République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal et Communications LSC

Thèse de Doctorat en Electronique

Présentée par :

M^r^ FERRAT Kamel

Magister en Electronique CRSTDLA Alger

Intitulée

Classification de la Parole Pathologique par Réseau de Neurones Artificiels

Soutenue Publiquement le 30/01/2014 devant le jury composé de :

Présidente :	HAMAMI Latifa	Professeur	ENP Alger
Rapporteur :	GUERTI Mhania	Professeur	ENP Alger
Examinateurs :	TIGZIRI Noura	Professeur	UMMTO Tizi-Ouzou
	BOUDRAA Bachir	Professeur	USTHB Bab Ezzouar Alger
	SAYOUD Halim	Professeur	USTHB Bab Ezzouar Alger
Invitée :	BENDRIS Soumaya	Orthophoniste CHU Bab El Oued Alger	

ENP 2014

DEDICACES

A la mémoire de mon père Moussa Ath Ferhat qui m'a appris les bons principes de la vie ;

A ma mère Fatima N Lhadj Lounis Ath Ferhat qui m'a toujours entouré de son affection. Que Dieu lui accorde sa sainte miséricorde, santé et longue vie, afin que je puisse la combler à mon tour ;

A mes chers frères et sœurs, ainsi qu'à toutes leurs familles ;

A ma femme pour son amour et son infaillible soutien ;

A mes trois enfants Manel, Lina et Moussa.

A vous tous, je suis fier de vous avoir

REMERCIEMENTS

Je souhaite de prime abord exprimer mon entière reconnaissance à ma Directrice de thèse GUERTI Mhania, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, pour son aide, ses précieux conseils, son apport méthodologique inestimable et pour toutes les corrections et commentaires dont elle m'a fait part dans la rédaction de mon travail de recherche. C'est grâce à ses orientations, ses encouragements et sa confiance que j'ai pu mener à terme ce travail. Qu'elle trouve ici ma profonde reconnaissance pour tout ce qu'elle a fait pour moi !

Comme je tiens vivement à remercier Mme HAMAMI Latifa, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, pour avoir accepté de présider mon jury de thèse. Je la remercie également pour sa disponibilité, sa gentillesse et ses encouragements lors de chacun de mes passages à l'ENP.

Mes vifs remerciements vont également à Mme TIGZIRI Noura, Professeur à l'Université Mouloud Mammeri de Tizi-Ouzou, Mr BOUDRAA Bachir et Mr SAYOUD Halim, Professeurs à l'Université des Sciences et Technologies Houari Boumediène Alger, qui me font l'honneur d'être membres du jury de ma thèse. Je tiens à leur exprimer toute ma reconnaissance pour avoir accepté de lire et d'évaluer ce travail.

Je remercie Mme BENDRIS Soumaya, Orthophoniste Praticienne au Centre Hospitalo-Universitaire de Bab El Oued Alger, pour avoir accepté l'invitation d'assister à ma soutenance.

Mes remerciements vont aussi pour le Directeur du CRSTDLA, le Professeur Benmalek Rachid, pour son soutien constant et sa disponibilité à tout moment, ainsi qu'à tous mes collègues chercheurs et administrateurs au CRSTDLA, Université d'Alger II.

Je ne peux oublier d'adresser de tout mon cœur toute ma reconnaissance aux patients et aux sujets témoins pour leur participation aux enregistrements.

Enfin, je voudrais exprimer toute ma gratitude à tous ceux qui d'une manière ou d'une autre m'ont apporté leur aide pour la réalisation de ce modeste travail.

الملخص

يتضمن عملنا الاستكشاف الآلي لأمراض الكلام بواسطة شبكات العصبونات الاصطناعية . الهدف هو تشخيص مختلف اضطرابات الكلام لاستغلالها في إعادة التأهيل الأرطفوني في الوسط الإستشفائي الجزائري. للقيام بذلك، تم تسجيل مدونة تتكون من 240 ملف صوتي و تحتوي على صوائت ممددة، وكلمات معزولة وجمل باللغة العربية منطوقة من طرف أشخاص عاديين، مرضى الباركينسون (اضطراب عصبي) ومرضى استئصال الحنجرة (اضطراب فيزيولوجي). في البداية، تم إجراء تحليل أكوستيكي لاستخراج أهم الخصائص الملائمة (التردد الأساسي، الجيتر، الشيمر، درجة تأثير الضجيج على نغمات الصوت "HNR"، نسبة مرور الموجة الصوتية على مستوى الصفر"TPZ"، درجة المقاطع الغير مجهورة في الصوت "DUF"، الشدة، ...). بعد ذلك، تم تطبيق الشبكة العصبية من النوع "TDNN" لغرض تمييز كل من هاتين الحالتين المضطربة مقارنة مع الكلام العادي.

حسب النتائج المحصل عليها، فإن الطريقة المقترحة تعطي نسبة عالية للتعرف الآلي لاضطرابات الكلام (95.00 % من التعرف للكلام الباركينسوني و 97.50 % للكلام المريئي عندما تأخذ على حدة و92,50 % للكلام الباركينسوني و 80.00 % للكلام المريئي عند مزج هذه الأصوات مع الكلام العادي).

يساهم بحثنا هذا في قيادة التشخيص التلقائي، إنشاء النظم الخبيرة التي تسمح تحديد اضطرابات الكلام (عسر الكلام، الرُتّة،...) بمستويات ملموسة و كذا المساعدة في التعليم الأرطفوني.

كلمات المفاتيح : شبكات العصبونات الاصطناعية، TDNN، التحليل الأكوستيكي ، اضطرابات الكلام.

Résumé

Notre travail concerne la classification automatique de la **P**arole **Path**ologique (**PP_{ath}**) par **R**éseaux de **N**eurones **A**rtificiels (**RNA**). Le but ciblé est la caractérisation de cette dernière, en vue de son exploitation en réhabilitation dans un milieu hospitalier algérien.

Nous avons élaboré un corpus de 240 fichiers sonores comprenant des voyelles soutenues, mots isolés et phrases en Arabe. Ce corpus a été enregistré par des locuteurs normaux et des patients aussi bien parkinsoniens (Pathologie Neurologique) que laryngectomisés (Pathologie Physiologique). Au préalable, nous avons effectué une analyse acoustique afin d'extraire les traits pertinents (fréquence fondamentale "F_0", Jitter, Shimmer, HNR, TPZ, DUF, Energie, …). Ensuite, un Réseau de Neurones à Décalages Temporels "**TDNN**" (**T**ime **D**elay **N**eural **N**etwork) a été appliqué afin de discriminer les deux types de PP_{ath} par rapport à la **P**arole **Norm**ale (**PN_{orm}**).

La méthode choisie nous a permis d'avoir des TR (%) appréciables des PP_{ath} par rapport à la PN_{orm}, lorsque ces dernières sont prises dans un contexte isolé (95.00 % pour les **P**aroles **Park**insoniennes (**PP_{ark}**) et 97.50 % pour les **P**aroles **Œso**phagiennes (**$PŒ_{so}$**)) et des taux respectifs de 92.50 % et 80.00 % pour les deux PP_{ath} lorsqu'elles sont mélangées.

Notre travail pourra contribuer à la conduite de diagnostics automatiques, l'établissement de systèmes experts aboutissant à des taux appréciables d'identification des anomalies vocales (dysphonies, dysarthries, …) et l'aide à l'enseignement en Orthophonie.

Mots clés : Réseaux de Neurones Artificiels, TDNN, Analyse Acoustique, Parole Pathologique.

Abstract

Our work concerns the automatic classification of the Pathological Speeches (PS) by Artificial Neural Networks (ANNs). The focused objective is the characterization of the PS for their exploitation in rehabilitation at Algerian hospitals.

To do this, a corpus of 240 sound speech files containing sustained vowels, Arabic isolated words and sentences was recorded by normal speakers, parkinsonian patients (Neurological pathology) and laryngectomized patients (physiological pathology). In the beginning, an acoustical analysis was performed to extract the relevant features (Fundamental frequency F_0, Jitter, Shimmer, HNR, ZCR, DUF, Energy ...). In the end, a Time Delay Neural Network "TDNN" was applied in order to discriminate both these PS compared to the Normal Speech (NS).

The chosen method allowed us to have significant TR (%) of PS compared to NS, when they are taken in an isolated context (95.00 % for the parkinsonian speech and 97.50 % for the esophageal speech), and respective rates of 92.50 % and 80.00 % for both PS when they are mixed.

This study will contribute to the conducting of automatic clinical diagnoses, the development of expert systems allowing appreciable levels of identifying of speech disorders (dysphonia, dysarthria, …), and to teaching aid in Speech Therapy.

Keywords : Artificial Neural Networks, TDNN, Acoustic Analysis, Pathological Speech.

Table des matières

CHAPITRE 3 : ANALYSE ACOUSTIQUE DE LA PAROLE PATHOLOGIQUE

CHAPITRE 4 : RESEAUX DE NEURONES ARTIFICIELS

CHAPITRE 5 : APPLICATION DES RNA A LA CLASSIFICATION DES PAROLES PATHOLOGIQUES

Liste des Abréviations

API	:	**A**lphabet **P**honétique **I**nternational
AS	:	**A**rabe **S**tandard
CV	:	**C**onsonne-**V**oyelle
DCT	:	**D**iscrete **C**osine **T**ransform
DTW	:	**D**ynamic **T**ime **W**arping
DUF	:	**D**egree of **U**nvoiced **F**rames
E	:	**E**nergie
FFT	:	**F**ast **F**ourier **T**ransform
FIR	:	**F**inite **I**mpulse **R**esponse
GOI	:	**G**lottal **O**pening **I**nstants (Instants d'Ouverture Glottale)
HMM	:	**H**idden **M**arkov **M**odels
HNR	:	**H**armonics to **N**oise **R**atio
IFFT	:	**I**nverse **F**ast **F**ourier **T**ransform
IOG	:	**I**nstants d'**O**uverture **G**lottale
J	:	**J**itter
$\mathbf{J_f}$	:	**J**itter **f**actor
LPC	:	**L**inear **P**redictive **C**oding
LM	:	**L**evenberg-**M**arquardt
MDVP	:	**M**ulti **D**imensional **V**oice **P**rogram
MFCC	:	**M**el **F**requency **C**epstral **C**oefficients
MLP	:	**M**ulti **L**ayer **P**erceptron (Perceptron Multi Couches)
MSE	:	**M**ean **S**quared **E**rror
MSW	:	**M**ean **S**quared **W**eights
NPC	:	**N**on **P**ris en **C**harge
NVP	:	**N**éo **V**ibrateur **P**haryngo-œsophagien
OCR	:	**O**ptical **C**haracter **R**ecognition
PC	:	**P**ris en **C**harge
PMC	:	**P**erceptron **M**ulti **C**ouches
$\mathbf{PN_{orm}}$	:	**P**arole **Norm**ale
$\mathbf{P\text{Œ}_{so}}$	:	**P**arole **Œso**phagienne
$\mathbf{PP_{ark}}$	:	**P**arole **Park**insonienne
$\mathbf{PP_{ath}}$	:	**P**arole **Path**ologique
RAL	:	**R**econnaissance **A**utomatique du **L**ocuteur
RAP	:	**R**econnaissance **A**utomatique de la **P**arole

RAV : **R**econnaissance **A**utomatique du **V**isage
RB : **R**égularisation **B**ayésienne
RNA : **R**éseaux de **N**eurones **A**rtificiels

S : **S**himmer
S_f : **S**himmer **f**actor
SVM : **S**upport **V**ector **M**achines

TAP : **T**raitement **A**utomatique de la **P**arole
TDNN : **T**ime **D**elay **N**eural **N**etworks
TPZ : **T**aux de **P**assage par **Z**éro
TR : **T**aux de **R**econnaissance

VC : **V**oyelle-**C**onsonne
VCV : **V**oyelle-**C**onsonne-**V**oyelle
VE : **V**ecteurs d'**E**ntrées

Liste des Figures

Liste des Tableaux

INTRODUCTION GENERALE

Le désir de communiquer oralement avec la machine faisait partie des rêves utopiques qui ont de tout temps hanté l'esprit humain. Ce désir est resté pendant longtemps à un état primitif, la science n'avait pas encore connu les progrès qui lui permettent de mieux analyser la parole et ainsi de mieux connaître ses caractéristiques pour pouvoir la manipuler. Cette grande aventure liant l'Homme à la Machine connaîtra, pendant deux siècles, une histoire passionnante qui a engendré beaucoup de travaux de recherche dans le domaine. Le développement connu par la théorie numérique du signal, des techniques de reconnaissance et de synthèse de la parole, de l'informatique et des sciences du langage en général, ont apporté leur contribution.

L'histoire de cette recherche nous a montré que le **T**raitement **A**utomatique de la **P**arole (**TAP**) fait appel à plusieurs intervenants : phonéticiens, linguistes, orthophonistes, médecins, technologues, etc. En d'autres termes, le TAP englobe plusieurs disciplines à la fois, telles que la Phonétique, la Linguistique, l'Electronique (Acoustique, Traitement du Signal), l'Informatique, la Psychologie (Neurosciences et Orthophonie), l'Intelligence Artificielle, les Mathématiques, etc. Son avantage réside principalement dans le fait qu'il reste une activité de recherche scientifique et technologique pluridisciplinaire, d'où son large impact dans la vie quotidienne de l'individu.

Pour analyser la parole, plusieurs méthodes ont vu le jour, s'intéressant à sa modélisation pour pouvoir la manipuler automatiquement. Dans leur sillage, une partie des recherches se sont intéressées à la modélisation de la **P**arole **Path**ologique (**PP_{ath}**). Ces recherches en PP_{ath} ont pour objectif essentiel la conduite de diagnostics automatiques permettant de caractériser de façon fiable les anomalies vocales. Ceci est assez complexe car une bonne rééducation des patients atteints de troubles de la voix et de la parole nécessite une connaissance précise de l'origine de ces troubles pour pouvoir y remédier en adaptant une technique de rééducation appropriée. A cet effet, l'exploration des voix et paroles pathologiques est un objectif de recherche clinique d'une importance particulière, car elle permet la mise en place de procédures et techniques d'évaluation physico-acoustiques des caractéristiques de la voix et de la parole afin de déterminer de façon objective leur écart par rapport aux valeurs normales. Il reste qu'en Algérie, peu de travaux ont été réalisés dans ce domaine. En effet, le milieu hospitalier algérien est confronté à un manque flagrant d'aide de la part du potentiel scientifique exerçant à l'université ou dans les centres de recherche. Ceci est d'autant plus important, car par une analyse acoustico-articulatoire des pathologies vocales, nous obtenons des données concrètes qui nous

permettent de caractériser objectivement la voix et la parole et d'estimer le degré d'une éventuelle détérioration par rapport à la norme et apporter ainsi les solutions nécessaires pour y remédier. De plus, la visualisation graphique de leurs prononciations stimule la curiosité des patients et augmente très souvent leur motivation, car ils perçoivent mieux la finalité de leur prise en charge et surtout l'évolution progressive de leur rééducation.

Dans le cadre de notre travail, nous nous sommes intéressés à la classification automatique de la **PP_{ath}** par rapport à la **P**arole **Norm**ale (**PN_{orm}**). Nous avons effectué, dans une première étape, une analyse acoustique portant sur deux cas de PP_{ath} : Un premier cas d'origine neurologique dont nous avons choisi la **P**arole **Park**insonienne (**PP_{ark}**) et un second cas d'origine physiologique et nous avons étudié la **P**arole **Œso**phagienne (**$PŒ_{so}$**). Dans une seconde étape, nous avons appliqué les **R**éseaux de **N**eurones **A**rtificiels (**RNA**) pour classifier les deux cas de PP_{ath} par rapport à la PN_{orm}. Pour cela, nous avons exploité le réseau de neurones dit à décalages temporels **TDNN** (**T**ime **D**elay **N**eural **N**etwork). L'avantage de ce type de réseau est qu'il permet de bien classifier les sons, en tenant compte de l'aspect dynamique de la parole et par conséquent, des phénomènes de la coarticulation (influence d'un son sur un autre contigu), connus comme assez pertinents lors d'un acte de parole.

Cette thèse est organisée en cinq chapitres :

- le premier expose quelques généralités sur les caractéristiques physico-acoustiques de la parole. Nous avons donné un aperçu sur les modes et lieux d'articulation et la production de la parole, ainsi que sur les paramètres acoustiques représentatifs du signal de parole. Des notions fondamentales sur l'Arabe Standard ont été également données ;
- le deuxième est consacré à la description de diverses pathologies de la parole. Nous avons donné une attention particulière à deux cas pathologiques PP_{ark} et $PŒ_{so}$, sur lesquels nous avons fait une analyse acoustique et une classification automatique par rapport à la PN_{orm} ;
- le troisième porte sur l'analyse acoustique des PP_{ark} et $PŒ_{so}$, en vue de leur caractérisation. Pour ce faire, nous avons exploité des paramètres spécifiques tels que le degré de perturbation de la fréquence fondamentale (Jitter), le degré de perturbation de l'intensité (Shimmer), l'influence du bruit sur les harmoniques du signal (**H**armonics to **N**oise **R**atio **HNR**), le nombre de trames non voisées (**D**egree of **U**nvoiced **F**rames **DUF**), le **T**aux de **P**assage par **Z**éro (**TPZ**) et l'énergie ;

- le quatrième est consacré aux généralités sur les RNA. Nous avons donné, en particulier, une importance aux réseaux à décalages temporels TDNN que nous avons appliqués dans le cadre de notre travail.
- le dernier chapitre concerne la classification automatique de la PP_{ath} par rapport à la PN_{orm} par RNA. Nous avons expliqué d'une manière détaillée l'architecture du système de classification que nous avons élaboré. Ensuite, une application à été réalisée sur les deux cas pathologiques choisis (PP_{ark} et $PŒ_{so}$). Les résultats obtenus ont été également exposés et commentés.

CHAPITRE 1

CARACTERISTIQUES PHYSICO-ACOUSTIQUES DE LA PAROLE

1.1. Introduction

La recherche en Traitement Automatique de la Parole (TAP) doit nécessairement passer par l'étude de la composante phonétique de la langue. Cette étude nous permettra de dégager les principales caractéristiques relatives à la production des différents sons et ainsi de cerner l'ensemble des paramètres physico-acoustiques représentatifs du signal de parole, en vue de les exploiter dans l'élaboration d'un système de reconnaissance ou de synthèse. Ces paramètres peuvent être exploités également pour la caractérisation des différentes paroles pathologiques par rapport à la normale. Connaître d'une manière précise la façon dont est produite la parole permettra de mieux manipuler et traiter celle-ci, afin de parvenir à discriminer efficacement une pathologie vocale, quelle que soit sa nature, par rapport à la normale.

Dans ce chapitre, nous avons exposé, d'une manière succincte, la théorie de la production de la parole. Nous avons ensuite montré les paramètres physico-acoustiques les plus importants en TAP. A la fin, nous avons décris les sons de l'Arabe Standard et les phénomènes spécifiques à cette langue.

1.2. Production de la parole

Par parole, nous entendons la catégorie de sons prononcés par l'être humain, soit le résultat de production de l'appareil phonatoire humain. Ce dernier est constitué d'un excitateur (Poumons + Cordes Vocales) qui est la source ou l'origine du son, et d'un ensemble de résonateurs (Cavités Vocales) qui est le volume dans lequel se propage l'excitation et qui modifient le signal source (Figure 1.1). Ces cavités permettent d'avoir les lieux d'articulation des sons ainsi que les modes d'articulations (occlusif, fricatif, etc.).

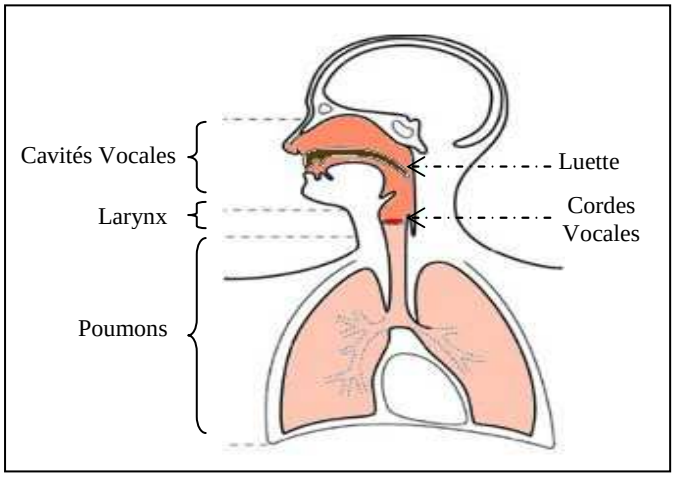


Figure 1.1: Schéma simplifié de l'appareil phonatoire humain

La production d'un acte de parole est déterminée par trois phases essentielles :

- la génération de l'air, qui met en jeu les poumons, le diaphragme et les différents muscles du thorax ;
- la vibration de cet air par les cordes vocales, à l'intérieur du larynx. Si ces dernières ne vibrent pas, nous entendons un son "non voisé" ou "sourd" comme [t]. Si elles se rapprochent et vibrent, nous obtenons un son "voisé", exemple [d] ;
- la résonance de cette vibration dans les cavités vocales (pharyngale, buccale, labiale et nasale), dont les configurations nous donnent le timbre de la voix. Ces différentes cavités jouent le rôle de résonateurs acoustiques. La majorité des voix se ressembleraient si le son produit est seulement laryngé (provenant uniquement de la vibration des cordes vocales). Or, ce sont les modifications de formes et de dimensions que subissent les cavités vocales pendant l'émission de la voix, qui donnent à celle-ci un timbre qui est particulier à chacun d'entre nous. C'est également dans ces cavités que prennent forme les différentes consonnes et voyelles de la langue.

1.2.1. Larynx

Le larynx est situé à la partie antérieure et médiane du cou, au-dessus de la trachée artère et en avant de l'œsophage. C'est une portion de tube qui prolonge la trachée vers l'arrière-bouche. Les cartilages du larynx s'articulent entre eux grâce à des muscles, ce qui assure sa fermeture et son ouverture permettant ainsi les deux fonctions de la phonation et de la respiration. Le larynx comporte deux plis vocaux bordant la glotte, appelés "Cordes vocales" (Figure 1.2). Ces dernières sont considérées comme un organe important du système phonatoire.

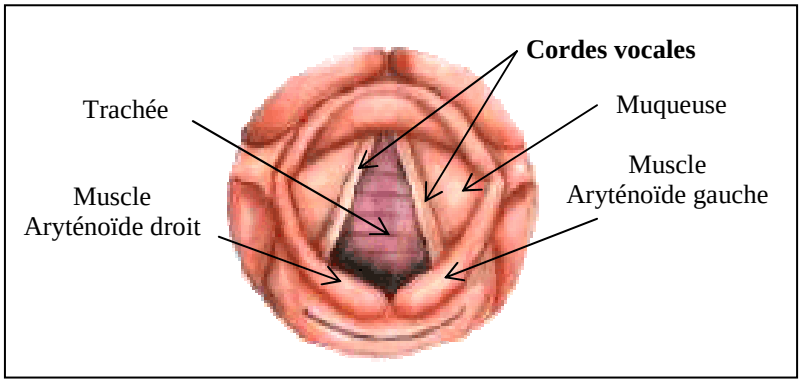


Figure 1.2 : Schéma du larynx

En vibrant, les cordes vocales permettent de produire tous les sons voisés notamment les voyelles. La quantité d'air qui remonte des poumons avec une pression et un débit

spécifiques, se faufile entre ces deux cordes et met leur bord interne en vibration, produisant ainsi le son vocal dit "laryngé".

La structure particulière de ces cordes, zone d'affrontement continu entre deux muscles soumis à des contraintes mécaniques importantes, explique le pourquoi de la fragilité de cette partie importante de l'appareil phonatoire. Cette zone est assez fréquemment le siège de diverses pathologies apparaissant précocement (entre 35 à 40 ans), sur des terrains prédisposants tels que "l'exploitation abusive de la voix" responsable de dysphonies (enseignants, chanteurs, ….) et l'irritation chronique (alcoolisme, tabagisme, …) pouvant s'aggraver et évoluer vers un cancer.

1.2.2. Cavités vocales

A la sortie du larynx, le son produit n'est pas encore une parole, proprement dite. Pour devenir parole, il doit être modelé en voyelles et consonnes. Ceci est réalisable lorsqu'il traverse les cavités vocales ou supraglottiques de l'appareil phonatoire, qui prennent des configurations spécifiques selon les mouvements de notre mâchoire inférieure, des lèvres et de la langue (Figure 1.3). Entre la cavité pharyngale et la cavité nasale se trouve une portion de muscle, qu'on appelle luette, velum ou uvule. Cette dernière s'ouvre pour permettre à une quantité d'air de s'échapper par le nez, lors de la prononciation des sons nasals. Dans la cavité buccale, se trouvent la langue, les dents supérieures et inférieures, les alvéoles, le palais dur derrière les alvéoles, et le palais mou proche du velum. Ces régions appelées lieux d'articulation permettent de discriminer les différentes consonnes de la langue.

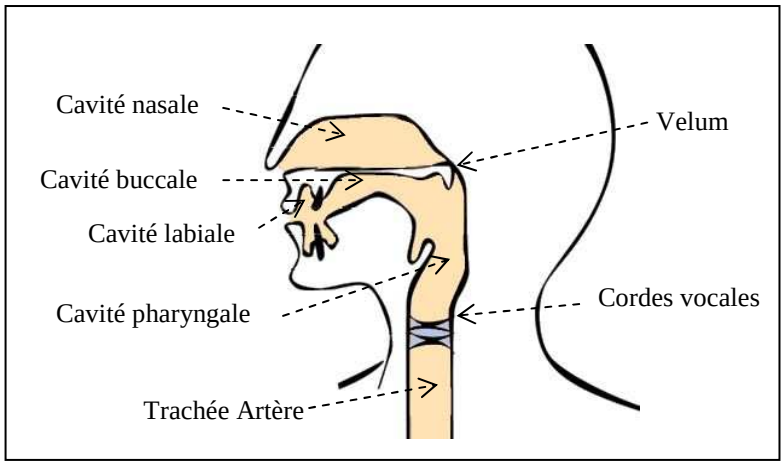


Figure 1.3 : Cavités Vocales de l'appareil phonatoire

1.3. Caractéristiques acoustiques de la parole

La caractérisation acoustique de la parole est un domaine très important de la phonétique. Sa fonction essentielle est de trouver puis d'extraire les paramètres acoustiques qui reflètent le mieux les différents processus physiologiques ou articulatoires qui rentrent dans la production d'un son quelconque de parole (Figure 1.4).

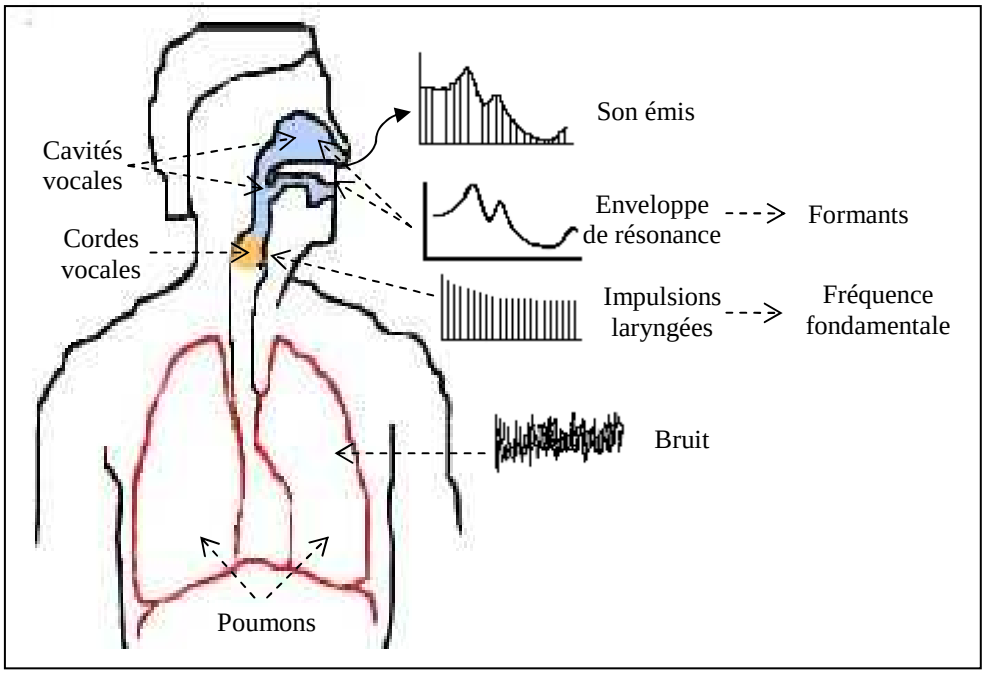


Figure 1.4 : Différents processus physico-acoustiques de production d'un acte de parole

Les paramètres acoustiques permettent de traiter, mesurer et analyser les différents phénomènes physiologiques entrant dans la production d'un acte de parole, depuis l'émission d'air par les poumons jusqu'à la sortie du son émis à l'extérieur de l'appareil phonatoire. Parmi les paramètres acoustiques les plus importants, nous pouvons citer :

1.3.1. Fréquence fondamentale

La fréquence fondamentale, notée F_0, représente le nombre de vibrations par seconde des cordes vocales. Elle vient du fait que lorsque nous prononçons certains sons tels que [b], [d] ou [z], nous faisons vibrer les cordes vocales à une certaine fréquence quasi périodique. Ce paramètre acoustique, appelé également pitch, peut varier généralement de :

- 80 à 200 Hz pour une voix masculine ;
- 150 à 450 Hz pour une voix féminine ;

- 350 à 600 Hz pour une voix d'enfant.

Les valeurs multiples de F_0 sont appelées : les harmoniques. Nous verrons plus loin qu'une richesse en harmoniques nous permet d'obtenir un timbre de voix très claire et qu'une pauvreté en harmoniques nous fait percevoir un timbre sombre d'une voix pathologique.

1.3.2. Formants et Transitions Formantiques

Lors du passage de l'air à travers les cavités vocales, il est amplifié et subit différentes transformations dues aux degrés d'ouverture et de fermeture au niveau de chaque cavité, à la position de la langue, des lèvres, etc. Ces cavités possèdent des fréquences de résonance qui renforcent certaines régions du spectre des sources excitatrices. En phonétique acoustique, les fréquences renforcées aux régions des fréquences de résonance correspondant aux cavités vocales sont désignées par le terme de "formants". Ainsi, ces derniers sont les paramètres acoustiques qui permettent d'étudier et d'expliquer les phénomènes physiologiques que subit le son laryngé lors de son passage à travers les différentes cavités vocales. Les valeurs des formants varient selon le volume de la cavité et la surface de l'ouverture du résonateur. De façon générale, un formant a une valeur de fréquence inversement proportionnelle au volume de la cavité. Plus le volume de cette dernière est grand, plus cette fréquence est basse, et vice versa.

Chaque son a ses formants caractéristiques. Sur un sonagramme, les formants sont représentés par des bandes noires (Figure 1.5). Au moins deux à trois formants sont nécessaires pour produire les différentes voyelles. Par contre, nous pouvons aller jusqu'à cinq formants au niveau des voyelles adjacentes pour discriminer entre les consonnes. Nous admettons généralement que la position fréquentielle des trois premiers formants caractérise le timbre vocalique : F_1 prend naissance dans la cavité résonante comprise entre le larynx et le dos de la langue; F_2 prend naissance dans la cavité résonante située entre le dos de la langue et les lèvres et enfin F_3 dépend de l'arrondissement des lèvres.

En phonétique acoustique, les transitions formantiques ont également leur importance dans la caractérisation des sons de parole. Ces transitions représentent les passages de la Voyelle vers la Consonne [VC] ou vice-versa, lors d'un acte de parole, en tenant compte de l'influence que subissent les formants durant cette période

transitoire. Ainsi, ces transitions permettent de représenter les phénomènes physiologiques apparaissant dans les cavités vocales lors de la phase de transition qui correspond au passage d'un son vers un autre son adjacent. Nous rappelons que dans un acte de parole, les sons ne sont pas isolés, ils se chevauchent et s'influencent les uns les autres (phénomènes de coarticulation). Ce sont des éléments essentiels de la reconnaissance phonétique. Ainsi, des études ont montré que le déficit de traitement phonologique chez les dyslexiques (troubles spécifiques de la lecture) serait le reflet d'un déficit de traitement et d'intégration de l'information linguistique lors de changements acoustiques rapides et brefs telles que les transitions formantiques [1]. Ces dernières véhiculent l'information probablement la plus importante que nous utilisons dans notre perception de très nombreuses consonnes. C'est cette raison d'ailleurs qui fait qu'une simple concaténation de diphones ou de syllabes pour synthétiser de la parole (passer automatiquement du texte vers la parole) est limitée du fait de l'inintelligibilité de la parole obtenue [2].

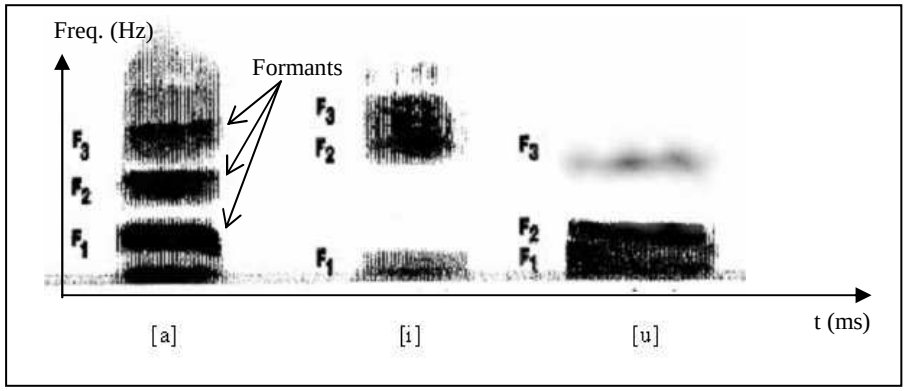


Figure 1.5 : Sonagrammes des voyelles [a], [i] et [u]

1.3.3. Intensité

L'intensité représente la qualité qui nous fait distinguer un son fort d'un son faible. Elle est liée à l'amplitude des vibrations sonores. Elle mesure l'énergie de l'acte phonatoire et dépend surtout de la pression d'air sous-glottique. Ainsi, l'amplitude d'une vibration peut être exprimée objectivement par le calcul des variations de pression d'air (exprimée en watt/cm^2). Nous utilisons toutefois plus fréquemment une unité de mesure relative, le décibel (dB). Une échelle illustre les niveaux de pression sonore auxquels l'être humain est soumis. Elle est graduée de 0 dB, seuil de perception de l'oreille humaine, à 120 dB considérée comme seuil de douleur. Une discussion normale a une

valeur d'intensité d'environ 60 dB. Une discussion forte peut aller jusqu'à 80 dB. Par contre, la parole chuchotée est à environ 10 dB.

1.3.4. Durée

La durée d'un son représente le laps de temps pendant lequel nous percevons ses vibrations par notre oreille. Elle est mesurée en millisecondes (ms) ou en secondes (sec). Une bonne détermination de ce paramètre acoustique est cruciale pour assurer le naturel de l'élocution en synthèse de la parole. Une durée erronée peut produire une parole chaotique et parfois difficilement intelligible, pouvant provoquer un changement du sens du mot ou de la phrase. C'est le cas en langue arabe, où ce paramètre est pertinent. Ainsi, les deux mots [ğamal] (chameau) et [ğamāl] (beauté) présentent deux sens différents même s'ils ne diffèrent que par la durée temporelle de la dernière voyelle.

En pathologie vocale, nous nous intéressons également au Temps Maximum de Phonation (TMP) qui représente la mesure du temps maximal d'émission vocale sur un [a] tenu, à une hauteur et une intensité confortables. En d'autres termes, il s'agit de la tenue d'un son le plus longtemps possible, après une inspiration maximale. La longueur du TMP dépend à la fois de la capacité pulmonaire et de la qualité d'accolement des cordes vocales. C'est un bon indicateur du rendement de la source vocale, car plus la fuite glottique est conséquente et plus le TMP est court. Dans un cas normal, la durée moyenne d'un [a] tenu varie entre 15 et 25 secondes.

1.3.5. Timbre vocal

C'est la couleur du son vocal à partir de laquelle, nous pouvons identifier une personne à la simple écoute de sa voix (par exemple, lors d'une conversation téléphonique, etc.). Le timbre vocal dépend de trois critères essentiels : l'accolement des cordes vocales, leur épaisseur et enfin les caractéristiques anatomiques des différentes cavités de résonance de l'appareil phonatoire. Par ailleurs, selon que les ouvertures glottiques se font plus ou moins rapidement, le spectre vocal est plus riche en aigus et inversement. Les cavités de résonances contribuent également à la couleur de la voix, car en modifiant leurs volumes, nous obtenons telle ou telle voyelle.

Le timbre vocal est l'un des paramètres acoustiques les plus caractéristiques des voix pathologiques. Selon que la voix est normale ou pathologique, nous distinguons les différents timbres :

- **clair :** caractéristique de la voix à l'état normale (voix naturelle), avec un accolement ferme et sans aucun dysfonctionnement des cordes vocales. C'est un timbre riche en harmoniques ;
- **serré :** le patient en forçant sur son larynx, qui a une position trop haute, entraîne une contraction et une diminution du volume de résonateurs. La pression expiratoire est très importante et provoque un accolement brutal des cordes vocales. La voix présente beaucoup trop d'harmoniques aigus ;
- **sombre :** la voix est exagérément grossie avec un larynx trop bas par manque de tonicité musculaire. Les résonateurs sont trop ouverts et sans tonicité. Les cordes vocales ne s'accolent pas suffisamment, entrainant des fuites d'air. Les cavités de résonance influent faiblement sur la voix et l'articulation devient floue ;
- **éraillé :** résultat de la superposition d'une vibration parasite irrégulière sur le son fondamental laryngé, provoqué par une lésion du bord libre d'une corde vocale. La vibration perd de sa souplesse et de sa régularité. La glotte n'est pas totalement fermée ;
- **voilé :** les cordes vocales présentent un défaut de fermeture modéré entraînant une perte des harmoniques aigus. La voix est de faible intensité. Elle manque de netteté et clarté ;
- **rauque :** la voix est grave, le signal de parole est apériodique. Les cordes vocales s'accolent mal. Elles semblent rigides du fait d'une altération de leur capacité vibratoire. La voix est émise dans le bas pharynx. En conséquence, nous avons une sensation d'effort avec un son rugueux et dur ;
- **soufflé :** il y a adjonction d'un bruit de souffle par défaut très important de fermeture glottique et la présence en excès d'harmoniques aigus.

1.4. Classification des sons de la parole

Nous ne pouvons procéder à une analyse acoustique de la parole sans avoir au préalable des connaissances en phonétique physiologique ou articulatoire. En effet, les caractéristiques acoustiques permettent de modéliser physiquement les phénomènes physiologiques, d'où la complémentarité physico-acoustique en analyse de la parole. L'objectif essentiel de la caractérisation physiologique ou articulatoire de la parole est de classer les différentes consonnes d'une langue donnée en groupes de caractéristiques appelées modes et lieux d'articulation.

1.4.1. Modes d'articulation

Le mode d'articulation est défini par un certain nombre de facteurs qui modifient la nature du courant d'air expiré. Parmi ces facteurs, nous pouvons citer les plus importants :

- passage libre ou mise en vibration de l'air au niveau de la glotte (son sourd ou sonore) ;
- passage par une voie unique ou deux voies différentes (son oral ou nasal) ;
- obstruction totale ou partielle du passage de l'air dans un lieu du conduit vocal (son occlusif ou fricatif).

Les consonnes occlusives sont produites par obstruction totale du conduit vocal (occlusion) de brève durée empêchant momentanément l'air de sortir (implosion), suivie d'une ouverture articulatoire expirant brutalement l'air emmagasiné dans le conduit vocal (explosion). Ils apparaissent sur le sonagraphe, sous forme d'un silence plus ou moins court correspondant à la phase de la tenue articulatoire de l'occlusion. Lorsque nous n'observons aucune amplitude d'énergie à basses fréquences dans cette zone de silence, la consonne est dite non voisée. Quand cette zone contient de l'énergie à basses fréquences, étalée le long d'une barre horizontale nommée "barre de voisement", la consonne est voisée. Cette durée de la tenue est suivie d'une barre verticale correspondant à la "barre d'explosion" (due au relâchement de l'occlusion) ou "Burst" (Figure 1.6).

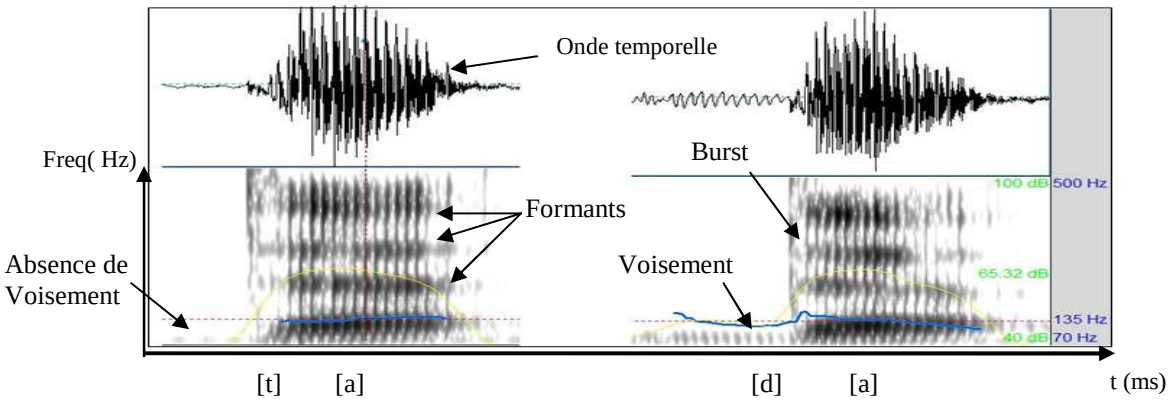


Figure 1.6 : Sonagrammes des occlusives sourde [t] / sonore [d] en contexte vocalique [a]

Les consonnes fricatives sont produites par un rétrécissement au lieu d'articulation du conduit vocal, lors du passage de l'air pulmonaire. Sur le sonagramme, elles apparaissent sous forme d'un bruit aléatoire. Les consonnes fricatives peuvent être voisées ou non voisées. Le voisement est caractérisé par une présence d'une barre horizontale d'énergie à basses fréquences. Il est représenté également par une courbe dite de voisement (Figure 1.7). L'absence de cette bande d'énergie correspond au trait sourd (non voisé) correspondant à la non de vibration des cordes vocales.

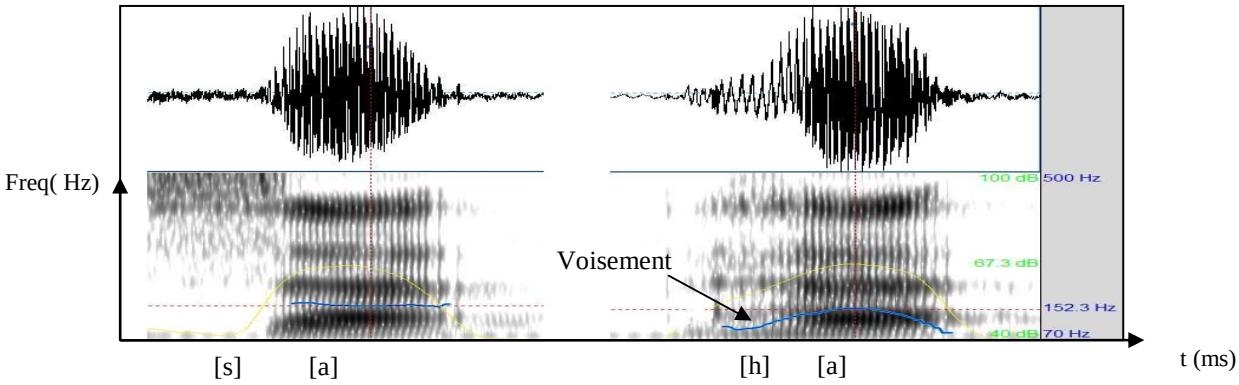


Figure 1.7: Sonagrammes des fricatives non voisée [s] / voisée [h] en contexte vocalique [a]

1.4.2. Lieux d'articulation

Le lieu d'articulation est l'endroit où se trouve un obstacle au passage de l'air dans les cavités phonatoires. En général, c'est la région où vient le plus souvent se placer la langue pour obstruer le passage du canal d'air (Tableau 1.1).

Tableau 1.1 : Lieux d'articulations selon les régions d'obstruction

Régions d'obstruction	**Lieux d'Articulations**
Lèvres	labiales ou bilabiales
Dents	Dentales
lèvres et dents	Labiodentales
Alvéoles	Alvéolaires
Palais	Palatales
voile du palais	Vélaires
Uvule ou luette	Uvulaires
Pharynx	Pharyngales
Glotte	Glottales

1.5. Sons de l'Arabe Standard (AS)

La langue Arabe comprend 28 Consonnes pour seulement 6 voyelles (3 voyelles brèves [a], [i] et [u] et 3 voyelles longues [ā], [ī] et [ū]). C'est une langue consonantique contrairement à l'Anglais ou le Français qui présentent beaucoup plus de voyelles. Elle se compose également d'un certain nombre de graphismes, tels que le [sukūn] symbolisé par un petit rond (°) apposé sur une consonne lorsque celle-ci n'est liée à aucune voyelle, le tanwīn marqué par le dédoublement des voyelles finales (◌ً, ◌ٍ, ◌ٌ), et les phénomènes propres à la langue (emphase, gémination, …) [3,4]. Le Tableau 1.2 montre la transcription phonétique que nous avons choisie dans le cadre de notre travail.

Tableau 1.2 : Transcription phonétique des sons de l'Arabe Standard

	Mode d'articulation		Transcription phonétique	Caractère arabe	Exemple en Arabe	Exemple en Français
Consonnes	occlusives	Voisées	b	ب	بلد	Pays
			d	د	دالة	Fonction
			ḍ	ض	ضجيج	Bruit
		non voisées	t	ت	تجربة	Expérience
			ṭ	ط	طريقة	Méthode
			k	ك	كلام	Parole
			q	ق	قرن	Siècle
			ʔ	ء	أداة	Outil
	fricatives	Voisées	ḏ	ذ	ذاكرة	Mémoire
			ẓ	ظ	ظهر	Dos
			ɛ	ع	عمل	Travail
			ɣ	غ	غناء	Richesse
			z	ز	زمن	Temps
			h	ه	هوية	Identité
		non voisées	f	ف	فصل	Chapitre
			ɵ	ث	ثلج	Neige
			s	س	سنة	Année
			ṣ	ص	صورة	Image
			š	ش	شبكة	Réseau
			x	خ	خليّة	Cellule
			ħ	ح	حجم	Taille
	nasales		m	م	ميدان	Domaine
			n	ن	نواة	Noyau
	vibrante		R	ر	رمز	Symbole
	latérale		l	ل	لغة	Langue
	semi-voyelles		w	و	وزن	Poids
			j	ي	يوم	Jour
	affriquée		ǧ	ج	جامعة	Université
Voyelles	voyelles brèves		a	◌َ		
			i	◌ِ		
			u	◌ُ		
	voyelles longues		ā	◌ا◌َ		
			ī	◌ي◌ِ		
			ū	◌ُو		

Le système vocalique de l'Arabe Standard (AS) se compose de trois voyelles brèves [a, u, i], appelées « ħarakāte », et trois voyelles longues [ā, ū, ī], notées également [a:, u:, i:], appelées « ħurūf al-madd ». Cette opposition temporelle brève/longue est fondamentale aux niveaux grammatical et sémantique. En effet, les deux mots [sabaqa] (devancer) et [sābaqa] (concourir) présentent deux sens différents même s'ils ne diffèrent que par la durée temporelle de la première voyelle. Le système vocalique de l'AS se présente comme suit (Figure 1.8).

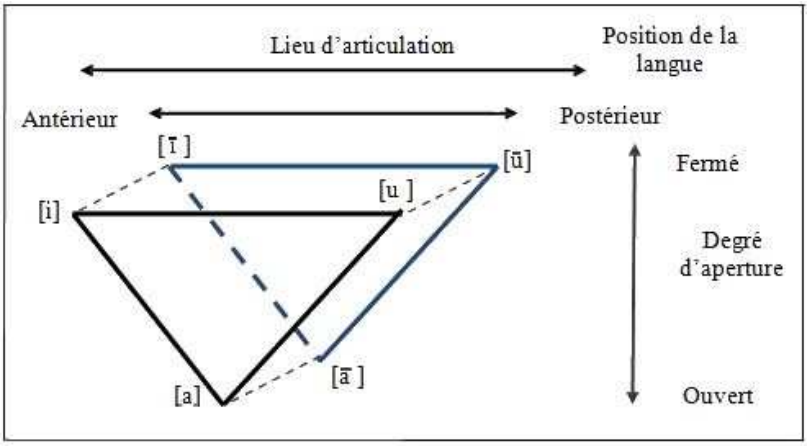


Figure 1.8 : Système vocalique de l'Arabe Standard [5]

1.5.1. Consonnes spécifiques

Les consonnes spécifiques de l'AS sont au nombre de huit (Tableau 1.3) : Quatre occlusives (une est voisée et trois sourdes) et quatre fricatives (deux voisées et deux sourdes). En dehors des emphatiques, toutes les autres consonnes spécifiques ont pour lieux d'articulation des régions postérieures de l'appareil phonatoire (uvulaire, pharyngale et glottale).

Tableau 1.3 : Modes et lieux d'articulation des sons spécifiques de l'Arabe Standard

Consonne	Caractère Arabe	Lieux d'articulation	Mode d'articulation			
			Voisé	Emphatique	Occlusif	Fricatif
[ṭ]	ط	dental	-	+	+	-
[ṣ]	ص	alvéolaire	-	+	-	+
[ḍ]	ض	Alvéodental	+	+	+	-
[ḏ]	ظ	Interdental	+	+	-	+
[q]	ق	Uvulaire	-	-	+	-
[ħ]	ح	Pharyngal	-	-	-	+
[ɛ]	ع	Pharyngal	+	-	-	+
[ʔ]	ء	Glottal	-	-	+	-

Les sonagrammes de quelques consonnes spécifiques nous montrent la différence entre les modes (occlusif/fricatif) et (voisé/non voisés) (Figure 1.9).

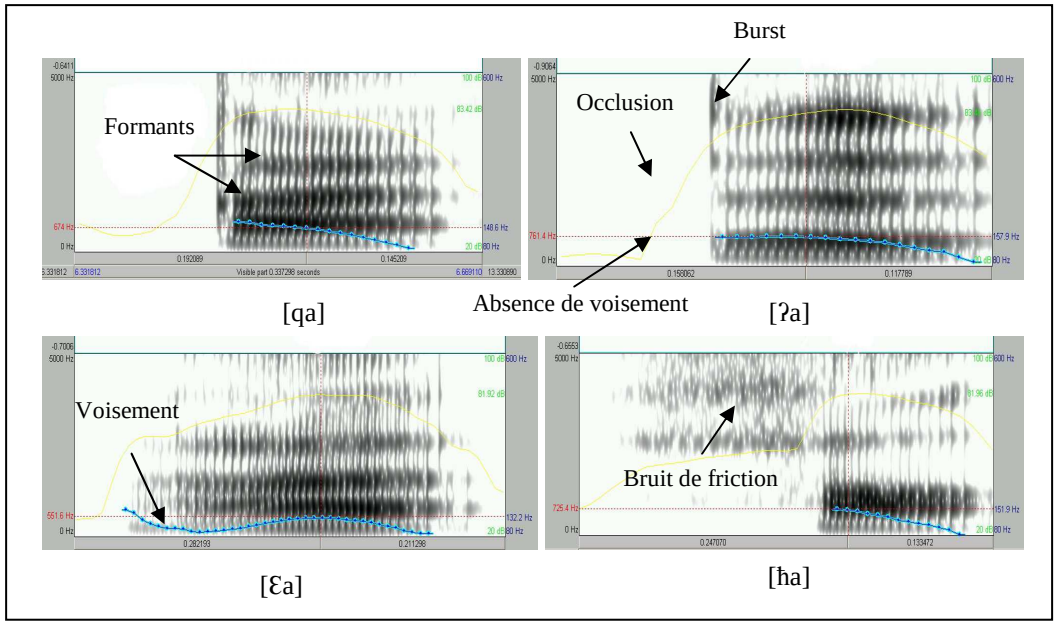


Figure 1.9 : Sonagrammes des occlusives [q] et [ʔ] et fricatives [Ɛ] et [ħ] en contexte vocalique [a]

1.5.2. Phénomènes d'emphase et de gémination

L'emphase et la gémination sont deux phénomènes physiologiques importants de la langue Arabe. Différentes études ont fait l'objet de multiples controverses du point de vue des mouvements et des positions des organes articulatoires lors de leur prononciation, ainsi que l'effet acoustique résultant [3, 4].

L'emphase correspond à la pharyngalisation des consonnes dont les lieux d'articulations sont antérieurs (devant de la langue). Les consonnes emphatiques présentent un même lieu d'articulation que leur opposées non emphatiques mais diffèrent par le deuxième lieu d'articulation pharyngale. Les consonnes emphatiques de l'AS sont respectivement (Tableau 1.3) :

- l'occlusive alvéodentale voisée [d̤] ;
- l'occlusive apicodentale non voisée [t̤] ;
- la fricative interdentale voisée [ð̤] ;
- la fricative alvéolaire non voisée [s̤].

Selon les anciens Grammairiens Arabes, la consonne emphatique [d̤] caractérise la langue Arabe. Son lieu d'articulation est latéral et correspond au bord extrême de la langue, du côté des molaires droites. Au Maghreb, cette consonne est généralement confondue dans la prononciation avec l'autre consonne fricative [ð̤].

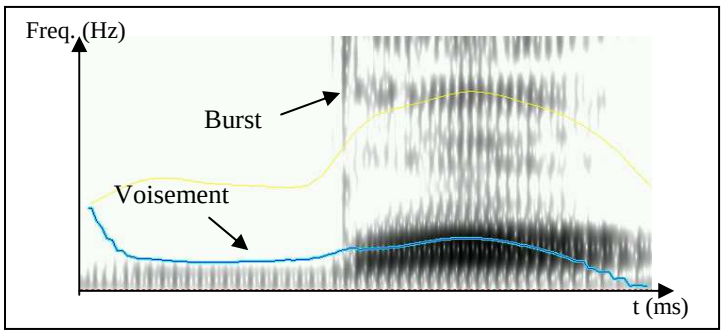


Figure 1.10 : Sonagramme de la consonne [ḍ] en contexte vocalique [a]

Sur le plan articulatoire, le phénomène d'emphase consiste en un report en arrière de la racine de la langue et en un abaissement et creusement du dos de la langue (Figure 1.11), et ainsi un élargissement de la cavité buccale et une constriction du pharynx [3].

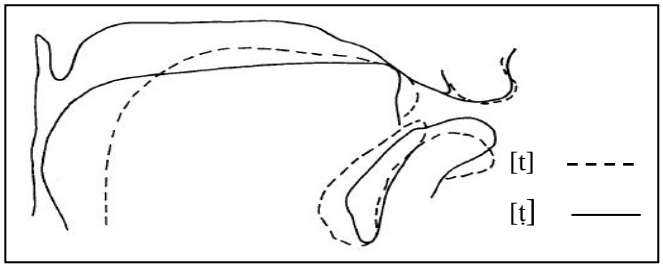


Figure 1.11 : Articulation de la consonne emphatique [ṭ] et son opposée [t]

Sur le plan acoustique, nous remarquons une chute du formant acoustique F_2 due à l'élargissement de la cavité buccale et une montée du formant acoustique F_1 due au rétrécissement de la cavité pharyngale (Figures 1.12).

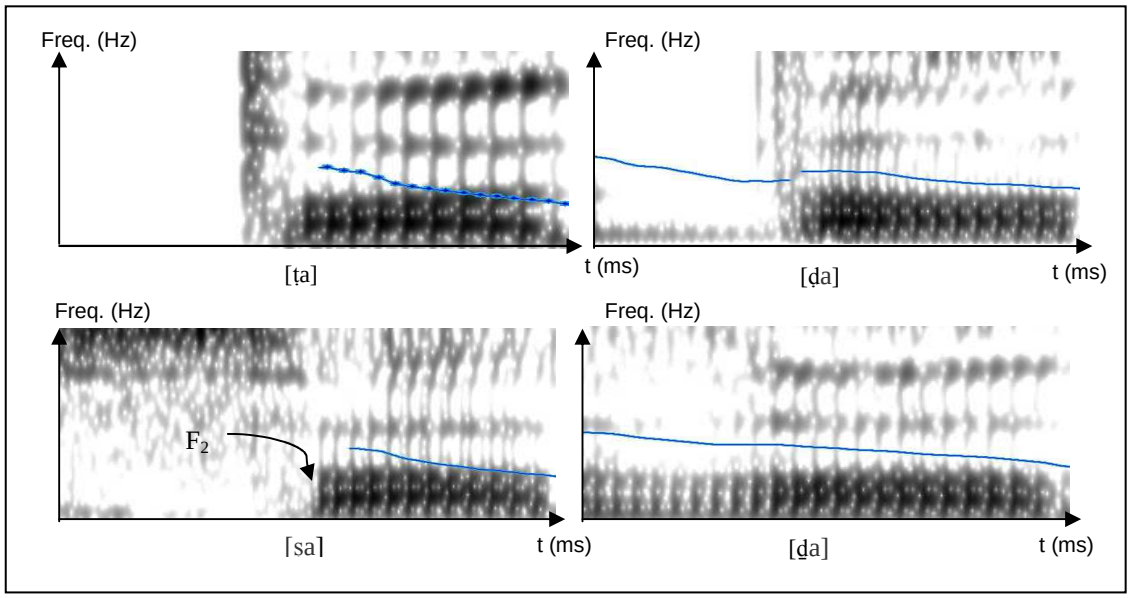


Figure 1.12 : Chute de F_2 en contexte emphatique [C_ea]

En présence d'une voyelle [i], nous remarquons également une chute du niveau de F_2 de l'emphatique occlusive, comme le montre la figure 1.13.

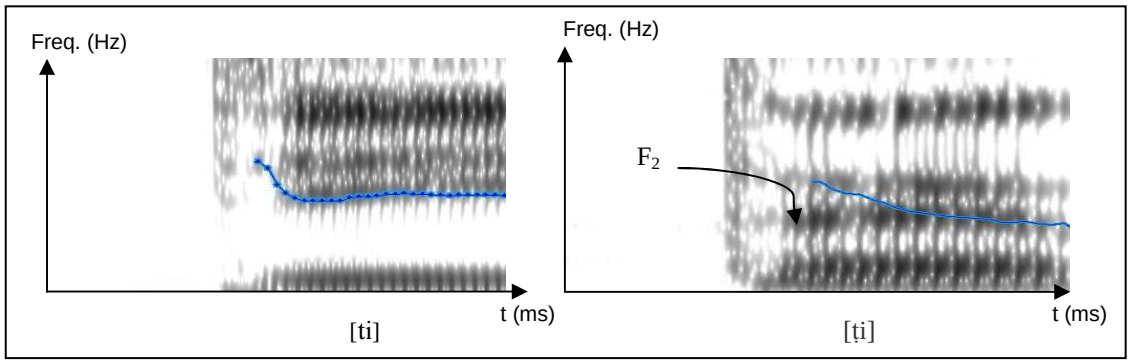


Figure 1.13 : Sonagrammes de l'emphatique occlusive [ṭ] et son opposée [t] en contexte [C_ei]

La même chute du niveau de F_2 est aussi observée pour l'emphatique fricative en contexte vocalique [i] (figure 1.14).

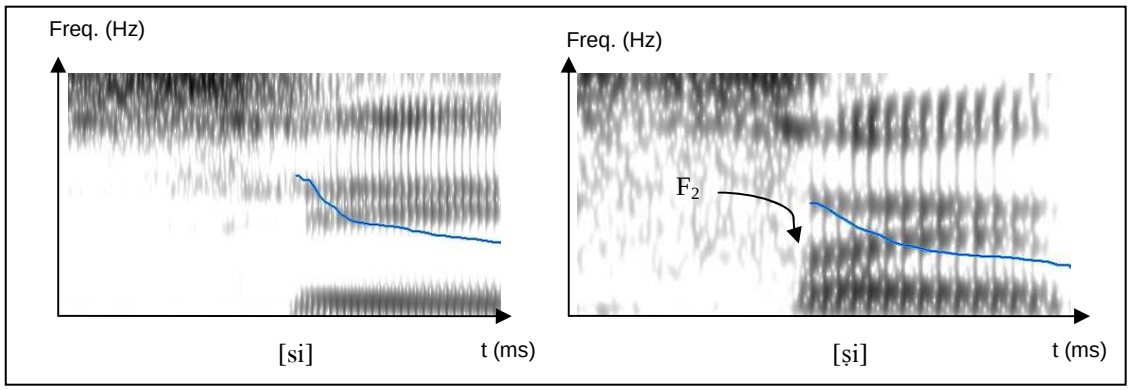


Figure 1.14 : Sonagrammes de l'emphatique fricative [ṣ] et son opposée [s], en contexte [C_ei]

La gémination correspond à la contraction de deux consonnes identiques en une seule dite géminée, comme dans l'exemple [Sabbaba] (il a provoqué) [4]. Une consonne peut être assimilée également dans la prononciation par la consonne qui la suit, provoquant une gémination, ainsi [man yakūn] devient dans la prononciation [mayyakūn]. Pour Sibawayh, ancien grammairien arabe, l'occurrence de deux consonnes identiques entraîne une lourdeur articulatoire. La réalisation de deux consonnes différentes étant plus légère que celle de deux consonnes (identiques) qui ont un même lieu d'articulation. En d'autres termes, il est lourd d'employer sa langue au sortir d'un lieu d'articulation pour l'y faire aussitôt revenir. Aussi, à cause de cette fatigue qu'apporte la réalisation de deux articulations identiques dans un même lieu, sans un intervalle de temps, cette réalisation est rejetée en faveur de la gémination des

deux consonnes identiques afin qu'il n'y ait qu'une seule élévation de la langue avec un moindre effort.

1.6. Conclusion

Nous avons donné un bref rappel sur la physiologie et l'acoustique de la parole. Dans la première partie, nous avons rappelé les différents modes et lieux d'articulation des consonnes. Dans la seconde, nous avons mis en valeur les paramètres acoustiques les plus exploités et qui permettent de caractériser le signal de parole. Nous avons enchaîné sur les sons de l'Arabe Standard où nous avons montré les caractéristiques physico-acoustiques les plus importantes des consonnes spécifiques.

Il reste que l'objectif essentiel de la présentation de ce chapitre est que l'étude anatomique et physiologique des différents organes constitutifs de l'appareil phonatoire et l'étude acoustique de la production vocale sont un préalable indispensable à l'approche, la compréhension, et la prise en charge des troubles de la voix et de la parole.

CHAPITRE 2

PATHOLOGIES DE LA PAROLE

2.1. Introduction

Dans ce chapitre, nous avons donné une brève description des diverses pathologies de la parole les plus fréquentes. Nous avons donné une attention particulière à deux cas pathologiques, à savoir les paroles parkinsonienne et œsophagienne, sur lesquelles nous avons fait une analyse acoustique et une classification automatique par rapport à la parole normale.

2.2. Pathologies de la parole

La parole humaine peut être atteinte par de nombreux dysfonctionnements. Nous pouvons citer parmi les plus importants, les défauts de prononciation et les lésions aux cordes vocales. Parmi les pathologies vocales les plus fréquentes, nous avons :

2.2.1. Dysphonie

D'une manière générale, la dysphonie (du grec *dys:* difficulté ou manque, *phonie*: parole) se définit par l'altération de l'un ou de plusieurs éléments du "trépied" acoustique de la voix : la fréquence fondamentale, l'intensité et le timbre. La voix peut être enrouée, voilée, soufflée ou devenir plus grave, avec une faible intensité. Lorsque la voix n'est plus audible ou réduite au chuchotement, on parle d'aphonie. Parmi les cas de dysphonie les plus connus, nous pouvons citer la paralysie récurrentielle des cordes vocales. Cette dernière est une pathologie fréquente qui concerne la paralysie d'une corde vocale. Si cette dernière est bloquée en ouverture, la fuite d'air sera importante avec une aphonie et des risques de fausses routes alimentaires. En revanche, si elle est bloquée en fermeture, nous rencontrons essentiellement un problème de dysphonie.

2.2.2. Dysarthrie

La dysarthrie (du grec *dys:* difficulté ou manque, *arthron:* jointure) est un trouble de la réalisation motrice de la parole lié à un mauvais fonctionnement des groupes musculaires responsables de la production de la parole, conséquence de lésions au niveau du système nerveux central. Ces lésions entraînent une atteinte neurologique touchant l'exécution motrice, comme dans le cas de la maladie de Parkinson.

2.2.3. Dyslalie

La dyslalie (du grec *dys:* difficulté ou de manque, *lalein:* parler) est un trouble de la communication caractérisé par des difficultés d'articulation de certains sons, dues à des

malformations physiques. La dyslalie peut être d'origine fonctionnelle, tels que le bégaiement et le sigmatisme, ou organique, telles que les fentes palatines.

2.2.3.1. Bégaiement

Le bégaiement est un trouble de la parole affectant le débit d'élocution. Il est caractérisé par des répétitions ou prolongations involontaires des sons, syllabes, ou mots, et par des pauses involontaires (blocages) où le malade est incapable de produire un son. La plupart des "bègues" en sont atteints dès leur prime enfance, néanmoins des adultes peuvent en souffrir après un accident ou un choc émotionnel, alors qu'ils n'ont jamais eu ce trouble auparavant.

2.2.3.2. Sigmatisme

Le sigmatisme est la mauvaise articulation des consonnes, en particulier les fricatives. C'est un des troubles dyslaliques les plus fréquents chez l'enfant, car ce type de consonnes nécessite une précision très importante de l'articulation. Selon son origine, nous pouvons classifier le sigmatisme en plusieurs classes :

- sigmatisme nasal, dû à un positionnement de la langue qui rend impossible le passage de l'air par la cavité buccale ;
- sigmatisme dorsal, dû à un soulèvement excessif de la langue ;
- sigmatisme occlusif, dû à un remplacement systématique de toute consonne fricative par la consonne occlusive dont le point d'articulation est le plus proche ;
- sigmatisme entraînant une confusion entre consonnes sourdes et sonores, dû à une non vibration des cordes vocales dans certaines consonnes voisées. Ce genre de sigmatisme est fréquent chez l'enfant sourd.

2.2.3.3. Fentes labio-palatines

Les fentes labio-palatines sont des malformations du bas du visage qui apparaissent tôt durant le développement embryonnaire. La fente labiale est une absence de fusion du tissu embryonnaire du visage aboutissant à une perte de substance de la lèvre supérieure. La fente palatine est une absence de substance de la voûte buccale aboutissant à une communication entre le nez et la bouche. Les fentes labiales sont plus fréquentes que les palatines. Les fentes labio-palatines ont évidemment de lourdes répercussions sur l'articulation des sons, empêchant une prononciation correcte des consonnes labiales (fentes labiales) et orales par nasalisation (fentes palatines).

Une manière de classer les pathologies vocales est de les scinder en deux groupes essentiels :

- les pathologies d'origine fonctionnelle ou neurologique : pour ce genre de pathologies, l'appareil phonatoire est intact. Le trouble provient de déficiences fonctionnelles neurologiques telles que l'aphasie, les maladies de Parkinson et Alzheimer, etc. Ce groupe de pathologies est plus connu sous le nom de "dysarthries". Une des pathologies d'origine neurologique est la maladie de Parkinson ;
- les pathologies d'origine organique ou physiologique : le trouble est causé par des déficiences au niveau de l'appareil phonatoire, telles qu'une ablation du larynx, une déformation de la langue, de la luette, une anomalie congénitale des cordes vocales, etc. Le larynx et en particulier les cordes vocales sont à la base de la production de la voix, une atteinte de ces organes causera inévitablement un trouble de la voix. Une des pathologies d'origine organique les plus connues est la Laryngectomie Totale.

2.3. Maladie de Parkinson

La maladie de Parkinson est connue comme l'une des plus importantes maladies neurodégénératives après la maladie d'Alzheimer. Elle touche plus de 4 millions de personnes dans le monde [6]. En Algérie, les estimations sont de 150 000 patients avec un enregistrement d'environ 2000 nouveaux cas par année, ce qui est considéré comme un taux assez important. Cette maladie se caractérise par la disparition d'un certain nombre de neurones qui sécrètent un neurotransmetteur appelé "dopamine", contenu dans la substance noire (ou locus niger) du cerveau (Figures 2.1 et 2.2).

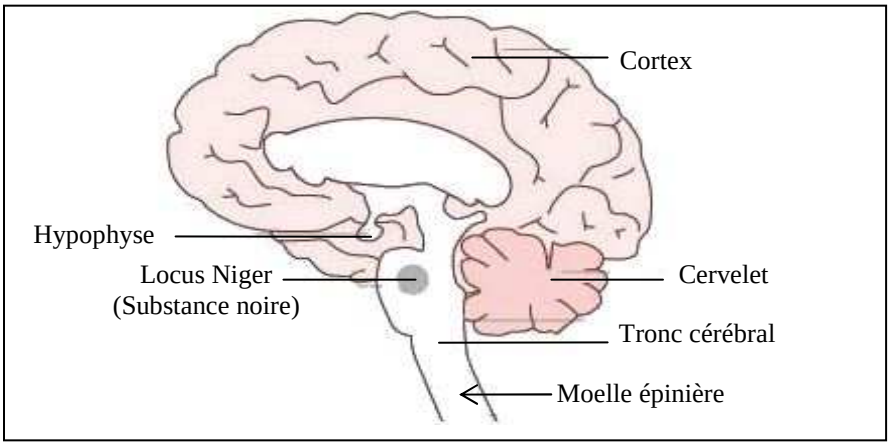


Figure 2.1 : Région de la substance noire dans le cerveau

Le neurotransmetteur est indispensable à la survie des cellules et au contrôle des mouvements nécessaires à l'équilibre général de l'organisme. Cette dégénérescence de neurones se caractérise par un tremblement, une lenteur des mouvements et une raideur, qui se traduisent par des troubles de la parole. Bien que cette maladie débute généralement entre 55 et 65 ans, 5 à 10 % des patients sont atteints à des âges bien moins avancés (entre 30 et 55 ans).

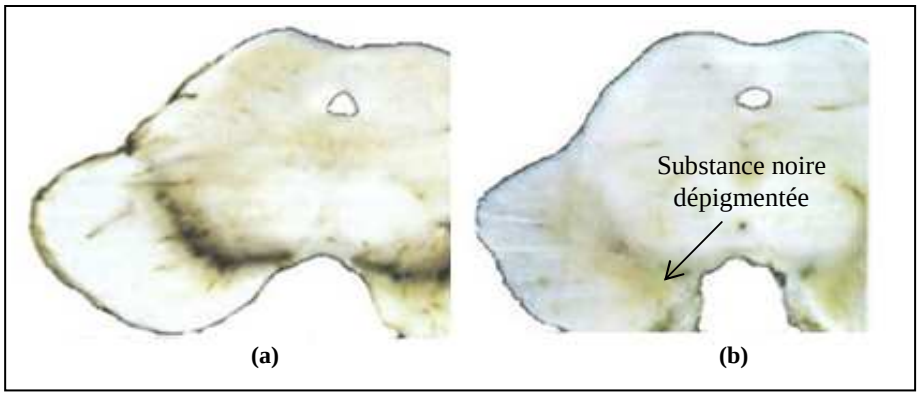


Figure 2.2 : Substance noire (locus niger) dans le cerveau
(a) Cas normal (b) Cas d'une dépigmentation, Maladie de Parkinson

Les causes de cette maladie ne sont pas encore déterminées. Elle pourrait être la conséquence de l'interaction entre une prédisposition génétique et des cofacteurs environnementaux. Depuis de nombreuses années, des toxiques environnementaux, métaux lourds et pesticides notamment, sont suspectés mais sans preuves d'une cause unique [7, 8].

2.3.1. Historique de la maladie de Parkinson

La première description de la maladie de Parkinson a été faite en 1817 par James Parkinson lui-même, dans son ouvrage intitulé "An essay on the Shaking Palsy" [9]. Il a regroupé l'association du tremblement de repos et de la marche festinante (accélération involontaire de la marche à petit pas, le corps penché en avant) en une seule entité qu'il a nommé "paralysie agitante".

Brièvement, nous pouvons résumer les grandes dates associées à cette maladie comme suit :

- en 1872, un clinicien et neurologue français du nom de J.M. Charcot proposa le terme de maladie de Parkinson, en décrivant, avec A. Trousseau et A. Vulpian, de façon explicite la démarche festinante et la rigidité dans les mouvements ;

- en 1895, E. Brissaud suspecta la substance noire (appelée également locus niger) comme la structure à l'origine des troubles ;
- en 1919, K. Tretiakoff, neurologue russe, affirma définitivement l'implication de la substance noire dans la maladie de Parkinson ;
- en 1951, W. Raab et W. Gigee découvrirent la présence d'un neurotransmetteur appelé "dopamine", contenue dans la substance noire ;
- en 1960, le rôle de la dopamine dans la maladie de Parkinson fut connu grâce aux travaux de H. Ehringer et O. Hornykiewicz, qui découvrirent un déficit dopaminergique du striatum chez les patients parkinsoniens ;
- en 2000, A. Carlsson, médecin et neurobiologiste suédois, a reçu le prix Nobel de médecine pour avoir démontré que la dopamine n'est pas seulement un précurseur de la noradrénaline et de l'adrénaline, mais également un neurotransmetteur important pour l'équilibre de la personne (Figure 2.3).

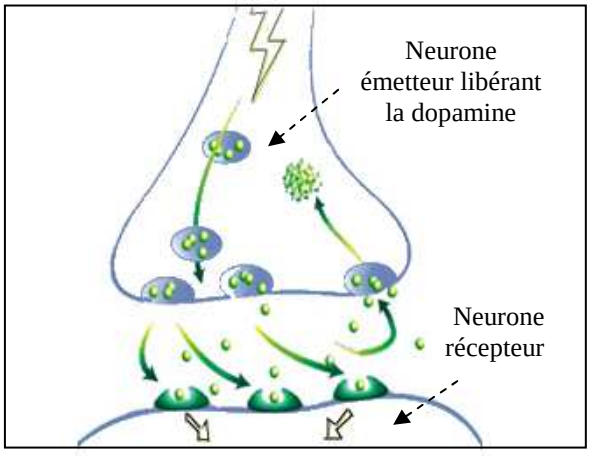


Figure 2.3 : Neurotransmission d'un neurone à un autre par la dopamine

De grandes personnalités sont porteuses de cette maladie. Nous pouvons citer entre autres, le célèbre boxeur Mohamed Ali Cassius Clay, Yasser Arafat, le Pape Jean Paul II, Adolph Hitler, etc.

2.3.2. Maladie de Parkinson et troubles de la parole

Un état de l'art des travaux réalisés dans le domaine nous montre une relation étroite entre la maladie de Parkinson et les troubles de la parole. Une étude de Logeman et al. montre que 45% des cas parkinsoniens présentent des troubles de la parole (dysarthrie hypokinétique), 89% des troubles de la voix (dysphonie parkinsonienne) et 20 % des troubles de la fluidité de la parole continue (pauses fréquentes et coupures de mots lors

d'un discours) [10]. De plus, du fait d'un défaut de fermeture correct du conduit vocal, les consonnes occlusives ressemblent généralement à des consonnes fricatives [11, 12]. L'occlusive étant réalisée de manière incomplète par les articulateurs, en l'absence d'une occlusion totale au lieu d'articulation de la consonne. Cette occlusion partielle provoque une fuite d'air, source de bruit de friction, qui nous fait percevoir une fricative à la place d'une occlusive. Ajouter à cela, les consonnes sourdes ont tendance à être sonorisées, notamment les occlusives dans un contexte vocalique. La dysarthrie parkinsonienne est caractérisée également par des erreurs sur le lieu d'articulation [13], des troubles du débit (accéléré ou ralenti) et une réduction des faits prosodiques (F_0 et intensité) [14-15]. D'une manière générale, nous observons une plage de variation de F_0 nettement plus réduite, avec un timbre souvent perçu comme rauque et soufflé par défaut d'accolement des cordes vocales.

En Algérie, peu d'importance est donnée à la rééducation orthophonique des patients. Seule une prise en charge par un médecin neurologue est assurée dans la majorité des hôpitaux.

2.4. Laryngectomie Totale

La Laryngectomie concerne l'ablation partielle ou totale du larynx, organe moteur de la phonation. L'intervention chirurgicale la plus répandue pour le traitement du cancer du larynx est une **L**aryngectomie **T**otale (**LT**), qui consiste à enlever complètement le larynx comprenant les cordes vocales et séparer les deux sources de la respiration et de la phonation par une trachéostomie (figure 2.4).

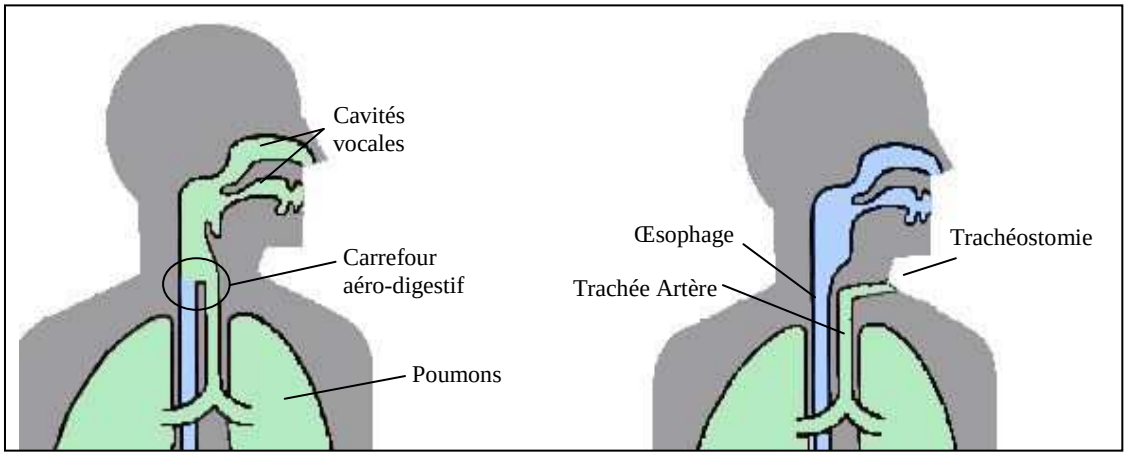


Figure 2.4 : Trachéostomie après Laryngectomie Totale

Dans le domaine des pathologies vocales, l'une des plus traumatisantes des maladies est le cancer du larynx. Ce dernier est lié principalement à la consommation du tabac fréquemment associée à un abus d'alcool. Le chirurgien pratique à la base du cou une ouverture qui sera permanente, que l'on appelle "trachéostome", et qui permet au patient de respirer en empruntant une nouvelle voie (figure 2.4). Ainsi, l'air ne peut plus passer par la bouche (ou le nez) et ne peut être exploité pour la phonation, comme dans le cas de la parole naturelle.

2.4.1. Historique de la Laryngectomie

Un bref historique des chirurgies laryngales et des techniques utilisées nous montre que les expériences dans le domaine dataient depuis très longtemps [16]. La 1ère Laryngectomie Partielle a été mise au point par Bowes en 1833, tandis que la 1ère Laryngectomie Totale a été effectuée par T. Billroth en 1873. Parmi les grands événements historiques de la laryngectomie, nous pouvons citer également quelques dates importantes :

- En 1951, un patient laryngectomisé J. Winter a mis au point la méthode d'injection dite "hollandaise" pour la rééducation par voix œsophagienne ;
- en 1978, M.I. Singer et E.D. Blom ont mis au point la 1ère "prothèse vocale miniaturisée", qui porte leur nom et dont nous admettons aujourd'hui, l'importance dans le cas de difficultés d'acquisition de la parole œsophagienne [17] ;
- en 1988, invention de la prothèse vocale miniaturisée, connue sous le nom de Provox®;
- à partir de 1997, introduction de la prothèse Provox de seconde génération (Provox®2), qui facilite l'occlusion manuelle du trachéostome par un simple mécanisme de valve, améliorant ainsi la phonation et les conditions d'hygiène.

L'évolution et le perfectionnement des techniques de laryngectomie aussi bien partielles que totales ont été guidés essentiellement, durant ces dernières années, par un double objectif : une meilleure chance de guérison de la maladie et une préservation maximale possible des fonctions de respiration, déglutition et phonation du patient [18].

2.4.2. Laryngectomie et troubles de la parole

Dans une phonation normale, trois conditions sont nécessaires pour parler :

- une source de pression d'air : le souffle d'air provenant des poumons ;

- une vibration de muscles élastiques, qui produisent le son de la voix : les cordes vocales ;
- des cavités vocales où se transforme le son laryngé en parole, avec la production des consonnes et voyelles.

Lors d'une LT, seule la dernière condition est partiellement satisfaite. La source d'air ne pourra plus venir des poumons et les cordes vocales n'existent plus. Donner de la parole à un laryngectomisé consiste donc à trouver une parole de remplacement, qui peut satisfaire les deux premières conditions manquantes. Pour cela, différents types de paroles de remplacement sont utilisées : la parole œsophagienne, la parole trachéo-œsophagienne ou implant phonatoire et la parole par prothèse externe. En Algérie, la méthode la plus utilisée est la parole œsophagienne, car elle est moins coûteuse que les prothèses et surtout elle arrive à donner au patient une parole assez intelligible qui lui permet de communiquer avec son entourage [19].

2.4.2.1. Parole Œsophagienne

La Parole Œsophagienne ($P\text{Œ}_{so}$) est l'approche prédominante de la réhabilitation de la parole de patients ayant subi une LT. Comme, le patient ne peut plus utiliser l'air provenant des poumons, c'est l'œsophage qui jouera désormais le rôle de réserve d'air, d'où le terme de "parole ou voix œsophagienne" (Figure 2.5).

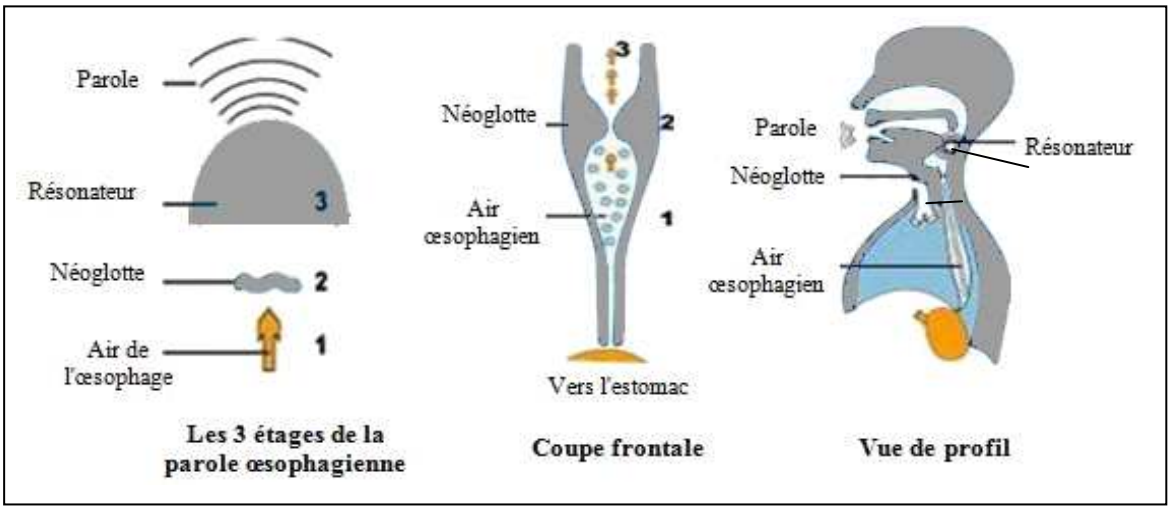


Figure 2.5 : Principe d'une Parole Œsophagienne

Il s'agit d'apprendre à produire un son d'éructation (un rot) dans l'entrée de l'œsophage, comme ce que nous faisons parfois involontairement après avoir mangé ou bu. De plus, à l'entrée de l'œsophage (sphincter supérieur) se trouve un petit

muscle qui peut vibrer en produisant un son, si une bouffée d'air remonte en le traversant. Ce néo-vibrateur exploité pour le remplacement des cordes vocales est généralement noté sous le terme de segment "**N**éo **V**ibrateur **P**haryngo-œsophagien" (**NVP**) ou simplement "Néoglotte". Nous retrouvons donc bien une analogie avec les trois conditions nécessaires à la parole : un souffle d'air provenant de l'œsophage, un muscle vibrant, souple et élastique qui est le segment NVP à l'entrée de l'œsophage et les cavités de résonance qui n'ont pas subi de changements significatifs. Il reste donc au patient à apprendre à exploiter ce nouveau mécanisme de la production de la parole.

2.4.2.2. Parole Trachéo-Œsophagienne ou Implant Phonatoire

Comparée à la parole œsophagienne, elle est plus facile à acquérir mais néanmoins plus coûteuse. Elle est obtenue grâce à la pose d'un implant phonatoire (prothèse interne), placé entre la partie supérieure de la trachée et l'œsophage (Figure 2.6). L'extrémité qui se trouve dans l'œsophage s'ouvre par un clapet de telle sorte que la communication ne soit possible que dans le sens trachée-œsophage. Ainsi, la salive, les boissons ou les aliments absorbés et descendant dans l'œsophage ne peuvent s'infiltrer à l'intérieur de l'implant, en direction de la trachée. Contrairement à la parole œsophagienne, le son est alimenté par l'air pulmonaire, en utilisant un mécanisme spécifique : pour parler, il suffit de boucher le trachéostome avec la main en même temps que l'on prononce des mots et des phrases. Quand on bouche l'orifice trachéal, le souffle est dévié vers l'œsophage par l'implant ; c'est pourquoi cette parole est appelée trachéo-œsophagienne.

L'utilisation de la prothèse interne permet de retrouver rapidement une voix et une parole d'assez bonne qualité puisqu'elle utilise le mécanisme naturel de la respiration. Le débit de parole est fluide et rapide. L'apprentissage peut-être de brève période. Il demeure néanmoins que les inconvénients sont aussi importants : Impossibilité de parler les mains libres, le trachéostome devant être fermé lors d'un acte de parole ; hygiène souvent peu satisfaisante du procédé ; nécessité du remplacement périodique de la prothèse (durée de vie de 8 mois environ). L'utilisation de la prothèse est notamment contre-indiquée lors d'une trop grande abondance de sécrétions bronchiques qui viendraient sans arrêt boucher la prothèse. De plus, certaines complications peuvent survenir : risque de rejet de la prothèse lors d'une toux violente ou d'un nettoyage maladroit, une surinfection de la muqueuse, des fausses-

routes de salive ou liquide qui s'infiltrent autour de la prothèse à partir de l'œsophage et glissent ensuite vers la trachée. Ceci explique le pourquoi de l'utilisation plus habituelle de la parole œsophagienne comme technique de rééducation. Cette technique est certes plus difficile à acquérir mais présentant moins de risques, liés aux inconvénients de la prothèse.

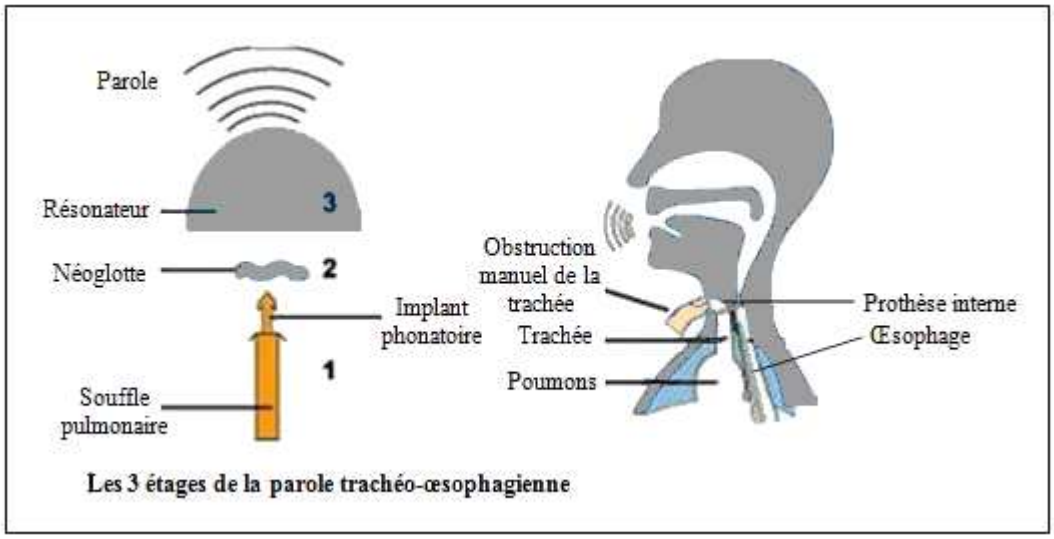


Figure 2.6 : Principe d'une parole trachéo-œsophagienne

2.4.3. Réhabilitation vocale par voie œsophagienne

Avant d'entamer toute rééducation, l'orthophoniste rééducateur devra nécessairement rappeler au patient, les conséquences anatomiques de la Laryngectomie : la trachée Artère étant déviée, la respiration ne se fera plus par la bouche ou par le nez. Ensuite, il lui expliquera le mécanisme de la nouvelle voix, qui consiste à produire un son vocal œsophagien en recréant un flux aérien qui fait vibrer le segment néo-vibrateur NVP. Pour cela, le patient doit faire pénétrer l'air dans l'œsophage et le faire ressortir en produisant une éructation volontaire, sonore et contrôlée. Les organes des cavités vocales (langue, luette, dents, lèvres, …) n'ont pas été touchés, l'articulation des sons continue à se faire comme avant. Pour introduire l'air dans l'œsophage et obtenir une émission sonore et contrôlée, plusieurs techniques sont possibles [20] :

- la déglutition, qui utilise le mécanisme naturel qui nous permet d'avaler. C’est la technique la plus ancienne. Malgré sa facilité d’exécution, elle est à éviter, car ses inconvénients sont nombreux : ballonnement digestif, voix saccadée et parole fractionnée ;

- l'inhalation qui consiste à aspirer l'air dans l'œsophage au cours d'une inspiration, en laissant la bouche entrouverte. Cette méthode entraîne une pénétration sonore de l'air dans l'œsophage et un bruit de souffle pulmonaire important ;
- la technique des blocages, qui utilise la pression glossopharyngienne pour refouler l'air de la bouche et du pharynx dans l'œsophage par le blocage des muscles buccaux. Ce blocage peut être labial [p], apico-dental [t] ou au niveau de la base de langue [k]. Cette technique faisant intervenir une pression abdominale, met en jeu plus d'énergie et donne une parole hachée et saccadée ;
- la méthode hollandaise, dite des consonnes injectantes, qui consiste en une injection d'air dans l'œsophage. Pendant l'articulation de consonnes telles que le [p] de la syllabe [pa], Il faut insister sur le serrage des lèvres, tête légèrement fléchie, l'entrée d'air se produit alors et il ressort de lui-même sans effort et avec peu d'énergie, lorsque l'on ouvre la bouche pour articuler la voyelle qui suit [20].

Pour avancer plus vite dans la rééducation, des cours sur la parole œsophagienne sont nécessairement prodigués aux patients pendant deux à trois semaines après l'opération. Les patients apprennent à connaître le nouveau mécanisme de la parole et les différentes techniques d'inhalation de l'air, d'éructation et de la vibration du segment NVP. Comme le locuteur œsophagien utilise un très faible réservoir d'air par inhalation (moins de 100 ml) comparé à un locuteur laryngien normal (environ 5 litres) [20], cela empêche logiquement ce locuteur à produire de longues suites de parole avec une seule charge d'air. Un travail assidu de la part du patient permet d'aboutir à l'utilisation réflexe de ce nouveau mécanisme de production de la voix. Des associations bénévoles sont même sollicitées, dans des pays comme l'Espagne, pour venir en aide aux patients pour leur donner des instructions sur la méthode d'injection de l'air [21].

La $PŒ_{so}$ comporte certes des paramètres acoustiques assez faibles, mais elle assure une réelle communication verbale avec une intelligibilité de la parole, au prix d'une rééducation durant au minimum six mois et comportant :

- l'acquisition d'un souffle buccal indépendant du souffle pulmonaire ;
- la vibration du segment NVP par les diverses techniques citées ci-dessus. La plus utilisée étant l'inhalation et ensuite l'éructation volontaire de sons à travers l'œsophage ;
- des productions syllabiques de variétés croissantes, permettant l'émission de mots puis de courtes phrases, au cours d'exercices conversationnels et un entraînement

articulatoire, au cours d'une rééducation réclamant autant de motivation pour le patient que de patience pour le rééducateur.

2.5. Conclusion

Dans ce chapitre, nous avons exposé les pathologies vocales les plus fréquentes. Nous avons décris, en particulier, deux cas de pathologies vocales : la pathologie d'origine fonctionnelle en prenant, comme exemple, la maladie de Parkinson et pour celle d'origine organique, nous avons présenté la Laryngectomie Totale. Une description détaillée de ces deux cas de pathologies nous a permis d'avoir une meilleure compréhension des troubles phonatoires que subissent les patients.

CHAPITRE 3

ANALYSE ACOUSTIQUE DE LA PAROLE PATHOLOGIQUE

3.1. Introduction

La voix et la parole ont connu un grand intérêt aussi bien dans l'étude de leurs physiologies, de leurs physiopathologies mais surtout de leurs moyens d'exploration et de mesure. Chaque mesure apporte des informations différentes sur les aspects de la production de la parole. Ainsi, la grande complexité des phénomènes acoustico-physiologiques de la production vocale rendent difficile l'élaboration d'une méthode unique de mesure et de quantification. En conséquence, aucune mesure ne suffit à elle seule à caractériser une Parole Pathologique (PP_{ath}).

L'analyse acoustique permet de faire des mesures et obtenir des indices de façon quantitative et objective. Nous l'opposons souvent au jugement perceptif, éminemment subjectif, variable d'un auditeur à l'autre ou changeant selon le contexte pour un même auditeur. De telles méthodes d'évaluation perceptive sont souvent inadéquates, car incapables de mettre en évidence de petites différences ou perturbations légères de la voix ou de la parole. Par contre, l'analyse acoustique s'avère comme un moyen plus crédible d'évaluer et mesurer les dysfonctionnements de la voix et de la parole.

Dans ce chapitre, nous avons présenté les différents paramètres acoustiques exploités en parole pathologique, en exposant particulièrement notre méthode d'extraction des paramètres acoustiques Jitter et Shimmer. Finalement, nous avons réalisé une analyse acoustique sur deux paroles pathologiques, l'une d'origine fonctionnelle (la parole Parkinsonienne) et l'autre d'origine organique (la parole œsophagienne).

3.2. Paramètres acoustiques de la Parole Pathologique (PP_{ath})

Une Parole Normale (PN_{orm}) est analysée essentiellement par l'observation des paramètres principaux :

- La fréquence fondamentale F_0 qui permet de mesurer les vibrations des cordes vocales ;
- les formants qui permettent d'étudier les effets que subissent les sons de parole lors de leurs passages à travers les cavités vocales ;
- la durée des sons pour étudier le débit d'air et la fluidité de la parole ;
- l'intensité qui permet de distinguer un son fort d'un son faible.

Cependant, l'analyse d'une Parole Pathologique (PP_{ath}) fait appel, en addition, à d'autres paramètres aussi importants tels que le degré de perturbation de F_0 (Jitter) et le

degré de perturbation de l'intensité (Shimmer), qui sont très exploités pour la caractérisation de la qualité de la PP_{ath} [22-25]. Ces deux paramètres sont habituellement mesurés sur les voyelles soutenues, et leurs valeurs au-dessus d'un certain seuil sont considérées comme étant liées à des PP_{ath} [26].

Dans le cadre de notre travail, nous avons mis au point les deux paramètres Jitter et Shimmer en utilisant le logiciel de programmation Matlab, version 2007. La méthode que nous avons adoptée pour déterminer ces deux paramètres est basée principalement sur la détection de pics glottiques par la technique dite "d'analyse cepstrale" (figure 3.3). Cette dernière permet une séparation distincte de la fréquence fondamentale produite par les cordes vocales et la fonction de filtre du conduit vocal (cavités vocales) [27]. Elle permet également de représenter les variations à court terme des paramètres spectraux du signal de parole et de fournir des informations détaillées sur F_0 [28-30]. De ce fait, nous avons exploité cette technique pour calculer le Jitter et le Shimmer, ainsi que le tracé des pics des Instants d'Ouverture Glottale (IOG) (Glottal Opening Instants GOI en Anglais). Rappelons que le cepstre d'un signal est défini comme la Transformée de Fourier Rapide Inverse (IFFT, Inverse Fast Fourier Transform en Anglais), calculée à partir du logarithme de l'amplitude de la Transformée de Fourier Rapide (FFT, Fast Fourier Transform en Anglais) d'une tranche du signal :

$$C(n) = IFFT\ (log|FFT(x(n))|) \tag{3.1}$$

$$x(n) = s(n) * h(n) \tag{3.2}$$

Avec :
x(n) : signal de parole ;
s(n) : signal d'excitation (source) ;
h(n) : fonction de transfert du conduit vocal (filtre).

Notons que l'espace de représentation du cepstre (espace *quéfrentiel*) est un axe temporel gradué en unités de "quéfrence" (anagramme de fréquence). Il est possible, par un filtrage temporel (*liftrage*), de séparer dans le signal, la contribution de la source *s* (la fréquence fondamentale) de celle du conduit vocal *h* (les formants). Pour estimer la contribution du conduit vocal dans le signal de parole, nous ne conservons que les premiers échantillons du cepstre *c(n)* qui correspondent en particulier aux informations sur les formants. Les échantillons du cepstre d'ordre plus élevé correspondent, en général, aux caractéristiques de la fréquence fondamentale des cordes vocales. Ainsi, pour obtenir une

estimation de F_0 à partir du cepstre, nous recherchons donc un pic dans la région des "quefrences supérieures", correspondant au signal d'excitation (figure 3.1).

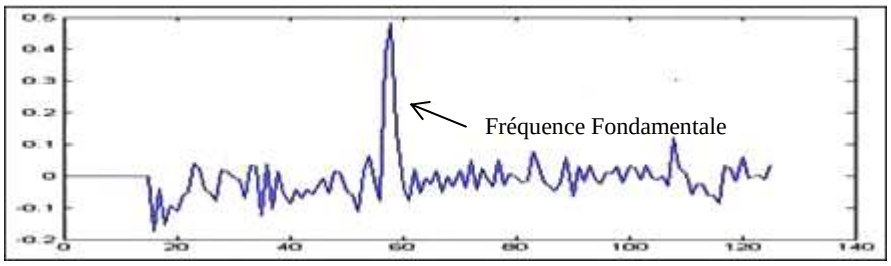


Figure 3.1 : Détermination de la fréquence fondamentale par cepstre

Après détection de F_0 et extraction de valeurs de crête à crête au cours du temps, nous exploitons les périodes locales et les amplitudes des pics glottiques pour extraire les valeurs du Jitter et du Shimmer (figure 3.2).

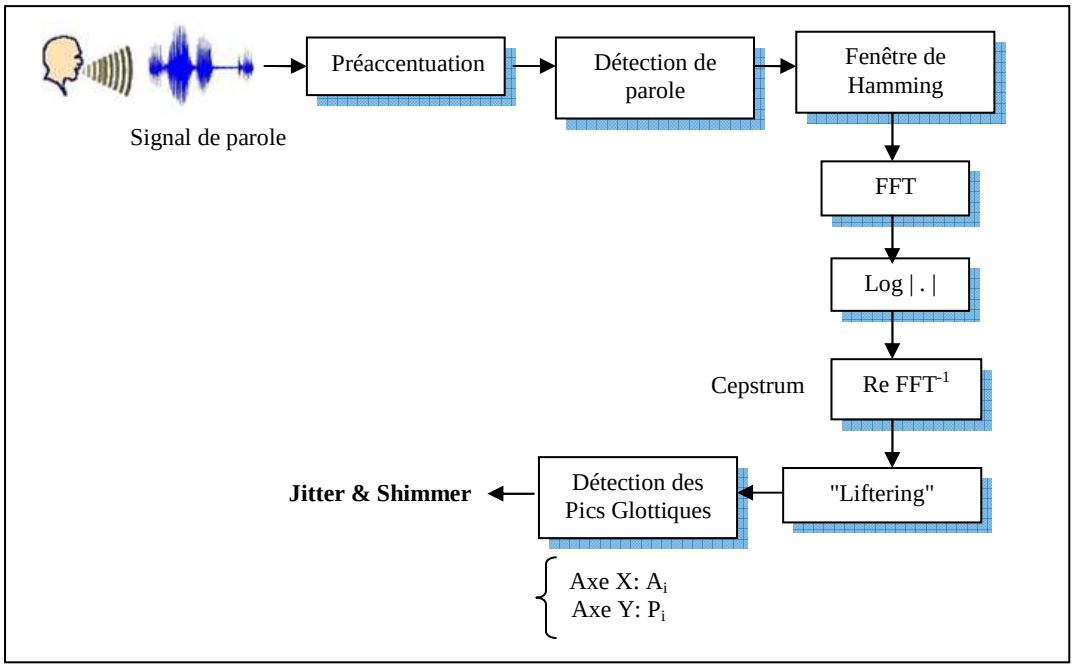


Figure 3.2 : Organigramme de calcul des paramètres Jitter et Shimmer

La Figure 3.3 montre les pics glottiques obtenus à partir d'un segment de la voyelle [a]. Les lignes verticales qui apparaissent sur l'onde temporelle représentent les pics des IOG.

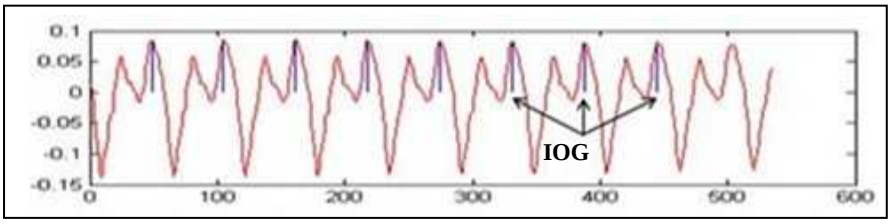


Figure 3.3 : Pics des Instants d'Ouverture Glottale (IOG)

3.2.1. Perturbation de F_0 (Jitter)

Le Jitter *J* réfère à la variation cycle par cycle de F_0 dans une trame du signal. Il se calcule par la moyenne de la différence de F_0 entre deux cycles consécutifs de vibrations (Figure 3.4).

$$J = \frac{1}{N-1} \sum_i |P(i) - P(i+1)| \quad (3.3)$$

Le Jitter factor J_f permet de normaliser le Jitter moyen en le comparant à F_0 moyenne. Il est un bon indice pour explorer la stabilité de la fréquence fondamentale.

$$J_f = 100. \frac{J}{\frac{1}{N} \sum_i P(i)} \quad (\%) \quad (3.4)$$

Avec :

$$P_{moy} = \frac{1}{N} \sum_i P(i) \quad (3.5)$$

alors :

$$J_f = 100. \frac{J}{Pmoy} \quad (\%) \quad (3.6)$$

P: Période ;
N: Nombre de Périodes.

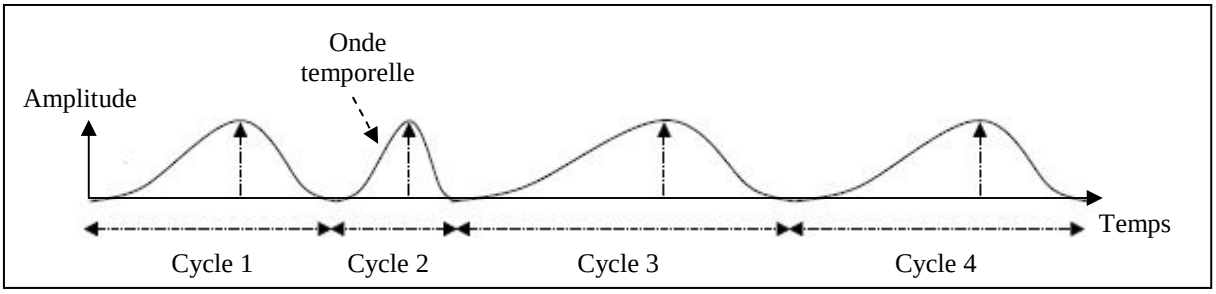


Figure 3.4 : Exemple d'apériodicité de la fréquence fondamentale

En cas de pathologie vocale, le Jitter augmente et une valeur supérieure à 1.04 % correspond à une parole pathologique [26]. En d'autres termes, le seuil normal/ pathologie correspond à 1.04 %.

3.2.2. Perturbation de l'intensité (Shimmer)

Le Shimmer *S* réfère à la variation cycle par cycle de l'intensité dans une trame du signal. Il se calcule par la moyenne de la différence d'amplitude entre deux cycles consécutifs de vibrations (Figure 3.5).

$$S = \frac{1}{N-1} \sum_i |A(i) - A(i+1)| \tag{3.7}$$

Comme pour le Jitter factor J_f, le Shimmer factor S_f est un bon indice pour explorer la stabilité de l'intensité. Il permet de normaliser le Shimmer moyen en le comparant à l'amplitude moyenne.

$$S_f = 100. \frac{S}{\frac{1}{N} \sum_i A(i)} \quad (\%) \tag{3.8}$$

Avec :

$$A_{moy} = \frac{1}{N} \sum_i A(i) \tag{3.9}$$

alors :

$$S_f = 100. \frac{S}{Amoy} \quad (\%) \tag{3.10}$$

Le Shimmer absolu moyen, exprimé en dB, est la moyenne des rapports d'amplitudes entre deux cycles consécutifs de vibrations. De même que pour le Jitter factor, le Shimmer factor relativise le Shimmer moyen et augmente en cas d'anomalie laryngée :

$$S_{dB} = \frac{1}{N-1} \sum_i |20\ log10\ (A(i)) - 20log10(A(i+1))| \tag{3.11}$$

$$S_{dB} = \frac{1}{N-1} \sum_i |20\ log10(A(i)/A(i+1))| \tag{3.12}$$

$$Sf_{dB} = 100. \frac{S}{\frac{1}{N} \sum_i 20Log10(A(i))} \tag{3.13}$$

Avec :

$$A_{dBmoy} = \frac{1}{N} \sum_i 20log10(A(i)) \tag{3.14}$$

alors :

$$S_{fdB} = 100. \frac{S_{dB}}{A_{dBmoy}} \quad (3.15)$$

A(i): Amplitude du pic P_i ;

N: Nombre de pics glottiques.

En cas de pathologie vocale, le Shimmer augmente et une valeur supérieure à 3.81% correspond à une parole pathologique [26]. En décibel, ce seuil correspond à 0.35 dB. Ainsi, le seuil normal/pathologie correspond à 3.81 % et 0.35 dB pour le Shimmer.

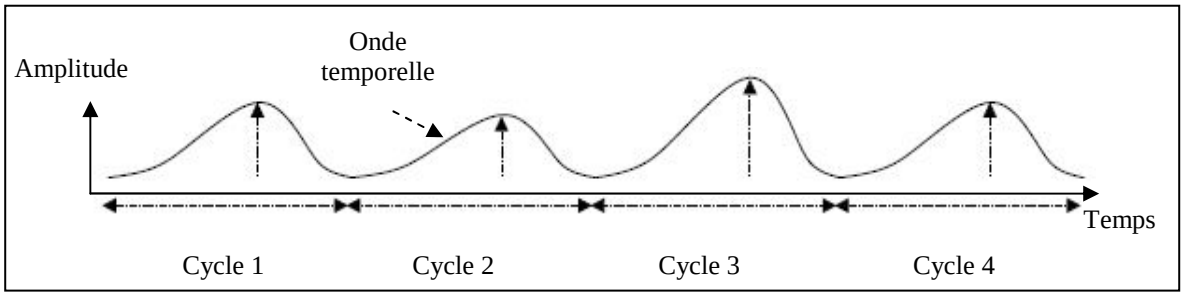


Figure 3.5 : Exemple d'instabilité de l'amplitude du signal

3.2.3. Rapport Harmoniques/Bruit

L'instabilité du signal glottique se manifeste comme un bruit qui lui est superposé. Elle est évaluée au moyen du rapport entre l'énergie harmonique dans le spectre du signal et l'énergie du bruit. Pour cela, nous utilisons la méthode dite "HNR" (Harmonics to Noise Ratio) proposée par Yumoto et al., qui permet de calculer le rapport "énergie des harmoniques / énergie du bruit" [31]. Ce rapport devrait diminuer en cas d'atteinte laryngée. Globalement, le HNR est défini par :

$$HNR = Energie\ (F_0 + Hi) / Energie\ (Bruit) \quad (3.16)$$

$$HNR_{dB} = 10\ log_{10}\ (E_h / E_b) = 10\ log_{10}\ (\frac{\sum_{n=0}^{N-1} (H[n])^2}{\sum_{n=0}^{N-1} (B[n])^2}) \quad (3.17)$$

Avec :

E_h: énergie des composantes harmoniques et périodiques du signal de parole ;

E_b: énergie des composantes apériodiques ou bruit ;

N: taille du signal.

3.2.4. Taux de Passage par Zéro

Le Taux de Passage par Zéro (TPZ) définit le nombre de fois où l'onde temporelle passe par le niveau zéro. Cette caractéristique est fréquemment utilisée pour la classification de la parole voisée/non voisée ou pour identifier les consonnes fricatives. En effet, les brusques variations du TPZ sont significatives de l'alternance voisée/non voisée donc de la présence de parole. Le TPZ est faible pour les zones voisées et très élevé pour les zones non voisées. Du fait de sa nature aléatoire, le bruit possède un TPZ supérieur à celui de la parole voisée.

$$TPZ = \frac{1}{N} \sum_{n=1}^{N-1} \begin{cases} 1 \ \ si \ (x_{n+1}.x_n \ < 0) \\ 0 \ \ sinon \end{cases} \tag{3.18}$$

En cas d'atteinte laryngée ou d'une parole bruitée, le TPZ augmente sensiblement par rapport à une parole normale. En pathologie vocale, nous utilisons le paramètre acoustique relatif au taux des trames non voisées contenues dans le signal DUF (Degree of Unvoiced Frames). Plus la valeur de ce paramètre est importante, plus le signal est bruité, par conséquent la parole est perturbée.

3.3. Analyse acoustique de la Parole Pathologique

L'analyse acoustique vise à extraire un ensemble de paramètres qui permettent d'expliquer et de caractériser l'ensemble des phénomènes physiologiques ou articulatoires intervenant dans la production de l'acte de parole. La recherche de techniques objectives d'analyse de la Parole Pathologique (PP_{ath}) certes difficile. En premier lieu parce que c'est toujours par notre oreille que nous entendons naturellement et analysons la production vocale. L'analyse perceptive reste donc la référence, bien qu'elle soit subjective, en évaluant le trouble vocal comme un tout. Aujourd'hui avec le développement d'outils et de logiciels informatiques assez performants, l'analyse acoustique tente d'apporter cette objectivité qui a toujours fait défaut dans la caractérisation de la voix et de la parole humaines. Ceci est d'autant plus important car cette analyse nous offre des données concrètes qui nous permettent :

- de caractériser objectivement la voix et la parole ;
- d'estimer le degré d'une éventuelle détérioration par rapport à la norme ;
- d'apporter ainsi les solutions nécessaires pour y remédier.

Dans le cadre de notre travail, nous avons appliqué une analyse acoustique sur deux types de pathologies : l'une d'origine fonctionnelle (Parole Parkinsonienne) et l'autre d'origine organique (Parole Œsophagienne).

3.3.1. Corpus d'analyse de la PP_{ath}

Le corpus choisi contient un ensemble de phrases, mots, syllabes et voyelles à l'état isolé, extraits du corpus conçu par le Professeur N. Zellal pour l'analyse des PP_{ath} en orthophonie [34]. L'avantage de ce corpus est qu'il tient compte essentiellement des consonnes spécifiques de la langue Arabe (emphatiques et postérieures) dans les différents contextes. Il est conçu, en grande partie, en Arabe Algérois. Ce qui est indiqué dans le cadre de notre travail, car les patients avec lesquels nous avons travaillé sont tous arabophones et utilisent l'Arabe Dialectale Algérois. Nous avons utilisé pour l'enregistrement le sonagraphe Kay 4300B, qui a la particularité d'éliminer, au maximum possible, les bruits environnants [26]. Nous avons donc obtenu des enregistrements moins bruités et d'assez bonne qualité. Nous avons également réalisé quelques enregistrements au niveau de l'hôpital de Beni-Messous (Alger), en utilisant le logiciel d'analyse acoustique Praat [35].

- **Parole Continue**

  Sourate El Ikhlas سورة الإخلاص

- **Phrases**

  [fatiḥa talƐab] (فتيحة تلعب)
  [ṣbāḥ lxīR] (صباح الخير)
  [ṭefla tbīƐ lḫubz] (طفلة تبيع الخبز)
  [ṭiflun saɣīR] (طفل صغير)
  [kāɣaṭ ɣāli] (كاغط غالي)

- **Mots**

  [maɣlūqa] (مغلوقة)
  [maḥRūqa] (محروقة)
  [maḫlūqa] (مخلوقة)
  [Ɛād] (عاد)
  [qād] (قاد)

- **Paires [CV]**

[ba] / [ta]	(بَ) / (تَ)
[qa] / [ga] / [ka]	(قَ) / (ڤَ) / (كَ)
[ʕa] / [ħa]	(عَ) / (حَ)
[ɣa] / [ẖa]	(غَ) / (خَ)
[ʔa] / [ha]	(أَ) / (هَ)

- **Consonnes emphatiques**

[ṭa] / [ṭi] / [ṭu]	(طَ) / (طِ) / (طُ)
[ḍa] / [ḍi] / [ḍu]	(ضَ) / (ضِ) / (ضُ)
[ṣa] / [ṣi] / [ṣu]	(صَ) / (صِ) / (صُ)
[ḏa] / [ḏi] / [ḏu]	(ظَ) / (ظِ) / (ظُ)
[ṣbāħ] / [ṣbāʕ]	(صباح) / (سباع)
[ṭāba] / [tāba]	(طَابَ) / (تَابَ)
[ṣabb] / [sabb]	(صبّ) / (سبّ)

- **Voyelles soutenues**

  [ā], [ī] et [ū]

3.3.2. Analyse acoustique de la Parole Parkinsonienne (PP_{ark})

En addition aux paramètres acoustiques ordinaires tels que la fréquence fondamentale, les formants et l'énergie (intensité), d'autres indices sont également mis en valeur, car ils permettent de discriminer la PP_{ath} de la PN_{orm}. Ces indices acoustiques sont le Jitter, le Shimmer, le HNR et le pourcentage de trames non voisées DUF (Degree of Unvoiced Frames). Le dépassement du seuil normal/ pathologique, pour un paramètre donné, est un moyen objectif qui permet de mettre en évidence les altérations vocales et de mesurer le degré de cette altération [32, 33].

3.3.2.1. Enregistrements du corpus parkinsonien

Cinq locuteurs parkinsoniens et trois locuteurs normaux ont participé aux expériences acoustiques. Quatre patients parkinsoniens n'ont pas bénéficié d'une rééducation par un orthophoniste (Tableau 3.1). Un seul patient a bénéficié d'une

rééducation (sessions pendant 2 ans). En outre, tous les patients présentent des troubles du langage à divers degrés.

Tableau 3.1 : Présentation des patients parkinsoniens

Patients	Age	Rééducation Oui(+)/Non(-)	Début de la maladie
1	82	-	2005
2	74	-	2004
3	68	-	2003
4	68	-	2001
5	61	+ (2 ans)	1992

3.3.2.2. Extraction des paramètres acoustiques

Nous illustrons notre analyse à l'aide d'une comparaison entre une parole normale de référence et une parole pathologique de deux patients dont le 1er n'a pas subi de rééducation orthophonique donc Non Pris en Charge (NPC) et le second Pris en Charge (PC) à travers des séances périodiques de rééducation (Tableaux 3.2, 3.3 et 3.4).

Tableau 3.2 : Paramètres acoustiques de la norme de référence

	Paramètres acoustiques					
	Durée (ms)	Intensité (dB)	F_0 (Hz)	Formants (Hz)		
Corpus				F1	F2	F3
[ba]	0.170	69	176	658	1965	2911
[ta]	0.184	71	165	622	1883	2813
[maħRūqa]	0.865	62	182	599	1498	2539
[fatiḥạ talɛab]	1.010	62	223	574	1898	2928
[ṣbāħ lxīR]	0.930	66	164	510	1641	2690

Tableau 3.3 : Valeurs des Paramètres acoustiques, Cas NPC

	Paramètres acoustiques					
	Durée (ms)	Intensité (dB)	F_0 (Hz)	Formants (Hz)		
Corpus				F1	F2	F3
[ba]	0.310	62	189	712	1839	2846
[ta]	0.385	64	175	784	1778	2898
[maħRūqa]	1.078	56	156	663	1327	2672
[fatiḥạ talɛab]	1.621	54	240	663	1641	2450
[ṣbāħ lxīR]	1.202	61	189	581	1531	2423

Tableau 3.4 : Valeurs des Paramètres acoustiques, Cas PC

	Paramètres acoustiques					
	Durée (ms)	Intensité (dB)	F_0 (Hz)	Formants (Hz)		
Corpus				F1	F2	F3
[ba]	0.206	67	311	503	1506	2536
[ta]	0.148	71	297	567	1701	2661
[maħRūqa]	0.607	62	306	570	1442	2373
[fatiħạ talɛab]	0.806	62	298	529	1701	2465
[ṣbāħ lxīR]	0.700	68	340	352	1447	2560

Nous avons comparé également les paramètres acoustiques pathologiques HNR (dB), Shimmer (%), Jitter (%) et DUF (%) pour les cas PC et NPC, par rapport à la norme de référence (Tableau 3.5).

Tableau 3.5 : Comparaison des paramètres acoustiques HNR, Shimmer, Jitter, DUF, Parole normale et pathologiques

	Paramètres acoustiques				
Cas	HNR (dB)	Shimmer (%)	Jitter (%)	Trames non voisées	DUF (%)
Normal	15.37	4.36	1.77	3.50	
NPC	07.19	12.57	3.04	7.14	
PC	13.03	6.02	2.64	8.21	

Nous avons relevé les sonagrammes des paramètres acoustiques (fréquence fondamentale, intensité, durée et formants), ainsi que les paramètres qui nous permettront de discriminer les différents cas choisis (figures 3.6 et 3.7).

Nous remarquons une perturbation assez perceptible de ces paramètres pour le cas NPC, avec notamment un niveau d'intensité faible de la voyelle [a]. Par contre, nous relevons une amélioration de ces paramètres pour le cas PC, avec particulièrement un niveau d'intensité plus élevé et une courbe de la fréquence fondamentale continue et sans altération ou discontinuité mais à un niveau plus aigu.

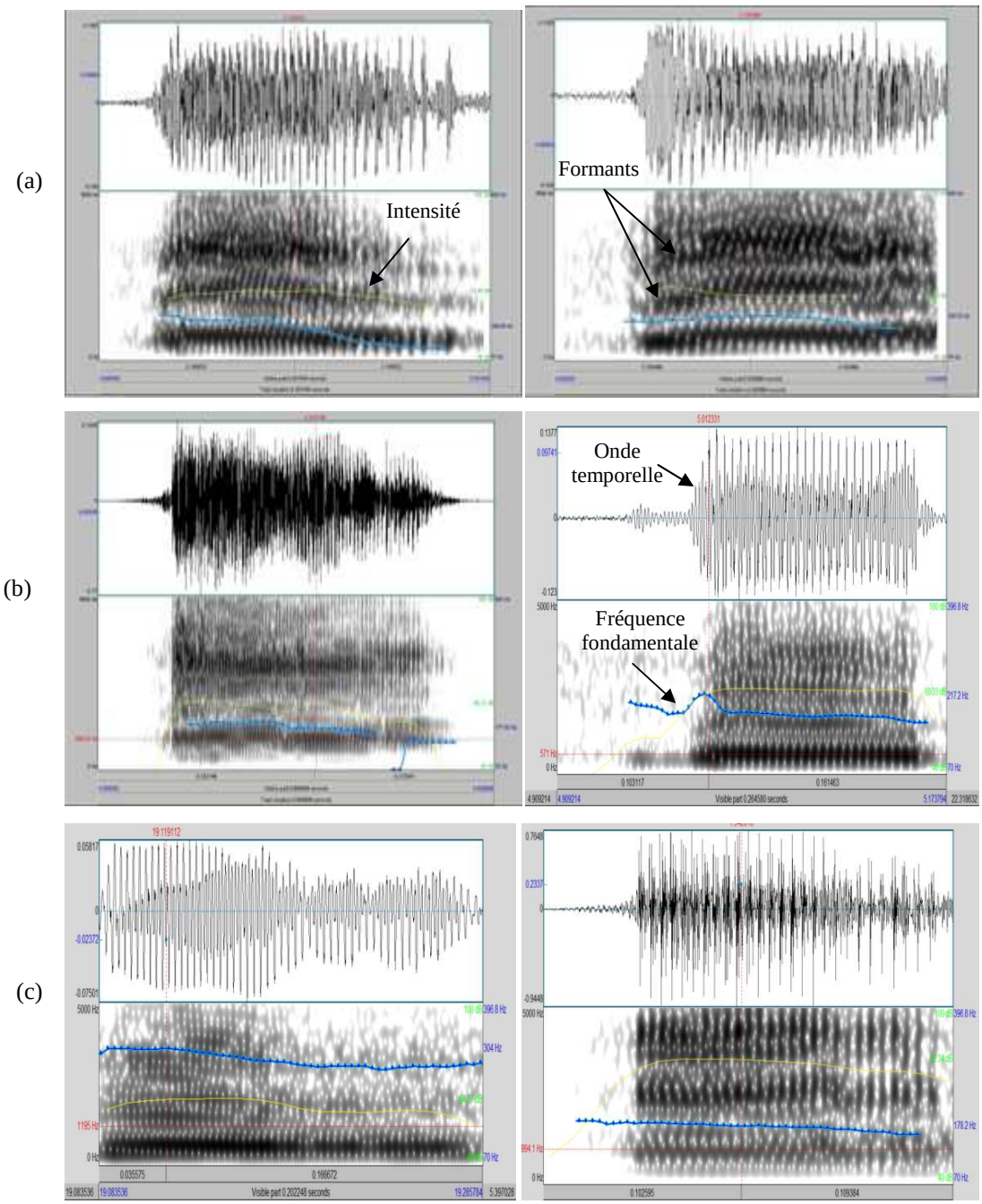


Figure 3.6 : Prononciation de la voyelle [a] et de la syllabe [ba],
(a) Cas normal, (b) Cas NPC et (c) Cas PC

En parole continue, nous remarquons une forte perturbation de la courbe de la fréquence fondamentale pour le cas NPC (Figure 3.7 b), alors que cette courbe montre une nette amélioration pour le cas PC (Figure 3.7 c). Ceci montre que la prise en charge orthophonique des patients parkinsoniens améliore sensiblement leurs voix.

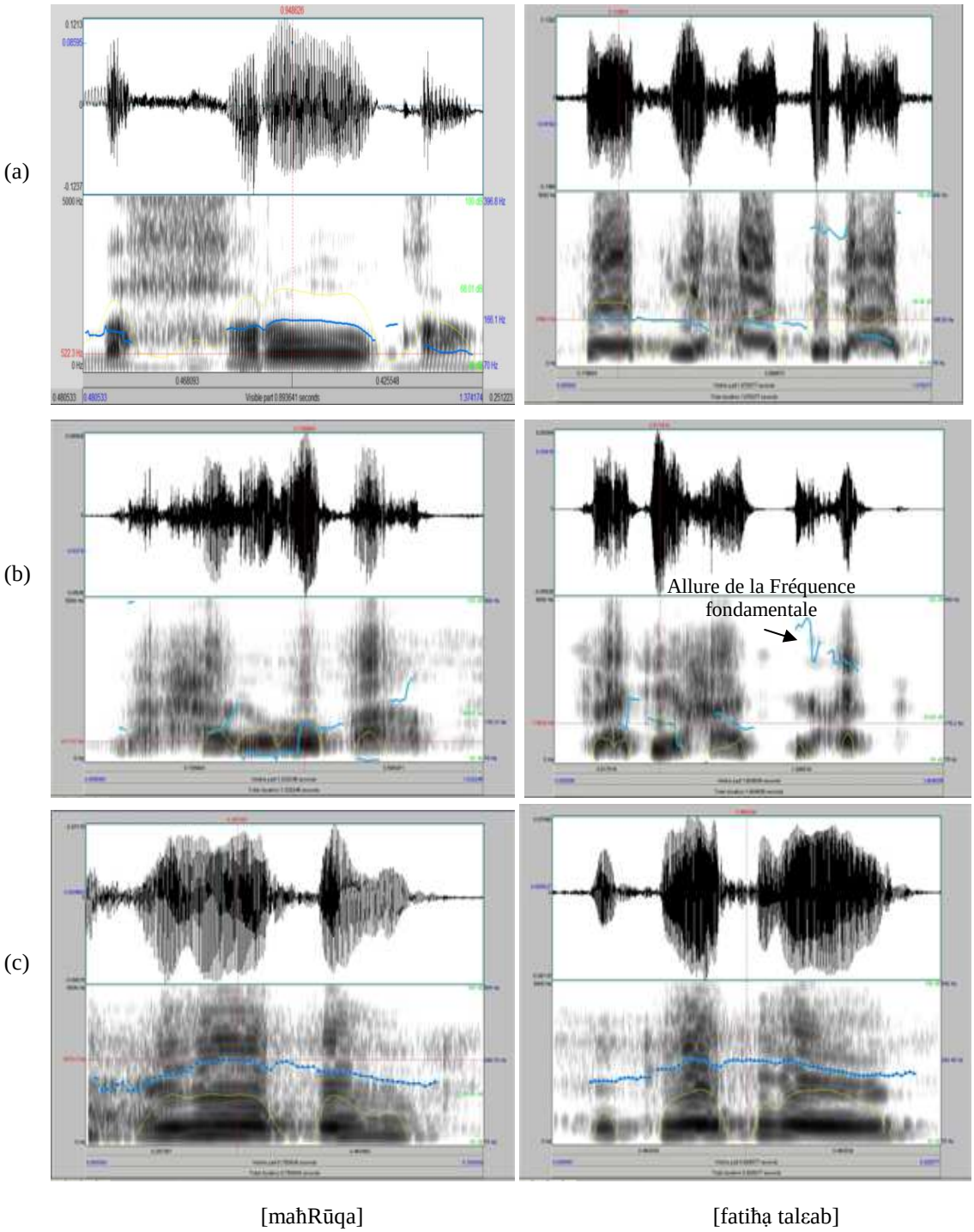


Figure 3.7 : Prononciation du mot [maħRūqa] et de la phrase [fatiħa talɛab]
(a) Cas normal (b) Cas NPC et (c) Cas PC

L'analyse acoustique des enregistrements des cas NPC nous montre :

- une valeur de F_0 légèrement élevé (240 Hz), comparée au cas normal. Nous pouvons considérer tout de même F_0 du cas pathologique dans les normes ;

- une valeur de F_1 plus importante par rapport au cas normal. Par contre, F_2 est moins élevé. Les valeurs de F_3 sont dans la majorité des cas moins élevées comparées à la norme de référence. L'altération des valeurs des formants montre un manque de synchronisation et de maîtrise des organes liés à la parole ;
- une valeur du Jitter (3.04 %) qui montre une forte perturbation de la vibration des cordes vocales ;
- une intensité faible et monotone de la voix du patient, avec une fuite d'air à travers les cordes vocales. De même, le tableau 3.5 montre une valeur du Shimmer très importante comparée à la norme (12.57 % contre 4.36 % pour le cas normal) ;
- une durée de prononciation des sons plus grande par rapport à la normale. Le patient prend beaucoup de temps pour prononcer les mots. Nous notons des pauses de durées assez importantes entre les mots et à l'intérieur des mots ;
- des valeurs de DUF révélant une fuite d'air assez importante lors de l'accolement partielle des cordes vocales (7.14 %, alors qu'elle n'est que de 3.50 % pour le cas normal). Ainsi, La prononciation des sons est fortement bruitée et le timbre de la PP_{ath} est sombre et voilé (influence importante du bruit sur les harmoniques) ;
- une élision de quelques sons lors de la prononciation. Exemple : le son fricatif emphatique sourd [ṣ] n'a pas été prononcé dans la phrase [ṣbāħ lxīR] ;
- une altération de la prononciation des consonnes occlusives sourdes, avec une confusion entre les sons non voisés [ta] et [ka] ;
- une lenteur assez perceptive dans le discours. Voix monotone, faible et voilée.

L'analyse acoustique des enregistrements du cas PC nous montre :

- une voix très aigue ($F_0 \geq 297$ Hz) comparée au cas NPC ;
- un niveau formantique plus bas. Le patient tend à emphatiser sa prononciation ;
- un niveau d'intensité plus proche de la normale. Cette amélioration est probablement due à la prise en charge orthophonique contrairement au cas NPC. De même, la valeur du Shimmer est proche de la normale (6.02 % contre 4.36 % pour le cas normal) ;
- une durée de prononciation de discours plus rapide. Nous sentons que le patient tend à terminer le plus vite possible son discours pour économiser ses efforts. Nous avons relevé des pauses entre les mots de durées proches du normal et une absence de

pauses à l'intérieur des mots en ce sens que ces derniers ne sont pas prononcés hachurés, comme pour les NPC ;

- des valeurs de DUF révélant une fuite d'air assez importante lors de l'accolement des cordes vocales, comme pour le cas NPC. Ce qui fait que les sons sont prononcées très bruités (8.21 % alors qu'elle n'est que de 3.50 % pour le cas normal) ;
- une perturbation moins importante de la vibration des cordes vocales (un Jitter de 2.64 % contre 3.04 % pour les cas NPC). Nous notons donc un timbre moins sombre. L'influence du bruit est moins perceptible car le patient tend à mieux maitriser la vibration de ses cordes vocales. Ceci est important à noter sachant que le patient est à un stade assez avancé de la maladie (début 1992) alors que le début de la maladie pour les autres cas se situe entre 2001 et 2005 ;
- une absence d'alternances entre sons aigus et sons graves. La voix est généralement très aigue ;
- une absence d'élision de sons lors de la prononciation ;
- une absence de perturbation lors de la prononciation des consonnes occlusives sourdes. Inexistence des confusions relevées pour les cas NPC ;
- une prononciation de discours plus rapide. Voix monotone, faible et voilée.

3.3.2.3. Interprétation des résultats

Les résultats trouvés nous montrent que les patients parkinsoniens présentent des troubles de la voix et de la parole assez importants. Nous avons relevé, entre autres, une confusion entre les sons, une élision de sons de parole lors du discours, une voix bruitée, faible et monotone, des valeurs de formants assez perturbées et donc un timbre sombre et voilé. Le patient ayant bénéficié d'une prise en charge (PC) présente moins de perturbation des paramètres acoustiques de la parole. Très peu de confusions entre les sons, timbre moins voilé et voix à intensité proche de la normale. Les résultats sont certes insuffisants comparés au cas normal, mais il y a une amélioration de la prosodie et de l'intelligibilité de la parole comparées à celles de ceux qui n'ont pas bénéficié d'une prise en charge orthophonique (NPC). Une comparaison de l'analyse acoustique du cas PC et des cas NPC montre l'importance de la prise en charge orthophonique des patients parkinsoniens. Une prise en charge qui fait défaut dans les hôpitaux algériens du fait que dans la majorité des cas, seule une prise en charge neurologique est assurée.

3.3.3. Analyse acoustique de la Parole Œsophagienne ($PŒ_{so}$)

L'analyse acoustique a été réalisée au moyen des logiciels de programmation Matlab et d'analyse acoustique Praat. Cette analyse concerne la fréquence fondamentale F_0 (Hz) avec l'écart-type et les valeurs extrêmes (max F_0 et min F_0), les formants (F_1, F_2, F_3), le Jitter (%), le Shimmer (% et dB), le HNR (dB), le degré de trames non voisées DUF (%) et l'énergie (dB). Le logiciel Praat a été utilisé afin d'extraire les paramètres suivants: F_0, F_1, F_2, F_3, DUF et HNR. Le Jitter, le Shimmer et l'énergie ont été extraits à partir de fonctions conçues sous Matlab 2007.

3.3.3.1. Enregistrements du corpus œsophagien

Nous avons enregistré le corpus au Laboratoire de Traitement Automatique de la Parole du CRSTDLA (Alger) entre Octobre 2008 et Septembre 2009. Les enregistrements ont été effectués avant le début de la rééducation, et après trois, six et onze mois de rééducation. L'intérêt de l'enregistrement de la parole, juste après chirurgie et avant rééducation, est de confirmer effectivement l'absence de composantes sonores, en mesurant précisément le degré de trames non voisées DUF (%), au cours de la prononciation des voyelles soutenues.

Les sons du corpus ont été prononcés par huit patients qui ont subi une laryngectomie totale, dont l'âge variant entre 47 à 59 ans, l'âge moyen étant de 55 ans. Tous les sujets ont bénéficié d'une rééducation vocale par un orthophoniste (sessions durant 11 mois). Ils ont appris la méthode de la $PŒ_{so}$ par la technique d'inhalation et ensuite d'éructation volontaire de sons à travers l'œsophage [19, 20]. Tous les sujets sont des hommes et communiquent en Arabe Dialectal Algérien (Algérois). Le même corpus a été prononcé par trois locuteurs normaux ne présentant aucune pathologie de la voix et de la parole.

3.3.3.2. Extraction des paramètres acoustiques

Un exemple de paramètres acoustiques extraits sur la voyelle soutenue [ā], prononcée par les patients avant et après 3, 6 et 11 mois de rééducation, est donné dans le Tableau 3.6.

Tableau 3.6 : Valeurs des paramètres acoustiques de la voyelle [ā], PŒ$_{so}$

Voyelle [ā]	**Normal**	**Avant Rééducation**	**3 mois**	**6 mois**	**11 mois**
Jitter (%)	00.247	-	12.271	6.013	01.745
Shimmer (%)	03.410	-	07.891	7.910	10.588
Shimmer (dB)	00.297	-	01.119	1.114	00.681
HNR (dB)	20.748	-	01.710	2.154	03.414
DUF (%)	00.000	88.57	46.930	41.370	13.115
Pitch F_0 (Hz)	129.36	-	061.15	062.39	104.01
F_0 Moyen (Hz)	129.08	-	060.42	068.74	099.03
Ecart-type F_0 (Hz)	0.911	-	2.144	15.140	9.955
F_{0min} (Hz)	127.92	-	059.26	049.97	094.41
F_{0max} (Hz)	131.41	-	064.73	102.73	113.46
F_1 (Hz)	646	1089	1099	768	970
F_2 (Hz)	1038	1590	1769	1870	1680
F_3 (Hz)	2806	3100	3311	3216	3051
Intensité (dB)	73.13	57.90	59.30	56.09	58.62

Une comparaison entre la prononciation des voyelles à l'état normal, avant rééducation et après 11 mois de rééducation est illustrée par les figures 3.8, 3.9 et 3.10. Nous remarquons une absence totale de voisement avant rééducation (les voyelles sont prononcées comme bruits) et une présence d'un voisement néoglottale après une période de 11 mois de rééducation.

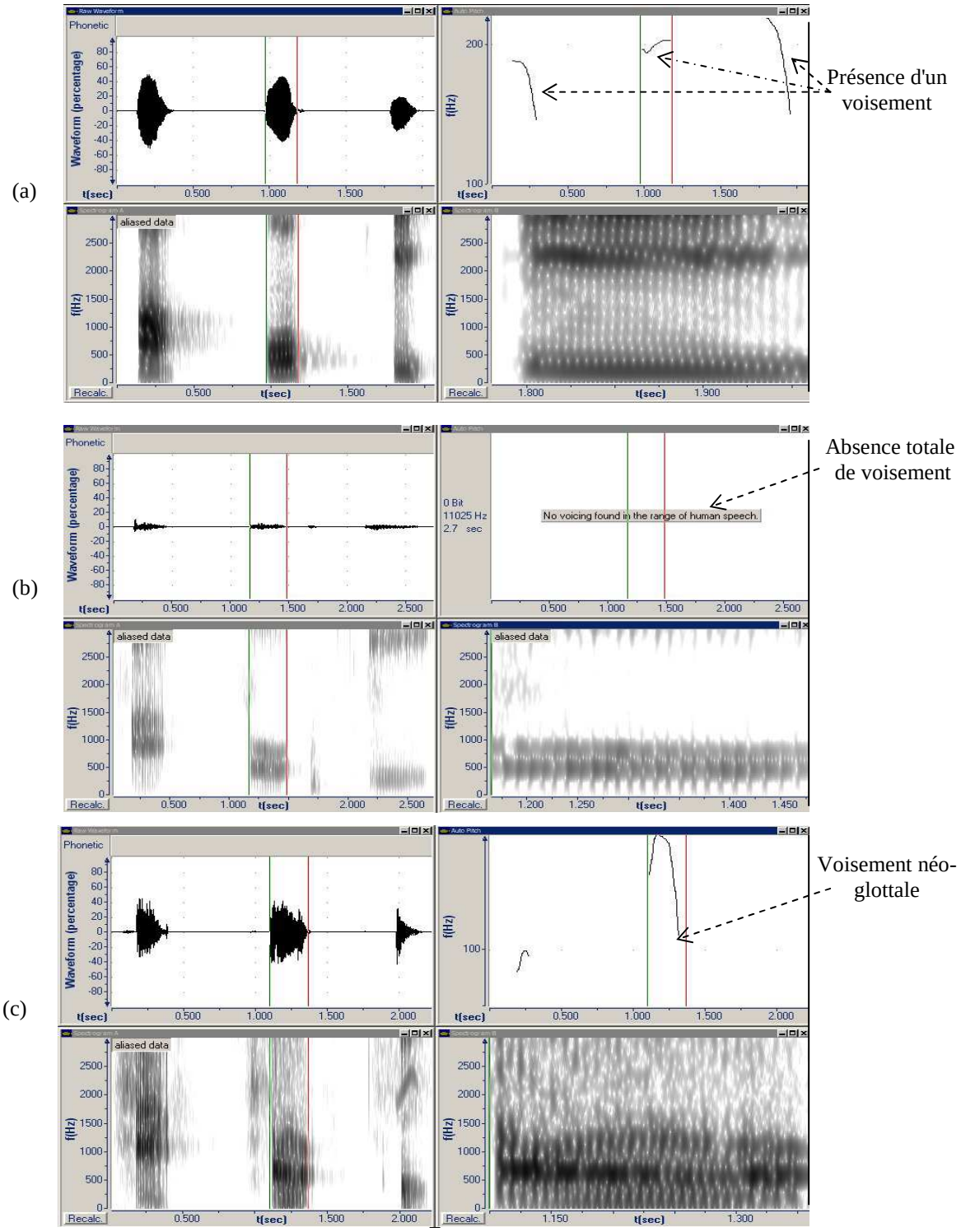


Figure 3.8 : Indice de voisement
a) Etat normal, (b) et (c) Avant et après 11 mois de rééducation $P\text{Œ}_{so}$

Une étude, à partir du logiciel d'analyse acoustique Praat, montre une richesse en harmoniques (timbre clair) pour le cas d'une voix normale, mais une absence totale

d'harmoniques avant rééducation et une faiblesse en harmoniques après une période de rééducation (Figure 3.9).

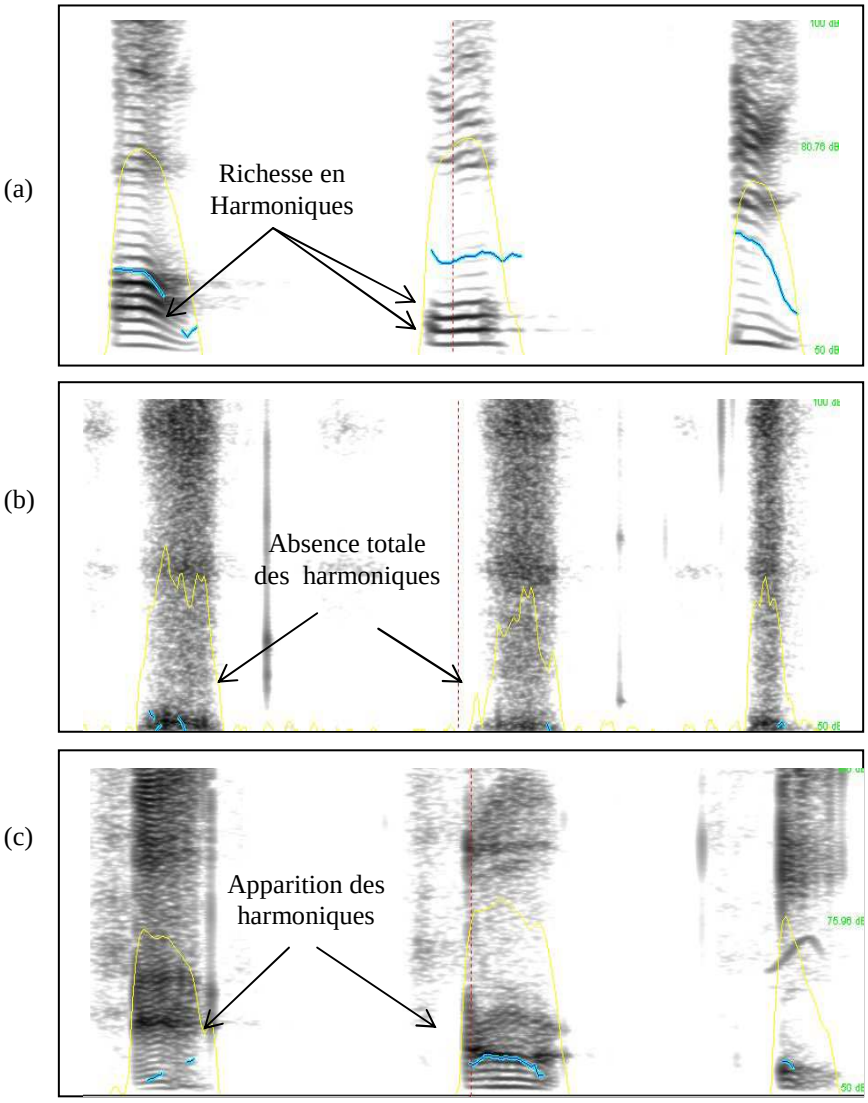


Figure 3.9 : Harmoniques des voyelles [a], [u], [i],
(a) Etat normal, (b) et (c) Avant et après 11 mois de rééducation $P\text{Œ}_{so}$

L'analyse acoustique des enregistrements du corpus prononcé par les patients laryngectomisés montre que :

- après 3 mois de rééducation, un voisement commence à apparaître (F_0 > 60 Hz). Nous constatons néanmoins une restriction générale du pitch (trames non voisées > 40 % après six mois de rééducation). Ainsi, la $P\text{Œ}_{so}$ est caractérisée par une valeur moyenne du pitch très faible. La gamme de valeurs du Jitter a tendance à diminuer au cours de la rééducation et nous notons une évolution

certes lente, mais perceptible des valeurs du Shimmer après six mois de rééducation.

- le HNR reste relativement faible par rapport à la normale. Les valeurs relevées pour le HNR pathologique sont encore loin du seuil normal. En outre, nous notons une augmentation des valeurs des formants F_1, F_2 et F_3 après ablation du larynx (figure 3.10), et des valeurs moins importantes de l'intensité. Le timbre vocal qui présente une réduction significative des composantes harmoniques commence à devenir riche après une période de 11 mois de rééducation. Néanmoins, la richesse en harmoniques est toujours loin du seuil de normalité.

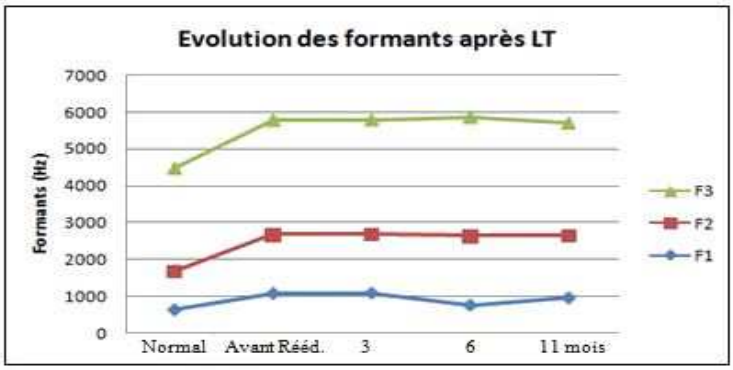


Figure 3.10 : Montée des formants de la voyelle [a] après Laryngectomie Totale

La figure 3.11 montre un exemple de confusion dans la prononciation de la consonne pharyngale voisée [ɛ] comme une pharyngale sourde [ħ].

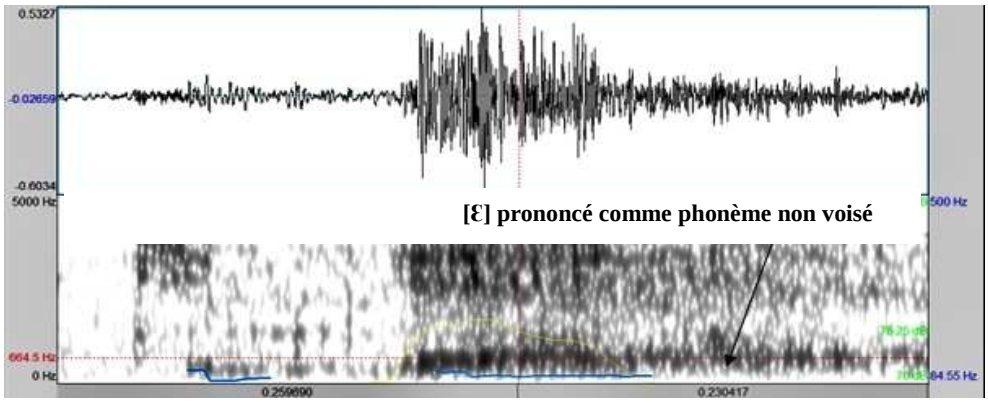


Figure 3.11 : Prononciation du [ɛ] dans le mot [tbīɛ] en $PŒ_{so}$

Dans la figure 3.12, ci-dessous, nous remarquons également une confusion dans la prononciation des consonnes postérieures, avec la consonne uvulaire voisée [ɣ] prononcée comme uvulaire sourde [x].

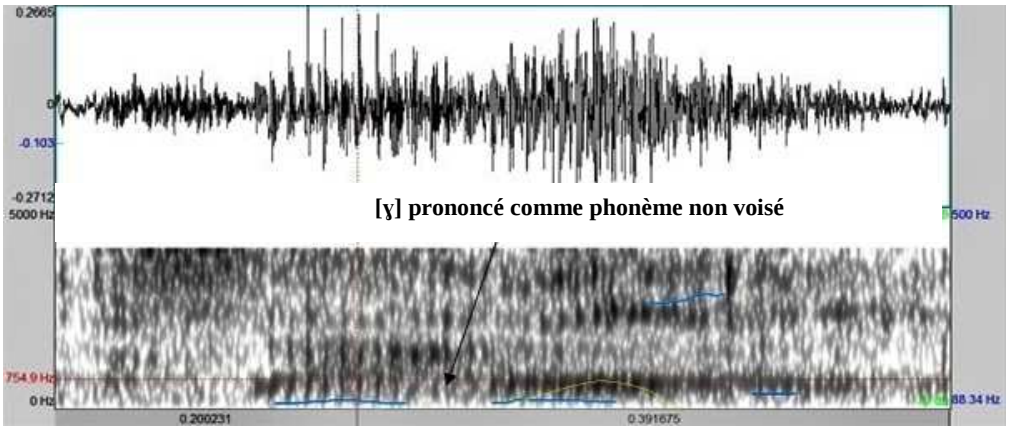


Figure 3.12 : Prononciation du [ɣ] dans le mot [saɣĩR] en $PŒ_{so}$

L'analyse des sonagrammes des sons prononcés par les patients montrent (figures 3.11 et 3.12) :

- une réelle confusion entre certaines consonnes ayant le même point d'articulation, mais diffèrent par l'indice voisé/non voisé. En conséquence, la consonne pharyngale voisée [ε] se prononce comme pharyngale sourde [ħ], et la consonne uvulaire voisée [ɣ] se prononce comme uvulaire sourde [x]. Pour les consonnes vélaires et dentales, cinq des huit patients sont capables de distinguer les oppositions entre consonnes voisées/non voisées, ayant le même point d'articulation, de façon correcte après onze mois de rééducation.
- de façon générale, nous avons relevé une difficulté de prononciation des sons emphatiques de l'AS. Ces sons sont soit ignorés, soit prononcés amputés du trait de l'emphase, c'est-à-dire [ṣ] comme [s], [ṭ] comme [t].
- au début de la rééducation, la parole œsophagienne est inintelligible. Elle se caractérise par plusieurs pauses dans les phrases longues (comme celles nécessaires pour l'injection d'air). Néanmoins, la parole devient plus fluide et moins tendue après rééducation. En outre, elle devient assez intelligible après onze mois de rééducation [19, 33].

3.3.3.3. Interprétation des résultats

L'interprétation des résultats obtenus lors de l'analyse acoustique nous montre :

- une faible capacité de voisement ($F_0 \leq 100$ Hz) qui peut être expliquée par la forme, le volume et l'élasticité du segment néo vibrateur NVP, qui diffèrent totalement de ceux des cordes vocales. Ce segment semble assez instable et pas

toujours périodique, en raison des caractéristiques anatomiques des structures vibrantes. Nous notons que ces résultats sont en accord avec quelques études précédentes rapportées dans d'autres langues [36, 37] ;

- des valeurs du Jitter et du Shimmer tendant vers des valeurs de seuil normal. Ceci est probablement dû au fait que le patient est en mesure de mieux contrôler ses efforts par une meilleure connaissance des divers organes de la phonation de son nouveau mécanisme de production de la parole ;
- une voix pathologique tellement bruitée au début, que le patient aura probablement besoin de plus de temps pour acquérir des valeurs HNR proches du seuil de normalité ;
- une augmentation des valeurs des formants qui peut être probablement expliquée par le fait que la distance entre le segment NVP et la première cavité de l'appareil vocal (cavité pharyngale) est modifiée par la présence du trachéostome. Ces résultats sont en accord avec ceux rapportés antérieurement dans d'autres langues [21, 36-38] ;
- une énergie moins importante après laryngectomie car la quantité d'air obtenue par éructation reste insuffisante (moins de 70 ml) par rapport à celle qui résulte des poumons dans la parole normale laryngée (environ 5000 ml) ;
- une confusion entre certaines consonnes, en accord avec des études précédentes rapportées dans la littérature concernant la laryngectomie [39, 40], contrairement à d'autres études qui rapportent un taux de confusion négligeable [41]. Dans notre étude, nous notons une confusion perceptible des consonnes postérieures ([ɛ], [ħ], [ɣ] et [x]) contrairement aux consonnes antérieures ([g], [k], [d], [t]). Ceci est très important à noter car les tests de rééducation orthophoniques exploités dans les hôpitaux algériens sont importés de France et donc plus adaptés aux consonnes antérieures car la langue française présente peu de consonnes postérieures. Cette confusion des consonnes postérieures peut être expliquée par le fait qu'à la différence des cordes vocales qui doivent être en adduction pour commencer à vibrer, le segment NVP doit être en position relaxe pour assumer les vibrations. Ainsi, avec la nouvelle configuration du conduit vocal, et par la présence d'une trachéostomie au niveau du cou, les patients semblent avoir des difficultés à prononcer correctement les consonnes pharyngales et uvulaires avec une relaxation du

segment contigu NVP. Il semble que les patients privilégient l'articulation au détriment du voisement ;

- une difficulté de prononciation des sons emphatiques de l'AS. Ces sons sont soit ignorés, soit prononcés amputés du trait de l'emphase c'est-à-dire [ṣ] comme [s], [ṭ] comme [t]. Ceci peut être expliqué par le phénomène d'emphase lui-même. Avec les changements subis par la cavité pharyngale et la présence du trachéostome (ouverture pour la respiration), le trait physiologique de pharyngalisation, nécessaire pour emphatiser les sons, est assez difficile à réaliser ;
- les patients sont unanimes à noter que la voyelle [i] est très difficile à prononcer, en particulier dans une parole continue. La voyelle antérieure haute [i] est articulée avec le dos de la langue soulevé et se rapprochant de la région alvéolaire. En conséquence de cette poussée vers l'avant de la masse de la langue, la partie inférieure du pharynx est élargie au cours de prononciation de la voyelle [i]. Avec les transformations subies par cette partie de l'appareil vocal après Laryngectomie Totale, il est possible que ce soient les contraintes articulatoires du [i] qui font que cette voyelle est plus difficile à prononcer que les autres voyelles [42].

3.4. Conclusion

Une analyse acoustique a été réalisée sur deux catégories de pathologies vocales : la parole parkinsonienne et la parole œsophagienne. Cette analyse a pour objectif essentiel la caractérisation physico-acoustique des PP_{ath}, permettant une meilleure connaissance des paramètres pertinents de discrimination des deux paroles parkinsonienne et œsophagienne. En effet, l'extraction de paramètres discriminants joue un rôle très important dans la fiabilité d'un système de reconnaissance automatique de PP_{ath}. Elle permet notamment de représenter fidèlement le signal de parole par un choix adéquat de la dimension et de la nature des vecteurs d'entrée du système de reconnaissance.

CHAPITRE 4

RESEAUX DE NEURONES ARTIFICIELS

4.1. Introduction

Ces dernières années, les modèles connexionnistes ou Réseaux de Neurones Artificiels (RNA) occupent une place notable parmi les nombreux modèles proposés pour résoudre le problème de la Reconnaissance Automatique de la Parole (RAP). L'idée principale de ces modèles est de s'inspirer de l'organisation des neurones biologiques humains et leurs interconnexions pour traiter l'information [43-46]. Les travaux effectués pour essayer de comprendre le comportement du cerveau humain ont mené à représenter celui-ci par un ensemble de composants structurels appelés neurones, massivement interconnectés entre eux. Le cerveau arrive à organiser ces neurones selon un assemblage complexe, non-linéaire et extrêmement parallèle, de manière à pouvoir accomplir des tâches très élaborées et très complexes. C'est la tentative de donner à l'ordinateur cette qualité d'organisation des neurones du cerveau humain qui a conduit à une modélisation électrique de celui-ci. C'est cette modélisation complexe que tentent de réaliser les RNA.

Ce chapitre s'articule essentiellement autour des principes de base des RNA, les techniques d'apprentissage existantes, avec une description détaillée de celles exploitées dans le cadre de notre travail. A la fin, nous avons présenté deux types de RNA (Perceptron Multi Couches et Réseaux à décalages temporels) que nous avons utilisé pour la classification automatique des Paroles Pathologiques.

4.2. Neurone biologique et neurone formel

De par leurs multiples interconnexions, leur mécanisme d'inhibition et d'activation, leur manière d'évoluer et de s'adapter tout au long de la vie d'un organisme vivant, les réseaux de neurones biologiques ont inspiré les RNA et continuent d'influencer le développement de nouveaux modèles d'applications en reconnaissance des formes (caractères, visages, images, parole, etc.).

4.2.1. Neurone biologique

La physiologie du cerveau montre que celui-ci est constitué de milliards de cellules interconnectées, appelées neurones. Chaque neurone pouvant recevoir les entrées (signaux sous forme d'impulsions électriques) de dizaines ou parfois de centaines de milliers d'autres neurones. Nous estimons que l'ensemble du cerveau humain contiendrait de l'ordre du million de milliard de synapses, ramifications de neurones permettant l'échange d'informations avec d'autres neurones adjacents (Figure 4.1). Ce

grand nombre de neurones et de connexions conduit à un enchevêtrement qui est, aujourd'hui encore, très difficile à cerner. Notons que la durée de chaque impulsion du signal transmis est de l'ordre de 1 ms et son amplitude d'environ 100 mV.

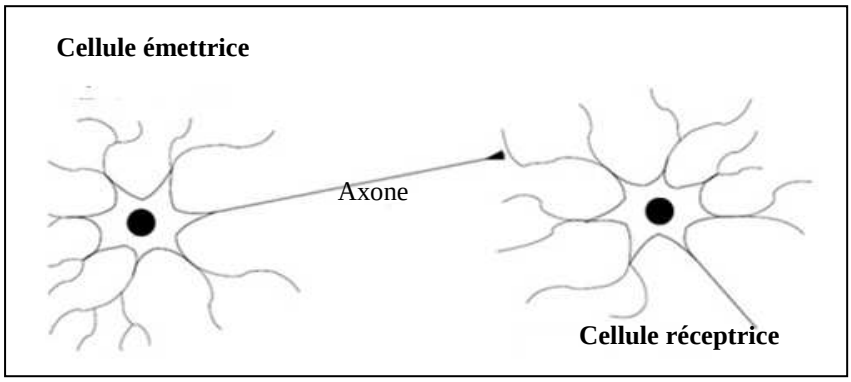


Figure 4.1 : Schéma simplifié d'une connexion entre deux neurones biologiques

La principale caractéristique des neurones biologiques est qu'ils permettent de véhiculer et de traiter des informations. Cette collecte de l'information est effectuée par les **dendrites** du neurone qui réceptionnent l'information des unités afférentes. Cette information est acheminée vers le noyau, également appelé soma. Une fois traitée, elle est répercutée ensuite en sortie de la cellule vers l'**axone** qui propage cette information vers d'autres cellules à travers des jonctions qui portent le nom de "**synapses**" (figure 4.2).

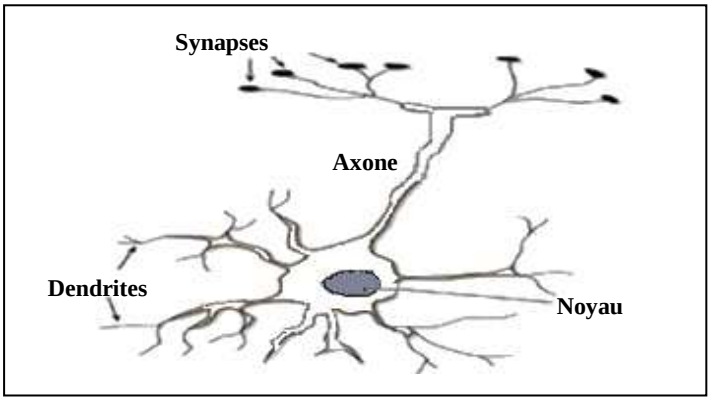


Figure 4.2. Représentation d'un neurone biologique

Les jonctions synaptiques sont de minuscules espaces qui séparent deux neurones et que l'influx nerveux doit franchir. Ce franchissement ne peut se faire que si le premier neurone sécrète une substance que l'on appelle neurotransmetteur, plus connu sous le nom de dopamine, qui sera à son tour reconnu par le neurone suivant, et ainsi de suite.

Une insuffisance de cette substance étant à l'origine de la maladie de Parkinson que nous avons étudié dans le cadre de ce travail.

4.2.2. Neurone formel

Un neurone formel est une représentation mathématique et informatique du neurone biologique, qui tente de reproduire son fonctionnement et son raisonnement intelligent de la meilleure façon possible [43-46]. Ainsi, les principales structures du neurone artificiel ont pratiquement toutes leurs équivalentes biologiques (Figure 4.3).

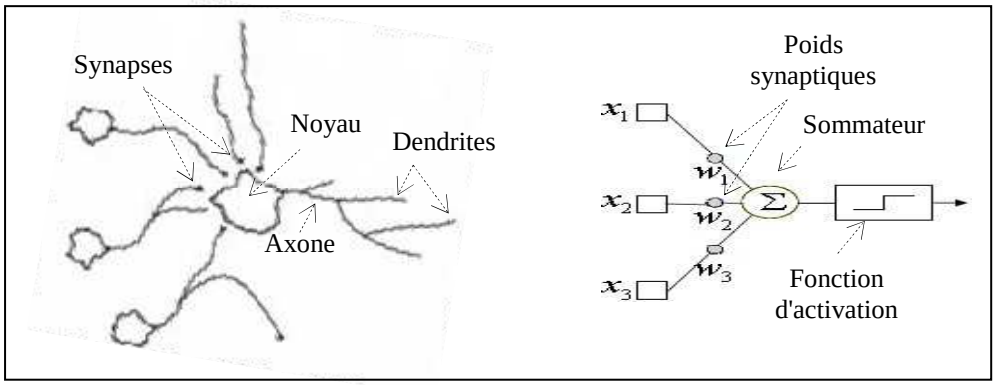


Figure 4.3. Mise en correspondance entre neurone biologique/neurone formel

Dès 1943, Mac Culloch et Pitts ont proposé un neurone formel simulant le neurone biologique et capable de mémoriser des fonctions booléennes simples [44-46]. Le fonctionnement de ce neurone formel tel que l'ont décrit, pour la première fois, Mac Culloch et Pitts se présente comme un composant calculatoire faisant la somme pondérée des signaux reçus en entrée, à laquelle nous appliquons une fonction dite d'activation, afin d'obtenir la réponse en sortie de la cellule (Figure 4.4).

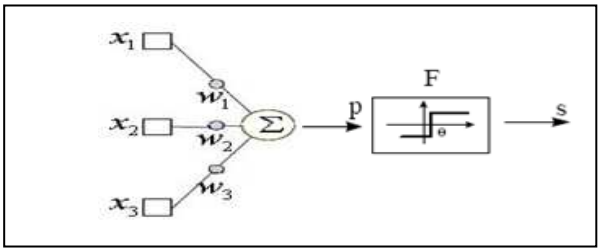


Figure 4.4. Fonctionnement de base d'un neurone formel (3 entrées et une sortie)

Par analogie au neurone biologique, le neurone formel doit être capable de recevoir en entrée différentes informations provenant des neurones environnants, analyser ces informations, de manière à envoyer en sortie une réponse, et enfin ajuster cette réponse

avant de l'envoyer aux neurones suivants. Pour ce faire, le neurone réalise trois opérations sur ses entrées :

- une pondération : multiplication de chaque entrée par un paramètre appelé poids de connexion ou poids synaptique. Plus la valeur d'un poids synaptique entre deux neurones est importante, plus l'intensité du signal entrant est forte, et donc, plus l'entrée correspondante est influente ;
- une sommation des entrées pondérées avec un ajout d'un seuil d'activation ou biais ;
- une activation : passage de cette somme dans une fonction, appelée fonction d'activation. Le résultat de cette fonction sera transmis en sortie du neurone pour être transféré vers les neurones suivants. Il est nécessaire que la valeur en sortie atteigne ou dépasse une certaine valeur seuil (biais du neurone) pour que l'information soit transmise.

La somme pondérée des signaux d'entrée d'un neurone *i* est :

$$p_i = \sum_{j=1}^{n} w_{ij}\, x_{ij} + b_i \qquad (4.1)$$

Avec :

w_{ij} : Poids synaptique associé à l'entrée *j* du neurone *i* ;
x_{ij} : entrée *j* du neurone *i* ;
b_i : biais du neurone *i*, appelé également seuil d'activation du neurone ou seuil interne de décharge ;
p_i : sortie du neurone *i*.

A partir de cette valeur, une fonction d'activation calcule la valeur de l'état du neurone qui sera ensuite transmise aux neurones avals, par la relation suivante :

$$s = f(p_i) \qquad (4.2)$$

Cette fonction d'activation est appelée également fonction de transfert ou de seuillage. Elle sert à introduire une non-linéarité dans le fonctionnement d'un neurone et présente généralement trois intervalles :

- en dessous du seuil, le neurone est non-actif ou inhibé (sa sortie vaut 0) ;
- aux alentours du seuil, une phase de transition ;
- au dessus du seuil, le neurone est actif ou excité (sa sortie vaut 1).

Pour ce faire, différentes fonctions d'activation sont exploitées telles que les fonctions seuil, sigmoïde, linéaire, gaussienne et tangente hyperbolique. La figure 4.5 montre quelques fonctions très utilisées en RAP.

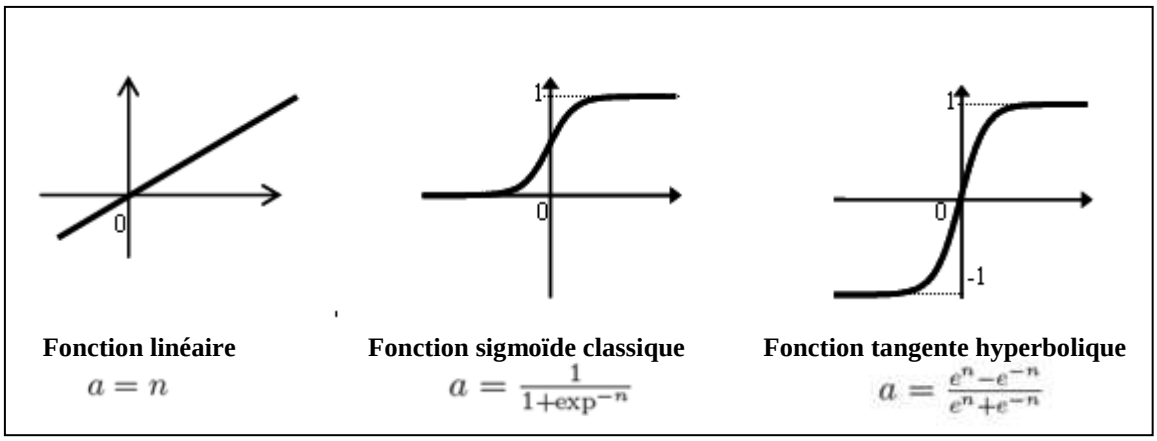


Figure 4.5 : Exemples de fonctions d'activation

4.3. Réseaux de Neurones Artificiels (RNA)

Un RNA est un maillage de plusieurs neurones, généralement organisés en couches. Autrement dit, c'est un graphe dont les sommets (neurones) ont une capacité à transformer un signal d'entrée en un signal de sortie. Les connexions entre les neurones servent à transférer les signaux d'un (ou plusieurs) sommet(s) vers un (ou plusieurs) autre(s). Les couches, autres que celles d'entrée et de sortie, ne sont pas visibles à l'extérieur du réseau, d'où leur appellation de "couches cachées" (figure 4.6). De façon générale, un RNA est défini essentiellement par :

- son architecture, c'est à dire le nombre de couches (entrée, sortie et cachées), le nombre de neurones par couche et le graphe de connexion ;
- son algorithme d'apprentissage, au cours duquel le réseau apprend, à partir d'exemples, à accomplir sa tache, tout comme le cerveau humain.

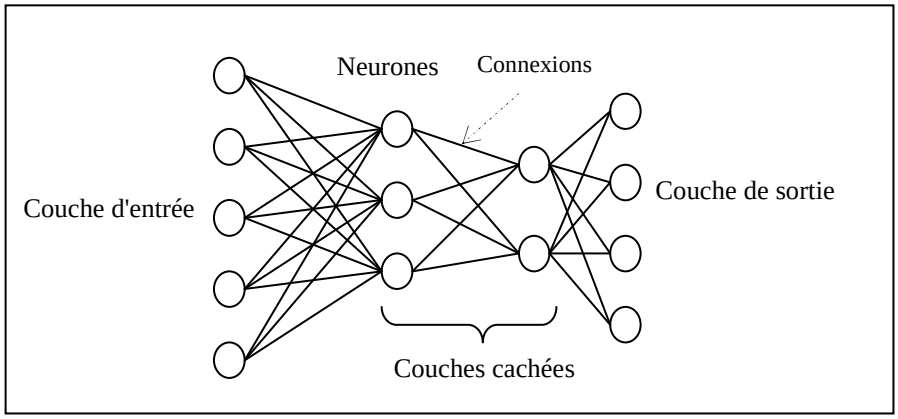


Figure 4.6 : Architecture simple d'un RNA

4.3.1. Historique sur les RNA

Il est difficile de résumer en quelques lignes plus d'un demi-siècle de recherche sur les Réseaux de Neurones, dont les étapes décisives sont jalonnées par des publications clés. Nous pouvons citer brièvement quelques grandes dates : Ce sont les scientifiques W. McCulloch et W. Pitts, l'un chercheur en neurologie et l'autre en psychologie cognitive, qui proposèrent en 1943 le tout premier modèle mathématique simplifié de neurone biologique, qu'ils ont appelé : le neurone formel [46]. Ce premier modèle reste l'élément de base des réseaux de neurones actuels. Il permet d'apprendre, de mémoriser des informations et de réaliser des fonctions logiques et arithmétiques. Le principe de fonctionnement de ce tout premier RNA est basé sur la notion de coefficient synaptique. Les entrées du neurone sont pondérées par des coefficients que l'on appelle "poids synaptiques". Mais les travaux de W. McCulloch et W. Pitts n'ont laissés aucune information pour adapter les coefficients synaptiques et il fallu attendre 1949 et D. Hebb, psychologue et neuropsychologue Canadien, qui donna un début de réponse grâce à ses travaux sur l'apprentissage, "The Organization of Behavior". S'inspirant du conditionnement de type pavlovien chez l'animal, D. Hebb propose une règle simple permettant de définir les coefficients synaptiques selon les liaisons des neurones: "*Il y a renforcement d'une connexion lorsque les deux neurones qu'elle relie sont simultanément excités"*. Cette règle, connue sous le nom de "Règle de Hebb", est encore utilisée aujourd'hui [46].

Notant qu'une année avant, en 1948, J. Von Neumann exposa une théorie sur les réseaux d'automates reproducteurs. Ces derniers sont des mécanismes se comportant de manière automatique, c'est-à-dire sans l'aide d'une intervention humaine, et surtout capables de se débrouiller seuls dans un environnement donné.

En 1958, F. Rosenblatt développe le modèle du Perceptron. Ce modèle possède deux couches de neurones : une couche de perception (servant à recueillir les entrées) et une couche de décision. C'est le premier modèle pour lequel un processus d'apprentissage a pu être défini. Ce modèle est ainsi capable d'apprendre à partir d'une base contenant un certain nombre de formes, à répondre (+1) si la forme d'entrée appartient à une classe A et (-1) si la forme appartient à une autre classe B. C'est le début de la notion de "l'apprentissage par exemple", qui est une caractéristique de l'intelligence humaine. Durant la même période, B. Widrow et T. Hoff développent le modèle linéaire Adaline

(Adaptive Linear Element) en s'inspirant du perceptron. Ce dernier sera, par la suite, le modèle de base des réseaux de neurones multicouches.

En 1969, Les recherches sur les réseaux de neurones ont été pratiquement abandonnées lorsque M.L. Minsky, chercheur en science cognitive et en intelligence artificielle, coécrit avec le mathématicien et informaticien S. Papert, un ouvrage "Perceptrons" mettant en avant les limites du modèle de Rosenblatt, en démontrant son incapacité à résoudre des problèmes non linéaires. Des critiques virulentes ont eu alors un effet catastrophique pour le domaine des réseaux de neurones, allant jusqu'à la suspension de toute subvention du gouvernement américain aux laboratoires travaillant dans ce domaine [46].

Durant toute la décennie 1970-1980, rares sont les travaux de recherche réalisés dans le domaine. Seuls quelques chercheurs ont continué à développer de nouvelles architectures et de nouveaux algorithmes plus puissants, tels que les mémoires associatives de T. Kohonen en 1972.

En 1982, les RNA ont connu un regain d'intérêt grâce aux travaux du physicien J. J. Hopfield. Ce dernier a développé un nouveau modèle basé sur des réseaux totalement connectés et s'inspirant de la règle de Hebb pour définir les notions d'attracteurs et de mémoire associative.

Il faudra tout de même attendre l'année 1984 pour que P.J. Werbos propose un modèle assez développé, qualifié de réseau multicouches et ne possédant pas les défauts démontrés par M.L. Minsky et S. Papert. Durant la même période, la machine de Boltzmann arrivera à traiter de manière satisfaisante les limites recensées du perceptron, battant en brèche les thèses de Minsky et Papert, grâce à l'ajout de couches dites "cachées".

En 1986, D. Rumelhart reprend le modèle multicouche de P.J. Werbos et développe un système d'apprentissage reposant sur la rétropropagation du gradient de l'erreur dans des systèmes à plusieurs couches. C'est ce nouveau développement, généralement attribué à D. Rumelhart et J. Mc Clelland, mais aussi découvert plus ou moins en même temps par P.J. Werbos et par Y. LeCun, qui a littéralement ressuscitée l'engouement pour le domaine des RNA. Depuis, une révolution survient dans le domaine des RNA : de nouvelles théories, de nouvelles structures et de nouveaux algorithmes de réseaux de neurones capables de traiter avec succès des phénomènes non-linéaires ont vu le jour. De par leur nature, ces réseaux constituent un domaine de recherche attractif et pluridisciplinaire englobant les mathématiques, l'informatique, la physique, le traitement

du signal, la psychologie, la neurobiologie, etc. La communauté gravitant autour de leur évolution, progresse sans cesse et multiplie leurs applications dans de nombreuses disciplines (traitement de la parole, imagerie, prévisions météorologiques et financières, sciences économiques, écologie et environnement, biologie et médecine...).

4.3.2. Architecture des RNA

Le choix de l'architecture influe à la fois sur les capacités de calcul du réseau et sur le type d'apprentissage susceptible d'être utilisé. Globalement, nous distinguons deux grands types d'architectures de réseaux de neurones :

- les réseaux de neurones non bouclés (acycliques, ou statiques) ;
- les réseaux de neurones bouclés (récurrents, ou dynamiques).

4.3.2.1. Réseau de neurones non bouclé

Un réseau de neurones non bouclé (Feed-forward, en Anglais) est un type de réseau dont l'information circule des entrées vers les sorties dans un sens unique, sans aucune rétroaction ou retour en arrière. Il est ainsi possible de représenter le réseau comme un graphe acyclique, c'est-à-dire qu'il est impossible de revenir à un neurone de départ quelconque en suivant les connexions.

Le réseau est ainsi organisé en trois couches : la couche d'entrée qui reçoit les données initiales (données d'entrée), la couche cachée qui fait propager l'information et enfin le neurone de sortie qui transmet à l'extérieur la valeur calculée par le réseau (Figure 4.7). Ce type de réseau est utilisé en classification, en reconnaissance des formes (caractères, parole, ...), en prédiction, etc. Un exemple connu de ce type de réseaux est le Perceptron Multi Couches PMC (MultiLayer Perceptron MLP, pour l'Anglais).

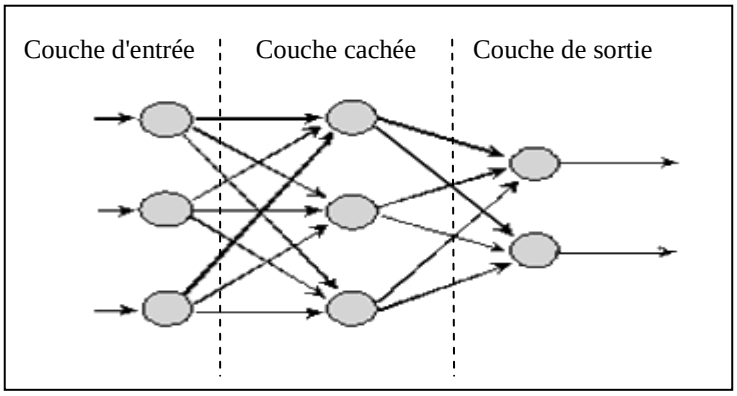


Figure 4.7 : Exemple de réseau de neurones non-bouclé

4.3.2.2. Réseau de neurones bouclé

Pour ce type de réseau, appelé également récurrent (Feedback, en Anglais), le graphe des connexions est cyclique: lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (cycle). Ces connexions récurrentes ramènent ainsi l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche (Figure 4.8). Les exemples de réseaux de neurones bouclés les plus connus sont: Les cartes auto-organisatrices de Kohonen et les Réseaux de Hopfield.

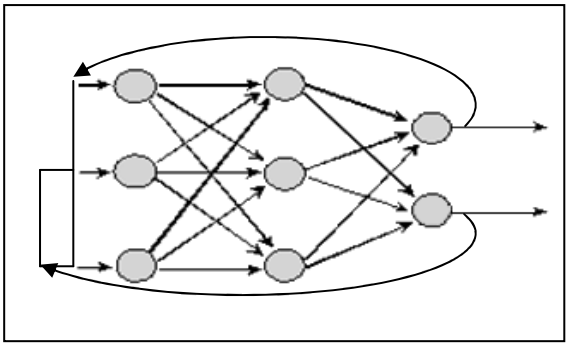


Figure 4.8 : Exemple de réseau de neurones bouclé

En pratique, ce type de réseau est utilisé pour effectuer des tâches de commande de processus, de modélisation de systèmes dynamiques ou de filtrage. Le réseau de Hopfield a la particularité d'être totalement connecté, c'est à dire que chaque neurone est relié à chacun des autres neurones, en ce sens qu'il n y ait aucune différenciation entre les neurones d'entrée et de sortie (Il n'y a plus de notions de couche comme dans le perceptron). Les connexions sont symétriques, c'est à dire que les connexions (i, j) et (j, i) sont affectées du même poids ($W_{i,j} = W_{j,i}$) *et* ($W_{i,i} = W_{j,j} = 0$). Le réseau évolue au cours du temps pour atteindre un état d'équilibre stable correspondant à une énergie minimale. L'idée de base du réseau de Hopfield est que le système nerveux recherche constamment des états stables, dans lesquels plus aucun neurone ne change son activation. Ces états stables permettent d'emmagasiner de l'information qui agit comme attracteur. Ainsi, pour obtenir cette information stockée, il faut fournir une partie de l'information (mémoire auto-associative).

4.3.3. Apprentissage d'un RNA

Parmi les plus importantes propriétés pour un RNA, la plus fondamentale est incontestablement sa capacité d'apprendre de son environnement et d'améliorer sa

performance à travers un processus d'apprentissage [44]. La capacité de mémorisation et de classification d'un objet ou d'une information n'étant pas acquise dès le départ, la plupart des RNA apprennent par l'exemple (un peu à la manière d'un enfant apprenant à reconnaître une image à partir d'exemples d'images qu'il a déjà vu). C'est donc une phase du développement d'un réseau de neurones durant laquelle son comportement est modifié jusqu'à l'obtention du comportement désiré. Cette notion d'apprentissage recouvre deux réalités souvent traitées de façon successive :

- la mémorisation, le fait d'assimiler sous une forme dense des exemples éventuellement nombreux ;
- la généralisation, le fait d'être capable, grâce aux exemples appris, de traiter des exemples distincts, encore non rencontrés, mais similaires.

Selon la manière dont le réseau apprend à partir d'exemples, nous pouvons subdiviser les techniques d'apprentissage en deux grandes catégories :

4.3.3.1. Apprentissage supervisé

Cette technique d'apprentissage consiste à calculer les paramètres de connexion entre les différentes couches du réseau de neurones, de telle manière que les sorties du réseau soient, pour les exemples utilisés, aussi proches que possible des sorties désirées. Les combinaisons d'entrées et de sorties désirées étant préalablement connues, il s'agit d'adapter les paramètres du réseau afin que pour chaque exemple, la sortie du réseau corresponde à une sortie désirée et connue. Ainsi, l'apprentissage dit *supervisé* force le réseau à converger vers un état final précis, chaque fois que nous lui présentons un motif en entrée [44, 47]. Pour ce faire, on utilise des algorithmes d'optimisation qui cherchent, de manière itérative, à minimiser une fonction dite de coût qui constitue une mesure de l'écart entre la réponse réelle du réseau et la réponse désirée. Ce type d'apprentissage est utilisé dans beaucoup d'architectures de RNA non bouclés, comme le MLP (Multi Layer Perceptron), RBF (Radial Basis Function), TDNN (Time Delay Neural Network), etc.

4.3.3.2. Apprentissage non supervisé

L'apprentissage non-supervisé est une méthode d'apprentissage automatique. Cette méthode se distingue de l'apprentissage supervisé par le fait qu'il n'y a pas de sortie *a priori*. Dans cet apprentissage, aucune donnée de sortie n'est fournie au système. Le réseau s'entraîne continuellement et sans besoin de supervision, c'est-à-dire, sans

que l'on ait besoin de le guider et de lui signifier comment il devrait se comporter. Ainsi, nous présentons une entrée au réseau et nous le laissons évoluer librement jusqu'à ce qu'il se stabilise. Une règle connue pour ce type d'apprentissage est la règle de Hebb qui revient à augmenter le poids de la connexion entre deux cellules si celles-ci sont simultanément actives et à le diminuer dans le cas contraire [48]. Ce type d'apprentissage non supervisé est utilisé dans beaucoup d'architectures de RNA bouclés tels que les réseaux de Hopfield, les cartes auto-organisatrices de Kohonen SOM (Self Organizing Map), etc.

4.3.4. Notions de rétropropagation et minimisation de fonction de coût

La rétropropagation (backpropagation, en Anglais) est une méthode qui a pour objectif le calcul de paramètres dits "poids synaptiques" W_i pour un réseau à apprentissage supervisé [49]. Elle consiste à minimiser une fonction dite de coût en ajustant ces poids. Le principe de cette fonction de coût est de propager un vecteur d'entrée, puis de calculer l'erreur en sortie par rapport à un vecteur de "sortie désirée" afin de modifier et de corriger les poids en fonction de cette erreur (figure 4.9). Cette procédure est itérée jusqu'à ce que les paramètres du modèle atteignent une stabilité avec une erreur de reconnaissance minimale. La fonction de coût la plus connue et la plus utilisée est la fonction dite "erreur quadratique moyenne" *MSE* (Mean Squared Error, en Anglais), que nous présentons brièvement comme suit :

Si nous avons une base de données d'apprentissage de *N* exemples, il faudra calculer pour chaque exemple *n* ($n \in N$), une erreur $e(n)$ représentant la différence entre la cible $d_i(n)$ (sortie désirée) et la valeur de sortie $y_i(n)$ (sortie réelle obtenue) :

$$e(n) = d_i(n) - y_i(n) \tag{4.3}$$

Ainsi, pour tous les exemples *N*, la fonction de coût *MSE* est donnée par l'équation:

$$MSE = \frac{1}{N} \sum_{n=1}^{N} e(n)^2 \tag{4.4}$$

Plus cette erreur est faible, plus le modèle reproduit fidèlement les observations utilisées pour l'apprentissage.

L'algorithme de rétropropagation que nous avons utilisé dans le cadre de notre travail est basé sur la régularisation bayésienne (RB) combinée à l'algorithme de Levenberg-Marquardt (LM), pour minimiser cette erreur d'apprentissage. Cette

combinaison a un grand avantage qui consiste en la capacité de généralisation du système (identification et classification de nouvelles données inconnues, présentées à l'entrée du réseau) et une convergence (obtention du résultat désiré en sortie) avec beaucoup moins d'itérations. Des études ont montré que l'association de la technique de la RB avec l'algorithme de Levenberg-Marquardt (LM) améliore sensiblement les résultats de l'apprentissage [50-52].

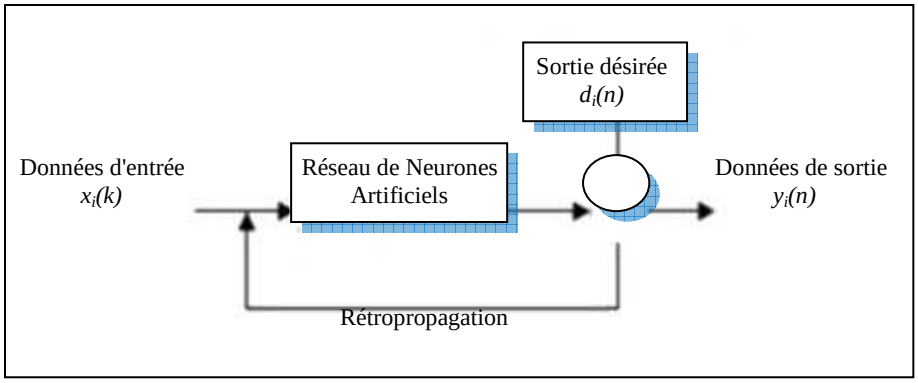


Figure 4.9. Rétropropagation par apprentissage supervisé

4.3.4.1. Régularisation Bayésienne (RB)

Un problème qui apparaît lors d'un apprentissage d'un réseau est le phénomène de sur-apprentissage ou sous-apprentissage. Par exemple, si le réseau de neurones apprend par cœur, il donnera de mauvais résultats quand nous lui présenterons brusquement des données différentes. Des techniques appropriées tentent d'optimiser la phase d'apprentissage afin que les phénomènes de sur et sous-apprentissage disparaissent. Le principe est de modifier les fonctions de coûts en leur ajoutant des coefficients supplémentaires qui permettent de réguler l'apprentissage du réseau. Une des techniques les plus connues est la Régularisation Bayésienne (RB) qui consiste à imposer des contraintes, en apportant une information supplémentaire, sur l'évolution des poids du réseau de neurones [53-55]. Cette méthode se base sur la modification de la fonction de coût *MSE* en lui ajoutant un terme dit de régularisation *MSW* (Mean Squared Weights) qui correspond à la somme des carrés des poids W_{ij} afin de pénaliser les valeurs absolues élevées des poids et éviter ainsi les risques de sur apprentissage. En d'autres termes, Cette technique force les paramètres (les poids) à ne pas prendre des valeurs élevées, et par conséquent à éviter le surajustement. Nous obtenons ainsi une nouvelle fonction de coût *MSE* régulée :

$$MSE_{reg} = \gamma MSE + (1-\gamma)MSW \quad (4.5)$$

Et,

$$MSW = \frac{1}{N}\sum_{j=1}^{N} W_j^2 \quad (4.6)$$

γ : paramètre dit de régularisation.

Cette modification provoque une diminution des valeurs des poids et force ainsi le réseau à avoir une bonne réponse tout en évitant le sur apprentissage. La RB donne, en général, des résultats très satisfaisants, en supposant que les poids suivent des distributions spécifiques (les paramètres sont estimés au fur et à mesure de l'apprentissage).

4.3.4.2. Algorithme de Levenberg-Marquardt (LM)

L'algorithme de Levenberg-Marquardt (LM) est l'un des algorithmes de second ordre les plus répandus pour les problèmes d'optimisation non-linéaire [52, 56, 57]. Il est aussi connu pour être le meilleur algorithme pour des problèmes d'optimisation appliqués à un RNA. Cet algorithme est une méthode itérative de minimisation de fonctions. L'un de ses avantages principaux est qu'il permet de converger très rapidement et avec beaucoup moins d'itérations contrairement aux autres algorithmes connus. Ainsi, il nécessite à peine une centaine d'itérations pour une convergence totale de la fonction de coût, par rapport à la méthode classique (simple gradient) qui peut aller jusqu'à plus d'un millier d'itérations.

Nous pouvons résumer l'algorithme LM au schéma itératif suivant :

$$\mathbf{q_i} = (J^T J + \lambda I)^{-1} J^T (\mathbf{y} - \mathbf{f}(\mathbf{p_i})) \quad (4.7)$$

A chaque itération, nous remplaçons $\mathbf{p_i}$ par une nouvelle estimation telle que :

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \mathbf{q}_i \quad (4.8)$$

Avec :

$\mathbf{P_i}$ vecteur de départ ;

J la jacobienne de **f** en **p** ;

I matrice identité (dite de régularisation) ;

$\mathbf{J^T J}$ matrice hessienne ;

λ pas d'apprentissage ajusté à chaque itération en le multipliant ou en le divisant par 10 (le pas de départ est fixé à $\lambda_0 = 0.005$ dans le cas de notre travail).

L'algorithme LM est décrit comme une interpolation entre la méthode de descente du gradient et l'itération de Gauss-Newton [56]. Ainsi, La descente du gradient dans notre cas correspond à:

$$\mathbf{q}_i = -\mu \nabla e(\mathbf{p}_i) \tag{4.9}$$

Par contre, la méthode de Gauss-Newton correspond à:

$$\mathbf{q}_i = -\left(\nabla^2 e(\mathbf{p}_i)\right)^{-1} \nabla e(\mathbf{p}_i) \tag{4.10}$$

Or,

$$\nabla e(\mathbf{p}) = \begin{pmatrix} \frac{\partial e}{\partial p_0} & \cdots & \frac{\partial e}{\partial p_{N-1}} \end{pmatrix} \tag{4.11}$$

Et

$$\frac{\partial e}{\partial p_j} = \frac{\partial}{\partial p_j} \sum_{i=0}^{M-1} (f_i(\mathbf{p}) - y_i)^2 \tag{4.12}$$

$$= \sum_{i=0}^{M-1} 2\,(f_i(\mathbf{p}) - y_i) \frac{\partial}{\partial p_j} (f_i(\mathbf{p}) - y_i) \tag{4.13}$$

$$= 2 \sum_{i=0}^{M-1} (f_i(\mathbf{p}) - y_i) \frac{\partial f_i}{\partial p_j}(\mathbf{p}) \tag{4.14}$$

Il vient ainsi au point p :

$$\nabla e = 2\,(J)^T (\mathbf{f} - \mathbf{y}) \tag{4.15}$$

De même, dérivant une seconde fois, il vient :

$$\frac{\partial^2 e}{\partial p_j^2} = 2 \sum_{i=0}^{M-1} \left(\frac{\partial f_i}{\partial p_j}\right)^2 + 2 \sum_{i=0}^{M-1} (f_i(\mathbf{p}) - y_i) \frac{\partial^2 f_i}{\partial p_j^2}(\mathbf{p}) \tag{4.16}$$

Et

$$\nabla^2 e(\mathbf{p}) = 2J^T J + 2 \sum_{i=0}^{M-1} (f_i(\mathbf{p}) - y_i) \nabla^2 f_i(\mathbf{p}) \tag{4.17}$$

Si nous prenons l'approximation linéaire de la fonction *f* ($\nabla^2 \mathbf{f} \simeq 0$) ou si le résidu (*fi(p)–yi)* est négligeable, le Hessien de *e* au point *p* devient :

$$\nabla^2 e \simeq 2J^T J \tag{4.18}$$

Nous pouvons donc écrire pour la descente du gradient :

$$\mathbf{q}_i = \mu' J^t \left(\mathbf{y} - \mathbf{f}(\mathbf{p})\right) \tag{4.19}$$

Et pour la méthode de Gauss-Newton :

$$\mathbf{q}_i \simeq \left(J^T J\right)^{-1} J^T \left(\mathbf{y} - \mathbf{f}(\mathbf{p})\right) \tag{4.20}$$

Si nous reprenons l'équation 4.7 de départ, nous pouvons constater que pour un λ grand, la méthode se rapproche de la descente du gradient (équation 4.19). A l'inverse, pour λ plus proche de zéro, nous suivons davantage la méthode de Gauss-Newton (équation 4.20). Ainsi, l'algorithme LM permet de réunir les deux méthodes de premier ordre (Descente de gradient) et de second ordre (Gauss-Newton) en une seule méthode unique.

4.4. Application des RNA en Reconnaissance Automatique de la Parole

Pour reconnaitre et classifier automatiquement une parole, plusieurs méthodes son utilisées. Parmi celles-ci, nous distinguons celles plus simples qui permettent de reconnaitre un vocabulaire limité et particulièrement les mots isolés, telle que la DTW (Dynamic Time Warping), où chaque mot du lexique est représenté par une réalisation de référence. Le processus de reconnaissance consiste à évaluer la distance d'une observation à chacune des références, par un processus d'alignement temporel. Parmi les méthodes à vocabulaire illimité les plus utilisées, nous citons les Modèles de Markov Cachés plus connus sous le nom de HMM (Hidden Markov Models), les SVM (Support Vector Machines) et les Réseaux de Neurones Artificiels (RNA) [58].

Dans l'approche markovienne, la reconnaissance de la forme à reconnaître s'effectue également par comparaison avec des formes de référence. A l'inverse des DTW où la forme est représentée par elle-même, les HMM représentent cette forme par un niveau plus abstrait correspondant à un modèle, composé d'un ensemble d'états et de transitions. La théorie des HMM décrit comment passer d'état en état à l'aide de probabilités de transitions et comment chaque élément de la séquence peut être émis par un état du HMM à l'aide de probabilités d'observations par état. Tous les états ont des transitions unidirectionnelles vers tous les autres états, y compris vers eux-mêmes. Ils sont cachés et chacun émet des "observations" qui, elles, sont observables. On ne travaille donc pas sur la séquence d'états, mais sur la séquence d'observations générées par les états. Une description détaillée des

principes de base et des différentes étapes d'application en Reconnaissance Automatique de la Parole (RAP) est assez disponible et peut être consultée sur internet et dans beaucoup d'ouvrages [59].

Pour les SVM, leur principal objectif est de déterminer si un élément appartient à une classe ou pas, d'où leur principale exploitation dans le domaine de la classification. Nous disposons d'un ensemble de données et nous cherchons à séparer ces données en deux groupes. Le premier est l'ensemble de données appartenant à une classe, ces données sont étiquetées généralement par (+) et un autre ensemble qui contient les éléments qui n'appartiennent pas à la classe donc étiquetées (-). L'algorithme SVM permet de trouver un hyperplan séparateur entre ces deux groupes. Pour optimiser la séparation, cet algorithme cherche l'hyperplan pour lequel la distance entre la frontière des deux groupes et les points les plus proches est maximale. En d'autres termes, le principe des SVM est de séparer les exemples de deux classes avec cet hyperplan tout en gardant le maximum de marge entre les exemples et ce même hyperplan. C'est le principe de maximisation de la marge ou distance du point le plus proche à l'hyperplan. Pour cela, les SVM sont également appelés "Maximum Margin Classifier" [60].

Enfin pour les RNA que nous avons utilisés dans le cadre de notre travail, les cellules sont structurées en couches successives capables d'échanger des informations au moyen de connexions qui les relient. Ces systèmes qui tentent de stocker et retrouver l'information de manière "similaire" au cerveau sont particulièrement adaptés à la Reconnaissance de Caractères OCR (Optical Character Recognition), la Reconnaissance Automatique de la Parole (RAP), du locuteur (RAL) et de visages RAV [44, 46, 61-65]. Les algorithmes connexionnistes possèdent une forte capacité discriminante et une bonne résistance aux bruits, l'un des plus grands facteurs de complexité que nous rencontrons dans l'élaboration d'un système de RAP, d'où l'idée de leur exploitation dans ce domaine. Beaucoup de techniques basées sur ces modèles sont proposées, quelques unes utilisent les paramètres statiques du signal (MLP) alors que d'autres tiennent compte de l'évolution dynamique du signal de parole (TDNN, Réseaux Récurrents, …).

Tout comme les autres systèmes de RAP, ceux à base RNA passent nécessairement par les deux étapes classiques (Figure 4.10) qui consistent en :

- une phase d'apprentissage permettant au système de lire les paramètres de référence, représentant les sons qui constituent le vocabulaire de l'application. Ces vecteurs de références sont obtenus à partir de modèles acoustiques qui permettent de caractériser les

différents sons prononcés. Le principe est de fournir au réseau une série d'exemples *x* et de résultats *y*, et ensuite trouver des coefficients spécifiques, appelés poids *W*, pour avoir un bon taux de reconnaissance et surtout une bonne généralisation. Grâce aux exemples appris, le système est capable de traiter des exemples distincts, encore non rencontrés, mais similaires.

- une phase de classification durant laquelle toute parole prononcée sera identifiée en comparaison avec les modèles de référence préalablement enregistrés. Les paramètres extraits lors de l'analyse acoustique seront introduits dans un classifieur automatique qui permettra de déterminer la nature de la parole prononcée. Pour ce faire, les vecteurs acoustiques, obtenus lors de la paramétrisation, seront comparés aux vecteurs du dictionnaire de référence (ou à leurs modèles obtenus lors de la phase d'apprentissage) en calculant une distance minimale, telles que la DTW (Dynamic Time Warping), la distance euclidienne, la distance de manhabolis, etc.

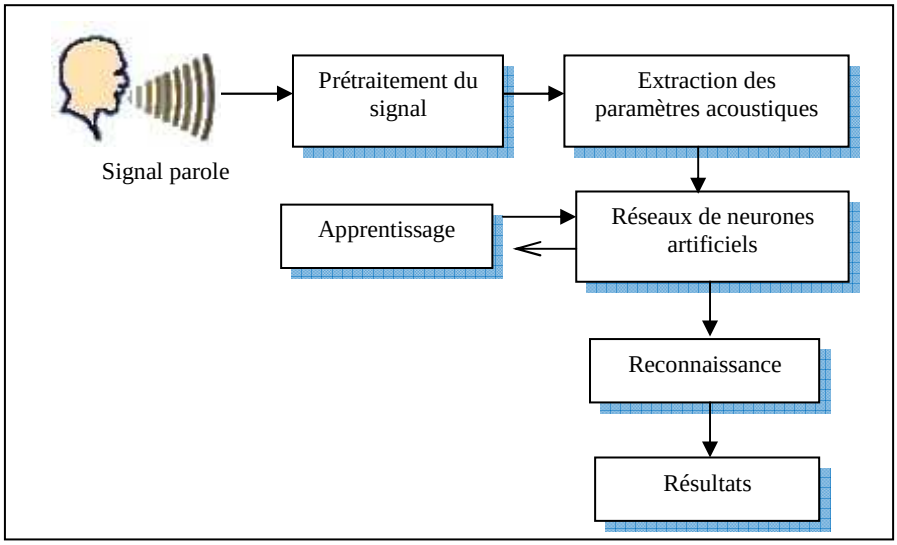


Figure 4.10. Structure d'un système standard de RAP basé sur les RNA

Il existe de nombreux types de RNA utilisés en RAP [63]. Parmi les plus utilisés, nous citons les réseaux de neurones multicouches MLP (MultiLayer Perceptron, en Anglais) également appelés PMC (Perceptron Multi Couches), et les réseaux de neurones à décalages temporels TDNN (Time Delay Neural Network, en Anglais). Nous donnons plus de détails sur ces deux types de réseaux, notamment le TDNN que nous avons exploité dans le cadre de notre travail. Pour les autres modèles, une documentation assez importante peut être exploitée sur internet. Nous avons donc jugé utile de ne pas les exposer dans cette thèse.

4.4.1. Perceptron Multi Couches

Le Perceptron Multi Couches MLP (Multi Layer Perceptron, en Anglais) est un modèle de réseau non bouclé dont les informations, ou activations, circulent dans un seul sens, c'est-à-dire des neurones d'une couche aux neurones de la couche suivante. Les neurones y sont ainsi organisés en couches successives : une couche d'entrée, une couche de sortie et entre les deux une ou plusieurs couches intermédiaires, appelées couches cachées (figure 4.11). Chaque neurone d'une couche reçoit des signaux de la couche précédente et transmet le résultat à la suivante, si elle existe. Les neurones d'une même couche ont la même fonction d'activation, mais ne sont pas interconnectés. Un neurone ne peut donc envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. L'information est ainsi transmise de manière unidirectionnelle du neurone *j* vers le neurone *i*, affectée du coefficient pondérateur w_{ij}.

La figure 4.11 montre un exemple d'un réseau MLP à une seule couche cachée avec :

- x_n n^{ième} entrée du réseau ;
- w_{jn} poids affecté à la connexion reliant le neurone *n* de la couche d'entrée au neurone *j* de la couche cachée ;
- w_{ij} poids affecté à la connexion reliant le neurone *j* de la couche cachée au neurone *i* de la couche de sortie ;
- f_j activation de la couche cachée. Elle est donnée par l'équation suivante :

$$f_j = F(a_i) = F(\sum_{n=1}^{N} w_{jn} . x_n) \qquad (4.\ 21)$$

Avec :

F( ) fonction d'activation ;

$x = (x_1, x_2, \ldots, x_N)$ vecteur d'entrée.

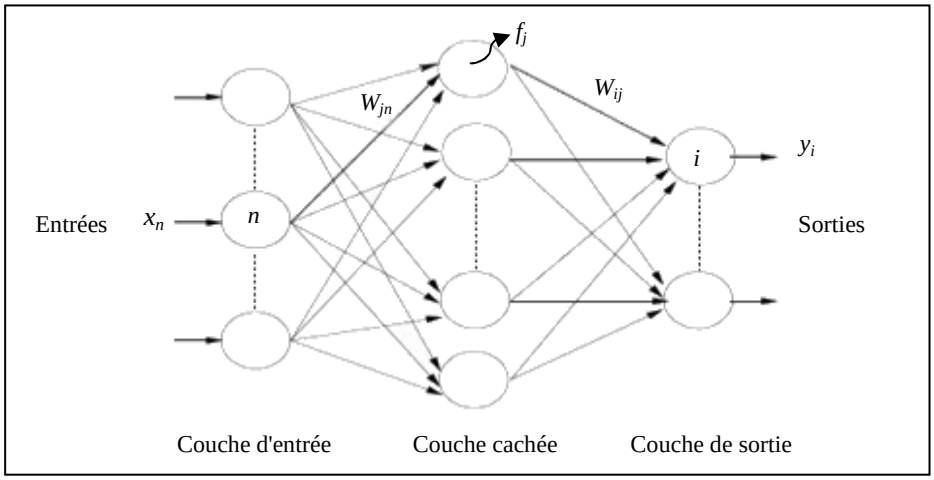


Figure 4.11 : Schéma d'un réseau MLP

De la même façon, l'activation de la $i^{ième}$ unité de la couche de sortie est obtenue en transformant le produit scalaire a_i avec la fonction d'activation :

$$y_i = F(a_i) = F(\sum_{j=1}^{L} w_{ij} . F(\sum_{n=1}^{N} w_{jn} . x_n)) \quad (4.\ 22)$$

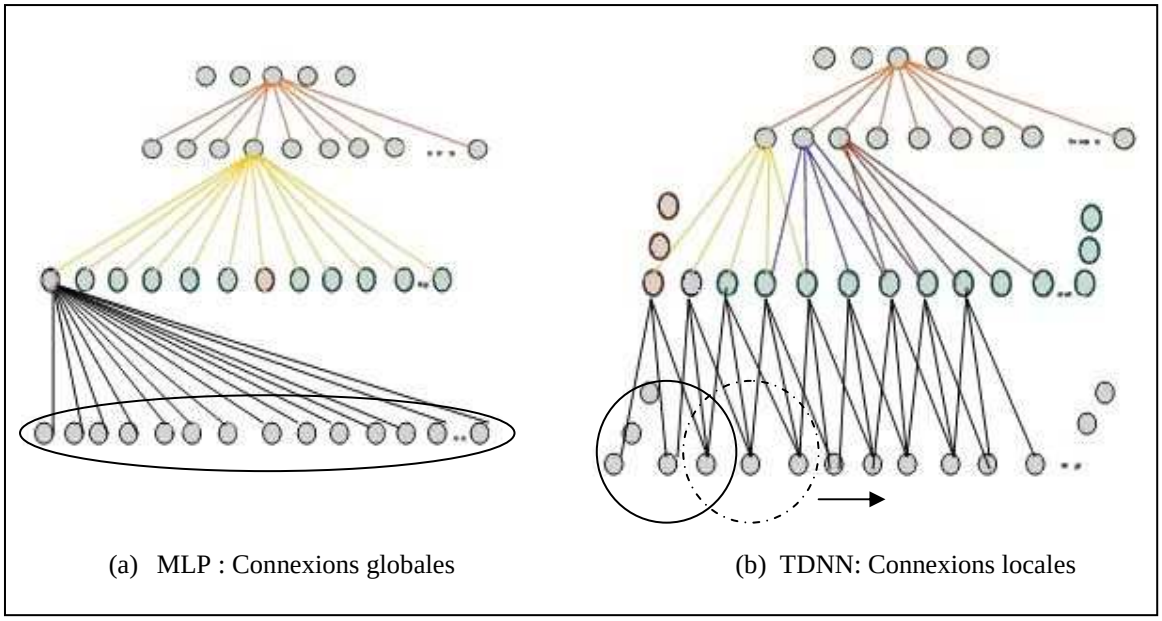
La fonction d'activation utilisée est généralement la fonction sigmoïde de la forme :

$$F(x) = \frac{1}{1+e^{-x}} \quad (4.\ 23)$$

4.4.2. Réseaux à décalages temporels TDNN

Cette architecture, connue sous le nom de TDNN (Time Delay Neural Network), a été utilisée pour la première fois par A. Waibel pour la reconnaissance de la parole [66]. Waibel a montré de très bons résultats pour la classification des consonnes japonaises [b], [d] et [g]. Il part du principe que pour une modélisation de signaux dynamiques tels que la parole, il est nécessaire d'introduire de la mémoire dans le réseau. Le TDNN se singularise d'un réseau de neurones classique, tel que le réseau MLP par le fait qu'il prend en compte la notion du temps, donc l'aspect dynamique de la parole tel que le phénomène de coarticulation. Les TDNN sont constitués comme les MLP d'une couche d'entrée, de couches cachées et d'une couche de sortie, mais ils se différencient de part l'organisation des liaisons inter-couches [66, 67]. Alors que le MLP prend en compte tous les neurones de la couche d'entrée en même temps, le TDNN ne prend qu'une fenêtre du spectre puis effectue un balayage temporel (Figure 4.12).

Le TDNN est basé principalement sur trois idées : poids partagés, fenêtres temporelles et délais (Figure 4.13). Dans les séquences de signaux comme la parole, le contexte est important et aide à la classification des consonnes et des voyelles. L'idée est d'augmenter le contexte et de partager donc les poids (nous parlons alors de réseaux de neurones à poids partagés) et chaque vecteur du contexte est traité par les mêmes fonctions. Ce principe de poids partagés permet de réduire le nombre de paramètres du réseau neuronal et induit ainsi une capacité de généralisation plus importante. Les poids sont partagés suivant la direction temporelle, c'est à dire que pour une caractéristique donnée, la fenêtre associée à celle-ci aura les mêmes poids selon la direction temporelle.



(a) MLP : Connexions globales (b) TDNN: Connexions locales

Figure 4.12 : Liaisons inter-couches d'un RNA de types (a) MLP et (b) TDNN

Dans un réseau TDNN, les neurones de la première couche cachée sont reliés aux neurones de la couche d'entrée par des connexions à retard, et les neurones de la deuxième couche cachée sont connectés à ceux de la première couche cachée par le même principe (figure 4.13). Deux importantes parties composent le réseau TDNN : l'extraction des caractéristiques et la classification.

La partie "Extraction des caractéristiques" se singularise par :

- le nombre de couches cachées (chaque couche a deux directions: direction temporelle et direction caractéristique) ;
- le nombre de neurones de chaque couche *i* selon la direction temporelle, *window_*t_i (fenêtre d'observation) ;
- le nombre de neurones de chaque couche selon la direction caractéristique, *nb_features* ;
- la taille de la fenêtre temporelle vue par chaque couche (sauf celle d'entrée) soit le nombre de neurones de la couche i vus par un neurone de la couche *i+1*, *field_*t_i (fenêtre de spécialisation) ;
- le délai temporel (nombre de neurones) entre deux fenêtres successives dans une couche donnée, *delay* (délai).

La partie "classification" se comporte comme un réseau de neurones de type MLP où chaque neurone de la couche est connecté à tous les neurones de la couche suivante

[68]. La première couche de cette partie "classification" correspond à la dernière couche de la partie "Extraction" du TDNN.

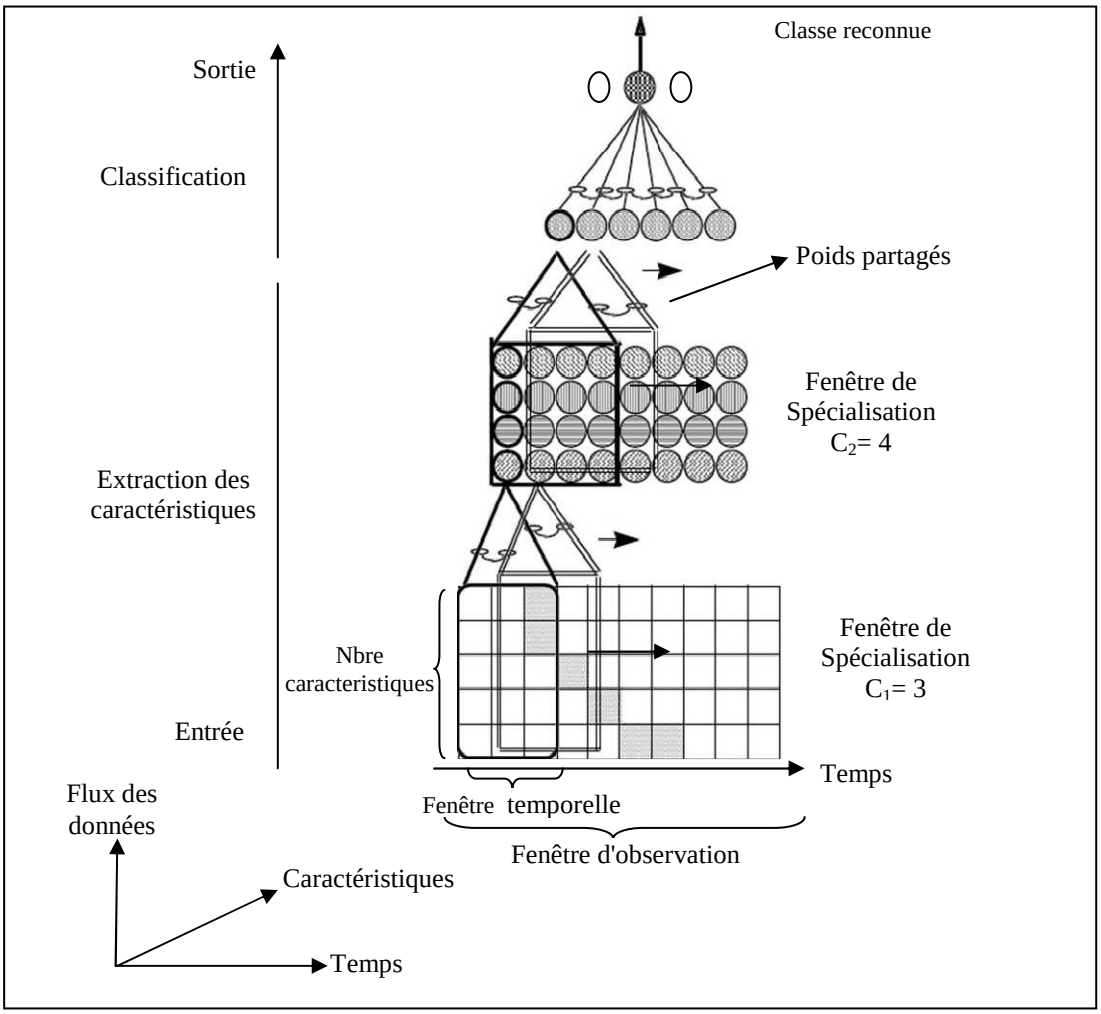


Figure 4.13 : Exemple de réseau à décalages temporels TDNN

Pour l'apprentissage du réseau TDNN, nous ajustons itérativement tous les poids de façon à réduire l'erreur *MSE* (Mean Squared Error), qui représente l'écart quadratique moyen entre la sortie réelle obtenue et la sortie désirée :

$$MSE = \frac{1}{2}\sum_i^n (d_i - y_i)^2 \qquad (4.24)$$

Avec :

d_i : sortie désirée pour le neurone i ;

y_i : sortie réelle obtenue par le réseau.

Apres initialisation aléatoire des poids de connexion, nous calculons la sortie du réseau en fonction de l'entrée (calcul du potentiel) par l'équation suivante :

$$y_i = f(p_i) \tag{4.25}$$

$$p_i = \sum_{i=1}^{N} \sum_{d}^{D} W_{ij}^{D} * e_i\,(t-d) \tag{4.26}$$

Avec :

p_i : Potentiel du neurone *i* ;

e_i : état du neurone i ;

t : l'instant présent ;

d : délai (retard).

W_{ij}^{D} : Poids de la connexion reliant le $i^{ème}$ neurone de la couche inférieure et le $j^{ième}$ neurone de la couche supérieure ;

$\sum_{i=1}^{N}$: Somme sur les neurones en entrée du neurone i ;

$\sum_{d}^{D}$: Somme sur les délais (fenêtres de spécialisation) ;

Les réseaux à décalages temporels TDNN ont été largement utilisés pour la reconnaissance de consonnes de plusieurs langues, dont le Japonais [66], l'Anglais [67], l'Arabe [69], le Français [70], le Hindi [71] et même le Malais [72]. Ils ont également été utilisés dans le cadre de la reconnaissance de la parole continue [64, 65, 73], la reconnaissance de l'écriture manuscrite [74], etc. Ces dernières années, ils ont été exploités avec succès dans d'autres domaines plus évolués tels que la reconnaissance des gestes de la main [75], la téléphonie mobile et les communications numériques avancées [76, 77] et plus récemment encore, l'inversion acoustico-articulatoire pour déterminer la forme du conduit buccal en fonction des sons prononcés [78]. L'objectif de notre travail est de les adapter pour la reconnaissance et la classification automatique de Paroles Pathologiques PP_{ath}.

4.5. Conclusion

Nous avons donné un aperçu global sur le principe de fonctionnement des RNA, en détaillant particulièrement une des phases les plus importantes de la conception d'un RNA, à savoir la phase d'apprentissage dont le bon choix permet une meilleur classification. Enfin, nous avons mis l'accent sur les réseaux à décalage temporel TDNN, que nous avons appliqué dans la classification automatique de PP_{ath} d'origine fonctionnelle (PP_{ark}) et d'origine organique ($PŒ_{so}$), par rapport à la parole normale (PN_{orm}).

CHAPITRE 5

APPLICATION DES RNA A LA CLASSIFICATION DES PAROLES PATHOLOGIQUES

5.1. Introduction

Ces dernières années, la reconnaissance de la parole pathologique PP_{ath} a reçu une attention particulière dans les recherches en TAP. Dans les récentes approches de classification automatique des PP_{ath}, plusieurs méthodes basées sur la reconnaissance de formes sont utilisées [79-84]. Dans leur sillage, des recherches sont orientées naturellement vers l'application des RNA. Ces derniers donnent de bons résultats en reconnaissance de formes, et la Reconnaissance Automatique de la Parole (RAP) est une technique traditionnellement connue comme un problème de reconnaissance de formes.

Dans ce chapitre, nous avons exposé la conception et l'architecture de notre système de classification élaboré. Nous avons appliqué ensuite les réseaux de neurones dynamiques TDNN pour reconnaitre et classifier automatiquement deux paroles pathologiques par rapport à la Parole Normale : la Parole Œsophagienne et la Parole Parkinsonienne. Les résultats et leurs interprétations sont discutés à la fin du chapitre.

5.2. Conception et architecture du système de classification élaboré

De façon générale, nous avons suivi les étapes suivantes: Prétraitement du signal, extraction des paramètres acoustiques, apprentissage puis discrimination automatique des PN_{orm} et PP_{ath} (Figure 5.1).

Une des plus importantes phases de notre système est le choix adéquat des paramètres acoustiques à exploiter comme vecteurs d'entrée. Pour mieux discriminer les PP_{ath} par rapport aux PN_{orm}, nous avons opté pour des paramètres assez représentatifs de la PP_{ath} et dont nous avons montré l'importance dans la partie analyse acoustique (Jitter, Shimmer, Energie, TPZ). A ces paramètres, nous avons ajouté les paramètres MFCC (Mel Frequency Cepstral Coefficients). Pour la mise au point des outils du prétraitement des fichiers sons (préaccentuation, détection de parole utile et fenêtrage), ainsi que l'apprentissage et la classification automatique de la PP_{ath} par rapport à la PN_{orm}, nous avons utilisé le langage de programmation Matlab 2007. Il est aussi important à noter que les vecteurs acoustiques MFCC, Jitter, Jitter factor, Shimmer, Shimmer factor, Energie et TPZ ont été conçus également à partir de ce langage. En ce qui concerne les paramètres du TDNN, nous avons utilisé les caractéristiques suivantes:

- une couche d'entrée comportant (14*18) neurones (14 neurones comme taille de la fenêtre d'observation et 18 neurones comme taille des caractéristiques) ;
- deux couches cachées de 12 et 8 trames respectivement ;

- une fenêtre d'observation correspondant à une taille de 160 ms, jugée suffisante pour prendre en compte la taille du phonème le plus long ou de la voyelle soutenue, soit donc 14 trames en prenant une fenêtre de 30 ms et un pas de 10 ms;
- une fenêtre de spécialisation de la couche d'entrée correspondant à C_1=3 ;
- une fenêtre de spécialisation de la 1^{ere} couche cachée correspondant à C_2 = 5 ;
- un délai temporel entre deux fenêtres successives, Delay=1.

Comme fonctions de transferts, nous avons utilisé la fonction tangente hyperbolique (type sigmoïde) pour chaque nœud des couches cachées et une fonction linéaire pure pour la couche de sortie.

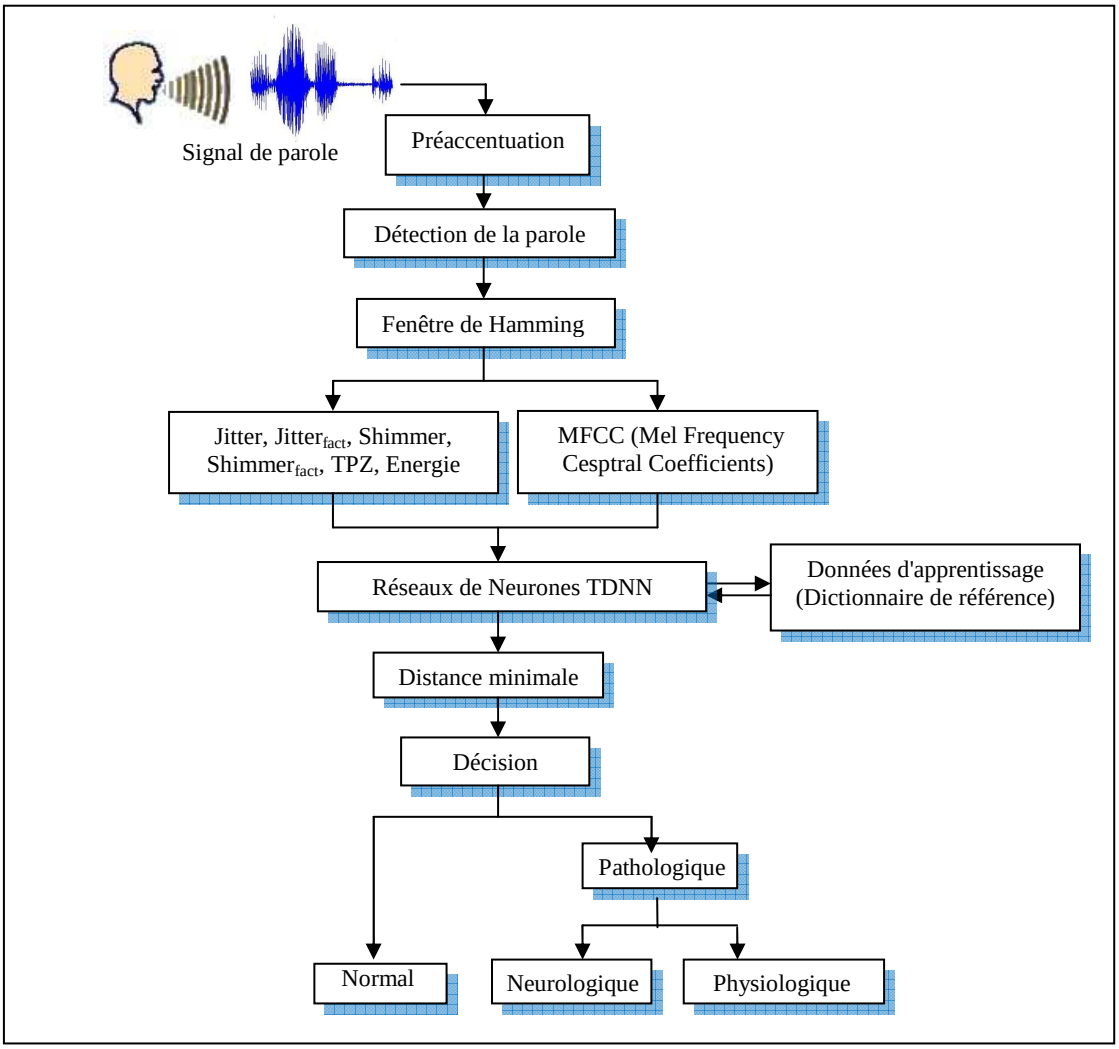


Figure 5.1: Organigramme de classification automatique de parole normale/pathologique

5.2.1. Enregistrements du corpus

Pour extraire les fichiers sons, nous avons exploité le même corpus utilisé pour l'analyse acoustique. Ce corpus comprend aussi bien des Paroles Parkinsoniennes PP_{ark}, Œsophagiennes $PŒ_{so}$, que Normales PN_{orm}. En plus des voyelles soutenues et des voyelles normales préalablement enregistrées, nous avons segmenté manuellement des mots, phrases et parole continue du corpus pour obtenir des fichiers de taille moyenne de 400 à 600 ms et en nombre assez suffisant. Nous avons divisé notre corpus des enregistrements sonores en deux groupes : un groupe concerne 120 fichiers sonores à exploiter lors de la phase d'apprentissage (équitablement répartis entre les PP_{ark}, $PŒ_{so}$ et PN_{orm}) et un autre également de 120 autres fichiers à exploiter lors de la phase de tests de classification. En effet, une fois le réseau de neurones entrainé, il est nécessaire de tester la fiabilité de notre système sur une autre base de données différente de celle utilisée pour l'apprentissage. Ce test permet d'apprécier les performances du système.

L'idéal en recherche est d'inclure le plus grand nombre de données afin de faire un traitement statistique le plus représentatif possible. Il reste que ce type de recherche appliquée à la pathologie est difficile. La recherche clinique pose un certain nombre de problèmes: les patients malades sont fatigables et souvent fragilisés psychologiquement; les conditions d'examen et d'enregistrement présentent des contraintes non négligeables pour les expérimentateurs [40].

Avant de passer à l'extraction des paramètres acoustiques représentatifs du signal de parole à étudier, nous devrons nécessairement faire subir à ce dernier quelques prétraitements importants qui nous permettent de récupérer le signal de parole "utile".

5.2.2. Prétraitement du signal de parole

Dans cette étape, le signal analogique capté au moyen d'un microphone sera transformé en composantes numériques plus faciles à traiter. Nous échantillonnons ce signal usuellement à une dizaine de kHz, afin de conserver l'information spectrale jusqu'à environ 5 kHz. Dans le cadre de notre travail, le signal de parole est échantillonné à une fréquence de 11025 Hz avec une précision de 16 bits. Comme les hautes fréquences sont souvent atténuées, nous les renforçons artificiellement à l'aide d'un filtre numérique. En fonction de l'application envisagée, la qualité demandée par la capture de la parole peut rapidement devenir très importante. Cette qualité dépend de la variabilité de la voix du locuteur dans le temps comme dans le cas de maladie (rhume, angine, …), des états émotionnels (joie, stress, angoisse …) et aussi de l'âge. De plus, les conditions d'acquisition du signal de

parole, telles que le bruit environnant et la fidélité des équipements du microphone (distorsions et bruits du filtrage du canal de transmission) jouent très fortement sur la qualité de la capture, et donc sur la qualité de l'analyse acoustique de la parole. Compte tenu de toutes ces contraintes, nous devrons subir un prétraitement pour le signal avant l'extraction des paramètres acoustiques, de manière à éliminer tous ces facteurs extérieurs qui pourraient influer sur les résultats à obtenir. Cette étape de prétraitement se résume en :

- une préaccentuation dont l'objectif est d'augmenter la quantité d'énergie dans les hautes fréquences et d'avoir une compensation de filtrage des effets de l'acquisition du signal (Figure 5.2). Pour cela, le signal de parole enregistré est appliqué à l'entrée d'un filtre de premier ordre FIR (Finite Impulse Response) de la forme :

$$H(z) = 1 - 0.95\, Z^{-1} \tag{5.1}$$

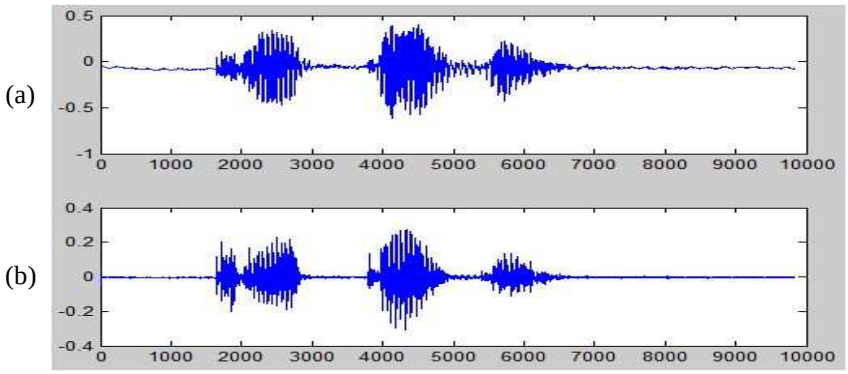


Figure 5.2 : Onde temporelle du mot [kataba]
(a) avant préaccentuation
(b) après préaccentuation

- une délimitation des débuts et fins de mots et élimination de toutes les portions du signal enregistré qui ne sont pas de la parole (Figure 5.3). Le défi consiste à éliminer ces échantillons inutiles à partir du signal sans perdre ou fausser l'information pertinente véhiculée par le signal de parole. Une fonction procédure, réalisée sous Matlab 2007, utilise un seuil minimal d'énergie moyenne calculé sur la base d'enregistrements de différents bruits d'environnement. Dès que l'énergie dépasse un seuil minimal dans une trame du signal (fenêtre de 30 ms), nous considérons que le début de parole commence à partir de cette trame et toutes les autres trames précédentes sont éliminées (figure 5.3). La même procédure est appliquée à la fin du signal de parole.

Il reste néanmoins que cette procédure donne de bons résultats pour le cas de mots isolés, mais reste très limitée dans le cas d'une parole continue car les frontières de mots sont très difficiles à distinguer (du fait des phénomènes de coarticulation), sauf si le locuteur marque explicitement une pose entre chaque mot.

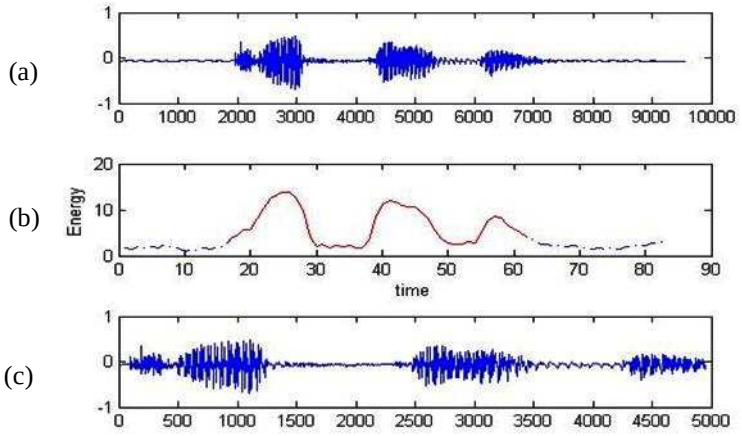


Figure 5.3 : Elimination des trames inutiles de début et fin du mot [kataba]

(a) signal de parole enregistré au départ
(b) énergie du signal de parole
(c) signal de parole après délimitation des frontières

- une extraction des paramètres acoustiques sur des portions de signal supposées stables, du fait que le signal de parole est connu comme très variable et donc non stationnaire. Dans notre application, le signal enregistré à une fréquence d'échantillonnage de 11025 Hz est segmenté en une succession de trames (fenêtres) de *N=330* échantillons chacune. Soit des fenêtres correspondant à un intervalle de temps de 330/11.025 ≈ 30 ms, car l'observation du signal de parole montre qu'il n'évolue pas ou peu sur des durées de cette taille. Les paramètres sont extraits avec un pas de recouvrement de 110 échantillons entre les trames, soit à un intervalle de temps de 10 ms. Pour le fenêtrage, nous avons choisi la fenêtre de Hamming, qui se présente sous la forme :

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \qquad (5.2)$$

N: taille de la fenêtre.

5.2.3. Extraction des vecteurs acoustiques

Le choix des paramètres est particulièrement important, car il s'agit de la base du système de classification. Du fait de la grande variabilité de la parole, les plus grands défis actuels en RAP est de mettre au point des algorithmes et techniques qui nous

permettent d'avoir des taux de reconnaissance appréciables, quelque soit le contexte et l'environnement. En d'autres termes, il s'agit de chercher à obtenir la forme la plus représentative possible du signal afin de réduire au maximum le taux d'erreur de reconnaissance. De même, le choix idéal des paramètres est également dicté par la facilité de mesure et enfin une robustesse aux distorsions et aux bruits. Cette étape, appelée modélisation acoustique, est très importante car elle contribue directement aux performances globales du système de classification. Dans le cadre de notre travail, nous avons utilisé respectivement comme vecteurs acoustiques le Jitter, le Jitter factor, le Shimmer, le Shimmer factor, l'énergie et le Taux de Passage par Zéro TPZ. En addition, nous avons ajouté les coefficients MFCC qui permettent une modélisation du signal de parole par des filtres conformes à notre système auditif [85, 86]. Les MFCC sont utilisés depuis longtemps dans l'identification du locuteur et des applications de reconnaissance, mais ont aussi montré des résultats prometteurs dans les récentes évaluations des voix et paroles pathologiques [83, 84]. Ces paramètres sont issus de l'hypothèse que le signal de parole est le résultat de la convolution entre un filtre (conduit vocal) et une excitation (cordes vocales). Une transformation homomorphique permet de transformer ce produit en une somme qui est ensuite filtrée pour obtenir les MFCC. Cette transformation homomorphique se décompose en deux étapes principales (Figure 5.4) :

- un passage dans le domaine spectral par calcul du module de la Transformée de Fourier Discrète DFT (Discret Fourier Transform, en Anglais). Le calcul de la DFT se fait sur des fenêtres glissantes ;
- un changement d'échelle pour rendre compte de la perception humaine. Pour cela, nous utilisons une échelle dite Mel, connue comme échelle perceptive du signal de parole. Nous rappelons que la correspondance entre une fréquence en Hz et une fréquence en Mel (Equation 5.3).

$$F_{mel} = 2595.\log(1 + \frac{F_{Hz}}{700}) \quad (5.3)$$

- enfin, une application de l'inverse de la Transformée en Cosinus Discrète IDCT (Inverse Discret Cosine Transform, en Anglais) nous permet de convertir le logarithme du spectre Mel en domaine temporel.

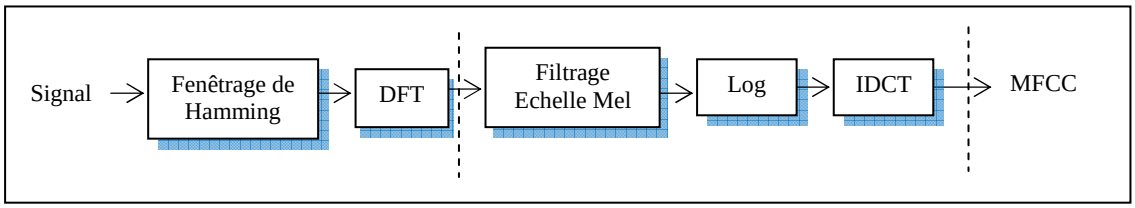


Figure 5.4 : Organigramme de calcul des paramètres MFCC

Les coefficients MFCC représentent l'information du conduit vocal. Ils ne sont pas corrélés et représentent une information proche des formants. Dans le cadre de notre travail, nous avons pris en compte uniquement les 12 premiers coefficients, sans tenir compte du coefficient relatif à l'énergie. De plus, les dérivées premières ΔMFCC et secondes ΔΔMFCC ne sont pas prises en considération car elles alourdissent considérablement l'apprentissage du réseau.

Pour l'extraction des vecteurs acoustiques, nous avons utilisé une fenêtre glissante de Hamming de 30 ms, avec un pas de recouvrement de 10 ms. Avant d'injecter ces vecteurs à l'entrée de notre système, ils ont été au préalable normalisés sur un intervalle [-1, +1]. Ces valeurs moyennes sont plus proches de régions de transitions d'une fonction sigmoïde qui permet un plus rapide apprentissage [51, 87]. Les valeurs des paramètres acoustiques d'entrée du réseau ne sont souvent pas de la même échelle et du même type. Il est donc nécessaire de normaliser les données en centrant et en réduisant les variables afin qu'elles aient le même impact sur le modèle. Pour cela, nous avons transposé les variables d'origine vers de nouvelles variables centrées et réduites :

$$\overline{x_i} = \frac{1}{N}\sum_{n=1}^{N} x_i \tag{5.4}$$

$$\sigma_i^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_i - \overline{x_i}\right)^2 \tag{5.5}$$

$$x_i = \frac{x_i - \overline{x_i}}{\sigma_i} \tag{5.6}$$

Avec :

N : nombre de trames ;
σ_i : variance du signal.

Rappelons qu'une variable centrée réduite a une espérance nulle, une variance égale à 1 et un écart type égal à 1. Ainsi nous obtenons des données indépendantes de l'unité ou de

l'échelle choisie et des variables ayant même moyenne et même dispersion. Cela équivaut à un changement d'unité qui n'a pas d'incidence sur les profils de variation. En d'autres termes, les valeurs des coefficients de corrélation entre les variables centrées réduites demeurent toutes identiques à ce qu'elles étaient avant l'opération de centrage et de réduction.

5.2.4. Phase d'apprentissage

Nous avons utilisé la méthode d'apprentissage supervisé basée sur la technique de Régularisation Bayésienne (RB) associée à l'algorithme de Levenberg-Marquardt (LM), afin de minimiser l'erreur quadratique de sortie et pour ajuster les poids synaptiques. Nous notons que les caractéristiques d'entrée sont normalisées entre [-1, 1], afin qu'elles aient le même impact sur le modèle. Nous avons arrêté l'algorithme de rétropropagation à un nombre fixé de 50 itérations de la base d'apprentissage. Ce nombre est jugé suffisant pour permettre une convergence du réseau.

5.2.5. Phase de Classification

Pour les tests de classification, nous avons suivi les étapes allant de la lecture des sons enregistrés jusqu'aux tests de classification, en incluant les différentes étapes de la détection des frontières des mots, de la préaccentuation, et de l'extraction des caractéristiques acoustiques. Pour valider nos tests de classification, nous avons enregistré un ensemble de fichiers sonores contenant des PN_{orm} et des PP_{ath}, différents de ceux utilisés pour l'apprentissage. Ces enregistrements ont été réalisés au laboratoire avec un milieu naturel contenant du bruit environnant. Il faudra tout de même noter que pour comparer les matrices des paramètres d'entrée d'apprentissage avec celles des paramètres d'entrée tests, ainsi que les vecteurs de sortie d'apprentissage avec ceux de sortie tests, nous devrons avoir des vecteurs et matrices de même dimensions. Pour cela, nous utilisons une fonction sous Matlab 2007, qui nous permet de choisir des tailles fixes pour tous les vecteurs et matrices correspondant à une durée de 160 ms, jugée comme taille suffisante pour englober les consonnes et les voyelles prononcées avec de longues durées. Cette taille est également importante dans l'extraction des paramètres Jitter, Jitter factor, Shimmer et Shimmer factor. Enfin, nous avons appliqué une procédure de distance minimale, afin de comparer la matrice des paramètres acoustiques du fichier test avec les matrices des paramètres acoustiques de l'ensemble des fichiers d'apprentissage.

Une simulation a été réalisée avec l'ensemble de fichiers tests, afin d'évaluer les performances de notre système lorsque les échantillons inconnus sont présentés à l'entrée. Le comportement de notre classificateur automatique est évalué en termes de pourcentage de classification correcte de l'ensemble de fichiers tests présentés à l'entrée. La méthode classique pour calculer le taux de reconnaissance (TR) des fichiers tests est donnée par l'équation classique :

$$TR\ (\%) = \left(\frac{Cas\ correctes}{Total}\right) . 100 \qquad (5.7)$$

5.3. Résultats Expérimentaux

Nous avons réalisé une classification automatique sur deux types de pathologies: l'une neurologique (PP_{ark}) et l'autre physiologique ($PŒ_{so}$). Dans une première étape, nous avons fait une classification automatique des PP_{ark} par rapport aux PN_{orm}. Dans une seconde étape, nous avons fait une classification automatique des $PŒ_{so}$ par rapport aux PN_{orm}. En dernière étape, nous avons fait une classification à partir d'un mélange de PP_{ark} , de $PŒ_{so}$ et de PN_{orm}. Dans les deux étapes, nous avons fait varier les **V**ecteurs d'**E**ntrée (**VE**) pour relever les paramètres acoustiques les plus pertinents pour la discrimination des paroles pathologiques.

5.3.1. Cas de la parole parkinsonienne (PP_{ark})

Nous injectons les paramètres acoustiques choisis à l'entrée du système de classification, en tenant compte des poids de chaque fichier de parole mémorisé durant la phase d'apprentissage. Pour évaluer les performances de notre système, une même procédure de classification a été appliquée en changeant à chaque fois les caractéristiques acoustiques d'entrée. Ainsi à chaque étape, nous faisons un apprentissage en utilisant une matrice donnée de vecteurs acoustiques extraits à partir d'un ensemble de 80 fichiers apprentissage (40 fichiers parkinsoniens et 40 fichiers normaux), puis nous passons à la classification automatique par "TDNN", en utilisant d'autres fichiers de paroles inconnus du dictionnaire d'apprentissage, soit 80 autres fichiers répartis équitablement entre les PP_{ark} et les PN_{orm}. Comme vecteurs acoustiques d'entrée, nous avons appliqué respectivement les paramètres Jitter (J), Jitter factor (J_f), Shimmer (S), Shimmer factor (S_f), Taux de Passage par Zéro (TPZ), Energie du signal (E) et enfin les coefficients MFCC. Nous obtenons les matrices de confusion suivantes :

Tableau 5.1 : Classification PP_{ark}, avec VE : (J, J_f, S, S_f, TPZ)

J, J_f, S, S_f, TPZ	**Normale**	**Parkinsonienne**
Normale	**29**	11
Parkinsonienne	14	**26**

Tableau 5.2 : Classification PP_{ark}, avec VE : (J, J_f, S, S_f, E)

J, J_f, S, S_f, E	**Normale**	**Parkinsonienne**
Normale	**37**	03
Parkinsonienne	02	**38**

Tableau 5.3 : Classification PP_{ark}, avec VE : (J, J_f, S, S_f, TPZ, E)

J, J_f, S, S_f, TPZ, E	**Normale**	**Parkinsonienne**
Normale	**34**	06
Parkinsonienne	04	**36**

Tableau 5.4 : Classification PP_{ark}, avec VE : (J, J_f, S, S_f, TPZ, E, MFCC)

J, J_f, S, S_f, TPZ, E, MFCC	**Normale**	**Parkinsonienne**
Normale	**32**	08
Parkinsonienne	01	**39**

En récapitulant les résultats, nous obtenons les résultats suivants, comme taux de reconnaissance TR (%) :

Tableau 5.5 : Taux de Reconnaissance (TR) des PP_{ark}, avec différents VE

	Paramètres acoustiques			
Cas	**J, J_f, S, S_f, TPZ**	**J, J_f, S, S_f, E**	**J, J_f, S, S_f, TPZ, E**	**J, J_f, S, S_f, TPZ, E, MFCC**
Normale (TR %)	72.50	92.50	85.00	80.00
Parkinsonienne (TR %)	65.00	95.00	90.00	97.50

VE : Vecteurs d'entrée.

Selon les résultats obtenus, nous pouvons dire que le système proposé, basé sur les RNA, donne un bon pourcentage de reconnaissance de la PP_{ark} par rapport à la PN_{orm}. De plus, en comparant ces différents résultats lorsque nous faisions varier les paramètres acoustiques, nous pouvons dire que les meilleurs résultats sont ceux obtenus en utilisant un choix de l'ensemble VE : (J, J_f, S, S_f, E), en tant que caractéristiques acoustiques. Ainsi, ces dernières caractéristiques sont pertinentes dans la discrimination de la PP_{ark} par rapport à la PN_{orm}. Ce qui montre clairement que la voix de patients parkinsoniens présente un pitch très irrégulier, une raucité de la voix, un volume de parole réduit et une intensité irrégulière. Ces résultats confirment les résultats des études acoustiques précédentes rapportées dans la littérature sur la maladie de Parkinson [13, 15, 88]. Encore plus important, notre système de classification répond positivement au bon choix des vecteurs acoustiques d'entrée.

5.3.2. Cas de la Parole Œsophagienne ($PŒ_{so}$)

Nous procédons de la même façon que pour le cas de la discrimination de la parole parkinsonienne par rapport à la parole normale. Pour évaluer les performances de notre système, une même procédure de classification a été appliquée en changeant à chaque fois les caractéristiques acoustiques d'entrée (VE). Ainsi à chaque étape, nous faisons un apprentissage en utilisant une matrice donnée de vecteurs acoustiques extraits à partir d'un ensemble de 80 fichiers apprentissage (40 fichiers $PŒ_{so}$ et 40 fichiers PN_{orm}), puis nous passons à la classification automatique, en utilisant d'autres fichiers de paroles inconnus du dictionnaire d'apprentissage, soit 80 autres fichiers répartis équitablement entre PN_{orm} et $PŒ_{so}$.

Tableau 5.6 : Classification $PŒ_{so}$, avec VE : (J, J_f, S, S_f, TPZ)

J, J_f, S, S_f, TPZ	**Normale**	**Œsophagienne**
Normale	**36**	04
Œsophagienne	02	**38**

Tableau 5.7 : Classification $PŒ_{so}$, avec VE : (J, J_f, S, S_f, E)

J, J_f, S, S_f, E	**Normale**	**Œsophagienne**
Normale	**39**	01
Œsophagienne	01	**39**

Tableau 5.8 : Classification PŒ$_{so}$, avec VE : (J, J_f, S, S_f, TPZ, E)

J, J_f, S, S_f, TPZ, E	**Normale**	**Œsophagienne**
Normale	**36**	04
Œsophagienne	01	**39**

Tableau 5.9 : Classification PŒ$_{so}$, avec VE : (J, J_f, S, S_f, TPZ, E, MFCC)

J, J_f, S, S_f, TPZ, E, MFCC	**Normale**	**Œsophagienne**
Normale	**35**	05
Œsophagienne	00	**40**

En récapitulant les résultats, nous obtenons les résultats suivants concernant le pourcentage de reconnaissance (Tableau 5.10).

Tableau 5.10 : Taux de Reconnaissance (TR) des PŒ$_{so}$ avec différents VE

	Paramètre acoustiques			
Cas	**J, J_f, S, S_f, TPZ**	**J, J_f, S, S_f, E**	**J, J_f, S, S_f, TPZ, E**	**J, J_f, S, S_f, TPZ, E, MFCC**
Normale (TR %)	90.00	97.50	90.00	87.50
Œsophagienne (TR %)	95.00	97.50	97.50	100.00

En comparant les résultats obtenus lorsque nous faisions varier les différents paramètres acoustiques, nous pouvons dire que les meilleurs résultats sont ceux obtenus en utilisant un choix de l'ensemble VE : (Jitter, $Jitter_{factor}$, Shimmer et $Shimmer_{factor}$, Energie) en tant que caractéristiques acoustiques. Soit la même matrice de paramètres acoustiques relevée pour le cas de la classification des PP_{ark}. La matrice de VE : (J, J_f, S, S_f, TPZ, E, MFCC) donne également un même pourcentage de classification (95.00%) avec un TR de 100% des voix pathologiques. Il reste que le TR des PN_{orm} est relativement faible (87.50%) et c'est également le cas lors de la classification de la PP_{ark} (80.00%). Nous pouvons ainsi déduire des résultats trouvés que les paramètres acoustiques Jitter, $Jitter_{factor}$, Shimmer, $Shimmer_{factor}$ et Energie sont très pertinents dans la discrimination des PP_{ath}. Ceci est conforme aux résultats trouvés dans l'analyse acoustique de la PP_{ath}.

5.3.3. Cas d'un mélange des corpus des PP_{ark} et $PŒ_{so}$

Nous avons essayé d'appliquer la même procédure de classification mais cette fois, en faisant mélanger l'ensemble des corpus des fichiers contenant les PP_{ark} et $PŒ_{so}$ ainsi que les fichiers de PN_{orm}. Ceci devient très difficile à discriminer, car les PP_{ath} que nous avons étudié précédemment présentent les mêmes vecteurs acoustiques pertinents VE pour obtenir le Taux le plus élevé de classification, d'où l'importance d'autres VE qui permettront de séparer les deux paroles pathologiques. Les résultats trouvés sont illustrés dans les tableaux suivants :

Tableau 5.11: Classification Ensemble des Pathologies, avec VE : (J, J_f, S, S_f, TPZ)

J, J_f, S, S_f, TPZ	**Normale**	**Parkinsonienne**	**Œsophagienne**
Normale	**29**	08	03
Parkinsonienne	14	**21**	05
Œsophagienne	01	10	**29**

Tableau5.12 : Classification Ensemble des Pathologies, avec VE : (J, J_f, S, S_f, E)

J, J_f, S, S_f, E	**Normale**	**Parkinsonienne**	**Œsophagienne**
Normale	**37**	03	00
Parkinsonienne	02	**38**	00
Œsophagienne	02	32	**06**

Tableau 5.13 : Classification Ensemble des Pathologies, avec VE : (J, J_f, S, S_f, TPZ, E)

J, J_f, S, S_f, TPZ, E	**Normale**	**Parkinsonienne**	**Œsophagienne**
Normale	**32**	05	03
Parkinsonienne	04	**30**	06
Œsophagienne	01	10	**29**

Tableau 5.14 : Classification Ensemble des Pathologies, avec
VE : (J, J_f, S, S_f, TPZ, E, MFCC)

J, J_f, S, S_f, TPZ, E, MFCC	**Normale**	**Parkinsonienne**	**Œsophagienne**
Normale	**30**	07	03
Parkinsonienne	02	**37**	01
Œsophagienne	00	08	**32**

En récapitulant les résultats, nous obtenons les résultats suivants (Tableau 5.15).

Tableau 5.15 : Taux de Reconnaissance (TR) des PP_{ark} et $PŒ_{so}$ mélangées avec les PN_{orm}

	Paramètres acoustiques			
Cas	**J, J_f, S, S_f, TPZ**	**J, J_f, S, S_f, E**	**J, J_f, S, S_f, TPZ, E**	**J, J_f, S, S_f, TPZ, E, MFCC**
Normale (TR %)	72.50	92.50	80.00	75.00
Parkinsonienne (TR %)	52.50	95.00	75.00	92.50
Œsophagienne (TR %)	72.50	15.00	72.50	80.00

La matrice de VE : (J, J_f, S et S_f, E), considérée comme assez représentative lorsque nous avons classifié isolément les deux groupes de PP_{ath}, donne des résultats médiocres lorsque nous mélangeons l'ensemble des fichiers pathologiques des deux groupes (uniquement 15% de reconnaissance de la PP_{ark}). Ce qui explique d'ailleurs le fait que les deux pathologies étudiées présentent beaucoup de traits acoustiques communs pertinents.

La figure 5.5 montre que la $PŒ_{so}$ est mieux caractérisée par le paramètre acoustique TPZ que par l'énergie. En changeant le vecteur acoustique Energie par le TPZ, le TR de la $PŒ_{so}$ passe brusquement d'un taux assez bas de 15.00% à un taux élevé de 72.50%. A l'inverse, la PP_{ark} est mieux représentée par l'énergie (95.00%) que par le TPZ (52.50%). Ceci est conforme à l'analyse acoustique que nous avons réalisée sur les deux paroles pathologiques $PŒ_{so}$ et PP_{ark}. La $PŒ_{so}$ présente une parole trop bruitée avec un fort taux de HNR, donc logiquement le TPZ est trop grand comparé à la PN_{orm} et à l'autre cas pathologique PP_{ark}. Cette dernière est caractérisée par une voix faible donc une valeur d'énergie assez basse.

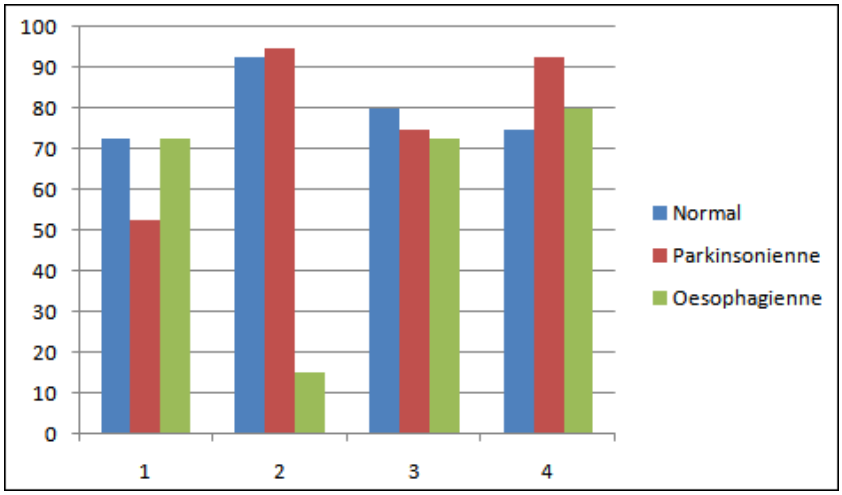


Figure 5.5 : Taux de Reconnaissance (TR) des PN_{orm}, $PŒ_{so}$ et PP_{ark} selon le choix des paramètres acoustiques

1: (J, J_f, S, S_f, TPZ)

2: (J, J_f, S, S_f, E)

3: (J, J_f, S, S_f, TPZ, E)

4: (J, J_f, S, S_f, TPZ, E, MFCC)

Le meilleur taux de reconnaissance est obtenu en exploitant la matrice comprenant les VE : (J, J_f, S, S_f, TPZ, E et MFCC). Ce qui est comme même très appréciable compte tenu du mélange que nous avons procédé sur les corpus pathologiques, en ce sens qu'il est assez difficile de départager les PP_{ath}. Ces dernières présentent des valeurs assez proches pour la majorité des paramètres acoustiques, tels que le Jitter, le Shimmer et l'Energie. Ce qui confirme donc la possible exploitation des paramètres MFCC dans la discrimination des paroles pathologiques, comme rapporté dans quelques études antérieures [82, 84, 89]. Pour la parole parkinsonienne, nous avons relevé lors de l'analyse acoustique que le niveau des valeurs formantiques est plus faible comparé au cas normal, car le patient tend à emphatiser sa prononciation. Par contre, nous avons relevé pour le cas de la parole œsophagienne, une augmentation des valeurs des formants qui peut être probablement expliquée par le fait que la distance entre le segment NVP et la première cavité de l'appareil vocal (cavité pharyngale) est modifiée par la présence du trachéostome. Ainsi, le paramètre acoustique relatif aux formants peut être considéré comme paramètre intéressant de discrimination des deux paroles pathologiques. Dans notre système de classification automatique, ceci est assuré par l'ajout du paramètre acoustique MFCC, qui a

effectivement permit une augmentation sensible du taux de classifications des deux paroles pathologiques par rapport à la parole normale. Notons que les MFCC ont été appliqués dans une étude récente pour la reconnaissance des voyelles à l'état isolé et après Laryngectomie Totale [90]. Selon les résultats de leur étude, les auteurs ont obtenu un Taux de Reconnaissance de la parole alaryngée de 98.00% en se basant sur les MFCC, alors que le taux n'est que de 75.00 % lorsqu'ils utilisent la paire formantique F_1-F_2. C'est dire toute l'importance de la prise en compte de ces coefficients MFCC dans la discrimination des paroles pathologiques. Les premiers coefficients donnent les caractéristiques de l'enveloppe spectrale. Ils révèlent ainsi les configurations des cavités de résonance du conduit vocal liées à la parole/au locuteur et permettent de donner des informations sur les formants.

5.4. Conclusion

Nous avons détaillé les différentes étapes pour reconnaître et classifier automatiquement deux PP_{ath} d'origine aussi bien fonctionnelle qu'organique. A travers les résultats expérimentaux, nous avons montré la contribution de la méthode des RNA pour la reconnaissance et la classification automatique de la PP_{ath}. Pour ce faire, nous avons appliqué les réseaux TDNN avec la technique d'apprentissage supervisé, exploitant les avantages de la Régularisation Bayésienne (RB), combinée à l'algorithme de Levenberg-Marquardt (LM). Cette méthode nous a permis d'avoir des TR (%) appréciables, notamment lorsque les différentes PP_{ath} sont prises isolement (95.00% pour la parole parkinsonienne et 97.50% pour la parole œsophagienne, en tenant compte des paramètres acoustiques VE : (Jitter, Jitter $_{factor}$, Shimmer et Shimmer $_{factor}$, Energie) à l'entrée du système de classification élaboré). Par contre, cette matrice de vecteur donne des résultats médiocres lorsque nous mélangeons les deux groupes de PP_{ath}. Dans ce dernier cas, les meilleurs résultats obtenus sont relatifs à la matrice comprenant les VE : (Jitter, Jitter $_{factor}$, Shimmer, Shimmer $_{factor}$, Taux de Passage par Zéro TPZ, Energie et Coefficients cepstraux MFCC).

Cette étude nous a montré que la reconnaissance de la parole pathologique est une tâche difficile à réaliser du fait de la complexité et de la variabilité de la parole humaine et surtout du rapprochement des valeurs des paramètres acoustiques pathologiques. Un choix adéquat de ces paramètres acoustiques permettra une diminution du taux d'erreur de classification et ainsi une meilleure discrimination automatique lors d'un mélange de diverses paroles pathologiques.

CONCLUSIONS GENERALES ET PERSPECTIVES

Notre travail a porté sur l'analyse acoustique et la Classification Automatique des Paroles Pathologiques (PP_{ath}) par rapport à la Parole Normale (PN_{orm}). Pour réaliser cet objectif, nous avons étudié la PP_{ath} sous ses deux aspects :

- fonctionnel ou neurologique, car pour ce genre de pathologies, l'appareil phonatoire est intact. Le trouble provient de déficiences fonctionnelles neurologiques telles que l'aphasie, les maladies de Parkinson et d'Alzheimer, etc. L'exemple que nous avons utilisé dans le cadre de notre étude est la Parole Parkinsonienne (PP_{ark}).
- organique ou physiologique, car le trouble est causé par de déficiences au niveau de l'appareil phonatoire telles qu'une ablation du larynx, une déformation de la langue, de la luette, une anomalie congénitale des cordes vocales, etc. Une des pathologies d'origine organique les plus connues est la Parole Œsophagienne ($PŒ_{so}$), obtenue après Laryngectomie Totale due à un cancer du larynx. C'est cet exemple que nous avons étudié dans le cadre de ce travail.

Dans une première étape, nous avons mené une analyse acoustique approfondie sur les deux cas (PP_{ark} et $PŒ_{so}$). Pour cela, nous avons extrait des paramètres spécifiques tels que le Jitter, le Shimmer, le HNR et l'intensité du signal de parole. Ces paramètres nous ont permis une caractérisation acoustique des deux cas. De telles études offrent ainsi aux spécialistes de la voix ou de la parole, des données objectives qui permettent d'estimer le degré de perturbation et d'apporter les solutions nécessaires pour y remédier.

Dans une dernière étape, nous avons réalisé une classification automatique de la PP_{ath} par un Réseau de Neurones Artificiels (RNA). Le type de réseau que nous avons utilisé est le réseau à délais temporels TDNN (Time Delay Neural Network). Cette méthode nous a permis de discriminer, avec des taux appréciables, les deux PP_{ath} par rapport à la PN_{orm}. Pour les paramètres acoustiques, nous avons exploité respectivement le Jitter, le $Jitter_{factor}$, le Shimmer, le $Shimmer_{factor}$, le Taux de Passage par Zéro, l'Energie et enfin les Coefficients cepstraux MFCC.

Ce travail a pour objectif la caractérisation des paroles pathologiques en vue de leur exploitation en réhabilitation de la parole, la conduite de diagnostics automatiques et l'établissement de systèmes expert permettant de caractériser de façon fiable les anomalies vocales en milieu hospitalier algérien. Le résultat attendu d'une telle étude est de fournir également une meilleure compréhension des effets des troubles neurologiques et physiologiques sur la production de la parole, d'un point de vue acoustique, enrichissant

pour le diagnostic des cliniciens mais également à des fins d'enseignement des troubles de la parole destinés aux orthophonistes et autres spécialistes du domaine. L'exploration des voix et paroles pathologiques étant un objectif de recherche clinique d'une importance particulière ces dernières années. Cette exploration permet la mise en place de procédures et techniques d'évaluation des caractéristiques de la voix et de la parole afin de déterminer de façon objective leur écart par rapport aux valeurs normales. Soulignons qu'en Algérie, peu de travaux ont été réalisés dans ce domaine.

En conclusion, le système de classification automatique élaboré dans le cadre de ce travail nous a permis d'avoir des TR (%) appréciables des PP_{ath} par rapport à la PN_{orm}, lorsque ces dernières sont prises dans un contexte isolé (95.00 % pour les **P**aroles **Park**insoniennes (**PP_{ark}**) et 97.50 % pour les **P**aroles **Œso**phagiennes (**$PŒ_{so}$**)) et des taux respectifs de 92.50 % et 80.00 % pour les deux PP_{ath} lorsqu'elles sont mélangées. Pour le cas du contexte isolé, les meilleurs résultats sont obtenus en exploitant les paramètres acoustiques Jitter, Jitter $_{factor}$, Shimmer et Shimmer $_{factor}$ et Energie. Par contre, pour le cas d'un mélange des PP_{ath}, les meilleurs résultats sont liés au choix des paramètres Jitter, Jitter $_{factor}$, Shimmer, Shimmer $_{factor}$, Taux de Passage par Zéro TPZ, Energie et Coefficients cepstraux MFCC. Il reste que cette étude nous a montré toute la difficulté pour classifier automatiquement un mélange de différentes paroles pathologiques, du fait de la complexité et de la variabilité de la parole humaine et surtout du rapprochement des valeurs des paramètres acoustiques pathologiques, tels le Jitter et le Shimmer.

En perspectives, d'autres paramètres peuvent contribuer sensiblement à l'amélioration du Taux de Reconnaissance. Dans le cadre d'un projet à long terme, il est également important de prendre, des échantillons de locuteurs et paroles pathologiques plus importants pour avoir des résultats plus représentatifs et plus fiables du système de classification élaboré.

Nous ne pouvons terminer notre travail sans souligner certaines lacunes relevées dans la prise en charge des patients et que nous avons jugé utile de mentionner à la fin de cette thèse :

- une absence de formation des orthophonistes dans la manipulation de logiciels d'analyse acoustique ;
- un manque flagrant de coopération entre l'orthophoniste en milieu hospitalier algérien, l'ingénieur, chercheur phonéticien, et acousticien dans le laboratoire de

recherche, et enfin le professeur enseignant à l'université. Une étroite collaboration entre ces institutions permettra une meilleure prise en charge des patients ;

- une utilisation exclusive de l'oreille (ouïe) pour évaluer l'effet de la réhabilitation vocale dans les hôpitaux algériens. L'évaluation de la voix pathologique est principalement basée sur la perception subjective des cliniciens sans aucune analyse acoustique de la PP_{ath}. Certes, cette dernière ne peut remplacer le travail traditionnel de l'orthophoniste, mais elle peut être un support de données objectives qui peuvent l'aider considérablement dans la rééducation des patients. Cette analyse rend objectif ce qui échappe parfois à l'audition de l'orthophoniste (jugement perceptif).

REFERENCES BIBLIOGRAPHIQUES

[1] C. Jacquier, Étude d'indices acoustiques dans le traitement temporel de la parole chez des adultes normo-lecteurs et des adultes dyslexiques, Thèse de Doctorat Neurosciences et Cognition, Université de Lyon, France, 2008.

[2] K. Ferrat and M. Guerti, Synthèse de la parole en Arabe Standard. Cas des phénomènes spécifiques à la langue, Colloque International en Traductologie et TAL, Université d'Oran, Algérie, 9-11 avril 2007.

[3] K. Ferrat. Acoustical study of the Tachdid and the Idgham in Standard Arabic-Application for speech synthesis. International Conference IEEE, Sciences of Electronic, Technologies of Information and Telecommunication, SETIT2005, Susa, Tunisia, 2005. http://www.setit.rnu.tn/last_edition/setit2005/trait-signal/200.pdf

[4] K. Ferrat, K. Baazi and M. Guerti, Etude Acoustico-Articulatoire de l'Emphase en Arabe Standard, Colloque International Traductologie et TAL, Université d'Oran, Algérie, 2007.

[5] M. Guerti, Contribution à la synthèse de la parole par diphones en Arabe Standard, Thèse de Magister en Electronique Acoustique et Physiologique de la Parole. Université d'Alger, Algérie, 1983.

[6] E.R. Dorsey, R. Constantinescu, J.P. Thompson, K.M. Biglan, R.G. Holloway, K. Kieburtz, F.J. Marshall, B.M. Ravina, G. Schifitto, A. Siderowf and CM. Tanner, Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030, Neurology, 68(5), pp.384-386, 2007.

[7] A. Ascherio, H. Chen, M. Weisskopf, E. O'Reilly, M. McCullough, E. Calle, M. Schwarzschild and M. Thun, Pesticide exposure and risk for Parkinson's disease, Annals of Neurology, 60, pp.197-203, 2006.

[8] A. Wang, S. Costello, M. Cockburn, X. Zhang, J. Bronstein and B. Ritz, Parkinson's disease risk from ambient exposure to pesticides, European Journal of Epidemiology, 26(7), pp.547-555, 2011. DOI : 10.1007/s10654-011-9574-5.

[9] J. Parkinson, An essay on the shaking palsy ([Reprod.]), Editions Whittingham and Rowland (London), 18717. http://gallica.bnf.fr/ark:/12148/bpt6k987658

[10] J.A. Logeman, H.B. Fisher and B. Boshes, Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients, Journal of Speech and Hearing Disorders, 43, pp.47-57, 1978.

[11] L.C. Gracco, V.L. Gracco, A. Löfqvist and K. Marek, An Aerodynamic Evaluation of Parkinsonian Dysarthria : Laryngeal and Supralaryngeal Manifestations, Haskins Laboratories Status Report on Speech research, SR-111-112, pp.103-110, 1992.

[12] S. Blanc and A. Charras, Application d'une grille d'auto-évaluation du Handicap Vocal (VHI) à la dysarthrie parkinsonienne : Normalisation, Validation, Mémoire de Licence en Orthophonie, Université de Lilles, France, juin 2005.

[13] K. Rigaldie, J.L. Nespoulous and N. Vigouroux, Dysprosody in Parkinson's Disease, An acoustic study based on tonal phonology and the INTSINT system, Speech Prosody 2004, Nara, Japan, March 23-26, 2004.

[14] S. Skodda and U. Schlegel, Speech rate and rhythm in Parkinson's disease, Journal of Movement Disorders, 23, pp.985 - 992, 2008.

[15] R.J. Holmes, J.M. Oates, D.J. Phyland and A.J. Hughes, Voice characteristics in the progression of Parkinson's disease, International Journal of Language & Communication Disorders, 35(2), pp.407- 418, 2000.

[16] R. Christophe, La Laryngectomie Totale. La chirurgie du larynx 20 ans après, 1O^{èmes} journées Evolution de la prise en charge des tumeurs du pharyngo-larynx, Clinique mutualiste de la Sagesse, Rennes, France, 1995.

[17] A.A. Makitie, R. Niemensivu, A. Juvas, L.M. Aaltonen, L. Back and H. Lehtonen, Postlaryngectomy voice restoration using a voice prosthesis: a single institution's ten-year experience, Annals of Otology, Rhinology, and Laryngology, 112(12), pp.1007-10, 2003.

[18] G. Mamelle, C. Domenge and E. Bretagne, Réinsertion et surveillance médicale du laryngectomisé, EMC (Encyclopedie Medico-Chirurgicale) / Elsever - Paris, Oto-Rhino- Laryngologie, Institut Gustave Roussy, France, 1998.

[19] K. Ferrat, Analyse acoustique et évaluation d'un cas de rééducation de laryngectomie totale en milieu hospitalier algérien, Actes des VIIèmes RJC Parole, ILPGA, Paris, France, pp.80-83, 2007.

[20] I. Hocevar-Boltezar and M. Zargi, Communication after laryngectomy, Radiology and Oncology, 35(4), pp.249-254, 2001.

[21] T. Cervera, J.L. Miralles and J. González-Alvarez, Acoustical analysis of Spanish vowels produced by laryngectomized subjects, Journal of Speech Language and Hearing Research, 44, pp.988-96, 2001.

[22] M. Brockmann, M.J. Drinnan, C. Storck and P.N. Carding, Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task, Journal of Voice, 25(1), pp. 44-53, 2011.

[23] F. Klingholz and F. Martin, Quantitative spectral evaluation of shimmer and jitter, Journal of Speech Language and Hearing Research, 28, pp.169-174, 1985.

[24] J. Munoz, E. Mendoza, M.D. Fresneda, G. Carballo and P. Lopez, Acoustic and perceptual indicators of normal and pathological voice, Folia phoniatrica et logopaedica, 55, pp.102-114, 2003.

[25] J. Kreiman and B.R. Gerratt, Perception of aperiodicity in pathological voice, Journal of the Acoustical Society of America, 117, pp.2201-2211, 2005.

[26] Kay Elemetrics, Multi-Dimensional Voice Program, Model 5105 Lincoln Park, NJ : Kay Elemetrics Corporation, 2008.

[27] A.V. Oppenheim and R.W. Shafer, From frequency to Quefrency : A history of the cepstrum, IEEE Signal Processing Magazine, pp.95-106, 2004.

[28] A.M. Noll, Cepstrum Pitch Determination, Journal of the Acoustical Society of America, 41(2), pp.293-309, 1967.

[29] W. Hess, Pitch and voicing determination of speech with an extension toward music signals, In Benesty M, Sondhi M, Huang Y (Eds.) Springer handbook of speech processing. Springer-Verlag, 2008.

[30] P.J. Murphy and O.O. Akande, Noise Estimation in Voice Signals Using Short-term Cepstral Analysis, Journal of the Acoustical Society of America, 121(3), pp.1679-1690, 2007.

[31] E. Yumoto, W.J. Gould and T. Bear, Harmonic-to-noise ratio as index of the degree of hoarseness, Journal of the Acoustical Society of America, 71(3), pp.1544-1550, 1982.

[32] K. Ferrat, Analyse acoustique et évaluation de la rééducation de la maladie de Parkinson dans le milieu hospitalier algérien, Revue Al-Lissaniyat, CRSTDLA, Algérie, 14-15, 2009, ISSN: 1112-4393.

[33] K. Ferrat and M. Guerti, Analyse acoustique de la parole œsophagienne en milieu hospitalier algérien, Proceedings Journées d'études Algéro-Françaises de doctorants en signal-image & applications, JEAFD2012, Ecole Nationale Polytechnique, Algérie, pp.41-46, décembre 2012.

[34] N. Zellal, Introduction à la phonétique orthophonique arabe, Editions OPU, Algérie, 1984.

[35] http://www.praat.org

[36] H. Liu, M. Wan, S. Wang, X. Wang and C. Lu, Acoustic characteristics of Mandarin esophageal speech, Journal of the Acoustical Society of America, 118, pp.1016-1025, 2005.

[37] M. Mięsikowska and L. Radziszewski, Acoustical analysis of Polish vowels of esophageal speakers, Measurements Automation and Monitoring (PAK), 57(12), pp. 1504-1507, 2011.

[38] R.A. Kazi, V.M.N. Prasad, J. Kangalingam, C.M. Nutting, P. Clarke, P. Rhys-Evans and K.J. Harrington, Assessment of the Formant Frequencies in Normal and Laryngectomized Individuals Using Linear Predictive Coding, Journal of Voice, 21(6), pp. 661-668, 2007.

[39] M.L. Ng and R. Chu, An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese, Folia phoniatrica et logopaedica, 61, pp.97-104, 2009.

[40] L. Crevier-Buchman, J. Vaissière, S. Maeda and D. Brasnu, Etude de l'intelligibilité des consonnes du français après laryngectomie partielle supracricoidienne, Revue de Laryngologie Otologie Rhinologie, 123, pp.307-310, 2002.

[41] H. Hirose, Voicing Distinction in esophageal Speech, Acta Oto-Laryngologica Supplementum, 524, pp.56-63, 1996.

[42] K. Ferrat and M. Guerti, A Study of sounds Produced by Algerian Esophageal Speakers, African Health Sciences, 4, pp.452-458, 2012. http://dx.doi.org/10.4314/ahs.v12i4.9

[43] K. Ferrat and M. Guerti, Apprentissage et Reconnaissance Automatique de la Parole par Réseaux de Neurones Artificiels, Revue Sciences de l'homme, ISSN:1112–8054, Laboratoire SLANCOM, Université d'Alger2, Algérie, 4, pp.57-71, 2011.

[44] G. Dreyfus, M. Samuelides, J.-M. Martinez, M. B. Gordon, F. Badran, S. Thiria and L. Hérault, Réseaux de neurones : Méthodologie et Application, Editions Eyrolles, 408 pages, 2e edition, France, ISBN 2-212-11 464-8, 2004.

[45] L. Fausett, Fundamentals of Neural Networks, Prentice-Hall, Englewood Cliffs, NJ, 1994.

[46] P. Borne, M. Benrejeb and J. Haggège, Les réseaux de neurones : présentation et applications, Editions Technip, Paris, France, 150 pages, 2007.

[47] D.K. Chaturvedi, Artificial neural network and supervised learning, Soft Computing Techniques and its Applications in Electrical Engineering, Springer Verlag, Berlin Heidelberg, pp.23-50, 2008.

[48] D.O. Hebb, The Organization of Behavior : A Neuropsychological Theory, Editions Wiley, New York, USA, 1949.

[49] P.J. Werbos, Backpropagation through time : What it does and how to do it, Proceedings of the IEEE, 78(10), pp.1550-1560, 1990.

[50] J.J. More, The Levenberg-Marquardt Algorithm : Implementation and Theory, Numerical Analysis, edited by. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, pp. 105-116, 1977.

[51] F.D. Foresee and M.T. Hagan, Gauss-Newton Approximation to Bayesian Regularization, Proceedings of the 1997 International Joint Conference on Neural Networks, pp. 1930-1935, 1997.

[52] M. Fun and M.T. Hagan. Levenberg-Marquardt Training for Modular Networks, International Conference on Neural Networks, pp. 468-473, 1996.

[53] D.J.C. MacKay, A practical Bayesian framework for backpropagation networks, Neural Computation, 4(3), pp.448–472, 1992.

[54] T.Y. Kwok and D.Y. Yeung, Bayesian regularization in constructive neural networks, Lecture Notes in Computer Science, 1112, Artificial Neural Networks - ICANN 96 International Conference Bochum, Germany, July 16-19, 1996.

[55] F. Burden and D. Winkler, Bayesian regularization of neural networks, Methods in Molecular Biology, 458, pp.25-44, 2008.

[56] D. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear parameters, SIAM Journal on Applied Mathematics, 11, pp.431-441, 1963.

[57] C. Kanzow, N. Yamashita and M. Fukushima, Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, Journal of Computational and Applied Mathematics, 173(2), pp.321-343, 2005.

[58] J.P. Haton, Reconnaissance automatique de la parole : du signal à son interprétation, Editions Dunod, Paris, France, 2006.

[59] S. Young HMMs and Related Speech Recognition Technologies, Springer Handbook of Speech Processing, pp. 539-558, 2008. http://link.springer.com/referencework/10.1007/978-3-540-49127-9/page/2.

[60] A.Ganapathiraju, J.E. Hamaker and J. Picone, Applications of support vector machines to speech recognition, Signal Processing, IEEE Transactions on, 52(8), pp. 2358-2355, 2004

[61] K. Ferrat and M. Guerti, Reconnaissance des sons spécifiques de l'Arabe Standard par Réseaux de Neurones Artificiels, 3eme Colloque International en Traductologie et TAL, Université Es Senia, Oran, Algérie, 17-18 janvier 2010.

[62] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, USA, 1995.

[63] R.P. Lippmann, Review of Neural Networks for Speech Recognition, Neural Computation, 1, pp.1-38, 1989.

[64] W. Gevaert, G. Tsenov and V. Mladenov, Neural Networks used for Speech Recognition, Journal of Automatic Control, University of Belgrade, Serbia, 20, pp.1-7, 2010.

[65] G. Dede and M.H. Sazlı, Speech recognition with artificial neural networks, Digital Signal Processing, 3(20), p.763-768, 2010.

[66] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. Lang, Phoneme recognition using time-delay networks, IEEE Transactions on Acoustics, Speech and Signal Processing, 37(3), pp.328–339, 1989.

[67] K. Lang and G. Hinton, The development of the Time Delay Neural Network Architecture for Speech Recognition, Carnegie Mellon University TR CMU-CS-88-152, USA, 1988.

[68] E. Poisson, C. Viard-Gaudin and P.M. Lallican, Multi-modular architecture based on convolutional neural networks for online handwritten character recognition, in Proc. of ICONIP'02 - 9th International Conference on Neural Information Processing, IEEE Neural Network Society, 5, pp.2444-2448, 2002.

[69] K. Ferrat and M. Guerti, Classification of the Arabic Emphatic Consonants using Time Delay Neural Network, International Journal of Computer Applications, Published by Foundation of Computer Science, New York, USA, 80(10), pp: 1-6, 2013. http://dx.doi.org/10.5120/13894-9341.

[70] L. Devillers, Reconnaissance monolocuteur des phonèmes français au moyen de réseaux à masques temporels, Proceedings des XXVIIIème Journées d'Etudes sur la Parole, Montréal, Canada, 1990.

[71] A. Dev, Effect of retroflex sounds on the recognition of Hindi voiced and unvoiced stops, AI & Society, 23(4), pp.603-612, 2009. DOI : 10.1007/s00146-008-0179-9

[72] B.F. Yong and H.N. Ting, Speaker-Independent Vowel Recognition for Malay Children Using Time-Delay Neural Network, 5th Kuala Lumpur International Conference on Biomedical Engineering 2011. IFMBE Proceedings , 35, pp 565-568, 2011.

[73] N. Morgan and H.A. Bourlard, Neural Networks for Statistical Recognition of Continuous Speech, Proceedings of the IEEE, 83(5), May 1995.

[74] 1. Guyon, Y. Le cun, J. Denker and W. Hubbard, Design of a neural network character recogniser for a touch terminal, Pattern Recognition, 24(2), pp.105119, 1991.

[75] Y. Ming-Hsuan, N. Ahuja and M. Tabb, Extraction of 2D motion trajectories and its application to hand gesture recognition, IEEE Transactions on Pattern analysis and Machine Intelligence, 24(8), pp.1061-1074, 2002.

[76] N. Sun-Kuk and P. Jae-Young, A Study on the Detection Algorithm of QPSK Signal Using TDNN, Lecture Notes in Computer Sciences, LCS Springer-Verlag, 3973, pp. 135-143, 2006.

[77] P. Le Callet, C. Viard-Gaudin and D. Barba, A convolutional neural network approach for video quality assessment, IEEE Transactions on Neural Networks, 17(5), pp.1316-1327, 2006.

[78] H. Behbood, S.A. Seyyedsalehi, H.R. Tohidypour, M. Najafi and S. Gharibzadeh, A novel neural-based model for acoustic-articulatory inversion mapping, Neural Computing and Applications, 21(5), pp.935-943, 2012.

[79] R.T. Ritchings, M.A. Mcgillion and C.J. Moore, Pathological voice quality assessment using artificial neural network, Medical Engineering and Physics, 24, pp.561-564, 2002.

[80] J. Wang and C. Jo, Performance of Gaussian Mixture Models as a classifier for pathological voice, Proceedings of the 11th Australasian International Conference on Speech Science and Technology. Univ. Auckland, New Zealand, pp.165-169, 2006.

[81] T. Ananthakrishna, K. Shama and U.C. Niranjan, k-means nearest neighbour classifier for voice pathology, Proceedings of the IEEE India Conference (INDICON), Indian Institute of Technology, Kharagpur, India, pp.352-354, 2004.

[82] J.I. Godino-Llorente and P. Gomez-Vilda, Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors, IEEE Transactions on Biomedical Engineering, 51, pp.380-384, 2004.

[83] K. Umapathy, S. Krishnan, V. Parsa and D.G. Jamieson, Discrimination of pathological voices using a time-frequency approach, IEEE Transactions on Biomedical Engineering, 52, pp.421-430, 2005.

[84] O. Chia Ai, M. Hariharan, S. Yaacob and L. Sin Chee, Classification of speech dysfluencies with MFCC and LPCC features, Expert Systems with Applications, 39(2), pp. 2157-2165, 2012.

[85] A. Dev and P. Bansal, Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order Autocorrelation Coefficients, International Journal of Computer Applications, 10, pp.36-38, 2010.

[86] V. Tiwari, MFCC and its applications in speaker recognition, International Journal on Emerging Technologies, 1, pp.19-22, 2010.

[87] M.H. Beale, M.T. Hagan and M.H. Demuth, Neural network toolbox 7 user's guide, Natick, MA: MathWorks Inc, USA, 2010.

[88] J. Locco, La production des occlusives dans la maladie de Parkinson, Thèse de Doctorat de l'Université de Provence. Aix-Marseille I, France, 2005.

[89] R. Fraile, N. Saenz-Lechon, J.I. Godino-Llorente, V. Osma-Ruiz and C. Fredouille, Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex, Folia phoniatrica et logopaedica, 61, pp.146-152, 2009.

[90] R.W. Pietruch and A.D. Grzanka, Vowel Recognition of Patients after Total Laryngectomy using Mel Frequency Cepstral Coefficients and Mouth Contour, Journal of Automatic Control, University of Belgrade, 20(1), pp. 33-38, 2010.