sum_{t=1}^{T} {}^{2}_{t}\right) \quad \textbf{(3.3)}$$

Nous étudions dans ce qui suit l'énergie de la parole voisée et spécialement la parole vocalique.

3.3.1.3. La durée

Au contraire de la fréquence et de l'intensité, on ne peut pas déterminer précisément la durée, parce qu'elle même reliée à la précision du processus « segmentation ». Elle fixe deux événements qui délimitent ses repères initial et final sur un signal de parole. Nous avons segmenté le signal parole par logiciel « WaveSurfer » [98]. Il permet la visualisation et la manipulation des sons. Nous nous en servirons principalement pour la visualisation des signaux audio sous forme de spectrogrammes et de « waveforms ». Il permet, entre autres, de réaliser de très bonnes segmentations des fichiers audio traités en utilisant les règles de l'acoustique de la parole.

3.3.1.4. Taux de passage par zéro (TPZ)

Le taux de passage par zéro (ZCR : Zéro Crossing Rate) représente le nombre de fois que le signal, dans sa représentation amplitude/temps, passe par la valeur zéro. Il est fréquemment employé pour des algorithmes de détection de sections voisées/non voisées

dans un signal. En effet, la partie non voisée possède généralement un taux de passage par zéro supérieur à celui de la partie voisée.

Le comptage du nombre de passages par zéro est très simple à effectuer. Dans un premier temps, il faut enlever la composante continue (ou la valeur moyenne du signal ; offset en anglais), produite par la majorité des équipements de captage et d'acquisition, pour centrer le signal autour de zéro. Ensuite, pour chaque trame, il suffit de dénombrer tous les changements de signe du signal. Le TPZ peut être estimé par la formule (3.4) pour un signal de parole x(n) sur une fenêtre de N échantillons :

$$Z(n)=\frac{1}{N}\sum_{n=1}^{N-1}\frac{|sgnx(n)-sgnx(n-1)|}{2} \qquad (3.4)$$

3.3.1.5. L'intonation

Mertens et al. définissent l'intonation comme suit : « En tant qu'aspect de la communication parlée, l'intonation peut s'étudier sous plusieurs angles : celui de sa substance sonore, celui de sa forme et celui de ses fonctions. Au niveau de la substance, l'intonation se manifeste par plusieurs propriétés sonores (changements de hauteur, accentuation, durée, débit, pauses, rythme, prises de souffle, qualité vocale). De plus, chaque langue se sert de son propre inventaire de formes intonatives et soumet leur utilisation à des contraintes syntaxiques. Mais il ne faut pas oublier que l'intonation a d'abord une fonction communicative : c'est un moyen linguistique pour transmettre certaines informations » [99].

Pour étudier l'intonation, nous nous sommes intéressés à sa forme. Donc, nous distinguons deux types d'intonation l'une ascendante (souvent en modalité d'interrogation) et l'autre descendante (généralement en modalité d'assertion). Ces deux contours intonatifs peuvent être décrits simplement par une régression linéaire qui est une droite reliant la fréquence fondamentale et le temps et dont l'équation est comme suit :

$$F_0=at+b \qquad (3.5)$$

La régression linéaire consiste à déterminer une estimation des valeurs a et b selon la méthode des moindres carrés [100].

- **Taux de suivi (FR : Follow up Rate) :** c'est un nouveau paramètre d'intonation que nous proposons. Le calcul du taux de suivi est effectué par l'accumulation des

répétitions dont le signe du coefficient « a » est le même pour le patient et le sujet de référence soit positif ou négatif. Puis, nous calculons son pourcentage.

3.3.1.6. Débit

Le débit est synonyme de la vitesse d'articulation (SR : Speaking Rate). Le débit de la parole est un paramètre qui intègre les notions de durée et de rythme. Il peut relater de la variabilité au sein d'un même énoncé ou de la variabilité entre différents styles de parole, différentes situations communicationnelles.

Il existe plusieurs approches pour déterminer le débit de parole, surtout en fonction des unités de référence choisies (phonèmes, syllabes, mots, énoncés...). Nous avons indiqué deux définitions simples utilisant comme unités de référence segment voisée et non voisée :

$$SR_1 = \frac{V}{V+U} \qquad (3.6)$$

$$SR_2 = \frac{V}{U} \qquad (3.7)$$

Où : V : la durée des segments voisés, U : la durée des segments non voisés.

3.3.2. Méthode d'analyse

Les caractéristiques prosodiques étudiées de la parole aphasique comprennent cinq propriétés acoustiques. La fréquence fondamentale F_0, intensité, durée, débit et l'intonation ont été mesurées pour tous les types des tâches de répétitions. L'analyse statistique a été effectuée comme indiqué par le schéma synoptique donné par la figure 3.1.

Pour les segments de voyelles extraits de parole aphasique, nous avons mesuré et analysé la variation du pitch et de l'énergie en calculant leur moyenne et leur écart type. Pour les deux grandeurs moyennes et écart-types (variance), le minimum, la moyenne et la valeur maximale seront extraits. À la fin, la durée des voyelles est prise en considération par sa valeur minimale, sa moyenne et sa valeur maximale.

Pour les segments voisés, nous calculons les paramètres suivants : fréquence fondamentale, énergie et débit (rythme) trois nouveaux paramètres : la moyenne, la variance et l'étendue. Ensuite, nous calculons pour la valeur moyenne du pitch sa valeur minimale, moyenne et sa valeur maximale. Nous répétons la procédure avec les deux paramètres

restants. Nous calculons aussi le débit SR par l'équation 3.6. Finalement, pour étudier l'intonation, le coefficient "a" de l'équation (3.5) de chaque répétition est exprimé en pourcentage avec un autre paramètre supplémentaire que nous proposons ; c'est le taux de suivi d'intonation FR.

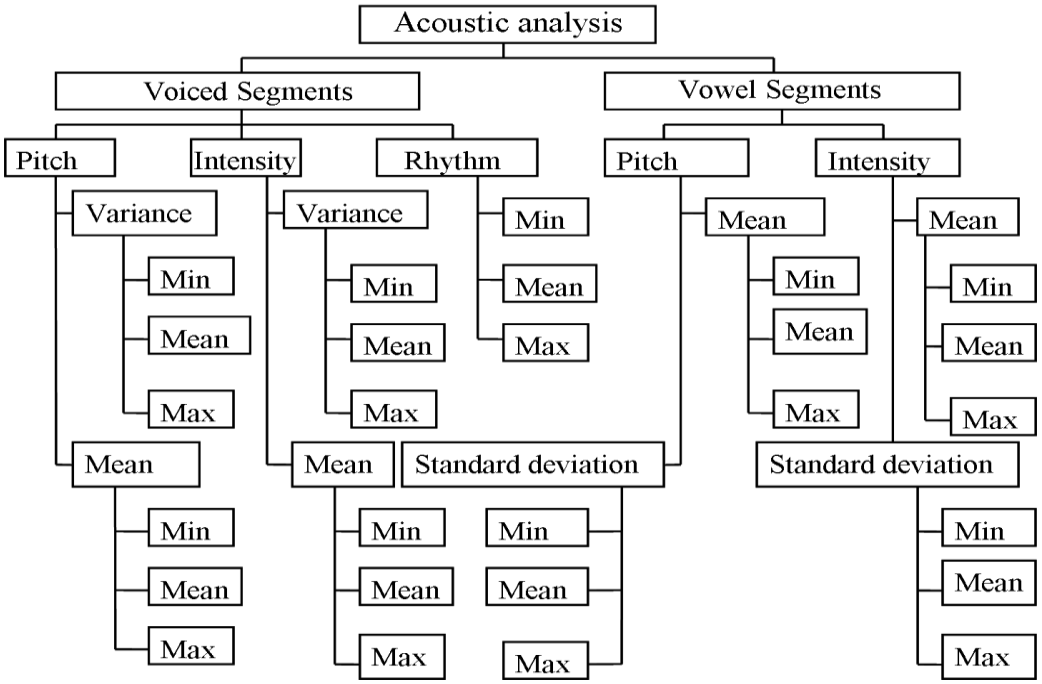

Figure 3. 1 : Schéma synoptique proposé pour l'analyse prosodique acoustique

La figure 3.2 présente l'interface graphique de cette analyse prosodique réalisée sur les intervalles voisés soit : voyelle ou tout simplement les segments voisés.

Nous effectuons une étude statistique afin de donner une vue générale sur l'homogénéité des données des patients. La variance qui est par exemple une mesure de dispersion autour de la moyenne est couramment utilisée.

Afin d'extraire l'information prosodique, nous avons utilisé deux types de segments de la parole : les segments vocaliques et les segments voisés / non voisés.

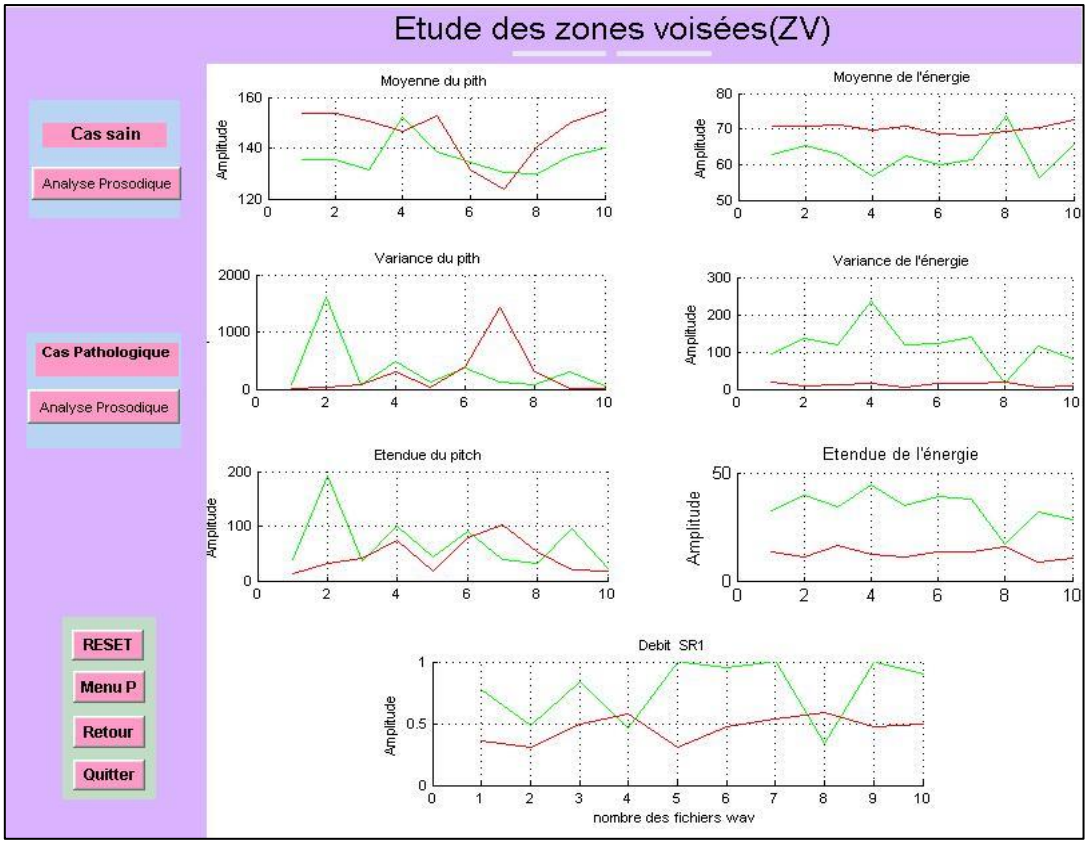

Figure 3. 2 : Le module d'analyse développé pour la parole voisée.

3.3.2.1. La segmentation vocalique [101, 102]

La segmentation du signal de parole vise à extraire des unités phonétiques plus ou moins précises. Dans notre cas, nous cherchons à extraire deux unités phonétiques : voyelle/ consonnes.

Nous avons utilisé une méthode statistique donnée par l'algorithme de **divergence forward-backward (DFB)** [101]**.** Une segmentation automatique indépendante du locuteur et de langues proposée par André Obrecht [101, 102]. La segmentation était effectuée en collaboration avec l'Institut des Systèmes Intelligents et de Robotique : ISIR [103, 104].

3.3.2.1. La segmentation voisée / non voisée

Les intervalles voisés et non voisés ont été segmentés automatiquement avec un script Matlab. Après l'extraction de la fréquence fondamentale, l'énergie et le taux de passage par zéro à l'aide d'un script Praat [105]. En fait, la segmentation voisée/non voisée est basée sur le calcul de F_0, l'énergie à court terme, et le calcul du taux de passage par zéro à travers des

fenêtres glissantes (Hamming en générale) de longueur N.

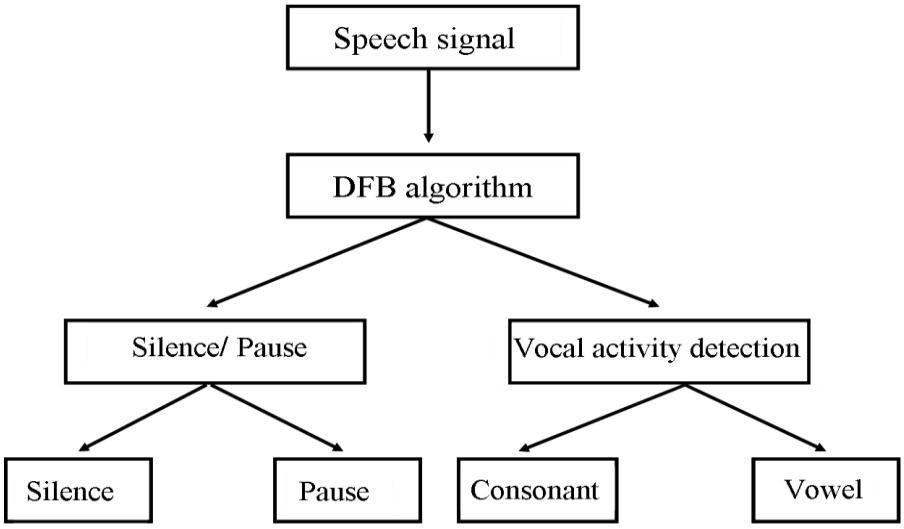

Figure 3. 3 : Schéma synoptique du système de segmentation vocalique.

3.3.3. Résultats et discussion

Nous avons calculé l'aphasiogramme de tous les patients avec l'aide de l'orthophoniste. Il nous a ainsi permis de bien connaître nos sujets et donc de décider pour le processus de déroulement de nos expériences.

Le tableau3.1, représenté graphiquement par la figure 3.4, montre le pourcentage des répétitions exactes à différents niveaux linguistiques (syllabes, logatomes...). Pour cette investigation primaire, nous avons sélectionné neuf patients désignés par PM^{ieme}.

Tableau 3. 1: le Pourcentage des correctes répétitions des patients aphasiques.

Type de tache de répétition	PM1	PM2	PM3	PM4	PM5	PM6	PM7	PM8	PM9
Syllabes	**77,77**	**88,88**	**100**	**100**	73,33	**44,44**	**18,18**	86,66	95,55
Logatomes	**88,09**	**85,77**	**93,00**	**100**	54,76	**33,33**	**00,00**	88,09	92,85
Mots simples	**99,88**	**90,00**	**97,67**	**100**	77,9	**49,46**	**14,77**	85,71	96,00
Mots complexes	**100**	**20,00**	**83,33**	**100**	88,89	**66,66**	**0**	100	--
phrases	**30,00**	**10,00**	**81,81**	**60,00**	40,00	**22,22**	**--**	44,44	--

En fait, une première lecture de l'aphasiogramme nous laisse observer que seuls les patients PM6 et PM7 ont un faible pourcentage de répétitions correctes dans l'ensemble des tâches de répétitions. Aussi, nous pouvons voir que la répétition des phrases est la tâche la

plus difficile pour tous les patients, sauf les patients PM3 et PM4 qui ont un pourcentage relativement élevé (PM3 et PM4 sont des patients presque guéris). Toutefois, il convient de préciser que l'aphasique ne récupère pas complètement ses facultés linguistiques.

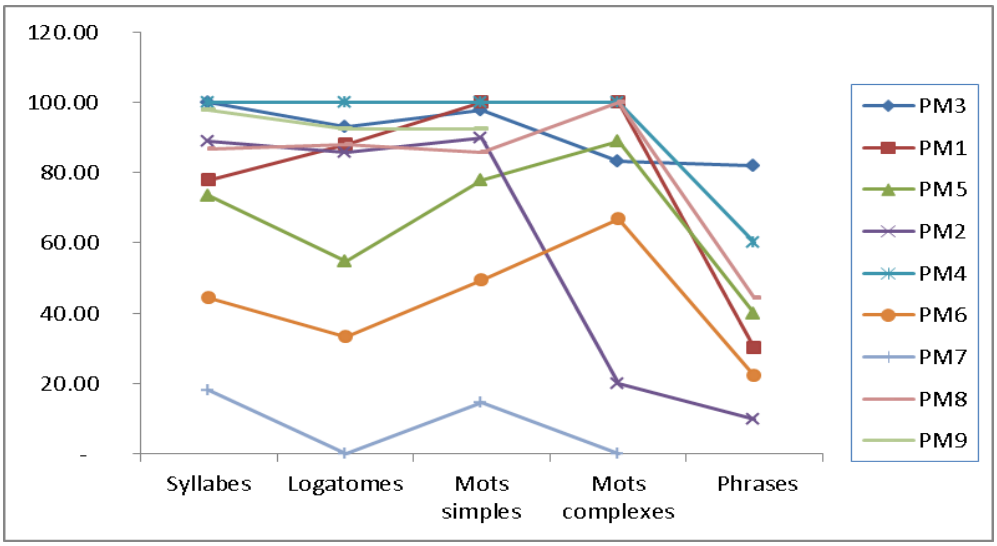

Figure 3. 4 : l'aphasiogramme des patients participants aux expériences prosodiques

Pour faire des comparaisons efficaces, deux listes expérimentales ont été générées : l'une contenant les patients convergents et l'autre contenant les patients divergents. Cette classification est faite d'après leurs réalisations en pourcentage pour la tache de répétions (voir tableau 3.1 et figure 3.4). Donc deux patients convergents dans la tache de répétions des syllabes par exemple sont deux patients qui ont deux taux des correctes répétions proches. Alors que les patients divergents ont des taux des correctes répétions divergents. L'organisation des patients participant aux expérimentations prosodiques est donnée par le tableau 3.2.

Tableau 3. 2 : Les patients participants dans les expériences prosodiques

Type de tache de répétions	Patients	
	Convergents	Divergents
Syllabes	PM3 et PM4	PM3 et PM7
Mots Complexes	PM3 et PM1	PM2 et PM6
Phrases	PM3 et PM1	PM3 et PM6

3.3.3.1. Étude de la fréquence Fo :

Les résultats de la variabilité de F_0 capturés par la moyenne, la variance et l'étendue sont représentés par les tableaux 3.3, 3.4 et 3.5 ainsi que par les figures 3.5 et 3.6. Ces résultats sont très diversifiés et n'indiquent pas quelque chose de spécial pouvant faire la différence entre patients convergents et patients divergents. Par conséquent, nous ne pouvons pas conclure en nous basant sur ce paramètre, mais il sera intéressant d'étudier à nouveau la fréquence fondamentale avec plus de données.

Tableau 3. 3 : Les statistiques du pitch des syllabes

Pitch		Syllabes			
		Patients Convergents		Patients Divergents	
		Patient PM3	Patient PM4	Patient PM3	Patient PM7
Moyenne	Min	107,10	153,48	107,10	122,43
	Moy	123,48	191,39	124,34	153,75
	Max	182,97	224,71	182,98	223,46
Variance	Min	5,82	54,66	5,82	17,93
	Moy	551,92	948,63	574,31	1006,23
	Max	6225,23	4450,22	6225,23	18787,21
Étendue	Min	6,37	24,66	6,37	17,18
	Moy	48,66	107,60	48,86	68,42
	Max	178,73	270,09	178,73	383,87

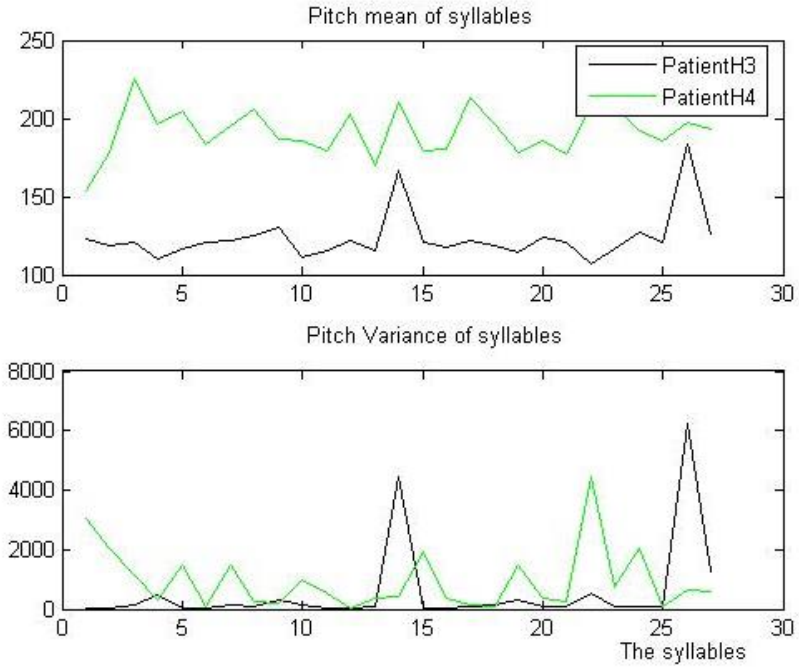

Figure 3. 5 : La moyenne et la variance du pitch des syllabes des patients convergents PM3 et PM4

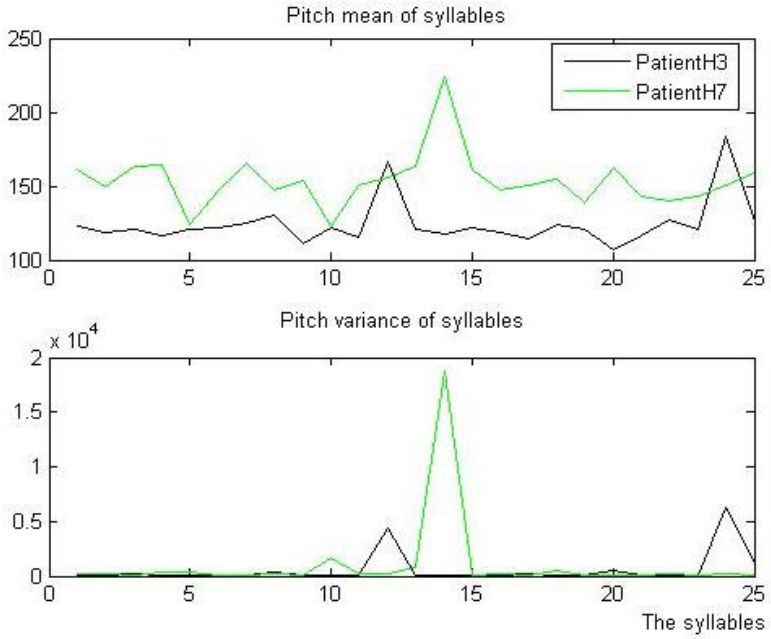

Figure 3. 6 : La moyenne et la variance du pitch des syllabes des patients divergents PM3 et PM7

Tableau 3. 4 : les statistiques du pitch des mots complexes

Pitch		Les mots complexes			
		Patients Convergents		Patients Divergents	
		Patient PM3	Patient PM1	Patient PM2	Patient PM6
Moyenne	Min	111,34	106,38	162,79	136,34
	Moy	133,91	114,38	214,80	159,41
	Max	222,10	126,62	250,29	184,82
Variance	Min	42,11	41,82	596,39	92,46
	Moy	1250,76	248,45	1391,09	315,29
	Max	7031,81	454,88	3923,47	551,90
Étendue	Min	21,59	30,44	84,31	32,83
	Moy	57,65	59,13	119,97	66,94
	Max	169,75	89,11	198,16	107,82

Tableau 3. 5 : Les statistiques du pitch des phrases

Pitch		Les phrases			
		Patients Convergents		Patients Divergents	
		Patient PM3	Patient PM4	Patient PM3	Patient PM6
Moyenne	Min	116,89	139,68	116,89	122,83
	Moy	142,82	186,17	139,10	165,56
	Max	192,11	250,40	176,65	219,70
Variance	Min	112,65	141,40	112,65	484,93
	Moy	1587,02	3223,26	1839,56	2348,84
	Max	4961,30	15528,82	4961,30	4780,97
Étendue	Min	37,83	51,79	37,83	72,55
	Moy	124,41	164,32	147,44	210,69
	Max	228,91	366,08	228,91	387,16

3.3.3.2. Étude de l'intonation

Contrairement à la fréquence F_0, l'intonation semble être un paramètre très pertinent comme nous pouvons le constater par l'analyse des tableaux 3.6, 3.7 et 3.8. Nous pouvons constater que pour les patients convergents, nous avons des pourcentages d'intonation proches et des pourcentages éloignés pour les patients divergents. Le taux de suivi donne une information précieuse indiquant fidèlement ce qui se passe réellement. En fait, il nous informe que malgré les différents pourcentages de l'intonation, les patients PM3 et PM4 sont convergents vraiment. Le taux FR est toujours plus élevé pour les patients convergents que pour les patients divergents.

Tableau 3. 6: les statistiques de l'intonation des syllabes

Intonation	Les syllabes			
	patients Convergents		patients Divergents	
	Patient PM3	Patient PM4	Patient PM3	Patient PM7
a (Positif)	33,33%	37,04%	4,00%	36,00%
a (Négatif)	66,67%	62,96%	96,00%	64,00%
Coefficient de suivi	74,07		60,00	

Tableau 3. 7: les statistiques de l'intonation des mots complexes

Intonation	Les mots Complexes			
	patients Convergents		patients Divergents	
	Patient PM3	Patient PM1	Patient PM2	Patient PM6
a (Positif)	16,67%	33,33%	12,50%	25,00%
a (Négatif)	83,33%	66,67%	87,50%	75,00%
Coefficient de suivi	83,33%		62,5%	

Tableau 3. 8 : les statistiques de l'intonation des phrases

Intonation	Les phrases			
	patients Convergents		patients Divergents	
	Patient PM3	Patient PM4	Patient PM3	Patient PM6
a (Positif)	14,29%	42,86%	40,00%	40,00%
a (Négatif)	85,71%	57,14%	60,00%	60,00%
Coefficient de suivi	71,43 %		60%	

3.3.3.3. Étude du débit

En ce qui concerne le débit de la parole exprimée par le rapport (SR). Les valeurs : minimale, moyenne et maximale ont été extraites. Elles sont données par les tableaux 3.9, 3.10 et 3.11. La variation du débit est représentée par la figure 3.7 à la figure 3.12.

Tableau 3. 9 : Les statistiques du taux de débit des syllabes

Débit		Syllabes			
		Patients Convergents		Patients Divergent	
		Patient PM3	Patient PM4	Patient PM3	Patient PM7
SR	Min	0,13	0,13	0,13	0,40
	Moy	0,63	0,68	0,62	0,76
	Max	1,00	1,00	1,00	1,00

Pour la répétition des syllabes et les mots complexes, nous pouvons voir dans les cas convergents étudiés que les courbes de taux débit se suivent et se rapprochent étroitement (Voir les figures 3.7, 3.9). Pour les cas divergents, les courbes de taux de débit s'éloignent l'une de l'autre et ne se suivent pas. (Voir les figures 3.8, 3.10). Pour les statistiques (minimum, moyen et maximums), exposées par tableaux 3.9 et 3.10, les variations sont complètement divergentes et ne se suivent pas.

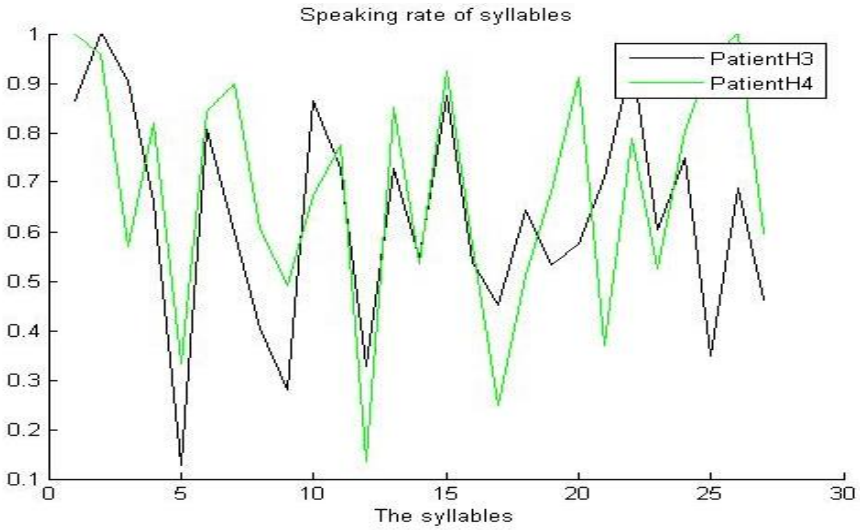

Figure 3. 7: Taux de débit des syllabes des patients convergents PM3 et PM4

Pour la variation SR des phrases, nous remarquons que les courbes sont hautement divergentes au niveau de la troisième et quatrième phrase pour les patients convergents PM3 et PM4 (voir figure 3.11). Cette anomalie est due à la présence d'une interaction entre le thérapeute et le patient au cours de la répétition des phrases concernées. Les données statistiques étaient très proches, soit pour les patients divergents ou pour les patients convergents (Voir tableau 3.11).

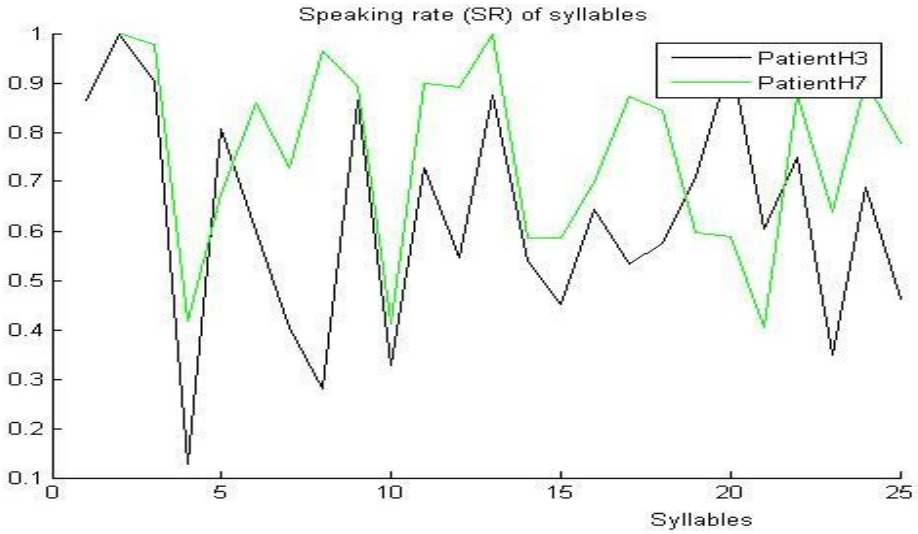

Figure 3. 8 : Taux de débit des syllabes des patients divergents PM3 et PM7

Tableau 3. 10 : Les statistiques du taux de débit des mots complexes

Débit		Les mots complexes			
		Patients Convergents		Patients Divergents	
		Patient PM3	Patient PM1	Patient PM2	Patient PM6
SR	Min	0,12	0,24	0,32	0,31
	Moy	0,35	0,46	0,47	0,51
	Max	0,57	0,75	0,71	0,75

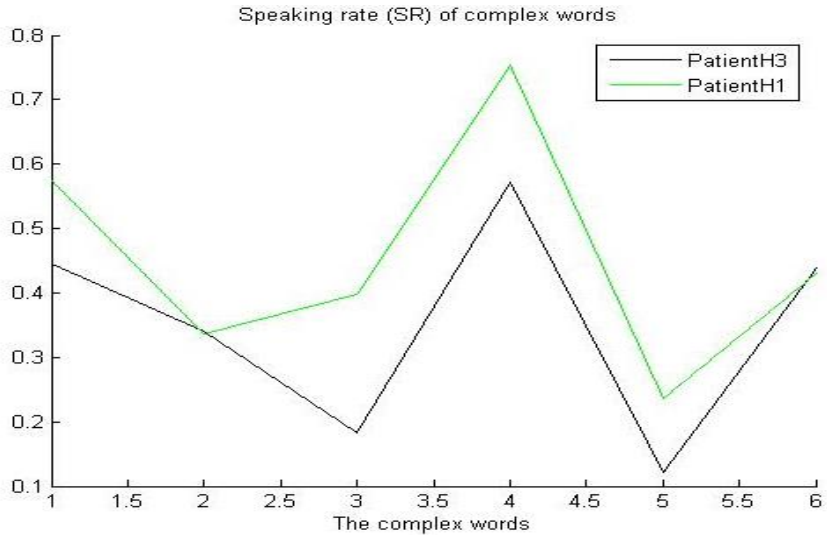

Figure 3. 9 : Taux de débit des mots complexes des deux patients convergents PM1 et PM3.

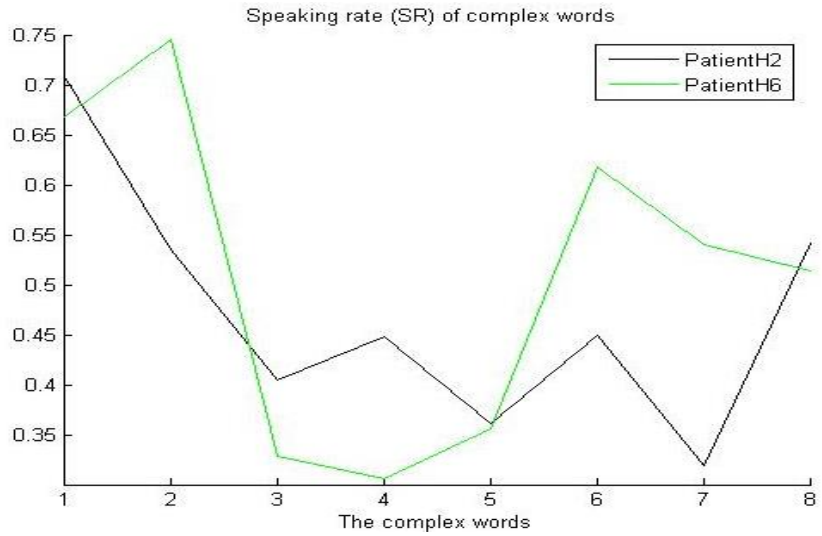

Figure 3. 10 : Taux de débit des mots complexes des deux patients divergents PM2 et PM6.

Tableau 3. 11 : Les statistiques du taux de débit pour les phrases

débit		Les phrases			
		Patients Convergents		Patients Divergents	
		Patient PM3	Patient PM4	Patient PM3	Patient PM6
SR	Min	0,33	0,22	0,36	0,36
	Moy	0,60	0,50	0,60	0,60
	Max	0,83	0,71	0,83	0,76

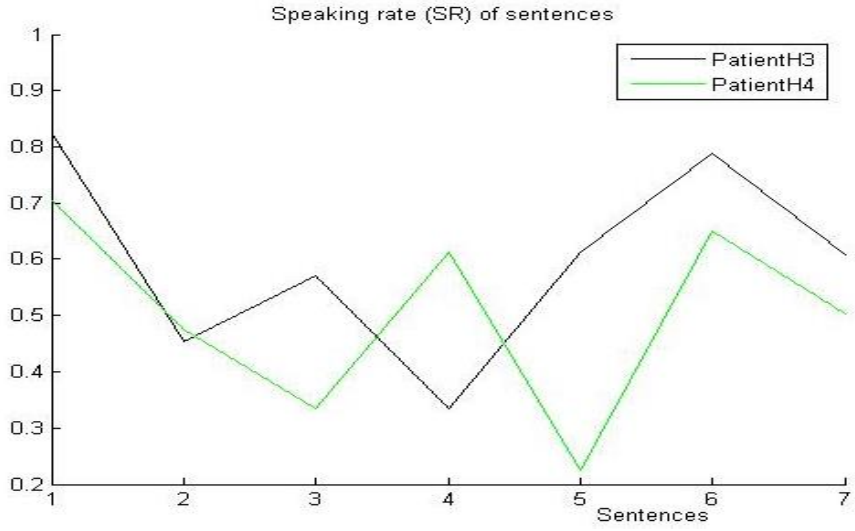

Figure 3. 11 : Taux de débit des phrases des patients convergents PM3 et PM4

Nous devons expliquer que les phrases constituaient la plus difficile épreuve de répétition pour tous les patients aphasiques soit pour ceux qui étaient au stade initiale ou ceux qui étaient au stade finale de la thérapie. C'est pourquoi l'épreuve considérée constitue

vraiment un indicateur de la récupération des patients. Par conséquent, nous avons décidé d'utiliser des phrases plus ou moins complexes relativement à l'évolution des cas étudiés. Notons enfin, que la répétition des phrases peut être plus utile que les autres tests pour l'étude des troubles de la dysprosodie (la dysprosodie fait référence à un trouble dans lequel une ou plusieurs des fonctions prosodiques sont soit compromises ou éliminés complètement [106]). En effet, la maladie est plus marquée si les stimuli sont plus longs et plus complexes.

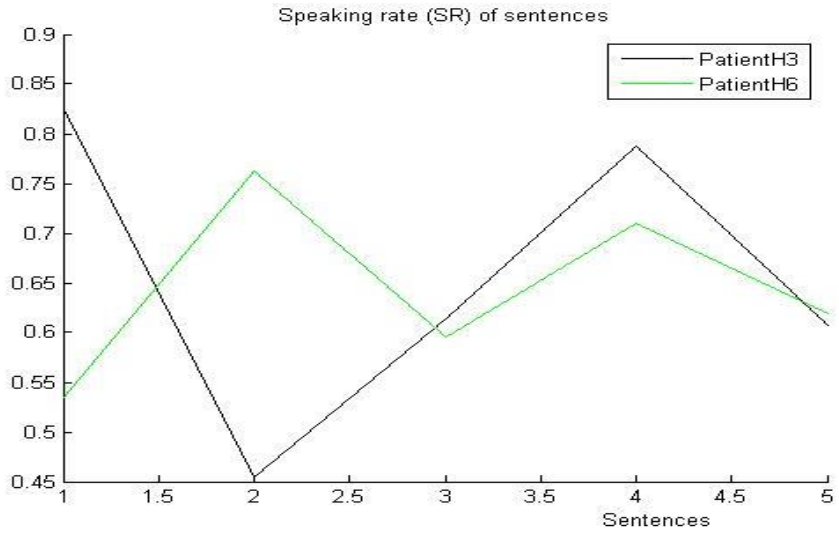

Figure 3. 12 : Taux de débit des phrases des patients divergents PM3 et PM4.

3.4. Analyse spectrale :

3.4.1. Analyse MFCC.

Pour la parole, une analyse cepstrale (figure 3.13) selon une échelle Mel est effectuée. Pour chaque trame une accentuation des aigus est réalisée (car les aigus sont toujours plus faibles en intensité que les graves) ; le fenêtrage est effectué pour garantir la stationnarité du signal de parole et pour atténuer les discontinuités. La fenêtre de Hamming est classiquement employée en parole.

Les énergies dans 24 filtres sont calculées après application du module de la FFT (Transformée de Fourier Discrète Rapide). Ces canaux sont répartis sur l'échelle de Mel : le nombre de coefficients utilisés pour décrire le signal est réduit tout en respectant une définition suffisante pour les fréquences qu'elles soient basses ou hautes, en respectant la perception humaine.

Précisons ci-dessous les quelques détails pratiques de mise en œuvre.

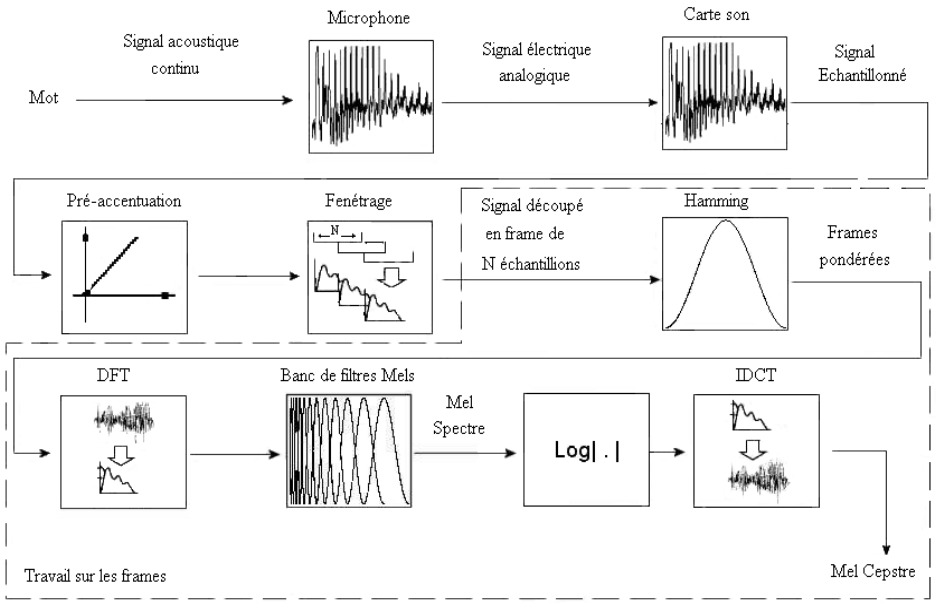

Figure 3. 13 : Analyse Spectrale

Les énergies des 24 canaux sont obtenues par l'application de filtres triangulaires se chevauchant et centrés sur les fréquences suivantes : 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1150, 1300, 1500, 1700, 2000, 2350, 2700, 3100, 3550, 4000, 4500, 5050, 5600, 6200 et 6850 Hertz.

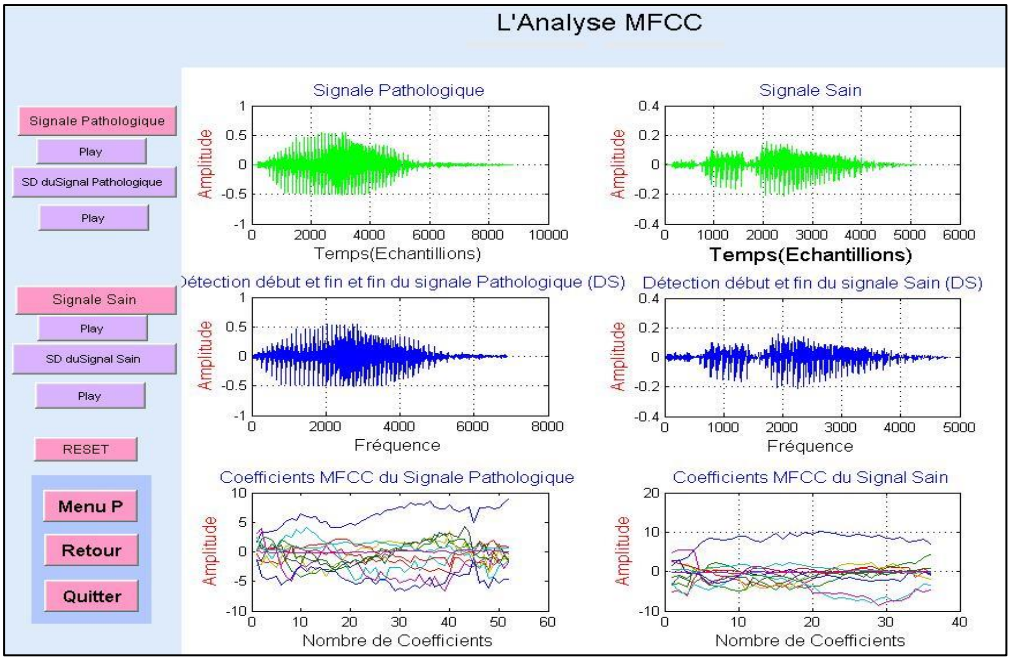

Figure 3. 14 : Analyse MFCC pour la parole aphasique : le logatome " cro"

L'analyse de chaque trame du signal donne un vecteur d'observation composé de 26 paramètres :

- L'énergie ;
- 12 coefficients cepstraux (MFCC) ;
- La dérivée de l'énergie ;
- 12 dérivées des coefficients cepstraux.

Cette suite de vecteurs subit ensuite une soustraction cepstrale. Chaque coefficient cepstral est diminué de la valeur moyenne des coefficients cepstraux. Ce traitement permet d'assurer une relative indépendance vis à vis du canal de transmission (microphone, ligne téléphonique.).

3.4.2. Analyse par spectrogramme

Un spectrogramme permet de visualiser l'évolution de l'énergie dans l'échelle des fréquences en fonction du temps. Le premier axe, l'axe horizontal (l'abscisse) représente l'axe du temps. Le second axe, l'axe vertical (l'ordonnée) représente l'axe des fréquences. Les bandes noires ou zones de noirceur correspondent à des zones de concentration d'énergie, Sur un spectrogramme, on peut donc repérer visuellement la plupart des corrélats acoustiques qui correspondent à la réalisation des phonèmes (silence, les bruits impulsionnels ou continus des sons apériodiques, les formants et leurs mouvements, etc.).

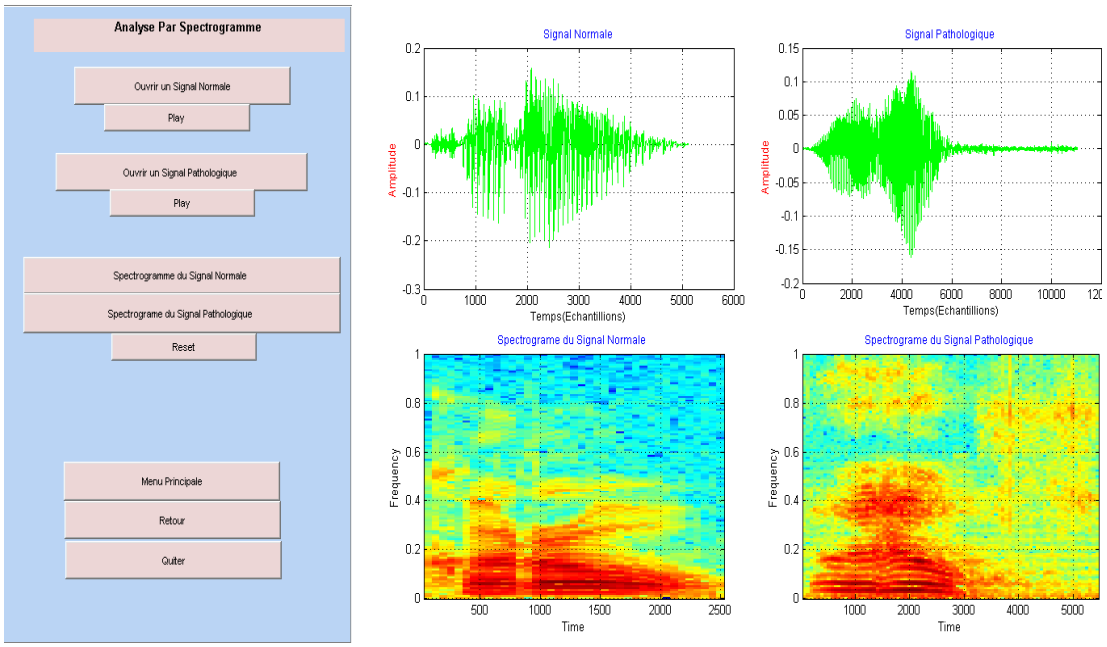

Figure 3. 15 : Analyse par spectrogramme.

Les étapes du calcul d'un spectrogramme sont :

- DFT d'une fenêtre de 4 à 32 ms qui se déplace de la moitié de sa durée. Avec une fréquence d'´échantillonnage de 16kHz, cette fenêtre a donc entre 64 et 512 points.
- Pour lisser le spectre on utilise en fait une DFT avec plus de points (au moins 256) ce qui permet d'interpoler le spectre plus finement.
- « zeropadding » : le signal de départ complété par des zéros. Si on utilise une DFT de 512 points et que la fenêtre a 64 points on complète par (512 - 64) zéros.
- Fenêtre plus courte que la période fondamentale (spectre à large bande : filtre de 300Hz ou plus) implique fort lissage fréquentiel et pas d'harmonique et donc permettent de visualiser les formants.
- Fenêtre plus longue (spectre à bande étroite : filtre de 45 Hz) implique faible lissage fréquentiel et visualisation des harmoniques (et donc F_0).

3.5. Analyse rythmique de la parole dysarthrique

3.5.1. Les paramètres de rythme

Dans cette étude, nous posons comme hypothèse que l'introduction de la métrique du rythme (les paramètres de Ramus et Grabe) [44, 94-96, 107] peut améliorer la précision du diagnostic et aider à identifier et classer les patients dysarthriques selon le degré de la sévérité de la maladie et, par conséquent, fournir une plus grande fiabilité à l'évaluation de l'intelligibilité.

Par conséquent, un de nos objectifs de ce travail est d'évaluer la gravité dysarthrique grâce à l'utilisation des mesures de rythme. Ramus et al., dans [96] fondent leur approche quantitative du rythme sur les caractéristiques purement phonétiques du signal de parole. Ils ont mesuré les durées vocaliques et la durée des intervalles consonantiques et ils ont calculé la variable %V qu'est la proportion des intervalles vocaliques au sein de la phrase, calculée comme la somme des durées des intervalles vocaliques divisée par la durée totale de la phrase. La variable ΔV est l'écart-type des durées des intervalles vocaliques au sein de la phrase. La variable ΔC est l'écart-type des durées des intervalles consonantiques (appelé aussi intervocaliques) au sein de la phrase.

Les mesures de rythmes sont basées sur des mesures acoustiques de la durée des intervalles vocaliques et consonantiques dans la parole continue. Elles prennent en compte la

variabilité de ces durées, et elles peuvent être calculées sous des formes brutes ou normalisées. Une liste de paramètres rythmiques utilisée dans nos expériences est donnée à la fin de cette section.

Ramus et al., ont soutenu que la combinaison de %V et soit ΔV ou ΔC fournissent le meilleur corrélat acoustique de classes rythmiques des langues. Ramus dans [97] a proposé qu'une procédure de normalisation peut être nécessaire à la fois pour ΔV et ΔC. À l'issue de cette discussion, nous constatons que les paramètres proposés pour quantifier le rythme à travers le signal acoustique rendent compte de certaines propriétés phonologiques comme la réduction vocalique et la complexité syllabique.

Rappelons que l'objectif du modèle Pairwise Variability Index (PVI) [95] rejoint sur le principe celui de Ramus [96] dans la mesure où il cherche à mesurer la complexité syllabique et la réduction vocalique. Ce modèle est basé en effet sur la mesure des durées des voyelles et des durées des intervalles entre les voyelles (à l'exclusion des pauses) dans un énoncé. Néanmoins, l'approche est différente de celle de Ramus [96] puisque le PVI prend en compte le niveau de variabilité en mesurant la moyenne des différences entre 2 intervalles vocaliques et respectivement 2 intervalles intervocaliques (soit : rPVIV et rPVIC) successifs dans la phrase.

Le calcul du PVI dans sa version brute (raw PVI), est défini comme suit :

$$\mathbf{rPVI}=\frac{\sum_{k=1}^{N-1}|d_k-d_{k+1}|}{N-1} \qquad (3.8)$$

N correspond au nombre d'intervalles vocaliques ou intervocaliques, et d est la durée de l'intervalle k. Nous avons également calculé le PVI dans sa version normalisée proposée par Grabe et Low [95] :

$$\mathbf{nPVI}=\frac{\sum_{k=1}^{N-1}\left|\frac{d_k-d_{k+1}}{d_k+d_{k+1}}\right|}{N-1} \qquad (3.9)$$

Dellwo [108] propose un nouveau paramètre « VarcoC », qui se base sur le calcul de la variation des intervalles consonantiques, c'est-à-dire l'écart type des durées des intervalles consonantiques divisé par la moyenne des durées des consonnes, multiplié par 100. Dellwo [108] teste VarcoC et ΔC afin de discriminer l'anglais et l'allemand du français. Il observe

que VarcoC discrimine clairement les langues en question et ce pour les différents types de tempo analysés. VarcoC varie en fonction des temps analysés mais la corrélation n'est pas systématique selon la langue. Cette nouvelle mesure d'intervalles normalisés est reprise par [107] et appliquée également aux voyelles.

Liss et al., [109] ont proposé un nouveau paramètre rythmique normalisé en fonction des durées combinées de successifs intervalles vocaliques et consonantiques désignés par VC, comme une approximation de la durée d'une syllabe : VarcoVC, l'écart type normalisé, nPVI-VC et rPVI-VC, les indices de variabilité par paires normalisés et non normalisés respectivement.

Nous avons pensé que cette approximation pour la durée d'une syllabe par la durée d'intervalle successif de voyelle et consonnes (exemple VC, VCC) peut être utilisé pour définir un nouveau paramètre de débit de parole (Speaking Rate : SR) au lieu de la définition classique du débit (le nombre de syllabes/la durée totale de l'énoncé de parole étudié). Le débit de parole est appelé SRVC est donné par la formule suivante :

$$\mathbf{SRVC} = \frac{\mathbf{N_{VC}}}{\mathbf{T_P}} \qquad (3.10)$$

Où :

N_{vc} : le nombre d'intervalles VC ; T_p : la durée totale de la phrase.

Enfin, Dellow [110] supposent que la quantification peut se faire en utilisant les segments voisés et non voisés au lieu des segments vocaliques. Par conséquent, nous avons refait le calcul de tous les paramètres évoqués en utilisant la segmentation voisée/non voisée. (Voir tableaux 3.12 et 3.13).

Tableau 3. 12 : Définitions des paramètres de rythme calculés pour la segmentation vocalique.

Paramètre	Description
%V	Le pourcentage des intervalles vocalique par rapport à la durée totale de la parole étudiée
ΔV	L'écart type des intervalles vocaliques
ΔC	L'écart type des intervalles consonantiques dans l'énoncé de la parole étudié
nPVI-V	L'index normalisé de variabilité par paire pour les intervalles vocaliques
rPVI-V	L'index non- normalisé de variabilité par paire pour les intervalles vocaliques
nPVI-C	L'index normalisé de variabilité par paire pour les intervalles consonantiques
rPVI-C	L'index non-normalisé de variabilité par paire pour les intervalles consonantiques

VarcoV	L'écart type devisée par la moyenne des intervalles vocaliques multiplié par 100
VarcoC	L'écart type devisé par la moyenne des intervalles consonantique multiplié par 100.
VarcoVC	L'écart type devisé par la moyenne des intervalles VC multiplié par 100.
nPVI-VC	L'index normalisé de variabilité par paire pour les intervalles VC
rPVI-VC	L'index non- normalisé de variabilité par paire pour les intervalles VC
SRVC	Le débit approximatif des intervalles VC : le nombre d'intervalles VC par la durée totale de l'énoncé de parole étudié.

Tableau 3. 13 : Définitions des paramètres de rythme calculés pour la segmentation voisée.

Paramètre	Description
%VO	Le pourcentage des intervalles voisés par rapport à la durée totale de l'énoncé de parole étudié
ΔVO	L'écart type des intervalles voisés dans l'énoncé de parole étudié
ΔUV	L'écart type des intervalles non-voisés dans l'énoncé de parole étudié
nPVI-VO	L'index normalisé de variabilité par paire pour les intervalles voisés
rPVI-VO	L'index non- normalisé de variabilité par paire pour les intervalles voisés
nPVI-UV	L'index normalisé de variabilité par paire pour les intervalles non- voisés
rPVI-UV	L'index non-normalisé de variabilité par paire pour les intervalles non- voisés
VarcoVO	L'écart type devisée par la moyenne des intervalles voisés multiplié par 100
VarcoUV	L'écart type devisé par la moyenne des intervalles non- voisés multiplié par 100.
VarcoVOUV	L'écart type devisé par la moyenne des intervalles VoUV multiplié par 100.
nPVI-VOUV	L'index normalisé de variabilité par paire pour les intervalles VOUV
rPVI-VOUV	L'index non- normalisé de variabilité par paire pour les intervalles VOUV
SRVOUV	Le débit approximatif des intervalles VOUV : le nombre d'intervalles VOUV par la durée totale de l'énoncé de parole étudié.

3.5.2. Analyse statistique : résultats expérimentaux et discussion

Avant de procéder à la classification des données, une analyse statistique est plus ou moins indispensable pour réduire la quantité des données à traiter et gagner plus en matière de temps de traitement. L'analyse d'ANOVA ou de Kruskal-Wallis permet de connaitre les paramètres significatifs pour discriminer une catégorie de l'autre alors que l'analyse DFA ou la régression logistique (LR method : **L**ogistic **R**egression method) permet de déterminer l'ensemble de paramètres les plus pertinents à la classification de la parole pathologique.

Nous avons mené plusieurs expériences sur la parole dysarthrique. Mais dans ce qui suit, nous allons exposer que les deux études publiés sur la classification parole dysarthrique

par rapport à la parole normale et sur la classification des niveaux de sévérité de la dysarthrie de la base de Nemours [43].

3.5.2.1. Méthodes statistiques[111, 112]

a) L'analyse de Variance : ANOVA

L'analyse de la variance (terme souvent abrégé par le terme anglais ANOVA : *ANalysis **Of VA**riance*) est un test statistique permettant de vérifier que plusieurs échantillons sont issus d'une même population. Le but de l'analyse ANOVA est de tester les différences significatives entre les moyennes du groupe. Cela se fait par l'analyse des écarts types des groupes de populations.

L'objectif est de savoir si une variable numérique a des valeurs significativement différentes selon plusieurs catégories (i.e., selon les valeurs d'un facteur = la variable). Plus précisément, on teste l'hypothèse nulle : il y a au moins deux catégories pour lesquelles les moyennes de la variable sont différentes.

b) L'analyse de Kruskal-Wallis

Au lieu d'utiliser l'analyse de variance (One-WayAnalysis : ANOVA), une approche non paramétrique : le test de Kruskal-Wallis a été préféré. Le test de Kruskal-Wallis (KW) est utilisé lorsqu'il faut décider si *k* échantillons indépendants sont issus de la même population. C'est donc un test d'identité. Les échantillons peuvent avoir des nombres d'observations différents. Le test de Kruskal-Wallis est non paramétrique : il ne fait aucune hypothèse sur la forme des distributions sous-jacentes. Le test ne repose pas sur les valeurs de la variable quantitative observée, mais sur les rangs qu'elles occupent dans la distribution. On commence par remplacer les N observations par leur rang : la plus petite valeur est remplacée par 1, la suivante par 2, etc. La plus grande valeur est remplacée par N. (N = nombre total d'observations dans les t échantillons).

La statistique du test de Kruskal-Wallis est construite à partir des **moyennes** des rangs des observations dans les différents échantillons. On remarquera donc la similitude de ce test avec l'ANOVA classique. L'ANOVA compare les **moyennes** des échantillons mais comme elle ne considère que des distributions normales et de même variance, elle teste en fait l'hypothèse selon laquelle ces distributions sous-jacentes sont identiques. Kruskal-Wallis

abandonne l'hypothèse de normalité et compare les **moyennes** des observations dans les différents échantillons.

c) La régression logistique : LR[113]

L'analyse LR est la mieux adaptée à notre objectif d'étude : identifier les paramètres de rythme les plus significatifs. Un autre but important est la classification. Comme il n'y a que deux groupes (DP et HC), la régression logistique est régulièrement utilisée car elle nécessite moins d'hypothèses et elle est plus robuste statistiquement. La régression ordinaire permet d'analyser une réponse quantitative variable (par exemple dans notre cas : maladie ou non maladie) en fonction d'une ou plusieurs variables explicatives (par exemple les paramètres rythmiques). Autrement dit, cette technique est utilisée pour des études ayant pour but de vérifier si des variables indépendantes peuvent prédire une variable dépendante dichotomique. Contrairement à la régression multiple et à l'analyse discriminante, cette technique n'exige pas une distribution normale des prédicteurs ni l'homogénéité des variances. Différents types de régression logistique existent, possédant chacun leur procédé statistique et conduisant à l'élaboration de différents modèles théoriques. Ainsi, seront abordés les types direct, séquentiel et automatisé («stepwise»). Toutefois, cette technique s'applique uniquement à de grands échantillons. Les prédicteurs (variables indépendantes) peuvent être des variables dichotomiques ou continues.

Le fonctionnement consiste à calculer les coefficients de régression de façon itérative. Cela signifie que le programme informatique, à partir de certaines valeurs de départ pour b_0 et b_1 de l'équation (3.12), vérifiera si les logs chances estimés sont bien ajustés aux données, corrigera les coefficients, réexaminera le bon ajustement des valeurs estimées, etc., jusqu'à ce qu'aucune correction des coefficients ne puisse atteindre un meilleur résultat.

La régression logistique repose sur l'hypothèse fondamentale suivante :

$$\ln \hat{Y}/(1-\hat{Y}) = \sum_{i=0}^{j} b_i X_i \qquad (3.11)$$

Où b_i : les coefficients de la régression ; Xi : les variables explicatives de la régression et Y : la variable à prédire.

Cette équation correspond au Log naturel de la probabilité de faire partie d'un groupe divisée par la probabilité de ne pas faire partie du groupe. Également, les prédicteurs choisis

doivent être spécifiques à un groupe ou l'autre, puisqu'il s'agit d'une variable dichotomique. En d'autres mots, les catégories doivent être mutuellement exclusives et exhaustives, car un prédicteur ne peut pas appartenir aux deux groupes à la fois. Cette technique s'avère très sensible à la multicollinéarité entre les prédicteurs, celle-ci se vérifiant à l'aide d'une matrice de corrélation. Ainsi, il est nécessaire d'examiner les corrélations entre les prédicteurs avant de procéder à l'élaboration du modèle. Lorsque certains prédicteurs sont fortement corrélés entre eux, il est préférable d'en éliminer quelques-uns puisqu'il s'agit probablement de variables redondantes. Dans le même ordre d'idées, la régression logistique présume que les réponses des différents cas sont indépendants les uns des autres. Il est alors prétendu que chaque réponse provient de cas différents, c'est-à-dire non reliés. Ainsi, si les variables résultantes sont formées par la période de temps pendant laquelle les mesures sont prises (avant et après traitement) ou si les variables proviennent d'un groupe par couplage (chaque sujet du groupe expérimental est jumelé à un sujet du groupe contrôle), la régression logistique devient inappropriée compte tenu des erreurs probables de corrélations.

d) L'analyse discriminante DFA.

L'analyse discriminante consiste à distinguer deux groupes ou plus sur la base d'un ensemble de variables.

DFA fournit une fonction discriminante (ou un ensemble de fonctions discriminantes pour plus de deux groupes de données), basées sur des combinaisons linéaires des meilleurs variables prédictives. L'équation de la fonction discriminante serait calculée comme suit :

$$\mathbf{D} = \mathbf{a} + \boldsymbol{c_1 X_1} + \boldsymbol{c_2 X_2} + \ldots\ldots + \boldsymbol{c_n X_n} \qquad \textbf{(3. 12)}$$

Où D est le score discriminant pour un sujet ou un cas, a est une constante, X_1 à X_n représentent les variables prédictives et c_1, ..., c_n sont les coefficients de la fonction discriminante (poids) pour chaque variable explicative. La procédure choisit automatiquement une fonction initiale qui sépare les groupes autant que possible. Il choisit ensuite une seconde fonction qui n'est pas corrélée à la première fonction et fournit autant que possible une séparation supplémentaire. La procédure se poursuit ajoutant des fonctions de cette façon jusqu'à atteindre le nombre maximum de fonctions tel que déterminé par le nombre de prédicteurs et les catégories de la variable dépendante.

Wilks Lambda méthode[111] a été utilisée pour sélectionner les variables à entrer dans

l'équation en fonction de leur capacité à réduire les valeurs Wilks Lambda. À chaque étape, les variables sont entrées dans l'analyse en fonction de leur capacité à réduire les Wilks lambda.

3.5.2.2. Analyse statistique pour La classification de la parole : dysarthrique / normale

Le corpus utilisé pour l'analyse statistique est constitué de huit phrases. Il est désigné par Set1. Pour chaque Patient Dysarthrique (DP : **D**ysarthric **P**atient) et pour l'orthophoniste qui est considère ici comme locuteur contrôle (HC : **H**ealthy **C**ontrol).

L'annexe A donne la liste des phrases de chaque locuteur. Une segmentation manuelle en consonne et voyelle est faite à l'aide du programme « wavesurfer » en utilisant le spectrogramme des signaux parole et les règles de segmentation [114]. Cette tache de segmentation est couteuse ce qui explique la quantité limitée de la parole étudiée.

Un résumé de l'évaluation de Frenchay est donné par le tableau 3.14. La moyenne et l'écart type des durées des intervalles vocaliques et consonantiques sont donnés par la Figure 3.16.

Tableau 3. 14 : l'évaluation Frenchay de la dysarthrie pour les patients de la base de données Nemours[115].

Patients	Niveau de sévérité (%)	Intelligibilité
KS (PS)	-	1
SC(PS)	49.2	3
BV(PS)	42.5	3
BK(PS)	41.8	3
RK(PMO)	32.4	2
RL(PMO)	26.7	4
JF(PMO)	21.5	4
LL (PM)	15.6	6
BB(PM)	10.3	8
MH(PM)	7.9	6
FB(PM)	7.1	-

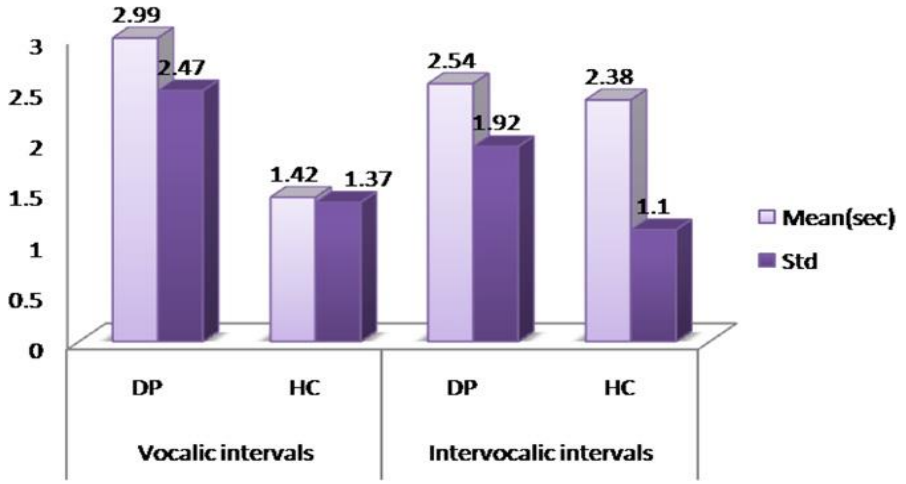

Figure 3. 16 : La moyenne (Mean) et la déviation standard (Std) des intervalles vocaliques et consonantiques

Ces mesures confirment clairement que les durées des deux intervalles sont plus élevées pour patients dysarthriques (DP) que pour le locuteur contrôle (HC). Ce résultat nous amène à effectuer une série d'expériences qui visent à créer des cartes bidimensionnelles (un paramètre par rapport à l'autre) qui pourrait être utile pour l'analyse de la pertinence de ces mesures. En outre, l'analyse de Kruskal-Wallis est également effectuée afin d'évaluer la signification statistique des paramètres rythmiques pour la classification de la parole dysarthrique contre la parole normale ou saine.

Le corpus utilisé pour extraire les intervalles voisés et non voisés est constitué de l'ensemble de toutes les phrases répétées par les patients et l'orthophoniste qu'est le sujet contrôle. Ce corpus est désigné par Set 3. Les mesures effectuées sur la durée des intervalles voisés et non voisés révèlent des différences notables entre les PD et HC. Comme le montre la figure 3.17, les patients dysarthriques ont tendance à produire des segments longs avec des valeurs supérieures de l'écart type que ceux du locuteur de référence. Les valeurs des intervalles produits par le patient KS (le cas le plus sévère) étaient de loin supérieures par rapport à celles des autres patients. Plus de détails sur les corpus utilisés (Set1 et Set3) se trouvent dans le chapitre suivant de l'évaluation objective de la parole dysarthrique.

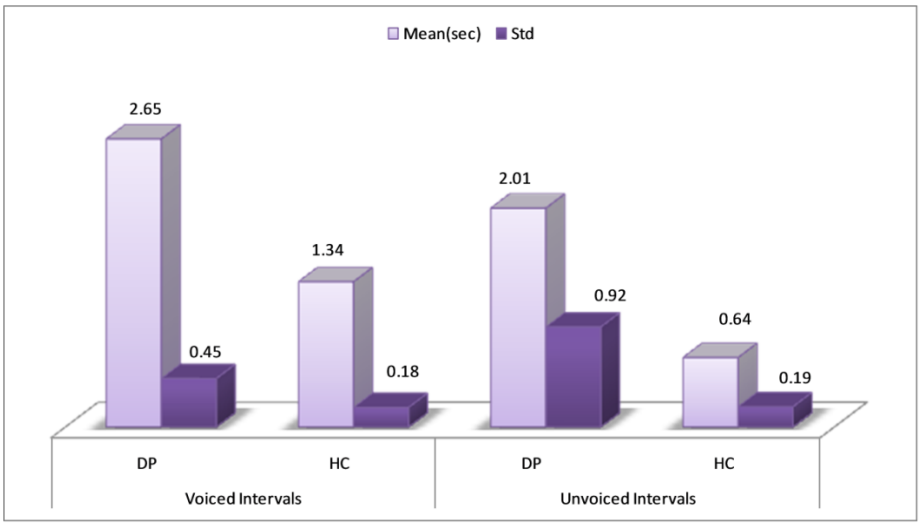

Figure 3. 17 : La moyenne (mean) et l'écart type (Std) des durées des intervalles voisés et non voisés.

Le test de Kruskal-Wallis avec un niveau de signification de 0,05 est effectué afin d'évaluer la signification statistique des paramètres de rythme pour les deux groupes de locuteurs DP et HC. Tableau 3.15 résume les résultats pour Set1 et Set3. Pour Set1 seulement, les paramètres suivants ne sont pas statistiquement significatifs : VarcoVC, nPVI-VC et % V. En ce qui concerne la segmentation voisée /non voisée, tous les paramètres sont statistiquement significatifs. Il est important de noter qu'il s'agit pratiquement du même résultat que nous avons obtenu en utilisant l'analyse ANOVA.

Par conséquent, une analyse de régression logistique est réalisée afin d'identifier la combinaison de paramètres qui prédisent le mieux la parole dysarthrique. À l'exclusion de la VarcoVC, nPVI-VC et V% pour les intervalles voyelles / consonnes, le critère de Wald a démontré que sept paramètres, parmi lesquels SRVC est le meilleur suivi par ΔC, nPVI-C, nPVI-V, rPVI-C, rPVI-VC et rPVI-V ont apporté une contribution significative à la prédiction de la parole dysarthrique.

Tableau 3. 15 : Résumé des résultats du test de Kruskal Wallis pour les deux paroles : dysarthrique et saine présentées par Set1 et Set 3.

Set 1			Set 3		
Paramètres	**Chi-Square**	**P**	**Paramètres**	**Chi-Square**	**P**
ΔC	8.229	.004	**ΔUV**	660.562	.000
ΔV	18.041	.000	**ΔVO**	593.302	.000
%V	1.039	.308	**%VO**	17.348	.000
nPVI-V	21.020	.000	**nPVI-VO**	56.073	.000

nPVI-C	16.062	.000	**nPVI-UV**	21.645	.000
rPVI-V	17.666	.000	**rPVI-VO**	501.605	.000
rPVI-C	6.437	.011	**rPVI-UV**	396.028	.000
VarcoV	21.101	.000	**VarcoVO**	143.818	.000
VarcoC	13.256	.000	**VarcoUV**	13.636	.000
nPVI-VC	2.741	.098	**nPVI-VOUV**	19.149	.000
rPVI-VC	17.147	.000	**rPVI-VOUV**	372.702	.000
VarcoVC	0.023	.879	**VarcoVOUV**	39.762	.000
SRVC	78.033	.000	**SRVOUV**	541.319	.000

En ce qui concerne la segmentation voisée et non voisée, huit prédicteurs sont déterminés :% SRVOUV, ΔUV, ΔVO, %VO, nPVI-UV, VarcoUV, rPVI-UV et rPVI-VO constituent un ensemble de paramètres pouvant distinguer de façon fiable entre les DP et HC (Chi-carré est : 1214,602, p = 0,000 avec 8 degrés de liberté).

Ce résultat nous conduit à mener une série d'expériences qui visent à créer des cartes bidimensionnelles (une métrique par rapport à l'autre) pouvant être utile pour l'analyse de la pertinence de ces mesures à la fois à la pratique clinique et à la recherche concernant la parole pathologique.

Dans la figure 3.18, l'espace des caractéristiques (nPVI-V, nPVI-C) montre que la parole normale (présentée par HC) est relativement bien regroupée autour des valeurs élevées de nPVI-V et les valeurs moyennes nPVI-C, tandis que les patients dysarthriques sont plus dispersés dans cet espace. Les positions particulières de RL, BK et RK sont assez surprenantes puisque le test Frenchay considère BK comme un cas grave alors que RL et RK les classe dans la catégorie modérée.

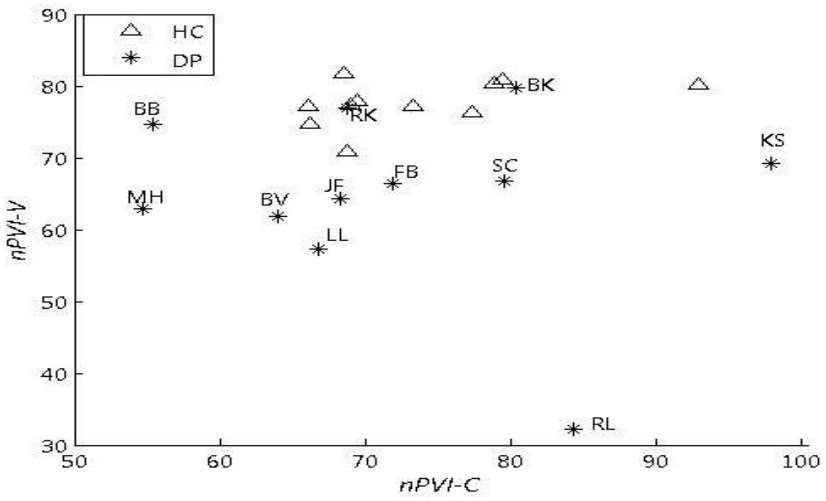

Figure 3. 18 : La distribution bidimensionnelle des DP et HC dans l'espace (nPVI-V, nPVI-C).

Nous pouvons remarquer de la figure 3.19 BV et BK qui sont considérés comme des cas graves et aussi les cas dysarthriques modérés sont proches de HC. Par conséquent, cette représentation semble insuffisante pour une bonne discrimination des sujets. Le cas le plus grave, KS, est bien discriminé par les deux représentations.

La figure 3.20, montre le tracé de la distribution 2-D de la DP et de HC dans le plan formé par l'écart type des intervalles non voisés et le pourcentage de la durée des intervalles voisés (ΔUV, VO%). Nous pouvons observer une distribution aléatoire des DP tandis que la parole HC est bien regroupée. Les cas les plus graves sont positionnés loin de HC. ΔUV relativement plus élevé. Le reste des DP sont près de HC. Dans toutes les représentations en 2-D que nous avons effectuées, nous avons constaté que BV, et BK dont les scores d'intelligibilité et de Frenchay respectivement sont 3 et 57, 5 et 58,2 (voir tableau 3.14) respectivement, sont toujours positionnés à proximité de la SC. En effet, en examinant la parole de BV, BK et du patient FB, nous avons constaté que la vitesse de la parole BV et BK était tout à fait normale et compréhensible dans la tâche répétition des phrases mais leurs lectures des paragraphes étaient lentes et difficiles à comprendre. Nous avons également noté que la parole de FB est moins altérée que celle des autres patients, mais elle n'est pas la plus proche à la parole de l'orthophoniste (c.-à-d. la parole de HC). En fait, son discours est très intelligible, mais son débit de parole est très lent.

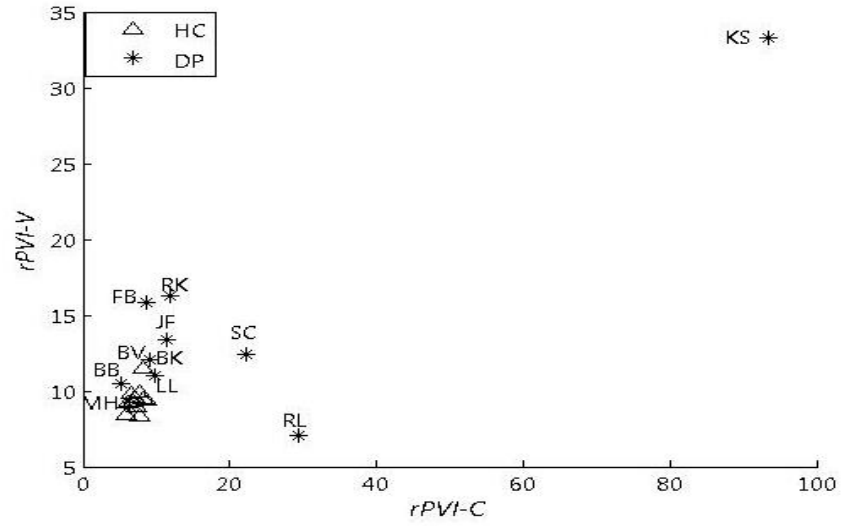

Figure 3. 19 : Les sujets dysarthriques et l'orthophoniste (HC) présentés dans l'espace paramètres (rPVI-C, rPVI-V).

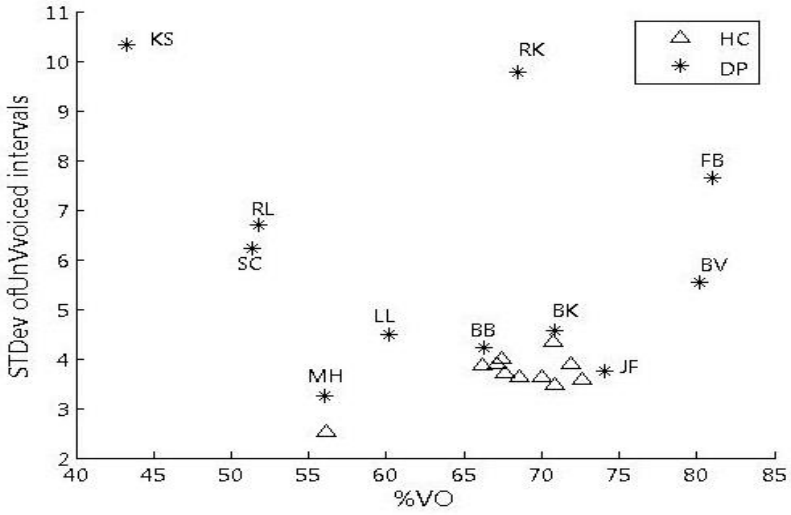

Figure 3. 20 : Les sujets Dysarthriques et l'orthophoniste (HC) représentés dans l'espace paramètres (%VO, ΔUV).

Nous avons également examiné les combinaisons du paramètre SRVC : le nouveau paramètre avec les autres paramètres. En utilisant les intervalles vocaliques et consonantiques les combinaisons sont : SRVC avec ΔC et ΔV. En utilisant les intervalles voisés et non voisés les combinaisons sont : SRVOUV avec ΔUV et rPVI-VO à travers les figures 3.21- 3.24.

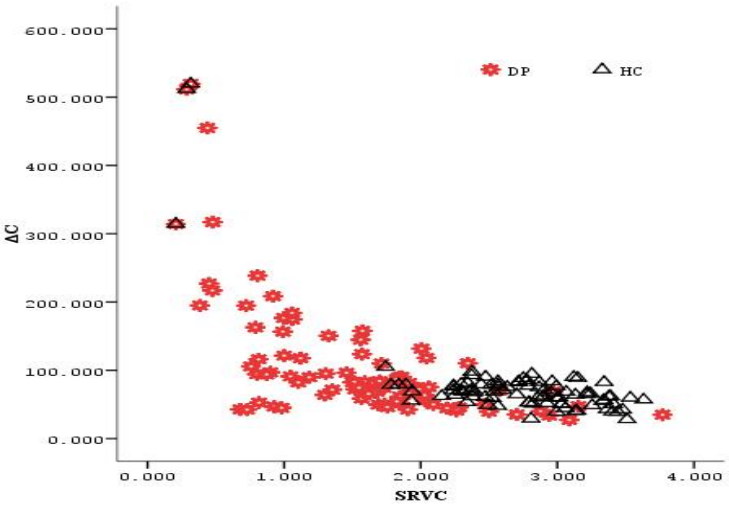

Figure 3. 21 : La distribution des DP et HC le long des deux dimensions : SRVC (l'axe des x) et ΔC (l'axe des y).

Figures 3.21 et 3.22 montrent la répartition des DP et HC le long des dimensions : SRVC (axe des x) et ΔV, respectivement ΔC (axe des y). Ces plans de paramètres montrent que les patients DP sont caractérisés par un niveau relativement élevé de ΔV et ΔC et un SRVC relativement faible. Les réalisations du thérapeute sont bien regroupées autour des valeurs nominales de ces paramètres.

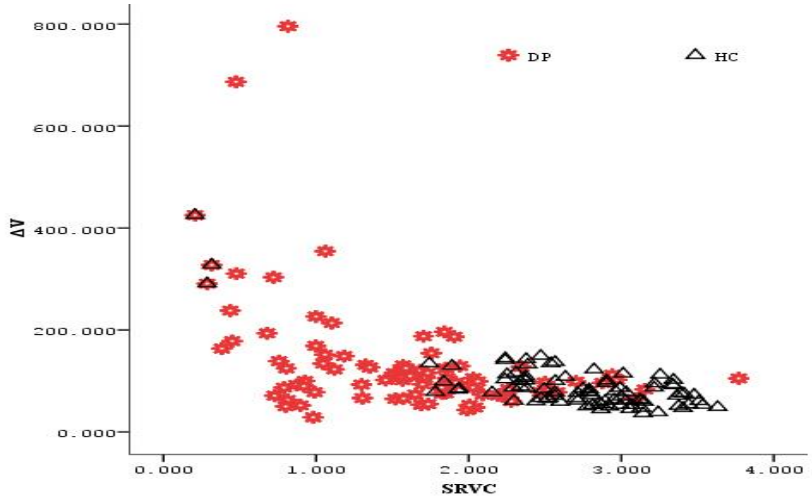

Figure 3. 22 : La distribution des DP et HC le long des deux dimensions : SRVC (l'axe des x) et ΔV (l'axe des y).

Figure 3.23 et 3.24 donnent La répartition 2-D des DP et de HC le long SRVOUV (l'axe des x) et ΔUV, rPVI-VO respectivement (l'axe des y) dimensions. Les patients dysarthriques peuvent être facilement distingués du HC. Mais, tout de même, nous constatons un chevauchement de la parole saine et dysarthrique. Nous pensons, que c'est du à la parole modérée.

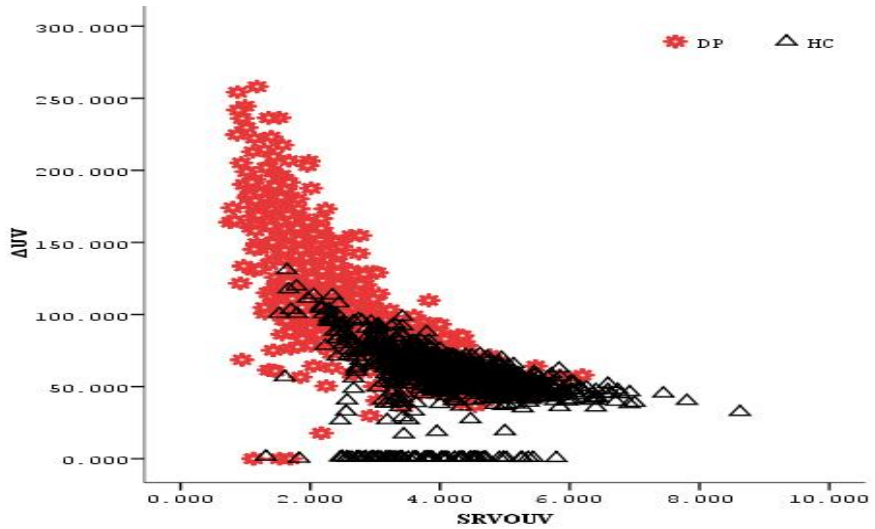

Figure 3. 23 : La distribution des DP et HC le long des deux dimensions : SRVOUV (l'axe des x) et ΔUV (l'axe des y).

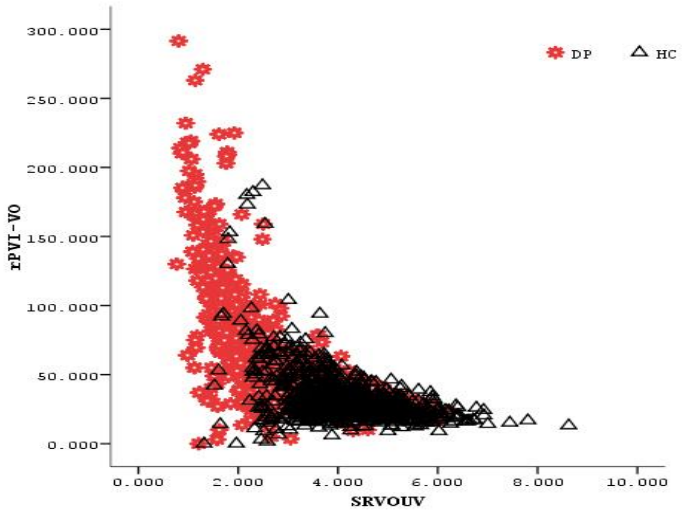

Figure 3. 24 : la distribution des DP et HC le long des deux dimensions : SRVOUV (l'axe des X) et rPVI-VO (l'axe des Y).

3.5.2.3. Analyse statistique pour la classification des niveaux de sévérité.

À ce niveau, nous introduisons un nouveau corpus désigné par Set 2. Nous voulons utiliser la segmentation vocalique automatique disponible avec la base de données Nemours et la comparer avec les autres types de segmentations (manuelle et automatique voisée). Ce corpus ne contient que la parole dysarthrique segmentée automatiquement (74 phrases pour chacun des 10 patients à l'exclusion du cas le plus grave, le patient KS). La segmentation a été développée en utilisant une segmentation automatique par HMM [116], mais les erreurs d'étiquetage n'ont pas été corrigées.

Les patients ont été codés selon le niveau de sévérité, comme indiqué dans le (**Tableau 2. 9: L'Évaluation Frenchay des Patients de Nemours (Frenchay Dysarthria Assessment : FDA)** : **P**atients **M**ild dont la parole est légèrement altérée (PM), les **P**atients a pathologie **MO**déré (PMO) et les **P**atients a pathologie **S**évère (PS) et la parole normale par la parole de l'orthophoniste (désignée toujours par HC). L'objectif était d'évaluer les niveaux de sévérité de la parole dysarthrique. En effet l'objectif est de déterminer si le trouble de la parole légèrement altérée se distingue de la parole normale, ou si la parole des patients à pathologie modérée qui se trouve entre la parole légèrement altérée et la parole sévèrement altérée pouvait être séparée ou distinguée de ces deux types de la parole pathologique. Les expériences ont été divisées en deux analyses. La première analyse a été réalisée avec la présence de la parole normale donc la présence du sujet de contrôle (l'orthophoniste : HC) et la deuxième analyse sans la parole normale. Les valeurs moyennes des paramètres

rythmiques sont présentées par les tableaux 3.16 et 3.117 (L'erreur standard de la moyenne en parenthèses (SE).

Tableau 3. 16 : les valeurs moyennes des paramètres rythmiques du Set 1 et Set 2 pour les différents niveaux de sévérité dysarthrique (l'erreur standard des valeurs moyenne en parenthèses).

	Set 1				Set 2		
Variable	**PM (SE)**	**PMO (SE)**	**PS (SE)**	**HC (SE)**	**PM (SE)**	**PMO (SE)**	**PS (SE)**
ΔC	62.53 (7.66)	92.18 (7.66)	111.31 (11.74)	79.06 (7.86)	80.80 (1.50)	120.78 (4.19)	106.42 (3.18)
ΔV	98.55 (5.40)	109.63 (14.10)	150.94 (30.42	90.94 (6.05)	112.42 (2.06)	138.95 (3.03)	139.69 (3.77)
%V	44.46 (1.12)	34.71 (1.78)	31.93 (2.14)	38.33 (0.66)	44.9 (0.53)	38.76 (0.68)	37.20 (0.87)
nPVI-V	65.29 (2.42)	57.43 (5.38)	69.43 (4.62)	77.39 (1.71)	72.97 (0.93)	69.77 (1.47)	79.54 (1.45)
nPVI-C	66.32 (2.25)	69.08 (3.92)	67.87 (4.16)	79.48 (1.62)	79.73 (0.86)	82.87 (1.47)	81.73 (1.25)
rPVI-V	118.37 (7.13)	121.80 (15.10)	161.65 (28.40)	102.61 (5.83)	127.77 (2.52)	162.93 (4.28)	169.63 (4.65)
rPVI-C	76.11 (4.42)	114.14 (10.90)	132.22 (15.77)	90.75 (6.30)	93.66 (1.93)	124.06 (4.43)	115.54 (3.70)
VarcoV	57.56 (2.52)	53.47 (5.39)	64.37 (5.13)	71.55 (1.78)	65.34 (0.89)	56.08 (1.13)	64.31 (1.10)
VarcoC	53.39 (2.13)	59.45 (4.03)	55.97 (2.91)	65.11 (1.25)	75.49(089)	86.84 (1.56)	90.09 (1.65)
nPVI_VC	55.38 (2.92)	58.51 (4.71)	78.19 (6.46)	69.28 (1.79)	64.57 (1.06)	60.44 (1.94)	71.86 (1.61)
rPVI_VC	212.10 (13.12)	334.51 (42.07)	644.41 (83.63)	264.42 (40.59)	220.84 (4.84)	275.08 (9.39)	300.71 (8.86)
VarcoVC	43.50 (1.68)	52.72 (3.42)	63.91 (4.82)	51.59 (1.36)	49.30 (0.77)	45.02 (1.18)	53.73 (1.10)
SRVC	2.15 (0.10)	1.40 (0.10)	1.32 (0.11)	2.68 (0.07)	2.08 (0.03)	1.28 (0.03)	1.49 (0.05)

Tableau 3. 17 : les valeurs moyennes des paramètres rythmiques du Set 3 pour les différents niveaux de sévérité dysarthrique (l'erreur standard des valeurs moyenne en parenthèses).

Variable	**PM (SE)**	**PMO (SE)**	**PS (SE)**	**HC (SE)**
ΔUV	92.64 (2.27)	95.34 (1.83)	100.27 (1.39)	55.29 (0.81)
ΔVO	53.72 (1.68)	75.78 (3.15)	70.85 (1.62)	30.64 (0.62)
%VO	69.31 (0.79)	63.60 (0.95)	66.77 (0.55)	68.16 (0.34)
nPVI-VO	90.91 (2.10)	90.91 (2.27)	93.95 (1.27)	84.46 (1.01)
nPVI-UV	110.63 (2.03)	105.51 (1.88)	110.95 (1.15)	104.70 (0.90)
rPVI-VO	57.10 (73.71)	62.84 (2.27)	63.90 (1.54)	33.19 (0.70)
rPVI-UV	73.71 (3.27)	67.11 (1.86)	101.55 (1.61)	44.60 (0.71)
VarcoVO	90.80 (2.24)	101.56 (2.58)	157.72 (1.27)	82.52 (1.01)
VarcoUV	157.22 (1.57)	150.92 (2.38)	78.43 (1.06)	153.75 (0.84)
nPVI_VOUV	73.91 (1.61)	77.31 (1.93)	98.05 (2.41)	73.22 (0.79)
rPVI_VOUV	85.09 (4.11)	96.45 (3.47)	98.05 (2.41)	56.42 (1.08)
VarcoVOUV	90.39 (1.66)	97.00 (1.90)	98.40 (1.23)	91.07 (0.85)
SRVOUV	3.06 (0.07)	2.69 (0.07)	2.77 (0.04)	4.14 (0.04)

a) Évaluation avec la présence de la parole normale :

La variable dépendante D'ANOVA est la sévérité. Cette dernière, montre une différence significative entre les quatre groupes pour tous les paramètres de rythme. Les résultats de l'ANOVA sont donnés par le tableau 3.18.

Tableau 3. 18 : Résumé des résultats d'ANOVA pour l'effet des niveaux de sévérité de la dysarthrie pour set 1 et set 3 avec la présence de la parole normale (HC).

Set 1			Set 3		
Var	**F(3,172)**	**P**	**Var**	**F(3,1612)**	**P**
ΔC	12.817	.000	**ΔUV**	420.405	.000
ΔV	11.577	.000	**ΔVO**	349.910	.000
%V	23.577	.000	**%VO**	32.175	.000
nPVI-V	8.985	.000	**nPVI-VO**	29.072	.000
nPVI-C	6.816	.000	**nPVI-UV**	15.304	.000
rPVI-V	11.054	.000	**rPVI-VO**	191.654	.000
rPVI-C	15.754	.000	**rPVI-UV**	160.976	.000
VarcoV	7.697	.000	**VarcoVO**	138.652	.000
VarcoC	5.697	.002	**VarcoUV**	12.404	.000
nPVI-VC	10.149	.000	**nPVI- VOUV**	17.673	.000
rPVI-VC	25.222	.000	**rPVI-VOUV**	160.390	.000
VarcoVC	11.920	.000	**VarcoVOUV**	55.790	.000
SRVC	68.338	.000	**SRVOUV**	306.609	.000

Le test post-hoc de Bonferroni pour Set 1 montre que ΔC, ΔV, rPVI-V, rPVI-C, nPVI-VC, VarcoVC et rPVI-VC ont un effet significatif pour séparer la dysarthrie légèrement altérée (les PM) et la dysarthrie modérément altérée(Les PMO). % V est statistiquement significative à séparer les cas légèrement altérés et les cas sévères des cas restants. nPVI-V, nPVI-C et VarcoV sont statistiquement significatives à séparer la parole saine de l'orthophoniste des patients dysarthriques PM et PMO. Le VarcoC est statistiquement significatif uniquement pour séparer la parole saine de la parole légèrement altérée. Le meilleur paramètre est SRVC, qui est statistiquement significatif pour séparer tous les niveaux de sévérité à l'exclusion de la distinction entre des deux niveaux : sévère et modéré. En ce qui concerne Set 3, des comparaisons multiples ont été effectuées pour tous les paramètres. Les deux paramètres, ΔVO et SRVOUV, sont statistiquement significatifs pour toutes les comparaisons entre les quatre groupes. ΔUV, nPVI-UV, nPVI-VO, nPVI-VOUV, VarcoVOUV et VarcoUV sont statistiquement significatifs à séparer les PS des PM, PMO et de l'orthophoniste. Pour séparer les groupes HC et PS des deux groupes restants, rPVI-VO, rPVI-UV VarcoVO et rPVI-VOUV sont statistiquement significatifs. %VO est statistiquement significatif uniquement à séparer les patients a parole légèrement altérée des patients appartenant aux niveaux modéré et sévère.

L'analyse discriminante pour Set 1 et Set 3 a été réalisée afin de déterminer les paramètres discriminants. Pour Set 1, cinq paramètres sont conçus pour être les plus pertinents, dont **SRVC** est le meilleur, suivi par **rPVI-VC, %V, VarcoV et VarcoC**. Pour

Set 3 (segmentation voisée/ non voisée), tous les paramètres de rythme sont sélectionnés en tant que paramètres prédicteurs. Mais seulement **nPVI-VO, rPVI-VOUV et nPVI-VOUV** sont exclus.

b) Évaluation sans la présence de la parole normale

Trois groupes de locuteurs dysarthriques ont été établis : locuteurs légèrement malades (Mild : PM), locuteurs à pathologie modérés (PMO) et locuteurs à pathologie sévères (SP). L'analyse Anova a été réalisée pour chacun des paramètres de rythme et a révélé l'importance statistique des paramètres suivants pour différencier entre les groupes de locuteurs de Set 1. Il s'agit de : ΔC, %V, rPVI-C, nPVI-VC, rPVI-VC, VarcoVC et SRVC. Pour Set 3 tous les paramètres sont statistiquement significatifs, sauf %VO. Pour Set 2, tous les paramètres sont statistiquement significatifs, à l'exception de nPVI-C. Tableau 3.18 présente la signification statistique des paramètres du rythme pour distinguer les différents niveaux de sévérité.

Un test post-hoc de Bonferroni pour les paramètres significatifs révèle pour Set 1 que ΔC, rPVI-C, %V et SRVC ont un effet significatif principal pour distinguer les patients PM des patients PS et PMO. Pour séparer les patients PM des deux autres groupes, ΔV,% V, RPVI-V, RPVI-C, VarcoC et RPVI-VC sont statistiquement significatifs. nPVI-V et nPVI-VC sont statistiquement significatifs pour distinguer les patients PS : des cas PM et PMO. Seulement VarcoV est statistiquement significatif à séparer le niveau modéré des niveaux légers et sévères. En ce qui concerne Set 3, les résultats montrent que deux paramètres, ΔVO et VarcoVO, sont statistiquement significatifs pour toutes les comparaisons entre les trois groupes. Les paramètres suivants : ΔUV, rPVI-UV, nPVI-VO, nPVI-VOUV, rPVI-VOUV, VarcoVO et VarcoUV sont statistiquement significatifs à séparer les locuteurs PS des locuteurs des deux niveaux restants. Pour séparer les locuteurs PS des locuteurs de PMO, nPVI-VO est statistiquement significatif. %VO est également statistiquement significatif en distinguant les patients PM des PMO. Les patients PS et PM peuvent être séparés les uns des autres par rPVI-VO, et le nouveau paramètre SRVOUV est statistiquement significatif pour séparer les patients PM des deux autres groupes.

Pour Set 1, trois paramètres ont été conçus pour être les plus pertinents par l'analyse DFA : SRVC (sa corrélation avec la première fonction canonique était 0.913), rPVI-VC et VarcoVC. Pour Set 2, sept paramètres ont été conçus pour être les prédicteurs les plus pertinents : ΔC, ΔV,% V, rPVI-C, VarcoC, VarcoV et VarcoVC. Pour Set 3, six mesures de rythme ont été sélectionnées en tant que paramètres de prédiction. Le paramètre le plus

pertinent avec la plus grande corrélation absolue dans la première fonction canonique était ΔVO, suivie par rPVI-VO, rPVI-UV, VarcoVO, VarcoUV et nPVI-VOUV.

Tableau 3. 19 : Résumé des résultats de l'analyse Anova pour la discrimination des différents niveaux de sévérité sans la présence de la parole normale (l'orthophoniste).

Set 1			Set 2		Set 3		
Var	**F(2,77)**	**P**	**F(2,738)**	**P**	**Var**	**F(2,731)**	**P**
ΔC	10.48	.000	62.27	.000	**ΔUV**	27.19	.000
ΔV	2.36	.101	33.20	.000	**ΔVO**	47.17	.000
%V	17.21	.000	34.20	.000	**%VO**	9.46	.000
nPVI-V	2.05	.135	13.50	.000	**nPVI-VO**	6.71	.001
nPVI-C	.17	.837	2.02	.133	**nPVI-UV**	7.46	.000
rPVI-V	1.82	.167	45.64	.000	**rPVI-VO**	10.68	.000
rPVI-C	8.05	.001	28.59	.000	**rPVI-UV**	8.65	.000
VarcoV	1.52	.224	17.96	.000	**VarcoVO**	21.94	.000
VarcoC	1.07	.348	42.18	.000	**VarcoUV**	9.63	.000
nPVI-VC	6.93	.002	13.12	.000	**nPVI-VOUV**	10.91	.000
rPVI-VC	20.21	.000	39.25	.000	**rPVI-VOUV**	15.60	.000
VarcoVC	9.95	.000	14.88	.000	**VarcoVOUV**	24.59	.000
SRVC	21.36	.000	140.43	.000	**SRVOUV**	21.32	.000

3.6. Conclusion

Nous avons présenté dans ce chapitre, divers paramètres et divers modélisations pouvant caractériser ou discriminer la parole pathologique de la parole normale. Notre objectif était d'essayer de trouver les paramètres prosodiques les plus pertinents pour bien caractériser la parole pathologique et qui peuvent être intégrés dans un système automatique de diagnostic et d'aide de l'orthophoniste.

En conclusion, les analyses préliminaires des résultats suggèrent que l'intonation et le débit de parole peuvent être utiles pour bien décrire et distinguer la parole pathologique de la parole normale. Mais effectivement pour arriver à des résultats réels, il nous faut plus de locuteurs patients que de locuteurs de référence et évidement plus de parole pathologique spontanée. Donc, il faut étudier la parole pathologique dans des situations où il y a les interactions quotidiennes et insister sur la prosodie.

Concernant les paramètres rythmiques, nous avons obtenu une variété de mesures du rythme extraits des phrases courtes dépourvues de sens produites par des patients dysarthriques avec divers degrés de sévérité. Ces mesures ont fourni des évaluations quantitatives préliminaires qui reflètent la sévérité de la parole pathologique. L'étude a

examiné les mesures avec présence et non-présence de la parole normale.

La première partie de l'évaluation était avec la parole normale. Analyse DFA désigne SRVC, rPVI-VC, %V, VarcoV et VarcoC pour Set 1 et toutes les mesures de rythme à l'exclusion de nPVI-VO, rPVI-VOUV et nPVI-VOUV pour Set 3 comme les prédicteurs les plus pertinents pour la distinction des différents niveaux de la sévérité de la parole.

Dans la deuxième partie de l'évaluation, les expériences impliquent seulement les patients dysarthriques. Les résultats sont différents. En effet, l'analyse DFA a identifié les paramètres prédictifs comme suit : pour Set 1 les paramètres sont : SRVC, rPVI-VC et VarcoVC, pour Set 2 ce sont ΔC, ΔV,% V, rPVI-C, VarcoC, VarcoV et VarcoVC et pour Set 3 ce sont : ΔVO suivie par rPVI-VO, rPVI-UV, VarcoVO, VarcoUV et nPVI-VOUV.

Au vu des résultats obtenus, il nous apparait important d'étudier plusieurs types de paramètres afin de les fusionner afin d'entreprendre efficacement le traitement automatique de la parole pathologique.

Chapitre 4 :

Parole Dysarthrique

(Nemours database) :

Résultats Expérimentaux

Chapitre 4 : Parole Dysarthrique (Nemours database) : Résultats Expérimentaux

4.1. Introduction

Diverses approches ont été proposées pour la classification des troubles de la parole liés à la dysarthrie. Ces méthodes se répartissent en trois grandes catégories. La première comprend les méthodes statistiques qui visent à déterminer une fonction de vraisemblance qui est définie comme la probabilité d'observer les données fournies au modèle. Par conséquent, la probabilité d'une observation est estimée en fonction du modèle, et l'estimation du maximum de vraisemblance est le paramètre d'intérêt dans le processus de classification [117, 118]. Le principal avantage des méthodes statistiques est leur taux de reconnaissance élevé. Cependant, leur inconvénient est le volume élevé des énoncés requis pour une phase d'apprentissage précise. La deuxième catégorie est basée sur des techniques informatiques bio-inspirées. Des réseaux de neurones ont été utilisés avec succès pour élaborer des modèles discriminants pour la parole pathologique [119]. Les avantages de classificateurs connexionnistes sont leur simplicité et leur facilité de mise en œuvre. Enfin, la troisième catégorie est basée sur une combinaison des deux premières catégories. L'idée est d'exploiter les avantages des deux méthodes. Ces modèles hybrides ont été étudiés de manière intensive à l'application normale de reconnaissance vocale ainsi que la classification de signaux biologiques [120]. Dans[121], une technique hybride combinant les modèles de Markov cachés (HMM) et les réseaux de neurones a été proposée pour l'analyse des signaux de parole dysarthriques en utilisant les paramètres MFCC. Nous pensons que l'approche hybride est plus précise que les HMM.

Dans ce chapitre, nous allons citer tout d'abord les méthodes de classification : les méthodes de la machine à support vecteurs (les SVM), les réseaux de neurones et le modèle des mélanges gaussiens (les GMM). Après, nous présentons les expérimentations et leurs résultats pour la classification parole dysarthrique/parole normale, la classification des niveaux de sévérité en utilisant surtout les paramètres de rythme décrite dans le chapitre 3. À la fin nous présentons le système de reconnaissance de la parole dysarthrique dépendant du locuteur en utilisant les résultats de la classification des niveaux de sévérité par les réseaux de neurones.

4.2. Les méthodes de classification

4.2.1. Les SVM multi-classes (machines à vecteurs de support)

La méthode des machines à vecteurs de support est une alternative récente pour la classification mais prometteuse [122]. Cette méthode repose sur l'existence d'un hyperplan séparateur dans un espace approprié. Les travaux de [123, 124], présentent les principes de base des SVM. Les SVM reposent sur deux notions : celle de marge maximale et celle de fonction noyau. Elles permettent de résoudre des problèmes de discrimination non linéaire. Nous allons voir comment un problème de classification se ramène à résoudre un problème d'optimisation quadratique. La marge est la distance entre la frontière de séparation et les échantillons les plus proches appelés vecteurs support. Dans un problème linéairement séparable les SVM trouvent une séparatrice qui maximise cette marge. Dans le cas d'un problème non linéaire on utilise une fonction noyau pour projeter les données dans un espace de plus grande dimension où elles seront linéairement séparables (voir les figures 4.1 et 4.2).

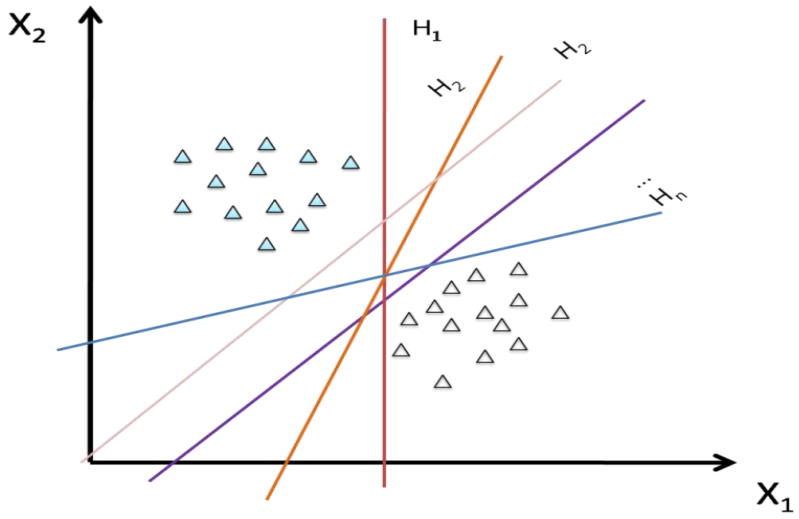

Figure 4. 1 : Pour un ensemble de points séparables linéairement, il y a une infinité d'hyperplans séparateurs.

Le problème posé est comment choisir parmi ces hyperplans ; l'hyperplan optimal qui sépare le mieux les points concernés comme nous pouvons le voir sur la figure 4.1. Selon la théorie de Vapnick [124], l'hyperplan optimal (optimum de la distance interclasse) est celui qui maximise la marge. Cette dernière étant définie comme la distance entre un hyperplan et les points échantillons les plus proches. Ces points particuliers sont appelés vecteurs support.

Les Figures 4.1 et 4.2 sont des exemples schématiques du classifieur SVM linéaire dans lequel les catégories des données sont séparées par une droite. Le cas linéairement séparable est peu intéressant. En effet, dans le monde réel, les problèmes de classification sont souvent

non linéaires. Pour résoudre ce problème, la méthode consiste à projeter les données dans un espace de dimensions supérieurs appelé espace de redescription. L'idée étant qu'en augmentant la dimensionnalité du problème on se retrouve dans le cas linéaire vu précédemment. L'idée clé est d'utiliser une fonction noyau notée K qui évite le calcul explicite du produit scalaire dans l'espace de redescription.

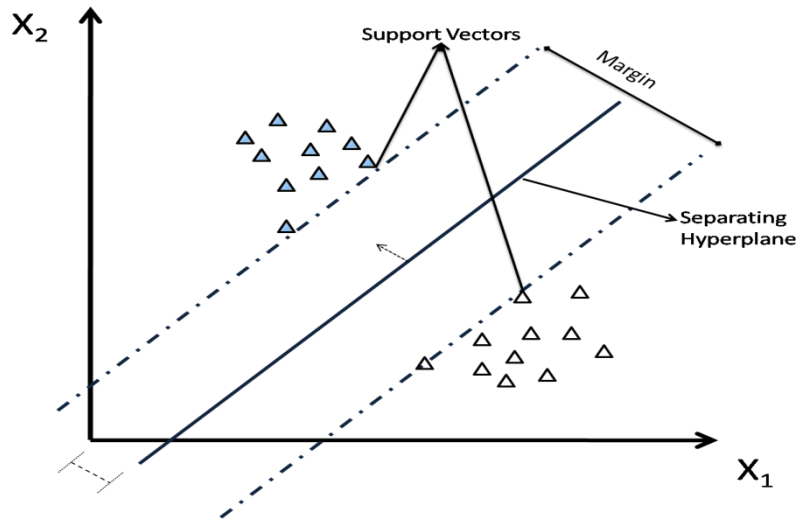

Figure 4. 2 : L'optimal hyperplan avec la marge maximale

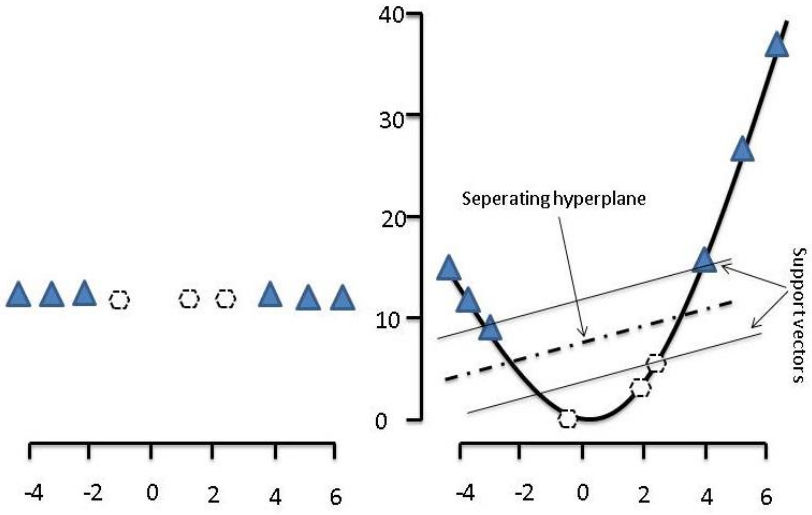

Figure 4. 3 : La transformation non linéaire qui projette les points dans un espace de dimension supérieur.

Nous avons opté pour le noyau gaussien (équation 4.1) qui est parmi les noyaux les plus utilisés. C'est un des noyaux dits de type radial. Le noyau gaussien une fonction à base

radial abrégée RBF[124-126]. Ce type de fonctions sont utilisés quand il n'y a pas de connaissance a priori sur les données (pas d'hypothèses sur la normalité).

$$\mathbf{K_\gamma(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}} \quad \textbf{(4. 1)}$$

Où : x_i, x_j sont des vecteurs de paramètres de deux échantillons de données. $\|\mathbf{x_i} - \mathbf{x_j}\|^2$ reconnue comme le carré de la distance euclidienne. $\boldsymbol{\gamma}$ est un paramètre libre. En générale, $\boldsymbol{\gamma}$ prend la valeur : 1/2.

4.2.2. Les réseaux de neurones

Nous proposons dans ce travail, un système modulaire constitué de réseaux de neurones simples (MLP), pour tenter une classification parole pathologique / parole normale et aussi la classification les différents niveaux de sévérité de la dysarthrie.

Les réseaux de neurones artificiels généralement se caractérisent par leur capacité d'approximer la plupart des fonctions. À cet effet, nous pouvons résoudre la plupart des taches. Autrement dit, leur capacité d'apprentissage à partir d'exemples, leur adaptabilité, leur robustesse aux données bruitées ou manquantes et en reconnaissance de la parole, par leur puissance de discrimination, à diviser l'espace de paramètres acoustiques en classes phonétiques [98]. De nombreuses implémentations de ces réseaux ont été proposées dans la littérature. La structure la plus répandue est celle du réseau multicouches (Multi-Layer Perceptron, MLP) [127].

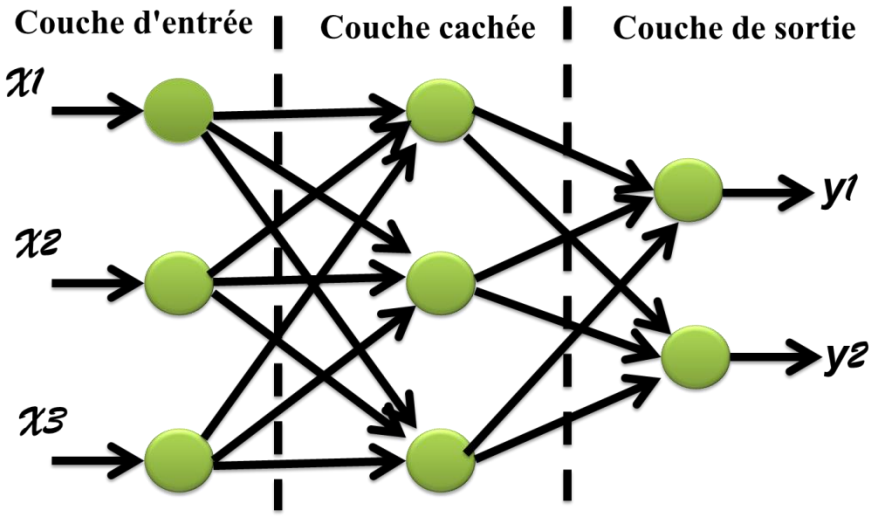

Figure 4. 4 : Exemple de réseau de neurones typiques

4.2.2.1. Le neurone formel

Le neurone formel ou artificiel est un dispositif à plusieurs entrées et une sortie. Les réseaux de neurones artificiels sont généralement l'ensemble de neurones interconnectés qui se transmettent l'information par l'intermédiaire de ces connexions (figure 4.4 et 4.5).

Chaque neurone formel réalise une somme pondérée des valeurs de ses entrées (voir figure 4.5).

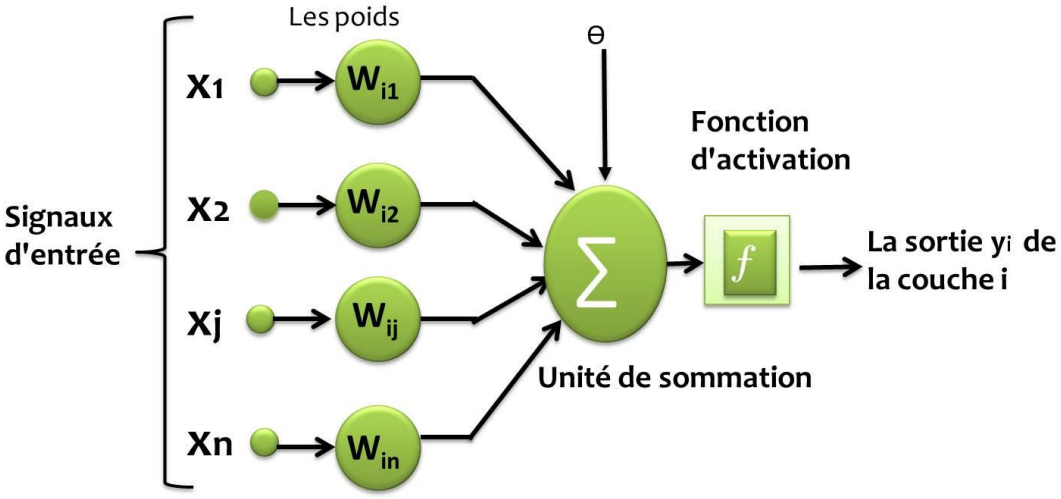

Figure 4. 5 : Schéma d'un neurone formel.

Sa sortie est donnée par la formule générale suivante :

$$y_i = f(\sum_{j=1}^{n} w_{ij} x_j - \theta) \qquad (4.2)$$

Où θ est le seuil d'activation du neurone. Tandis que f est La fonction d'activation ; elle régularise les valeurs de sortie et prend en général ses valeurs dans l'intervalle $[0 - 1]$. Il existe plusieurs fonctions d'activation (Voir Figure 4.6) :

4.2.2.2. Les perceptrons

Les perceptrons sont des réseaux sans contre réaction (Feed-forward networks). Ils permettent aux signaux de circuler dans un seul sens ; de l'entrée à la sortie. Ils sont définis à partir d'une répartition des neurones formels en couches : les sorties des neurones de la couche i constituent les entrées de la couche i+1. Généralement, l'entrée du perceptron est le vecteur de l'observation à classer ; à chaque sortie de neurone correspond une classe. La classe reconnue correspond au neurone de sortie fortement activée. La figure 4.7 montre la structure d'un perceptron à trois couches dont une couche cachée, appliqué à un problème de

reconnaissance de mots isolés. Ce modèle est issu des travaux de F. Rosenblatt [128] sur le perceptron monocouche.

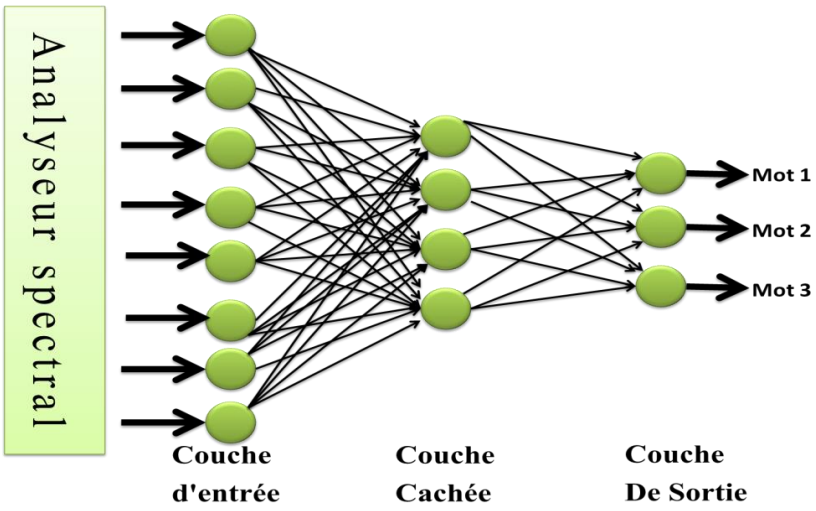

Figure 4. 6 : Structure d'un perceptron a trois couches

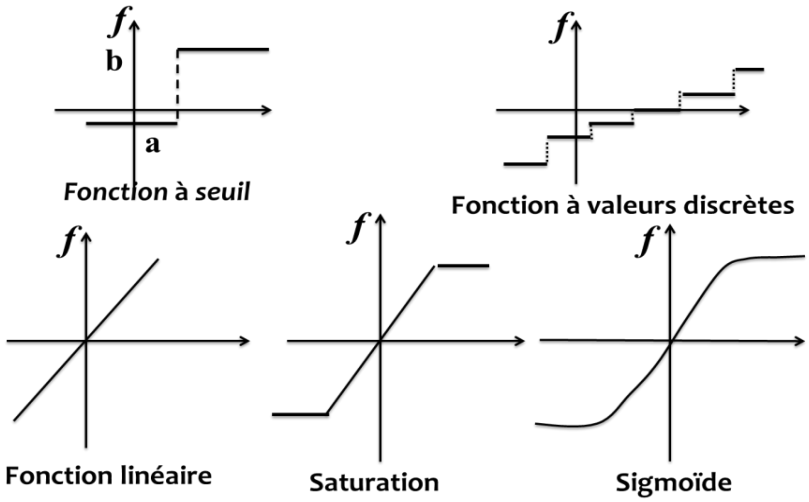

Figure 4. 7 : Exemples de fonctions d'activation.

4.2.3. Modèle de mélange de gaussiennes (MMG).

Un modèle de mélange gaussien (MMG) est une fonction de densité de probabilité paramétrique représentée comme une somme pondérée des densités Gaussiennes [129]. Les MMG sont couramment utilisés comme un modèle paramétrique de la distribution de probabilité de mesures continues ou des caractéristiques d'un système biométrique, tels que les caractéristiques spectrales du conduit vocal liées à un système de reconnaissance du locuteur. Les paramètres GMM sont estimés à partir des données d'apprentissage à l'aide de

l'algorithme EM (Expectation Maximization : EM) ou l'estimation par maximum a posteriori (MAP). Un modèle de mélange gaussien est une somme pondérée de densités gaussiennes de p composantes telle que donnée par l'équation (4.3) [130, 131].

Si y est un vecteur d'observation de d dimension alors :

$$\mathbf{P}\left(\frac{\mathbf{y}}{\boldsymbol{\lambda}}\right) = \sum_{\mathbf{i=1}}^{\mathbf{M}} \omega_i \, \mathbf{g}(\mathrm{y}/\mu_i, \boldsymbol{\Sigma_i}) \qquad \textbf{(4.3)}$$

$\mathbf{g}(\mathrm{y}/\mu_\mathrm{i}, \boldsymbol{\Sigma_i})$ Représente les densités des composantes Gaussiennes où i=1,2,….M.

μ_i Représente le vecteur de la moyenne de la composante IIème du mélange.

$\boldsymbol{\Sigma_i}$ Représente la matrice de covariance de la composante IIème du mélange.

ω_i Représente les poids du mélange où i=1,2,….M.

M désigne le nombre de lois gaussiennes multidimensionnelles dans le mélange.

Chaque composante est une densité gaussienne de d variables de la forme :

$$\mathbf{g}(\mathbf{y}/\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}) = \frac{\mathbf{1}}{(\mathbf{2\pi})^{\mathbf{d/2}}|\boldsymbol{\Sigma_i}|^{\mathbf{1/2}}} \mathbf{exp}\left\{-\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{y}-\boldsymbol{\mu_i})'\boldsymbol{\Sigma_i}(\mathbf{y}-\boldsymbol{\mu_i})\right\} \qquad \textbf{(4.4)}$$

4.3. Évaluation objective de la parole dysarthrique

4.3.1. Corpus

Le matériau linguistique étudié est constitué de trois séries :

4.3.1.1. Série 1 : Set1

La segmentation phonétique manuelle en particulier pour la parole pathologique est un travail fastidieux. Pour cette raison, seules huit phrases pour chaque locuteur dysarthrique et locuteur de contrôle ont été segmentées et étiquetées manuellement à des intervalles consonantiques et vocaliques (en totalité 176 phrases). C'est le même corpus utilisé par Selouani et al., et Dahmani et al. [43, 45, 46, 132]. Le début et la fin de chaque phonème au sein de chaque phrase sont marqués par l'écoute de la forme d'onde et en utilisant le spectrogramme à large bande comme un guide. Nous avons suivi les critères de segmentation cités par Mattys et al. [107, 133]. La segmentation des intervalles VC est basée sur la segmentation vocalique déjà faite. En fait, les durées des intervalles de CV sont la somme des durées des intervalles successifs vocalique et consonantique (intervocalique). (L'annexe A présente l'ensemble de phrases utilisées dans cette série).

4.3.1.2. Serie2 : Set2

Set 2 désignée pour la segmentation automatique. Set 2 contient uniquement la parole pathologique. Effectivement, la segmentation automatique dans la base de données Nemours a été réalisée uniquement pour les locuteurs dysarthriques, à l'exclusion du cas le plus sévère KS (74 phrases pour chacun des 10 patients à l'exclusion KS : 740 phrases). La segmentation a été développée en utilisant la transcription phonétique automatique par les HMM [116], la segmentation a été délivrée avec la base de données sans la correction des erreurs de la transcription.

4.3.1.3. Serie3 : Set3

La troisième série Set 3 de stimuli est l'ensemble des phrases de la base de données Nemours segmentées en intervalles voisées et non voisées (1628 phrases). Les intervalles voisés et non voisés ont été segmentés automatiquement avec un script Matlab (www.mathworks.com) après l'extraction de F_0, fichiers RMS et le taux de passage par zéro (TPZ) à l'aide d'un script Praat. En fait, la segmentation voisée / non voisée est basée sur les mesures à court terme du pitch F_0, de l'énergie et du TPZ à travers une fenêtre glissante de longueur N. Le calcul de la durée des intervalles VO (**VO**iced segments) et les intervalles UV (**U**n**V**oiced segments) nous permettent de calculer les intervalles VOUV ou UVVO (**Vo**iced **U**n**v**oiced ou **U**n**v**oiced **Vo**iced) (par analogie avec les intervalles CV ou VC).

4.3.2. Classification de la parole dysarthrique / parole normale.

Les SVM sont utilisés dans les deux phases d'apprentissage et de décision. La classification a été réalisée en utilisant seulement les paramètres prédicteurs dans le test fermé et le test ouvert (« Closed and Open test» respectivement). Dans le test fermé (Closed test), les données de test sont les mêmes données utilisées pour la phase d'apprentissage. Un test ouvert (Open test), les données à tester sont différentes des données utilisées dans la phase d'apprentissage. Dans notre cas, nous avons opté pour deux tests ouverts. Le premier test ouvert utilise uniquement l'ensemble de données de test qui est totalement différent de l'ensemble de données d'apprentissage désigné dans nos figures par : open test : test Set. Tandis que, le deuxième test ouvert utilise la totalité des données : l'ensemble de données d'apprentissage et l'ensemble de données de test (**100 %** des données) et que nous désignons en anglais par : Open Test : Learning set.

En général, un test ouvert donne une indication plus fiable de l'efficacité de la classification. Par conséquent, nous avons divisé nous-mêmes les données en deux sous-

ensembles disjoints : l'ensemble d'apprentissage (Learning Set) avec **62,5%** de l'ensemble des données utilisées pour élaborer le modèle de prévision. L'ensemble de test (Test Set) avec les données restantes qui sert à mesurer la performance du système de classification ou de reconnaissance.

4.3.1.1. La méthode LR

Pour Set 1 et Set 3 le coefficient R^2 de Nagelkerke est 0.780 et 0,706 respectivement. L'utilisation de R^2, le coefficient de détermination, également appelé le coefficient de corrélation multiple, est bien établie. IL est la valeur absolue du coefficient de corrélation linéaire entre Y la variable à prédire (la maladie ou non et X la variable explicative de la régression (les paramètres de rythme). Il est utile en tant que mesure du succès de prédiction de la variable dépendante des variables indépendantes. R^2 qui varie entre 0 et 1, Il indique une forte relation entre la prédiction et le regroupement. Plus R^2 se rapproche de la valeur 1, meilleure est l'adéquation du modèle aux données. Un R^2 faible signifie que le modèle a un faible pouvoir prédictif [134].

Avec le test ouvert, le succès de prédiction global est estimé à **87,5%** (**93,9%** pour les DP et **81,8%** pour les HC) pour Set 1 et à **86,5%** (**83,5%** pour les DP et **89,4%** pour les HC) pour Set 3.

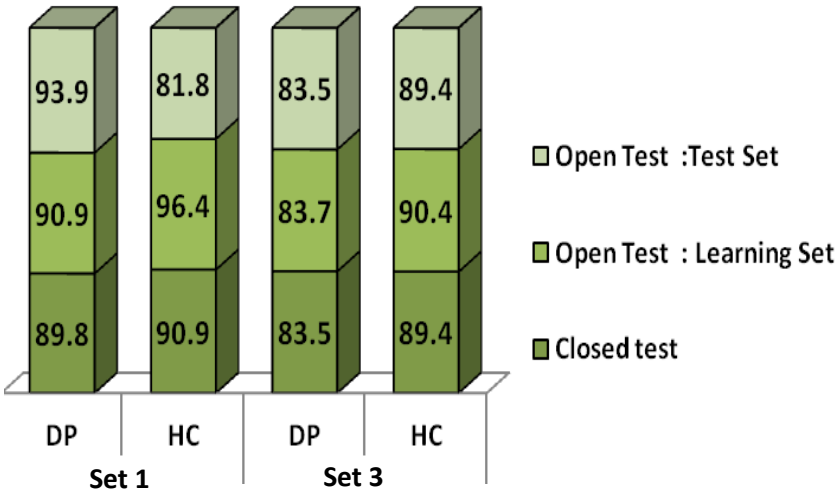

Figure 4. 8 : Les résultats de la classification DP / HC par la méthode LR avec les paramètres prédicteurs : SRVC, nPVI-V, nPVI-C, rPVI-V, rPVI-C et ΔC pour Set 1 (SRVOUV, ΔVO, ΔVO, %VO, nPVI-UV, rPVI-UV, VarcoUV et rPVI-VO pour Set 3).

Nous avons accompli les mêmes expériences avec seulement le nouveau paramètre rythmique SRVC. Les résultats sont satisfaisants : nous considérons le test ouvert, le pourcentage global de la classification correcte a diminué que de **2,7%** par rapport aux

résultats obtenus avec tous les paramètres pertinents pour Set 1 et de **11,5%** pour Set 3. Les résultats obtenus avec le test fermé montrent également une légère baisse. Ainsi, son taux global de bonne classification pour Set 1 est de **84,8%** (**84,8%** pour les DP et **84,8%** pour les HC) et pour Set 3 est de **75,0%** (**74,4%** pour les DP et **75,5%** pour les HC) avec le seulement le paramètre SRVOUV. Ces résultats indiquent clairement sa forte contribution à la classification correcte de la parole pathologique par rapport à la parole normale. Un résumé de tous les résultats est donné par les figures 4.8 et 4.9.

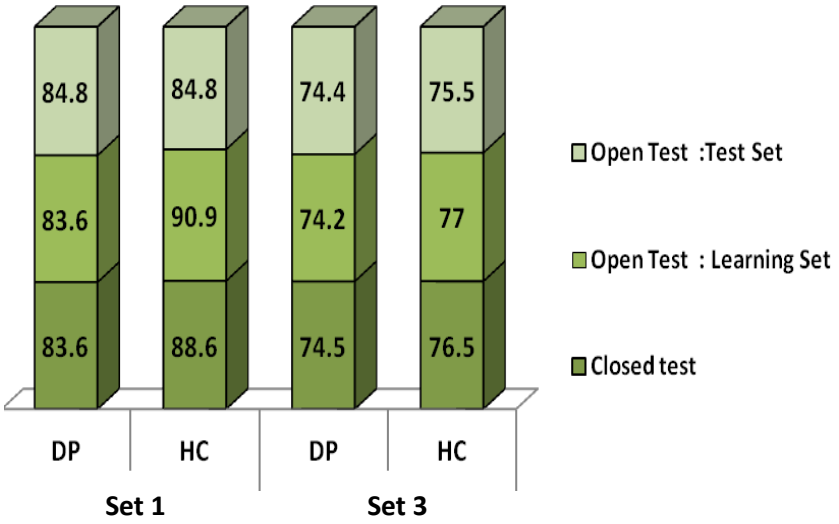

Figure 4. 9 : les résultats de la classification DP/ HC par la méthode LR en utilisant le nouveau paramètre : SRVC. (Pour set 1) et SRVOUV (Pour Set3).

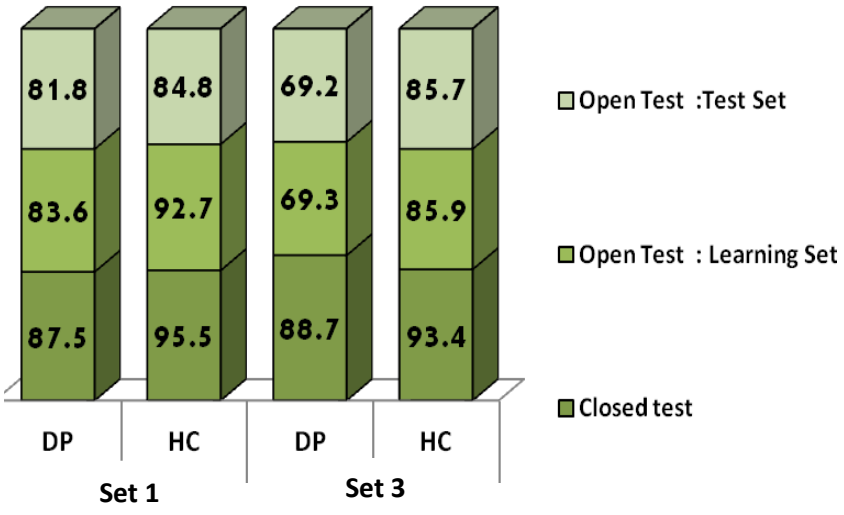

Figure 4. 10 : résultats pour la classification de DP / HC par la méthode SVM en utilisant les paramètres prédicteurs : SRVC, nPVI-V, nPVI-C, rPVI-V, rPVI-C et ΔC pour Set 1 (SRVOUV, ΔVO, ΔVO, %VO, nPVI-UV, rPVI-UV, VarcoUV and rPVI-VO pour Set 3).

4.3.1.2. La méthode SVM

Nous pouvons voir à travers la figure 4.10, qu'en utilisant des variables prédictives identifiées par la méthode LR, le taux de la classification correcte est de **83,3%** (81,8% pour les DP et **84,8%** pour les HC). En ce qui concerne Set 3, les résultats SVM étaient les meilleures pour le test fermé. Notons une amélioration par rapport aux résultats LR de **5,6%** et de **6,2%** dans le taux global de bonne classification en mode test fermé avec variables prédictives et avec uniquement le paramètre SRVC respectivement (voir l'annexe C). Dans le test ouvert, le taux global de la classification correcte a diminué pour Set 1 à **83,3%** (**81,8%** pour les DP et **84,8%** pour les HC) et pour Set 3 à **77,5%** (**69,2%** pour les DP et **85,7%** pour les HC).

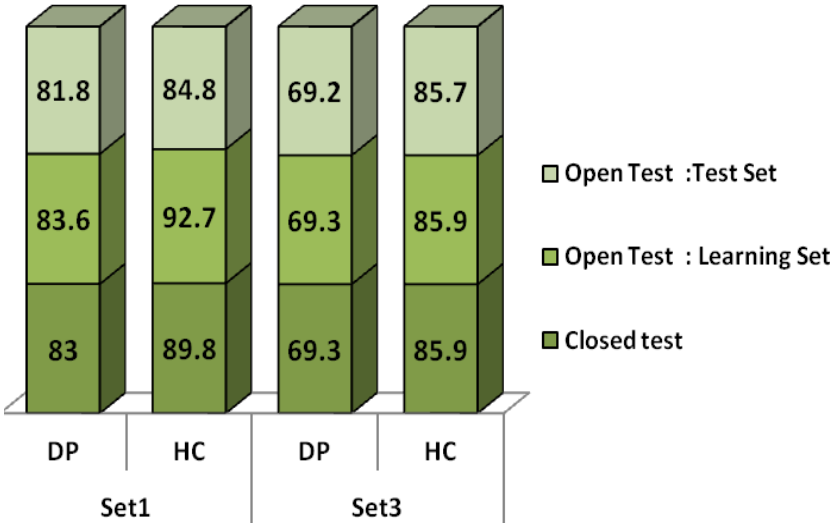

Figure 4. 11 : les résultats de la classification DP / HC par la méthode SVM en utilisant le nouveau paramètre prédicteur : SRVC pour Set 1 (SRVOUV, pour Set 3)

En utilisant uniquement le paramètre que nous proposons SRVC, les SVM atteignent **83,3%** (**81,8%** pour les DP et **84,8%** pour les HC). Par analogie pour Set 2, le paramètre SRVOUV atteint **77,5%** du taux global de classification correcte (**69,2%** pour les DP et **85,7%** pour les HC). (Voir figure 4.11).

4.3.3. Évaluation objective de la sévérité dysarthrique par les SVM.

Les sujets ont été codés selon le niveau de sévérité, comme indiqué dans le tableau 2.9 (**Tableau 2. 9: L'Évaluation Frenchay des Patients de Nemours (Frenchay Dysarthria Assessment : FDA)**) (voir aussi le tableau 4.1). Les patients à pathologie légère (MP), les patients à pathologie modérée (MOP) et les patients à pathologie sévère (SP) et le niveau de la parole normale était désigné par HC (HC : Healthy Control). L'objectif était d'évaluer les niveaux de sévérité de la

parole dysarthrique. Nous avons obtenu une variété de mesures du rythme des segments vocaliques / consonantiques extraits des phrases courtes, non sensées produites par les locuteurs avec des degrés variables de dysarthrie de la base de données Nemours. Ces mesures ont fourni des évaluations quantitatives préliminaires qui reflètent la sévérité de la parole de dysarthrie. L'étude a examiné les mesures avec présence et non-présence de la parole normale.

4.3.3.1. Classification des niveaux de sévérité avec la présence de la parole normale.

La première partie de l'évaluation était avec la parole normale (voir figures 4.12 et 4.13). Analyse DFA conçoit SRVC, rPVI-VC, %V, VarcoV et VarcoC comme paramètres pertinents pour Set 1. Pour Set 3, tous les paramètres de rythme à l'exclusion de nPVI-VO, rPVI-VOUV et nPVI-VOUV étaient discriminants des niveaux de sévérité. Les résultats de classification sont très prometteurs. Les SVM reconnaissent **78,1%** de la parole légèrement altérée (PM) pour Set 1 et **96,6%** de la parole normale (HC) pour Set 3. Quant à la parole sévèrement altérée, la méthode SVM atteint une bonne précision de reconnaissance avec **88,7%** et **71,9%** pour les deux Set 3 et Set 1 respectivement. La classe des patients dont la parole modérément altérée, est mal reconnue par rapport aux deux classes PM et PS. Le faible pourcentage de reconnaissance correcte est de **42,7%** pour Set 3.

Tous les résultats de toutes les classifications correctes et erronées peuvent être vus dans les tableaux AC.1 et AC.2 (annexe C). Cependant, ici, nous résumons seulement les classifications erronées du test fermé pour Set 3 car elle représente de mieux l'allure générale des résultats. La parole modérément altérée est mal classée comme parole légèrement altérée, normale et sévèrement altérée (**24,5%** en PM, **20,9%** en HC et **11,8%** comme PS). La parole sévèrement altérée est mal classée le plus souvent comme parole légèrement altérée et comme parole normale (**15%** en PM, **11,5%** en HC et **7,1%** en MOP). La parole légèrement altérée comme nous nous y attendions, est mal reconnue comme parole normale et comme parole modérément altérée (**10,6%** en HC et **3%** en MOP). La parole normale est mal classée comme parole légèrement altérée (**2,8%).**

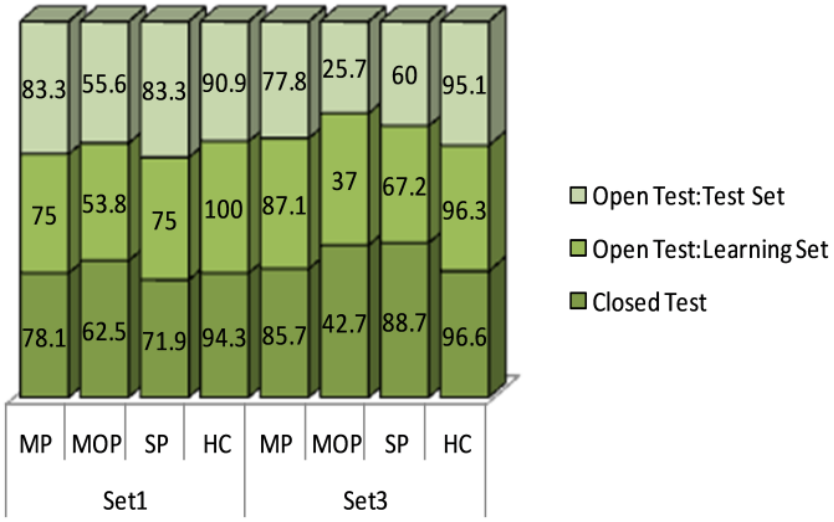

Figure 4. 12 : Les résultats de la classification des niveaux de sévérité de la dysarthrie avec la présence de HC par la méthode SVM en utilisant les paramètres prédicteurs : SRVC, rPVI-VC, %V, VarcoV et VarcoC pour Set 1 (nPVI-VO, rPVI-VOUV et nPVI-VOUV pour Set 3)

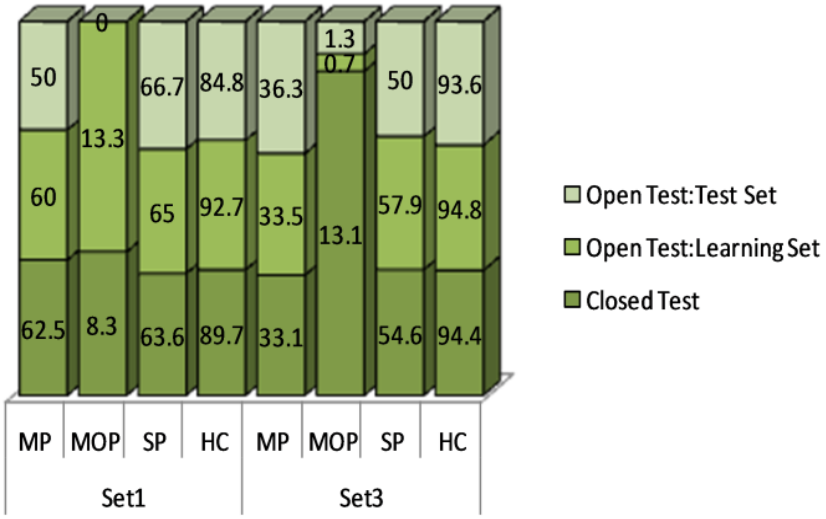

Figure 4. 13 : les résultats de la classification des niveaux de sévérité de la dysarthrie avec la présence de HC par la méthode SVM en utilisant seulement le paramètre prédicteur : SRVC pour Set 1 (SRVOUV, pour Set 3).

Tableau 4. 1 : Résumé des scores de l'évaluation de Frenchay des patients dysarthriques de Nemours.

Patient	**KS (PS)**	**SC (PS)**	**BV (PS)**	**BK (PS)**	**RK (PMO)**	**RL (PMO)**	**JF (PMO)**	**LL (PM)**	**BB (PM)**	**MH (PM)**	**FB (PM)**
Sévérité (%)	-	49.5	42.5	41.8	32.4	86.7	21.5	15.6	10.3	7.9	7.1

4.3.3.2. Classification des niveaux de sévérité sans la présence de la parole normale.

Comme c'est entendu, dans la deuxième partie de l'évaluation, les expériences n'impliquent que la parole pathologique. Les résultats sont différents. En effet, l'analyse DFA a identifié les paramètres prédicteurs pertinents comme suit : pour Set 1 les paramètres sont : SRVC, rPVI-VC et VarcoVC, pour Set 2 sont ΔC, ΔV, %V, rPVI-C, VarcoC, VarcoV et VarcoVC et pour Set 3 sont : ΔVO suivie par rPVI-VO, rPVI-UV, VarcoVO, VarcoUV et nPVI-VOUV. Les taux de reconnaissance les plus élevés sont pour la classification des patients PM. Avec Set 1, la méthode SVM reconnait **100%** des patients PM. Ce taux a diminué à **95,9%** et **93,2%** pour Set 2 et Set 3, respectivement (voir la figure 4.14). Pour les cas PMO et PS, les résultats sont autour de **50%**. La plupart des mauvaises classifications confondent les cas PM avec les cas PMO et PS (les détails de ces résultats sont présentés dans les tableaux AC.1 et AC.2 de l'annexe C).

En se basant sur les résultats ci-dessus, nous pouvons conclure que le problème réside beaucoup plus dans le classement de la parole de la classe MOP. Cette confusion peut être due à des déficits dans l'évaluation perceptuelle ou peut-être simplement à cause des caractéristiques de la parole modérément altérée (MOP), qui peut être confondue avec la parole légèrement altérée ou normale ou parfois même avec la parole sévèrement altérée. Ces résultats confirment la capacité des paramètres de rythmes de faire la distinction entre les niveaux de sévérité, même avec la présence de la parole normale. Bien que le SRVC ou SRVOUV réalisent une faible précision dans la classification correcte des PM et des PMO, il reste une caractéristique pertinente. En effet, parce que le paramètre SRVC est sélectionné par l'analyse DFA comme la meilleure variable pour séparer les trois niveaux de sévérité. Nous avons examiné sa seule influence pour la classification des niveaux de sévérité. Les résultats étaient globalement satisfaisants. Nous avons enregistré que SRVC n'est pas en mesure de distinguer la parole des PM de la parole normale (HC) (**33,1% et 13,1%** respectivement avec Set 3). Bien que, nous pouvons remarquer à partir de la figure 4.13 (et le tableau AC.2 de l'annexe C) que les résultats du SRVC pour PS et HC sont meilleurs que les résultats des PM et des MOP (**54,6%** pour les PS et **94,4%** pour HC).

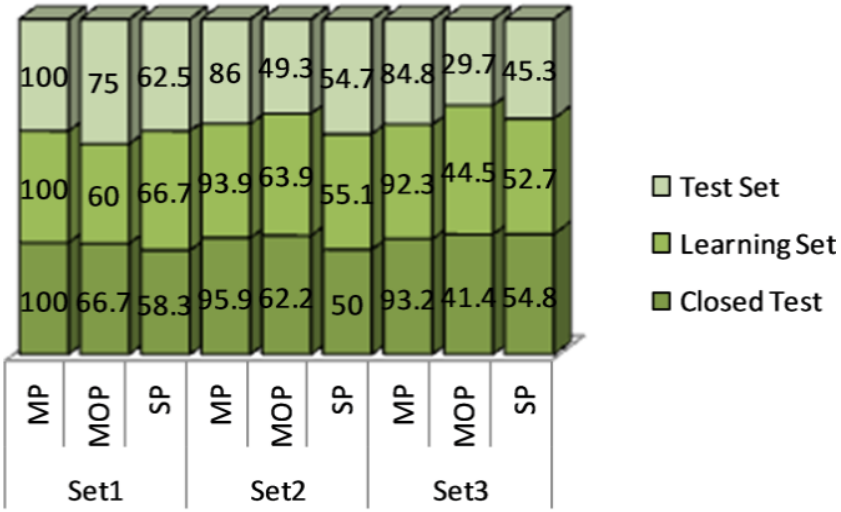

Figure 4. 14 : les résultats des SVM pour la classification des niveaux de sévérité sans la présence de la parole normale HC en utilisant les paramètres prédicteurs : SRVC, rPVI-VC et VarcoVC pour Set 1 (ΔC, ΔV, %V, rPVI-C, VarcoC, VarcoV et VarcoVC pour Set 2, et pour Set 3 : ΔVO, rPIV-VO, rPIV-UV, VarcoVO, VarcoUV et nPIV-VOUV.

4.3.4. Le Système connexionniste hybride pour la classification et la reconnaissance de la parole dysarthrique

Nous présentons une approche hybride qui combine les distributions postérieures des classes, la structure hiérarchique perceptron multicouches (MLP) et les HMM pour effectuer une reconnaissance de la parole dysarthrique après avoir évaluer les différents niveaux de la sévérité de la dysarthrie[45]. Pour l'évaluation des niveaux de sévérité, nous comparons le système hybride avec des systèmes de référence, à savoir les modèles de mélanges gaussiens (GMM), simple MLP et la structure hiérarchique MLP standard. Pour la reconnaissance, un système de base HMM est utilisé, et les résultats obtenus par les modèles dépendants du locuteur sont comparés avec les autres résultats des autres systèmes.

4.3.4.1. Estimation des probabilités a posteriori.

Le problème de classification des vecteurs d'observations dans l'un des K classes peut être formalisé par la règle de Bayes. Une classe spécifique est déterminée si une probabilité à posteriori du vecteur appartenant à la classe k est plus grande que celles des autres classes. Cette probabilité a posteriori est donnée par :

$$P(k/x)=\frac{P(K)P(x/k)}{P(x)} \qquad (4.5)$$

Où P(k) est une probabilité a priori de la classe k et P (k/x) est la fonction de densité de probabilité (**P**robability **D**ensity **F**unction : **pdf**) de x convenu à la classe k. Dans notre

application, la pdf du vecteur des caractéristiques est représentée par un modèle général gaussien avec K classes :

$$P(x)= \sum_{k=1}^{K} \aleph (x; \mu_k,\Sigma_k) \quad (4.6)$$

Où $\aleph(x ; \mu_k, \Sigma_k)$ est la distribution gaussienne de dimension d. $\mu_k \epsilon R^d$ et $\Sigma_k \epsilon R^{dxd}$ représentent respectivement le vecteur moyen et la matrice covariance. La distribution gaussienne peut être écrite comme suit :

$$\aleph(x; \mu_k,\Sigma_k)= (2\pi)^{-d/2}|\Sigma_k|^{-\frac{1}{2}}\exp(\frac{-1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) \quad (4.7)$$

Puis, en utilisant (4.5) et (4.6), la probabilité a posteriori P(k/x) peut s'écrire sous la forme :

$$P(k/x) = \frac{\aleph(x, \mu_k, \Sigma_k)}{\sum_{k'=1}^{K} \aleph (x, \mu_{k'}, \Sigma_{k'})} \quad (4.8)$$

La probabilité a priori de toutes les classes est considérée comme un facteur d'échelle. En utilisant la moyenne $\mu_k=(\mu_{1,k},.....,\mu_{d,k})^T$ et l'inverse de la matrice covariance $\Sigma_k^{-1}=|\sigma_{ijk}|$ et en utilisant (4.7), nous pouvons écrire :

$$\aleph(X,\mu_k,\Sigma_k)=(2\pi)^{-\frac{d}{2}}|\Sigma_k|^{-\frac{1}{2}}$$
$$x \exp(-\frac{1}{2}\sum_{j=1}^{d}\sum_{i=1}^{j}(2-\delta_{ij})\sigma_{ijk}x_ix_j+\sum_{j=1}^{d}\sum_{i=1}^{j}\sigma_{ijk}\mu_{jk}x_i-\frac{1}{2}\sum_{j=1}^{d}\sum_{i=1}^{j}\sigma_{ijk}\mu_{jk}\mu_{ik}) \quad (4.9)$$

Où δ_{ij} est le symbole de Kronecker. Considérons la fonction ψ définie comme le logarithme de $\aleph(x,\mu_k,\Sigma_k)$ et qui peut être écrite comme suit :

$$\psi=\log (\aleph(x,\mu_k,\Sigma_k)=\alpha_k^T\tilde{x} \quad (4.10)$$

Où α_k^T et $\tilde{x}\epsilon R^G$ sont définis par les équations suivantes :

$$\alpha_k=(\alpha_{0,k}\sum_{i=1}^{d}\sigma_{j1k}\mu_{jk}, ..., \sum_{j=1}^{d}\sigma_{jdk}\mu_{jk} -\frac{1}{2}\sigma_{11k}-\sigma_{12k}.,.....\sigma_{1dk},.....-\frac{1}{2}(2-\delta_{ij})\sigma_{1jk},.....-\frac{1}{2}\sigma_{ddk})^T \quad (4.11)$$

Où $\alpha_{0,k}=\sum_{j=1}^{d}\sum_{i=1}^{d}\sigma_{ijk}\mu_{jk}\mu_{ik}-\frac{d}{2}\log(2\pi)-\frac{1}{2}(\Sigma_k)$ (4.12)

Et $\tilde{x} = (1, x^T, x_1^2, x_1x_2, \dots\dots, x_1x_dx_2^2, x_2x_3, \dots\dots x_2x_d, \dots\dots.x_d^2)^T$ (4.13)

La dimension G est définie par :

$$G=1+\frac{d(d+3)}{2} \quad (4.14)$$

Notre approche consiste à utiliser α_k comme la fonction d'activation MLP dans la structure hybride. En effet, la probabilité a posteriori de (4.5), en tenant compte de (4.10), en introduisant le logarithme ; peut-être écrite comme suit :

$$P(k/x) = \frac{\exp(\alpha_k^T \tilde{x})}{\sum_{k=1}^{K} \exp(\alpha_k^T \tilde{x})} \quad (4.15)$$

4.3.4.2. Structure hiérarchique de MLP

L'approche connexionniste présentée ici, propose d'évaluer les niveaux de sévérité de la dysarthrie en utilisant un expert de neurones (MLP). Cette structure s'est avérée très efficace dans le cas de la reconnaissance des phonèmes [135]. Dans notre application, une classification binaire des sous-tâches est attribuée individuellement à chaque couche MLP qui est formée indépendamment pour distinguer deux niveaux de sévérité. Au cours de la phase d'apprentissage, un flux de données segmentées est présenté à l'entrée du réseau. Puisque l'apprentissage est supervisé, les données sont étiquetées par rapport au niveau de sévérité de la dysarthrie. Ce vecteur d'entrée est prétraité selon l'équation (4.13). Par conséquent, la couche d'entrée de chaque MLP est composée de G unités calculées par (4.14). La fonction d'activation est définie par l'équation (4.11). On se référera par E^1 et O^1 à l'entrée et à la sortie de la première couche et par E^2 et O^2 d'entrée et de sortie de la seconde couche. Les équations suivantes définissent les relations entre les entrées et les sorties de chaque couche :

$$E_k^1 = x \; ; \qquad O_k^1 = \frac{\exp(\sum_{g=1}^{G} \alpha_{gk}^T E_g^1)}{\sum_{k=1}^{k} \exp(\sum_{g=1}^{G} \alpha_{gk}^T E_g^1)} \quad (4.16)$$

α_k^T jouera le rôle de poids dans les réseaux neuronaux classiques. Pour la seconde couche, on peut écrire :

$$E_k^2 = O_k^1 \; ; \qquad O_k^2 = \frac{\exp(\alpha_k^T O^1)}{\sum_{k=1}^{k} \exp(\alpha_k^T O^1)} \quad (4.17)$$

4.3.4.3. Configuration des taches de la classification

Le réseau de neurones utilisé est composé de trois MLP. Ces MLP utilisent les fonctions activations que génère la distribution a posteriori pour la discrimination binaire des tâches. Comme l'illustre la figure 4.15, la première tâche du MLP consiste à distinguer entre la parole normal (Healthy Control : HC) et la parole dysarthriques quel que soit le niveau de sévérité. Le deuxième MLP traite uniquement la parole dysarthrique en vue de son

classement dans deux niveaux : locuteurs à pathologie légère ou locuteurs à pathologie sévère. Le troisième MLP effectue une classification des cas modérément légers et des cas sévères. Quatre niveaux de classification sont pris en compte : L_0, L_1, L_2 et L_3. Le niveau L_0 correspond au HC et L_3 pour les cas sévères de la dysarthrie (PS). Pour extraire les caractéristiques MFCC, une fenêtre de Hamming glissante de 30 ms avec un recouvrement de 40%. Comme le nombre d'entrées MLP est fixe et le nombre de segments varie, nous avons réalisé une compression basée sur la moyenne des paramètres sur chaque cinq segments. Ce nombre a été trouvé optimal pour les cas étudiés après de vastes essais de validation croisée. Ces tests permettent également de sélectionner les paramètres les plus pertinents de rythme qui constituent les entrées du MLP utilisant le standard algorithme de rétro-propagation. Pour un MLP simple, cinq ΔV (des cinq trames moyennés) %V et n-PVI qui se trouvent être les meilleurs paramètres par le même test de la validation croisée. Par conséquent, le nombre paramètres de rythme est sept pour chaque phrase prononcée. Le nombre d'entrées du système hybride est donc de 72 paramètres (13 MFCC x 5+ sept mesures du rythme).

4.3.4.4. Évaluation des performances du système de la classification

Les résultats sont donnés dans par tableau 4.2. Une comparaison de la performance du système hybride est faite avec des systèmes de référence, un système à base des modèles de mélanges gaussiens (GMM), un système MLP simple et un système MLP standard hiérarchique. Les deux systèmes MLP : simple et standard hiérarchiques utilisent l'algorithme de rétro-propagation avec sigmoïde comme fonction d'activation. La tâche principale de ces systèmes est d'effectuer la discrimination entre les quatre niveaux de sévérité. Le taux de classification globale des deux systèmes : hybride et MLP hiérarchique est calculée comme le produit des taux des trois MLP constituant les deux systèmes concernés. Les résultats montrent que le système hybride proposé surpasse ceux de référence. Une amélioration de plus de 3% a été atteinte en comparaison avec les GMM en utilisant la même analyse acoustique. L'influence de l'introduction des paramètres rythmiques, en plus des paramètres MFCC, était très importante (des améliorations variant de 3% à 6% ont été obtenues) ce qui confirme la pertinence de ces paramètres pour l'évaluation de la parole dysarthrique. Nous notons que BV, dont le score de Frenchay était 42,5% (tableau 4.1), est toujours mal classé. En effet, à l'examen de la parole de BV, nous avons constaté et cité dans le chapitre 3 que la vitesse de la parole du sujet BV tout à fait normale et sa parole presque intelligible, mais avec nasalité. Nous notons également que FB est classé comme un cas presque guéri par

l'évaluation de Frenchay (tableau 4.1) mais ses paroles sont pour la plupart classés dans la catégorie L2. En fait, sa parole est très intelligible, mais son débit de parole est très lent.

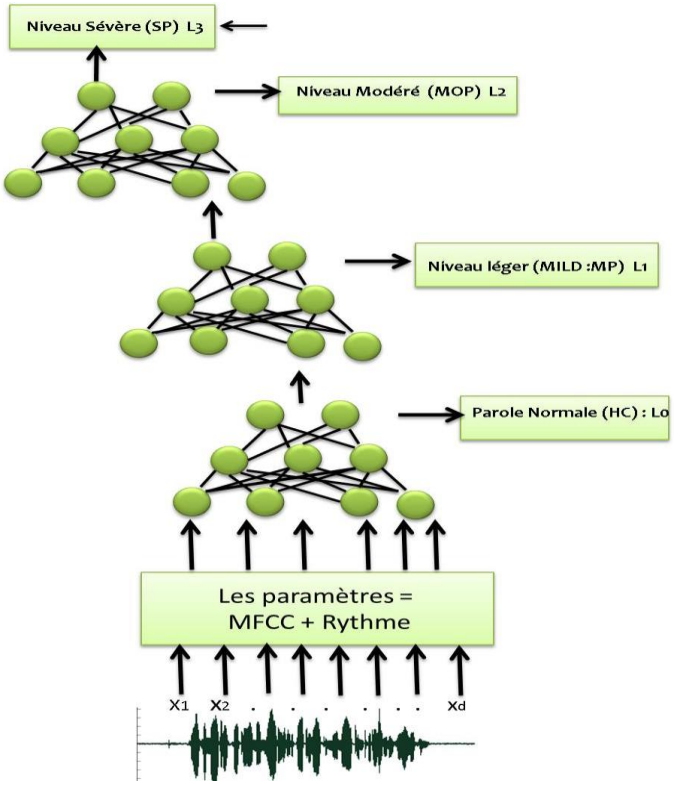

Figure 4. 15 : Structure **hiérarchique du MLP pour la classification de la parole dysarthrique**

Tableau 4. 2 : Comparaison de la méthode hybride proposée avec les systèmes de référence.

Système	**GMM**	**MLP simple**	**MLP Hiérarchique**	**Hybride**
MFCC	80.75	78.58	80.14	**82.64**
MFCC +Paramètres rythmique	83.05	81.42	82.80	**86.35**

4.3.4.5. Évaluation des performances du système de la reconnaissance

Afin d'évaluer le système de reconnaissance de la parole dépendant du locuteur (SD : Spearker Dependant) en utilisant les données des niveaux de sévérité (voir figure 4.16), nous employons la plateforme HTK à base des HMM [46, 136] (voir aussi chapitre 5 pour beaucoup plus de détails sur les HMM et le logiciel HTK). Chaque phonème est représenté par un modèle HMM de 5 états avec deux états non émetteurs (1^{er} et 5^{eme}état). Le même nombre et paramètres MFCC et rythmique sont utilisés dans Selouani et al.[45] (Voir aussi le paragraphe *4.3.4.3*). Soit donc un vecteur de 72 paramètres (65 MFCC + 7 paramètres de rythme). Le modèle de langage utilisé est une bigramme qui dépend du nombre des

statistiques générées à partir de la transcription phonétique. Les statistiques sont générées en utilisant la fonction Hlstats de HTK [136]. Les modèles pour chaque locuteur sont des HMM triphones de gauche-droite avec des probabilités d'émission modélisées par une combinaison linéaire de gaussiennes à matrice de covariance diagonale. En raison de la quantité limitée des données d'apprentissage, les paramètres acoustiques du modèle HMM en contexte dépendant de chaque locuteur, ont été initialisés aléatoirement en utilisant la parole normale de l'orthophoniste comme un système de référence.

Les données de trois locuteurs dysarthriques (BB, JF, BK) sont utilisées pour l'évaluation du système de reconnaissance automatique de la parole dépendant du locuteur (RAP-DL ou SD- ASR). Différentes fenêtre de Hamming sont testées. Les longueurs des fenêtres testées sont de 15, 20, 25 et 30 ms. Pour chaque locuteur, la parole de l'apprentissage est composée de 50 phrases (300 mots) et celle du test est composée de 24 phrases (144 mots).

Deux séries d'expériences ont été réalisées. La première série implique le système ASR purement dépendant du locuteur (SD-HMM) sans évaluation préalable des niveaux de la sévérité. La deuxième série d'expériences est réalisée en utilisant le même modèle HMM à un groupe de locuteurs appartenant au même niveau de sévérité. Nous nous sommes référés à ce dernier système par (SD-NN-HMM) car il utilise la structure connexionniste pour la classification des niveaux de sévérité (Neural Networks). Dans la première série, nous utilisons exclusivement un seul SD-HMM pour chaque locuteur. Dans la deuxième série d'expériences, chaque niveau de sévérité (L_0, L_1, L_2 et L_3) possède ses propres HMM créés à l'aide des données de parole de tous les locuteurs appartenant à cette catégorie. Par exemple, pour le locuteur BB le système SD de L1 sera utilisé.

Le tableau 4.3 montre les taux de la reconnaissance pour différentes longueurs des fenêtres Hamming et le meilleur résultat (en gras) obtenus pour les locuteurs BB, BK, et JF. Ces résultats montrent que la taille de la fenêtre joue un rôle important ce qui confirme les conclusions faites dans Selouani et al.[117]. Par exemple, il était montré qu'il est préférable d'augmenter la longueur de la fenêtre du traitement pour les cas sévères dont le rythme de la parole est très lent.

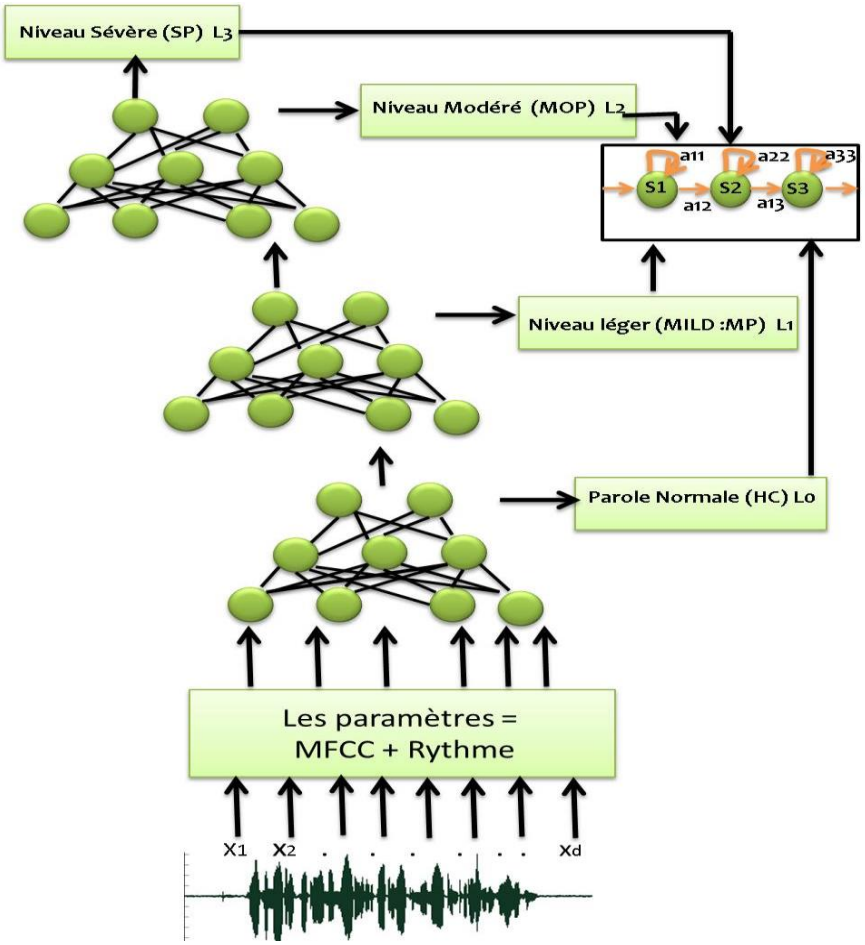

Figure 4. 16 : Structure hiérarchique du MLP pour la reconnaissance de la parole dysarthrique

Tableau 4. 3 : Comparaison des taux de reconnaissances des deux systèmes SD-HMM et SD-NN-HMM relativement aux tailles des fenêtres Hamming.

Système	Locuteur	15 ms	20 ms	30 ms	35 ms
SD-HMM	**BB**	62.5	63.89	65.28	**68.66**
	BK	52.08	55.56	**56.86**	54.17
	JF	61.74	63.82	64.54	**67.14**
SD-NN-HMM	**BB**	60.15	61.68	**63.24**	62.55
	BK	51.28	52.68	**53.64**	53.26
	JF	58.96	60.45	**62.51**	60.73

Les meilleures tailles de fenêtres sont 30 ms et 25 ms pour SD-HMM et SD-NN-HMM respectivement. Le taux de reconnaissance moyen pour les trois locuteurs lors de l'utilisation de la classification préalable des niveaux de sévérité dysarthrique est d'environ 60%, ce qui est considéré comme un résultat très satisfaisant (le cas de la parole pathologique). Les

systèmes purement SD atteignent 64,5% de taux de reconnaissance correcte. Ceci nous amène à conclure que les systèmes SD restent meilleurs que les systèmes basés sur SD-NN-HMM. Toutefois, il convient de mentionner que les systèmes SD-ASR induisent des coûts plus élevés, car ils ont besoin de plus de données et de plus de temps pour l'apprentissage. Mais le système SD-NN-HMM peut être considéré comme une bonne alternative pour résoudre le problème du manque de données pathologiques. En effet, le problème crucial du traitement automatique de la parole pathologique est le déficit important dans la quantité et la qualité des ressources linguistiques.

4.4. Conclusion

Ce chapitre a présenté les méthodes de classification utilisées. Ensuite, les résultats de l'évaluation objective de la parole dysarthrique ont été discutés. Effectivement, nous avons donné un bref aperçu sur la théorie des SVM multi-classes, les GMM et les réseaux de neurones qui ont été utilisés en hybridation avec un réseau bayésien pour distinguer entre les niveaux de sévérité de la dysarthrie ainsi que pour classifier la parole pathologique et normale.

En fait, l'évaluation de la sévérité de la parole dysarthrique était par presque tous les paramètres de rythme existants. Ces paramètres ont été proposés récemment par plusieurs études soit dans le domaine de l'identification des langues ou dans l'évaluation de la parole dysarthrique. Nous avons également proposé une nouvelle mesure de rythme, soit la SRVC : taux de segments de CV qui était sélectionné comme le paramètre le plus pertinent par les analyses DFA et LR. Il est aussi important de noter que cette évaluation a été réalisée grâce à l'utilisation de méthodes d'analyse statistiques et de classification différentes et des données de différents types et tailles utilisant différentes techniques de segmentation (manuelle, automatique, segmentation vocaliques / consonantiques et segmentation voisée / non voisée).

Les résultats des SVM concernant la classification de la parole pathologique contre la parole normale ou ceux de la classification des niveaux de sévérité étaient très satisfaisants. Ils démontrent surtout la pertinence des mesures de rythme dans la caractérisation de la sévérité de la parole dysarthrique. Enfin, Les résultats de segmentation voisée et non voisée sont très prometteurs, comme le sont ceux de la segmentation automatique vocalique, bien que ces séries Set 2 et Set3 ont été utilisées sans aucune correction des erreurs de cette

segmentation. Leurs résultats sont intéressants surtout que la segmentation manuelle de la parole reste toujours un travail très fastidieux et couteux.

Nous avons présenté ensuite le système hybride utilisant une approche connexionniste pour estimer les distributions de probabilités postérieures pour la classification de la parole dysarthrique. L'objectif est de procéder à une évaluation automatique des niveaux de sévérité de dysarthrie. Les entrées sont des vecteurs composés des paramètres classiques MFCC et de quelques paramètres de rythme extraits de la segmentation vocalique /consonantique. Nous pensons que l'approche hybride proposée est pertinente si nous visons à établir une démarche objective pour l'évaluation automatique des troubles de la dysarthrie. Un tel outil sera très utile pour les cliniciens et peut être utilisé pour prévenir les jugements subjectifs inexacts.

Il convient de souligner les limites de la base de données Nemours en raison de son petit nombre relatif des patients participants. C'est pourquoi, à la fin, nous présentons un système hybride utilisant le système connexionniste déjà décrit pour estimer les niveaux de sévérité de dysarthrie et un système dépendant du locuteur pour reconnaître la parole dysarthrique. Les données de chaque niveau de sévérité classé par la structure hiérarchique connexionniste sont les données d'apprentissage des modèles acoustique pour l'ensemble des locuteurs appartenant au même niveau. En effet, les modèles acoustiques d'un locuteur peuvent être entrainés en utilisant les données de l'ensemble de locuteurs appartenant au même niveau de sévérité. Les résultats du système SD-NN-HMM n'étaient pas meilleurs comparativement au système non hybride SD-HMM mais cependant restent satisfaisants. Ceci devra être cependant confirmé en utilisant un échantillon plus large de patients dysarthriques ayant différentes étiologies neurologiques afin d'étudier les effets des paramètres rythmiques sur les différentes tâches de classification ou de la reconnaissance.

Chapitre 5 :

Parole Aphasique (ADAD) :

Résultats Expérimentaux

Chapitre 5 : Parole Aphasique (ADAD Database) : Les résultats expérimentaux

5.1. Introduction

L'objectif de notre travail est de proposer un système de RAP pathologique, capable de reconnaitre de la parole pathologique continue. Donc, dans ce chapitre, nous présentons dans une première partie de la théorie sur les HMM et la plateforme HTK. Aussi de la théorie sur les méthodes d'adaptation des locuteurs : MLLR *(Maximum Likelihood Linear Regression*) et MAP *(Maximum a Posteriori*) permettant de réaliser l'adaptation acoustique pour remédier au problème du manque de données de la parole pathologique et surtout la parole aphasique.

Avant d'entamer la réalisation du système de reconnaissance, nous décrirons d'abord le corpus d'étude puis nous donnerons les différentes conventions de transcription. Ces dernières jouent un rôle très important dans la mesure où elles nous facilitent la réalisation des modèles acoustiques à adapter. En outre, nous mettrons en évidence les caractéristiques de la langue arabe dialectale algérienne présentant de l'intérêt pour la reconnaissance de la parole aphasique. Cela inclut, les caractéristiques acoustiques et les caractéristiques de langage du dialecte algérien qui doivent être soigneusement prises en considération.

Ensuite, nous allons présenter le système de base ou de référence et ses versions améliorées réalisées pour la reconnaissance de la parole aphasique de la base de données ADAD.

5.2. Les modèles de Markov cachés (HMM)

HMM est l'abréviation anglaise qui sera utilisée couramment pour parler des modèles de Markov cachés (Hidden Markov Modèles : HMM). Nous présentons dans ce qui suit la base théorique nécessaire à l'utilisation de ce type de modèle pour la reconnaissance automatique de la parole.

Les modèles de Markov cachés (HMM) représentent un cadre simple et efficace pour la modélisation des séquences de vecteurs spectraux variant dans le temps. En conséquence, la quasi-totalité des systèmes de reconnaissance automatique de la parole (RAP) les utilisent[137, 138]. Par la suite, mous décrivons la plate-forme HTK utilisée pour valider notre système [139].

5.2.1. Un système de référence de RAP à base des HMM.

Les principales composantes d'un système de reconnaissance de la parole continue à grand vocabulaire de reconnaissance sont illustrées par la figure 5.1. Les HMM sont utilisés par les systèmes de reconnaissance de la parole pour faciliter l'identification des mots représentés par le signal acoustique d'entrée.

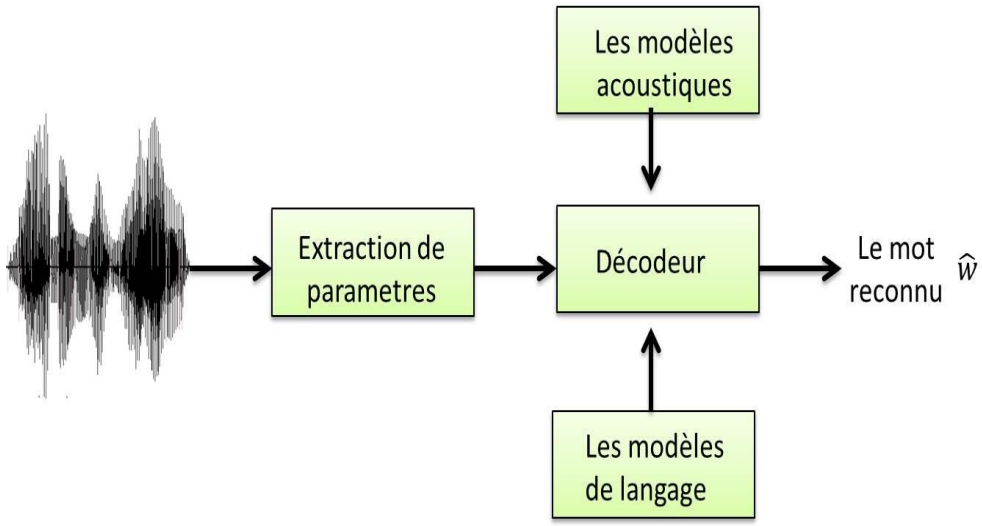

Figure 5. 1 : Système de reconnaissance automatique de la parole (RAP)

Le signal de parole d'entrée est converti en une séquence de vecteurs acoustiques de taille fixe : $O_{1:T} = o_1 \ldots\ldots o_T$ dans un processus appelé extraction de paramètres acoustiques. Le décodeur tente alors de trouver la séquence de mots : $W_{1:L} = w_1 \ldots\ldots w_L$ la plus probable à être générée par O, c'est à dire le décodeur essaie de trouver $\widehat{w}$ comme suit :

$$\widehat{\boldsymbol{w}} = \boldsymbol{arg\,max_w}\{\boldsymbol{P\langle W|O\rangle}\} \quad \textbf{(5.1)}$$

Cependant, puisque $P\langle W|O\rangle$est difficile à modéliser directement, la règle de Bayes est utilisée pour transformer l'équation 5.1 au problème équivalent permettant de trouver :

$$\widehat{\boldsymbol{w}} = \boldsymbol{arg\,max_w}\{\boldsymbol{P\langle O|W\rangle\ P(W)}\} \quad \textbf{(5.2)}$$

La probabilité P⟨O|W⟩ est déterminée par un modèle acoustique. La probabilité a priori P(w) est déterminée par un modèle de langage.

L'unité de base du son représentée par le modèle acoustique est le monophone ou le phonème (phone en anglais). Par exemple, le mot «jeeja» (poulet) se compose de trois

phonèmes /j/e/e/j/a/. Environ 42 de phonèmes sont nécessaires pour le dialecte de l'Est algérien.

Pour tout mot *w* donné, le modèle acoustique correspondant est synthétisé par la concaténation des modèles de monophones pour faire des mots tels que défini par un dictionnaire de prononciation. Les paramètres de ces modèles de monophones sont estimés à partir des données d'apprentissage consistant en signaux de la parole et de leurs transcriptions orthographiques. Le modèle de langage est typiquement un modèle N-gramme dans lequel la probabilité de chaque mot est conditionnée uniquement par ses *N - 1* prédécesseurs.

Comme indiqué plus haut, chaque mot parlé *w* est décomposé en une séquence de sons de base K_w monophones dits de base. Cette séquence est appelée sa prononciation : $q_{1:Kw}^{(w)} = q_1 \ldots\ldots q_{K_w}$. Afin de permettre la possibilité de plusieurs prononciations, la probabilité $P\langle O|W\rangle$ peut être calculée sur plusieurs prononciations comme suit :

$$p(O/w) = \sum_Q P(O/Q)P(Q/w) \quad (5.3)$$

Où la sommation est faite sur toutes les séquences de prononciation valides pour *w*, *Q* est une séquence particulière de prononciations,

$$P(Q/w) = \prod_{l=1}^{L} P(q^{(w_l)}/w_l) \quad (5.4)$$

Où chaque $q^{(w_l)}$ est une prononciation pour le mot *w*.

Chaque monophone *q* est représenté par un HMM de densité continue de la forme illustrée à la figure 5.2 avec les paramètres probabilité de transition a_{ij} et les distributions d'observation de sortie *b(j).* En fait, un HMM effectue une transition de son état actuel à un de ses états connectés à chaque pas de temps. Pour plus de détails sur la théorie des HMM, consulter [140].

Une simple gaussienne multivariée sera prise en charge pour la distribution de sortie :

$$b_j(o) = N(o;\mu^{(j)},\sum^{(j)}) \quad (5.5)$$

Où μ(j) est la moyenne et Σ (j) est la covariance de l'état s_j. Étant donné que le la dimensionnalité du vecteur acoustique y est relativement élevée, les covariances sont souvent limités à être diagonale. Compte tenu du composite HMM, *Q* est formé par la concaténation de tous les monophones *q (w1), ..., q (w),* alors la probabilité acoustique est donnée par :

$$\boldsymbol{p}\left(\frac{\boldsymbol{o}}{\boldsymbol{Q}}\right) = \sum_{\boldsymbol{\theta}} \boldsymbol{p}(\boldsymbol{\theta}, \frac{\boldsymbol{o}}{\boldsymbol{Q}}) \quad (5.6)$$

Où $\theta = \theta_0, \ldots, \theta_{T+1}$ est une séquence des états du Modèle Q et donc, nous aurons

l'équation 5.7 :

$$p\left(\theta,\frac{o}{Q}\right) = a_{\theta_0\theta_1}\prod_{t=1} b_{\theta_t}(o_t)a_{\theta_t\theta_{t+1}} \quad (5.7)$$

Dans cette équation, θ_0 *et* θ_{T+1} correspondent aux états : d'entrée 1 et de sortie 5 non émettrices de la Figure 5.2. Ceux-ci sont fournis pour simplifier le processus de la concaténation des modèles de monophones pour faire des mots.

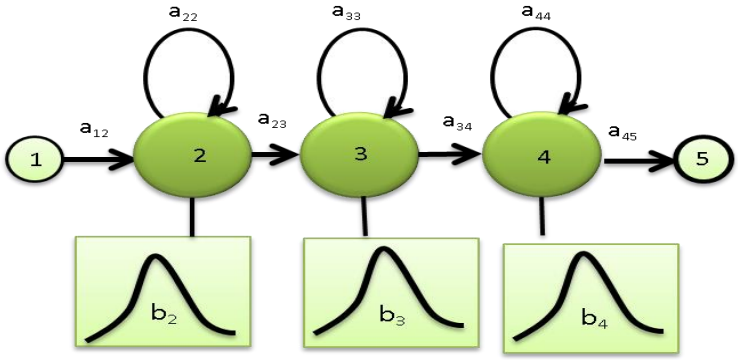

Figure 5. 2 : Le model acoustique phone des HMM

En conclusion, Un HMM est semblable à une machine à états finis dans laquelle les transitions et les émissions sont stochastiques. Chaque état représente un son qui émet de manière probabiliste un vecteur d'observations. Les transitions représentent les différentes possibilités d'enchaîner les sons. Les paramètres du modèle acoustique d'un HMM constitué d'un nombre fini d'états et est défini traditionnellement comme suit :

$$\lambda = [\{a_{ij}\},\{b_j()\}] \quad (5.8)$$

Et :

- Un nombre N d'états numérotés de 1 à N,
- La loi de probabilité initiale λ^0 :

$$\lambda^0 = (\lambda^0_1,.....\lambda^0_N) \quad (5.9)$$

Avec $\lambda^0_i \geq 0$ et $\sum_i \lambda^0_i = 1$

Un choix commun pour le jeu de paramètres initial λ^0 est à imposer à la moyenne globale et à la covariance des distributions gaussiennes et à toutes les probabilités de transition d'être égal. On obtient ainsi ce qu'on appelle un modèle de démarrage à plat « flat start model).

La matrice de transition A, dont les éléments a_{ij} représentent la probabilité de passer de l'état i, à un instant quelconque, à l'état j à l'instant suivant :

$$\sum_{j=1}^{N} a_{ij} = 1 \quad (5.10)$$

Avec $1 \leq i, j \geq N$

La matrice d'observation notée $B = [b_j()]$ caractérise les lois d'observations sur chaque état. Les lois sont en général des mélanges de lois gaussiennes. Chaque (bj) est alors paramétré par(λ_ι, μ_ι, Σ_ι, L=1,.....,L).

L'apprentissage des paramètres d'un tel modèle est réalisé à l'aide de l'algorithme de Baum-Welch et la reconnaissance, fondée sur la recherche du meilleur chemin, est effectuée grâce à l'algorithme de Viterbi.

5.2.2. Modélisation phonétique

Les modèles des mots sont obtenus par la concaténation de plusieurs HMM. Ces HMM correspondent à la suite des unités caractérisant les mots en question. Le problème majeur avec ce qui est la décomposition d'un mot d'un vocabulaire quelconque en une séquence de monophones ou de phonèmes indépendants du contexte ne rend pas compte de la grande variation du contexte qui existe dans le monde réel (l'effet de la coarticulation).

Une façon simple de résoudre ce problème est d'utiliser un modèle de monophones appelé « triphones » ou phonème en contexte dépendant ; il tient compte de deux phonèmes, le précédent et le suivant. Ce modèle est à 3 états par phonème, en faisant l'hypothèse que l'état du milieu modélise la partie stationnaire du phonème et les états extérieurs modélisent la coarticulation avec les phonèmes voisins (figure 5.3)[141].

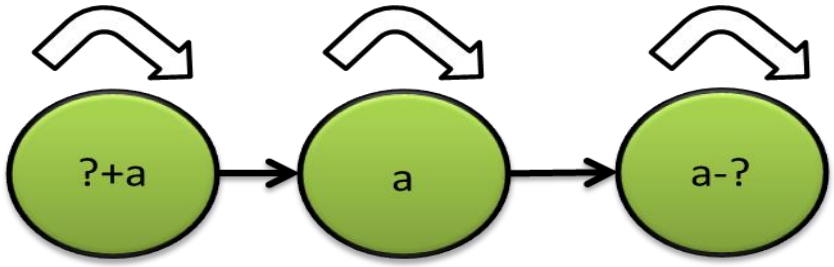

Figure 5. 3 : Exemple d'un modèle de phonème à trois états : le phonème est « a ».

Et s'il y a N phonèmes de base, il existe un potentiel de N^3 triphones. Pour éviter le problème d'éparpillement de données qui en résultent, l'ensemble des triphones logique L peut être associé à un ensemble réduit de modèles physiques P par le regroupement et l'attachement de l'ensemble les paramètres de chaque groupe. Ce processus de mise en correspondance est illustré par la figure 5.4 où la notation x-q q q-y désigne le triphone correspondant à q le phonème dans le contexte d'un phonème x précédent et d'un autre y suivant.

Le regroupement des modèles logiques à des modèles physiques fonctionne généralement au niveau des états du HMM plutôt qu'au niveau du modèle acoustique. Il est plus simple et permet une estimation plus robuste des modèles physiques. Le choix des états à regrouper est généralement fait en utilisant des arbres de décisions [142].

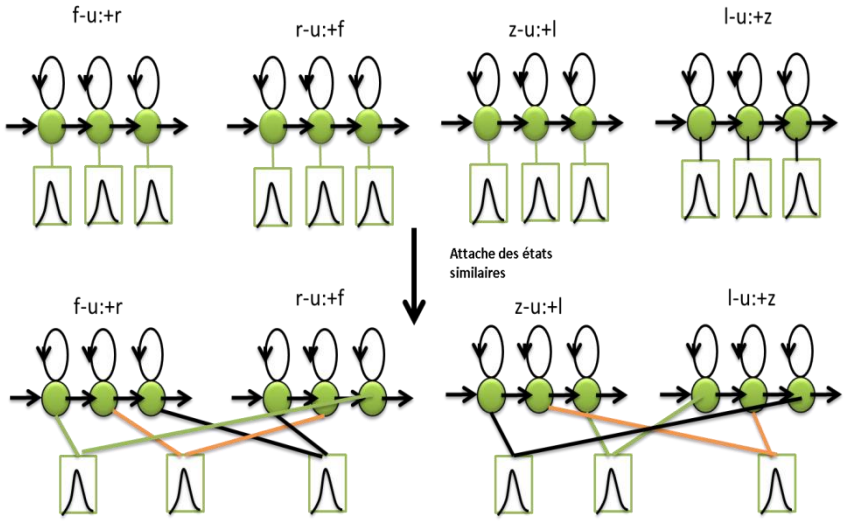

Figure 5. 4 : Formation des modèles de phonèmes aux états attachés « tied-state phone »

Chaque position d'un état de chaque phonème *q* a une arborescence binaire qui lui est associée. Chaque nœud de l'arbre porte une question concernant le contexte. Pour regrouper un état *i* d'un phonème *q*, tous les états *i* de tous les modèles logiques dérivés de q sont collectés dans un seul ensemble au niveau du nœud racine de l'arbre. Selon la réponse à la question à chaque nœud, l'ensemble des états est successivement divisé jusqu'à ce que tous les états aient des groupes distincts (aient des retombées sur les nœuds feuilles 1,2…5). Tous les états de chaque groupe final sont ensuite liés ou attachés pour former un modèle physique (voir figure 5.5).

Pour résumer, les modèles acoustiques de base d'un système de reconnaissance de la parole sont généralement constitués d'un ensemble des HMM liés à trois états avec des distributions de sortie gaussiennes. Ce noyau est généralement construit par les étapes suivantes :

- Un ensemble à plat de démarrage (flat start) monophone est créé, dans lequel chaque phonème de base est un HMM monophonique à une gaussienne avec une moyennes et covariance égales à la moyenne et la covariance des données d'apprentissage.
- Les paramètres des gaussiennes monophones sont réestimés utilisant 3 ou 4 itérations de l'algorithme EM [143].

- Chaque gaussienne d'un monophone q est cloné une seule fois pour chaque triphone distinct x - q + y qui apparaît dans les données d'apprentissage.
- L'ensemble de la formation triphones-données qui en résulte est à nouveau réestimé en utilisant l'algorithme EM et les comptages d'occupation des états de la dernière itération sont enregistrés.
- Un arbre de décision est créé pour chaque état pour chaque phonème de base, les triphones des données d'apprentissage sont mis en correspondance dans un ensemble plus restreint de triphones d'états liés (tied state) et réestimés itérativement en utilisant toujours l'algorithme EM.

Le résultat final est l'ensemble requis des modèles acoustiques à l'état lié dépendant du contexte.

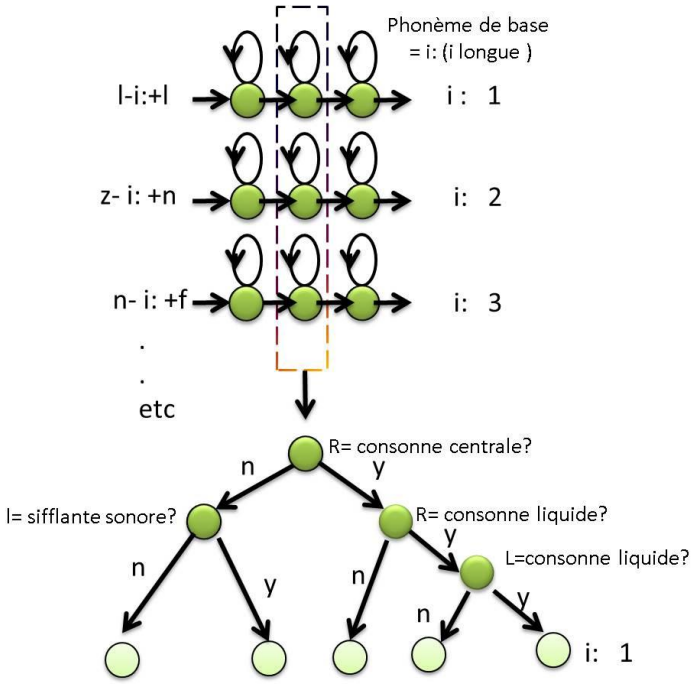

Figure 5. 5 : Clustering ou regroupement suivant l'arbre de décision [144] : adapté aux lexiques de la base ADAD.

Les modèles des mots sont alors obtenus par la concaténation de plusieurs HMM. Ces HMM correspondent à la suite des unités caractérisant les mots en question.

Nous nous inspirons de cette approche pour définir le modèle de notre système de reconnaissance continue pour la parole aphasique.

5.2.3. Utilisation de modèle de mélange de gaussiennes

L'une des extensions les plus couramment utilisés à un HMM standard est de modéliser l'observation de sortie d'un état par un modèle de mélange. Le modèle de l'HMM à une simple gaussienne suppose donc que les vecteurs des caractéristiques observées sont symétriques et unimodale ; mais cela est rarement le cas. Dans la pratique, les différences d'accent et du sexe des locuteurs ont tendance à créer plusieurs modes dans les données. Pour résoudre ce problème, l'observation de sortie d'un état à distribution gaussienne unique peut être remplacée par un mélange de gaussiennes qui est une distribution très flexible capable de modéliser, par exemple, des données à distribution asymétrique et multimodales. L'une des extensions les plus couramment utilisés à HMM standards est de modéliser la répartition état-sortie comme un modèle de mélange. La probabilité pour l'état s_j est maintenant obtenue en additionnant toutes les vraisemblances pondérées par leur probabilités à priori C_{jm} donné par l'équation suivante :

$$b_j(o) = \sum_{m=1}^{M} C_{jm} N(o, \mu^{(jm)}, \sum^{(jm)}) \qquad (5.11)$$

Où C_{jm} est la probabilité à priori pour la composante m de l'état s_j. Ces probabilités à priori doivent satisfaire les contraintes standards des probabilités :

$$\sum_{m=1}^{M} C_{jm} = 1 \qquad \text{et } C_{jm} \geqq 0 \qquad (5.12)$$

Pour la moyenne de la composante gaussienne m de l'état s_j :

$$\hat{\mu}^{(jm)} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T(r)} \gamma_t^{(rjm)} o_t^{(r)}}{\sum_{r=1}^{R} \sum_{t=1}^{T(r)} \gamma_t^{(rjm)}} \qquad (5.13)$$

Où $\gamma_t^{(rjm)} = p(\theta_t = S_{jm} | O^{(r)}; \lambda)$ est la probabilité que la composante m de l'état s_j génère une observation à l'instant t de la séquence $X^{(r)}$ et R le nombre des séquences d'apprentissage.

5.2.4. Adaptation au locuteur : MLLR et MAP

Notre objectif est de bénéficier d'une petite quantité de données de parole dialectale qui est phonétiquement transcrite et bénéficier ainsi des ressources vocales de la parole normale existantes. Notre approche est : d'abord de former un modèle acoustique à partir des données de la parole normale, puis adapter ce modèle avec les données de parole dialectales pathologique en utilisant des techniques d'adaptation des modèles acoustiques.

Il existe deux méthodes d'adaptation des modèles acoustiques bien connues qui sont : la régression linéaire du maximum de vraisemblance (Maximum Likelihood Linear Regression : MLLR) [145, 146] et Maximum A Posteriori (Maximum A-Posteriori : MAP) [147].

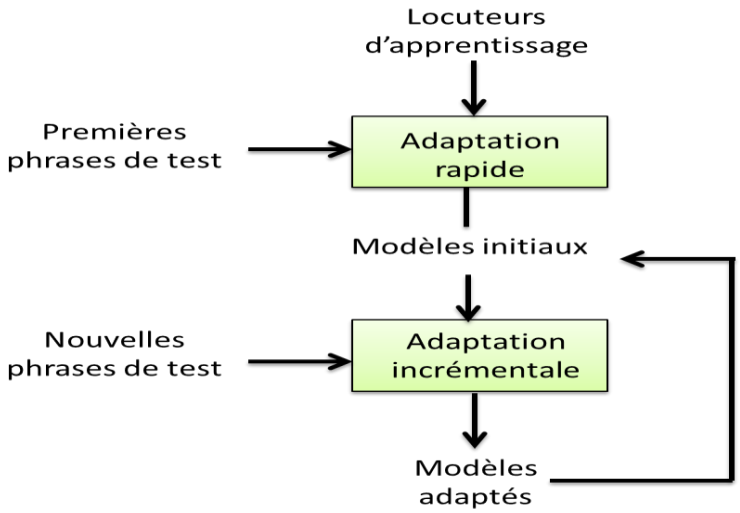

Figure 5. 6: L'adaptation des modèles acoustiques

5.2.4.1. L'adaptation MLLR

L'adaptation MLLR consiste à trouver une matrice de transformation linéaire permettant de convertir le vecteur moyen μ de la gaussienne m de l'état s_j d'un HMM en un vecteur moyen adapté μ_{MLLR} La transformation réalise une combinaison linéaire des gaussiennes du modèle initial pour calculer les nouvelles gaussiennes. Elle peut être simplifiée par l'équation :

$$\mu_{MLLR} = a\,\mu + b \qquad (5.14)$$

Ou a et b les paramètres de la transformation linéaire.

La transformation est calculée de manière à maximiser la vraisemblance des données d'adaptation par rapport au modèle, selon un schéma EM (Expectation/Maximization). Pour augmenter la robustesse de la méthode, les gaussiennes sont regroupées en classes de régression selon un critère de distance dans l'espace des caractéristiques. Les gaussiennes d'une même classe de régression subissent la même transformation[145].

5.2.4.2. L'adaptation MAP

Le principe de l'adaptation MAP consiste, en la connaissance préalable de la distribution initiale des paramètres du modèle [147]. Les données limitées d'adaptation seraient pour modifier les paramètres du modèle initial avec cette préalable connaissance. Ceci empêche une grande modification des paramètres initiaux à moins qu'une quantité suffisante des données d'adaptation soit disponible.

$$\lambda_{MAP}=\arg\max_{\lambda} p\left(\frac{O}{\lambda}\right) p(\lambda) \quad (5.15)$$

Où λ est le modèle des paramètres initial et λ_{MAP} est le model des paramètres adapté en utilisant MAP.

MLLR peut être utilisée comme processus d'adaptation de modèles acoustiques, même lorsque la quantité de données d'adaptation est très petite ; MAP exige une plus grande quantité de données d'adaptation que MLLR pour être efficace, mais si la quantité de données d'adaptation utilisée est suffisante, l'algorithme MAP sera plus performant que MLLR.

5.2.5. La précision de la reconnaissance (Word Error Rate : WER)

Elle a été évaluée comme métrique en termes de taux d'erreur de mots(WER). Le **taux d'erreur de mots**, ou en anglais *Word Error Rate :* **WER**, est une unité de mesure classique pour mesurer les performances d'un système de reconnaissance de la parole.

Le **WER** est dérivé de la distance de Levenshtein[148] en travaillant au niveau des mots au lieu des caractères. Il indique le taux de mots incorrectement reconnus par rapport à un texte de référence. Plus le taux est faible (minimum 0.0) plus la reconnaissance est bonne. Le taux maximum n'est pas borné et peut dépasser 1.0 en cas de très mauvaise reconnaissance s'il y a beaucoup d'insertions.

Après avoir aligné de manière optimale la référence avec le texte reconnu. Le taux d'erreur de mots est donné par :

$$\mathbf{WER} = \frac{\mathbf{S+D+I}}{\mathbf{N}} \quad (5.16)$$

Où :

- N est le nombre de mots de référence,
- S est le nombre de substitutions (mots incorrectement reconnus),
- D est le nombre de suppressions (mots omis),
- I est le nombre d'insertions (mots ajoutés),
- H est le nombre de mots correctement reconnu.

• Le **taux de reconnaissance de mots**, ou *Word Accuracy* (**WAcc**) en anglais, est défini ainsi :

$$W_{Acc} = 1 - WER = \frac{N-S-D-I}{N} = \frac{H-I}{N} \quad (5.17)$$

5.2.6. La plate-forme HTK

La plate-forme **HTK** (Hidden Markov Model Toolkit), développée par l'Université de Cambridge[149] fournit tous les outils logiciels nécessaires à la réalisation des systèmes de reconnaissance de la parole fondés sur des modèles HMM.

Notre choix s'est porté vers cette plate-forme pour plusieurs raisons. Les deux principales raisons sont :

- HTK permet de construire la chaîne complète : apprentissage et reconnaissance de la parole.
- HTK est un système largement utilisé dans le domaine du traitement automatique de la parole. ceci peut permettre à comparer facilement les résultats obtenus et les données utilisées.

5.3. Évaluation objective de la parole aphasique

5.3.1. Description du système de référence

Dans le cadre de la reconnaissance de la parole continue indépendamment du locuteur, notre objectif est de développer un système de référence basé sur les modèles de Markov cachés [140, 150-154] en utilisant la parole normale. Ensuite une normalisation doit être faite au niveau du corpus de la parole normale et pathologique. La parole pathologique normalisée est utilisée pour adapter le système de référence basé sur la parole normale pour avoir un système plus performant pour reconnaitre la parole pathologique. Le système de référence est mis au point à partir de la plate-forme HTK [155] et sur la base de données des locuteurs normaux ou sains. Sa réalisation est décomposée en deux étapes. Un système de base est tout d'abord construit avec des modèles phonétiques indépendants du contexte ; un certain nombre de choix sont faits sur ce système, comme le nombre d'états des modèles, le type de densités de probabilité d'émission associées aux états et la représentation du signal par des coefficients cepstraux et des coefficients différentiels. Un système de référence plus performant est ensuite développé à partir du système de base ; il est construit avec des modèles phonétiques dépendants du contexte. Les résultats en reconnaissance de notre système de référence le situent à un niveau de performance satisfaisant. Les traitements qui nous semblent pertinents

dans le domaine de l'adaptation au locuteur seront intégrés au système de référence et évalués au moyen d'expériences de décodage acoustico-phonétique.

5.3.2. La normalisation phonétique

Afin de bénéficier des ressources existantes de la parole normale, une approche de modélisation acoustique est proposée. Elle est basée sur l'état de l'art des techniques d'adaptation des modèles acoustiques, y compris MLLR et MAP. Malheureusement, comme, la parole normale et pathologique ne partagent pas le même ensemble de phonèmes ; il n'est pas possible d'appliquer les techniques d'adaptation de leurs modèles acoustiques respectives. Effectivement, les phonèmes de la parole pathologique sont en générale altérés et déformés. Afin de rendre l'adaptation possible, nous allons proposer une approche de normalisation des ensembles de phonèmes de la parole normale et la parole pathologique de l'ADAD.

Nous citerons certaines erreurs ou quelques troubles de parole connues chez les patients aphasiques. Nous tenons surtout à préciser que le patient parfois déforme totalement le mot prononcé mais l'orthophoniste le compte comme une répétition correcte. Effectivement, le mot prononcé par le patient est proche du mot à répéter soit morphologiquement soit dans le sens. Le patient est conscient du sens du mot à répéter mais ne réussit pas à le prononcer. En général, ces altérations consistent en :

- des erreurs de substitutions comme : la lettre s(س) au lieu de ch, s(ش) ;
- l'omission d'une consonne ou d'une syllabe finale du mot à répéter ou juste prononcer tout simplement le début du mot ;
- le changement d'ordre de deux phonèmes : désintégration phonétique ;
- la présence d'anomalies de débit verbal (la fluence) ;
- la tendance à produire les fricatives comme des occlusives : /s/ produit [t], par exemple. Le patient a des difficultés à distinguer en production les différents modes d'articulation (occlusive, fricative, affriquée) ;
- la tendance à produire tous les sons trop tendus, ce qui exige un effort articulatoire global très important.

En conclusion, toute altération, hésitation, substitution, omission, ou autres troubles sont transcrits comme phonèmes correctement prononcés et donc vont apprendre au système de reconnaissance les déficits phonétiques aphasiques.

Nous avons résumé les principales différences phonétiques entre les sous dialectes régionaux. Les régions d'où viennent les patients sont les suivantes : Annaba, Skikda, Tébessa, Sétif qui sont des régions correspondant à la partie Est d'Algérie.

Cela a signifié que le parler dialectal des aphasiques comprend des écarts par rapport à la langue standard qui sont d'ordre phonologique. Ceci peut donc être source d'un problème de compréhension.

Nous avons considéré les régions d'Annaba, de Tébessa, et de Skikda comme une seule région que nous avons dénommée Région d'Annaba contre celle de Sétif et ses environs. Nous pouvons voir les caractéristiques de ces deux sous dialectes résumées dans le tableau 5.1.

Tableau 5. 1 : Normalisation phonétique dialectale

phonèmes différents	/j/ (ج)	/dh/ (ذ)/	/th /(ث)	dh_/ (ظ, ض)	/g/	/i: /	/i/	/t/
Région d'Annaba	z (ز)	/d/ (د)	/t/	/d/ (د)	/q/ (ق)	/ey/	//	/ts/
Sétif	/j/ (ج)	/dh/ (ذ)/	(ث)	/d / (د)	/g/	/i: /	/i/	/t/

Dans la Région d'Annaba, on remplace le phonème par /j/ (ج) par / z/ dans certains mots ; par exemple : mzouez au lieu de mezouej qui veut dire il est marié, aazouja par aazouza qui veut dire vieille femme. À Skikda ou à Constantine, on trouve la lettre t se prononce comme tch ou tsi respectivement. En général, une confusion se trouve aussi entre : la lettre /p/ et la lettre /b/ et entre la lettre /f/ et /v/ selon que la personne est cultivée ou non. À la fin en rajoute une dernière normalisation concernant les lettres emphatiques. Que la lettre soit emphatique ou non, elle est considérée comme une variation de prononciation et non pas comme une différence. Par exemple pour le mot ba_ba_ est considéré comme une deuxième prononciation pour baba ou encore baba_ ou encore ba_ba. Donc cette normalisation ignore l'emphase comme une différence et la considère comme une variation.

Le système vocalique est totalement diffèrent d'une ville à l'autre. Une nouvelle normalisation est considérée est donné par le tableau 5.3.

5.3.2.1. Le nouveau système de transcription

Nous avons précisé dans le chapitre 2 que nous avons opté pour le code de transcription proposé par Nacera Zellal dans [75]. Mais, pour des raisons de simplicité et pour des raisons techniques concernant la lecture des chaines de caractères dans HTK, nous avons apporté bon nombre de modifications de ce code. Nous proposons un nouveau code qui est une combinaison du code SAMPA, le code ACA utilisé dans les conversations arabes sur internet (ACA : Arabic Chat Alphabets) et aussi le code phonétique utilisé par Zellal. Le résultat est un nouveau système de transcription plus pratique pour le dialecte algérien donné par le tableau 5.2 et 5.3.

Tableau 5. 2: Le Nouveau système proposé pour la transcription pour les consonnes.

SY Ar	Code Sampa	Symbol	Mot-clé dialectal	Signification en français
	[P]	[P]	[Plaas]	Place
ب	[b]	[b]	[bèèba]	Papa
/	[v]	[v]	[viilo]	Vélo
م	[m]	[m]	[muus]	couteau
ف	[f]	[f]	[fiil]	éléphant
ت	[t]	[t]	[ta?'], [tmar]	De, dattes
ط		[t_]	[t_a_a_qa]	fenêtre
د	[d]	[d]	[denya]	Vie
		[d_]	[d_oor]	Tourne-toi
ر	[r]	[r]	[riix]	Vent
		[r_]	[r_a_a_x]	Il est parti
س	[s]	[s]	[senna]	Dent
ص		[s_]	[s_ghéér]	Petit
ز		[z]	[ziin]	Beauté
ل	[l]	[l]	[liil]	Nuit
ن	[n]	[n]	[niif]	Nez

ث		[th]	[the?'leb]	Chacal en arabe classique
ش		[ch]	[chuuf]	Vois
/		[dj]	[djbel]	montagne
ج		[j]	[jbel] dans l'est algerien	Montagne
/		[tch]	[tchuuf]	Tu vois
ك		[k]	[kursi]	chaise
/		[g]	[gantra]	Pont
خ		[kh]	[khyt_]	Fil
غ		[gh]	[gha_a_r]	Trou
ح		[x]	[xeel]	temps
ع		[?']	[?'a_yn]	Œil
ه		[h]	[huuwwa]	Il
ء		[?]	[s ?el]	Il a demandé
ذ		[dh]	[dhkar]	Male
ظ/ض		[dh_]	[dh_rab]	Il a tapé
ق		[q]	[qa_lb]	coeur
و		[w]	[wèèlu]	Rien
ي		[y]	[yuum]	Jour

Tableau 5. 3 le nouveau système de transcription pour les voyelles

Code Sampa	Symbol	Mot-clé dialectal	Signification en français
	[a]	[kla]	Il a mangé
	[a_]	[t_a_b]	Il est cuit
	[è]=e	[chèèf]	Il a vu
	[é]=y	[khéér]	Bien

	[e]	[rajel]	Homme
	[i]	[benti]	Ma fille
	[o]	[fooq]	Dessus
	[u]	[?'_umro]	Son âge
	[eu]	[?'euchch]	Nid

- **Signes diacritiques :** deviennent comme suit :

Emphase : c_ v_ ; voyelle nasalisée : vn ; voyelle longue : vv : gémination vv,cc

5.3.2.2. Le corpus linguistique avec le nouveau système de transcription

- **Parole Spontanée**

wechesmeuk (quel est votre nom ?)

gedeeh fi ?'omrek (quel âge vous avez ?)

wiinkhlaqti (où vous êtes né ?)

wechtakhdem (qu'est que vous faite comme travail ?)

?axkilii ?'la khadmetek(parlez-moi de votre travail)

?axkilii ?'la ?'ayltek(parlez-moi de votre famille)

?axkilii ?'la ma_rdak(parlez-moi de votre maladie)

- **Séries Automatiques**

gulliiyeemesmana (citez-moi les jours de la semaine)

gullichhoora (citez-moi les mois de l'année)

?axseblihattal?'achra(comptez jusqu'à dix ou vingt)

Répétitions

a) Syllabes

Tableau 5. 4 : Répétition des syllabes

1	ba	7	ab	12	du	18	ud	24	fy	30	yf	36	ry	41	yr
2	bo	8	ob	13	ko	19	ok	25	fi	31	if	37	za	42	az
3	ti	9	it	14	ga_	20	a_g	26	su	32	us	38	kha_	43	a_kh
4	ly	10	yl	15	ra_	21	a_r	27	s_u	33	us_	39	gha_	44	a_gh
5	?'a	11	a?'	16	cha	22	ach	28	qa	34	aq	40	ha_	45	a_h
6	?a			17	ya	23	ay	29	xa	35	ax	46	chu	47	uch

b) Logatomes

Tableau 5. 5 : Répétition des logatomes

1	kro	7	fra	13	sky	19	xko	25	khli	31	ska	37	?leuf
2	sbi	8	bli	14	s_t_a_	20	baan	26	xro	32	?'fa	38	fha
3	dh_ry	9	tru	15	kla	21	suun	27	kwa	33	ghna	39	xna
4	blo	10	flu	16	bro	22	t_yyn	28	t_ra_	34	?'t_a_	40	xfa
5	gro	11	xyy	17	fri	23	chlu	29	sla	35	ghsi	41	?'qa_
6	xfy	12	t_qa_	18	xma	24	ghra_	30	ghza	36	?'ra_	42	?'da

c) Les mots

Les mots sont deux catégories : simple et complexe. Les mots simples se divisent en mots simples à consonnes et en mots simples à voyelles (voir les tableaux 5.6 et 5.7 et 5.8).

Tableau 5. 6 : Répétition des mots contenant les consonnes.

1	gh	gha_a_r	22	ba_ghla	42	t_	t_yyn	63	xa_t_t_
2	l	liil	23	lliil	43	d_	d_a_a_r	64	?'ad_d_ ou ?'a_dh_dh_

3	f	fuur	24	ruuf	44	r	ra_a_x	65	xa_rr
4	s	siil	25	lbees	45	r	riix	66	deer
5	z	zuul	26	luuz	46	y	yuum	67	jeeyy
6	s_	s_yyf	27	kha_a_s_s_	47	k	kiif	68	fiik
7	ch	chaaff	28	feuchch	48	g	gee?’	69	deugg
8	m	ma et malika	29	leem	49	kh	kheef	70	deekh
9	n	na_a_r	30	been	50	q	qa_a_m	71	xma_q
10	b	beeb	31	beeba	51	?’	?’eem	72	fee?’
11	t	tiliifon	32	meet	52	?	?enam		
12	d	diir	33	reed	53	h	heem	73	ddeeh
13	fl	fla	34	deufl	54	f_r_	fra_?’	74	dh_a_fr
14	bl	bleed	35	lbla	55	bgh	bgha_	75	teubgh
15	t_r	t_ra	36	fatr	56	d_r	d_ra		
16	kl	kla			57	gl	gleub		
17	kr_	kra_	37	fa_kr	58	ghr_	ghrob	76	s_oghr
18	sb	sbar	38	xa_sb	59	sk	skeun	77	meusk
19	kt	kteeb	39	breukt	60	ght_	ght_a_	78	dboght
20	bt_	bt_a_	40	?’eubt	61	fn	fna	79	jeufn
21	sm	sma	41	?eusm	62	qr	qra_	80	xaqr

Tableau 5. 7 : Répétition des mots contenant les voyelles

81	oo (o:)	looto	85	a_	qa_hwa	89	ø(eu) kheubz
82	Uu (u:)	kuury	86	e:(ee)	jeeja	90	ô(on) zonqa_
83	an(â)	sandooq	87	?	?eulf	91	eu ?’euchch
84	é: (yy)	t_ yyr	88	i:(ii)	BIIR	92	t_omoobiil

Tableau 5. 8 : Répétition des mots complexes

1	ba_s_tyla	2	ksebtha	3	sukka_rtek	4	ttiiliifuun
5	chbektha	6	sbaaxelkhiir	7	chekchuukaxa_mra_	8	amentniyaak
9	lb_aar_a_a_s_yuun	10	salaamu?'alikom	11	?eumseulkhiir	12	smaxly
13	s_a_xxa_	14	beuslaama	15	Chekhchoukhaxa_mra_	16	?'em

a) Phrases

1. lliil been (la nuit vient)
2. lbalonlas_far b?`iid (le balon jaune est loin)
3. chchemes charqet wrajjbaal fel?`achwa (le soleil s'est levé derrière les montagnes le soir)
4. lmesfuuf elli yumma weklaatu ls_abriina ?awskhu:n waxluw (le masfouf (genre de couscous) qui a fait manger maman à sabrina est chaud et sucré)
5. lebgar ta?` jaarii serxo felgaba legriba ?elli menewra haadii ddaar (les vaches de mon voisin sont au proche foret derrière cette maison)
6. chra fiil (il a achté un éléphant)
7. wa?'laah babaah ghaab (pourquoi son père est-il absent)
8. jaami jaa felwaqt (il n'est jamais à temps)
9. jaami lgaah felwaqt (il ne l'a jamais trouvé à temps)
10. ?'t_ak kull es_waared jiibhem ehna (il t'a donné tout l'argent, ramener les ici)
11. metel?'bch eltem luukaantelgaw laxwaayej elliitesxa_qqhem (ne jouez pas la bas si vous trouvez les choses dont vous avez besoin).
12. lmaa baared (l'eau est froide)
13. nnwaad_r twadro (les lunetes sont perdues

14. lqahowa ta?’ s_bah hada morah (ce café du matin est amer)
15. dar ?ammi hmed lmabniya belberik lexmar melixa (la maison de mon oncle qui est construite avec le brike est belle)
16. train ta?` qsant_ina yuus_al nis_af see?’a qbal ma yuus_al ta?’ st_if (le train de Constantine arrive demi-heure avant celui de Sétif)
17. mohamed chchra furmaj wo khobz wo rax lljazar yachri llxam (mohamed a acheté fromage et pain et il est allé au boucher pour acheter de la viande)
18. mraa tiliifonii (une femme téléphone)
19. rajal y?’om felbxar (un homme nage dans la mer)
20. t_afela tajeri (une petite fille coure)
21. mraa taqeraa felketab (une femme lit dans le livre)
22. t’afela tazegii felnewaar (une petite fille arrose les fleurs)
23. jazaaryegat_a?’ fellexam (le boucher coupe la viande)
24. liinda taakol fettofaax (lynda mange la pomme)

5.3.3. Caractéristiques du système de base

Les choix qui ont été faits pour le système de base ou de référence de la RAP sont les suivants :

- Représentation centième des secondes du signal par 12 MFCC, l'énergie, 13 coefficients différentiels du premier ordre, et 13 coefficients différentiels du second ordre soit 39 coefficients ;
- Unités acoustiques : 43 phonèmes indépendants du contexte ;
- Topologie des modèles : modèle de Bakis à 3 états émetteurs ;
- Probabilité d'émission modélisée par une combinaison linéaire de gaussiennes à matrice de covariance diagonale ;
- Modèle de langage : bigramme phonétique.

Tous les modèles ont la même topologie, et les probabilités d'émission de tous les états sont représentées par un nombre identique de mélange de gaussiennes. Les modèles

acoustiques et le modèle de langage sont appris sur le corpus de répétitions de syllabes, logatomes, mots simples et complexe et aussi les phrases. L'apprentissage est réalisé en modèles isolés avec les outils HInit pour l'initialisation par les k-moyennes segmentales et HRest pour la réestimation par la procédure de Baum-Welch.

5.3.4. Les résultats du système de référence de la parole pathologique

Les nombres optimisées des états liés et gaussiennes par état ont été fixés à 250 et à12 respectivement. La recherche a été effectuée dans l'intervalle de 62 à 1000 états liés : TS et de 1 à 18 gaussiennes, comme indiqué dans le tableau 5.9. Pour chaque combinaison des états liés (Tied States : TS) et des densités gaussiennes, un test de reconnaissance a été effectué en utilisant un ensemble de test composé de 15 patients aphasiques.

Les résultats du décodage de la base de données de la parole normale qu'est une partie de la base totale aphasique ADAD. Pour des raisons pratiques ; nous allons designer cette base de données par : HDAD (Healthy Dialectal Algerian Database). Donc, en utilisant la parole normale aussi comme données de test, la valeur absolue du WER obtenue est de **14.2 %** comme c'est montré dans le tableau 5.9.

Tableau 5. 9: WER en fonction du nombre de TS et le nombre de gaussiennes pour la base de données de la parole normale (HDAD).

Gaussienne	**Les états attachés ou liés (tied stats)**				
	62	**125**	**250**	**500**	**1000**
1	28 %	27.2 %	17.5 %	16 %	25.3%
12	23.4%	21.5%	**14.2%**	15.8 %	27.5 %
18	18.7 %	17.2.1%	14.8 %	15.5 %	30.6%

5.3.5. Les résultats du système de référence après la normalisation phonétique

Après la normalisation des phonèmes de la parole normale et la parole aphasique, il est possible d'utiliser des modèles acoustiques de la parole normale au lieu de ceux de la parole aphasique. L'ensemble du corpus de la parole normale (50 locuteurs entre femmes et hommes adultes) a été utilisé pour former les modèles acoustiques avec un nombre variable de TS (62-1000) et des densités gaussiennes (de 1 à 18).

La précision de la reconnaissance de la parole a été évaluée sur la parole aphasique qui se compose de 15 patients aphasiques. Le WER absolue a été évalué à 17.5%. Nous nous attendions à l'augmentation du WER, selon que les modèles acoustiques initiales de la parole normale sont des modèles acoustiques indépendants du contexte.

Tableau 5. 10 : WER en fonction du nombre de TS et le nombre de gaussiennes pour la base de données Aphasiques ADAD.

Gaussienne	Les états attachés (tied stats)				
	62	**125**	**250**	**500**	**1000**
1	30 %	30.2 %	29.9 %	30.4 %	27.5%
12	27.4%	28.2%	31.2%	27.7 %	26.8.5 %
18	25.5 %	29.5%	29.6 %	**17.5 %**	30.6%

Comme nous l'avons constaté, les modèles acoustiques formés avec les données de la parole normale peuvent être adaptés pour reconnaître la parole pathologique. Cette approche peut être très utile lorsque les ressources de la parole pathologique ne sont pas disponibles ou très limitées.

5.3.6. Regroupement ou pooling des données normales et pathologiques.

Un regroupement de données de la parole normale et pathologique a été utilisé pour apprendre les modèles acoustiques du nouveau système. Nous avons crié une banque de données de la parole qui se compose de presque12 heures de la parole normale ou saine conjointement avec presque 5 heures de la parole aphasique. Enfin, les résultats de la reconnaissance de la parole n'ont indiqué pratiquement aucune réduction dans le WER par rapport au système de référence de la parole pathologique. Le WER a même augmenté, comme indiqué dans le tableau 5.11 de 17.5 % à 20.2 %. Effectivement, la quantité de données normales était plus que le double de la quantité des données dialectales. Par conséquent, les données normales ont beaucoup influencé les caractéristiques acoustiques des données de parole aphasique.

5.3.7. Le système de reconnaissance automatique adapté

Trois techniques d'adaptation de modélisation acoustique ont été évaluées :

5.3.7.1. Adaptation MLLR

Adaptation MLLR a permis d'obtenir de meilleurs résultats lors de l'adaptation de tous les paramètres des modèles acoustiques : les poids, les moyennes et la variance des mélanges gaussiennes ainsi que les probabilités de transitions. Trois itérations de MLLR ont été

appliquées sur les moyennes du mélange des gaussiennes. Dans chaque itération, les moyennes ont été transformées en utilisant la matrice MLLR. Le système adapté a abouti à un WER absolu de 13.5% une réduction de 0.7% du WER du système de base ADAD comme indiqué dans le tableau 5.11.

5.3.7.2. Adaptation MAP

L'adaptation MAP aussi a donné de bonnes performances. Le WER absolu était de 14.6 % (voir le tableau 5.11).

5.3.7.3. La méthode d'adaptation hybride : MLLR-MAP

En fait, les deux processus d'adaptation peuvent être combinés pour améliorer les performances, en utilisant les moyennes MLLR transformées *a priori* pour l'adaptation MAP par la suite. Par conséquent, nous avons combiné les deux MLLR et MAP, en commençant par deux itérations de MLLR suivies par l'adaptation MAP. Le WER absolu était 12.1 % ; le meilleur de tous les systèmes réalisés. Une réduction de 2.1% du WER réalisé avec le système de base ADAD comme le montre le tableau dessus.

Tableau 5. 11: Les résultats du décodage de l'ADAD utilisant l'ADHD et les différentes méthodes d'adaptation

Modèles Acoustiques	WER
ADAD	17.5%
HDAD	14.2%
ADAD / HDAD : POOLING	20.2 %
ADAD / HDAD Adaptation MLLR	13.5 %
ADAD / HDAD Adaptation MAP	14.6 %
ADAD / HDAD Adaptation MAP+MLLR	12.1%

5.3.8. Application du système adapté MLLR-MAP : Automatisation de l'aphasiogramme

Afin de rendre notre travail utile pour l'orthophoniste lors du diagnostic et de la rééducation du patient, nous avons réalisé la reconnaissance sur les différentes taches de répétition chacune à part pour imiter le travail de l'orthophoniste. Le résultat est l'automatisation de l'aphasiogramme. Un exemple d'un aphasiogramme automatisé est donné par la figure 5.7.

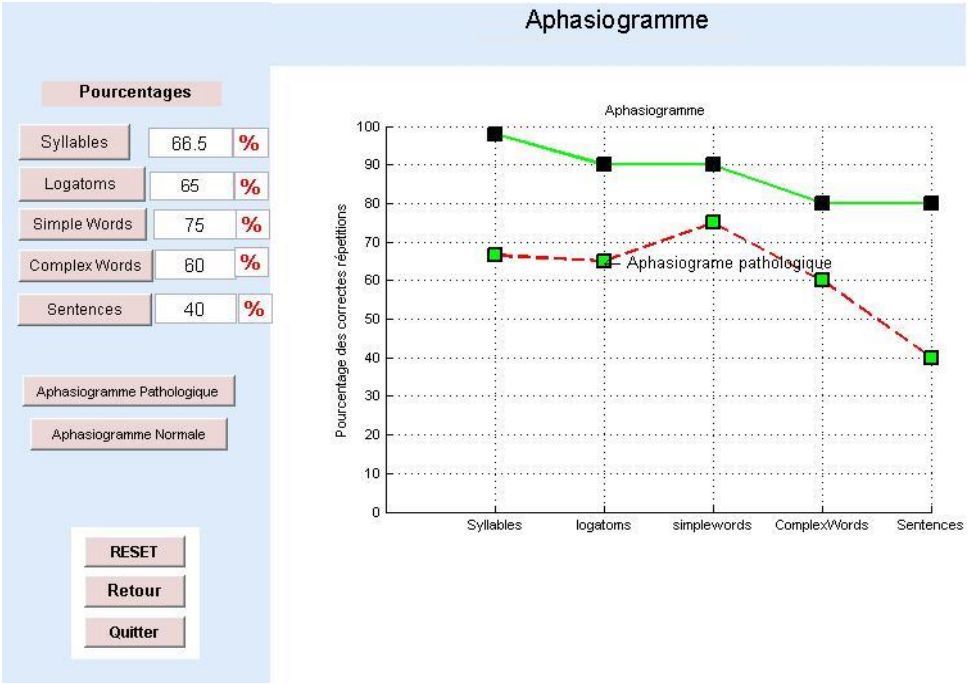

Figure 5. 7 : la comparaison entre un aphasiogramme d'un patient aphasique et celui d'un sujet normal.

Nous avons adapté la parole des patients aphasiques au fur et à mesure à la parole normale d'un locuteur normale. Pour apprendre les modèles acoustiques du système de référence, nous avons laissé un locuteur normal répéter chaque tache 20 fois (20 répétitions de chaque syllabe, de chaque logatome…). En tenant compte de la normalisation phonétique et la technique de l'adaptation hybride MAP+MLLR, nous avons comparé la parole de chaque patient à la parole normale de référence. Les résultats étaient satisfaisants dans le sens où les aphasiogrammes automatisés des différents patients n'étaient pas loin de ceux réalisés subjectivement par l'orthophoniste. Les résultats étaient satisfaisants toujours en considérant la difficulté de reconnaitre la parole pathologique due à plusieurs facteurs.

5.4. CONCLUSION

Ce dernier chapitre se focalise sur la reconnaissance de la parole aphasique à travers la base de données ADAD. Grâce aux données de notre corpus du bilan aphasique, nous avons essayé de constituer un système de reconnaissance pour la parole pathologique en bénéficiant de la parole normale que nous avons fort exploitée. Nous avons réussi à atteindre l'objectif de cette partie de notre travail à savoir l'automatisation de l'aphasiogramme.

Effectivement, la technique présentée dans ce chapitre est basée sur le formalisme des

HMM (classiquement utilisé en RAP). Nous avons donc proposé une optimisation des HMM en termes de ressources linguistiques de la parole normales et des techniques d'adaptation des locuteurs (MLLR et MAP) qui ont permet d'atteindre un niveau de performance intéressant. Les résultats présentés dans ce chapitre montrent que notre approche permet une réduction relative du taux d'erreur de 2.1 % (WER 12.1 % avec la combinaison de MLLR/MAP) comparée à un système de référence HMM classique pour la parole pathologique.

Une suite logique de ce travail serait d'envisager différentes formes d'adaptation avec une base de données plus large de parole normale et de parole pathologique de meilleure qualité et surtout une étude linguistique morphologique phonétique des différents dialectes algériens. Nous avons essayé de faire une sorte normalisation au niveau du dialecte de l'est algérien mais cet effort reste insuffisant. Cette étude doit être faite par les spécialistes.

CONCLUSION GÉNÉRALE

Chapitre 6 : Conclusion Générale

6.1. Revue des réalisations

Dans ce travail, nous avons proposé et présenté une approche basée sur la paramétrisation rythmique et prosodique spécialement pour l'analyse de la parole pathologique. Les paramètres rythmiques basés sur les caractéristiques temporelles d'intervalles vocaliques et intervocaliques et les indices de variabilité par paires. Les résultats sont très encourageants et démontrent la capacité des paramètres rythmiques de séparer la parole pathologique de la parole normale.

Le corpus pour la parole dysarthrique était divisé en trois sous-ensembles. Set 1 consiste en huit phrases pour chaque locuteur segmentées manuellement en voyelles et consonnes (176 phrases). Set 2 contient l'ensemble des phrases (740 phrases) répétées uniquement par les patients dysarthriques segmentées automatiquement en voyelles et consones. Set 3 consiste en 74 phrases pour chaque patient et pour l'orthophoniste segmentées automatiquement en segments voisés et non voisés (1628 phrases).

Le présent document pose la question suivante : est ce que les mesures de rythme proposées sont capables de discerner les niveaux de sévérité de la parole dysarthriques ? Les résultats de différentes expériences de classification montrent que les mesures de rythme peuvent être utilisées efficacement pour caractériser la sévérité dysarthrique.

Les paramètres pris en considération sont proposés récemment dans le domaine d'identification des langues et le domaine de la classification de la parole pathologique[39, 41, 43-46, 74, 94, 96, 97, 110, 156, 157]. Ils sont pour la segmentation vocalique : %V, ΔV, ΔC, nPVI-V, rPVI-V, nPVI-C, rPVI-C, VarcoV, VarcoC, VarcoVC, nPVI-VC, rPVI-VC. Par analogie, nous avons proposé des paramètres rythmiques basés sur la segmentation voisée et non voisée : %VO, ΔVO, ΔUV, nPVI-VO, rPVI-VO, nPVI-UV, rPVI-UV, VarcoVO, VarcoUV, VarcoVOUV, nPVI-VOUV, rPVI-VOUV.

En plus de ces paramètres, nous proposons un nouveau paramètre que nous appelons SRVC (SRVOUV pour la segmentation voisée et non voisée). Un nouveau paramètre de débit de parole (Speaking Rate : SR) au lieu de la définition classique du débit (le nombre de

syllabes / la durée totale de l'énoncé de parole étudié). Il s'agit d'une approximation de la durée d'une syllabe par la durée d'intervalles successifs de voyelles et consonnes (exemple VC, VCC).

La première section des expériences, rapporte les résultats obtenus suite à l'étude statistique pour déterminer la signification des paramètres et discerner les plus pertinents. Pour valider les résultats, une classification rythmique par le biais des représentations bidimensionnels a été faite. Les résultats de classification montrent la pertinence des mesures rythmiques pour distinguer la parole normale de la parole dysarthrique. La majorité des paramètres avaient obtenu plus de 90% de taux de réussite dans le classement de la parole dans leurs groupes appropriés.

Toujours, suite à l'étude statistique faite sur les mesures de rythme. Une série d'expériences de classification de la parole pathologique vs. La parole normale d'une part et la classification des niveaux de sévérité d'autre part était effectuée. Nous avons réalisé deux systèmes : classification par la méthode SVM et un système hybride basé sur la classe des distributions à posteriori combinés avec une structure hiérarchique de perceptrons multicouches.

Les résultats démontrent la pertinence des mesures du rythme et l'efficacité du système de classification hybride proposés pour discriminer les niveaux de sévérité de la dysarthrie. La méthode de classification SVM donne de bons taux de classification correcte surtout pour la classification de parole pathologique / parole normale. Les résultats de la classification des niveaux de sévérité étaient satisfaisants mais le problème réside dans la classification de la parole modérément altérée. La parole modérément altérée se situe entre la parole légèrement altérée et la parole sévèrement altérée ce qui explique la difficulté de son classement.

Le nouveau paramètre que nous proposons était désigné par l'analyse statistique comme le paramètre le plus pertinent. Effectivement, les taux de classification en l'utilisant séparément des autres paramètres montrent son efficacité à distinguer surtout la parole pathologique de la parole normale (81.8% pour reconnaitre les patients dysarthriques).

Nos expériences ont également montré que l'approche connexionniste proposée était efficace. En utilisant seulement les paramètres MFCC nous avons obtenu un taux de

classification correcte de 82.64 et avec la combinaison des paramètres de rythme, une augmentation de 3.71 % a été réalisée (86.35 %).

Le résultat principal est que les mesures de rythme sont sensibles aux différences entre les groupes de locuteurs dysarthriques. L'implication méthodologique de ce résultat est que, pour certains cas, le test Frenchay subjectif risque de mal classer certains sujets. Nous suggérons d'inclure les mesures du rythme dans la conception de l'évaluation objective de la dysarthrie.

Concernant l'évaluation objective de la parole aphasique, nous avons construit une base de données pathologique au niveau de deux établissements hospitaliers spécialisés : EHS d'Annaba (Seraidi) et de l'EHS de Sétif (Rass-Elma). Des ressources linguistiques basées sur le bilan diagnostique des patients aphasiques comprend plusieurs scripts : parole spontanée qui consiste en une entrevue du patient, des séries automatiques (compter de 1 à 10, récitation des mois de l'année …), répétition de syllabes, logatomes, mots simples et mots complexes, et la répétition de phrases plus ou moins compliquées etc…. Les enregistrements n'étaient pas vraiment faciles à entreprendre dû à plusieurs facteurs. En effet, en 2009, la date du début de la période de collecte des données aphasiques, il n'existait que trois EHS sur toute l'Algérie où nous pouvions trouver de l'orthophonie et où nous pouvions trouver des aphasiques. La prise en charge des aphasiques n'est pas totale ce qui veut dire que le patient peut sortir à n'importe quelle moment. Effectivement, il y avait des patients avec lesquels nous avons commencé les enregistrements mais qui étaient déclarés sortants à la prochaine séance d'enregistrement sans prendre en considération l'avis de l'orthophoniste. Le facteur le plus important est la fatigue du patient aphasique. En général, le patient aphasique se fatigue rapidement. Il faut également signaler que les conditions d'enregistrement n'étaient pas idéales car le local de consultation était situé dans un milieu hospitalier[42, 52].

Nous avons également proposé une normalisation phonétique pour la reconnaissance de la parole aphasique en utilisant la parole normale. En fait, la parole normale et la parole pathologique ainsi que les différents dialectes algériens ne partagent pas le même ensemble de phonèmes. Par conséquent, afin d'utiliser la parole normale pour la reconnaissance de la parole pathologique, nous avons proposé une normalisation de l'ensemble des phonèmes pour les différents dialectes des patients. Après la normalisation, nous avons appliqué les techniques d'adaptation des modèles acoustiques comme la régression linéaire du maximum de vraisemblance (MLLR) et Maximum A Posteriori (MAP). Les résultats de la

reconnaissance pathologique ont montré une réduction significative du WER par rapport aux autres systèmes de référence. Ainsi, une réduction du taux d'erreur de mots WER de presque 7% avec le système amélioré adapté par une combinaison de MLLR et MAP a été obtenue.

6.2. Perceptives

Il convient de souligner les limites de la base de données Nemours ainsi que la base l'aphasique ADAD en raison de son nombre relativement faible de participants. Les recherches à venir avec un échantillon plus large de locuteurs ayant des étiologies neurologiques dysarthriques ou aphasiques différentes seront nécessaires pour étudier les effets des paramètres de rythme. Aussi, il faut varier les types de scripts de parole en considérant par exemple les tâches de la parole spontanée, la lecture, la répétition ou le chant. En effet, nous pensons que le choix du type de tâche ou de script de parole est vraiment crucial avant de se prononcer sur le pouvoir discriminant des paramètres rythmiques.

Les travaux futurs devraient également doit se concentrer sur la nécessité d'automatiser les mesures rythmiques et de fusionner plus de paramètres acoustiques et plus de méthodes d'analyse et de classification. Diversifier les analyses acoustico-phonétiques et prosodiques permettrait de fournir de plus amples informations susceptibles d'aider au diagnostic et à l'évaluation pour une éventuelle intervention clinique ou à l'évaluation objective de l'état d'avancement de la rééducation. Un travail de groupe est très nécessaire dans le domaine de la parole en général et surtout la parole pathologique. Effectivement. Nous avons besoin de la contribution du neurologue, de l'orthophoniste, le linguiste, le psycholinguiste, le phonéticien...etc.

BIBLIOGRAPHIE

[1] O. Bälter, O. Engwall, A.-M. Öster *et al.*, "Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction." pp. 36-43.

[2] H. T. Bunnell, D. Yarrington, and J. B. Polikoff, "STAR: articulation training for young children." pp. 85-88.

[3] HTTP://WWW.VIDEOVOICE.COM/.

[4] J. J. Mahshie, "Feedback considerations for speech training systems." pp. 153-156.

[5] R. Nickerson, D. Kalikow, and K. Stevens, "Computer-aided speech training for the deaf," *Journal of Speech and Hearing Disorders,* vol. 41, no. 1, pp. 120-132, 1976.

[6] M. W. Richard D. Steele , Robert T. Wertz , Maria K. Kleczewska , Gloria S. Carlson, "Computer-based visual communication in aphasia.," *Neuropsychologia,* vol. Volume 27, no. 4, pp. 409-426, 1989.

[7] B. Shanon, "Classification and identification in an aphasic patient," *Brain and language,* vol. 5, no. 2, pp. 188-194, 1978.

[8] I. Tobolcea, "Logopaedic interventions for dyslalia's therapy," Spanda Press, Iasi, Romania, 2002.

[9] M. Weinrich, "Computerized visual communication as an alternative communication system and therapeutic tool," *Journal of neurolinguistics,* vol. 6, no. 2, pp. 159-176, 1991.

[10] C. Adnene, B. Lamia, and M. Mounir, "Analysis of pathological voices by speech processing." pp. 365-367.

[11] L. Salhi, T. Mourad, and A. Cherif, "Voice disorders identification using multilayer neural network," *pathology,* vol. 2, no. 7, pp. 8, 2008.

[12] M. Alsulaiman, "Speech recognition for medically disordered voice," in 24th International Conference on Computers and Their Applications in Industry and Engineering, Hawaii, 2011, pp. PP 233-7.

[13] M. Alsulaiman, "Voice Pathology Assessment Systems for Dysphonic Patients: Detection, Classification, and Speech Recognition," *IETE Journal of Research,* vol. 60, no. 2, pp. 156-167, 2014.

[14] M. Alsulaiman, G. Muhammad, and Z. Ali, "Classification of Vocal Fold Diseases Using RASTA-PLP." pp. 22-25.

[15] P. T. Hosseini, F. Almasganj, and M. R. Darabad, "Pathological voice classifcation using local discriminant basis and genetic algorithm." pp. 872-876.

[16] G. Muhammad, M. Alsulaiman, A. Mahmood *et al.*, "Automatic voice disorder classification using vowel formants." pp. 1-6.

[17] G. Muhammad, T. A. Mesallam, K. H. Malki *et al.*, "Formant analysis in dysphonic patients and automatic Arabic digit speech recognition," *Biomedical engineering online,* vol. 10, no. 1, pp. 41, 2011.

[18] S. Lotfi, and C. Adnène, "A Speech Processing Interface for Analysis of Pathological Voices."

[19] G. Muhammad, K. AlMalki, T. Mesallam *et al.*, "Automatic Arabic digit speech recognition and formant analysis for voicing disordered people." pp. 699-702.

[20] G. Muhammad, T. A. Mesallam, K. H. Malki *et al.*, "Multidirectional regression (MDR)-based features for automatic voice disorder detection," *Journal of Voice,* vol. 26, no. 6, pp. 817. e19-817. e27, 2012.

[21] M. H. AMAL, "PLACE DE LA RÉÉDUCATION ORTHOPHONIQUE DANS LA PRISE EN CHARGE ES DYSPHONIES D."

[22] S. de CORBIERE, É. Fresnel, and C. FRECHE, "La voix: la corde vocale et sa pathologie."

[23] G. Muhammad, Z. Ali, M. Alsulaiman *et al.*, “Vocal Fold Disorder Detection by applying LBP Operator on Dysphonic Speech Signal.”

[24] G. Muhammad, and M. Melhem, “Pathological voice detection and binary classification using MPEG-7 audio features,” *Biomedical Signal Processing and Control,* vol. 11, pp. 1-9, 2014.

[25] K. Elemetrics, "Kay elemetrics corp. disordered voice database," Model, 1994.

[26] A. S. M. Saudi, A. A. Youssif, and A. Z. Ghalwash, “Computer aided recognition of vocal folds disorders by means of rasta-plp,” *Computer and Information Science,* vol. 5, no. 2, pp. p39, 2012.

[27] H. Adam, “Voice onset time production in Palestinian Arabic-speaking aphasics–preliminary results.”

[28] M. Habiba, "Les Statistiques d’0rdre Supérieur: Application au Traitement du Signal Parole Pathologique (Patients à Audition Déficiente – Patients Trachéotomisés)," 2007.

[29] Z. Benselama, M. Guerti, and M. Bencherif, “Arabic speech pathology therapy computer aided system,” *Journal of Computer Science,* vol. 3, no. 9, pp. 685, 2007.

[30] A. Ghelis, M. Guerti, and C. Fredouille, "Statistical modeling for dysphonic classification." pp. 1-4.

[31] D. Messaoud, “Pathologie du langage oral chez un locuteur arabophone: cas de la substitution des phonèmes,” Université d'Alger2, 2012.

[32] H. Fodhil, “Les troubles d’Articulé Dentaire chez un locuteur Arabophone,” Université d'Alger2, 2012.

[33] K. Ferrat, and M. Guerti, “A study of sounds produced by Algerian esophageal speakers,” *African health sciences,* vol. 12, no. 4, pp. 452-458, 2013.

[34] M. Ahmed, “Application des Réseaux de Neurones Aux Troubles du Langage Oral de l’Arabe,” Ecole normale supérieure de Bouzaréah, 2013.

[35] F. AMARA, and M. FEZARI, “Voice Pathologies Classification Using GMM And SVM Classifiers,” *RECENT ADVANCES in BIOLOGY, MEDICAL PHYSICS, MEDICAL CHEMISTRY, BIOCHEMISTRY and BIOMEDICAL ENGINEERING*, pp. 65, 2013.

[36] M. FEZARI, and F. AMARA, “VOICE DISORDER DETECTION BASED ON AUTOMATIC SPEAKER IDENTIFICATION TECHNIQUES,” 2013.

[37] M. FEZARI, F. AMARA, and I. M. El-EMARY, “Acoustic Analysis for Detection of Voice Disorders Using Adaptive Features and Classifiers.”

[38] M. Pützer, and J. Koreman, “A German database of patterns of pathological vocal fold vibration,” *Phonus,* vol. 3, pp. 143-153, 1997.

[39] E. Grabe, “The IViE labelling guide,” *URL http://www. phon. ox. ac. uk/IViE/guide. html*, 2001.

[40] W. Grabe, “Reading research and its implications for reading assessment,” *Fairness and validation in language assessment*, pp. 226-262, 2000.

[41] F. Ramus, M. Nespor, and J. Mehler, “Correlates of linguistic rhythm in the speech signal,” *Cognition,* vol. 73, no. 3, pp. 265-292, 1999.

[42] H. Dahmani, S. Selouani, M. Chetouani *et al.*, "Prosody Modelling of Speech Aphasia: Case Study of Algerian Patients." pp. 1-6.

[43] H. Dahmani, S.-A. Selouani, D. O’shaughnessy *et al.*, “Assessment of dysarthric speech through rhythm metrics,” *Journal of King Saud University-Computer and Information Sciences,* vol. 25, no. 1, pp. 43-49, 2013.

[44] J. M. Liss, L. White, S. L. Mattys *et al.*, “Quantifying speech rhythm abnormalities in the dysarthrias,” *J Speech Lang Hear Res,* vol. 52, no. 5, pp. 1334-52, Oct, 2009.

[45] S.-A. Selouani, H. Dahmani, R. Amami *et al.*, "Dysarthric Speech Classification Using Hierarchical Multilayer Perceptrons and Posterior Rhythmic Features." pp. 437-444.

[46] S.-A. Selouani, H. Dahmani, R. Amami *et al.*, “Using speech rhythm knowledge to improve dysarthric speech recognition,” *International Journal of Speech Technology,* vol. 15, no. 1, pp. 57-64, 2012.

[47] R. D. K. a. G. W. Yunjung Kim, “An Acoustic Study of the relationships among neurologic Disease, Dysarthria Type, and Severity of Dysarthria,” *Journal of Speech & Hearing Disorders,* vol. 54:, no. 417-429., Apr 01, 2011.
[48] N. Zellal, *Orthophonie: essai de définition de l'orthophonie; une étude en aphasie*: Office des Publications universitaires, 1982.
[49] N. ZELLAL, "Introduction à la phonétique orthophonique arabe," *office des publications universitaire*, 1984.
[50] N. Zellal, *De la recherche en orthophonie: études de cas*: Office des publications universitaires, 1992.
[51] N. ZELLAL, “L'Orthophonie n'est pas la psychologie elle en est encore mois l'apanage,” 2013.
[52] S. A. S. H. DAHMANI , M. CHETOUANI and N.DOGHMANE , Paris, Juillet 2007., “Ressources linguistiques pour l’assistance aux aphasiques d’une région de l’est algérien.,” in Actes des ParoleVIIèmes RJC, Paris Juillet 2007, pp. p. 68-71.
[53] http://www.aphasia.org/.
[54] G. BERGOUNIOUX, “L'APHASIE DES APHASIQUES,” *La découverte et ses récits en sciences humaines: Champollion, Freud et les autres*, pp. 113, 1998.
[55] A. Trousseau, “De l'aphasie, maladie décrite récemment sous le nom impropre d'aphémie,” *Gazette des Hôpitaux,* vol. 37, pp. 13-14, 1864.
[56] A. Ombredane, “L'aphasie et l'elaboration de la pensee explicite,” 1951.
[57] P. P. Broca, “Broca on anthropology,” *Anthropological Review,* vol. 6, pp. 45-46, 1868.
[58] H. Goodglass, *Understanding aphasia*: Academic Press, 1993.
[59] C. Wernicke, *Der aphasische symptomenkomplex*: Springer, 1974.
[60] M. Louis, “Etude longitudinale de la dysprosodie d'un cas d'Aphasie Progressive Primaire: analyse des variables temporelles,” Aix Marseille 1, 2003.
[61] N. Z. D. YOUCEF, “CONTRIBUTION A LA RECHERCHE EN ORTHOPHONIE L'APHASIE EN MILIEU HOSPITALIER ALGERIEN ETUDE PSYCHOLOGIQUE ET LINGUISTIQUE,” 1986.
[62] J. Makhoul, B. Zawaydeh, F. Choi *et al.*, “‘2005 BBN/AUB DARPA Babylon Levantine Arabic Speech and Transcripts,” *Linguistic Data Consortium, Philadelphia*, 2005.
[63] A. Messaoudi, J. Gauvain, and L. Lamel, "Arabic broadcast news transcription using a one million word vocalized vocabulary." pp. I-I.
[64] M. Farahat, K. H. Malki, T. A. Mesallam *et al.*, “Development of the Arabic Version of Dysphagia Handicap Index (DHI),” *Dysphagia*, pp. 1-9, 2014.
[65] R. Hamdi, “La variation rythmique dans les dialectes arabes,” Lyon 2, 2007.
[66] A. Youssi, “The Moroccan triglossia: facts and implications,” *International journal of the sociology of language,* vol. 112, no. 1, pp. 29-44, 1995.
[67] S. Garmadi, “La situation linguistique actuelle en Tunisie: problèmes et perspectives,” *Revue tunisienne des sciences sociales,* vol. 13, pp. 13-24, 1968.
[68] N. Zellal, *Test orthophonique pour enfants en langue arabe: phonologie et parole*: Office des publications universitaires, 1991.
[69] F. Laroussi, *Plurilinguisme et identités au Maghreb*: Publication Univ Rouen Havre, 1997.
[70] http://www.phon.ucl.ac.uk/home/sampa/arabic.htm.
[71] S. Baloul, “Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé,” Le Mans, 2003.
[72] M. Djoudi, D. Fohr, and J. Haton, “SAPHA(Un système expert pour le décodage acoustico-phonétique de l'arabe standard),” 1989.
[73] M. Djoudi, J. Haton, and D. Fohr, “MARS(un système de reconnaissance de l'arabe moderne).”
[74] S. Ghazali, R. Hamdi, and M. Barkat, "Speech rhythm variation in Arabic dialects."
[75] N. Zellal, *Protocole neuropsycholinguistique du «MTA» - Version plurilingue algérienne* Université d'Alger, septembre 2002.

[76] Nacira ZELLAL, *Introduction à la phonétique orthophonique arabe*, Alger, Algérie: office des publications universitaire, 1984.

[77] B. Ducarne, and M. Barbeau, "Examen clinique et modes de rééducation des troubles visuels d'origine cérébrale," *Rev. Neurol,* vol. 137, pp. 11,693-707, 1981.

[78] P. Auzou, M. Gaillard, C. Özsancak *et al.*, "Evaluation clinique de la dysarthrie," *Isbergues, L'Ortho-Editions*, 1998.

[79] D. Duez, "Organisation temporelle de la parole et dysarthrie parkinsonienne," *Ozsancak C., Auzou P. Les troubles de la parole et de la déglutition dans la maladie de Parkinson. Solal, Marseille*, pp. 195-211, 2005.

[80] P. M. Enderby, *Frenchay dysarthria assessment*: Pro-ed Austin, TX, 1983.

[81] C. Özsancak, A. Parais, and P. Auzou, "Évaluation perceptive de la dysarthrie: présentation et validation d'une grille clinique: Etude préliminaire," *Revue neurologique,* vol. 158, no. 4, pp. 431-438, 2002.

[82] P. Auzou, C. Özsancak, S. Pinto *et al.*, *Les dysarthries*: Groupe de Boeck, 2007.

[83] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech, Language, and Hearing Research,* vol. 12, no. 2, pp. 246-269, 1969.

[84] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor speech disorders*: Saunders Philadelphia, 1975.

[85] J. M. P. X. M. Pedal, S. M. Peters, J. E. Leonzio, and H. T.Bunnell,, "the Nemours database of dysarthric speech." pp. pp. 1962 - 1965.

[86] J. P. Xavier Menédez-Pidal, shirley H.jenny E.leonzio, H.T.bunnell "Tne Nemours database of dysarthric speech." pp. pages 1962-1965.

[87] R. D. Kent, Weismer, G., Kent, J. F., & Rosenbek, J. C, "Toward phonetic intelligibility testing in dysarthria," *Journal of Speech & Hearing Disorders,* vol. 54, pp. 482-499, 1989.

[88] J. S. Garofolo, and L. D. Consortium, *TIMIT: acoustic-phonetic continuous speech corpus*: Linguistic Data Consortium, 1993.

[89] E. Grabe, " The IViE Labelling guide," *http://www.phon.ox.ac.uk/~esther/ivyweb/guide.html.*, 2002.

[90] E. Grabe, and . "Variation adds to prosodic typology," *In B. Bel and I. Marlin (eds)*, pp. 127-132, 2002.

[91] E. L. Grabe, E. L., "Durational variability in speech and the rhythm class hypothesis," *Papers in Laboratory Phonology,* vol. 7, 2002.

[92] F. Ramus, Nespor, M., & Mehler, J., "Correlates of linguistic rhythmin the speech signal.," *Cognition,* vol. 73, pp. 265-292., 1999.

[93] F. Ramus, "Language discrimination by newborns: teasing apart phonotactic, rhythmic and intonational cues," *Annual Review of Language Acquisition,* vol. 2, pp. 85-115, 2002.

[94] E. Grabe, "Variation adds to prosodic typology."

[95] E. L. Grabe, E. L " Durational variability in speech and the rhythm class hypothesis," *Papers in Laboratory Phonology,* vol. 7, 2002.

[96] F. Ramus, Nespor, M., & Mehler, J, "Correlates of linguistic rhythmin the speech signal," *Cognition,* vol. 73, pp. 265-292, 1999.

[97] F. Ramus, "Language discrimination by newborns: teasing apart phonotactic, rhythmic andintonational cues," *Annual Review of Language Acquisition,* vol. 2, pp. 85-115, 2002.

[98] R. L. Watrous, and L. Shastri, "Learning phonetic features using connectionist networks," *The Journal of the Acoustical Society of America,* vol. 81, no. S1, pp. S93-S94, 2005.

[99] P. Mertens, "L'intonation du français. De la description linguistique à la reconnaissance automatique," *status: published*, 1987.

[100] P. Tchébychev, *Sur l'interpolation par la méthode des moindres carrés*: Eggers et Comp., 1859.

[101] R. Andre-Obrecht, and N. Parlangeau, “Apport d'une segmentation statistique du signal de parole dans un système de décodage acoustico phonétique basé sur les connaissances,” *Le Journal de Physique IV,* vol. 4, no. C5, pp. C5-477-C5-480, 1994.
[102] J. Farinas, J.-L. Rouas, F. Pellegrino *et al.*, “2-Extraction automatique de paramètres prosodiques pour l'identification automatique des langues,” 2005.
[103] M. Chetouani. http://www.isir.upmc.fr/?op=view_profil&id=11.
[104] F. Ringeval, M. Chetouani, and J.-L. Zarader, “Extraction Automatique de la Prosodie des Enfants Autistes: premiers pas.”
[105] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot international,* vol. 5, no. 9/10, pp. 341-345, 2002.
[106] L. Pietrosemoli, and E. Mora, "Dysprosody in three patients with vascular cerebral damage." pp. 571-574.
[107] L. White, & Mattys, S. L., “Calibrating rhythm: First language and second language studies,” *Journal of Phonetics,* vol. 35, pp. 501-522, 2007a.
[108] V. Dellwo, " Rhythm and speech rate: Avariation coefficient for delta C." pp. pp. 231-241.
[109] J. M. Liss, L. White, S. L. Mattys *et al.*, “Quantifying Speech Rhythm Abnormalities in the Dysarthrias,” *J Speech Lang Hear Res,* vol. 52, no. 5, pp. 1334-1352, October 1, 2009, 2009.
[110] V. Dellwo, "Rhythm and speech rate: Avariation coefficient for delta C." pp. 231-241, Piliscsaba 2003.
[111] D. C. Howell, M. Rogier, V. Yzerbyt *et al.*, *Méthodes statistiques en sciences humaines*: De Boeck Université Brussels, 1998.
[112] J. Larson-Hall, *A Guide to Doing Statistics in Second Language Research Using SPSS*, 2010.
[113] G. Saporta, *Probabilités, analyse des données et statistique*: Editions Technip, 2011.
[114] E. Lhote, “Enseigner loral en interaction: Percevoir, écouter,” 1995.
[115] J. B. B. Polikoff, H.T, "The Nemours database of dysarthric speech: A perceptual analysis."
[116] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters *et al.*, "The Nemours database of dysarthric speech." pp. 1962-1965.
[117] S.-A. Selouani, M. S. Yakoub, and D. O'Shaughnessy, “Alternative speech communication system for persons with severe speech disorders,” *EURASIP Journal on Advances in Signal Processing,* vol. 2009, pp. 6, 2009.
[118] H. Tolba, and A. El Torgoman, "Towards the improvement of automatic recognition of dysarthric speech." pp. 277-281.
[119] F. Rudzicz, "Phonological features in discriminative classification of dysarthric speech." pp. 4605-4608.
[120] F. O. Tsuji T., Ichinobe H., Kaneko M. , “A log-linearized Gaussian mixture network and its application to EEG pattern classification,” in IEEE transactions on Systems Man,and Cybernetics, 1999, pp. pp.60-72.
[121] P. D. Polur, and G. E. Miller, “Investigation of an HMM/ANN hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals,” *Medical engineering & physics,* vol. 28, no. 8, pp. 741-748, 2006.
[122] O. Chapelle, V. Vapnik, O. Bousquet *et al.*, “Choosing multiple parameters for support vector machines,” *Machine learning,* vol. 46, no. 1-3, pp. 131-159, 2002.
[123] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data mining and knowledge discovery,* vol. 2, no. 2, pp. 121-167, 1998.
[124] V. Vapnik, *The nature of statistical learning theory*: springer, 2000.
[125] L. Bottou, *Large-scale kernel machines*: MIT Press, 2007.
[126] L. Bottou, and C.-J. Lin, “Support vector machine solvers,” *Large scale kernel machines*, pp. 301-320, 2007.
[127] J. Haton, “Modèles neuronaux et hybrides en reconnaissance de la parole: état des recherches,” *fondements et perspectives en traitement automatique de la parole, éditions H. Méloni*, pp. 139-154, 1995.

[128] F. Rosenblatt, “Principles of neurodynamics,” 1962.
[129] J. Saunders, "Real-time Discrimination of Broadcast Speech/Music." pp. pages 993-996.
[130] E. S. e. M. Slaney, "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. ." pp. pages 1331-1334.
[131] T. B. E. Wold, D. Keislar, et J. Wheater. , *Classification, Search and Retrieval of Audio.*, 1999.
[132] H. Dahmani, S. A. Selouani, N. Doghmane *et al.*, “On the relevance of using rhythmic metrics and SVM to assess dysarthric severity,” *International Journal of Biometrics,* vol. 6, no. 3, pp. 248-271, 2014.
[133] L. White, & Mattys, S. L, "Rhythmic typology and variation in first and second languages," *Segmental and prosodic issues in romance phonology: Current issues in linguistic theory series*, J. M. M.-J. S. E. P. Prieto, ed., pp. 237-257, Amsterdam: John Benjamins, 2007b.
[134] N. J. Nagelkerke, *Maximum likelihood estimation of functional relationships*: Springer, 1992.
[135] P. Schwarz, P. Matějka, and J. Černocký, "Hierarchical structures of neural networks for phoneme." pp. 325-328.
[136] S. J. Young, and S. Young, *The HTK hidden Markov model toolkit: Design and philosophy*: Citeseer, 1993.
[137] J. Baker, “The DRAGON system--An overview,” *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 23, no. 1, pp. 24-29, 1975.
[138] F. Jelinek, “Speech recognition by statistical methods,” *Proceedings of the IEEE,* vol. 64, pp. 532-556, 1976.
[139] C. Barras, “Reconnaissance de la parole continue: adaptation au locuteur et contrôle temporel dans les modèles de Markov cachés,” *Université de Paris VI, Paris, Thèse de Doctorat*, 1996.
[140] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, 1989.
[141] R. Bakis., “Continuous Speech Recognition via Centisecond Acoustic States,” in Dans 91st Meeting of the Acoustical Society of America, Washington, USA, avril 1976.
[142] S. J. Young, J. Odell, and P. C. Woodland, "Tree-based state tying for high accuracy acoustic modelling." pp. 307-312.
[143] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1-38, 1977.
[144] S. Young, G. Evermann, M. Gales *et al.*, “The HTK book (for HTK version 3.4),” *Cambridge university engineering department,* vol. 2, no. 2, pp. 2-3, 2006.
[145] C. Leggetter, and P. Woodland, "Flexible speaker adaptation using maximum likelihood linear regression." pp. 110-115.
[146] C. J. Leggetter, and P. Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models,” *Computer Speech & Language,* vol. 9, no. 2, pp. 171-185, 1995.
[147] J.-L. Gauvain, and C.-H. Lee, “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,” *Speech and audio processing, ieee transactions on,* vol. 2, no. 2, pp. 291-298, 1994.
[148] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals." p. 707.
[149] S. Young., “The HTK Hidden Markov Model Toolkit : Design and Philosophy. ,” *Rapport technique 152, Cambridge University Engineering Department, UK*, 1994.
[150] Calliope, *La parole et son traitement automatique*: Masson, 1989.
[151] X. Huang, A. Acero, H.-W. Hon *et al.*, *Spoken language processing: A guide to theory, algorithm, and system development*: Prentice Hall PTR, 2001.
[152] F. Jelinek, *Statistical methods for speech recognition*: MIT press, 1997.

[153] D. Jurafsky, “en JH Martin.(2009),” *Speech and Language Processing: An Introduction to natural*.

[154] L. R. Rabiner, and B.-H. Juang, *Fundamentals of speech recognition*: PTR Prentice Hall Englewood Cliffs, 1993.

[155] S. Young, G. Evermann, D. Kershaw *et al.*, *The HTK book*: Entropic Cambridge Research Laboratory Cambridge, 1997.

[156] V. Dellwo, & Wagner, P, " Relations between language rhythm and speech rate.." pp. 471-474.

[157] L. E. Ling, E. Grabe, and F. Nolan, “Q uantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English,” *Language and speech,* vol. 43, no. 4, pp. 377-401, 2000.

ANNEXE

1. Annexe A : le matériau linguistique dysarthrique

Le Matériau linguistique pour Set 1 : les locuteurs dysarthriques sont identifiés par un code à deux lettres (e.g., BB.wav, RK.wav, SC.wav, etc.) et les productions parallèles du locuteur adulte normale ont le code JP ajouté au début du nom de fichier (e.g., JPBB1.wav, JPBB2.wav, etc.)

BB1 (JPBB1): The bash is pairing the bath

BB2 (JPBB2): The bad is sleeping the bin

BB3 (JPBB3): The two is weeping the bit

BB4 (JPBB4): The kong is licking the chin

BB5 (JPBB5): The bat is shooing the chew

BB6 (JPBB6): The pat is wearing the sue

BB7 (JPBB7): The thin is surging the vat

BB8 (JPBB8): The pin is knowing the cop

RL1 (JPRL1): The con is bearing the butt

RL2 (JPRL2): The bait is waking the bet

RL3 (JPRL3): The rot is weeping the boat

RL4 (JPRL4): The back is heaping the yacht

RL5 (JPRL5): The com is licking the vat

RL6 (JPRL6): The lot is surfing the fate

RL7 (JPRL7): The phase is chewing the dive

RL8 (JPRL8): The faith is sitting the fade

BK1 (JPBK1): The sin is sitting the who

BK2 (JPBK2): The goo is surfing the batch

BK3 (JPBK3): The Bert is sinning the die

BK4 (JPBK4): The fife is waking the bad

BK5 (JPBK5): The watt is waning the dive

BK6 (JPBK6): The back is stewing the com

BK7 (JPBK7): The bin is pairing the tin

BK8 (JPBK8): The coo is singing the thin

BV1 (JPBV1): The fat is surging the bat

BV2 (JPBV2): The chew is waking the fine

BV3 (JPBV3): The fife is owing the cop

BV4 (JPBV4): The bit is sipping the mat

BV5 (JPBV5): The inn is shooing the din

BV6 (JPBV6): The two is knowing the moo

BV7 (JPBV7): The faith is mowing the Jew

BV8 (JPBV8): The thin is sinning the knew

FB1 (JPFB1): The yacht is lifting the bin

FB2 (JPFB2): The dial is suing the die

FB3 (JPFB3): The five is weeping the butt

FB4 (JPFB4): The bathe is going the thin

FB5 (JPFB5): The shoe is pairing the beet

FB6 (JPFB6): The boat is heaping the fat

FB7 (JPFB7): The dive is reaping the bat

FB8 (JPFB8): The bite is listing the bet

JF1 (JPJF1): The fay is sitting the Bert

JF2(JPJF2): The thin is shooing the dew

JF3 (JPJF3): The tin is bearing the bit

JF4 (JPJF4): The yacht is knowing the cop

JF5 (JPJF5): The badge is weighing the bat

JF6 (JPJF6): The coo is stewing the fake

JF7 (JPJF7): The com is living the base

JF8 (JPJF8): The beet is mowing the butt

KS1 (JPKS1): The fife is lifting the boat

KS2 (JPKS2): The bake is shooing the five

KS3 (JPKS3): The bathe is reaping the dial

KS4 (JPKS4): The watt is tearing the phase

KS5 (JPKS5): The con is licking the bash

KS6 (JPKS6): The shoe is suing the goo

KS7 (JPKS7): The back is stewing the thin

KS8 (JPKS8): The vat is living the Jew

LL1 (JPLL1) : The fate is stewing the sue

LL2 (JPLL2) : The gin is going the dew

LL3 (JPLL3) : The con is knowing the cob

LL4 (JPLL4) : The back is sitting the tin

LL5 (JPLL5) : The lot is surging the inn

LL6 (JPLL6) : The fife is licking the bin

LL7 (JPLL7) : The chin is daring the bathe

LL8 (JPLL8) : The rot is sweeping the dial

MH1 (JPMH1) : The cop is suing the badge

MH2 (JPMH2) : The fate is serving the phase

MH3 (JPMH3) : The bathe is shooing the rot

MH4 (JPMH4) : The goo is wearing the bet

MH5 (JPMH5) : The beet is lifting the lot

MH6 (JPMH6) : The vat is weeping the five

MH7 (JPMH7) : The bite is knowing the batch

MH8 (JPMH8) : The boat is owing the boot

RK1 (JPRK1): The dew is bearing the fight

RK2 (JPRK2): The bet is going the thin

RK3 (JPRK3): The shin is shooing the yacht

RK4 (JPRK4): The rot is wading the coo

RK5 (JPRK5): The fin is weighing the bait

RK6 (JPRK6): The dial is singing the zoo

RK7 (JPRK7): The chin is surging the fay

RK8 (JPRK8): The com is pairing the kong

SC1 (JPSC1): The bin is sleeping the con

SC2 (JPSC2): The fat is reaping the dime

SC3 (JPSC3): The badge is waking the bad

SC4 (JPSC4): The fin is shooing the face

SC5 (JPSC5): The sue is sipping the sin

SC6 (JPSC6): The batch is listing the lot

SC7 (JPSC7): The yacht is living the gin

SC8 (JPSC8): The zoo is daring the bash

2. Annexe B : La transcription phonétique de Nemours

La transcription phonétique de chaque mot utilisé pour construire les phrases de la base de données Nemours :

cob , kcl k aa bcl b	bath, bcl b ae th	beige, bcl b ey zh
cop, kcl k aa pcl p	bass, bcl b ae s	fade, f ey dcl d
com, kcl k aa m	bash, bcl b ae sh	fate, f ey tcl t
con, kcl k aa n	batch, bcl b ae tcl ch	fake, f ey kcl k
kong, kcl k aa ng	badge, bcl b ae dcl jh	faith, f ey th
bad, bcl b ae dcl d	bait, bcl b ey tcl t	face, f ey s
bat, bcl b ae tcl t	bake, bcl b ey kcl k	phase, f ey z
bag, bcl b ae gcl g	bathe, bcl b ey dh	fay, f ey
back, bcl b ae kcl k	base, bcl b ey s	fight, f ay tcl t

fine, f ay n

five, f ay v

fife, f ay f

dime, dcl d ay m

dive, dcl d ay v

dial, dcl d ay el

die, dcl d ay

dew, dcl d uw

two, tcl t uw

goo, gcl g uw

coo, kcl k uw

moo, m uw

knew, n uw

sue, s uw

zoo, z uw

shoe, sh uw

who, hh uw

chew, tcl ch uw

Jew, dcl jh uw

bin, bcl b ih n

pin, pcl p ih n

din, dcl d ih n

tin, tcl t ih n

fin, f ih n

thin, th ih n

sin, s ih n

shin, sh ih n

chin, tcl ch ih n

gin, dcl jh ih n

inn, ih n

rot, r aa tcl t

lot, l aa tcl t

watt, w aa tcl t

yacht, y aa tcl t

bat, bcl b ae tcl t

pat, pcl p ae tcl t

mat, m ae tcl t

vat, v ae tcl t

fat, f ae tcl t

bet, bcl b eh tcl t

beet, bcl b iy tcl t

bit, bcl b ih tcl t

boat, bcl b ow tcl t

butt, bcl b ah tcl t

boot, bcl b uw tcl t

Bert, bcl b er tcl t

bite, bcl b ay tcl t

wading, w ey dcl d ix ng

waiting, w ey tcl t ix ng

waking, w ey kcl k ix ng

waning, w ey n ix ng

waving, w ey v ix ng

weighing, w ey ix ng

reaping, r iy pcl p ix ng

leaping, l iy pcl p ix ng

weeping, w iy pcl p ix ng

heaping, hh iy pcl p ix ng

sweeping, s w iy pcl p ix ng

sleeping, s l iy pcl p ix ng

licking, l ih kcl k ix ng

living, l ih v ix ng

lifting, l ih f tcl t ix ng

listing, l ih s tcl t ix ng

bearing, bcl b eh r ix ng

pairing, pcl p eh r ix ng

daring, dcl d eh r ix ng

tearing, tcl t eh r ix ng

wearing, w eh r ix ng

suing, s uw ix ng	sinning, s ih n ix ng	surging, s er dcl jh ix ng
shooing, sh uw ix ng	singing, s ih ng ix ng	searching, s er tcl ch ix ng
chewing, tcl ch uw ix ng	going, gcl g ow ix ng	surfing, s er f ix ng
stewing, s tcl t uw ix ng	mowing, m uw v ix ng	serving, s er v ix ng
sipping, s ih pcl p ix ng	knowing, n ow ix ng	the, dh ax
sitting, s ih tcl t ix ng	owing, ow ix ng	is, ih z

3. Annexe C : Les résultats de la classification SVM /DFA pour les niveaux de sévérité

Tableau AC 1 : Résumé des résultats de la classification SVM pour les niveaux de sévérité sans la présence de la parole normale.

		Test fermé				Test ouvert							
						Ensemble d'apprentissage				L'ensemble de test			
		MP	MOP	SP	%	MP	MOP	SP	%	MP	MOP	SP	%
Set1	MP	32	0	0	100	20	0	0	100	12	0	0	100
	MOP	6	16	2	66.7	5	9	1	60.0	2	6	1	75.0
	SP	8	2	14	58.3	5	0	1	66.7	3	1	15	62.5
Set2	MP	284	7	5	95.9	184	8	4	93.9	86	5	9	86.0
	MOP	67	138	17	62.2	39	94	14	63.9	26	37	12	49.3
	SP	75	111	36	50.0	52	14	81	55.1	27	7	41	54.7
Set3	MP	273	12	8	93.2	179	6	9	92.3	84	7	8	84.8
	MOP	103	91	26	41.4	64	65	17	44.5	32	22	20	29.7
	SP	77	23	121	54.8	54	15	77	52.7	28	13	34	45.3

Tableau AC 2 : Résumé des résultats de la classification SVM pour les niveaux de sévérité Avec la présence de la parole normale.

		Test fermé					Test ouvert										Closed Test*					Open Test*									
							L'ensemble d'apprentissage					L'Ensemble de test										Learning Set *					Test Set*				
		MP	MOP	SP	HC	%	MP	MOP	SP	HC	%	MP	MOP	SP	HC	%	MP	MOP	SP	HC	%	MP	MOP	SP	HC	%	MP	MOP	SP	HC	%
Set1	MP	25	0	0	7	78.1	15	0	0	5	75.0	10	0	0	2	83.3	20	0	0	12	62.5	12	0	0	8	60.0	6	0	0	6	50.0
	MOP	6	55	3	0	62.5	5	8	2	0	53.8	2	5	2	0	55.6	10	2	10	2	8.3	7	2	5	1	13.3	3	0	5	1	00.0
	SP	7	1	23	1	71.9	4	0	15	11	75.0	3	0	10	0	83.3	5	3	21	3	63.6	3	2	13	2	65.0	1	1	8	2	66.7
	HC	2	0	3	83	94.3	0	0	0	55	100	0	0	3	30	90.9	6	0	3	79	89.7	4	0	0	51	92.7	2	0	3	28	84.8
Set3	MP	251	9	2	31	85.7	169	4	1	20	87.1	77	6	2	14	77.8	97	8	90	179	33.1	65	3	4	122	33.5	36	0	6	57	36.3
	MOP	54	94	26	46	42.7	42	54	15	35	37.0	27	19	16	12	25.7	20	20	11	102	13.1	17	1	56	72	0.7	11	1	25	37	1.3
	SP	44	21	196	34	88.7	31	12	131	21	67.2	16	9	60	15	60.0	40	11	161	83	54.6	24	1	153	57	57.9	23	1	50	26	50.0

* Results with ONLY the SRVC feature.

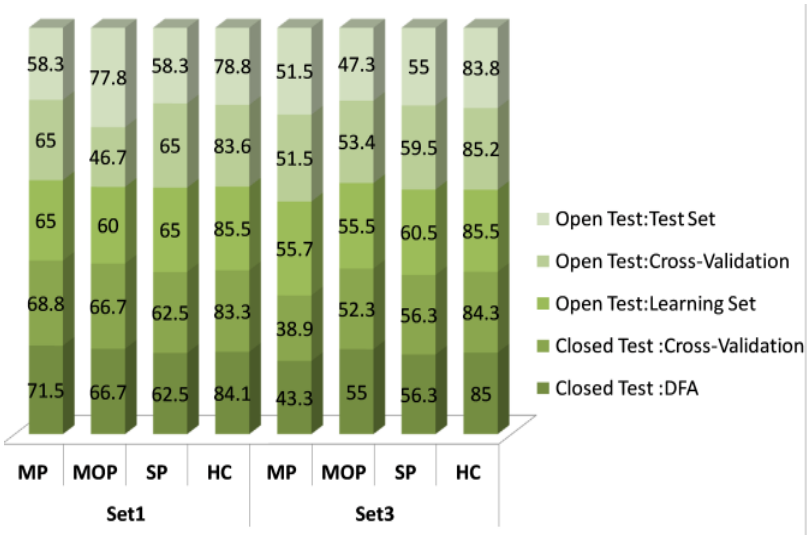

Figure AC 1: Résultats de la méthode DFA pour la classification des niveaux de sévérité avec la présence de la parole normale de HC par les mesures prédictives suivants: SRVC, rPVI-VC, %V, VarcoV and VarcoC pour Set 1 (toutes les paramètres à l'exclusion nPVI-VO, rPVI-VOUV and nPVI-VOUV pour Set 3).

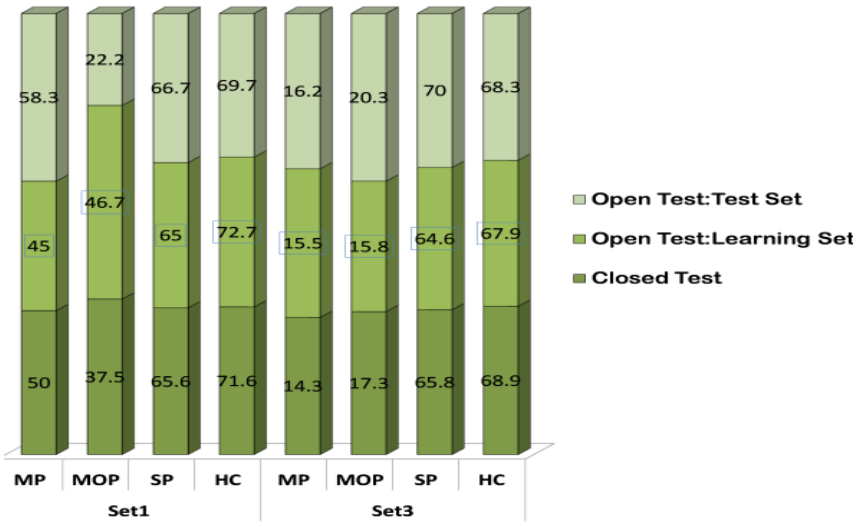

Figure AC 2 Résultats de la méthode DFA pour la classification des niveaux de sévérité avec la présence de la parole normale de HC par le seul paramètre SRVC pour Set 1 (SRVOUV, pour Set 3)

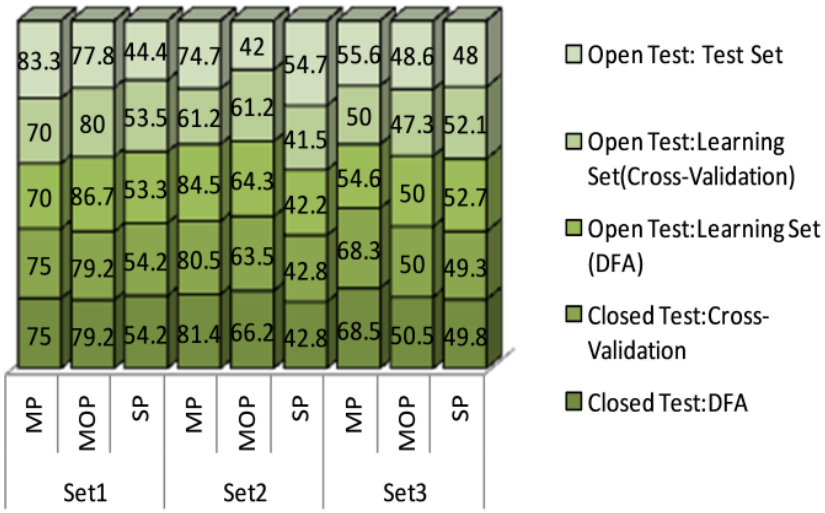

Figure AC 3 : Résultats de la méthode DFA pour la classification des niveaux de sévérité sans la présence de la parole normale de HC par les mesures prédictives suivants : SRVC, rPVI-VC and VarcoVC pour Set 1 (ΔC, ΔV, %V, rPVI-C, VarcoC, VarcoV and VarcoVC pour Set 2 et ΔVO, rPIV-VO, rPIV-UV, VarcoVO, VarcoUV et nPVI-V)

4. Annexe D : Corpus avec le système de transcription DE N.ZELLAL

4.1 Parole Spontanée

Wechesmeuk (quel est votre nom ?)

gedeeh fi ?'omrek (quel âge vous avez ?)

wiin khlaqti (où vous avez né ?)

wech takhdem (qu'est que vous faite comme travail ?)

µaxkilii 3la khadmetek (parlez-moi de votre travail)

µaħkilii 3la 3ayltek (parlez-moi de votre famille)

µaxkilii 3la mạrdak (parlez-moi de votre maladie)

4.2. Séries Automatiques

guulli ijeem esmana (citez-moi les jours de la semaine)

guulli chho:ra (citez-moi les mois de l'année)

µaħseblii hattal3achra (comptez jusqu'à dix ou vingt)

4.3. Répétition

4.3.1. Syllabes

Tableau AD 1 : Répétition des syllabes

1	ba	**7**	ab	**12**	du	**18**	ud	**24**	fé	**30**	éf	**36**	ṛé	**41**	é ṛ
2	bo	**8**	ob	**13**	ko	**19**	ok	**25**	fi	**31**	if	**37**	za	**42**	az
3	ti	**9**	it	**14**	ġa	**20**	ạg	**26**	su	**32**	us	**38**	xạ	**43**	ạx
4	lé	**10**	él	**15**	ṛạ	**21**	ạṛ	**27**	ṣu	**33**	uṣ	**39**	Rạ	**44**	ạR
5	3a	**11**	A3	**16**	cha	**22**	ach	**28**	qa	**34**	aq	**40**	hạ	**45**	ạh
6	µa			**17**	ya	**23**	ay	**29**	ħa	**35**	aħ	**46**	chu	**47**	uch

4.3.2. Logatomes

Tableau AD 2 : Répétition des logatomes

1	kro	**7**	fra	**13**	sky	**19**	xko	**25**	khli	**31**	ska	**37**	?leuf
2	sbi	**8**	bli	**14**	s_t_a_	**20**	baan	**26**	xro	**32**	?'fa	**38**	fha
3	dh_r_y	**9**	tru	**15**	kla	**21**	suun	**27**	Kwa	**33**	ghna	**39**	xna
4	blo	**10**	flu	**16**	bro	**22**	t_yyn	**28**	t_r_a_	**34**	?'t_a_	**40**	xfa
5	gro	**11**	xyy	**17**	fri	**23**	chlu	**29**	sla	**35**	ghsi	**41**	?'qa_
6	xfy	**12**	t_qa_	**18**	xma	**24**	ghr_a_	**30**	ghza	**36**	?'r_a_	**42**	?'da

4.3.3. Les mots

Les mots sont deux catégories: une simple et l'autre complexe. Les mots simple se divisent à des mots simples à consonnes et des mots simples à voyelles.

Tableau AD 3 : Répétition des mots contenants les consonnes.

1	gh	gha_a_r	**22**	ba_ghla	**42**	t_	t_één	**63**	xa_t_t_
2	l	liil	**23**	lliil	**43**	d_	d_a_a_r	**64**	?'ad_d_ ou ?'a_dh_dh_

3	f	fuur	**24**	ruuf	**44**	r	ra_a_x	**65**	xa_rr
4	s	siil	**25**	lbees	**45**	r	riix	**66**	dèèr
5	z	zuul	**26**	luuz	**46**	y	yuum	**67**	jeeyy
6	s_	s_ééf	**27**	kha_a_s_s_	**47**	k	kiif	**68**	fi:k
7	ch	chaaff	**28**	feuchch	**48**	g	gee?’	**69**	deugg
8	m	ma et malika	**29**	leem	**49**	kh	kheef	**70**	deekh
9	n	naar	**30**	been	**50**	q	qa_a_m	**71**	xma_q
10	b	beeb	**31**	beeba	**51**	?’	?’ eem	**72**	fee?’
11	t	tiliifon	**32**	meet	**52**	?	?enam		
12	d	diir	**33**	reed	**53**	h	heem	**73**	ddeeh
13	fl	fla	**34**	deufl	**54**	fr	fra_?’	**74**	dh_a_fr
14	bl	bleed	**35**	lbla	**55**	bgh	bgha_	**75**	teubgh
15	t_r_	t_ra	**36**	f_atr	**56**	d_r	d_ra		
16	kl	kla			**57**	gl	gleub		
17	kr_	kra_	**37**	fa_kr	**58**	ghr	ghrob	**76**	s_oghr
18	sb	sbar	**38**	xa_sb	**59**	sk	skeun	**77**	meusk
19	kt	kteeb	**39**	breukt	**60**	ght_	ght_a_	**78**	dboght
20	bt_	bt_a_	**40**	?’eubt	**61**	fn	fna	**79**	jeufn
21	sm	sma	**41**	?eusm	**62**	qr	qra_	**80**	xaqr

Mots simples : (81) ?eumseulkhiir (bonsoir)–(82) smaxly(pardonnez-moi) – (83)s_a_xxa_ (salut) –(84)beuslaama (au revoir).

Tableau AD 4: Répétition des mots contenants les voyelles

85	oo (o:) loomo	**89**	a_ qa_hwa	**93**	ø(eu) kheubz
86	Uu (u:) kuury	**90**	e:(ee) jeeja	**94**	ô(on) zonqa_

87	an(â) sandooq	**91**	? ?eulf	**95**	eu ?'euchch
88	é: (yy) t_ yyr_	**92**	i:(ii) biir	**96**	t_omoobiil

4.3.4. Mots complexes

Tableau AD 5 : Répétition des mots complexes.

1	ba_s_tyla	**3**	sukka_rtek	**5**	chbektha	**7**	chekchuukaxa_m_ra_	**9**	lb_aara_a_syjuun
2	ksebtha	**4**	ttiiliifuun	**6**	sbaaxelkhiir	**8**	?'em	**10**	salaamu?'alikom

4.3.5. Phrases

1. lli:l been (la nuit vient)
2. lbalon las_far b3i:d (le balon jaune est loin)
3. SSemes charqet wra jjbaal fel 3achwa (le soleil s'est levé derrière les montagnes le soir)
4. lmesfu:f elli jumma weklaatu lşabriina ?aw skhu:n waxluw (le masfouf (genre de couscous) qui a fait manger maman à sabrina est chaud et sucré)
5. Lebgar ta?` jaarii serxo felgaba legriba ?elli menewra haadiiddaar (les vaches de mon voisin sont à la proche foret derrière cette maison)
6. Chra fiil (il a achté un éléphant)
7. Wa?'laah babaah ghaab (pourquoi son père est-il absent)
8. jaami ja felwaqt (il n'est jamais à temps)
9. jaami lgaah felwaqt (il ne l'a jamais trouvé à temps)
10. ?'t_ak kulles_waared jiibhemehna (il t'a donné tout l'argent, ramener les ici)
11. matel?'bch eltem luukaan telgaw laxwaayjejellii tesxa_qqhem (ne jouez pas la bas si vous trouvez les choses dont vous avez besoin).
12. Lmaa baared (l'eau est froide)
13. nwaad_r twadro (les lunetes sont perdues
14. Lqahowa ta?' s_bah hada morah (ce café du matin est amer)
15. dar ?'ammi hmed lmabniya belberik lexmar melixa (la maison de mon oncle qui construite avec le brike est belle)
16. tran ta?' qsant_ina yuus_al nis_af see?'a qbal ma yu:s_al ta?' st_if (le train de Constantine arrive demi-heure avant celui de Sétif)

17. mohamed chchra: furmaj wo khobz wo rax lljazar yachri llxam (mohamed a acheté du fromage et pain et il est allé au boucher pour acheter de la viande)
18. mraa tiliifonii (une femme téléphone)
19. rajal ya?'om fel bxar (un homme nage dans la mer)
20. T_afela tajeri: (une petite fille coure)
21. Mraa taqeraa fel ketab (une femme lit dans le livre)
22. T'afela tazegii fel newaar (une petite fille arrose les fleurs)
23. jazaar yegat_a?' fel lexam (le boucher coupe la viande)
24. liineda taakol fettofaax (lynda mange la pomme)