REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

مخبر الإشارة و الإتصالات
Signal & Communications Lab.

Département d'Electronique
Laboratoire Signal et Communications

Thèse de Doctorat en Electronique

Présentée par :

Mr KABACHE Mahraz

Méthodes Acoustique et Neuronale Appliquées à la Laryngectomie et à la Paralysie Laryngée Unilatérale en Vue d'une Evaluation Objective de la Rééducation

Présentée et Soutenue le 13/11/2023:

Composition du jury :

M^{me} Latifa HAMAMI	Prof	ENP	Présidente
M^{me} Mhania GUERTI	Prof	ENP	Directrice de thèse
M^{r} Mourad ADNANE	Prof	ENP	Examinateur
M^{me} Nadjia BENBLIDIA	Prof	USD-Blida	Examinatrice
M^{me} Siham SAYOUD	Prof	USTHB	Examinatrice

ENP 2023

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique

Département d'Electronique
Laboratoire Signal et Communications

Thèse de Doctorat en Electronique

Présentée par :

Mr KABACHE Mahraz

Méthodes Acoustique et Neuronale Appliquées à la Laryngectomie et à la Paralysie Laryngée Unilatérale en Vue d'une Evaluation Objective de la Rééducation

Présentée et Soutenue le 13/11/2023:

Composition du jury :

M^{me} Latifa HAMAMI	Prof	ENP	Présidente
M^{me} Mhania GUERTI	Prof	ENP	Directrice de thèse
M^{r} Mourad ADNANE	Prof	ENP	Examinateur
M^{me} Nadjia BENBLIDIA	Prof	USD-Blida	Examinatrice
M^{me} Siham SAYOUD	Prof	USTHB	Examinatrice

ENP 2023

Remerciements

Je souhaite exprimer mon entière reconnaissance à ma Directrice de thèse Mme ***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, qui me fait l'honneur pour avoir accepté de présider mon jury de thèse.

Mes vifs remerciements vont également à Monsieur ***ADNANE Mourad***, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, qui me fait l'honneur d'avoir accepté de lire et d'évaluer ce travail.

Je tiens également à présenter mes remerciements à Mme ***BENBLIDIA Nadjia***, Professeur à l'Université Saâd Dahleb de Blida, qui me fait l'honneur d'avoir accepté de se joindre à ce jury comme examinatrice.

Mes vifs remerciements vont également à Mme ***FALEK Leila***, Professeur à l'Université des Sciences et de la Technologie Houari Boumediène - Alger, qui me font l'honneur d'être membres du jury de ma thèse.

Je tiens à exprimer toute ma gratitude à Mme ***SAYOUD Siham,*** Professeur à l'Université des Sciences et de la Technologie Houari Boumediène - Alger, qui me fait l'honneur d'être examinatrice et évaluer notre travail.

Je remercie vivement les Orthophonistes Praticiennes au niveau des Centres Hospitalo-Universitaires de Bab El Oued et Beni Messous - Alger, pour leurs aides durant toute la période d'enregistrement des corpus.

Je souhaite aussi remercier l'équipe administrative de l'ENP pour leurs efforts, dans le but de nous offrir une excellente prise en charge dans l'élaboration de ce travail de recherche.

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

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

في هذا العمل، نقترح طريقتين للتقييم الموضوعي لجودة الصوت المرضي: الأولى تستند إلى التحليل الصوتي، والطريقة الثانية هي التقييم بواسطة نظام الكشف الالي للأصوات المرضية بناءً على الشبكات العصبية المتكررة. لذلك اخترنا مرضين للعلاج: شلل الاوتار الصوتية أحادي الجانب واستئصال الحنجرة الشامل على مستوى مصلحة الأنف والأذن والحنجرة في كل من مستشفى باب الواد وبني مسوس الجزائر العاصمة.

في هذا العمل قمنا بتحليل إشارة الصوت اعتمادا على استقرار سعة وتواتر اهتزازات الاوتار الصوتية. العناصر الصوتية المعتمدة في هذه الدراسة هي التواتر الأساسي Fo، البواني، شدة الصوت، معامل التغيير للتواتر الأساسي، نسبة الجهر، Jitter ، Shimmer، HPR، CPP، Breathy Voice (H1-H2)، نسبة الإشارة على الضجيج (HNR) وكذلك المدة القصوى للنطق (TMP). تسجيلات المدونة تمت في مصلحة طب الآذن الانف والحنجرة بكل من مستشفى باب الواد وبني مسوس الجزائر العاصمة. قمنا كذلك باستخدام طريقة أخرى التقييم الموضعية والتي تعتمد على الشبكات العصبونية المتكررة وهذا بأنشاء نضام للتعرف الالي على الأصوات المضطربة.

دراسة النتائج المتحصل عليها خلال فترة التأهيل تبين توافق كبير لعناصر التحليل الصوتي بين الصوت المضطرب والصوت المرجعي وهذا ما تم تأكيده من خلال نسبة التعرف المقبولة جدا المتحصل عليها (78.94 % LT 88.65 % PRU). وعليه يمكن ان نستنتج ان استعمال السمع كطريقة وحيدة للتقييم في المستشفى الجزائري غير كافي حيث يجب توظيف نتائج التحليل الصوتي الفيزيائي مع التحليل السمعي من اجل وضع بروتكول علاجي من اجل تكفل فعال وهادف للصوت المضطرب.

الكلمات المفتاحية: التحليل الصوتي الموضوعي، اضطرابات الصوات، التعرف والتصنيف الالي. الشبكة العصبية المتكررة.

Abstract:

In this work, we propose two methods of objective evaluation of pathological voice quality: the first is based on acoustic analysis, the second method is an evaluation by an automatic detection system of pathological voices based on Recurring Neural Networks. For this we have chosen two pathologies to treat: a Unilateral Laryngectomy Paralysis and a Total Laryngectomy at the level of the ENT services of the hospitals of Bab El Oued and Beni Messous-Algiers.

In this work, an acoustic analysis of the vocal signal is based on measurements of the instability of the amplitude and frequency of the vibrations of the vocal cords. The acoustic parameters chosen for this study are: the average fundamental frequency Fo, the formants, the intensity, the Coefficient of Variation of Fo, The Percentage of Voicing, the jitter, the shimmer, HPR, Breathy Voice (H_1-H_2), CPP, and the Harmonic to Noise Ratio (HNR). For the aerodynamic parameters we used Maximum Phonation Time (MTP). The study of the results obtained by the acoustic analysis of the pathological voice during the rehabilitation phase shows a strong correlation of the acoustic parameters between the pathological voice and the reference one. These results are confirmed by the very acceptable accuracy rate by our detection system (88.65% and 78.94 for PRU and LT). The exclusive use of hearing to evaluate the effect of voice rehabilitation in the Algerian hospital environment remains insufficient. It is important to correlate the perceptual information with the interpreted acoustic measurements, in a manner to be able to develop a therapeutic platform for effective and objective management of voice pathology. The combination of the two acoustic and neuronal methods in the development of a computer application, allows the speech therapist to evaluate automatically and objectively the process of rehabilitation over time.

Keywords: Objective Acoustic Analysis, Voice Pathology, Automatic Detection and Classification, Recurrent Neural Networks.

Résumé :

Dans ce travail, nous proposons deux méthodes d'évaluation objective de la qualité de la voix pathologique : la première est basée sur une analyse acoustique, la seconde méthode est une évaluation par un système de détection automatique des voix pathologiques à base de Réseaux Neurones Récurrents. Nous avons choisi pour cela deux pathologies à traiter : une Paralysie Laryngée Unilatérale (PLU) et une Laryngectomie Totale (LT) au niveau des services ORL des hôpitaux de *Bab El Oued et Beni Messous*-Alger.

Les paramètres acoustiques choisis pour cette étude sont : la fréquence fondamentale moyenne Fo, les formants, l'intensité, le Coefficient de la Variation de Fo, le pourcentage de voisement, le jitter, le Shimmer, HPR, Breathy Voice (H_1-H_2), CPP, ainsi que le Rapport de l'énergie du spectre Harmonique et celle du spectre de Bruit (HNR). Pour les paramètres aérodynamiques, nous avons utilisé le Temps Maximal de Phonation (TMP).

L'étude des résultats obtenus durant la phase de rééducation montre une forte corrélation des paramètres acoustiques entre les voix pathologiques et celles de la norme de référence. Ces résultats sont confirmés par le taux de précision très acceptable par notre système neuronal de détection (88.65% et 78.94 pour la PRU et la LT).

L'utilisation exclusive et abusive de l'ouïe pour évaluer l'effet de la rééducation vocale dans le milieu hospitalier algérien reste insuffisante. Il est important de corréler les données perceptives avec les mesures acoustiques interprétées, de façon à pouvoir élaborer une plateforme thérapeutique pour une prise en charge efficace et objective de la pathologie de la voix.

La combinaison des deux méthodes acoustique et neuronale dans le développement d'une application informatique, permet à l'orthophoniste rééducateur d'évaluer automatiquement et objectivement le processus de rééducation au cours du temps.

Mots clefs — Analyse Acoustique objective, Pathologies de la voix, détection et classification automatique, les Réseaux de Neurones Récurrents.

Chapitre 2 : Méthodes Objectives d'évaluation de la qualité de la voix

Chapitre 3 : Evaluation Objective de la Voix Pathologique par l'Analyse Acoustique

Chapitre 4 : Evaluation Objective par les Réseaux de Neurones Récurrents

Liste des Figures

Liste des Tableaux

Liste des abréviations

ASR	:	**A**utomatic **S**peech **R**ecognition
APQ	:	**A**mplitude **P**erturbation **Q**uotient
CoV	:	**Co**efficient de **V**ariation
CNN	:	**C**onvolutional **N**eural **N**etworks
CPP	:	**C**epstral **P**eak **P**rominence
DCT	:	**D**iscrete **C**osine **T**ransform
DFT	:	**D**iscret **F**ourier **T**ransform
DL	:	**D**eep **L**earning
DNN	:	**D**ynamic **N**eural **N**etwork
DUV	:	**D**egree of **U**nvoiced **V**oice
EGG	:	**E**lectro-**G**lotto-**G**raph
FFT	:	**F**ast **F**ourier **T**ransform
GMM	:	**G**aussian **M**ixture **M**odels
GRNN	:	**G**eneral **R**egression **N**eural **N**etwork
KNN	:	**K**-**N**earest **N**eighbors
HMM	:	**H**idden **M**arkov **M**odels
HNR	:	**H**armonics to **N**oise **R**atio
HPR	:	**H**igh-frequency **P**ower **R**atio
IA	:	**I**ntelligence **A**rtificielle
IFFT	:	**I**nverse **F**ast **F**ourier **T**ransform
LSTM	:	**L**ong **S**hort-**T**erm **M**emory
LP	:	**L**aryngectomie **P**artielle
LT	:	**L**aryngectomie **T**otale
MDVP	:	**M**ulti **D**imensional **V**oice **P**rogram
MFCC	:	**M**el **F**requency **C**epstral **C**oefficients
ML	:	**M**achine **L**earning
MLP	:	**M**ulti **L**ayer **P**erceptron
PNN	:	**P**robabilistic **N**eural **N**etwork
PR	:	**P**aralysie **R**ecurrentielle
PRU	:	**P**aralysie **R**ecurrentielle **U**nilatérale
RAP	:	**R**elative **A**verage **P**erturbation

ReLu	**:**	**R**ectified **L**inear **U**nit
RNA	**:**	**R**éseaux de **N**eurones **A**rtificiels
RNN	**:**	***R**ecurrent **N**eural **N**etwork*
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
TVP	**:**	**T**aux de **V**rai **P**ositif
TVN	**:**	**T**aux de **V**rai **N**égatif
TMP	**:**	**T**emps **M**aximal de **P**honation.

Introduction Générale

Dans notre société, la communication est une faculté fondamentale de l'être humain et représente une clé pour la réussite sociale et professionnelle. Elle peut être réalisée sous la forme de différentes modalités : orale, écrite, gestuelle, etc. La communication parlée reste malgré tout, le centre de la communication humaine en permettant une communication simple et efficace pouvant contenir différents messages (informer, demander, exprimer un avis/sentiment, etc.). Elle permet aussi grâce au riche vocabulaire et style de parole de véhiculer et traduire plusieurs nuances parfois difficiles à exprimer autrement. Outre un moyen de communication, ses caractéristiques sont le reflet et la représentation de notre identité et personnalité humaine. Malgré la révolution numérique des dernières décennies qui a permis l'introduction de nouveaux outils de communication (messagerie électronique, messagerie instantanée, réseaux sociaux, etc.), la parole n'a pas perdu son statut de moyen principal et incontournable de communication.

C'est la raison pour laquelle, tout trouble de la voix peut avoir des conséquences importantes sur la vie quotidienne des personnes concernées allant jusqu'à engendrer des comportements de repli sur soi et d'isolement de la société. Ces comportements peuvent être liés à des causes physiques où la personne atteinte voit ses capacités de parole diminuer limitant ses aptitudes professionnelles. Aussi, les troubles de la parole, entraînant des productions marquées et symptomatiques, peuvent décourager les personnes concernées à parler en public affectant aussi bien leur vie professionnelle que sociale.

La dysphonie est un trouble de la voix qui résulte d'une lésion organique et/ou d'une dysfonction de production, principalement liée aux cordes vocales. Elle est généralement diagnostiquée chez des patients dont le travail au quotidien repose essentiellement sur la voix : enseignants, chanteurs, etc.

Cadre de la thèse

L'évaluation de la qualité de la voix est un enjeu important pour la laryngo-phoniatrie dans le but de valider la pertinence et l'efficacité des traitements proposés, qu'il s'agisse de la rééducation ou de la phono-chirurgie. Dans ce sens, le jugement à l'oreille, connu également sous la terminologie d'évaluation subjective ou perceptive, est la seule méthode d'analyse et d'évaluation de la voix pathologique utilisée en milieu clinique algérien [1,2,3]. Dans cette méthode, l'orthophoniste rééducateur est le seul chargé d'évaluer par écoute la qualité de la voix, ce qui entraîne une évaluation perceptive non fiable, vue que l'analyse perceptive fiable

impliquant plusieurs auditeurs experts et plusieurs sessions d'écoute s'avèrent finalement consommatrices en temps et en ressources humaines, et ne permettant pas une utilisation régulière en routine clinique [3,4,5].

Un autre inconvénient majeur de l'évaluation subjective est la variabilité inter et intra auditeurs dans la perception de la voix par un jury d'experts. Cette variabilité peut être influencée par le contexte, par l'état émotionnel ou l'attention de l'auditeur. En revanche, l'évaluation objective de la voix, basée sur l'analyse acoustique permet de faire des mesures sur le signal vocal pour obtenir des paramètres, des indices, dégager des tendances, et ce, de façon quantitative et objective. Cette méthode instrumentale peut aider l'orthophoniste dans l'évaluation de la qualité de la voix au cours de la période de rééducation.

Dans ce sens, il existe des travaux de recherches qui ont été réalisés visant à évaluer objectivement par analyse acoustique, la voix en milieu clinique algérien. Néanmoins, ils restent insuffisants par rapport au protocole et les conditions d'enregistrement de la voix pathologique qui nécessitent des précautions techniques à l'évaluation acoustique [1,2,3]. D'autre part, et à nos connaissances, il n'existe pas de travaux de recherche qui ont élaboré des systèmes de détection ou de classification des voix pathologiques par les Réseaux de Neurones à Mémoire Longue et Courte Terme LSTM en milieu hospitalier algérien. D'autres outils d'apprentissage automatique comme les Machines à Support de Vecteurs (SVM) et les Réseaux Neuronaux Artificiels classiques, étaient déjà utilisés dans des travaux similaires [6].

Objectif de la thèse

Dans ce travail, nous proposons deux méthodes d'évaluation objective de la qualité de la voix pathologique : la première est basée sur une analyse acoustique, la seconde méthode est une évaluation par un système de détection automatique des voix pathologiques à base de Réseaux Neurones Récurrents. Nous avons choisi pour cela deux pathologies à traiter : une Paralysie Laryngée Unilatérale et une Laryngectomie Totale au niveau des services ORL des hôpitaux de *Bab El Oued et Beni Messous*-Alger. Le principal objectif de ce travail est de montrer que l'évaluation objective, avec ses résultats, pourra aider l'orthophoniste dans la rééducation, évaluer et contribuer de façon objective, l'évolution de cette rééducation au cours du temps.

Problématique de la recherche

Nous allons répondre dans ce travail à la problématique suivante : Quel est le degré de convergence des paramètres acoustiques de la voix après rééducation et la voix normale ? La méthode objective basée sur l'analyse acoustique du signal vocal peut-elle contribuer à l'évaluation de la voix après rééducation au niveau du service ORL ? La méthode de rééducation adoptée au niveau des services ORL est-elle adaptée à la population pathologique algérienne ? les systèmes de détection et de classification automatique de voix pathologiques pourront-ils aider l'orthophoniste dans la rééducation vocale ?

Organisation de la thèse

Cette thèse est organisée en quatre chapitres :

Dans le premier chapitre, nous présentons quelques aspects essentiels de la phonation. Son but est de comprendre le fonctionnement de l'activité phonatoire, et donc de décrire les principaux organes jouant un rôle direct dans la phonation. Nous présentons ensuite, une classification des différents types de dysphonies.

Le deuxième chapitre est une présentation des méthodes d'évaluation de la voix. Nous détaillons l'évaluation objective basée sur l'analyse acoustique qui est le but de notre sujet. Cette présentation est suivie par une introduction aux Réseaux de neurones artificiels. Nous donnons, en particulier, une importance aux Réseaux Récurrents de type LSTM que nous appliquons dans le cadre de notre travail.

Le troisième chapitre est consacré à l'évaluation objective des dysphonies basée sur l'analyse acoustique. Nous avons commencé par une mise en place du dispositif expérimental, tels que le choix du matériel, les conditions d'enregistrement, le corpus et le protocole d'enregistrement. Ensuite, une analyse acoustique est effectuée sur des voix pathologiques : avant, en cours et après rééducation, choisie pour cette étude. Nous avons terminé ce chapitre par des interprétations et commentaires des résultats obtenus afin de dégager une évaluation objective globale de la voix après la phase de rééducation.

Dans le dernier chapitre, nous appliquons l'une des méthodes de classification automatique de la voix pathologique : Réseaux de Neurone Récurrents, pour évaluer automatiquement le processus de la rééducation. Nous achevons cette thèse par des conclusions, perspectives et des références concernant ce domaine.

Chapitre 1

Phonation et Dysphonie

1.1. Introduction

L'émission d'un son humain, autrement dit la voix, est un processus complexe qui demande une coordination efficace de la part de tous les organes mis en jeu : la soufflerie, le vibrateur et les résonateurs. En effet, l'origine du son provient de l'air expulsé par l'appareil respiratoire, qui sera ensuite modifié par sa mise en vibration en traversant les cordes vocales et modulé en atteignant les cavités de résonance.

Dans ce chapitre, nous présentons en premier lieu quelques aspects essentiels de la phonation. Le but est de comprendre le fonctionnement de l'activité phonatoire, et donc de décrire les principaux organes jouant un rôle direct dans la phonation. Ensuite, nous exposons une classification des différents types de dysphonies en mettant l'accent sur les deux pathologies vocales étudiées dans notre travail.

1.2. Phonation

La physiologie de la phonation correspond à l'ensemble des mécanismes qui permettent l'apparition d'une vibration au niveau du bord libre des *cordes vocales*. Il s'agit du mécanisme sonore initial qui est ensuite soumis au filtrage du pharyngo-bucco-nasal pour être transformé en parole.

1.3. Appareil phonatoire humain

Le mécanisme phonatoire humain se compose de trois organes ayant chacun une fonction dans la production de la parole : la Soufflerie pulmonaire, le Vibrateur laryngé et les résonateurs (cavités pharyngo- bucco-nasales) (Figure 1.1).

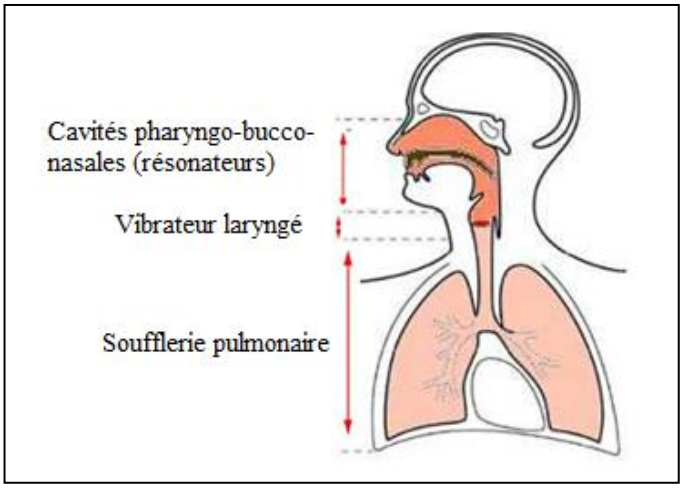


Figure 1.1 : Les trois organes du mécanisme de phonation [7]

1.3.1. Soufflerie pulmonaire

Les poumons et la trachée artère constituent l'étage sous glottique de la phonation ou *la Soufflerie* qui fournit la source d'énergie nécessaire à l'ensemble de l'appareil phonatoire. L'effort pulmonaire est régi par les mouvements du diaphragme. La parole est essentiellement produite lors de l'expiration de l'air, qui est à l'origine de la formation d'une surpression en dessous du larynx, appelée *Pression sous-glottique*.

1.3.2. Vibrateur laryngé

Le larynx, situé au carrefour des voies respiratoire et œsophagienne, est la partie de l'appareil phonatoire qui abrite des replis de membranes muqueuses appelées *cordes vocales (ou plis vocaux)*. L'espace compris entre les deux cordes vocales est appelé *la glotte*.

L'interaction fluide-structure entre l'écoulement d'air provenant des poumons et l'élasticité des cordes vocales permet donc, de produire des sons voisés. Ce voisement se traduit par l'apparition d'un mouvement de vibration périodique des cordes vocales. La source de débit acoustique ainsi créé par la modulation de l'écoulement glottique a une *fréquence fondamentale*, notée F_0, correspondant à la fréquence de vibration des cordes vocales.

1.3.3. Résonateurs

Les voies aériennes supérieures se composent des différentes cavités de l'appareil phonatoire (*conduit vocal*). L'ensemble des éléments du conduit vocal (langue, palais, lèvres, dents, épiglotte) sont impliqués dans l'articulation, ainsi que le voile du palais qui permet lorsqu'il est abaissé de faire communiquer le volume du conduit vocal avec celui de la cavité nasale. L'articulation est la modulation du signal de la source voisée par les résonances acoustiques des cavités supra-glottiques (figure 1.2).

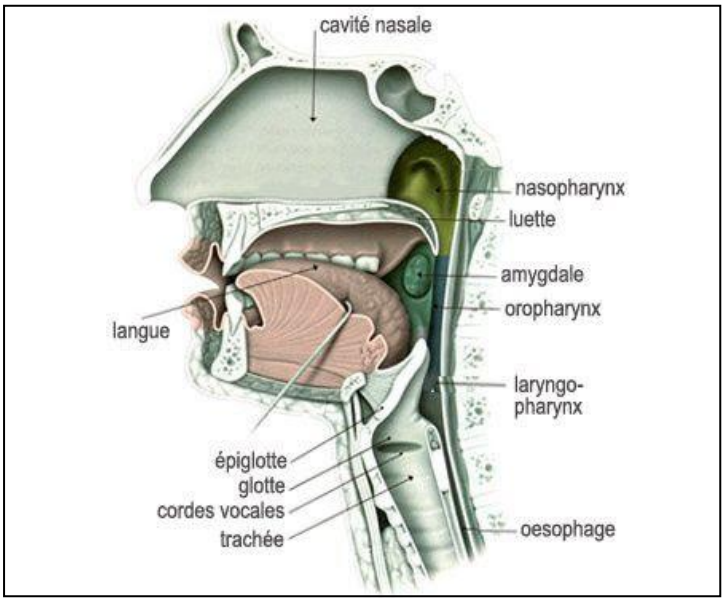


Figure 1.2 : Coupe de l'Appareil Phonatoire humain [8]

1.4. Larynx

Le larynx est le vibrateur de l'appareil phonatoire, situé entre l'appareil respiratoire (soufflerie) et les cavités de résonance. C'est un assemblage de cartilages articulés, reliés entre eux par des ligaments et des muscles, l'ensemble étant tapissé d'une muqueuse. Les principales fonctions du larynx sont : ***La protection*** des voies aériennes au cours de la ***déglutition*** pour en éviter les fausses routes, la ***respiration*** et ***la phonation*** :

- ***la phonation*** est possible grâce aux cordes vocales dont la muqueuse vibre sous l'effet de l'air expulsé de la cage thoracique (souffle expiratoire). Cette vibration va varier en fonction de la tension et donc de la longueur des cordes vocales qui est sous le contrôle de deux muscles, les *muscles crico-thyroïdiens* et *thyro-aryténoïdiens*. Le larynx peut faire varier deux paramètres du son : *l'intensité* en augmentant la pressions sous glottique, la *fréquence* du son, en faisant varier la fréquence de vibration des cordes vocales.

- ***la respiration*** est possible grâce au passage d'air dans la colonne laryngée et en particulier au niveau de l'espace situé entre les cordes vocales et la commissure postérieure du larynx, c'est-à-dire l'espace glottique. Au cours de l'inspiration, les cordes vocales sont en *abduction*, permettant d'ouvrir le larynx et le passage de l'air. Au cours de l'expiration les cordes vocales se rapprochent sous l'action des muscles

adducteurs du larynx. Le larynx intervient aussi au cours des efforts à glotte fermée, pour permettre de maintenir une pression sous glottique importante.

- ***la déglutition***, la fermeture et l'ascension du larynx protègent les voies aériennes et libèrent le cricoïde, permettant d'orienter préférentiellement le bol alimentaire de la base de langue vers la bouche œsophagienne qui se relâche alors. Le cartilage épiglottique est alors plaqué sur la partie haute du larynx (figure 1.3).

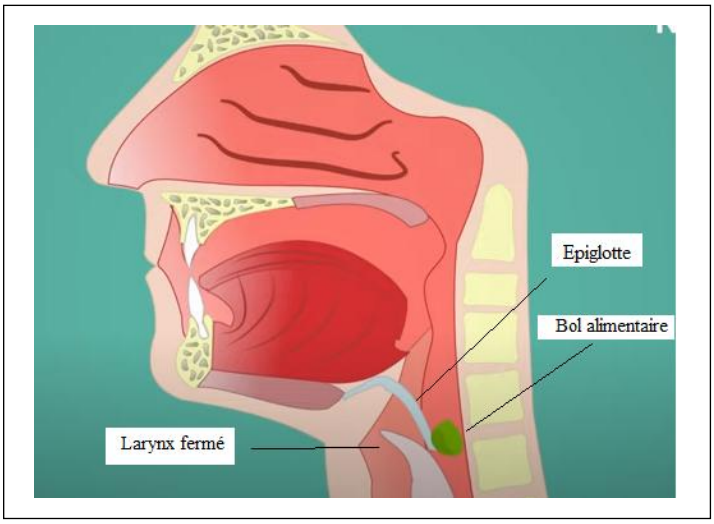


Figure 1.3 : Protection des voies respiratoires (fermeture du trachée) permettant le le passage des aliments dans l'œsophage

1.4.1. Muscles du larynx

Les muscles qui connectent les différents cartilages du larynx permettent de contrôler la tension et l'écartement des cordes vocales. La figure 1.4 présente les cartilages du larynx et les muscles intervenant dans la production de parole.

1.4.1.1. Muscles tenseurs des cordes vocales

La contraction des muscles *thyro-aryténoïdiens* rigidifie les cordes vocales ce qui entraîne une augmentation de la fréquence fondamentale de la phonation. L'action des muscles *crico-thyroïdiens* provoque le basculement du cartilage thyroïde (pomme d'Adam) vers l'avant. Ce mouvement provoque une augmentation de la longueur et de la tension longitudinale des cordes vocales. Tout comme la contraction des muscles thyro-aryténoïdiens, cette stratégie est utilisée pour augmenter la fréquence fondamentale de la phonation [9].

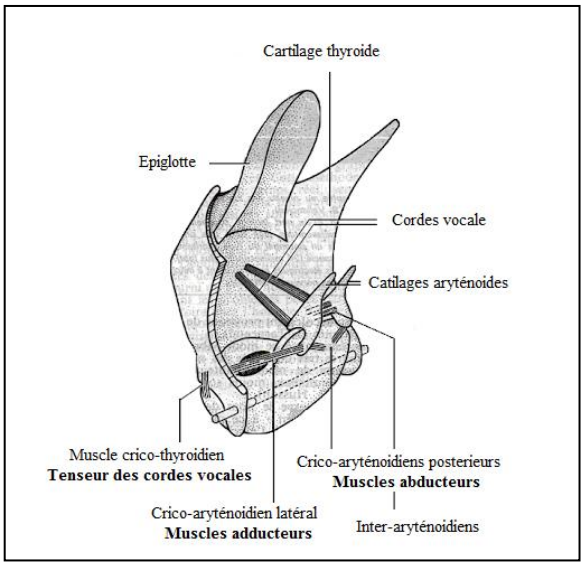


Figure 1.4 : Schéma représentatif des muscles de larynx [8]

1.4.1.2. Muscles adducteurs et abducteurs

Les cordes vocales sont écartées (abduction) lors de la respiration, afin de laisser passer l'air, et rapprochées lors de la phonation (adduction). Le niveau d'adduction des cordes vocales a un impact direct sur le timbre de la production vocale. Le geste d'adduction des cordes vocales est assuré par l'action des muscles *inter-aryténoïdiens*, dont la contraction a pour effet de rapprocher les deux cartilages *aryténoïdes*, et par la contraction des muscles *crico-aryténoïdien*s latéraux, qui fait pivoter les aryténoïdes. L'abduction des cordes vocales est réalisée par le biais des muscles *crico-aryténoïdiens* postérieurs dont la contraction provoque un basculement des aryténoïdes, qui a pour effet d'écarter les cordes vocales l'une de l'autre [9].

1.4.2. Innervation du larynx

Le Larynx est innervé par deux nerfs de chaque côté : les nerfs laryngés supérieurs et les nerfs laryngés inferieurs ou nerfs récurrents. Les deux paires de nerfs naissent du X^{ème} nerf pneumogastrique, appelé aussi nerf vague (Figure 1.5). L'innervation cordale est donc mixte : sensitive avec les nerfs laryngés supérieurs, et motrice avec les nerfs récurrents.

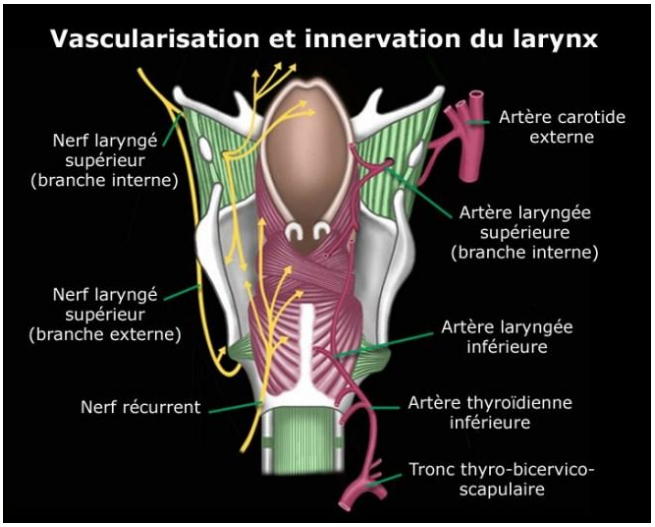


Figure 1.5 : Innervation du Larynx [10]

1.4.3. Fonctionnement du larynx dans la phonation

Lors de la respiration calme, qui est un phénomène automatique et passif, les cordes vocales sont ouvertes. Lors de l'émission vocale, qui se produit sur la phase d'expiration, les cordes vocales vont d'abord se rapprocher en position de fermeture, grâce aux cartilages aryténoïdes. La pression de la colonne d'air expiratoire (pression sous glottique) se heurte à un obstacle (fermeture des cordes). Elle va augmenter et contraindre, les bords libres des cordes vocales à s'écarter légèrement, laissant passer une petite quantité d'air ou puff. Ce puff d'air aussitôt libéré, les bords libres vont à nouveau se rapprocher, à la fois : sous l'action de la diminution de la *pression sous glottique*, par *effet Bernoulli* (effet de rétro-aspiration de la muqueuse cordale) et grâce à *l'élasticité propre* des cordes vocales [11].

Le phénomène va se reproduire de façon périodique car la pression sous glottique augmente à nouveau, les cordes vocales étant refermées, créant ainsi une nouvelle vibration. L'énergie aérienne se transforme en énergie acoustique. Les puffs d'air libérés successivement vont créer le son laryngé, assimilé à une impulsion acoustique, qui a une structure discontinue. Son rythme détermine la fréquence de la voix, son amplitude l'intensité, sa forme le timbre. Ces petits mouvements très rapides de fermeture ouverture des cordes vocales représentent la fréquence fondamentale (F_0) de la voix.

La figure 1.6 montre un cycle complet de fonctionnement : à gauche, les bords de la muqueuse recouvrant les cordes vocales vus de face (coupe frontale), au milieu la glotte vue de dessus, et à droite une courbe représentant le déroulement du cycle [12].

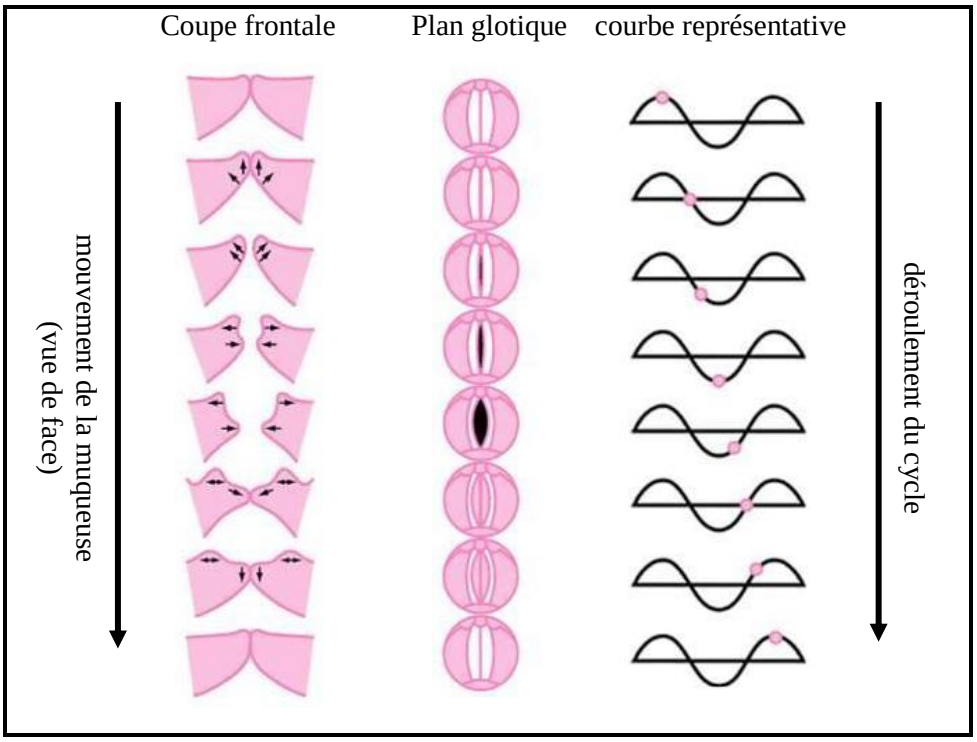


Figure 1.6 : Cycle vibratoire des Cordes Vocales [12]

1.5. Dysphonie

Le Huche & Allali définissent la *dysphonie* comme un trouble momentané ou durable de la fonction vocale ressentie comme telle par le sujet lui-même et son entourage. L'établissement même du statut de *dysphonique* dépend beaucoup du vécu subjectif du patient. La dysphonie se traduit par une diminution du confort vocal et par une altération, fréquente mais non systématique, d'un ou plusieurs paramètres acoustiques de la voix (timbre, intensité, Pitch) [13].

1.5.1. Dysphonies d'origines neurologiques

Cette première famille regroupe les dysphonies provoquées par l'état neuromoteur du patient. La dysphonie peut être provoquée schématiquement par de multiples facteurs tels que l'hypotonie ou l'hypertonie de la musculature laryngée et respiratoire ou encore des tremblements, qui ont pour conséquence de moduler la hauteur,

l'intensité et le timbre de la voix. Elle peut également être provoquée par un mauvais contrôle de la fermeture de la glotte, conséquence de spasmes ou de paralysies [13].

L'hypotonie a pour conséquence une faible intensité de la voix et un abaissement de la Fo. *L'hypertonie,* qui se manifeste par la difficulté à initialiser un acte volontaire du larynx, se traduit par des hésitations au démarrage du voisement, des émissions vocales discontinues, une augmentation de la Fo, un timbre sourd par manque d'harmonique, et voilé par suite à un mauvais accolement des cordes vocales. Les tremblements, qui peuvent être de fréquence variable en fonction de leur origine, rendent la voix chevrotante. Nous pouvons leur associer des instabilités de Fo en voix tenue [13].

Les dysphonies spasmodiques (ou dystonies laryngées) provoquent des changements brutaux de la hauteur de la voix qui peut s'interrompre, repartir, glisser et chevroter. Elles sont caractérisées par un timbre désagréable et être, au pire, inintelligible. Dans les paralysies laryngées, une corde vocale demeure en position plus ou moins ouverte à la suite d'un mauvais contrôle neuromoteur. La voix est soufflée et rauque avec une fuite d'air importante, entraînant un essoufflement en fin de phrase et une voix projetée continue impossible.

1.5.2. Dysphonies d'origines morphologiques

Cette seconde famille regroupe les changements anatomiques de la glotte provoqués par l'apparition de *nodules, polypes et kystes* qui sont des lésions bénignes des cordes vocales provoquées généralement par un forçage vocal permanent ou brutal (Figure 1.7). La voix est plus grave, rauque, soufflée. Son timbre est voilé, sourd et éraillé. Le changement de structure de la glotte peut être également provoqué par des laryngites qui sont des inflammations de l'ensemble des cordes vocales occasionnées par des infections, favorisées par l'effort vocal et qui peuvent s'installer de manière permanente (œdème de Reinke). La voix est plus grave, avec des difficultés dans les aigus, rauque et peu timbrée. Elle peut même disparaître totalement (extinction de voix). Enfin, les plus importants changements anatomiques de la glotte sont provoqués par des traumatismes chirurgicaux à la suite de l'ablation d'un cancer cordal. La voix est très dégradée, grave, de faible intensité, mais intelligible sauf dans le bruit. Le timbre est très rauque, granuleux, soufflé en rapport avec la fuite glottique [14].

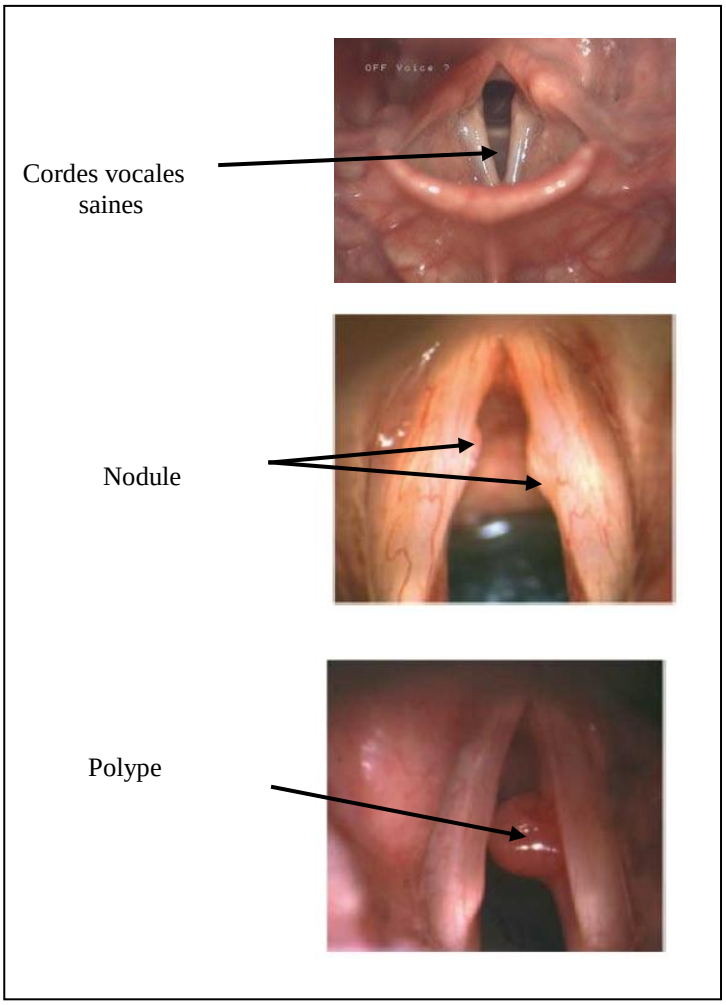


Figure 1.7 : Lésions bénignes des cordes vocales [15]

1.6. Paralysie Récurrentielle (PR)

Les immobilités laryngées sont définies comme une diminution ou un arrêt complet du mouvement *d'adduction et/ou d'abduction* du larynx. En fonction de leur topographie laryngée (unie ou bilatérale, position plutôt en abduction ou en adduction), elles exposeront à un risque vital par gêne respiratoire ou par troubles de la déglutition et à un risque fonctionnel portant sur les différentes fonctions du larynx : la phonation, la déglutition, la respiration et la gestion des efforts à glotte fermée. Elle correspond à une atteinte du nerf laryngé inférieur (nerf récurrent), par compression (tumeur de la bouche de l'œsophage par exemple), section ou étirement de ses fibres (lors d'une thyroïdectomie par exemple) [16].

Il existe deux types de paralysies récurrentielles qui peuvent être :

- unilatérales ;
- bilatérales.

1.6.1. Paralysie Récurrentielle Unilatérale

La **P**aralysie **R**écurrentielle **U**nilatérale PRU est la forme la plus fréquente (Figure 1.8). Elle consiste en l'atteinte d'un des nerfs récurrents induisant une paralysie de la corde vocale. Cette atteinte est, d'ailleurs, plus fréquemment localisée à gauche, puisque le trajet récurrentiel gauche est beaucoup plus long que le droit et qu'il traverse davantage de structures anatomiques. Les conséquences sur le larynx sont une dysphonie plus ou moins marquée, notamment selon la position de la corde vocale paralysée, et des troubles de la déglutition non systématique, mais touchant essentiellement les liquides [15]. La PRU peut être une atteinte isolée ou associée.

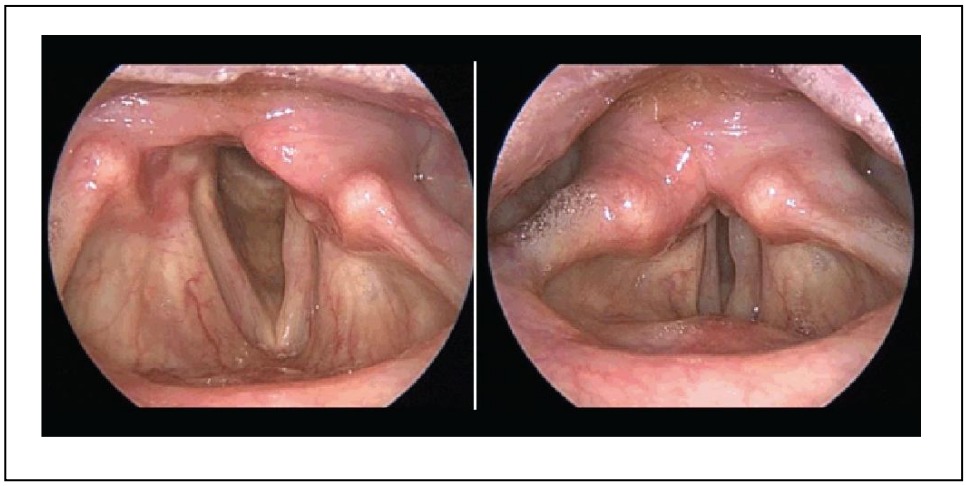


Figure 1.8 : Paralysie Unilatérale de la corde vocale gauche [15]

1.6.1.1. Paralysie Récurrentielle Unilatérale isolée

La morphologie d'un nerf récurrent est complexe, et sa lésion peut être située dans un quelconque point de son long parcours, et ces lésions se diffèrent l'une de l'autre selon qu'il s'agit d'une atteinte du nerf récurrent droit ou gauche.

Les Paralysies Récurrentielles Unilatérales isolées peuvent être à l'origine d'une :

- **atteinte traumatique :** les PRU sont essentiellement liées à la blessure chirurgicale d'un nerf récurrent lors d'intervention portant sur le cou (au cours de la chirurgie de la glande thyroïde et des parathyroïdes) et sur le médiastin (le nerf récurrent

gauche peut être lésé lors de la chirurgie broncho-pulmonaire gauche, de la crosse aortique : canal artériel, anévrysme de la crosse) ;

- **Compression:** du nerf récurrent qui peut être due au développement d'un nodule qui peut entraver le fonctionnement du nerf par compression, s'il est très volumineux ou par invasion maligne s'il est de nature cancéreuse. D'autres causes possibles de compression sont : le cancer de l'œsophage, le cancer bronchique,
- **Standard:** de nombreuses maladies infectieuses peuvent entraîner une Paralysie Récurrentielle. Citons plus particulièrement le *zona*. Les névrites toxiques (plomb, arsenic).

1.6.1.2. Paralysie Récurrentielle Unilatérale associée

La PRU peut être aussi associée à d'autres atteintes nerveuses comme la tumeur (endocrânienne, de la base de crâne), à une inflammation (polyradiculonévrite du Guillain-Barré, zona pharyngo-laryngé, …) ou encore à une atteinte bulbaire vasculaire, dégénérative ou tumorale [15].

1.6.2. Signes de la Paralysie Récurrentielle Unilatérale

Les principaux signes de la Paralysie Récurrentielle Unilatérales sont [13, 15] :

- une **voix modifiée** : soufflée, rauque ou éraillée, l'intensité de la voix limitée et une altération du timbre, etc. ;
- un **essoufflement** à la parole ;
- des **fausses routes,** principalement aux liquides ou avec certains aliments (ex : riz, petits grains, biscuits secs, etc.) ;
- Une sensation de sécrétions au fond de la gorge, une envie fréquente de se *racler la gorge*.

1.6.3. Rééducation vocale de la Paralysie Récurrentielle

D'après Aronsone, un traitement orthophonique peut être défini comme un effort positif, doit convenir aux besoins sociaux professionnels du patient. Il s'agit pour le patient de reconstruire la meilleure voix possible en fonction de ses capacités anatomiques, physiologiques et psychologiques [17].

Cet effort doit être fait par le patient et l'orthophoniste : l'orthophoniste repère les défaillances de la voix et propose au patient les stratégies les mieux adaptées, en

parallèle, le rôle du patient est d'accepter le travail proposé et de mettre en pratique les différentes techniques.

L'objectif général de la prise en charge de la paralysie unilatérale, est de travailler, non seulement la phonation mais aussi, selon les cas, la respiration et la déglutition. Au début de chaque prise en charge, une prise de conscience corporelle pourra permettre au patient de mieux comprendre et de mieux sentir les différents muscles mis en jeu lors de l'émission vocale. C'est la base nécessaire à la connaissance de son corps, de ses possibilités et de ses limites.

La corde vocale paralysée va rapidement s'incurver et s'atrophier. Il s'agit de commencer la prise en charge rapidement, le plus tôt possible, pour empêcher ou ralentir l'atrophie et faciliter la fermeture de la glotte. Pour cela le travail de la rééducation permettra à la corde vocale saine de dépasser la ligne médiane pour assurer une vibration passive de la corde vocale paralysée [18].

La rééducation orthophonique de la pathologie vocale s'appuie en général sur trois principaux axes de travail: *relaxation*, *souffle et travail vocal* [19].

1.6.3.1. Travail de relaxation

La relaxation est indispensable pour toute rééducation vocale. La plus utilisée est la méthode de relaxation avec les yeux ouverts de Le Huche, qui a adapté des exercices de psychanalystes ayant pour but la régression, avec les yeux fermés.

Le Huche a modifié l'objectif pour travailler la relaxation dans une visée phoniatrique. Ce travail est basé sur la sensation de poids, de tension, avec des contractions suivies de relâchements, avec une durée d'exercice de 7 à 10 minutes. Les yeux sont volontairement gardés ouverts pour enlever le côté anxiogène que la situation pourrait avoir. Le patient est allongé sur le dos, un coussin est placé sous la courbure de sa colonne vertébrale ainsi qu'un autre sous ses mollets afin de les surélever. Ses bras sont longs du corps et ses mains sont ouvertes vers le haut. Pendant un certain temps (10 secondes à 2 minutes), le patient prend conscience de sa propre position. Il s'installe en quelque sorte et vérifie qu'il repose confortablement sur le dos et que les parties droite et gauche de son corps s'appuient de façon égale sur le plan horizontal. Il rectifie au besoin la position de ses épaules ou de son bassin.

L'orthophoniste appliquera sur le patient des techniques de relaxation [19] : introduction des soupirs, Crispation-relaxation de la main et de l'avant-bras droits, crispation-relaxation de la jambe et du pied droits, crispation-relaxation de la jambe et du pied gauches, crispation-relaxation de la main et de l'avant-bras gauche, etc.

1.6.3.2. Travail de souffle

En second lieu, l'orthophoniste doit s'assurer que son patient maîtrise bien la relaxation, pour passer à la deuxième étape qui concerne des exercices de respiration. Il commence d'abord à apprendre au patient une respiration correcte qui est essentielle pour la production d'une voix de bonne qualité et pour une articulation satisfaisante. Elle joue aussi un rôle dans l'intonation de la voix et le rythme de la parole. Pour une respiration correcte, il est nécessaire de respirer avec le diaphragme (muscle situé au niveau des côtes et de l'abdomen) et non avec les épaules et le haut de la poitrine, les poumons auront ainsi plus de place pour se dilater et la quantité d'air emmagasinée sera plus importante. Cette technique se pratique dans différentes positions, soit allongée assis ou debout [19].

1.6.3.3. Travail vocal

En fin, après que le sujet soit bien détendu, on passe ensuite aux exercices de la voix, qui est la dernière étape de la rééducation. Cela ne veut pas dire que les étapes précédentes soient dépassées, le rééducateur pourrait toujours revenir à elles. Les exercices de vocalisation consistent en des séries d'émission vocales relativement brèves et en des exercices plus complexes, à partir de textes parlés ou chantés [19].

1.7. Laryngectomie

L'ablation totale ou partielle du larynx appelée *laryngectomie* est un acte chirurgical résultat d'un cancer. Il peut s'agir d'un cancer limité au larynx ou d'un cancer du larynx étendu au pharynx.

Les cancers du larynx ou du pharynx sont des affections fréquentes chez les hommes que chez les femmes (dix fois plus). Ils sont favorisés principalement par la consommation d'alcool et l'usage du tabac [20,21].

Cliniquement, le cancer du larynx se manifeste à son début de façon variable. Il peut s'agir d'une altération progressive du timbre vocal d'abord, puis prendre peu à peu

l'aspect à la fois rauque et mat de la voix [21]. Il peut se traduire par une dysphagie douloureuse (gène à la déglutition). Le principal caractère de ces signes est leur persistance et leur aggravation progressive d'où la nécessité d'un examen laryngoscopie lorsque ces signes persistent et en particulier devant toute dysphonie ou toute gène de déglutition qui durent plus de trois semaines. La laryngectomie peut être classée en laryngectomie partielle ou bien totale

1.7.1. Laryngectomie Partielle

La Laryngectomie Partielle (LP) préserve en général une partie suffisante du larynx pour permettre au patient de parler correctement après l'opération. Dans certains cas, la voix reste un peu rauque ou faible mais le circuit aéro-digestif est préservé.

1.7.2. Laryngectomie Totale

La Laryngectomie Totale (LT) par l'ablation du larynx a pour conséquence directe, la séparation de la voie respiratoire et de la voie digestive (puisqu'il n'y a plus de carrefour). Il faut donc reconstruire de manière à permettre au patient de manger par la bouche sans risquer une fausse route (passage d'aliment ou de salive dans la trachée). Cette reconstruction consiste donc à rétablir une continuité entre la bouche et l'œsophage (appelée néo-pharynx) en faisant un entonnoir avec la muqueuse de la gorge. La voie respiratoire est rétablie en suturant la trachée au niveau de la base du cou, dans la cicatrice appelée un *trachéostome*. C'est par cet orifice que le patient respirera de manière définitive (Figure 1.9). Après la Laryngectomie Totale, le patient ne peut plus parler en voix laryngée utilisant l'air pulmonaire. Il pourra apprendre une autre technique pour parler [22].

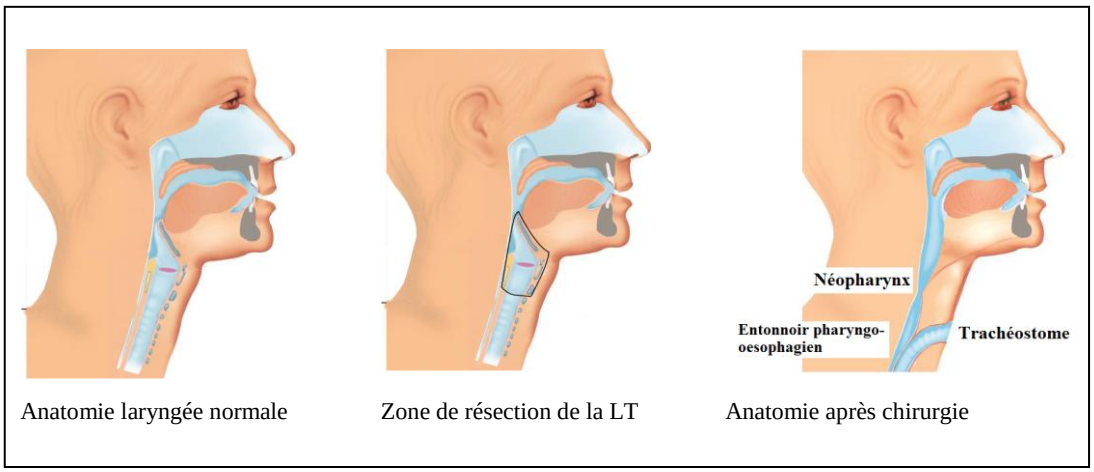


Figure 1.9 : Principe de l'intervention chirurgicale dans la LT [23]

1.7.3. Réhabilitation vocale de la LT

Il existe différentes méthodes de réhabilitation vocale depuis l'apparition de la laryngectomie totale au début du XIX^{ème} siècle [24] :

- voix *Œsophagienne* obtenue lors d'une éructation faisant vibrer la *néo-glotte*, ou *néo pharynx*. Cette technique constitue le moyen de réhabilitation, adopté par le service d'ORL et de chirurgie cervico-faciale du CHU de Beni Messous, Alger ;
- voix *Trachéo-œsophagienne*, obtenue à l'aide d'un implant phonatoire permettant l'issue d'air provenant des poumons jusqu'à la néo-glotte afin de la faire vibrer ;
- voix *Prothétique* créée à partir d'un matériel électronique, appelé *électro-larynx*.

1.7.3.1. Voix œsophagienne

L'apprentissage de la voix *œsophagienne* peut commencer dès que les sutures ont cicatrisé, entre la 3ème et la 6ème semaine après l'intervention. La précocité du démarrage de la rééducation donne au patient de meilleures chances pour acquérir le bon geste d'éructation, sans la mise en place de mauvaises habitudes (voix chouchoutée).

La voix œsophagienne est la technique prédominante et la plus utilisée de la réhabilitation vocale. Le patient ne peut plus utiliser l'air provenant des poumons, et c'est l'œsophage qui les remplacera. C'est pourquoi, cette technique est nommée *voix œsophagienne*. Il s'agit d'apprendre au patient à produire un son dans l'entrée de l'œsophage. C'est un son d'éructation (rot) comme ce que nous faisons parfois involontairement après avoir mangé ou bu. Nous retrouvons les trois conditions

nécessaires à la production de la parole : la soufflerie qui est l'œsophage, la *Néo-glotte*, un muscle vibrant (un rétrécissement), formé par les fibres musculaires suturées restantes après ablation du larynx, situé à la partie supérieure de l'œsophage. Les cavités de résonance n'ont pas subi de changement [24,25].

Pour parler, le patient doit faire pénétrer de l'air de la bouche à travers l'entrée de l'œsophage. L'action des muscles abdominaux et diaphragmatique entraîne une ouverture réflexe de la Néo-glotte provoquant la remontée de l'air et la mise en vibration de la muqueuse de la bouche œsophagienne. L'air est ainsi sonorisé par le Néo-vibrateur avant d'être libéré. La technique suppose de la part du patient une maîtrise parfaite de la coordination entre respiration et éructation, de la quantité d'air à comprimer dans l'œsophage, ainsi qu'une posture adéquate [25,26] (Figure 1.10).

La voix œsophagienne a l'avantage de ne nécessiter aucun matériel. Elle permet non seulement une liberté de mouvement du visage ou des mains pendant la phonation, mais également d'éviter toute contrainte d'appareillage et d'entretien par la suite. Elle nécessite par contre un apprentissage souvent long et jugé difficile par les patients. Les qualités acoustiques de cette voix sont souvent très satisfaisantes, mais le débit de parole est parfois haché par les injections d'air et l'intensité vocale jugée trop faible [26].

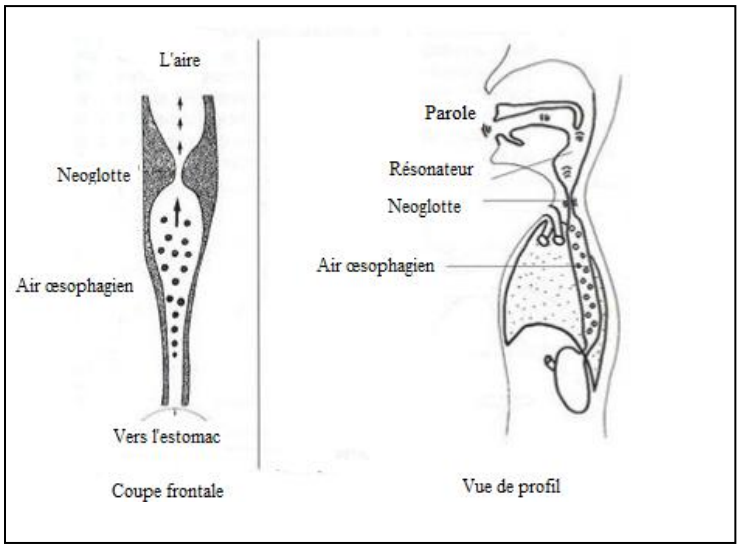


Figure 1.10 : Principe de production de la voix œsophagienne [23]

1.7.3.2. Voix Trachéo-œsophagienne

Comparée à la voix œsophagienne, la voix Trachéo-œsophagienne est plus facile à acquérir. Elle est obtenue grâce à la pose d'un implant phonatoire (prothèse interne). Ce dernier est placé entre la partie supérieure de la trachée et l'œsophage (Figure 1.11). Elle est ouverte aux deux extrémités et est munie côté œsophage d'un petit clapet qui s'ouvre vers l'œsophage et se ferme dans le sens œsophage-trachée empêchant ainsi le passage de la salive et de l'alimentation vers les poumons. La valve ne fonctionne donc que dans un sens poumons-œsophage [27].

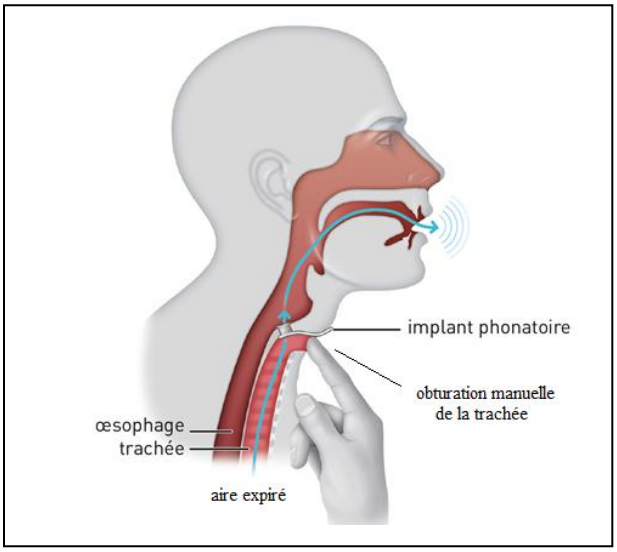


Figure 1.11 : Principe de la voix Trachéo-Œsophagienne [28]

Contrairement à la voix œsophagienne, le son est produit par l'air pulmonaire. Pour parler, il suffit de boucher le *Trachéostome* avec la main en même temps que l'on prononce des mots ou des phrases. Quand on bouche l'orifice trachéal, le souffle est dévié vers l'œsophage par l'implant, c'est pourquoi cette voix est appelée *Trachéo-œsophagienne* [27,28].

Le principal avantage de cette voix de remplacement est assez proche de la voix laryngée en fluidité et intensité. La mise en place d'un implant phonatoire lors de la Laryngectomie Totale, permet de parler rapidement après l'intervention sans passer par un véritable apprentissage. Néanmoins, les inconvénients de cette technique sont importants : nécessité de fermer l'implant pour parler donc impossibilité de parler avec des mains libres, nécessité d'un entretien quotidien, changements périodique de la

prothèse (entre 6 et 10 mois, en général) et risque de mycose autour et derrière l'implant, etc. [28].

1.7.3.3. Voix Prothétique

Ce mode de réhabilitation nécessite un appareil appelé *électrolarynx* ou *laryngophon*e, à transmission électrique vibratoire. L'extrémité est appliquée sur le cou, et la simple articulation du mot est amplifiée par l'appareil (Figure 1.12). L'usage du laryngophone n'est possible que si la voix chuchotée est correcte. La voix obtenue est plus *robotique* mais permet une très bonne compréhension [29].

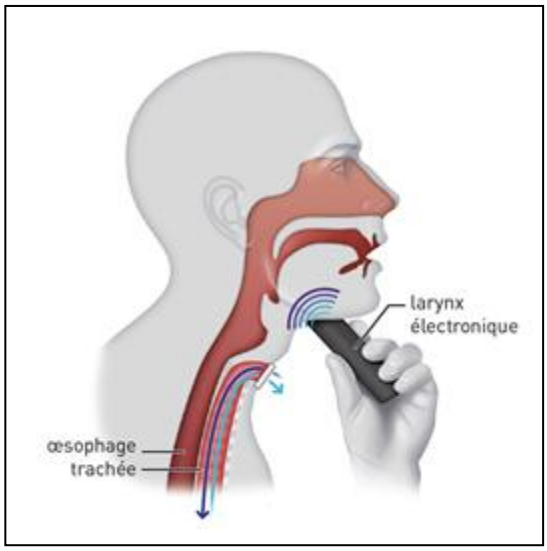


Figure 1.12 : Schéma représentant la technique de la Voix Prothétique [29]

1.8. Conclusion

L'étude anatomique et physiologique de l'appareil phonatoire est un préalable indispensable à l'approche, la compréhension, et la prise en charge des troubles de la voix. C'est pour cette raison qu'elle est présentée dans ce chapitre. Cependant, cette étude sera en permanence orientée vers la fonction phonatoire, et l'accent sera mis sur les éléments anatomiques et physiologiques importants pour cette fonction.

Chapitre 2

Méthodes Objectives d'Évaluation de la Qualité de la Voix

2.1. Introduction

La voix est un phénomène complexe multidimensionnel, son évaluation doit être subjective et objective pour une meilleure fiabilité aux résultats. Dans ce chapitre, nous allons aborder, d'abord, quelques méthodes d'évaluation subjective de la voix qui coïncident avec les méthodes d'évaluation objectives, dans la réalisation d'une évaluation complète et fiable de la voix. Nous expliquons ensuite en détail l'évaluation objective par l'analyse acoustique qui est le but de notre travail.

Dans la seconde partie de ce chapitre, nous allons présenter les Réseaux de Neurones Récurrents RNN, l'une des techniques de l'évaluation objective, nous exposons leurs principes de fonctionnement. Nous allons étudier avec détail les Réseaux Récurrents à Mémoire à Long et à Court Terme (LSTM), une variante de RNN.

2.2. Evaluation de la qualité de la voix

L'évaluation de la qualité de la voix est un enjeu important pour le phoniatre ou l'orthophoniste dans le but de valider la pertinence et l'efficacité des traitements proposés qu'il s'agisse de la rééducation ou de la phono-chirurgie. Deux principales méthodes en général sont utilisées pour l'évaluation de qualité de la voix :

- Une évaluation *Subjective* de la voix, par le patient lui-même avec une échelle spécifique de handicap ressenti (***V**oice **H**andicap **I**ndex : VHI*) et une évaluation par le praticien selon une échelle perceptuelle de GRBAS ;
- Une évaluation dite *Objective* regroupant une *Analyse Acoustique* et Aérodynamique du signal vocal ainsi que des méthodes de classification automatique de voix pathologiques.

2.2.1. Evaluation Subjective de la voix

De nombreux orthophonistes, phoniatres sont d'accord qu'une évaluation subjective de la voix doit s'appuyer sur : des informations recueillies à partir d'une anamnèse, une échelle d'auto-évaluation et une échelle d'analyse perceptive [15].

2.2.1.1. Données de l'anamnèse

L'objectif de l'anamnèse, dans le cadre d'un bilan vocal, est de recueillir toutes les données nécessaires pour situer le problème vocal dans son contexte. L'anamnèse porte sur [15] :

- **Histoire de la voix du patient**: les problèmes vocaux qu'il a déjà rencontrés. En effet, bien souvent, le trouble vocal s'inscrit dans la durée. Ainsi, certains phoniatres supposent que de nombreux adultes dysphoniques ont déjà rencontré des problèmes avec leur voix lorsqu'ils étaient enfants.

- **Histoire médicale du patient**: maladie respiratoire (asthme, insuffisance respiratoire), interventions passées (chirurgie abdominale, thoracique cervicale, faciale, cérébrale…), maladie neurologique et cérébrale, conduites adductives pouvant porter atteinte à la fonction laryngée (tabagisme, alcoolémie), troubles hormonaux, troubles d'ordre psychiatrique. En effet, il est essentiel de connaître toutes les pathologies, séquelles, traitements et autres, pouvant avoir des répercussions sur la voix.

- **Le contexte social et professionnel**: (situation sociale du patient, utilisation quotidienne de sa voix, conditions de travail, etc.).

2.2.1.2. Echelle auto-évaluation

L'auto-évaluation est une technique qui permet à un individu de porter un jugement sur ses capacités vocales. Des échelles d'auto-évaluation de la fonction vocale et de la qualité de vie relative à la voix ont été développées et validées.

Il existe plusieurs questionnaires de qualité de vie globale, avec des groupes de questions relatifs chacun à un domaine particulier de la vie. Actuellement trois questionnaires validés de qualité de vie, spécifiquement orientés vers la qualité de la voix et vers les conséquences de la dysphonie dans la vie quotidienne sont disponibles à savoir le *Voice Outcome Survey (VOS)*, le *Voice-Related Quality of Life (V-RQOL)* et le Voice *Handicap Index (VHI)* [15,30,31]. Ce dernier, est l'outil d'évaluation le plus utilisé [32], il est constitué de 30 items regroupés en trois sous-échelles de dix items chacune : *fonctionnelle* (impact du trouble vocal sur les activités quotidiennes), *émotionnelle* (impact psychologique) et *physique* (perceptions personnelles des caractéristiques physiques de la voix). Une grille de réponses à cinq degrés de sévérité est proposée allant de 0 (non, jamais de problème) à 4 (oui, toujours un problème). Le score de chaque sous-échelle varie entre 0 et 40, le VHI total varie entre 0 et 120. Plus le score est élevé, plus l'handicap lié au problème vocal est sévère. Un score inférieur ou égal à 10 points est considéré comme normal [30,33].

L'avantage de ces questionnaires de qualité de vie est d'obtenir l'avis subjectif du patient sur son état. Ils ont pour but de révéler les différents handicaps ressentis par le patient lui-même, qui ne correspondent pas toujours à ceux perçus par le médecin ou le chirurgien.

2.2.1.3. Échelle d'évaluation perceptuelle

L'évaluation perceptive a pour but d'analyser les voix d'un point de vue esthétique, phonétique et physiologique. L'évaluation perceptive est la méthode la plus utilisée en pratique clinique algérien pour évaluer la voix ; elle est toujours considérée comme la méthode de référence.

De nombreuses méthodes d'analyses perceptives ont été proposées pour l'évaluation de la qualité de la voix [34,35,36]. Parmi ces méthodes, c'est l'échelle *GRBAS* de Hirano (1981) qui est couramment utilisée en pratique phoniatrique [35]. Elle repose sur l'appréciation de cinq caractères de la voix pathologique, chacun noté selon quatre niveaux (***0*** représente la voix normale, ***1*** une dysphonie légère, ***2*** une dysphonie moyenne et ***3*** une dysphonie sévère) :

- ***G*** ou *Grade* : appréciation globale de la qualité de la voix ;
- ***R*** ou *Roughness* : impression audible d'irrégularités des cycles vibratoires interprétée par la raucité de la voix ;
- ***B*** ou *Breathness* : impression audible de fuite d'air en phonation (*souffle*) ;
- ***A*** ou *Asthenicity* : fatigue vocale ou voix hypotonique ;
- ***S*** ou *Strain* : voix forcée ou voix hypertonique.

L'évaluation du grade général, même réduit à une échelle à quatre niveaux est entachée d'une *variabilité* importante *inter et même intra auditeur*. L'évaluation d'un échantillon de voix peut varier d'un niveau pour le même auditeur à différents moments, et dans les mêmes conditions.

La variabilité inter auditeur est encore plus importante par le fait qu'ils sont de culture et d'école cliniques différentes et que chaque phoniatre définit à l'usage, et parfois inconsciemment, ses propres critères subjectifs. L'évaluation inter auditeur, à l'écoute absolue, c'est-à-dire sans référence de comparaison, peut atteindre plus de 50 % d'erreur [6]. Les scores s'améliorent énormément avec des jurys d'écoute, mais

ces derniers sont très lourds à mettre en œuvre et impossibles en pratique clinique de routine.

2.2.1.4. Représentations graphiques du signal sonore

Plusieurs représentations graphiques du signal peuvent être utiles à l'évaluation et l'analyse du signal vocal. Même si, l'étude qualitative de ces graphiques n'est pas un moyen suffisant pour évaluer la qualité de la voix, il est indispensable de savoir interpréter et mettre en relation les informations visuelles qu'ils apportent.

2.2.1.4.1. Enveloppe du son

L'enveloppe d'un son est une courbe représentant l'amplitude, autrement dit, l'intensité, ou encore l'énergie acoustique de ce son, en fonction du temps (Figure 2.1).

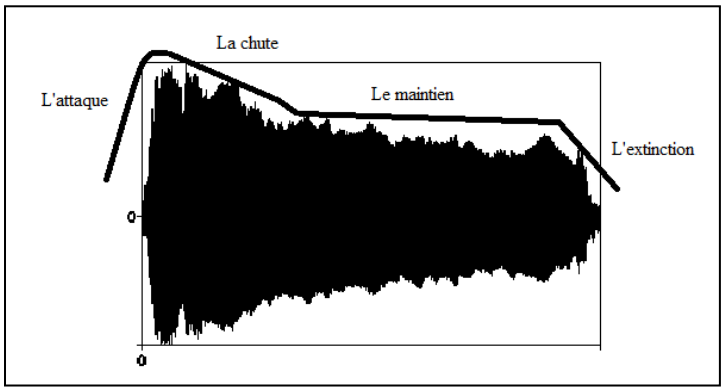


Figure 2.1 : Enveloppe d'une voyelle [a] tenue pour une voix féminine normale

En observant une plus petite portion du signal, il est possible d'observer les périodes du son et leur régularité (Figure 2.2).

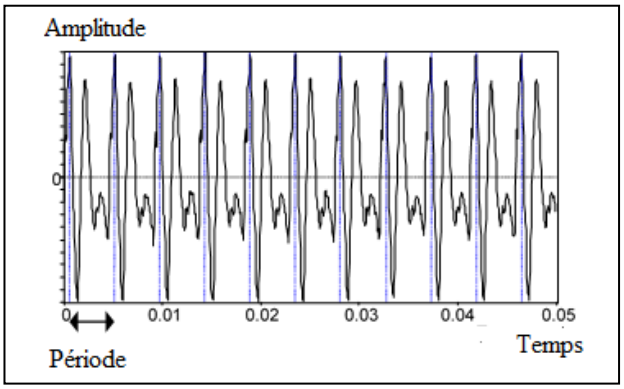


Figure 2.2 : Fenêtre très courte de l'enveloppe d'une voyelle [a] tenue

2.2.1.4.2. Spectrogramme

L'enveloppe d'un son est un aspect classique de courbe en deux dimensions. Le *spectrogramme,* lui, représente le son en trois dimensions. En abscisse, on trouve le temps, et en ordonnée, la fréquence. La troisième dimension, l'intensité, est symbolisée par l'aspect plus ou moins foncé du tracé.

Cet outil se révèle précis, informatif et fiable pour analyser les caractéristiques de la production sonore. L'idée du spectrogramme est très ancienne. W. Koenig [37] montre en temps réel la distribution de l'énergie en fréquence et en temps de différents échantillons de sons sur un *spectrographe sonore*.

L'ouvrage *Speech Sciences* [38] consacre un chapitre *Acoustic phonetics* à de nombreux exemples de spectrogramme. De son côté, Baken et al. [39] consacrent le chapitre *Sound Spectrography* au spectrogramme, en détaillant la théorie associée et le paramétrage des calculs et de la représentation graphique. L'ouvrage de Cornut et al. [40] donne de nombreuses illustrations à base de spectrogramme en lien avec la voix chantée.

Le spectrogramme a beaucoup intéressé la communauté scientifique ces dernières années comme un champ d'application possible de la reconnaissance des formes, en vue de l'aide au diagnostic. L'analyse des contrastes, profils et changements de formes dans les spectrogrammes a permis à Sharma et al. [41] de séparer 50 enfants normophoniques et 50 enfants avec troubles spécifiques du langage. De manière similaire, Laiba et al. [42] ont pu exploiter les images des spectrogrammes dans le cadre de la maladie de Parkinson pour séparer 150 témoins de 150 patients, avec une fiabilité élevée, sur la base de voyelles, de mots et de monologues.

La figure 2.3 représente le spectrogramme d'une voix normale féminine (a) et celui d'une voix normale d'homme (b). Les harmoniques sont présents sur toute la portion visible. Ces harmoniques sont relativement nets sur les deux graphiques. L'écartement différent des harmoniques sur les deux spectrogrammes s'explique par la différence de fréquence fondamentale entre les deux voix.

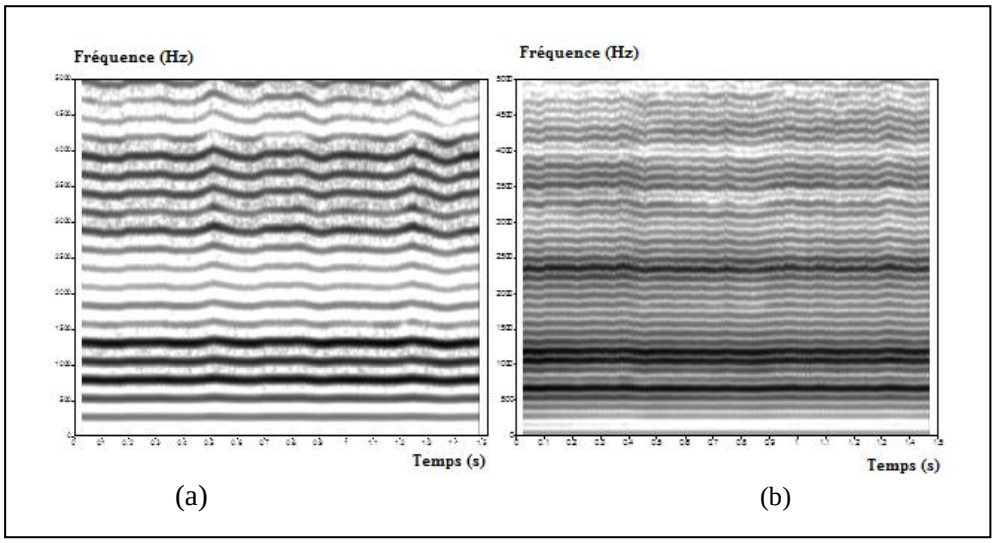


Figure 2.3 : Spectrogrammes de [a]tenue de deux voix normales (a) : Voix féminine, (b) : Voix masculine [39]

Le spectrogramme de la figure 2.4 présente une voix féminine soufflée d'une patiente ayant une PRU. Les harmoniques sont partiellement effacés par le souffle à partir du second formant, situé ici vers 1300 Hz.

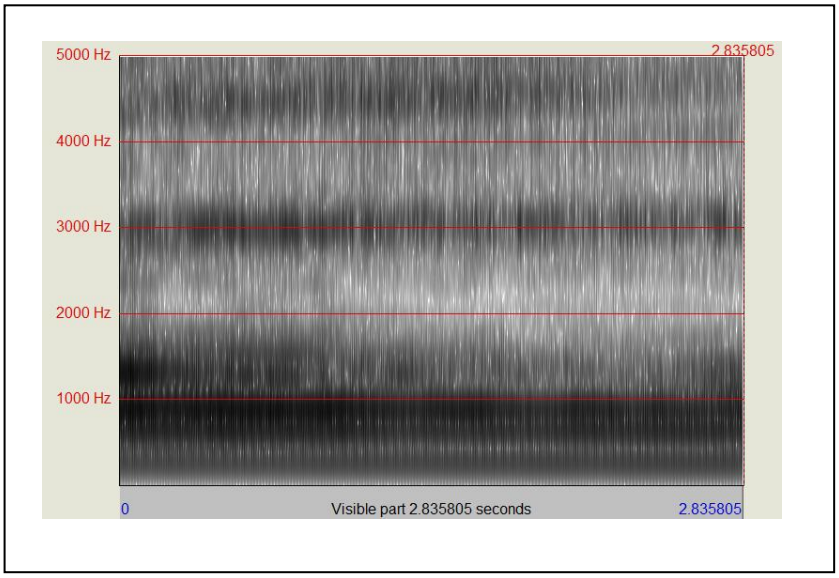


Figure 2.4 : Spectrogramme de [a]tenue d'une voix féminine soufflée (PRU 9)

Sur le spectrogramme de la figure 2.5, on remarque, après la période stable où les harmoniques sont bien dessinés, une nette bitonalité est apparue sur le spectrogramme, qui se traduit à l'oreille par l'impression d'entendre deux sons de hauteurs différentes en même temps.

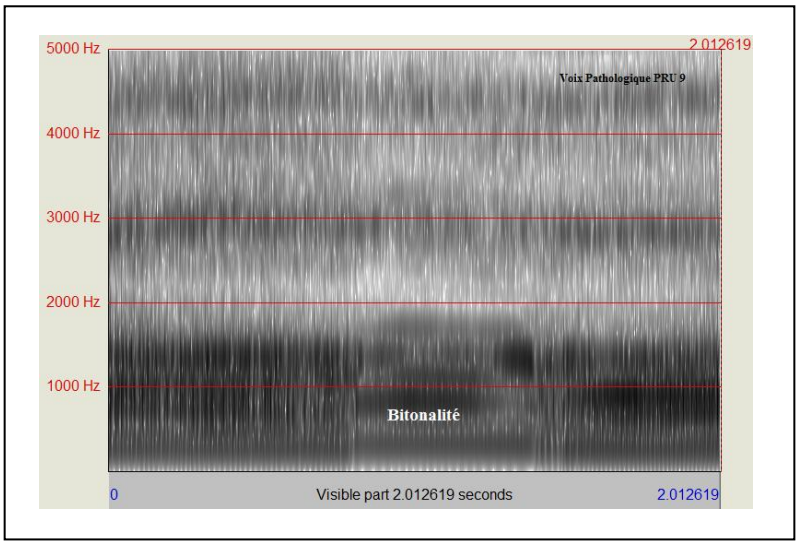


Figure 2.5 : Spectrogrammes de [a] tenue d'une PRU avec bitonalité

2.2.2. Evaluation Objective par l'Analyse Acoustique

L'évaluation objective de la voix basée sur l'analyse acoustique et aérodynamique est une méthode d'exploration fournissant des mesures quantitatives objectives sur un échantillon vocal. Les principaux paramètres altérés dans le cadre d'une dysphonie sont la hauteur (ou fréquence fondamentale), la sonie (ou intensité), et le timbre qui donne toute la couleur à la voix. D'autres paramètres reflètent les perturbations aérodynamiques de la parole comme le débit d'air.

2.2.2.1. Mesures acoustiques de la voix

L'instabilité de la vibration de laryngée est une cause essentielle des dysphonies. Sa mesure a donc une grande importance dans leur évaluation. Sur le plan instrumental, plusieurs paramètres peuvent apporter des informations sur *la stabilité en fréquence et en amplitude de la vibration laryngée.*

2.2.2.1.1. Fréquence Fondamentale moyenne

La fréquence fondamentale (F_0) reste le meilleur indicateur des caractéristiques biomécaniques des cordes vocales. La F_0 moyenne apporte une mesure globale de la hauteur de la voix du sujet (voix aiguë, grave).

La fréquence moyenne peut être calculée par la relation suivante :

$$F_0 \text{ moyenne (en Hz)} = \frac{1}{N}\sum_{i=1}^{N} F0_i \qquad (2.1)$$

L'instabilité de la F_0 se traduit par des variations de fréquence au cours du temps. Elle se mesure par *l'écart type* de la F_0, qui correspond à l'ampleur en Hz des variations de F_0 autour de la moyenne.

Le ***C**oefficient de **V**ariation* CoV permet de relativiser *l'écart type* en le comparant à la F_0 moyenne. Il correspond donc à l'ampleur en % des variations de F_0 par rapport à la F_0 moyenne. Ainsi, à un écart type de 4.9 Hz pour une F_0 moyenne de 180 Hz correspond un coefficient de variation de 2.7 %, valeur importante. Le même écart type pour une F_0 moyenne de 500 Hz fournit un coefficient de variation de 0.98 %, valeur beaucoup plus normale. Le coefficient de variation de la F_0 est donc le meilleur indice pour explorer la stabilité de la fréquence fondamentale.

L'écart type de F_0 et son CoV sont calculés par les formules mathématiques suivantes :

$$\text{Ecart type (Hz)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (F0_i - F0_{moy})^2} \qquad (2.2)$$

$$\text{CoV (\%)} = \frac{Ecart\ type}{F_0\ moyenne}\ x100 \qquad (2.3)$$

2.2.2.1.2. Stabilité à Court Terme de F_0

Les fluctuations à Court Terme, c'est-à-dire d'une durée de l'ordre d'un cycle glottique, caractérisent surtout les atteintes morphologiques des cordes vocales. On désigne sous le terme générique de *Jitter*, les différentes mesures des variations à Court Terme, d'une période à l'autre, de la fréquence fondamentale (Figure 2.6).

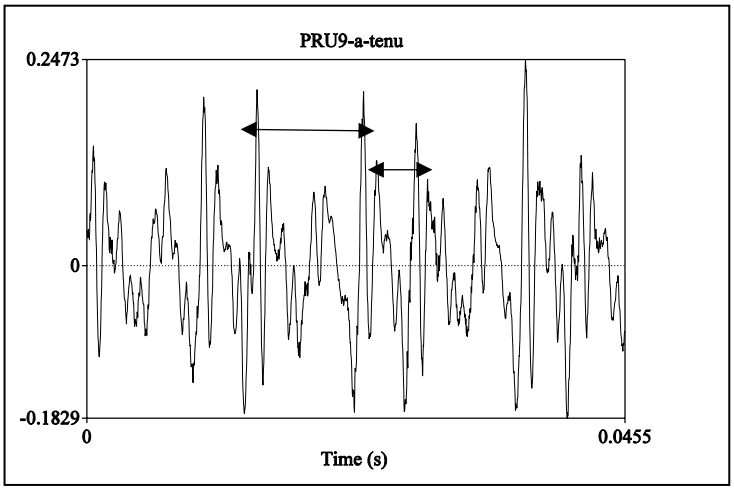


Figure 2.6 : Variations de la vibration de [a] tenue pour un cas de la PRU

Chez le sujet sain, cette variation cyclique est physiologique et inévitable en raison de petites irrégularités mécaniques, tissulaires et fonctionnelles (force et pression légèrement différentes d'un cycle à l'autre, par exemple) [43].

La mesure du Jitter a plusieurs définitions mathématiques et sa valeur dépend de la technique de mesure de la fréquence fondamentale. Sur un même échantillon, elle n'est pas la même avec différents logiciels de mesure. Il existe plusieurs représentations du Jitter [39] :

- le *Jitter absolu* est la moyenne (sur une durée de l'ordre d'une seconde) de la différence de période entre deux cycles vibratoires du larynx consécutifs. Ces variations de fréquence sont mesurées très précisément *cycle à cycle*.

$$Jitter\ absolu\ (s) = \frac{1}{N-1}\sum_{i=1}^{N-1}|To_i - To_{i+1}| \qquad (2.4)$$

Avec N : Nombre de périodes dans une seconde

- Le Jitter absolu est plus grand si la période de la voix est plus longue. Il est donc plus intéressant d'utiliser un *Jitter factor* en divisant la valeur moyenne de la perturbation par la période moyenne. C'est un bon indice pour évaluer la stabilité de la Fo. Il est considéré comme l'indice le plus significatif de la raucité de la voix [39].

$$Jitter\ factor\ (\%) = 100x\ \frac{Jitter\ absole\ (s)}{To_{moy}} \qquad (2.5)$$

Y. Koike et al. ont proposé une autre mesure de Jitter, appelée ***R**elative **A**verage **P**erturbation (RAP)* [44]. Ce paramètre tient compte des variations régulières et volontaires de la fréquence, telles que les variations prosodiques. Dans la formule du Jitter absolu, on compare la durée de chaque période avec celle de la période suivante. Dans le calcul du RAP, on compare la durée de chaque période T_i à une moyenne des trois périodes successives T_{i-1}, T_i et T_{i+1}. Cette *moyenne sur trois points* a pour but d'atténuer les variations volontaires à plus long terme de la fréquence fondamentale.

$$RAP = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{T0_{i-1}+T0_i+T0_{i+1}}{3} - T_i\right|}{T_{0\,moy}} \qquad (2.6)$$

2.2.2.1.3. Stabilité à Court Terme de l'amplitude F_0

Par analogie avec les mesures des variations à court terme de la fréquence fondamentale, le *Shimmer* est une mesure de perturbation à Court Terme de l'amplitude du cycle vibratoire : c'est la différence d'amplitude de cycle à cycle (Figure 2.7).

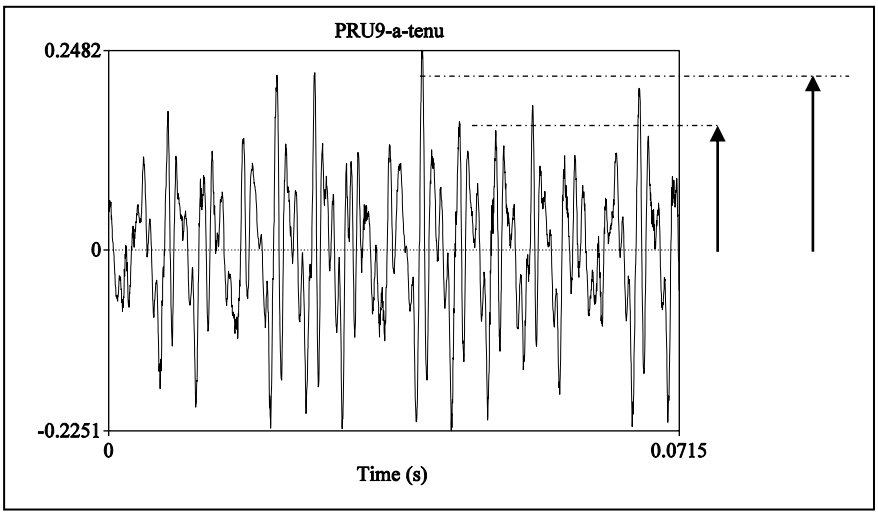


Figure 2.7 : Variation de l'amplitude sur deux périodes consécutives de [a] tenue pour un cas de PRU

- le *Shimmer moyen* exprimée en dB est la moyenne des rapports d'amplitudes entre deux cycles de vibration consécutifs du larynx. Ces variations d'amplitude sont mesurées très précisément *cycle à cycle*.

$$Shimmer\ moyen\ (dB) = \frac{1}{n-1}\sum_{i=1}^{n-1}\left|20.log\frac{A_i}{A_{i+1}}\right| \qquad (2.7)$$

- le *Shimmer factor est* le Shimmer moyen rapporté à l'amplitude moyenne du signal.

$$\textit{Shimmer factor (\%)} = \frac{Shimmer\ moyen}{20\ log\ \sum_{i=1}^{n} \frac{A_i}{n}} \text{ x100} \qquad (2.8)$$

- Les auteurs qui ont mis au point l'algorithme du RAP ont également proposé une mesure analogue des variations d'amplitude, *l'**A**mplitude **P**erturbation **Q**uotient* (APQ) [39]. Dans la formule du Shimmer, on compare l'amplitude d'une période à celle de la période suivante. Pour obtenir l'APQ, on compare l'amplitude d'une période Ti du signal à l'amplitude moyenne des pics des périodes T_{i-5} à T_{i+5} (soit 11 périodes). On calcule la moyenne des valeurs absolues des différences calculées sur l'ensemble du signal, et on divise enfin par la valeur moyenne des pics d'amplitude du signal. La formule présente des similarités avec celle du RAP :

$$APQ = \frac{\frac{1}{N-10}\sum_{i=6}^{N-5}\left|\frac{A_{i-5}+\cdots A_i+\cdots A_{i+5}}{11}\right|}{A_{moy}} \qquad (2.9)$$

2.2.2.2. Souffle de la voix

C'est également un élément très important dans l'évaluation d'une dysphonie. Le souffle de la voix est considéré comme un bruit se superposant au signal vocal de la source laryngienne. Plusieurs paramètres peuvent être utilisés pour évaluer la composante du souffle dans la voix.

2.2.2.2.1. Rapport Harmonique sur bruit

Le rapport entre l'énergie du spectre Harmonique et celle du spectre de Bruit (**H**armonics to **N**oise **R**atio: HNR) est l'un des paramètres pour quantifier le souffle de la voix. Ce dernier peut être un bruit d'écoulement aérodynamique créé par une constriction du conduit vocal ou par un débit d'air trop important. Le bruit peut être dû également à l'instabilité du signal glottique. Le HNR ne donne donc, par principe, une bonne évaluation du bruit de souffle qu'avec une fréquence de vibration stable [45].

Il a été montré que plus la dysphonie est importante, et plus les harmoniques du signal vocal ont tendance à être remplacés par du bruit. Le HNR a pour fonction d'évaluer l'émergence des harmoniques d'un signal par rapport au bruit [45].

Au cours d'une étude de la fonction vocale de patients atteints d'une paralysie laryngée, il a été constaté que le paramètre acoustique HNR est lié à l'insuffisance de l'occlusion glottique (fuite d'air glottique), alors que le Jitter et le Shimmer étaient associés à l'irrégularité des vibrations [46]. On peut conclure en disant que le HNR, d'un point de vue théorique, est un paramètre global, puisqu'il tient compte du souffle et de l'apériodicité, mais qu'il semble aussi que ce paramètre soit plus particulièrement sensible à la composante de souffle.

2.2.2.2.2. High-frequency Power Ratio

Le paramètre HPR (**H**igh-frequency **P**ower **R**atio) est une mesure qui a été introduite par Shoji & coll. [47]. Le HPR permet une comparaison de l'énergie totale de la bande de fréquence 0-6000 Hz à celle de la bande de fréquence 6000-20000 Hz. Ces auteurs ont montré que le seuil de 6000 Hz permet de différencier de manière fiable une voix normale d'une voix soufflée. Cette mesure s'appuie sur les travaux de Baken [39] qui explique que sur le spectrogramme d'une voix produite par un locuteur sain, on constate une diminution progressive de l'énergie des harmoniques au fur et à mesure que l'on monte dans les fréquences. Or on sait aussi que le souffle se répartit sur toutes les fréquences, qu'elles soient basses ou hautes. L'énergie de la composante de souffle apparaîtra donc mieux dans les aigus. C'est pourquoi, la quantité d'énergie présente dans les plus hautes fréquences, chez les sujets sains, est plutôt faible, tandis qu'elle reste élevée chez les sujets pathologiques. L'analyse spectrale montre un spectre de raies bien défini pour un signal vocal de bonne qualité et un spectre continu massif pour une voix soufflée dans les hautes fréquences (Figure 2.8).

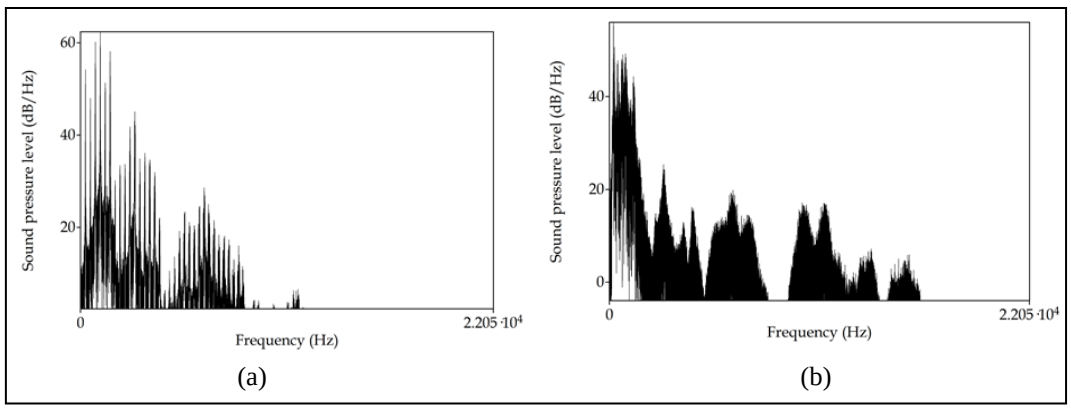


Figure 2.8 : Spectre de la voyelle [a] tenue en fonction de la fréquence (a) : Voix Saine de référence ; (b) : Voix Pathologique soufflée (PRU)

2.2.2.2.3. Différence d'amplitude entre les deux premiers harmoniques

Lors d'un cycle de vibration dit normal, les plis vocaux ont une phase d'ouverture, puis une phase de fermeture, durant laquelle le passage de l'air est bloqué. En revanche, pour la voix soufflée, les plis vocaux ne se ferment pas complètement et laissent donc passer un flot d'air continu entre les plis vocaux. Le tout entraînant ainsi le passage d'un flux d'air plus important par rapport à une voix normale [48,49].

Acoustiquement, cette qualité de voix se traduit par la présence d'une forte amplitude du premier harmonique H_1 par rapport au deuxième H_2 [50,51].

La voix des patients qui présentent une PR se caractérise par une voix soufflée. Cette qualité de voix peut être évaluée par la différence, en amplitude, entre les deux premiers harmoniques H_1–H_2. Cette différence est importante pour une voix pathologique et à peu près nulle pour une voix normale (Figure 2.9).

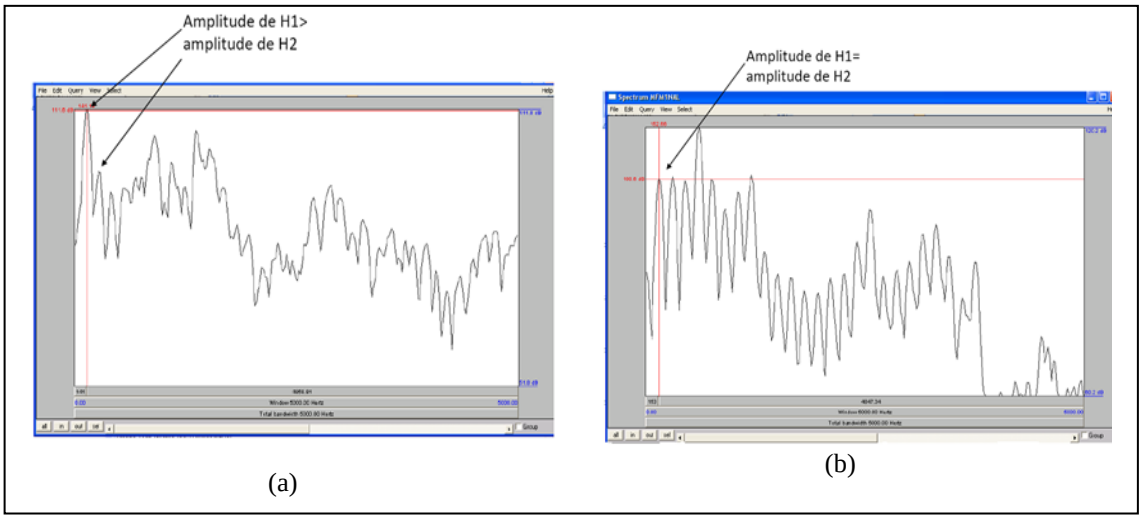


Figure 2.9 : Différence d'amplitude entre le premier et le deuxième harmonique (a) : Voix Pathologique, (b) : Voix Normale

2.2.2.2.4. Peak de Proéminence Cepstrale

Le Peak de Proéminence Cepstrale CPP (**C**epstral **P**eak **P**rominence) est un paramètre acoustique décrit initialement par Hillenbrand et Houd en 1994 [52]. Cette valeur permet de quantifier le degré d'harmonie d'un échantillon vocal donné [53,54].

Le Cepstre utilisé dans l'analyse de la voix et de la parole est donné par la transformée de Fourier inverse du spectre acoustique. Ce processus peut être compris

intuitivement comme un spectre d'un spectre. Tout d'abord, la forme d'onde est transformée par Fourier dans le domaine spectral. Ensuite, le logarithme de ce spectre est pris, et une autre transformée de Fourier (inverse) est effectuée dans le domaine Cepstrale.

Les pics harmoniques périodiques dans le spectre sont représentés comme un seul grand pic (et ses harmoniques) dans le Cepstre autour d'une quéfrence correspondant à la période du signal vocal (Figure 2.10). La hauteur (c'est-à-dire la proéminence) de ce pic par rapport à une ligne de régression à travers le Cepstre global est appelée la *Proéminence du Pic Cepstral* ou CPP et est généralement rapportée en unités de décibels. Les valeurs du CPP se situent donc dans une plage continue, où des valeurs plus faibles sont généralement corrélées à des niveaux plus élevés de dysphonie.

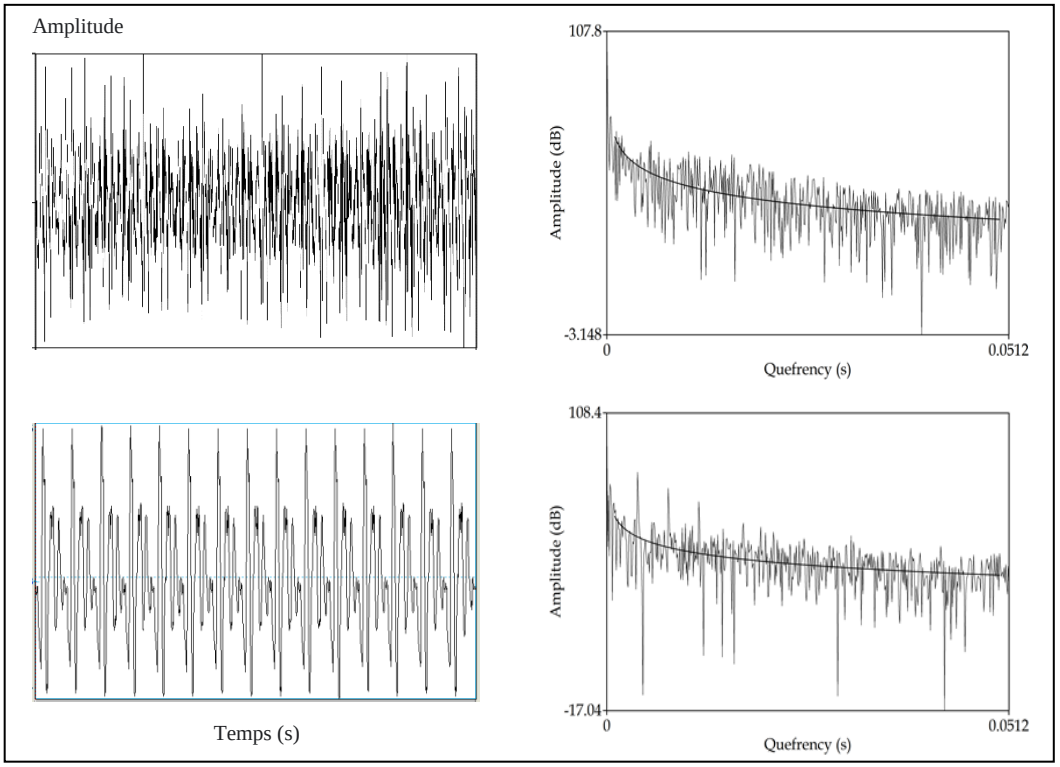


Figure 2.10 : Oscillogramme et Cepstrum d'une Voix Pathologique (PRU) en haut et Voix Normale de référence en bas.

Il existe également une variante du CPP dite *lissée* ou CPPs (**S**mooth **C**epstral **P**eak **P**rominence) dans laquelle des opérations de lissage sont effectuées à la fois dans les domaines temporel et Cepstral [55]. Plus le signal vocal sera périodique, plus la valeur du CPPs sera élevée [56].

2.2.2.3. Mesures aérodynamiques

L'analyse aérodynamique consiste à évaluer la capacité d'un sujet à produire et gérer le souffle phonatoire. Les évaluations aérodynamiques sont d'une grande richesse dans l'étude des dysfonctionnements laryngiens mais restent malheureusement peu accessibles à cause du manque de matériels bien étudiés pour ce faire, et de leur coût. Cela s'explique par le fait que les capteurs aérodynamiques, en particulier de débit, sont plus complexes qu'un simple microphone. Les paramètres aérodynamiques sont: la pression sous glottique, les débits d'air et le Temps Maximale de Phonation.

2.2.2.3.1. Pression sous-glottique

La pression sous-glottique est un facteur important de contrôle de l'intensité vocale et de la hauteur. Elle est produite grâce aux poumons, et est influencée par une plus ou moins grande résistance des cordes vocales au passage du flux d'air pulmonaire, lors de la phonation. Ainsi, en cas de béance glottique, les poumons doivent produire une pression plus importante pour obtenir une intensité phonatoire suffisante malgré la déperdition d'air. La pression sous-glottique est plus importante, en moyenne, chez des sujets dysphoniques que chez des sujets sains. La pression sous-glottique au cours de la phonation est mesurable par deux façons [57] :

- directement, au moyen d'une seringue hypodermique reliée à un capteur de pression. Il s'agit d'un moyen très efficace mais très invasif ;
- indirectement, par une estimation fondée sur la mesure de la pression intra-orale, que l'on capte au moyen d'un tuyau souple relié à un capteur de pression.

2.2.2.3.2. Débit d'air buccal

Le débit oral d'air expiré est la vitesse d'un volume d'air ou la quantité de l'air expiré par la bouche par unité de temps, exprimé en l/s, ml/s ou dm^3 /s [39]. C'est un paramètre physique primordial pour la phonation, car il traduit la consommation d'air nécessaire pour la production vocale et intervient dans l'évaluation du rendement

énergétique de celle-ci. Les constrictions le long du conduit vocal produisent un orifice déterminant le débit : le débit de l'air est modulé par la glotte et les constricteurs supra-glottiques (c'est-à-dire la langue, le palais, le nez, les lèvres et les dents). Il se différentie du débit d'air nasal, écoulement de l'air par le nez, mesurable pour des voyelles et consonnes nasales par exemple.

L'équipement utilisé pour le mesurer varie selon les auteurs : spiromètre, débitmètre et pneumotachographes, ce dernier équipement faisant l'objet de la grande majorité des études. Le pneumotachographe consiste en un débitmètre et un transducteur différentiel de pression. À l'une de ses extrémités, il se termine par un masque posé sur le visage du sujet de manière imperméable grâce à un joint étanche en caoutchouc. Le sujet tient le masque par son manche. À son autre extrémité, il est connecté à un amplificateur et à une station informatique traitant les données de débit résultantes [58] (Figure 2.11).

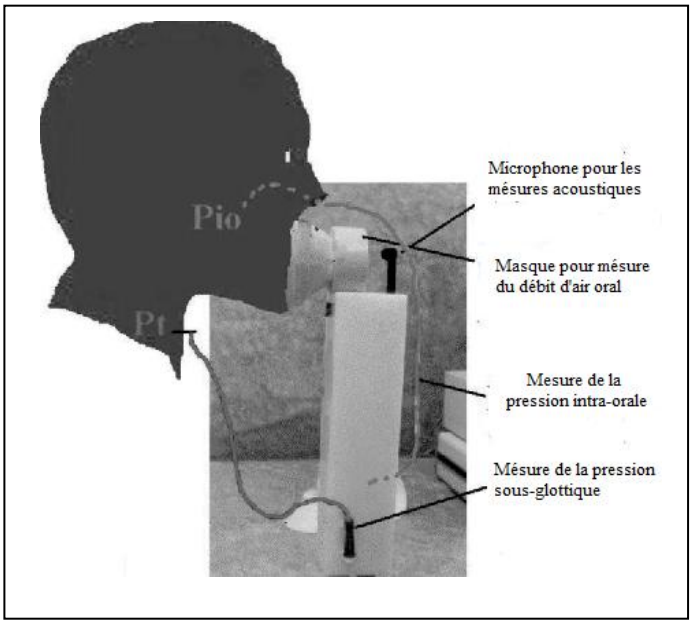


Figure 2.11 : Position du sujet pour l'enregistrement des données aérodynamiques [58]

2.2.2.3.3. Temps Maximum de Phonation

Comme dans le cas du débit d'air buccal, on comprend facilement que le ***T**emps **M**aximum de **P**honation* (TMP) dépend à la fois de la capacité pulmonaire et de la présence éventuelle d'une fuite glottique lors de la phonation. Une faible capacité

pulmonaire ou une déperdition d'air glottique entraînent nécessairement une réduction du TMP. D'autres facteurs influencent la valeur du temps maximum de phonation, il risque de diminuer si l'intensité vocale est élevée, car le maintien de cette intensité demande une pression sous-glottique importante [59,60].

On peut mesurer le temps maximum de phonation au moyen d'un chronomètre, mais pour des raisons de rapidité, nous avons préféré le mesurer visuellement dans la fenêtre de l'éditeur de sons du logiciel Praat.

2.2.2.4. Choix des échantillons destinés à l'analyse objective

Les évaluations acoustiques se font sur le phonème [a] tenu. Les avantages en sont multiples. C'est le phonème le plus spontané dans toutes les langues du monde (car le plus simple à articuler). La voyelle [a] fait partie des voyelles les plus énergétiques, avec le contenu harmonique très riche. Il minimise l'influence de la charge acoustique du conduit vocal sur la vibration du larynx. Il permet surtout d'évaluer la stabilité et le bruit du vibrateur en régime permanent [39,61]. A. Giovanni et al. soulignent qu'il est important de travailler sur un *signal stabilisé pendant le temps des mesures*. La fréquence et l'intensité du signal doivent être aussi stables que possible lorsque l'on veut analyser les micro-perturbations de ces deux paramètres [61].

La voyelle [a] tenue sert à la fois, à l'analyse acoustique et à l'analyse aérodynamique, pour mesurer le TMP. Pour cela, le patient étant censé maintenir la phonation jusqu'à ce qu'il n'ait plus de souffle du tout. L'attaque, et l'extinction du [a] tenue doivent donc être éliminés avant l'analyse, pour ne conserver que sa partie stable [1,2].

2.2.3. Evaluation Objective par Réseaux de Neurones

L'Intelligence Artificielle (IA) est en plein développement ces dernières années. Comprendre les dernières avancées dans ce domaine revient à étudier deux concepts très populaires tels que l'Apprentissage Automatique ou *Machine Learning* (ML) et l'Apprentissage Profond ou *Deep Learning* (DL).

Le ML est un champ d'étude de l'IA qui vise à donner aux machines la capacité d'*apprendre* à partir de données, via des modèles mathématiques. Plus précisément, il

s'agit du procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'apprentissage.

Le but de cette phase est l'obtention des paramètres d'un modèle qui atteindront les meilleures performances, notamment lors de la réalisation de la tâche attribuée au modèle. Une fois l'apprentissage réalisé, le modèle pourra ensuite être déployé en production.

Le DL est un sous-ensemble du domaine de ML qui utilise les Réseaux de Neurones Artificiels (RNA) (Figure 2.12). Le Big Data Analytics et le DL sont deux domaines intéressants de la science des données. Le Big Data est devenu important parce que de nombreux domaines utilisent des quantités massives d'informations, utiles pour résoudre des problèmes, tels que l'intelligence nationale, la cyber-sécurité, la détection des fraudes, le marketing, l'informatique médicale, la vidéo et parole, etc.

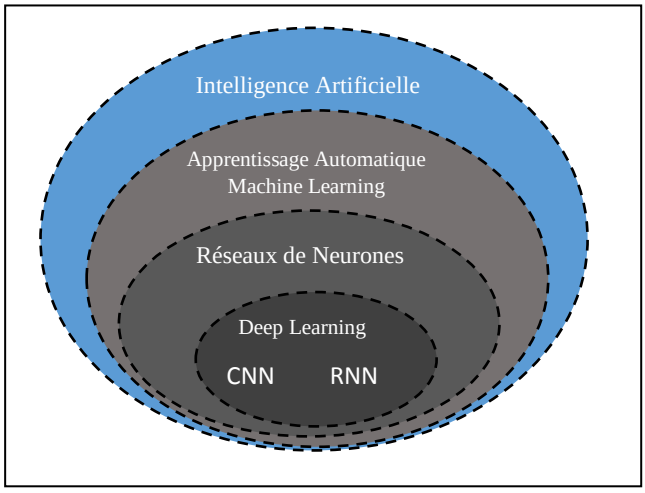


Figure 2.12 : Intelligence Artificielle et ses Dérivées

Un des principaux avantages du DL est l'analyse et l'apprentissage des quantités massives de données**.** Ce qui représente un outil précieux pour le Big Data Analytics où les données brutes sont en grande partie non marquées et non-classées. Les solutions de DL ont donné des résultats remarquables dans différentes applications de ML**,** y compris la Reconnaissance Automatique de la Parole, la vision par ordinateur, et le Traitement Automatiques du Langage naturel, etc.

2.2.3.1. Différents types de ML

Dans le domaine du ML, il existe deux principaux types d'apprentissages: *supervisé* et *non supervisé*.

2.2.3.1.1. 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.

L'apprentissage supervisé est généralement effectué dans le contexte de la classification et de la régression :

- **Classification:** Un problème de classification survient lorsque la variable de sortie est une catégorie, telle que pathologique et non pathologique ;
- **Régression**: Un problème de régression se pose lorsque la variable de sortie est une valeur réelle, telle que les prévisions des prix en bourses.

2.2.3.1.2. Apprentissage non supervisé

L'apprentissage non supervisé consiste à ne disposer que de données d'entrée (X) et pas de variables de sortie correspondantes. 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. On l'appelle apprentissage non supervisé car, contrairement à l'apprentissage supervisé ci-dessus, il n'y a pas de réponse correcte ni d'enseignant. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données.

2.2.3.2. Apprentissage Profond

Dans les algorithmes d'apprentissage classiques, des caractéristiques doivent être extraites des données brutes afin d'effectuer la tâche d'apprentissage. Le but étant

d'avoir une représentation plus haut niveau des données. L'extraction de caractéristiques à partir des données brutes demande de bonnes connaissances sur celles-ci et sur la tâche d'apprentissage. Cette opération est relativement coûteuse à la mise en place, dépend du contexte et une mauvaise extraction des caractéristiques mène à de très mauvaises performances en terme d'apprentissage.

L'idée des architectures profondes consiste à intégrer cette extraction de caractéristiques, normalement faite *manuellement*, par un processus d'apprentissage dans les premières couches du réseau de neurones. Dans un réseau de neurone MLP, les couches intermédiaires permettent de transformer la représentation des données d'entrée en une représentation plus haut niveau. Durant la phase d'apprentissage, chaque couche d'un MLP apprend une représentation de son entrée qui doit être intéressante pour les couches suivantes. Les informations contenues dans chacune de ces couches vont devenir de plus en plus haut niveau. Le terme profond réfère donc au nombre de couches des réseaux de neurones profonds entre l'entrée et la sortie. Un réseau avec une seule couche cachée est appelé réseau *peu profond*, et un réseau avec plus de 2 couches cachées est dit *profond*.

2.2.3.3. Réseaux de Neurones Artificiels

Les Réseaux de Neurones sont des modèles d'apprentissage automatique capables de représenter une relation entre des données d'un espace d'entée X et un espace de sortie Y. Ils sont utilisés dans de nombreux domaines, comme la vision assistée par ordinateur, le traitement du langage naturel, traitement automatique de la parole, etc. L'unité de calcul de base est le neurone. Celui-ci prend en entrée plusieurs signaux et les interprète pour envoyer un nouveau signal vers d'autres neurones ou vers la sortie du réseau de neurones, c'est-à-dire la sortie du modèle.

2.2.3.3.1. Modèle du perceptron

Dans sa version la plus simple, le perceptron est un réseau de neurones composé de seulement un neurone, qui prend n entrées. Chacune de ses entrées i est pondérée par un poids noté w_i. Le neurone peut prendre les états *1* ou *0* respectivement actif ou non-actif, en fonction de ses entrées pondérées, et d'un biais noté b. Cet état représente la sortie du modèle (figure 2.13). La sortie d'un perceptron pour un vecteur $\boldsymbol{x}$ en entrée est calculée tel que [62]:

$$y_i = f(\sum_{j=1}^{n} w_{ij}\, x_{ij} + b_i) \qquad (2.10)$$

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
y_i : sortie du neurone i.

La fonction d'activation *f* est appelée également fonction de transfert, elle sert à introduire une non-linéarité dans le fonctionnement d'un neurone.

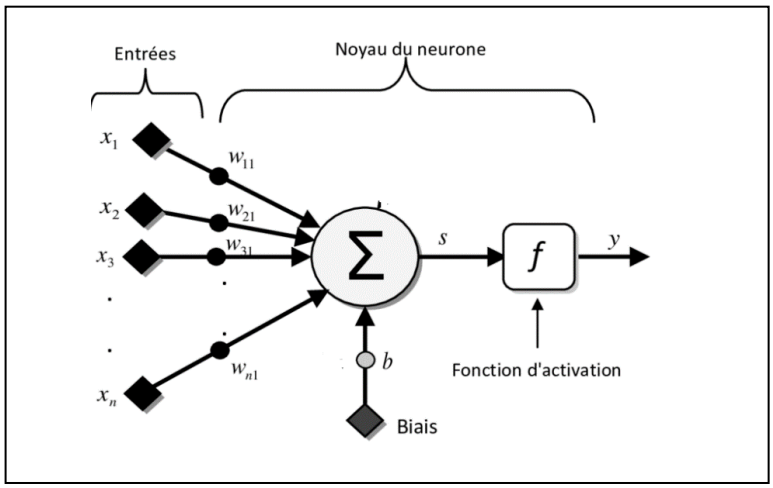


Figure 2.13 : Modèle de neurone formel (perceptron) [62]

2.2.3.3.2. Perceptron Multi-Couches

Un perceptron multicouche MLP (***M**ulti **L**ayer **P**erceptrons*) est composé de plusieurs couches de neurones et de connexions (Figure 2.14). Ce nombre est au moins égal à deux, signifiant ainsi que le réseau possède deux couches de poids connexionnistes, une couche de sortie et une cachée. Le nombre de couches cachées détermine la complexité des frontières des différents sous-espaces que le réseau pourra représenter. La complexité de l'approximation est également déterminée par le nombre de neurones de chaque couche puisque ce nombre détermine le nombre maximal d'informations que le réseau peut extraire du signal traité. La couche d'entrée, correspondant le plus souvent à un vecteur de données issu d'une phase de prétraitement, n'est pas véritablement considéré comme appartenant au réseau.

Dans le réseau MLP, les informations, ou activations, circulent dans un seul sens, c'est-à-dire des neurones d'une couche aux neurones de la couche suivante. 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.

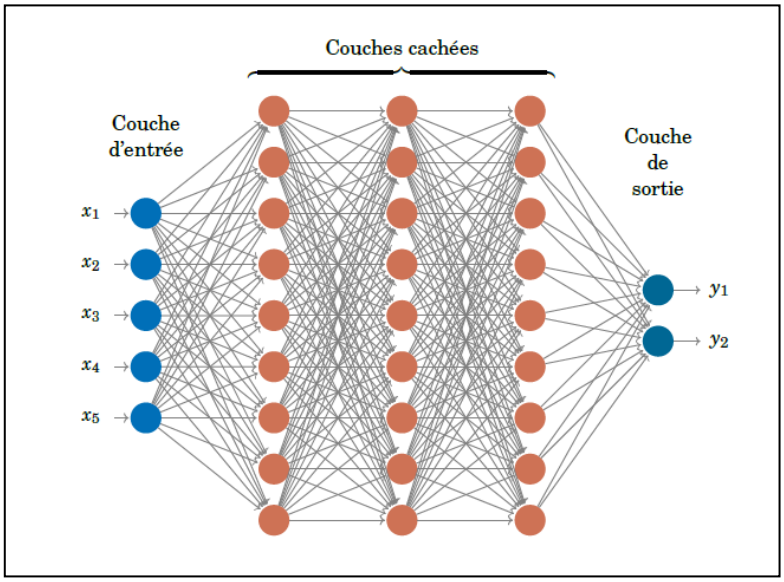


Figure 2.14 : Réseau MLP [62]

2.2.3.4. Réseaux de Neurones Profonds

L'apprentissage en profondeur est un sous-domaine de l'apprentissage automatique et les réseaux de neurones constituent l'épine dorsale des algorithmes d'apprentissage en profondeur. Un Deep Neural Network (DL), ou réseau de neurones profond, est composé d'au moins deux couches de neurones cachées. En fait, le nombre de couches, ou la profondeur, permet aux réseaux de neurones de résoudre des problèmes complexes [63].

2.2.3.4.1. Réseau de Neurones Récurrents

Les Réseaux de Neurones Récurrents RNN *(**R**ecurrent **N**eural **N**etworks)* font partie des algorithmes de DL. Ils ont une architecture d'apprentissage profond capable de prédire des séries temporelles ou des séquences, c'est-à-dire des signaux variables dans le temps.

Cette architecture est largement utilisée dans des applications récentes et dans de nombreux domaines. Nous citons par exemples, la prévision de la météo, le traitement automatique du langage naturel pour faire de la traduction automatique, la

reconnaissance automatique de la parole, l'analyse de sentiments à partir d'un texte ou d'un fichier audio, etc.

Un RNN est très similaire à un réseau de neurones classique, dans lequel le flux des activations se dirige dans un sens unique, depuis la couche d'entrée vers la couche de sortie. Son architecture est composée d'une couche d'entrée, de couches cachées et d'une couche de sortie. En revanche, le RNN se distingue par la capacité de revenir en arrière dans le réseau, les données de sortie sont multipliées par un poids et réinjectées dans l'entrée (Figure 2.15).

Expliquons le principe de fonctionnement du RNN à partir d'un RNN constitué par un seul neurone. Ce neurone reçoit des entrées et produit une sortie, à chaque étape temporelle ***t***, ce neurone récurrent reçoit le vecteur d'entrée ***x(t)*** ainsi que sa propre sortie produite à l'étape temporelle précédente ***y(t-1)***.

Chaque neurone récurrent possède deux types de poids:

- des poids reliant les neurones de l'entrée à la sortie comme pour un réseau de neurones classique (W_x) ;
- des poids entre la sortie et l'entrée de la couche elle-même, qui sont les connexions récurrentes (W_y).

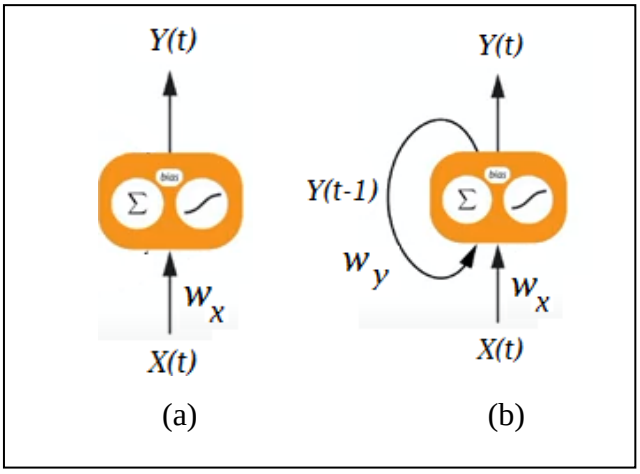


Figure 2.15 : Un neurone simple (a) et un neurone récurrent (b) [64]

Nous pouvons représenter ce réseau le long d'un axe du temps. On dit alors qu'on a déplié le réseau dans le temps. La sortie y_{t4} est une fonction de x_{t4} et de y_{t3}, qui est une fonction de x_{t3} et de y_{t2}, qui elle-même est une fonction de x_{t2} et de y_{t1}, ... Par conséquent, y_{t4} est une fonction de toutes les entrées depuis l'instant t=0. Lors de la première étape temporelle, à t=0, les sorties précédentes n'existent pas (en général, sont supposées nulles). En fait, la sortie d'un neurone récurrent étant une fonction de toutes les entrées des étapes précédentes. On considère que ce neurone possède une forme de mémoire (Figure 2.16).

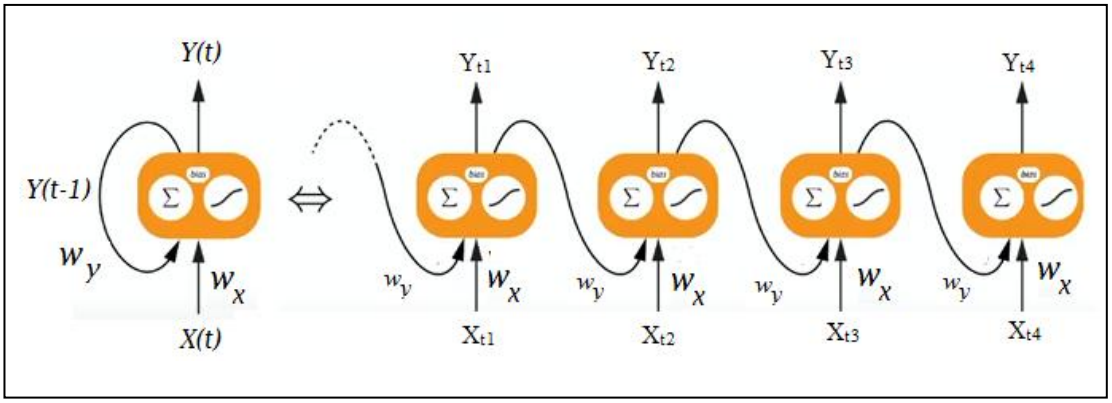


Figure 2.16 : Principe de fonctionnement d'un neurone Récurrent [64]

Le problème avec les RNN est qu'au fur et à mesure que le temps passe et qu'ils sont alimentées par de plus en plus de nouvelles données, ils commencent à *oublier* les données précédentes qu'ils ont vues, car ils se diluent entre les nouvelles données, la fonction de transformation à partir de l'activation et la multiplication des poids. Cela signifie qu'ils ont une bonne mémoire à court terme, mais un problème lorsqu'ils essaient de se souvenir de choses qui se sont produites il y a un certain temps (données qu'ils ont vues à plusieurs reprises dans le passé). Nous avons besoin d'une sorte de mémoire à long terme, ce que les LSTM (***L**ong **S**hort-**T**erm **M**emory*) fournissent.

2.2.3.4.2. Réseaux Récurrents LSTM

Un réseau récurrent à Mémoire Court et Long Terme (LSTM), est l'architecture de réseau de neurones récurrents la plus utilisée en pratique, et qui permet de répondre au problème de disparition de gradient dans le réseau RNN de base. Le réseau LSTM a été proposé par Sepp Hochreiter et Jürgen Schmidhuber en 1997. Ils ont une structure

cellulaire plus complexe qu'un neurone récurrent normal, ce qui leur permet de mieux réguler la façon d'apprendre ou d'oublier à partir des différentes sources d'entrée [65].

La base du fonctionnement d'une cellule LSTM est le Réseau Récurent RNN standard (Figure 2.17).

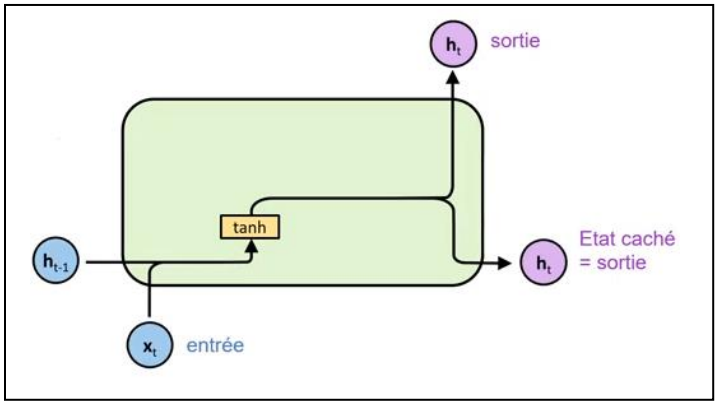


Figure 2.17 : Cellule d'un Réseau RNN [65]

L'état caché de ce dernier est actualisé par la formule suivante :

$$h_t = tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \qquad (2.11)$$

Avec: *tanh: fonction d'activation tangente hyperbolique*

W_{xh} *: les poids synaptiques entre l'entrée et la sortie du neurone*

W_{hh} *: les poids synaptiques entre la sortie et l'entrée de la couche elle-même, qui sont les connexions récurrentes*

Le plus souvent, ce réseau ne gère pas bien les dépendances à long terme à cause de la disparition du gradient. Pour cela, on introduit une nouvelle variable $\mathbf{c_t}$ qui permettra de se souvenir des informations pendant de longues périodes (Figure 2.18)

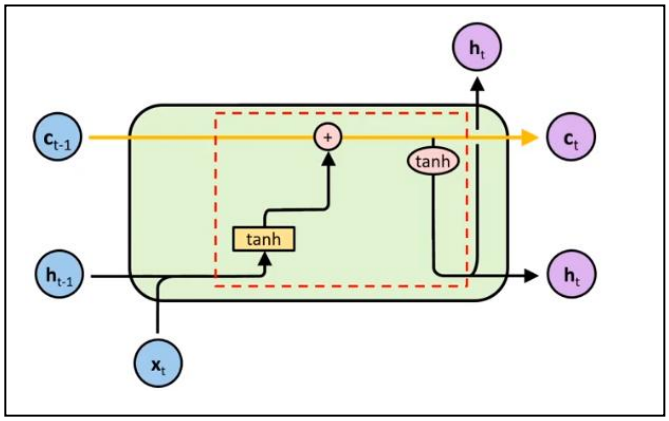


Figure 2.18 : Introduction d'une Mémoire dans la cellule RNN de base

La mise à jour des variables de l'état caché $\boldsymbol{c_t}$ et la sortie $\boldsymbol{h_t}$ est réalisée selon les équations suivantes :

$$c_t = c_{t-1} + tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \qquad (2.12)$$

$$h_t = tanh\,(c_t) \qquad (2.13)$$

L'inconvénient de cette architecture est que le passé est toujours important avec le présent. L'idée est donc de permettre au réseau de garder ou d'oublier les informations du passé selon qu'elles soient déterminantes ou pas.

La solution adoptée est de pondérer la mémoire du passé représentée par $\boldsymbol{c_t}$ par un réseau de neurones avec une seule couche cachée Sigmoïde. Ce réseau va apprendre à oublier les informations non pertinentes. Il est appelé Porte d'Oubli (*Forget Gate*) (Figure 2.19).

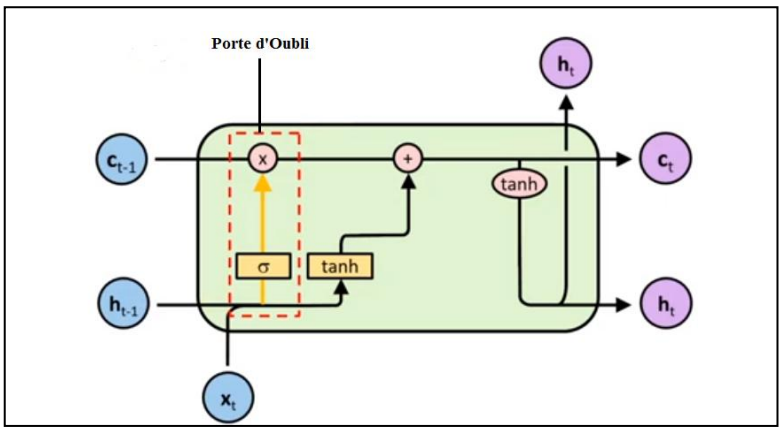


Figure 2.19 : Porte d'oubli ajoutée dans la cellule RNN de base

Les équations (2.14), (2.15) et (2.16) permettent de définir : la porte d'oubli $\boldsymbol{f_t}$, l'état caché de la cellule $\boldsymbol{c_t}$ et sa sortie $\boldsymbol{h_t}$.

$$f_t = \sigma\,(W_{hf}\, h_{t-1} + W_{xf}\, x_t) \tag{2.14}$$

$$c_t = f_t * c_{t-1} + \tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \tag{2.15}$$

$$h_t = \tanh\,(c_t) \tag{2.16}$$

Avec: σ*: fonction d'activation Sigmoïde.*

Une autre porte est ajoutée pour pondérer la mise à jour de l'état de la cellule $\boldsymbol{c_t}$ par la sortie ***tanh*** prenant comme entrée $\boldsymbol{h_{t-1}}$ et $\boldsymbol{x_t}$. Cette porte appelée Porte d'entrée (*Input Gate)*, va apprendre à utiliser, à ignorer ou moduler les informations d'entrées selon leur importance (Figure 2.20).

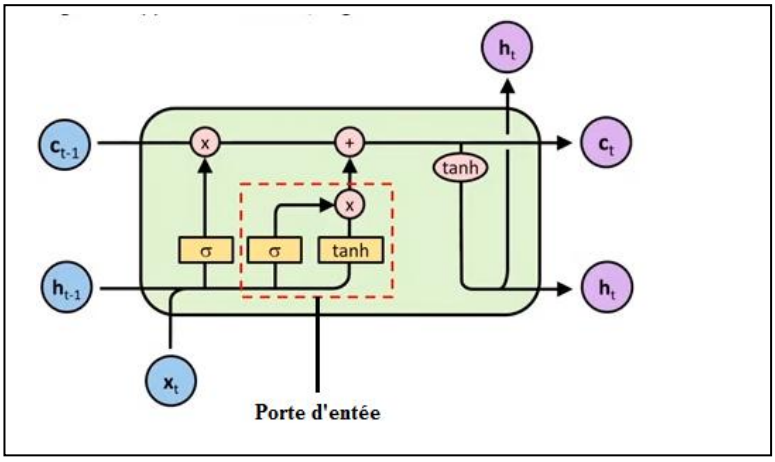


Figure 2.20 : Introduction d'une porte d'entrée dans la cellule RNN

Les équations (2.17), (2.18) et (2.19) permettent de définir : la porte d'entrée $\boldsymbol{i_t}$, l'état caché de la cellule $\boldsymbol{c_t}$ et sa sortie $\boldsymbol{h_t}$.

$$i_t = \sigma\,(W_{hi}\, h_{t-1} + W_{xi}\, x_t) \tag{2.17}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \tag{2.18}$$

$$h_t = \tanh\,(c_t) \tag{2.19}$$

De la même manière, une Porte de Sortie (*Output Gate*) est ajoutée pour pondérer la mise à jour de l'état caché $\boldsymbol{h_t}$. Cela permet de décider quelles informations, l'état caché va porter (Figure 2.21).

Les équations (2.20), (2.21) et (2.22) pour définir : la porte de sortie o_t, l'état caché de la cellule LSTM c_t et sa sortie h_t.

$$o_t = \sigma\,(W_{ho}\, h_{t-1} + W_{xo}\, x_t) \tag{2.20}$$

$$c_t = ft * c_{t-1} + i_t * tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \tag{2.21}$$

$$h_t = o_t * tanh\,(c_t) \tag{2.22}$$

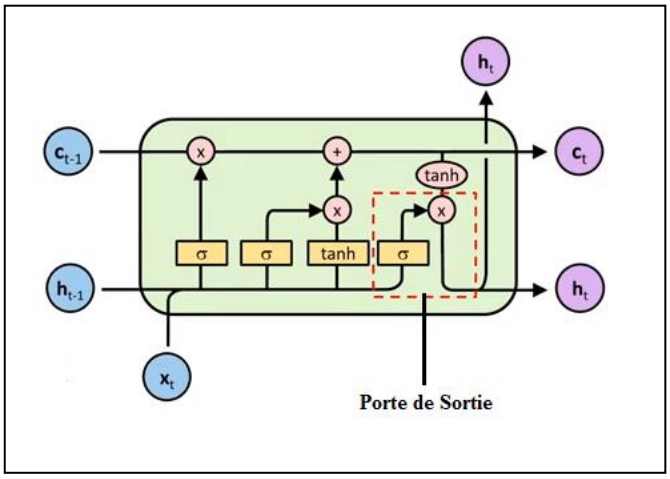


Figure 2.21 : Introduction d'une Porte de sortie dans la cellule RNN [65]

En résumé, une cellule LSTM est composée de trois portes : porte d'oubli f_t, porte d'entrée i_t et porte de sortie o_t. Elle est constituée également de deux types de sorties appelées états : état caché c_t et état de la cellule h_t (Figure 2.22).

La *Porte d'Oubli* décide de quelle information doit être conservée ou jetée: l'information de l'état caché précédent est concaténé à la donnée en entrée, puis on y applique la fonction sigmoïde afin de normaliser les valeurs entre 0 et 1. Si la sortie de la sigmoïde est proche de 0, cela signifie que l'on doit oublier l'information et si on est proche de 1 alors il faut la mémoriser pour la suite

La *Porte d'Entrée* a pour rôle d'extraire l'information de la donnée courante. Nous appliquons en parallèle une sigmoïde aux deux données concaténées h_{t-1} et x_t avec la fonction ***tanh***.

- *Sigmoïde* va renvoyer un vecteur pour lequel une coordonnée proche de 0 signifie que la coordonnée en position équivalente dans le vecteur concaténé n'est pas importante. A l'inverse, une coordonnée proche de 1 sera jugée importante ;
- *Tanh* va simplement normaliser les valeurs (les écraser) entre -1 et 1 pour éviter les problèmes de surcharge de l'ordinateur en calculs ;
- Le produit des deux permettra donc de ne garder que les informations importantes, les autres étant quasiment remplacées par 0.

Avant d'aborder la dernière porte (porte de sortie) o_t, nous expliquons d'abord l'état caché c_t, car la valeur de la sortie est calculée en fonction de cet état caché.

L'état de la cellule se calcule à partir de la porte d'oubli et de la porte d'entrée : on multiplie coordonnée à coordonnée la sortie de l'oubli avec l'ancien état de la cellule. Cela permet d'oublier certaines informations de l'état précédent qui ne servent pas pour la nouvelle prédiction à faire. Ensuite, on additionne le tout (coordonnée à coordonnée) avec la sortie de la porte d'entrée, ce qui permet d'enregistrer dans l'état de la cellule ce que le LSTM (parmi les entrées et l'état caché précédent) a jugé pertinent.

Dans la dernière étape, la porte de sortie doit décider qui sera le prochain état caché, qui contient des informations sur les entrées précédentes du réseau. Pour ce faire, le nouvel état de la cellule calculé juste avant est normalisé entre -1 et 1 grâce à ***tanh***. Le vecteur concaténé de l'entrée courante avec l'état caché précédent passe, pour sa part, dans une fonction sigmoïde dont le but est de décider des informations à conserver (proche de 0 signifie que l'on oublie, et proche de 1 que l'on va conserver cette coordonnée de l'état de la cellule).

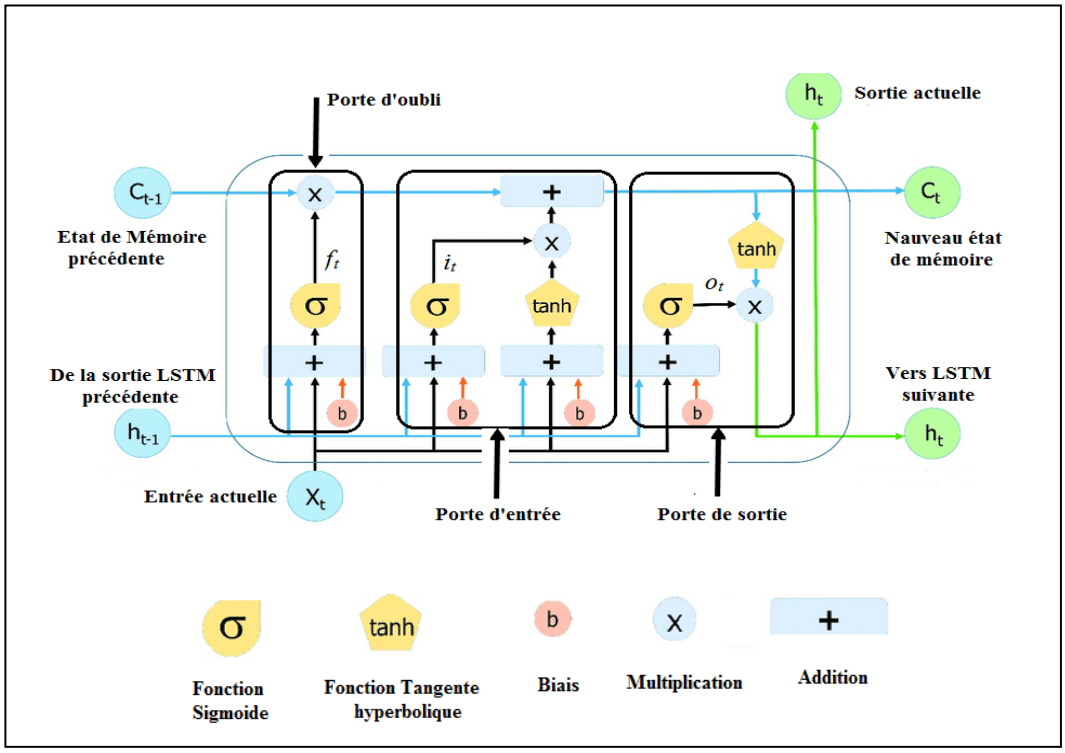


Figure 2.22 : Fonctionnent d'un neurone LSTM à l'instant t [65]

Les équations mathématiques ci-dessous résument le comportement global de la cellule LSTM. Les équations (2.23), (2.24) et (2.25) permettent de définir trois portails : une porte d'oubli, une porte d'entrée et une porte de sortie. L'équation (2.26) calcule le nouvel état de la cellule, et l'équation (2.27) calcule la nouvelle sortie.

$$f_t = \sigma\,(W_{hf}\, h_{t-1} + W_{xf}\, x_t) \tag{2.23}$$

$$i_t = \sigma\,(W_{hi}\, h_{t-1} + W_{xi}\, x_t) \tag{2.24}$$

$$o_t = \sigma\,(W_{ho}\, h_{t-1} + W_{xo}\, x_t) \tag{2.25}$$

$$c_t = f_t * c_{t-1} + i_t * tanh\,(W_{hh}\, h_{t-1} + W_{xh}\, x_t) \tag{2.26}$$

$$h_t = o_t * tanh\,(c_t) \tag{2.27}$$

2.3. Conclusion

L'évaluation de la qualité de la voix a toujours été la préoccupation clinique principale des phoniatres. Comme dans les autres disciplines médicales, ils ont été attentifs à toutes les techniques et méthodes qui seraient susceptibles de leur donner des

informations complémentaires, pour aider au diagnostic et évaluer les effets des traitements chirurgicaux ou médicamenteux ou les progrès d'une rééducation.

Nous avons donné, dans ce chapitre un aperçu global sur le principe de fonctionnement de deux méthodes objectives d'évaluation de voix pathologiques : l'analyse acoustique basée sur des mesures physiques et les Réseaux de Neurones, en particulier les Réseaux Récurrents à Mémoire Long et Court Terme LSTM, que nous avons appliqué dans notre travail.

Chapitre 3

Evaluation Objective de la Voix Pathologique par l'Analyse Acoustique

3.1. Introduction

Le but de la première partie de ce chapitre est de présenter la mise en place du dispositif expérimental pour une évaluation objective par l'analyse acoustique, tels que le choix du matériel, conditions d'enregistrement, le corpus utilisé et le protocole d'enregistrement. Ensuite, des mesures acoustiques sont effectuées sur des voix pathologiques : avant, en cours et après rééducation. Nous avons choisi pour cette étude deux pathologies, une de type neurologique et l'autre morphologique.

Dans la seconde partie, nous présentons les différents résultats obtenus. Ces derniers sont commentés et interprétés afin de dégager une relation technique et des explications scientifiques entre les mesures effectuées et les phénomènes physiopathologiques.

3.2. Population choisie

Pour la dysphonie de type neurologique, nous avons choisi la PRU. La population retenue est constituée de neuf patientes algériennes âgées entre 42 et 56 ans (Figure 3.1), dont six, présentent une paralysie gauche, et qui ont suivies un protocole de rééducation au niveau du service ORL du centre Hospitalo-Universitaire de Bab El Oued, Alger. (Table. 3.1).

Tableau.3.1 : Population Pathologique pour la PRU

Sujets Pathologiques	Sexe	Age	Type de Pathologie	
			PRU Droite	PRU Gauche
PRU 1	F	42	x	
PRU 2	F	56		x
PRU 3	F	43		x
PRU 4	F	47		x
PRU 5	F	53		x
PRU 6	F	52		x
PRU 7	F	53	x	
PRU 8	F	45	x	
PRU 9	F	48		x

Un enregistrement est effectué durant une période de réhabilitation vocale entre 20 et 32 semaines. Toutes les patientes ont bénéficié de séances de rééducation orthophoniques qui visent l'amélioration de la dysphonie, la dyspnée et de la déglutition, à raison d'une séance par semaine, avec une durée de 30 min par séance.

Le nombre de séances était en fonction de l'évaluation clinique et de la vitesse d'apprentissage du patient. Seules les patientes qui avaient suivi un protocole de réhabilitation régulier ont été incluses dans cette étude. Le même corpus a été prononcé par 3 locutrices normales âgées entre 40 et 50 ans, ne présentant pas de troubles de la voix (norme de référence).

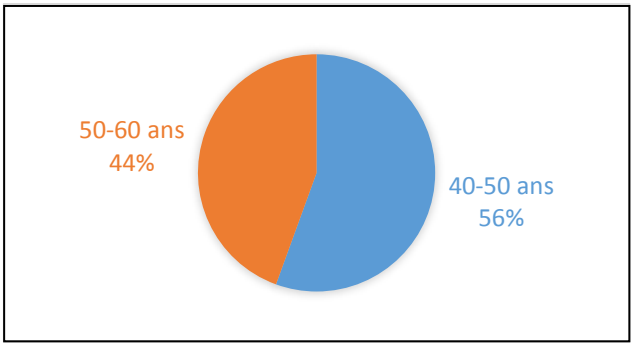


Figure 3.1 : Répartition des patientes de la PRU en fonction des tranches d'âge

La population choisie pour la deuxième pathologie est constituée de neuf patients hommes algériens âgés entre 47 et 68 ans, traités par Laryngectomie Totale pour un cancer avancé du larynx au service ORL de l'Hôpital Universitaire de Béni Messous-Alger (Figure 3.2). Le tabac était la cause de tous les neufs patients. Ils ont reçu leurs traitements complémentaires (soins médicaux), 6 semaines après l'intervention. Une période de rééducation vocale entre 9 et 12 mois est réalisée au sein de l'unité de rééducation orthophonique.

La technique de récupération de la voix adoptée après l'opération chirurgicale est la parole œsophagienne. Lors de l'opération chirurgicale, une ouverture permanente, appelée *trachéostome*, est pratiquée à la base du cou. Cette ouverture permet au patient de respirer en générant donc cette nouvelle voix. Pour la voix de référence, le même corpus a été prononcé par 3 locuteurs hommes algériens, normaux âgés entre 40 et 58 ans, ne présentant pas de troubles de la voix.

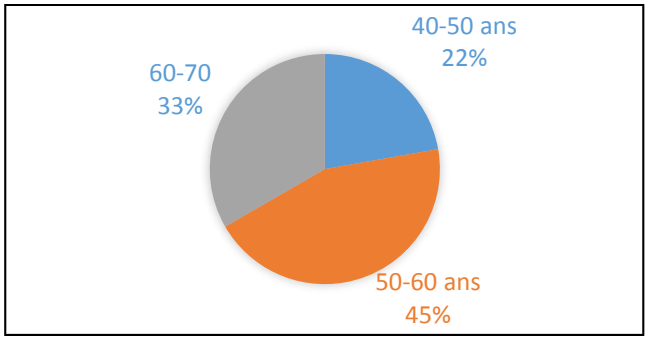


Figure 3.2 : Répartition des patients de la LT en fonction des tranches d'âge

3.3. Matériel d'enregistrement

Le matériel d'enregistrement détermine la qualité du son, et influe sur l'analyse acoustique et donc sur l'évaluation objective de la voix pathologique.

3.3.1. Microphone utilisé

. Il est déconseillé d'utiliser un microphone bas de gamme connecté directement sur l'entrée micro de l'ordinateur : le résultat pourra être convenable à l'oreille mais les logiciels d'analyse fournissent des mesures altérées et non fiables aux évaluations de la voix

Dans notre travail, l'enregistrement est effectué par un microphone électrodynamique à bobine mobile de type *Sennheiser e815S* unidirectionnel de type *Cardioïde*. La réponse de fréquence varie entre 80 et 15 kHz pour une sensibilité de 1.5 mV/Pa, ce qui permet au microphone d'atténuer efficacement les bruits de manipulation, et assurer une excellente réjection des sons parasites environnants ainsi que la réverbération de la salle.

3.3.2. Interface audio externe

Il est indispensable de connecter le microphone sur un préamplificateur externe et non directement à l'ordinateur vu que ces dispositifs périphériques génèrent beaucoup moins de bruit et possèdent un excellent traitement numérique du signal [1]. Pour cela, le corpus de voix a été enregistré avec une carte son externe *M-audio pro* reliée à un Port USB, qui donne un Rapport Signal sur Bruit (SNR) de 100 dB avec 16 bits de résolution, et les fréquences d'échantillonnage jusqu'à 96 kHz. La fréquence d'échantillonnage choisie est de 44.1kHz permettant de préserver le maximum

d'informations portées par les ondes acoustiques. Les enregistrements sont effectués avec le logiciel du son *Sound Forge* version 10.

3.4. Conditions et protocole d'enregistrement

Les enregistrements des voix ont été réalisés dans une salle acoustiquement calme afin d'éliminer les sources sonores parasites. Lors de l'enregistrement des épreuves, la distance entre le microphone et la bouche du patient était fixée à 5 cm. Le microphone est placé à 45° latéralement par rapport à la bouche, son gain a été ajusté pour éviter la saturation et d'assurer une qualité optimale de l'enregistrement. Le patient doit prendre une forte inspiration et tenir la voyelle [a] à hauteur et à intensité confortables, le plus longtemps possible. Aucune démonstration n'est préalablement donnée au patient, pour ne pas influencer la hauteur.

3.5. Utilisation du logiciel Praat en pratique clinique

Praat est un logiciel d'analyse de la parole développé par P. Boersma et D. Weenik, de l'Institut des Sciences Phonétiques de l'Université d'Amsterdam. Il est disponible sur les plates-formes Windows, Macintosh et Unix [66]. Le logiciel propose de nombreux modes de représentation graphique du signal sonore, tels que les enveloppes sonores, les spectres instantanés, ou encore les spectrogrammes à bande large et à bande étroite. Ces différents outils sont couramment employés dans l'étude de la pathologie vocale.

Praat représente l'outil parmi les plus utilisés pour l'extraction des paramètres acoustiques et pathologiques de la parole et la visualisation des courbes respectives [67]. Toutes les données acoustiques du Praat sont quantifiées et permettent une évaluation précise des indicateurs pathologiques. De nombreuses études ont montré une forte corrélation entre les paramètres acoustiques de Praat et d'autres logiciels d'analyse comme le Vocalab et le **M**ulti-**D**imensional **V**oice **P**rogram (MDVP) [68, 69, 70]. Toutefois, Praat est limité par l'absence de traitement de l'information en temps réel.

3.6. Analyse acoustique

Les paramètres acoustiques étudiés ont été extraits à l'aide du logiciel Paat version 6.1. Nous avons choisi une durée de deux secondes correspondant à la partie la plus stable du signal enregistré [39]. Comme la dysphonie concerne essentiellement la vibration des cordes vocales, nous avons choisi, dans ce travail, les mesures basées sur

l'instabilité de la fréquence et l'amplitude de la vibration laryngée. Les paramètres acoustiques retenus pour cette étude sont : la fréquence fondamentale moyenne F_0 avec son **C**oefficient de **V**ariation (CoV) qui mesure l'instabilité globale de la voix sur tout l'échantillon, les Formants, le Jitter qui évalue l'instabilité à Court Terme de F_0, l'intensité, le Shimmer qui évalue l'instabilité à Court Terme de l'amplitude de F_0, ainsi que le pourcentage de voisement DUV (**D**egree of **U**nvoiced **V**oice).

Pour les paramètres aérodynamiques et de l'estimation du bruit, nous avons utilisé TMP, HNR, HPR, Breathy Voice (H_1-H_2) et le Pic de Proéminence Cepstral CPP. Pour les deux derniers paramètres, nous avons utilisé l'application *Voice Sauce* implémentée sous Matlab version 2019a.

Nous avons choisi, pour tous les échantillons, de placer la fenêtre d'analyse à 0,5 seconde du début du son, ce qui est, largement suffisant pour supprimer l'attaque du signal. L'analyse porte ensuite sur une portion de 2 secondes, la plus stable (Figure 3.3).

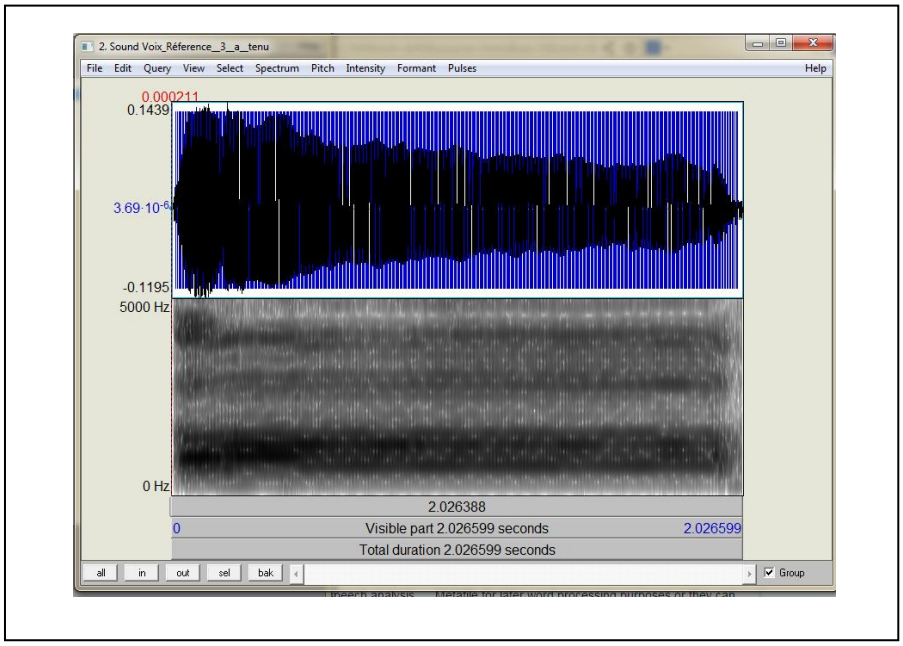


Figure 3.3 : Fenêtre de deux secondes de la voyelle [a] tenue

3.7. Résultats de l'analyse pour la PRU

La figure 3.4 représente l'oscillogramme de la voyelle tenue [a] prise parmi les voix normales de référence (Figure 3.4.a), une portion sélectionnée du même signal (Figure 3.4.b) ainsi que la variation du pitch en fonction du temps (Figure 3.4.c). La

périodicité à court terme du [a] est très bien marquée, ce qui explique la stabilité de la vibration laryngée (Pitch).

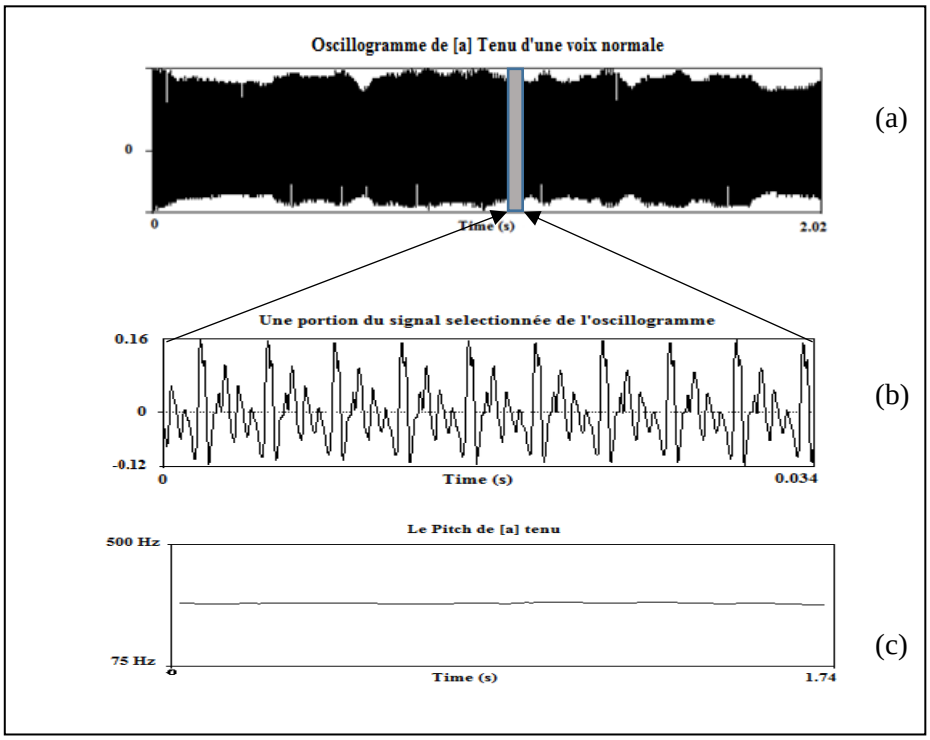


Figure 3.4: La [a] tenue d'une Voix normale,

(a) : Oscillogramme,
(b) : Une portion sélectionnée de l'oscillogramme,
(c) : Le pitch de la même voix

Pour une voix pathologique avant la rééducation, d'une patiente atteinte d'une PLU, la périodicité du signal et son amplitude sont altérées et elles sont nettement visibles sur l'oscillogramme à Court Terme (Figure 3.5.b), ceci donne une voix avec un important degré de raucité perceptible à l'écoute. L'instabilité de F_0 est expliquée par des irrégularités des vibrations des cordes vocales (Figure 3.5.c).

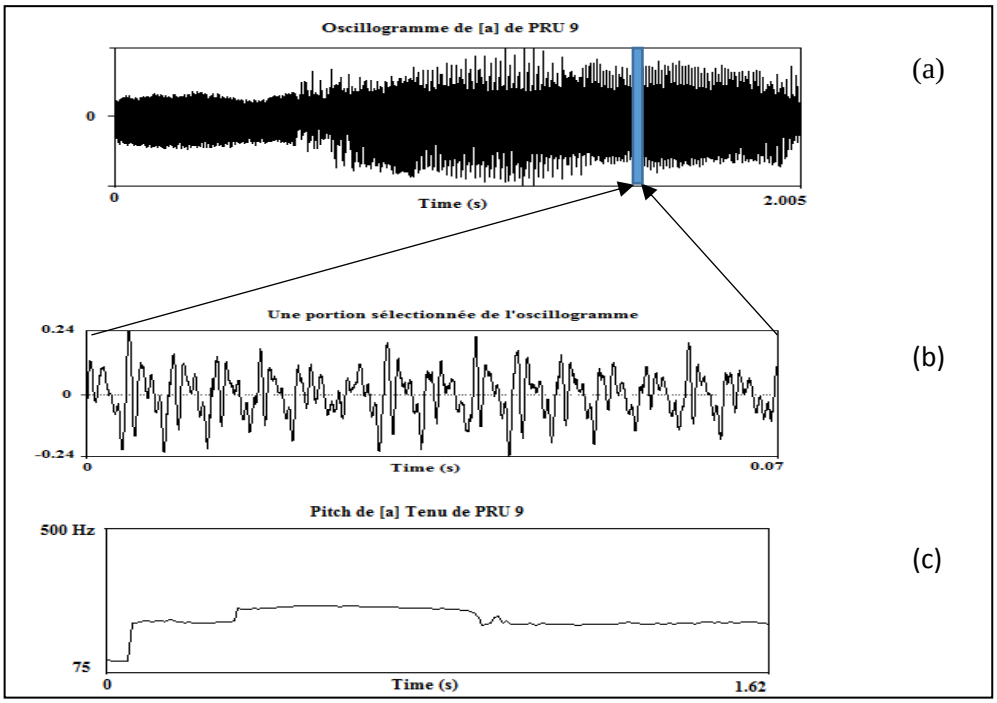


Figure 3.5 : La [a] tenue d'une Voix Pathologique PRU 9,

(a) : Oscillogramme,
(b) : Une portion sélectionnée de l'oscillogramme,
(c) : Le pitch de la même voix.

La figure 3.6 présente un cas de PRU en fin de rééducation qui présente une nette stabilité de la vibration laryngée en fréquence et en amplitude (Figure 3.6. b) avec un Pitch visiblement très stable (Figure 3.6. c).

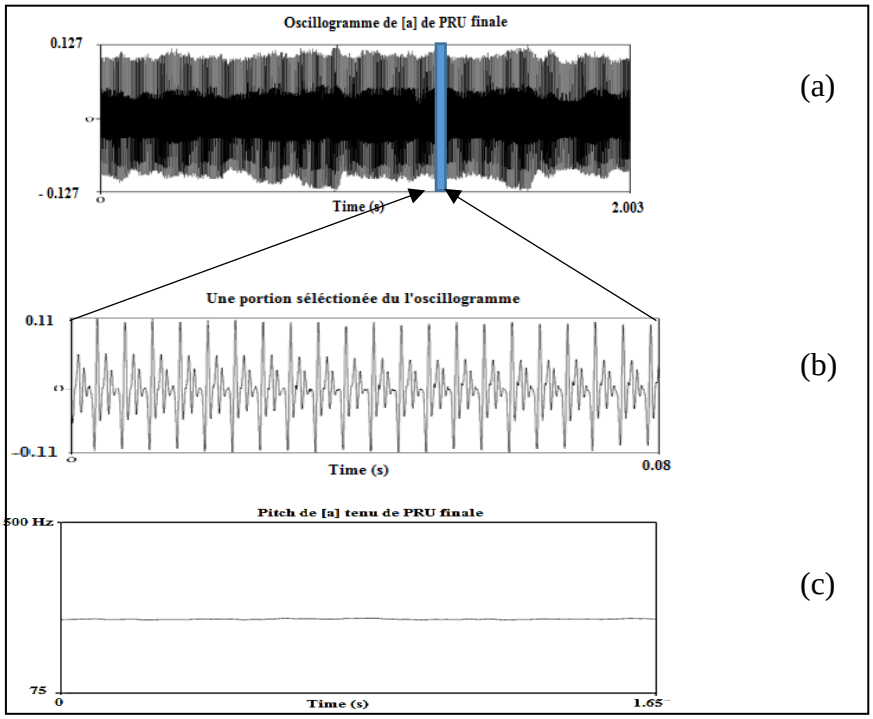


Figure 3.6 : La [a] tenue d'une Voix de la PRU après rééducation,
(a) : Oscillogramme,
(b) : Une portion sélectionnée de l'oscillogramme,
(c) : Le pitch de la même voix.

Les spectrogrammes de la figure 3.7 présentent une comparaison entre deux voix féminines, l'une atteinte d'une PRU gauche (patiente PRU 9) et l'autre celle d'une voix normale (voix de référence 2). L'analyse du spectrogramme de la voix dysphonique (en bas) montre que les premiers formants sont noyés dans le bruit représenté par la couleur blanche et avec un degré important pour le troisième et le quatrième formant. Ces taches blanches correspondent à une voix soufflée. En revanche, les formants dans la voix normale (en haut) sont nettement visibles et facilement repérables.

Nous avons également constaté que des harmoniques viennent s'interposer entre ceux du son fondamental dans le spectrogramme de la voix pathologique. Il s'agit d'une bitonalité particulièrement nette, qui se traduit à l'oreille par la perception d'un changement involontaire de hauteur (Figure 3.7).

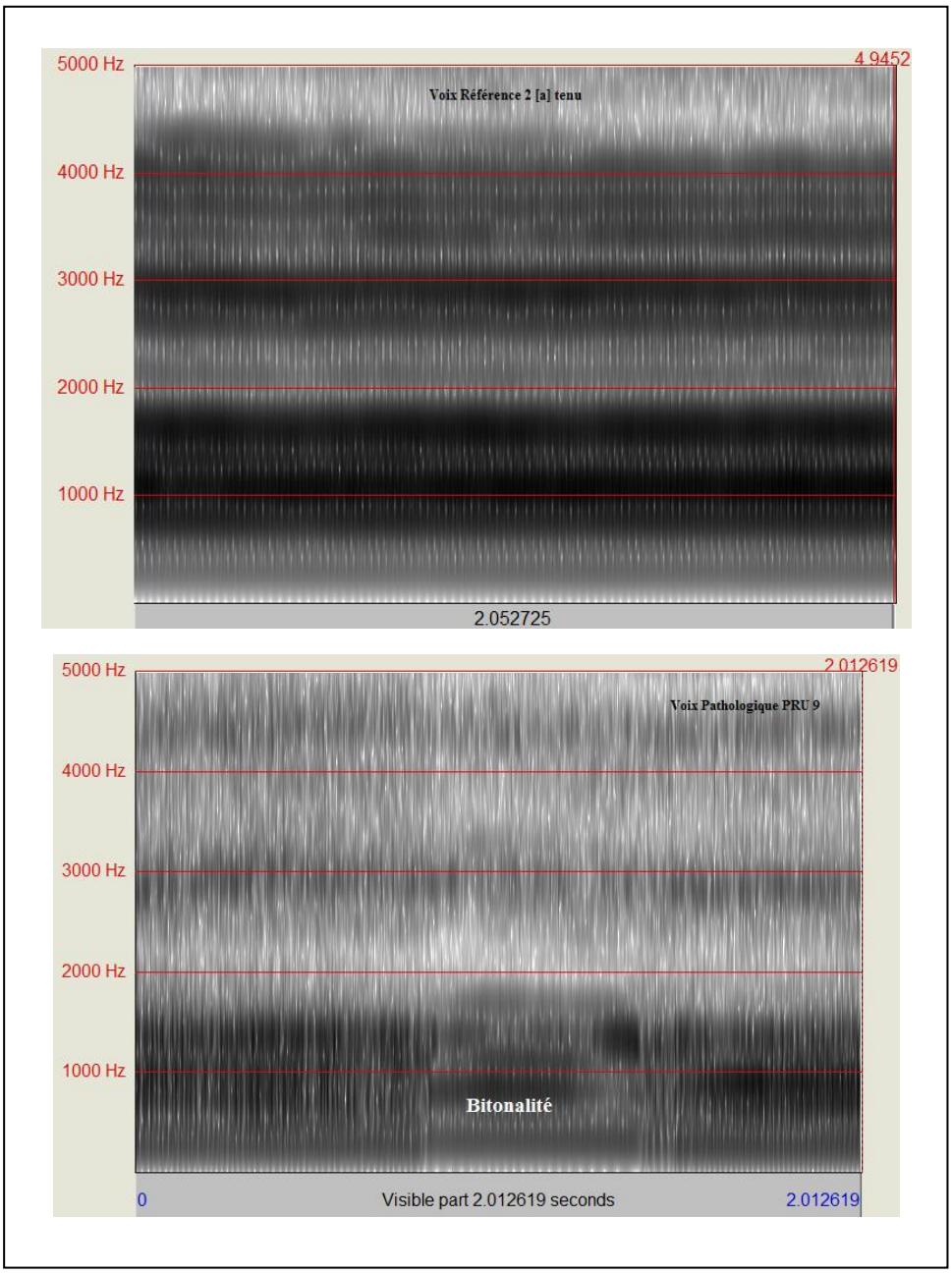


Figure 3.7 : Comparaison des spectrogrammes de Voix Normale (en haut) et Voix d'une PRU 9

Les tableaux 3.2, 3.3 et 3.4 présentent les résultats obtenus après l'analyse acoustique à l'aide du logiciel Praat. Nous avons effectué l'analyse acoustique des voix pathologiques avant, en cours et à la fin de la rééducation. Pour chaque paramètre acoustique, nous avons pris la moyenne des résultats. Nous avons comparé les résultats obtenus avec une norme de référence constituée de trois voix normales.

Tableau 3.2. Resultats de l'Analyse Acoustique des paramétres de la stabilité de F_0

	Paramètres acoustiques	Phase de rééducation			Voix de référence (Valeur moyenne)
		Avant rééducation	En cours	Fin	
Paramètres de la stabilité de F_0	Fo moyenne (Hz)	172.41	203.59	224.89	254.35
	écart type de Fo (Hz)	7.82	4.16	2.33	1.97
	CoV de Fo (%)	4.53	2.04	1.03	0.77
	Jitter absolu (µs)	89.83	25.80	15.27	9.2
	Jitter factor (%)	0.69	0.67	0.27	0.19

Tableau 3.3 : Resultats de l'analyse acoustique des paramétres de la stabilité de l'amplitude de F_0

	Paramètres acoustiques	Phase de rééducation			Voix de référence (Valeur moyenne)
		Avant rééducation	En cours	Fin	
Paramètres de la stabilité de l'amplitude de F_0	Intensité moyenne (dB)	63.05	64.21	70.11	80.60
	Ecart type de l'intensité (dB)	1.79	0.76	1.21	0.55
	CoV de l'amplitude F_0 (%)	2.83	1.18	1.72	0.68
	Shimmer absolu (dB)	0.58	0.25	0.14	0.09
	Shimmer (%) factor	4.15	2.83	4.86	1.42

Tableau 3.4. Resultats de l'analyse du Bruit et l'analyse aérodynamique

	Paramètres acoustiques	Phase de rééducation			Voix de référence (Valeur moyenne)
		Avant rééducation	En cours	Fin	
Analyse du Bruit et Aérodynamique	HNR (dB)	17.57	18.71	24.08	25.48
	HPR (dB)	-34.03	-33.44	-32.79	-26.88
	TMP (s)	5.53	8.10	10.60	12.80
	H_1-H_2 (dB)	5.82	3.34	1.92	1.65
	CCPs (dB)	8.22	11.04	15.24	16.41

L'évolution de la valeur moyenne de F_0 avec son Coéfficient de Variation CoV, au cours de la période de rééducation est présentée dans la figure 3.8. La valeur moyenne de la F_0 de 172 Hz est normale pour une voix féminine. Par contre, son écart type de 7.82 Hz et son CoV de 4.53%, confirmant une importante instabilité de la F_0. Le CoV de la F_0 est un indice important pour explorer la stabilité de la fréquence fondamentale. A la fin de rééducation, nous observons une nette amélioration de la moyenne F_0 (224.89 Hz), son écart type de 2.33 Hz et son CoV est de 1.03 % montre la bonne stabilité de la Fo comparée à la valeur de référence.

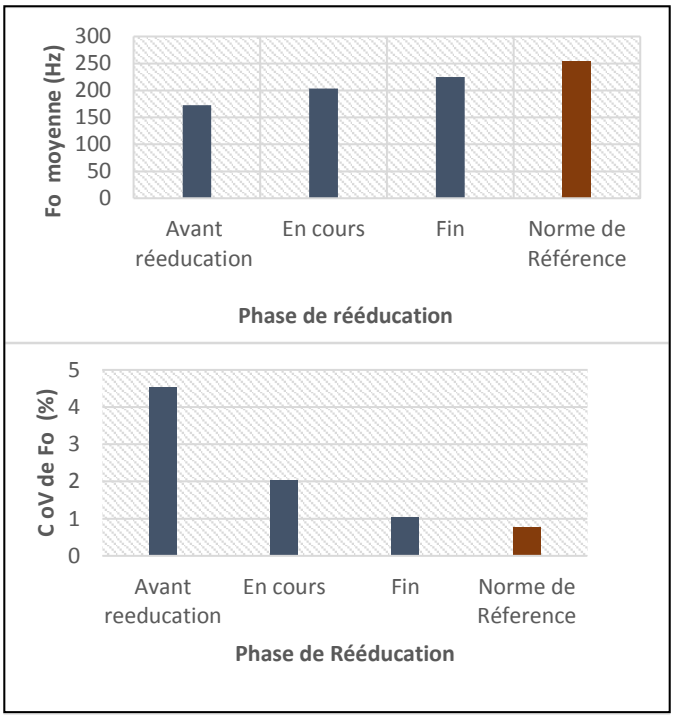


Figure 3.8 : Evolution de F_0 (en haut) et son CoV (en bas) au cours de la Période de Rééducation

La figure 3.9 représente l'évolution de Jitter factor au cours de la phase de rééducation. L'instabilité à Court Terme de la F_0 se traduit par des variations de fréquence entre chaque cycle d'oscillation, et elle est mesurée par le Jitter factor. Avant la rééducation, un jitter factor de 0.69 % pour une F_0 moyenne de 172 Hz est élevé en le comparant à la norme de référence, ce qui témoigne la présence d'une instabilité à court terme du vibrateur laryngien. En revanche, à la fin de la rééducation, un Jitter factor de 0.27 a été mesuré pour une F_0 moyenne de 224.89 Hz, valeur très normale comparée à la norme de référence.

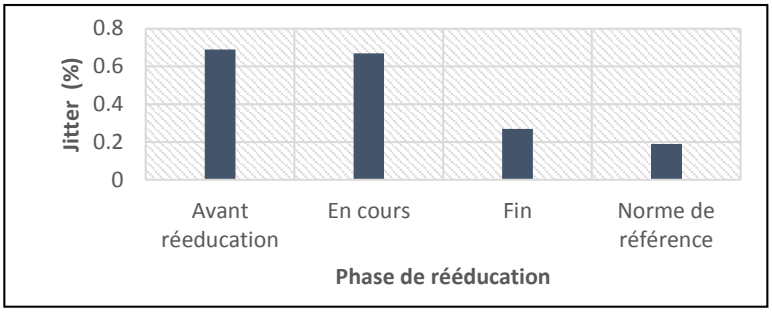


Figure 3.9 : Evolution de Jitter factor au cours de la phase de rééducation

La moyenne de l'intensité de 63 dB avant la rééducation et de 64 dB au cours de la rééducation montre une voix faible. Une intensité moyenne de 70 dB en fin de rééducation, est une valeur moyenne et proche de la norme de référence (Figure 3.10).

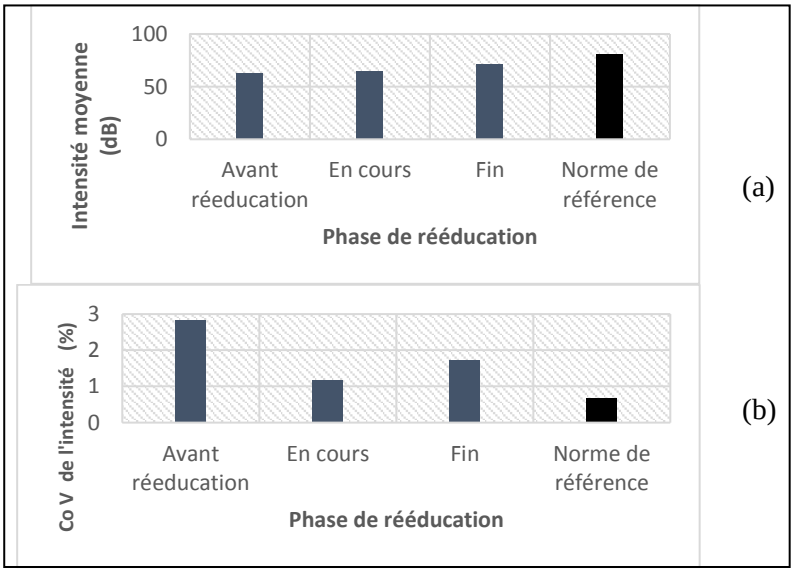


Figure 3.10 : Evolution de l'intensité moyenne
(a) et son CoV (b), au cours de la Période de rééducation

Le CoV de l'intensité (Figure 3.10.b) et le Shimmer (Figure 3.11) restent élevés avec des valeurs variables durant la phase de rééducation. Cette variation est expliquée par l'instabilité à court terme de l'amplitude de la vibration laryngée.

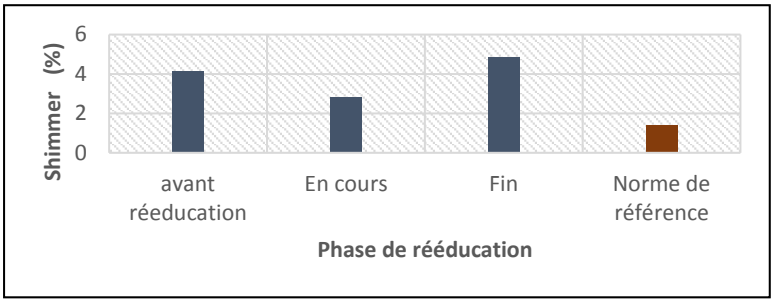


Figure 3.11 : Evolution de Shimmer factor au cours de la Période de rééducation

Le souffle de la voix est considéré comme un bruit se superposant au signal vocal de la source laryngienne. L'analyse spectrale montre un spectre de raies bien défini pour un signal vocal normal de bonne qualité (Figure 3.12.a), un spectre continu massif pour un signal vocal pathologique avant la rééducation (Figure 3.12.b) et une diminution significative du bruit de souffle à la fin de la rééducation (Figure 3.12.c).

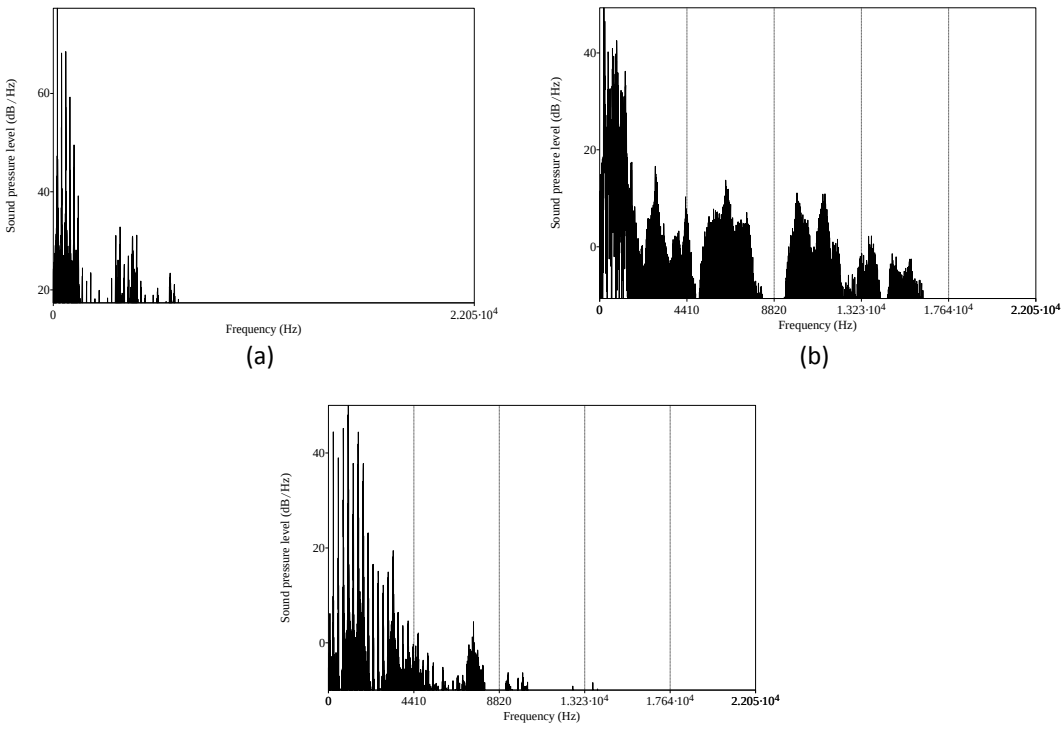


Figure 3.12 : Spectre de la voyelle [a] en décibel en fonction de la fréquence
(a) : voix normale,
(b) : avant la rééducation,
(c) : fin de la rééducation.

Le HNR est le rapport entre l'énergie du spectre Harmonique et celle du spectre de Bruit. Ce bruit peut être un bruit d'écoulement aérodynamique créé par une constriction du conduit vocal ou par un débit d'air trop important. Le HNR n'a relevé aucune modification significative durant la première phase de rééducation (17.57 dB et 18.71 dB). Une nette amélioration est enregistrée en fin de rééducation où le HNR passe de 18.71 dB à 24.08 dB, valeur considérée normale comparée à la norme de référence (Figure 3.13).

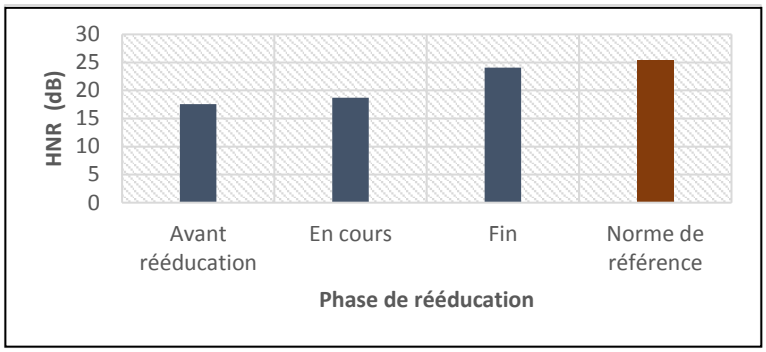


Figure 3.13 : Evolution de HNR au cours de la Période de rééducation

Pour le paramètre TMP, nous avons enregistré une très faible valeur (5.53 secondes) avant la rééducation. Il a évolué durant la phase de rééducation, avec une valeur de 10.6 secondes, valeur moyenne proche de la norme de référence (Figure 3.14).

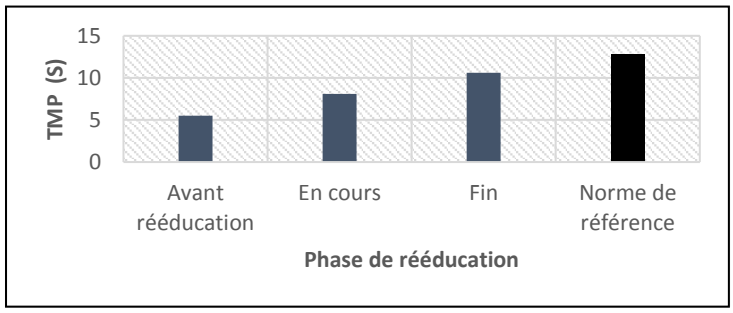


Figure 3.14 : Evolution de TMP au cours de la Période de rééducation

3.8. Résultats de l'analyse pour la LT

Le tableau 3.5 présente les résultats obtenus après l'analyse acoustique pour le cas de la LT. Nous avons présenté pour chaque paramètre de l'analyse acoustique, les valeurs minimale et maximale. Une comparaison de la valeur moyenne de chaque paramètre acoustique a été effectuée avec la moyenne de la norme de référence (sujet sain).

Tableau 3.5 : Résultats de l'analyse acoustique pour la LT

Paramètres Acoustiques	Valeurs			Voix de référence (Valeur moyenne)
	Min	Max	moyenne	
F0 moyenne (Hz)	62.80	110.3	80.4	121.53
F1 (Hz)	685.2	854.3	781.4	564.11
F2 (Hz)	1025.3	1331.2	1265.5	841.48
F3 (Hz)	2757.9	2987.4	2883.2	2326.69
Jitter factor (%)	0.12	3.04	2.07	0.70
Intensité moyenne (dB)	52.11	62.89	56.11	72.12
Shimmer factor (%)	7.14	11.68	9.99	6.61
DUV (%)	0.86	0.97	0.93	0
HNR (dB)	7.27	14.33	11.31	19.14
TMP (s)	2.35	6.65	4.45	10.02

La figure 3.15 présente la variation de F_0 en fonction du temps. Nous remarquons l'instabilité de F_0 pour la voix pathologique, résultat de remplacement des cordes vocales par la Néo-glotte comme source de vibration.

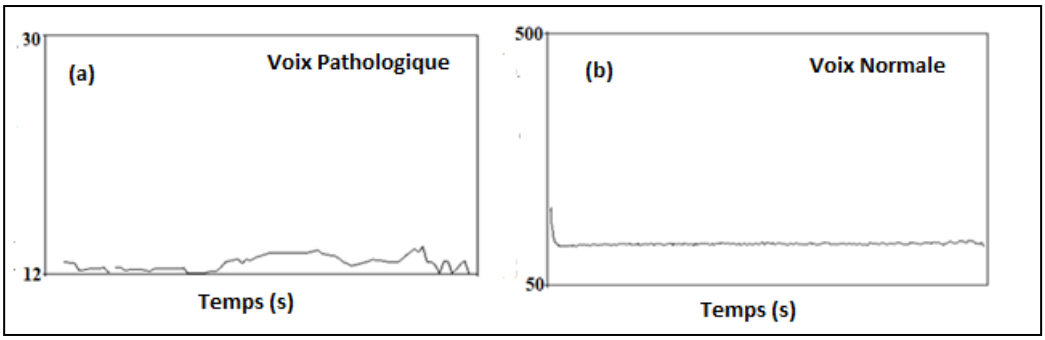


Figure 3.15 : Variation de la Fréquence Fondamentale
(a) : Voix Pathologique,
(b) : Voix Normale

Dans la figure 3.16, une comparaison de F_0 et les formants entre la voix normale et celle pathologique ont été présentés. Après la période de la rééducation, nous remarquons une faible valeur de F_0 comparée à la voix normale, ce qui donne perceptuellement, une voix grave par rapport aux personnes normales.

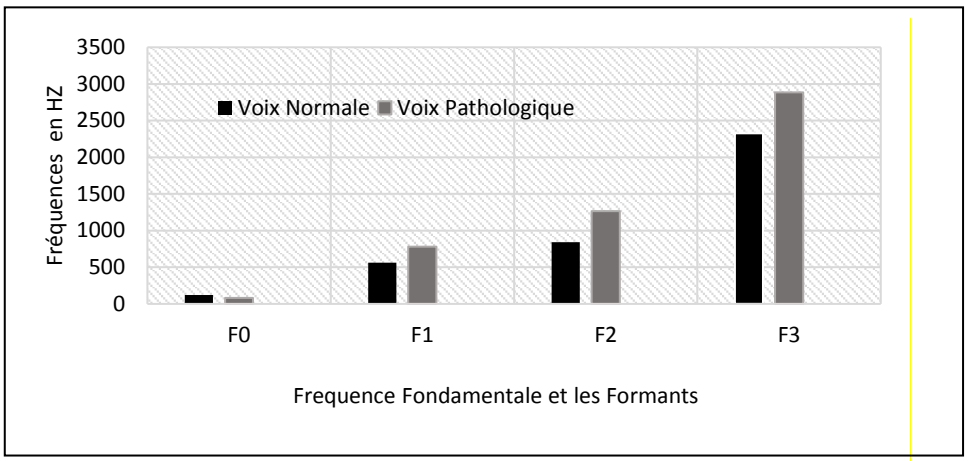


Figure 3.16 : Comparaison de la F_0 et les Formants entre voix pathologique et voix normale

La figure 3.17 illustre les paramètres acoustiques qui présentent l'instabilité de l'amplitude et la fréquence de la F_0 (Jitter, Shimmer et DUV).

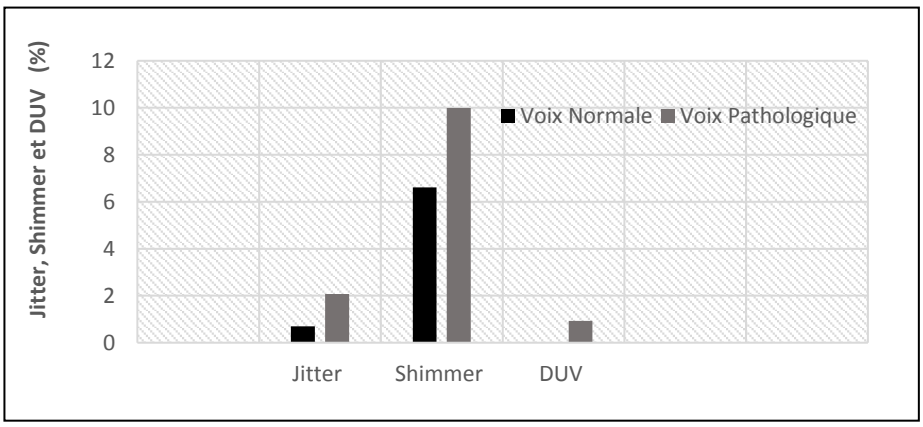


Figure 3.17 : Comparaison des paramètres: Jitter, Shimmer et DUV entre voix pathologique et voix normale

Le souffle de la voix est considéré comme un bruit se superposant au signal vocal (la nouvelle voix). L'analyse spectrale montre un spectre de raies bien défini pour un signal vocal normal de bonne qualité et un spectre continu massif pour un signal de la voix laryngectomisée. Un écart significatif du HNR a été enregistré entre les deux voix

pathologique et normale, ce qui donne une voix soufflée avec un timbre altéré (Figure 3.18).

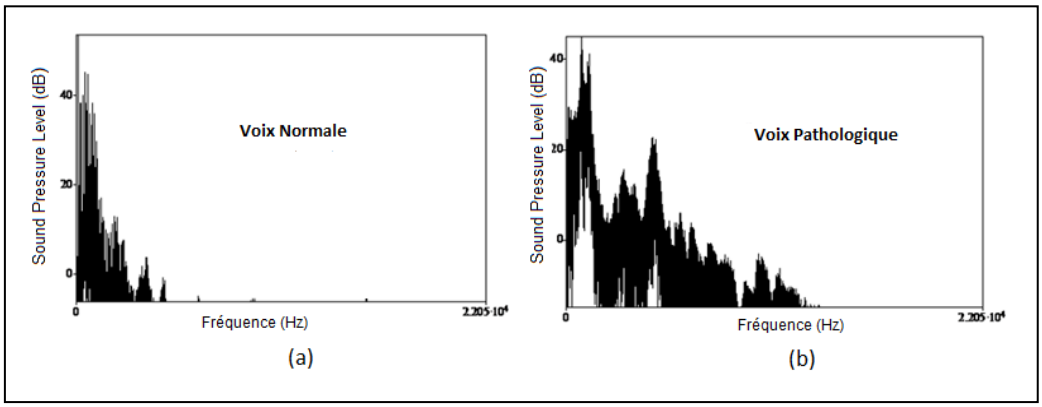


Figure 3.18 : Spectre à bande étroite
(a) : Voix Normale,
(b) : Voix Pathologique

3.9. Discussion

Le but de cette étude est d'évaluer les résultats de la rééducation vocale dans la prise en charge des dysphonies de type Paralysie Laryngée et Laryngectomie Totale avec une méthode objective, basée sur l'analyse acoustique. L'objectif est de préciser l'apport de cette analyse sur la prise en charge thérapeutique de ce type de pathologies.

Pour la PLU et après la période de rééducation, globalement, l'analyse des résultats obtenus montre une stabilité générale de la vibration laryngée et donc l'absence de problèmes de contrôle pneumo-phonique. Ce constat global est confirmé par l'évaluation subjective à l'écoute par l'orthophoniste rééducateur. Néanmoins, l'étude des résultats obtenus par l'analyse acoustique durant la phase de rééducation, nous a permis de faire ressortir quelques constatations importantes.

Nous avons enregistré une amélioration significative de la valeur moyenne de Pitch F_0 avec son CoV. Toutefois, nous avons constaté une très faible amélioration de Jitter dans la première phase de rééducation, ce qui témoigne d'une instabilité à court terme du vibrateur laryngien durant cette période. Cela peut être expliqué par l'adaptation un peu lente des patientes avec la méthode de rééducation adoptée par le service ORL de l'Hôpital de Bab El Oued qui utilise la méthode de François Le Huche comme technique de rééducation [18].

La comparaison des résultats obtenus pour le HNR, HPR, H_1-H_2, CPP et TMP par rapport à la norme de référence dans la première phase de rééducation démontre une importante fuite glottique provoquée par un mauvais accolement des cordes vocales. Ces paramètres ont été améliorés en fin de rééducation. En revanche, nous avons enregistré une augmentation du CoV et le Shimmer à la fin de la rééducation, ce qui pose des questions sur cette hausse de valeurs de ces deux paramètres. Ces résultats montrent toujours la présence d'une instabilité à court terme de l'amplitude de la vibration des cordes vocales en fin de rééducation, malgré l'amélioration de l'intensité. Cela peut être expliqué par une rééducation basée sur un travail de respiration plus que le travail de vocalisation ou de tenue de voyelles, nécessaire pour la vibration stable des cordes vocales [13].

Globalement, pour la PRU, nous avons constaté une forte corrélation entre l'évaluation objective basée sur les paramètres acoustiques et l'évaluation subjective. Néanmoins, nous avons souligné des valeurs insuffisantes pour certains paramètres acoustiques cités précédemment, ce qui a été confirmée par l'orthophoniste rééducateur, qui affirme que la majorité des patientes refusent d'appliquer tout le protocole de rééducation comme les exercices de chant, ce qui entraîne une faible stabilité de la vibration laryngienne.

Pour la LT et après la période de la rééducation, nous avons remarqué des valeurs de F_0 faibles et instables par rapport à la voix normale. La valeur de F_0 et le pourcentage de voisement DUV montrent que le voisement acquis grâce à la Néo-glotte (ou pseudo-glotte) est perceptible mais reste encore loin du voisement normal des cordes vocales. Les résultats obtenus pour ces deux paramètres peuvent être expliqués par la forme et l'élasticité de la Néo-glotte qui différent totalement de ceux des cordes vocales.

Les valeurs relevées pour le Jitter, nous montrent une évolution faible de ce paramètre vers la référence. Cela est expliqué par le changement du vibrateur laryngé naturel par le segment pharyngo-œsophagien utilisé comme Néo-glotte, le comportement vibratoire est devenu donc, très différent de la phonation laryngée. Le segment pharyngo-œsophagien semble vibrer avec plus d'apériodicité, ce qui peut être attribué à un mauvais contrôle volontaire sur la Néo-glotte.

La valeur de 56 dB de l'intensité moyenne est considérée comme valeur acceptable car, généralement le niveau sonore d'une voix œsophagienne est compris entre 55 et 65 dB [1], intervalle qui est inférieur à celui de la voix normale car le Néo vibrateur est alimenté par l'air œsophagien, donc par une soufflerie très limitée.

Le Shimmer de la voix normale est de 6.61 %, la moyenne de nos patients est de 9.99 % avec des extrêmes de 7.14 à 11.68 % qui parait supérieure à la valeur normale. Cependant, cette valeur est acceptable vu que le signal de la voix œsophagienne est une réalisation aléatoire, donc l'hypothèse de stationnarité garante de la fiabilité des résultats n'est pas remplie.

Le rapport harmoniques/bruit HNR a pour fonction d'évaluer l'émergence des harmoniques d'un signal par rapport au bruit. L'analyse temporelle des échantillons de la voix laryngectomisée a mis en évidence l'existence d'un bruit précédant le début de la prononciation, dû à l'effort intense de respiration du malade pour produire la parole œsophagienne.

Pour le TMP, nous constatons un temps inférieur à 5 secondes, ce qui est très court par rapport à la référence. Plusieurs facteurs peuvent influencer la valeur du TMP. Logiquement, il risque de diminuer si l'intensité vocale est élevée, car le maintien de cette intensité demande une pression sous-glottique importante. En plus, la voix laryngectomisée utilise l'air œsophagien, qui est une soufflerie très limitée.

Malgré que la structure formantique n'est pas touchée par la laryngectomie, nous avons relevé une forte augmentation des valeurs des formants après ablation du larynx. Ceci peut être expliqué par le fait que la distance entre la Néo-glotte et la première cavité est modifiée (raccourcissement du conduit vocal).

Globalement, on peut résumer que la LT affecte considérablement les mécanismes de phonation normale en supprimant le vibrateur naturel qui est la glotte et la source d'énergie qui est la soufflerie pulmonaire. Il en résulte donc, une voix chuchotée, rauque souvent difficilement perceptible.

D'après les résultats obtenus, la voix œsophagienne est caractérisée par une réduction significative du pitch, de l'intensité, du débit de parole, des variations de la hauteur avec un rapport signal sur bruit faible aboutissant perceptuellement à une voix rauque.

3.10. Conclusion

Une évaluation objective basée sur l'analyse acoustique a été réalisée sur deux catégories de pathologies vocales : la Paralysie Récurrentielle Unilatérale et la Laryngectomie Totale. Nous avons constaté pour la PRU, une forte corrélation entre les résultats pathologiques et ceux de la norme de référence. Néanmoins, nous avons enregistré des résultats insuffisants pour la LT, comparés à la voix normale de référence. Cela peut être expliqué par le choix de la méthode de l'évaluation objective qui est basée sur les paramètres de l'instabilité de la vibration de la pseudo-glotte.

Chapitre 4

Evaluation Objective de la Voix Pathologique par les Réseaux de Neurones Récurrents

4.1. Introduction

Les nouvelles technologies dans l'apprentissage et les méthodes de classification ont joué un rôle important dans l'évolution et le développement des méthodes et moyens de diagnostic des troubles vocaux. Elles permettent une évaluation objective de la qualité des voix pathologiques. Les systèmes de détection ou d'évaluation du niveau de sévérité des voix pathologiques s'inspirent largement du domaine du Traitement Automatique de la Parole (TAP) et des approches de classification [71].

Parmi les nombreux modèles proposés pour résoudre le problème de la détection et la classification des voix pathologiques, nous trouvons les modèles connexionnistes, ou *Réseaux Neuronaux*. Ils ont été utilisés depuis longtemps dans des problèmes difficiles de classification et de reconnaissance des formes que l'on rencontre précisément en détection et classification de voix pathologiques.

Dans ce chapitre nous allons développer un système de détection et d'évaluation automatique des voix pathologiques en utilisant les Réseaux de Neurones Récurrents. Nous utiliserons dans ce système, une analyse acoustique discriminante basée sur des paramètres pathologiques utilisés précédemment. L'objectif est de montrer que l'utilisation des RN dans le processus d'évaluation de la rééducation avec l'introduction des indices pathologiques reflétant le dysfonctionnement des cordes vocales, dans la phase d'extraction des vecteurs acoustiques, peut améliorer et faciliter considérablement l'évaluation de la voix au cours de la rééducation.

4.2. Méthodes de classification des voix pathologiques

Les méthodes de classification sont très utilisées en TAP, elles permettent de grouper des objets (observations ou individus) dans des classes de manière à ce que les objets appartenant à la même classe sont plus similaires entre eux qu'aux objets appartenant aux autres classes. Un système d'évaluation automatique peut discriminer entre les échantillons normaux et pathologiques et de classer les pathologies de la voix. Le processus de différenciation entre les sujets normaux et pathologiques est un problème à deux classes appelé *détection* de pathologie. En revanche, la discrimination entre les différents types de pathologies est un problème multi-classes, appelé *classification* des pathologies.

Parmi les méthodes utilisées dans la détection et la classification automatique de la voix pathologique, nous citons les modèles connexionnistes ou les **R**éseaux de **N**eurones **A**rtificiels (RNA), modèles à base de mixtures de gaussiennes adaptées à partir d'un modèle générique de parole GMM (**G**aussian **M**ixture **M**odels), le classificateur de k plus proche voisin (**K-N**earest **N**eighbors : KNN), les machines à support de vecteurs (**S**upport **V**ector **M**achines : SVM), etc.

Le tableau 4.1 résume des travaux de recherches sur des systèmes de détection et de classification de parole et voix pathologiques, nous avons précisé leurs méthodologies sur l'analyse acoustique utilisée, le type de classifieur ainsi que le corpus choisi.

Tableau 4.1 : Modèles de classifieurs utilisés en détection et classification de paroles et voix pathologiques

Référence	Paramètres acoustiques	modèle de classificateur	Corpus utilisé	Précision (%)
[72]	MFCC	GMM	voyelles [a], [i] et [u]	99
[73]	MFCC	SVM	Voyelle [a] de 1.5 sec	93
[74]	MFCC, Jitter, Shimmer et HNR	SVM	voyelles [a], [i] et [u]	71
[75]	MFCC et signaux EGG	Combinaison SVM et GMM	Voyelles tenues [a], [i] [u], parole continue	96.96
[76]	MFCC	Modèle hybride SVM /GMM	Voyelles tenues	96.1
[77]	HNR, bande critique du spectre énergétique	KNN	Voyelle [a] tenue	93.4
[78]	Prétraitement du signal Vocal, signaux EGG	DNN	Voyelles tenues [a], [i], [u] et Parole continue	71.36

4.3. Elaboration du système d'évaluation par les RN

La classification de la voix, que ce soit pour identifier un mot, un locuteur, une particularité (langue, accent, genre, émotion, etc.) ou une pathologie vocale, se décompose en plusieurs phases.

La première phase est celle de l'extraction de paramètres vocaux à partir du signal brut. Après cette étape s'ensuit la phase de construction d'un modèle de classement automatique, faisant généralement appel à l'apprentissage automatique supervisé. Cette

phase commence par l'entrainement du modèle de classification à partir des vecteurs caractéristiques labellisés (dont la classe est précisée) extraits de données servant à l'apprentissage. Après l'étape d'apprentissage du modèle, a lieu une étape de test pendant laquelle de nouvelles données sont classées.

4.3.1. Choix du modèle de classification

Ces dernières années, les Réseaux de Neurones Artificiels occupent une place importante pour résoudre le problème de la Reconnaissance Automatique de la Parole (RAP) et notamment dans la détection, classification et l'évaluation de la voix pathologique.

Une technique d'identification des voix pathologiques par les réseaux de neurones multicouches MLP (**M**ulti **L**ayer **P**erceptron) est présentée. Les auteurs ont utilisé le Pitch, les trois premiers formants, la transformée en ondelettes et l'énergie pour extraire les vecteurs acoustiques. Un taux de classification varie entre 80 % et 100 % a été réalisé en fonction de type d'analyse acoustique [79].

Dans le but de tester l'efficacité d'extraire les paramètres des signaux d'impédance (**E**lectro-**G**lotto-**G**raph : EGG) dans les domaines temporel et fréquentiel à court et à long terme, Ritchings et al. ont développé un système d'évaluation de la qualité de la voix pathologique par les RN de type MLP, avec une précision de 92 % [80].

Dans [74] et [81], Teixeira et al. ont proposé un système de détection vocale en gardant les mêmes caractéristiques dans ses deux publications mais en changeant les classificateurs. Dans [74], ils ont utilisé SVM avec Jitter, Shimmer et HNR et la précision était de 71 %. Dans [81], ils ont utilisé MLP avec Jitter, Shimmer et HNR et le rapport de précision était 100 % et 90 % pour les voix féminines et masculines respectivement.

Des nouveaux modèles connexionnistes couramment utilisés dans la reconnaissance d'image et vidéo ont été également appliqués dans la détection et la classification des pathologies vocales, les auteurs dans [82] ont comparé deux classificateurs connexionnistes : les Réseaux de Neurones Convolutifs CNN (**C**onvolutional **N**eural **N**etwork) et Les Réseaux de Neurones Récurrents RNN (**R**ecurrent **N**eural **N**etwork), pour une base de données dépassant 2000 sujets entre

normaux et pathologiques avec 72 pathologies vocales. Un taux de détection de 87.11 % a été enregistré pour CNN contre 86.52 % pour RNN.

Dans [83], une méthode de détection et de classification des voix pathologiques a été proposée en utilisant trois modèles connexionnistes : les perceptrons multicouches MLP, les réseaux de neurones de régression généralisée GRNN (**G**eneral **R**egression **N**eural **N**etwork) et les réseaux de neurones probabilistes PNN (**P**robabilistic **N**eural **N**etwork). Les résultats obtenus montrent que le modèle neuronal (PNN) se comporte de la même manière que celui de GRNN. Les auteurs ont constaté que le MLP avec les paramètres MFCC fonctionne mieux que PNN et GRNN dans la classification des voix pathologiques.

Dans notre travail, nous avons choisi d'utiliser un système de détection de la voix pathologique, basé sur un réseau RNN-LSTM. Ces derniers ont la capacité de traiter des séquences de taille variable qu'on trouve principalement dans la détection et la classification de voix pathologiques. L'avantage de ces Réseaux est que les valeurs d'entrée transmises au réseau passent non seulement par plusieurs couches LSTM, mais se propagent également dans le temps dans une cellule LSTM afin d'éviter les problèmes liés à une dépendance à long terme [84].

4.3.2. Base de données

La base de données utilisée dans ce travail est constituée des enregistrements sonores de la voyelle [a] tenue pour les différents paramètres pathologiques, répartie en deux fichiers. Le premier de taille de 756 échantillons de voix normales et pathologiques est utilisé pour l'apprentissage et la validation du processus de l'entrainement du système, 80 % de données de ce fichier sont utilisées pour l'apprentissage et 20 % pour la validation. Le deuxième fichier pour la détection, est formé de 384 échantillons entre normaux et pathologiques.

4.3.3. Paramétrisation du signal vocal

Dans la RAP, il est nécessaire d'effectuer une analyse des données de manière à déterminer les caractéristiques discriminantes. Ces caractéristiques constituent l'entrée du réseau de neurones. Cette étape a des conséquences à la fois sur la taille du réseau (et donc le temps de simulation), sur les performances du système (pouvoir de

séparation, taux de détection) et sur le temps de développement (temps d'apprentissage).

La représentation paramétrique du signal vocal doit être précédé d'abord par une phase de mise en forme du signal, appelée aussi *Prétraitement*. Pour cela, quelques étapes sont effectuées avant tout traitement (Figure 4.1).

4.3.3.1. Prétraitement du signal vocal

Dans cette étape, le signal analogique capté par le microphone est transformé en composantes numériques. L'information acoustique pertinente du signal de parole se situe principalement dans la bande passante [50 Hz - 5 kHz], la fréquence d'échantillonnage *Fe* devrait au moins être égale à 10 kHz, selon le théorème de *Shannon*, mais elle peut varier en fonction du domaine d'application. Dans le cadre de l'évaluation de la qualité de la voix, le signal de parole est généralement échantillonné à une fréquence de 44.1 kHz avec une résolution de 16 bits.

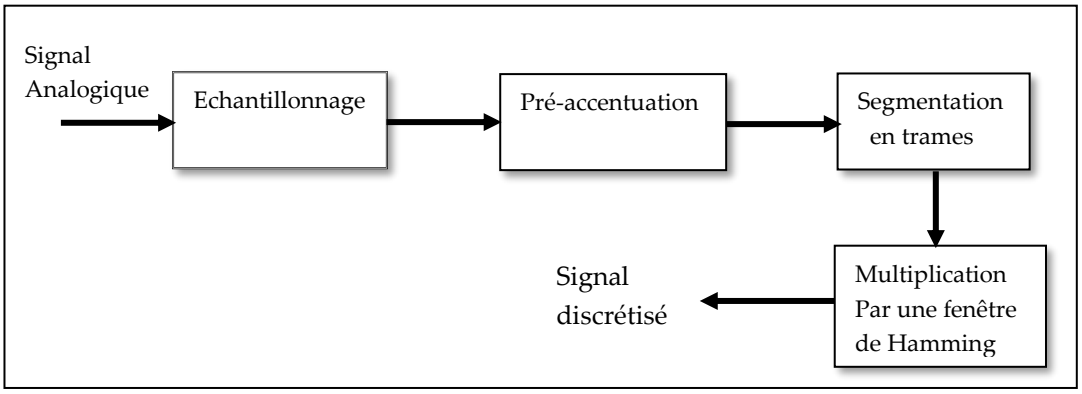


Figure 4.1: Prétraitement du signal vocal.

Après la numérisation du signal vocal, une pré-accentuation est effectuée afin d'accentuer les hautes fréquences qui sont moins énergétiques que les basses fréquences. On utilise généralement un filtre dit de pré-accentuation de transmittance :

$$H(z) = 1 - 0.95Z^{-1} \qquad (4.1)$$

Le signal vocal résultant qui est fortement non stationnaire est décomposé ensuite en une succession de tranches élémentaires supposées stationnaires appelées *fenêtres d'analyse* ou *trames*. Le découpage en trames est appliqué toutes les 10 ms sur des fenêtres de 25 ms (par glissement et recouvrement des fenêtres d'analyse). Ce découpage en trames produit des discontinuités aux frontières des trames, qui se

manifestent par des lobes secondaires dans le spectre. Ce phénomène s'appelle *Effets de bord*. Pour compenser ces effets, nous appliquons une fenêtre de Hamming à chacune de ces tranches.

$$H(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (4.2)$$

Avec : *N longueur ou taille de la fenêtre et* $0 \leq n \leq N-1$

4.3.3.2. Extraction multi variables des paramètres acoustiques

L'étape de prétraitement du signal d'entrée est suivie par une analyse acoustique multi-variables. Afin d'avoir une discrimination optimale de notre système, nous avons pris en compte les points suivants dans l'extraction des paramètres acoustiques (Tab.4.2) :

- La stabilité de la fréquence et de l'amplitude de la vibration laryngée F_0 de la voix, par les paramètres pathologiques : Jitter, Shimmer;
- L'analyse du bruit qui permet d'évaluer le souffle de la voix, notamment les fuites glottiques dans le cas de la PRU. Nous avons choisi les paramètres : HNR, HPR, H_1-H_2 et le CPP.
- Les Coefficients Cepstraux de fréquence à l'échelle de Mel (MFCC).

Tableau 4.2 : Paramètres acoustiques utilisés pour chaque pathologie

	Paramètres acoustiques		Nombre de Coefficients
	PRU	LT	
Paramètre Cepstrale	MFCC	MFCC	12
Paramètres de la Stabilité de F_0	Jitter	Jitter	1
	Shimmer	Shimmer	1
Paramètres du Bruit	HNR	HNR	1
	HPR	-	1
	H_1-H_2	-	1
	CPP	-	1

Les Coefficients MFCC sont les paramètres les plus répandus, dans le traitement de la parole, utilisés pour la reconnaissance automatique de la voix, du locuteur, ainsi que la détection et la classification des pathologies de la voix [72,73,74,75]. Le principe de

calcul des MFCC est issu des recherches psycho-acoustiques sur la perception des différentes bandes de fréquences par l'oreille humaine. La sensibilité de l'oreille diminue avec l'accroissement des fréquences. Par analogie on utilise généralement des bancs de filtres dont la répartition reproduit cette sensibilité (échelle de Mel). L'intérêt principal de ces coefficients est d'extraire des informations pertinentes en nombre limité en s'appuyant à la fois sur l'hypothèse de la production de la parole où le signal vocal est le résultat de la convolution entre une excitation (cordes vocales) et un filtre (conduit vocal), et sur la perception de la parole (échelle des Mels) (figure 4.2). L'extraction des MFCC se décompose en trois étapes principales :

- après l'étape de prétraitement du signal vocal citée précédemment, un passage dans le domaine spectral est effectué par le calcul de la Transformée de Fourier Discrète DFT (Discret Fourier Transform) ;
- un passage à l'échelle de Mel. La conversion de *hertz* en *mels* se fait par la formule de transformation suivante :

$$mel = 2595 \log_{10}(1 + \frac{f}{700}) \quad (4.3)$$

- enfin, nous convertissons le spectre logarithmique de Mel en domaine temporel par l'application de l'inverse de la Transformée en Cosinus Discrète IDCT (Inverse Discret Cosine Transform). Pour chaque échantillon de voix, nous obtenons 12 coefficients MFCC pour chaque trame.

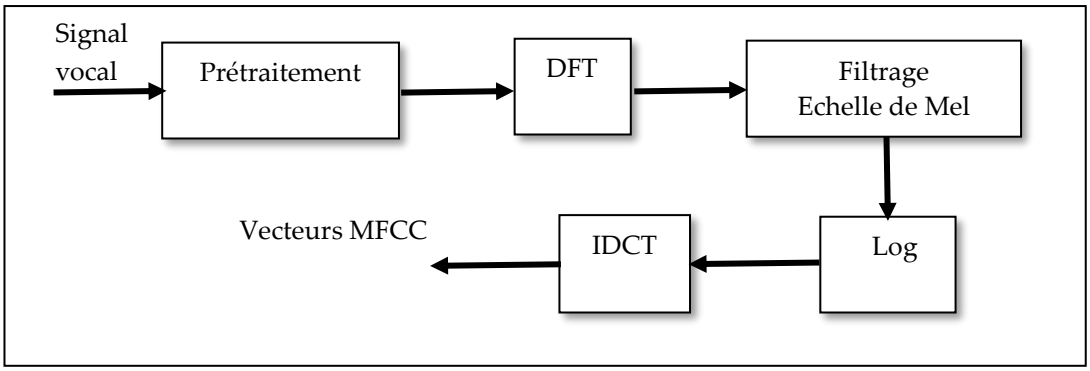


Figure 4.2 : Calcul des Coefficients MFCC

L'organigramme ci-dessous résume les différentes phases permettant de réaliser notre système, il fonctionne selon deux principales étapes (Figure 4.3) : La première est

celle de l'extraction de paramètres acoustiques à partir du corpus utilisé, suivie par une étape de classification (décision) pendant laquelle de nouvelles données sont classées.

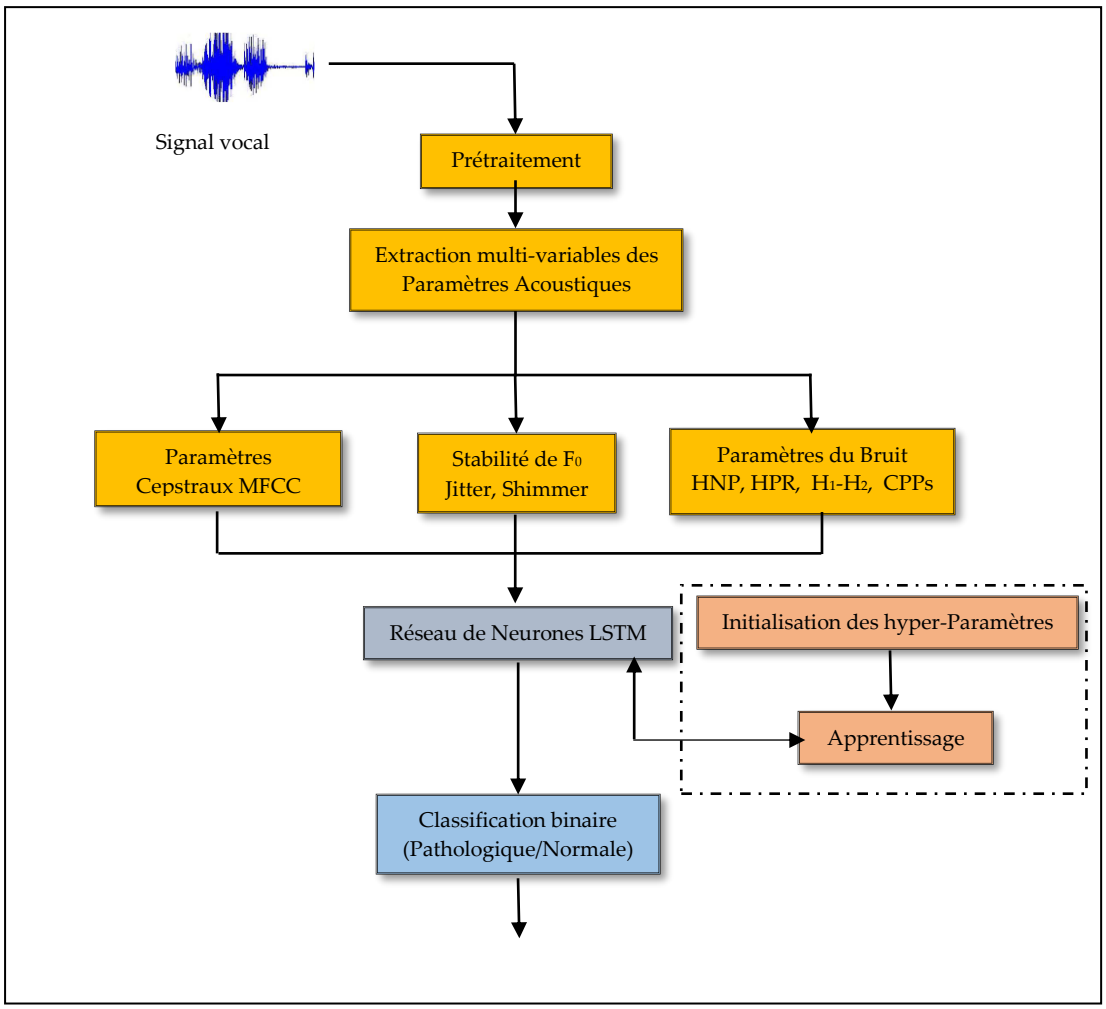


Figure 4.3 : Organigramme du système de détection automatique de voix pathologiques

4.3.4. Architecture du système

L'architecture du système de détection de la voix pathologique que nous avons choisi dans ce travail est constituée de 6 couches: une couche d'entrée, une couche de sortie et 4 couches cachées :

- le nombre *d'entrées* de la première *couche* est fixé à 18 pour la PRU et 15 pour la LT. Les Données d'entrées correspondent aux coefficients de l'analyse acoustique pour chaque trame (fenêtre), représentées par une matrice, dont le nombre de lignes est fixe (18 ou 15) et celui de colonnes, selon la taille de l'échantillon (nombre de trames) ;

- la *première couche cachée* est limitée à 100 neurones. Elle correspond à la couche des cellules LSTM [85, 86] ;
- une autre couche de neurones cachée appelée *Dropout* pour éviter le sur-apprentissage ou apprentissage par cœur (*Overfitting*). Le Dropout est une technique où des neurones sélectionnés au hasard sont ignorés (temporairement) pendant l'apprentissage. Cela signifie que leur contribution à l'activation des neurones qui leur succède est temporairement supprimée lors de la phase de propagation et toutes les mises à jour de poids ne sont pas appliquées au neurone lors de la phase de retro-propagation. Lors de la phase d'apprentissage, pour chaque itération, un neurone est gardé avec une probabilité p, sinon il est supprimé [85] ;
- la *troisième couche cachée* a deux neurones complètement connectés *(Fully Connected)*. Elle représente le nombre de classes pour assurer la non-linéarité des activités du réseau, permettant aux modèles d'apprendre plus rapidement et plus efficacement [87];
- la *quatrième couche cachée* est une fonction d'activation *Softmax* à deux neurones pour normaliser les probabilités ainsi calculées dans la couche précédente [78] ;
- la *couche de sortie* contient un seul neurone pour la décision (Pathologique P ou Normale N) [82, 83] (Figure 4.4).

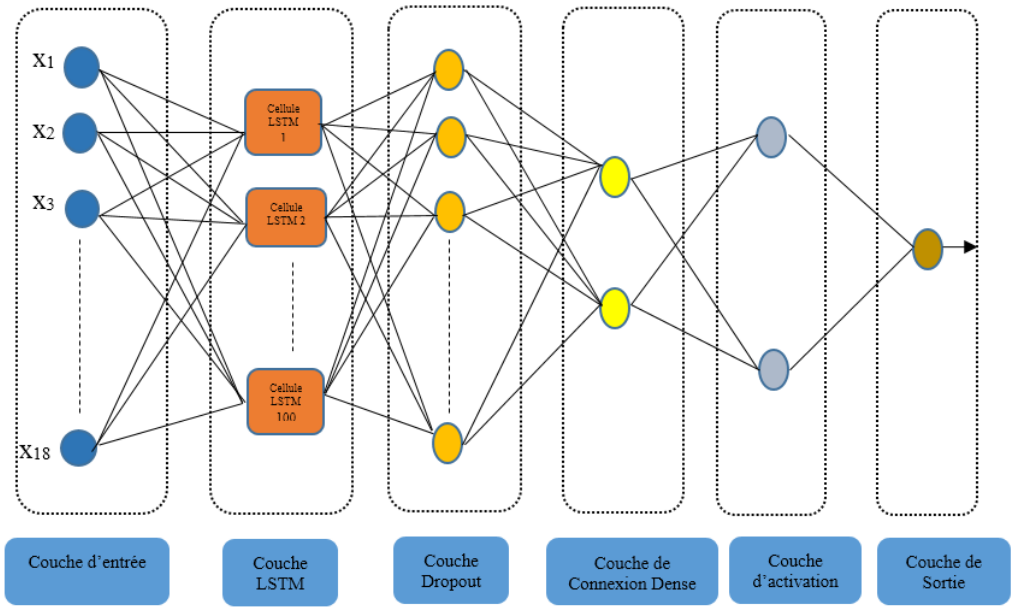


Figure 4.4 : Architecture du système de détection de la voix pathologique Cas de PRU

Le tableau 4.3 résume les étapes avec leurs caractéristiques de conception de notre système de détection. Nous avons commencé par l'élaboration de la banque de données pour la phase d'apprentissage et de détection, suivie par une analyse acoustique de ces données, et en dernière étape la détermination de l'architecture du réseau neuronal.

Tableau 4.3 : Caractéristiques du système de Détection

Taille du Corpus d'apprentissage et de détection	Apprentissage Validation Détection	600 156 384 (194 PRU et 190 LT)
Phase d'analyse acoustique	Fréquence d'échantillonnage	44.1 KHz
	Filtre de préaccentuation	$H(n) = 1 - 0.95Z^{-1}$
	Durée de segmentation en trames	25 ms avec 10 ms de recouvrement
	Nombre de trames par séquence	8 à 22
	Nombre de coefficients MFCC	12
	Dimension du vecteur acoustique d'entrée	15 coefficients pour LT et 18 PRU
Architecture du Réseau	Couche d'entrée	15 entrées pour LT et 18 PRU
	Couche cachée 1	LSTM avec 100 neurones
	Couche cachée 2	Dropout avec 100 neurones
	Couche cachée 3	Fully Connected à 2 neurones
	Couche cachée 4	Fonction Softmax à 2 neurones
	Couche de sortie	01 neurone pour décision P ou N

4.3.5. Fonctions d'activation utilisées

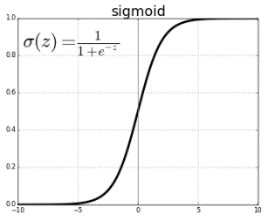
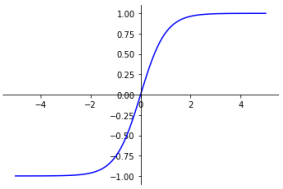
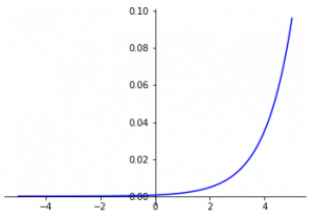
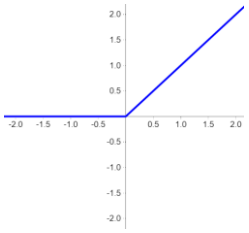
La fonction d'activation sert à modifier de manière *non-linéaire* les données. Cette non-linéarité permet de modifier spatialement leur représentation. Après que le neurone a effectué le produit entre ses entrées et ses poids, il calcule la somme pondérée des entrées et ajoute le biais ensuite, il applique également une non-linéarité sur ce résultat pour décider si le neurone est activé ou non. Cette fonction non linéaire s'appelle la *fonction d'activation*.

Dans les réseaux de neurones profonds, les fonctions d'activations sont des fonctions non-linéaires puisque l'application récurrente d'une même fonction linéaire n'aurait aucun effet. Ceci permet de séparer les données non-linéairement séparables et donc d'effectuer des classifications plus poussées qu'en apprentissage automatique classique. Autre point très important, les données traitées par les neurones peuvent

atteindre des valeurs très grandes et rendant les calculs beaucoup plus complexes. Afin d'y remédier, les fonctions d'activation non linéaires réduisent la valeur de sortie d'un neurone le plus souvent sous forme d'une simple probabilité [88].

Dans ce travail, nous avons utilisé 4 fonctions d'activation : Sigmoïde logistique (Sigmoid) et Tangente hyperbolique (Tanh) pour la couche cachée LSTM, la fonction Relu (**R**ectified **L**inear **U**nit) est utilisée pour la couche Fully Connected et la fonction Softmax pour la dernière couche cachée [78, 89, 90] (tableau 4.4).

Tableau 4.4 : Caractéristiques mathématiques des fonctions d'activation

Fonction d'activation	Equation mathématique	représentation
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$ ou $x \epsilon R$	
Tanh	$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ ou $x \epsilon R$	
Softmax	$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x^j}}$ ou $x \epsilon R$	
Relu	$f(x) = \begin{cases} 0 \; si \; x < 0 \\ x \; xi \; x \geq 0 \end{cases}$ ou $x \epsilon R$	

La fonction sigmoïde est utilisée dans la couche des cellules LSTM par les trois portes : d'entrée, de sortie et d'oubli. Elle génère une valeur comprise entre 0 et 1, elle peut soit ne laisser passer aucun signal, soit compléter le flux d'information *via* les portes.

La fonction tangente hyperbolique ou *tanh* ressemble à la fonction sigmoïde mais son résultat peut être compris entre −1 et 1. Nous avons utilisé la *tanh* dans la couche de neurones LSTM car elle permet de limiter les grandes valeurs négatives et positives entre −1 et 1.

La fonction *ReLU* résout le problème du risque de disparition du gradient (*vanishing gradient problem*). C'est la non-linéarité utilisée par défaut des réseaux de neurones multicouches.

La fonction *Softmax* est une non-linéarité utilisée pour la dernière couche d'un réseau neuronal profond créé pour une tâche de classification à *k* classes car elle permet de sortir *k* probabilités de prédiction dont la somme totale est égale à 1.

4.3.6. Rétro-propagation, Fonction du Coût et Algorithme d'Optimisation

Le principe de Rétro-propagation de l'erreur, consiste à présenter au réseau un vecteur d'entrée, de procéder au calcul de la sortie par la propagation à travers les couches, de la couche d'entrée vers celle de la sortie en passant par les couches cachées. Cette sortie obtenue est comparée à la sortie désirée, une erreur est alors obtenue. Ensuite on calcule le gradient de l'erreur qui se propage de la couche de sortie vers la couche d'entrée, d'où le terme de rétro-propagation. Cela permet la modification des poids du réseau.

L'objectif de la méthode de rétro-propagation est d'adapter les poids synaptiques *w* de façon à minimiser une fonction dite de *Coût*. Dans le cadre de notre travail, la fonction du coût utilisée est l'*Entropie Cro*isée. Pour un classifieur de *n* classes, cette fonction est donnée par l'équation suivante [91] :

$$L\,(\,\hat{y}, y) = -\sum_i^n y_i \, log\hat{y_i} \qquad (4.4)$$

Avec $y_i$et $\hat{y_i}$: les sorties obtenue et désirée respectivement.

Pour minimiser cette fonction du coût et mettre à jour les poids de réseau itératifs en fonction des données d'apprentissage, nous avons choisi l'algorithme d'optimisation *Adam* (Adaptative Momentum estimation), largement adopté pour les applications d'apprentissage en profondeur [92].

Adam calcul la moyenne exponentielle et les carrés du gradient pour chaque paramètre. Le taux d'apprentissage est ensuite multiplié par la moyenne du gradient et en le divisant par la racine carrée de la moyenne de l'exponentielle des gradients. Ensuite une mise à jour est ajoutée [92].

$$v_t = \beta_1 \cdot v_{t-1} + (1 - \beta_1).g_t \cdot \quad (4.5)$$

$$s_t = \beta_2 \cdot s_{t-1} + (1 - \beta_2) \cdot g_t^{\,2} \quad (4.6)$$

$$\Delta w_t = -\varphi \frac{v_t}{\sqrt{s_t + \epsilon}}\, g_t \quad (4.7)$$

$$w_{t+1} = w_t + \Delta w_t \quad (4.8)$$

Avec:

φ : *Taux d'apprentissage initial.*
g_t : *Le gradient à l'instant t.*
v_t : *La moyenne exponentielle des gradients.*
s_t : *La moyenne exponentielle des carrées des gradients*
β_1, β_2 *et* ϵ *: Paramètres initialisés généralement à 0.9, 0.99 et* 10^{-8} *respectivement.*

4.3.7. Initialisation des paramètres de l'apprentissage

Définir les paramètres d'apprentissage (appelés aussi hyper-paramètres) est une étape importante et difficile dans la conception d'un système de classification à base de réseau de neurones profond. Elle influe sur la durée d'apprentissage, la convergence du réseau et les performances du système.

Nous commençons par le choix des poids synaptiques initiaux des couches cachées du réseau. L'initialisation avec les poids appropriés peut faire la différence entre un réseau convergent dans un délai raisonnable et un réseau non convergent, malgré un nombre d'itérations qui implique plusieurs jours. Plus les poids initiaux sont proches de leur valeur finale et plus la convergence est rapide. En effet, quand ces poids sont trop faibles, ceci entraîne un apprentissage très long. Pour choisir les valeurs de poids, nous avons utilisé la méthode de Glorot et Bengio publiée en 2010, qui est une méthode de convergence du gradient plus rapide que celles utilisées avec des valeurs aléatoires des poids [93].

Glorot et Bengio proposent une initialisation appelée *initialisation de Xavier* (ou de Glorot, respectivement prénom et nom du chercheur). Nous choisissons de garder la

méthode Glorot, par défaut de la bibliothèque logicielle qui est la distribution uniforme dont l'intervalle d'existence est [-d, d] [93]:

$$d = \sqrt{\frac{6}{nombre\ d'unités\ d'entrée + nombre\ d'unités\ de\ sortie}} \tag{4.9}$$

Le taux d'apprentissage est fixé à 0.001, le Dopout (abandon de neurone) est de 0.5 [78], le nombre d'Epoch est 50 et la taille du lot (Batch size) est fixé à 25 (choix empirique). Pour rappel, en Deep Learning, le Batch size désigne la répartition des données d'apprentissage en lots. Un Epoch représente une passe en avant et en arrière une seule fois de toutes les données d'apprentissage.

La figure 4.5 illustre l'entraînement du réseau, elle présente l'évolution du Taux de détection (Accuracy) et le Coût (Loss) pour les données de validation, en fonction du nombre d'Epoch (ou nombre d'itération). Nous avons testé plusieurs valeurs pour ce paramètre, 50 Epoch est suffisant pour une bonne convergence du réseau et qu'aucune baisse de performance n'apparaît au-delà.

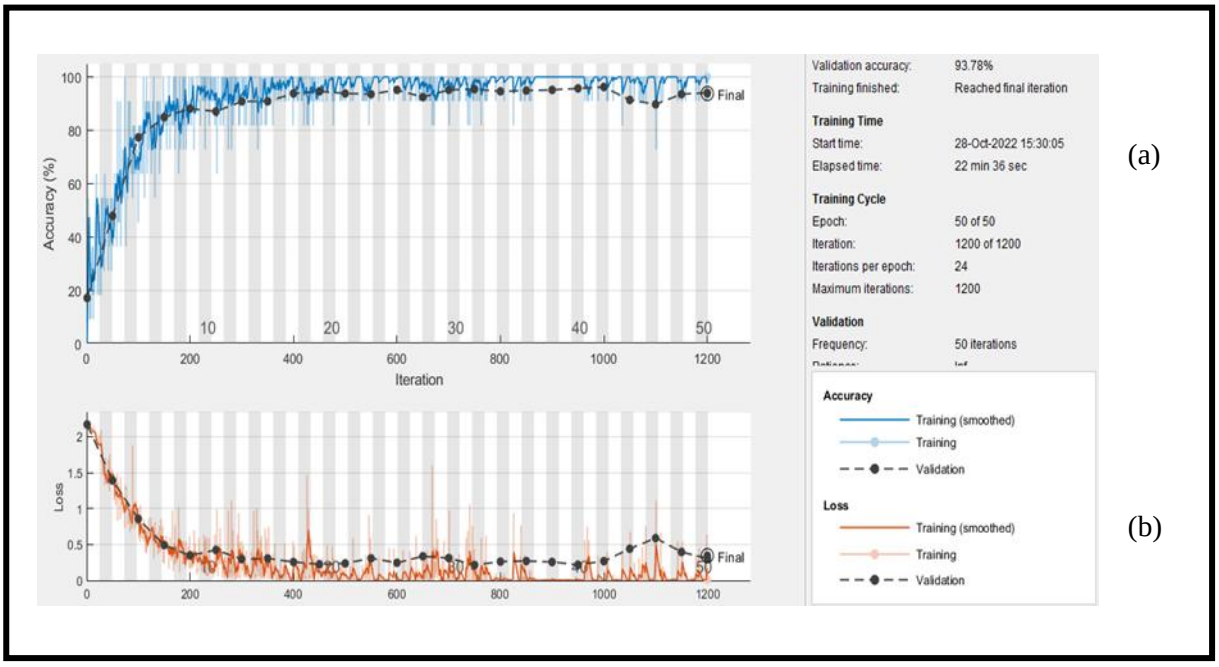


Figure 4.5 : Evolution du Taux de précision (Accuracy) (a) et la fonction du Coût (b) lors de l'apprentissage en fonction du nombre d'Epoch

4.3.8. Matrice de confusion et évaluation des performances

Les performances de détection ou de classification de voix pathologiques sont représentées par un tableau à deux dimensions appelé *Matrice de Confusion*. Les voix réelles sont disposées en lignes et les voix prédites en colonnes (Tab.4.5) [94, 95].

Tableau 4.5 : Matrice de confusion

		Voix Détectées		Total
		Voix Normale	Voix Pathologique	
Voix réelles	Voix Normale P	Vrai Positif VP	Faux Négatif FN	P= VP+FN
	Voix Pathologique N	Faux Positif FP	Vrai Négatif VN	N= FP+VN

Si une voix est positive (P) et qu'elle est détectée comme positive, c'est-à-dire une voix normale correctement détectée, elle est comptée comme un **V**rai **P**ositif (VP). Si elle est détectée comme négative, donc elle est considérée comme un **F**aux **N**égatif (FN). Si une voix est négative (N) et qu'elle est détectée négative, elle est considérée comme un **V**rai **N**égatif (VN), si elle est détectée comme positive, donc elle considérée comme un **F**aux **P**ositif (FP).

Afin de mesurer les performances d'un classifieur de voix pathologiques, trois principaux indices sont pris en considération : la Précision, la Sensibilité et la Spécificité [74,94,95].

- ***Précision*** (AC : Accuracy en anglais) est l'une des mesures couramment utilisées pour la performance de détection et de classification. Elle est définie comme un rapport entre les voix correctement détectées et le nombre total de voix.

$$AC\ (\%) = \frac{VP+VN}{VP+FP+VN+FN} * 100 = \frac{VP+VN}{P+N} * 100 \qquad (4.10)$$

- ***Sensibilité***, représente le **T**aux de **V**rais **P**ositif (TVR). C'est la capacité d'un classifieur de détecter les échantillons positifs correctement classés par rapport au nombre total d'échantillons positifs. Il est estimé par l'équation suivante :

$$TVP\ (\%) = \frac{VP}{VP+FN} * 100 = \frac{VP}{P} * 100 \qquad (4.11)$$

- ***Spécificité***, concerne les échantillons négatifs ou pathologiques, elle représente le **T**aux de **V**rais **N**égatif (TVN). C'est la capacité d'un classifieur de détecter les échantillons négatifs correctement classés par rapport au nombre total d'échantillons négatifs. Il est estimé par l'équation suivante :

$$TVN(\%) = \frac{VN}{VN+FP} * 100 = \frac{VN}{N} * 100 \qquad (4.12)$$

Généralement, on peut considérer la sensibilité et la spécificité comme deux types de précision, où la première pour les échantillons positifs réels et la seconde pour les échantillons négatifs réels. La sensibilité dépend de VP et FN qui sont dans la même ligne de la matrice de confusion, et de même, l'indice de spécificité dépend de VN et FP qui sont dans la même ligne.

4.4. Résultats expérimentaux

Nous avons réalisé dans ce travail, deux systèmes de détection pour deux pathologies différentes : PRU et LT. Vue la particularité dysphonique de chacune de ces dernières, chaque système est spécialisé dans la détection d'une seule pathologie.

4.4.1. Cas de la pathologie PRU

Les tableaux 4.6 et 4.7 illustrent les matrices de confusions obtenues par notre système de détection, pour le cas de la PRU avant et après la rééducation en indiquant le taux global de détection représenté par la précision ainsi que la sensibilité et la spécificité du système. Nous avons remarqué une détection (spécificité) totale (100 %) de voix pathologique avant la rééducation avec un taux élevé de précision par le système (95.87 %).

Après rééducation, nous avons observé une confusion ente les voix normales et pathologiques traduite par un taux de spécificité de 83.33 % ce qui a diminué la précision du système (88.65 %).

La sensibilité de notre système de détection est considérée comme très élevée (92.72 %) vu le facteur de variabilité interlocuteur et intra-locuteur du corpus qui peut causer des difficultés en terme de performance pour les systèmes de Reconnaissance Automatique de la Parole d'une manière générale.

Tableau 4.6 : Matrice de confusion et taux de détection obtenu pour la PRU avant rééducation

		Voix Détectées		Total	Précision AC (%)	Sensibilité TVP (%)	Spécificité TVN (%)
		Voix Normale	Voix Pathologique				
Voix réelles	Voix Normale P	**102**	8	110	**95.87**	**92.72**	**100**
	Voix Pathologique N	00	**84**	84			

Tableau 4.7 : Matrice de confusion et taux de détection obtenu pour la PRU après la rééducation

		Voix Détectées		Total	Précision AC (%)	Sensibilité TVP (%)	Spécificité TVN (%)
		Voix Normale	Voix Pathologique				
Voix réelles	Voix Normale P	**102**	8	110	**88.65**	**92.72**	**83.33**
	Voix Pathologique N	14	**70**	84			

L'analyse de ces données confirme les résultats obtenus précédemment par l'analyse acoustique objective. La différence significative entre la spécificité et la sensibilité du système de détection est expliquée par les écarts entre les voix de références (normales) et pathologiques après rééducation des différents paramètres acoustiques.

4.4.2. Cas de la pathologie LT

Les tableaux 4.8 et 4.9 présentent le taux global de détection pour le cas de la LT, représenté par la précision ainsi que la sensibilité et la spécificité du système à partir des matrices de confusions obtenues par notre système de détection.

Tableau 4.8 : Matrice de confusion et taux de détection obtenu pour la LT avant la rééducation

		Voix Détectée		Total	Précision AC (%)	Sensibilité TVP (%)	Spécificité TVN (%)
		Voix Normale	Voix Pathologique				
Voix réelles	Voix Normale P	**87**	15	102	**92.10**	**85.29**	**100**
	Voix Pathologique N	00	**88**	88			

Tableau 4.9 : Matrice de confusion et taux de détection obtenu pour la LT après la rééducation

		Voix Détectée		Total	Précision AC (%)	Sensibilité TVP (%)	Spécificité TVN (%)
		Voix Normale	Voix Pathologique				
Voix réelles	Voix Normale P	**87**	15	102	**78.94**	**85.29**	**71.59**
	Voix Pathologique N	25	**63**	88			

L'analyse de ces résultats montre une dégradation de la performance du système de détection par rapport à la PRU où nous avons enregistré un taux de précision de 78.94 %. Ce résultat peut être expliqué par la particularité de cette pathologie morphologique où le noyau du mécanisme de phonation a été remplacé (le vibrateur laryngé). Les résultats de l'analyse acoustique réalisés précédemment sur la LT ont également confirmé cette constatation par l'obtention des écarts plus ou moins significatifs entre la voix de LT et celle de la norme de référence.

4.5. Influence du type d'analyse

Nombreux travaux de recherches scientifiques ont montré que, le choix de l'analyse acoustique utilisée dans les systèmes de Reconnaissance Automatique de la Parole en générale et ceux qui sont dédiés principalement aux voix pathologiques peuvent améliorer considérablement les performances de ces systèmes [96,97].

Afin de tester les performances de notre système de détection pour les différentes techniques d'analyse acoustique, l'apprentissage et les tests ont été effectués en utilisant

les coefficients MFCC combinés avec Jitter, Shimmer, HNR, HPR, H_1-H_2 et le CPP (Table.4.10).

Tableau 4.10 : Performance du système en fonction du type d'Analyse Acoustique

Type d'analyse acoustique	Précision (%)	
	PRU	LT
MFCC	84.56	73.35
MFCC, Jitter, Shimmer	86.31	77.05
MFCC, Jitter, Shimmer, HNR,HPR, H_1-H_2, CPP	**88.65**	**78.94**

Les résultats obtenus montrent une discrimination importante de notre système de détection pour l'analyse acoustique multi-variables de MFCC combinée avec Jitter, Shimmer, HNR, HPR, H_1-H_2 et CPP.

Nous avons remarqué également que, l'introduction des paramètres du bruit dans l'extraction du vecteur acoustique a amélioré considérablement les performances du système, notamment pour la PRU vu que cette pathologie se caractérise par une voix soufflée, résultats d'un mauvais accolement des deux cordes vocales induisant un passage d'un flux d'air plus important par rapport à une voix normale.

4.6. Conclusion

Nous avons présenté dans ce chapitre un système d'évaluation de la qualité des voix pathologiques. Le système fonctionne sur la base d'une détection automatique de la voix normale/pathologique entre les différents échantillons introduits. Le noyau classificateur de ce système est un Réseau de Neurone Récurrent de type LSTM.

Lors de la conception de notre système, nous avons utilisé une analyse acoustique Muli-variables très représentative et discriminante, celle de l'analyse MFCC combinée avec les paramètres acoustiques de la stabilité laryngée (Jitter et Shimmer) et ceux du Bruit (HNR, HPR, H_1-H_2 et CPP), afin d'avoir une bonne modélisation du signal vocal.

Cette étude nous a montré que la détection automatique de la voix pathologique est une tâche difficile à réaliser du fait de la complexité et de la variabilité de la voix. De ce fait, un choix adéquat des paramètres de l'analyse acoustique permettra une

meilleure discrimination automatique lors d'un mélange de diverses voix pathologiques.

L’analyse des résultats obtenus nous a permis de déduire que l’évaluation objective par les réseaux de neurones profonds donne des résultats très intéressants. Néanmoins, la confusion entre les voix normales et pathologiques dans la phase de détection, pose un véritable problème aux orthophonistes rééducateurs, en terme de l’évaluation de la qualité de la voix, ce qui rend son utilisation seule dans le processus de rééducation orthophonique inefficace.

Discussions sur les deux méthodes

Nous avons déjà évoqué lors de la description de la méthode d'évaluation subjective, la limite principale de cette technique perceptive qui concerne le manque de fidélité *intra et inter-évaluateurs*, c'est à-dire la stabilité du jugement entre différents juges et la stabilité dans le temps du juge lui-même (test-retest). En effet, sans entraînement préalable ni référence, le pourcentage de variation entre les auditeurs pourra atteindre 50 % [98].

Comme dans tous les systèmes de Reconnaissance Automatique de la Parole, la variabilité inter et intra-locuteur pose de nombreux problèmes aux chercheurs, lors de la conception des systèmes de détection ou classification de voix pathologiques. Cette *variabilité* est dite *intra-locuteur* si un même message prononcé deux fois par un même locuteur dans des conditions identiques produit deux formes spectrales différentes. Elle est due à l'état émotionnel, le débit de parole, le degré d'articulation, etc. Par ailleurs, le même message prononcé par deux locuteurs différents engendre des variations beaucoup plus grandes, classées dans les variations dites *inter-locuteurs.* De ce fait, les résultats obtenus par des réseaux de neurones récurrents, dans notre système de classification (détection) automatique de la voix pathologique, doivent être pris en précaution, à cause de cette variabilité qui conduit à une confusion entre les voix normales et pathologiques dans la phase de détection.

Egalement pour la méthode d'évaluation par l'analyse acoustique, et à cause de cette variabilité, les mesures acoustiques effectuées sur deux [a] tenus par un même locuteur, par exemple, sont différentes. Plus étonnamment encore, les paramètres mesurés sur un même [a] diffèrent selon l'endroit où est placée la fenêtre d'analyse.

Une autre critique pouvant être faite envers les deux outils d'évaluation, qui est d'ordre technique, elle réside dans leur forte dépendance au matériel d'enregistrement utilisé. Nous avons montré que certains paramètres analysant le bruit de la voix comme le HPR et HNR, nécessitent une très bonne qualité du signal vocal. Saenz-Lechon et al. ont montré qu'il est possible de distinguer une voix normale comme voix dysphonique si ces deux populations sont enregistrées avec des microphones différents. En effet, les systèmes automatiques fonctionnent en aveugle et donc modélisent toute forme de variation (et donc des conditions d'enregistrement) [99].

Malgré que les résultats obtenus dans ce travail par l'analyse acoustique sont très intéressants, mais ils restent insuffisants, s'ils ne sont pas soutenu par des mesures

aérodynamiques. En effet, la voix et la parole sont la conséquence acoustique de phénomènes aérodynamiques qui se produisent dans le conduit vocal en fonction des mouvements des organes articulateurs. Les paramètres aérodynamiques comme la pressions sous-glottique et les débits d'air nasal et oral, doivent être impliqués dans l'évaluation objective des dysphonies. Le débit d'air oral, associé aux paramètres acoustiques, permet de bien évaluer le rendement laryngien. Il permet également d'évaluer précisément la fuite glottique en phonation, en relation avec le bruit de souffle, dans les cas où l'accolement total des cordes est empêché par une paralysie laryngée.

L'avantage de l'analyse acoustique par rapport à la méthode neuronale, réside dans la possibilité de déterminer l'étape où le travail de rééducation est en deçà de l'objectif. L'interprétation des résultats de l'analyse acoustique permet à l'orthophoniste rééducateur de corriger et rattraper le processus de rééducation dans l'étape concernée. Si, par exemple, pour des valeurs de shimmer et HNR proches de la norme de référence, et si nous mesurons une faible valeur de l'intensité, dans ce cas, c'est le travail de souffle qui nécessitera une prise en charge particulière. En revanche, la méthode neuronale comme la plupart des classifieurs, souffre d'un problème métrologique, c'est-à-dire qu'elle associe un échantillon vocal à une catégorie (ex: voix normale ou dysphonique), tache intéressante mais qui peut être insuffisante si l'on souhaite obtenir une valeur analogique pour observer et suivre le processus de rééducation. Ces systèmes, malgré leur simplicité d'utilisation, fonctionnent comme des boites noires. Il est très difficile de savoir sur quels éléments a porté la décision. Ce qui ne convient pas nécessairement au clinicien ou au phonéticien qui cherche à comprendre et expliquer ses observations.

De nombreuses études ont montré une forte corrélation entre les résultats des deux méthodes objective et subjective. Les cliniciens Eadie et Doyle soulignent que la précision du diagnostique ne pourra atteindre 100 % que lorsque les évaluations perceptive et acoustique sont combinées [100].

Schindler et al. ont eux aussi mesuré la corrélation entre des données perceptives et des mesures objectives (TMP, jitter, shimmer, HNR, F0 moyenne), dans des groupes dysphoniques de différentes étiologies. Les corrélations significatives mesurées variaient selon l'étiologie de la dysphonie. Les auteurs en concluent que les patients ne sont pas

attentifs aux mêmes aspects de leurs voix en fonction de l'étiologie de leurs dysphonies, ce qui donne de bonnes indications pour la prise en charge [101].

Toutes ces données et d'après le travail réalisé dans cet axe de recherche, et dans le but d'évaluer le processus de rééducation des dysphonies en milieu hospitalier algérien, nous observons une tendance à combiner les deux méthodes subjective et objective, pour une évaluation complète à l'aide de différents outils complémentaires : anamnèse, mesures objectives acoustiques, aérodynamiques et évaluation subjective par le patient et par le thérapeute en intégrant aussi les systèmes de classification automatique comme outil d'évaluation. Ces deux méthodes objective et subjective doivent être *complémentaires* et leurs résultats seront indispensables à la réalisation d'un bilan vocal complet du patient dysphonique.

Conclusions Générales et Perspectives

Ce travail s'inscrit dans le cadre de l'évaluation objective des dysphonies qui résultent d'une lésion neurologique et organique liée aux cordes vocales. Nous avons choisi pour cela deux pathologies à traiter, neurologique et organique qui sont respectivement : la Paralysie Laryngée Unilatérale et la Laryngectomie Totale.

Nous avons utilisé dans cette étude deux méthodes d'évaluation objectives : la première est analytique ou physique basée sur les paramètres d'analyse acoustique porteurs d'informations sur les dysfonctionnements vocaux, en particulier les paramètres mesurant l'instabilité de la vibration des cordes vocales. Dans la seconde méthode, nous avons également évalué objectivement la qualité de la voix, par l'élaboration d'un système de détection automatique de voix pathologique par les réseaux de neurones artificiels. Le système fonctionne sur la base d'une détection automatique de la voix normale/pathologique entre les différents échantillons introduits. Le noyau du classificateur de ce système est un Réseau de Neurone Récurrent de type LSTM.

L'objectif de ce travail est l'évaluation objective de la qualité de la voix pathologique en vue de leur exploitation dans le processus de la réhabilitation de la parole pour une prise en charge fiable des dysphonies dans le domaine des troubles de la communication parlée en milieu hospitalier algérien.

Vue la particularité de l'analyse acoustique sur l'évaluation de la qualité des voix dysphoniques, qui nécessite des mesures fines pour avoir une bonne qualité du signal sonore, nous avons donné une grande importance au choix du matériel d'enregistrement avec son utilisation optimale.

Globalement, et après la période de rééducation, l'évaluation objective basée sur l'analyse acoustique montre une stabilité générale de la vibration laryngée pour la PLU et donc l'absence de problèmes de contrôle pneumo-phonique. Ce constat global est affirmée par l'évaluation subjective à l'écoute par l'orthophoniste rééducateur. Néanmoins, l'analyse de certains paramètres acoustiques comme le Shimmer a permis de constater que la rééducation a été basée sur un travail de respiration plus que le travail de vocalisation ou de tenue de voyelles qui permet de faire vibrer les cordes vocales dans de bonnes conditions.

Pour la Laryngectomie Totale, nous avons enregistré une stabilité de la vibration de la Néoglotte acceptable, mais reste encore loin par rapport à la vibration normale des cordes vocales. La technique de rééducation basée sur la parole œsophagienne donne de bons résultats

si le patient maîtrise les étapes de la production de la parole œsophagienne et les applique avec une grande volonté. Cependant, le choix de la méthode de l'évaluation objective basée sur les paramètres de l'instabilité de la vibration de la pseudo-glotte reste insuffisante vue qu'on utilise plus le vibrateur naturel qui est la glotte et la source d'énergie qui est la soufflerie pulmonaire.

L'application des réseaux de neurones récurrents LSTM, dans un système de détection automatique de la voix pathologique, nous a permis d'avoir des résultats appréciables. Cependant, le problème de confusion entre les voix normales et pathologiques dans la phase de détection, rend l'utilisation des réseaux de neurones seule, comme méthode d'évaluation objective dans le processus de rééducation orthophonique inefficace, mais elle pourra aider l'orthophoniste dans la détection et l'évaluation de la voix pathologique avec les deux autres méthodes : objective par l'analyse acoustique et subjective par l'écoute.

Globalement, Ce travail nous a permis de dire que l'utilisation exclusive et abusive de l'ouïe pour évaluer la phase de la rééducation vocale dans le milieu hospitalier algérien reste insuffisante. Il est important de corréler les données perceptives avec les méthodes objectives, de façon à pouvoir élaborer un projet thérapeutique au plus près des attentes et des difficultés du patient.

La limite de notre méthodologie réside dans l'utilisation d'appareillages de mesure. Cependant, l'évaluation instrumentale analytique a été conçue, à l'origine, pour fournir une réponse, sous la forme d'une ou plusieurs mesures. Pour le cas des paralysies laryngées, l'immobilité d'une corde vocale se traduit par une importante fuite glottique et pour mesurer cette fuite d'air, le meilleur instrument reste le débitmètre qui peut fournir le débit d'air avant et après chirurgie, offrant directement une estimation chiffrée du taux de fermeture de la glotte en phonation.

En perspectives, nous proposons d'élaborer un protocole d'évaluation au niveau des services ORL, issue d'une coopération entre l'orthophoniste praticien en milieu hospitalier et l'ingénieur ou chercheur phonéticien. Ce protocole sera suivi par une application informatique pour automatiser l'évaluation de la qualité des dysfonctionnements de la parole. Ainsi, nous proposons introduire d'autres paramètres dans l'analyse acoustique, telles que les mesures aérodynamiques qui permettent de mesurer avec des capteurs, les pressions ou les débits d'air pour évaluer le degré de la fuite glottique.

Références Bibliographiques

[1] M. Kabache, M. Guerti, "Multi parametric method for the objective Acoustic Evaluation of the Voice Produced by laryngectomy patients". *Instrumentation, mésure et métrologie*, vol.20, no.3, pp.137-142, 2021.

[2] M. Kabache, M. Guerti, "Acoustic Analysis of Voice Signal of Patients with Unilateral Laryngeal Paralysis a view to objective evaluation after rehabilitation", *Revue de Traitement de Signal*, vol.38, no. 5, pp.1339-1344, 2021.

[3] K. Ferrat, , M. Guerti, "A study of sounds produced by Algerian esophageal speakers*"*, *African Health Sciences*, vol. 12, no. 4, 2012.

[4] Yu, Ping, M. Ouaknine, J. Revis, A. Giovanni, "Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements", *Journal of Voice*, vol. 15, no. 4, pp. 529–542, 2001.

[5] E. Saltürk, T. Özdemir, Z. Lütfi Kumral, E. Karabacakoğlu, H. Kumral, G. Yildiz, Y. Mersinlioğlu, , G. Atar, G. Berkiten, Y. Yildirim, "Subjective and objective voice evaluation in Sjögren's syndrome", *Logopedics Phoniatrics Vocology*, vol. 42, no. 1, pp. 9-11, 2017.

[6] J. Kreiman, B. Gerratti, R. Kempster, G. B. Erman, G. S. Berke, "Perceptual Evaluation of Voice Quality: Review, Tutorial, and a Framework for Future Research*"*, *Journal of Speech and Hearing Research*, vol 36, pp. 21-40, 1996.

[7] A. Van Doorslaer. https://quizlet.com/259788982/schema-de-lappareil-phonatoire-diagram/.2021.

[8] Le cerveau à tous les niveaux, https://lecerveau.mcgill.ca/flash/capsules/outil_bleu21.html, 2021

[9] F. Le Huche, A. Allali, "La voix : Anatomie et physiologie des organes de la voix et de la parole*"*, Tome 1, 4éme éd., *collection phoniatrie, Elsevier Masson*, Paris, 2001.

[10] V.L.C. Research. https://blogglophys.wordpress.com/2015/06/23/larynx.2021

[11] H. David, M. Farland, "L'anatomie en orthophonie parole, déglutition et audition*"*, 3éme édition, *Elsevier Masson*, Paris, 2016.

[12] M. Perarad, M. Miclot, M. Serazin. https://cdn.website-editor.net/50befd41f5384db9b59f3b7296cd351f/files/uploaded/phonaperard.pdf 2021

[13] F. Le Huche, A. Allali, " La voix : pathologies vocales d'origine fonctionnelle*"*, Tome 2, 3éme éd, *collection phoniatrie, Elsevier Masson*, Paris, 2010.

[14] F. Le Huche, , A. Allali, "La voix : pathologies vocales d'origine organique*",* Paris, Tome 3, 2éme éd, *collection phoniatrie, Elsevier Masson*, Paris, 2010.

[15] M. Omari "Les paralysies laryngées*"*, Thèse de doctorat en médecine Université de sidi Mohamed, Maroc, 2017.

[16] J.M. Kremer, E. Lederlé, C. Maeder, "Intervention dans les troubles : parole, voix, déglutition et déficiences auditives*"*, Guide de L'Orthophoniste, vol. 4, *éd. Lavoisier*, 2016.

[17] A. E. Aronson, "Les troubles cliniques de la voix*"*, vol. 1, *éd. Elsevier Masson*, Paris, 1983.

[18] F. Le Huche, A. ALLALI, "Défauts de mobilité Laryngée et réhabilitation fonctionnelle", *collection Voix, Parole, Langage, éd. Solal*, 2007.

[19] F. Le Huche, A. Allali, "La voix, Thérapeutique des troubles vocaux", *collection phoniatrie, Tome 4, éd. Elsevier Masson*, Paris, 2002.

[20] E. Babin, "Le cancer de la gorge et la laryngectomie*:* la découration", *éd. L'Harmattan*. Paris, 2011.

[21] J. M. Prades, E. Reyt, "Cancers du larynx". *EMC* - Oto-Rhino-Laryngol. vol. 8, pp. 1–15, 2013.

[22] G. Heuillet-Martin, L. Conrad, "Du silence à la voix: nouveau manuel de rééducation après laryngectomie totale". *éd. Solal*, 1997.

[23] A. Fanny, "La réhabilitation vocale après Laryngectomie, état des lieux et établissement d'un support d'éducation thérapeutique", *Thèse de Doctorat en médicine. Université Angers*, France, 2017.

[24] C. G. Tang, C. F. Sinclair. "Voice Restoration After Total Laryngectomy". *Otolaryngol. Clin. North Am*., vol. 48, pp. 687–702, 2015.

[25] J. Algaba "Voice rehabilitation for laryngectomized patients". Laryngol Otol Rhinol, vol. 108, pp.139-142, 1987.

[26] N. Ouattassi, "La réhabilitation vocale après Laryngectomie, Les aspects acoustiques de la voix oesophagienne: développement d'une application informatique d'analyse acoustique de la voix", *Thèse de Doctorat en médicine, Université de Fès*. Maroc, 2011.

[27] E. Blom, M. Singer, R. Hamaker, "A prospective study of tracheoesophageal speech". *Archives of Otorhinolaryngology-Head & Neck Surgery*, vol. 112, pp. 440-447, 1986.

[28] Institut National du Cancer. https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancers-de-la-sphere-ORL-voies-aerodigestives-superieures/Focus-la-tracheotomie-et-la-tracheostomie France, 2019.

[29] L. Traissac, F. Devars, M. Gioux, J. Petit, A. Benjebria, C. Henry, "Vocal rehabilitation of the total laryngectomized patient by phonatory implant. Current results*"*. *Rev. Laryngol. Otol. Rhinol*. vol. 108, pp. 157–159, 1987.

[30] D.M. Hartl "Méthodes actuelles d'évaluation des dysphonies", *Ann Otolaryngol Chir Cervicofac*, vol. 122, no. 4, pp. 163-172, 2005.

[31] L. Crevier-Buchman, S. Brihaye-Arpin, A. Sauvignet, C. Tessier, Monfrais- C. Pfauwadel, D. Brasnu "Dysphonies non organiques (dysfonctionnelles). *EMC Oto-Rhino-Laryngologie, Elsevier*, vol. 21, no. 2, pp. 1-12, 2006.

[32] B-H. Jacobson, A. Johnson, C. Crywalski, A. Silberglent, G. Jacobson, M-S. bennnger, C. Newmen, "The voice handicap index (VHI: development and validation". *American Journal of Speech Pathology*, vol. 6, no. 3, pp. 66-70, 1997.

[33] F. Bouwers, Frederik G. Dikkers. "A Retrospective Study Concerning the Psychosocial Impact of Voice Disorders: Voice Handicap Index Change in Patients with Benign Voice Disorders After Treatment (Measured with the Dutch Version of the VHI)". *Journal of Voice*, vol. 23, no. 2, pp. 218-224, 2009.

[34] B. Hammarberg, , B. Fritzell, J. Gauffin, , J. Sundberg, L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities", *Acta-Oto-laryngologica*, vol. 90, no.1 pp. 441-451, 1980.

[35] M. Hirano, "Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice. Clinical Evaluation of Voice", *Springer Verlag, Wien*, 1981.

[36] V. Woisard-Bassols "Bilan clinique de la voix". *Encyd Méd Chir (Editions Scientifiques et Médicales Elsevier SAS, Otorhino-laryngologie*, 20-753-A-10, 2000.

[37] W. Koenig, "The Sound Spectrograph". *Thirty-First Meeting of the Acoustical Society of America*, 1947.

[38] R. D. Kent, "Speech Sciences". Edition *Singular*, 1997.

[39] R. J. Baken, R. Orlikoff, "Clinical measurement of speech and voice", Second Edition. *Singular*, 1999.

[40] G. Cornut, S. Arom, "Moyens d'investigation et pédagogie de la voix chantée", *Actes du colloque*, Lyon, Symétrie, 2001.

[41] G. Sharma, D. Prasad, K. Empathy, S. Krishnan,. "Screening and analysis of specific language impairment in young children by analyzing the textures of speech signal", *In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 964-967, 2020.

[42] L. Zahid, M. Maqsood, M. Y. Durrani, M. Bakhtyar , J. Baber , H. Jamal, I. Mehmoud, Y. Song "A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease", *IEEE Access*, vol. 8, pp. 35482-35495, 2020.

[43] R. J. Baken, "Irregularity of vocal period and amplitude: a first approach to the fractal analysis of voice". *Journal of Voice*, vol. 4, no. 3, pp. 185–197, 1990.

[44] Y. Koike, H. Takahashi, T. Calcaterra, "Acoustic Measures for Detecting Laryngeal Pathology", *Acta-Oto-laryngologica*., vol. 84, no. 1-6, pp. 105-117, 1977.

[45] E. Yumoto, W.J. Gould, "Harmonics to noise ratio as an index of the degree of hoartheness", *JASA*, vol 71, no. 6, pp. 1544-1550, 1982.

[46] G. De Krom, "A Cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals", *Journal of Speech and Hearing Research*, vol. 36, no. 2, pp. 254–266, 1993.

[47] K. Shoji, E. Regenbogen, J. Daw Yu, S. Blaugrund, "High-Frequency Power Ration of Breathy Voice", *Laryngoscope*, vol. 102, no. 3, pp. 267-271, 1992.

[48] B. Fritzell, B. Hammarberg, J. Gauffin, I. Karlsson, J. Sundberg, "Breathiness and insufficient vocal fold closure". *Journal of Phonetics*, vol. 14, pp. 549-553. 1986

[49] J. P. Jeannon, "Vocim analysis of laryngeal images: is breathiness related to the glottic area? ", *Clinical Otolatyngology & Allied Sciences*, vol. 23, pp. 351-353, 1998.

[50] D.H. Klatt & L.C. Klatt "Analysis, synthesis and perception of voice quality variations among male and female talkers", *Journal of the Acoustical Society of America*, vol. 87, pp. 820–856, 1990.

[51] R. Shrivastav, C. M. Sapienza, "Objective measures of breathy voice quality obtained using an auditory model". *Journal of the Acoustical Society of America*, vol. 114, pp. 2217–2224, 2003.

[52] J. Hillenbrand, R. A. Houde, "Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech". *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 2, pp. 311-321, 1994.

[53] A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, "Discriminating Pathological Voice From Healthy Voice Using Cepstral Peak Prominence Smoothed Distribution in Sustained 41 Vowel". *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646-654, 2018.

[54] Y. D. Heman-Ackah, R. T. Sataloff, G. Laureyns, D. Lurie, D. Michael, R. Heuer, A. Rubin, R. Eller, S. Chandran, M. Abaza, K. Lyons, V. Divi, J. Lott, J. Johnson, J. Hillenbrand, "*Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia*", *Journal of Voice*, vol. 28, no. 6, pp. 783-788, 2014.

[55] R. Fraile, J. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis". *Biomedical Signal Processing and Control*, vol. 14, pp. 42‑54, 2014.

[56] C. Batthyany, Y. Maryn, I. Trauwaen, E. Caelenberghe, J. van Dinther, A. Zarowski, F. Wuyts. "A Case of Specificity: How Does the Acoustic Voice Quality Index Perform in Normophonic Subjects?" *Applied Sciences*, vol. 9 no.12, 2527. 2019.

[57] A. Giovanni, C. Heim, D. Demolin, J. Triglia, "Estimated subglottic pressure in normal and dysphonic subjects", *Ann. Otol. Rhinol. Laryngol*, vol. 109, pp. 500- 504, 2000.

[58] L. Claire Pillot, "Pression sous-glottique et débit oral d'air expiré comme aides à la pose du diagnostic de dysodie ; implications pour la rééducation vocale". *Entretiens d'orthophonie*, pp.32-45, 2011.

[59] M. Hirano, Y. KOIKE, H. Von LEDEN, "Maximum phonation time and air usage during phonation". *Folia Phoniatrica*, vol. 20, pp. 185-201. 1968.

[60] A. M. Johnson, A. Goldfine, "Intrasubject Reliability of Maximum Phonation Time". *Journal of Voice*, vol. 30, no. 6, pp. 775.e1-775.e4, 2016

[61] A. GiovanniI, D. Robert, N. Estublier, , B. Teston, "Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements", *Folia Phoniatrica et Logopeadica*, vol. 48, pp. 175-185, 1996.

[62] M. Kabache "Application des Réseaux de Neurones à la reconnaissance des phonèmes spécifiques de l'Arabe Standard ", *Thèse magister*, CRSTDLA-ENS, Alger, 2006.

[63] G. Gelly, "Réseaux de neurones récurrents pour le traitement automatique de la parole", *Thèse de doctorat*, université Parie Saclay-Paris Sud, 2017.

[64] https://www.youtube.com/watch?v=64Yl3K-4FzM.

[65] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[66] P. Boersma, D. Weenink, "Praat: doing phonetics by computer", version 6.1, http://www.fon.hum.uva.nl/praat/. 2019.

[67] E. Sicard, , L. Meyrieux, M. Moreau, A. Remacle, "L'analyse acoustique des voix d'enfants de 5 ans: Proposition de valeurs de référence pour les logiciels PRAAT et VOCALAB ". *Journées de Phonétique Clinique*, Belgique, 2019.

[68] H. Oguz, M.A. Kilic, M.A. Şafak, " Comparison of results in two acoustic analysis programs: praat and mdvp", *Turkish Journal of Medical Sciences, vol.* 41, no. 5, pp. 835-841, 2011.

[69] A. Ofer, W. Michael, A. Noam, "A clinical comparison between two acoustic analysis softwares: MDVP and Praat", *Biomedical Signal Processing and Control*, vol.4, pp. 202-205, 2009.

[70] Y. Maryn, , P. Corthals, , M. De Bodt, , P. Van Cauwenberge, D. Deliyski, "Perturbation measures of voice: a comparative study between multi-dimensional voice program and praat". *Folia Phoniatrica et Logopaedica*, vol. 61, pp. 217-226, 2009.

[71] C. Fredouille, G. Pouchoulin, J.-F. Bonastre., M. Azzarello, A. Giovanni, et A. Ghio, "Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia)", *Proceedings of Interspeech'05*, Lisboa, Portugal, vol. 90, pp. 149-152, 2005.

[72] O. Eskidere, A. G. urhanlı, "Voice disorder classification based on multitaper mel frequency cepstral coefficients features", *Computational and mathematical methods in medicine*, vol. 4, pp. 1-15, 2015.

[73] C.M. Vikram, K. Umarani, "Pathological Voice Analysis To Detect Neurological Disorders Using MFCC & SVM", *International Journal of Advanced Electrical and Electronics Engineering*, vol. 2, no. 4, 2013.

[74] F. Teixeira, J. Fernandes, V. Guedes, A. Junior, J.P. Teixeira, "Classification of control/pathologic subjects with support vector machines," *Procedia Computer Science*, vol. 138, pp. 272–279, 2018.

[75] F. Amara, M. Fezari, H. Bourouba "An Improved GMM-SVM System based on Distance Metric for Voice Pathology Detection", *Applied Mathematics & Information Sciences.* vol. 10, no. 3, pp. 1061-1070, 2016

[76] W. Xiang, Z. Jianping, Y. Yonghong, "Discrimination between pathological and normal voices using GMM/SVM approach", *Journal of Voice*, vol. 25, no. 1, 2011.

[77] S. Kumara, K. Anantha, N. U. Cholayya, "Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology", *Journal on Advances in Signal Processing*, 2007

[78] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska_, Z. Galaz_, R. Burget, Z. Smekal, "Voice Pathology Detection Using Deep Learning: a Preliminary Study", *arxiv*, 1907.05905, 2019.

[79] L. Salhi, , T. Mourad, , A. Cherif, "Voice Disorders Identification Using Multilayer Neural Network". *The International Arab Journal of Information Technology*", vol. 7, no. 2, 177-185, 2010.

[80] RT. Ritchings, M. McGillion, CJ. Moore, "*Pathological Voice qualité assement using artificial neural network*", Medical Engineering & Physics, 2002, vol. 24, pp. 561-564.

[81] J. P. Teixeira, P. O. Fernandes, and N. Alves, "*Vocal Acoustic Analysis - Classification of Dysphonic Voices with Artificial Neural Networks*", Procedia Computer Science, vol. 121, pp. 19–26, 2017.

[82] A.S. Sidra, R. Munaf, H. Samreen Z. Hira, "Comparative Analysis of CNN and RNN for Voice Pathology Detection", *BioMed Research International*, 2021, .

[83] V. Srinivasan, V. Ramalingam, P. Arulmozhi, "Artificial neural network based Pathological Voice classification using MFCC features", *International Journal of Science, Environment and Technology*, vol. 3, no 1, pp. 291 – 302. 2014.

[84] V. Guedes, A. Junior, J. Fernandes, F. Teixeira, J. P. Teixeira "Long Short Term Memory on Chronic Laryngitis Classification", *Joana Procedia Computer Science* vol. 138, pp. 250–257, 2018.

[85] D. Sztahó, G. Kiss, T. M. Gábri, "Deep Learning Solution for Pathological Voice Detection using LSTM-based Autoencoder Hybrid with Multi-Task Learning", 14th *International Conference on Bio-inspired Systems and Signal Processing, BIOSIGNALS*, vol. 4, pp. 135-141, 2021.

[86] G. Vibhuti, "Voice Disorder Detection Using Long Short Term Memory (LSTM) Model", *arXiv*:1812.01779, 2018.

[87] C. Etienne, "Apprentissage profond appliquée à la reconnaissance des émotions dans la voix", *Thèse de doctorat*, Université Paris-Saclay, 2018.

[88] F. Simon, "Deep Learning, les fonctions d'activation", 2018 https://www.supinfo.com/articles/single/7923-deep-learning-fonctions.

[89] X. Glorot, A. Bordes, Y. Bengio, "Deep sparse rectifier neural networks" in Aistats, vol. 15, pp. 275, 2011.

[90] C. Etienne, G. Fidanza, A. Petrovskii, L. Devillers, B. Schmauch. "CNN+LSTM Architecture for Speech Emotion Recognition with Data Augmentation". *In Proc. Workshop on Speech, Music and Mind*, pp. 21–25, 2018.

[91] I. Goodfellow, Y. Bengio, A. Courville, "*Deep Learning*" *Edition. MIT Press*, 2016. https://www.deeplearningbook.org.

[92] D.P. Kingma and J.L. Ba, "ADAM: A method for stochastic optimization", *Arxiv*, 14126980v9, 2015.

[93] X. Glorot and Y. Bengio, "Understanding the Difficulty of Training Deep Feedforward Neural Networks", In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–356. Italy, 2010.

[94] B. Sabir, F. Rouda, Y. Khazri, B. Touri, and M. Moussetad, "Improved algorithm for pathological and normal voices identification", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 1, pp. 238-243, 2017.

[95] A. Tharwat, "Classification assessment methods", *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168-192, 2016.

[96] M. Kabache, M. Guerti, " Analyse acoustique multivaraible appliquée à la reconnaissance des emphatiques de l'arabe standard ", *revue al – Lisaniyyat*, , vol. 18, no. 17, pp. 83-99, 2011.

[97] M. Kabache, M. Guerti, " Reconnaissance des phonèmes emphatiques de l'arabe standard par une approche connecxioniste modulaire ", *ICCMD'06*, Annaba, Alegria, pp. 66, 2006.

[98] B. Teston. "L'évaluation instrumentale des dysphonies. Etat actuel et perspectives", Giovanni A. Le bilan d'une dysphonie, *Solal*, pp.105-169, 2004.

[99] N. Saenz-Lechon, J. Godino-Llorente, V. Osma-Ruiz, P. Gomez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection". *Biomedical Signal Processing and Control*, vol.1 no.2, pp. 120-128, 2006.

[100] T. L. Eadie, P. C. Doyle, "Classification of dysphonic voice: Acoustic and auditory perceptual measures", *Journal of Voice*, vol.19 no.1, 2005.

[101] A. Schindler, F. Mozzanica, M. Vedrody, P. Maruzzi, F. Ottaviani, “ Correlation between the Voice Handicap Index and voice measurements in four groups of patients with dysphonia”. *Otolaryngology - Head and Neck Surgery, vol.* 141, pp. 762-769, 2009.