

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد - تلمسان -

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE



MEMOIRE

Présenté pour l'obtention du **diplôme** de **MASTER**

En : Génie biomédical

Spécialité : Instrumentation biomédicale

Par : Mekhellet Ahmed Yacine

Sujet

Détection et Classification des Troubles de la Voix

Soutenu publiquement, le 20/09/2023, devant le jury composé de :

M. KERAÏ Salim	MCA	Université de Tlemcen	Président
M. BENALI Redouane	MCA	Université de Tlemcen	Examineur
M. TAOUALI Sid Ahmed	MCB	Université de Tlemcen	Encadreur

Année universitaire : 2022/20

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à Dieu, dont la grâce et la guidance ont éclairé chaque étape de mon parcours académique.

M. TAOLI Sid Ahmed, vous avez été bien plus qu'un encadreur pour moi. Vos conseils éclairés, votre disponibilité constante, et votre passion pour la recherche m'ont inspiré à donner le meilleur de moi-même. Votre patience inépuisable et votre capacité à trouver des solutions aux obstacles m'ont été d'une aide inestimable. Je vous suis sincèrement reconnaissant pour tout ce que vous avez fait.

Je souhaite également exprimer ma reconnaissance envers les membres du jury qui ont examiné ce mémoire : **M. KERAÏ Salim**, en tant que président, et **M. BENALI Redouane**, en tant qu'examineur. Votre expertise et vos commentaires constructifs ont grandement enrichi ce travail.

Enfin, je tiens à remercier chaleureusement toutes les personnes qui ont apporté leur soutien, que ce soit sous forme de discussions inspirantes, de relectures attentives ou de conseils bienveillants. Votre encouragement constant et votre confiance en moi ont été des sources de motivation essentielles tout au long de ce voyage académique.

Ce mémoire n'aurait pas pu voir le jour sans l'apport exceptionnel de chacune de ces personnes, et je leur suis profondément reconnaissant pour leur précieuse contribution à cette réalisation.

Dédicace

À mes chers parents, ***Mekhellet Mokhtar et Guiraa Meriem,***

Vos sacrifices, votre amour inconditionnel, et votre soutien indéfectible ont été le fondement sur lequel j'ai bâti mon parcours académique. Votre foi en moi m'a toujours poussé à donner le meilleur de moi-même, et votre amour inconditionnel a été le moteur de ma réussite. Vos sacrifices ont pavé la voie vers mes rêves. Cette dédicace est une modeste expression de ma reconnaissance éternelle pour tout ce que vous avez fait. Vous êtes mes piliers, ma force, et mon modèle de persévérance.

À mon frère ***Anes,*** ma sœur ***Sirine*** et ma tante ***Khanssa,***

Votre soutien infaillible et vos sourires lumineux ont été mes réconforts dans les moments de doute. Vous êtes mon inspiration quotidienne. Cette dédicace est un témoignage de mon amour infini pour vous.

À tous mes amis, anciens et ceux que j'ai rencontrés à l'université,

Votre amitié, vos encouragements, et les moments de partage ont rendu cette aventure académique encore plus précieuse. Vous avez été mes compagnons de route, mes partenaires d'étude, et mes sources d'inspiration. Cette dédicace est une expression de ma gratitude pour votre amitié sincère.

À toutes celles et ceux qui ont guidé ma quête du savoir, votre précieuse contribution m'éclaire toujours.

Je tiens également à rendre hommage à notre cher ami ***Hamza-Cherif Ibrahim,*** qui, bien que ne soit plus parmi nous, a laissé une empreinte indélébile dans nos cœurs et a contribué à rendre cette expérience encore plus significative. Son souvenir restera à jamais gravé dans nos souvenirs et notre dédicace à l'amitié sincère.

Résumé :

Ce mémoire se consacre à la classification des voix pathologiques, un domaine vital pour la santé vocale. À travers trois chapitres, il explore les fondamentaux du signal vocal, présente une méthodologie intégrant le traitement du signal, y compris l'analyse en ondelettes et l'extraction des paramètres, puis dévoile les résultats prometteurs obtenus grâce à un modèle SVM pour la classification.

L'objectif majeur est d'améliorer le diagnostic et la prise en charge des troubles vocaux en fournissant des outils plus performants aux professionnels de la santé, en se basant sur la base de données VOICED. Ce travail souligne l'importance cruciale de la santé vocale et la nécessité d'investir dans des méthodes de diagnostic plus précises et accessibles, tout en ouvrant de nouvelles perspectives pour les soins et la qualité de vie des individus touchés.

Mots clés : Classification, Voix pathologiques, Signal vocal, Analyse en ondelettes, Extraction de paramètres, Méthodologie, Modèle SVM, La base de données VOICED.

Abstract:

This thesis is dedicated to the classification of pathological voices, a vital area in vocal health. Over three chapters, it delves into the fundamentals of vocal signal, presents a methodology that integrates signal processing, including wavelet analysis and parameter extraction, and unveils promising results achieved using an SVM model for classification.

The primary objective is to enhance the diagnosis and management of vocal disorders by providing more effective tools to healthcare professionals, based on the VOICED database. This work underscores the critical importance of vocal health and the necessity of investing in more precise and accessible diagnostic methods, while also opening up new prospects for care and the quality of life of affected individuals.

Keywords: Classification, Pathological Voices, Vocal Signal, Wavelet Analysis, Parameter Extraction, Methodology, SVM Model, VOICED Database.

هذه المذكرة مخصصة لتصنيف الأصوات المرضية، وهي منطقة حيوية في الصحة الصوتية. على مدى ثلاثة فصول، يتعمق في أساسيات الإشارة الصوتية، ويقدم منهجية تدمج معالجة الإشارات، بما في ذلك تحليلها واستخراج المعلومات، ويكشف النقاب عن النتائج الواعدة التي تحققت باستخدام نموذج SVM للتصنيف.

الهدف الأساسي هو تعزيز تشخيص وإدارة الاضطرابات الصوتية من خلال توفير أدوات أكثر فعالية لأخصائي الرعاية الصحية، بناء على قاعدة البيانات. يؤكد هذا العمل على الأهمية الحاسمة للصحة الصوتية وضرورة الاستثمار في طرق تشخيص أكثر دقة ويمكن الوصول إليها، مع فتح آفاق جديدة للرعاية ونوعية حياة الأفراد المتضررين.

الكلمات المفتاحية: التصنيف، الأصوات المرضية، الإشارة الصوتية، استخراج المعلومات، المنهجية، نموذج SVM، قاعدة البيانات.

Table des Matières

Remerciements	I
Dédicace	II
Résumé	III
Table des Matières	V
Liste des Figures	VIII
Liste des Tableaux	IX
Tableau des Abréviations	X
Introduction Générale	a
Chapitre 1 : Généralités sur le signal vocal	1
1. Introduction	2
2. Système de production de la parole	2
2.1 Partie sub-glottique (l'appareil respiratoire)	2
2.1.1 Le diaphragme	2
2.1.2 Les poumons.....	3
2.1.3 La trachée	3
2.2 Partie glottique (Le larynx)	3
2.2.1 Les cartilages du larynx.....	4
2.2.1.1 Cartilages impairs.....	4
2.2.1.1.1 Cartilage thyroïdien	4
2.2.1.1.2 Cartilage cricoïde	4
2.2.1.1.3 L'épiglotte.....	4
2.2.1.2 Cartilages pairs	4
2.2.1.2.1 Les cartilages aryténoïdes	4
2.2.1.2.2 Les cartilages cornéliens	4
2.2.1.2.3 Les cartilages cunéiformes	4
2.2.2 Les muscles du larynx	5
2.2.3 Les cordes vocales.....	6
2.3 Partie supra-glottique	7
3. La production de la parole	7
3.1 Sons vocaux	8
3.2 Sons non vocaux	8
4. Troubles de la voix	9

4.1	Définitions.....	9
4.2	Les pathologies	10
4.2.1	Polypes	10
4.2.2	Nodules.....	11
4.2.3	Kystes	12
4.2.4	Tumeur cancérigène	12
4.2.5	Reflux laryngite.....	12
4.2.6	Œdème de Reinke.....	12
4.2.7	Paralysie des cordes vocales.....	13
4.2.8	Traumatismes chirurgicaux	14
4.2.9	Hypercinésie dysphonique de la voix (hyperkinétique).....	14
4.2.10	Hypokinésie dysphonique de la voix (hypokinétique)	14
5.	Méthodes d'identification des troubles de la voix	14
5.1	Identification subjective	14
5.2	Identification objective	15
6.	Paramètres de la voix.....	16
6.1	Fréquence Fondamentale (F0)	16
6.2	Pitch	16
6.3	Hauteur.....	17
6.4	Intensité.....	17
6.5	Timbre.....	17
6.6	Formants.....	17
6.7	Harmoniques	17
6.8	Jitter et Shimmer	17
7.	Conclusion.....	18
Chapitre 2 : La méthodologie de la classification des voix pathologiques.....		19
1.	Introduction.....	20
2.	Transformée de Fourier (TF).....	20
2.1	Transformée de Fourier à court terme (STFT).....	21
3.	Transformée en ondelettes	22
3.1	Familles d'ondelettes	24
3.2	Choix de la meilleure famille.....	24
3.3	Transformée par ondelette continu (TOC).....	25
3.3.1	Décomposition de l'ondelette continue.....	25
3.4	Transformée par ondelette discrète	27
4.	Machines à vecteurs de supports	28

4.1	Principe de fonctionnement	28
4.2	Linéarité et non linéarité	30
4.2.1	SVM linéaire	30
4.2.2	SVM non-linéaire	30
5.	Méthode proposée	31
5.1	Base de données	32
5.2	Décomposition des signaux en ondelette	32
5.3	Extraction des paramètres	34
5.4	Classification (normale ou pathologique)	36
6.	Conclusion.....	37
Chapitre 3 : Résultats et Discussion		38
1.	Introduction.....	39
2.	Travaux effectués	39
2.1	Identification Des Voix Pathologiques Par Réseaux De Neurones	39
2.2	Identification des troubles de la voix par l'analyse temps-fréquence et les machines à vecteurs de support	40
2.3	Détection et classification des voix pathologiques utilisant l'analyse par ondelette discrète basée sur les caractéristiques statistiques d'ordre supérieures	40
2.4	Développement d'un système de détection de la voix pathologique.....	41
2.5	Identification des troubles de la voix à l'aide des techniques d'apprentissage automatique	42
3.	Les étapes de la classification.....	42
3.1	Collecte de Données.....	42
3.2	L'analyse des signaux et l'extraction de caractéristiques	43
3.3	Classification.....	43
3.3.1	Phase d'apprentissage.....	44
3.3.2	Phase de test	44
3.4	Matrice de confusion.....	45
4.	Résultats et discussion	47
4.1	Matrice de Confusion en SVM	47
4.2	Taux de classification.....	48
4.3	Etude comparative des différents résultats des travaux réalisés :	49
5.	Conclusion.....	49
Conclusion Général		50
Bibliographie.....		52

Liste des Figures

Figure 1.1 : Schéma de l'appareil respiratoire.	3
Figure 1.2 : Anatomie du Larynx.	5
Figure 1.3 : Schéma des muscles du Larynx.	6
Figure 1.4 : Les cordes vocales.	7
Figure 1.5 : signal sonore et non sonore.....	9
Figure 1.6 : Polype de la corde vocale	11
Figure 1.7 : Nodules sur la corde vocale.....	11
Figure 1.8 : Tumeur cancéreuse sur les cordes vocales	12
Figure 1.9 : Œdème de Reinke.	13
Figure 1.10 : Paralyse des cordes vocales.....	13
Figure 2.1 : Passage temporel vers fréquentiel par la TF	21
Figure 2.2 : Passage temporel vers temps-fréquence par STFT.....	22
Figure 2.3 : Passage temporel vers temps-échelle par WT	23
Figure 2.4 : Transformée en ondelettes continu pour une voix pathologique	26
Figure 2.5 : Transformée en ondelettes continu pour une voix normale.....	26
Figure 2.6 : Exemple de classification par SVM.....	29
Figure 2.8 : Exemple de transformation des données non linéaires dans un espace de dimension supérieure.....	31
Figure 2.9 : Méthode Proposée de la classification des voix Pathologiques.....	31
Figure 2.10 : Exemple de décomposition en ondelette discrète à 4 niveaux de voix normale	33
Figure 2.11 : Exemple de décomposition en ondelette discrète à 4 niveaux de voix pathologique	34
Figure 3.1 : Principe d'identification de la parole.....	39
Figure 3.2 : Organigramme d'extraction d'attributs	40
Figure 3.3 : Méthode de classification des voix pathologiques en utilisant l'analyse EMD-DWT	41
Figure 3.4 : Schéma de la reconnaissance vocale simplifiée	42
Figure 3.5 : Classification des données en pathologiques et normales	43
Figure 3.6 : Résultats de la matrice de confusion.....	47

Liste des Tableaux

Tableau 2.1 : Familles d'ondelettes	24
Tableau 2.2 : Les paramètres d'un signal vocal normal.....	36
Tableau 2.3 : Les paramètres d'un signal vocal pathologique	36
Tableau 3.1 : Le nombre d'exemples utilisés pour les bases d'apprentissage et de test .	44
Tableau 3.2 : Matrice de Confusion	45
Tableau 3.3 : Résultats de classification obtenus grâce au SVM.....	48
Tableau 3.4 : Etude comparative de notre travail avec des autres réalisés	49

Tableau des Abréviations

Abréviations	Références
TF	Transformée de Fourier
STFT	Transformée de Fourier à Court Terme
WT	Transformation d'ondelettes (Wavelet Transformation)
TOC	Transformée par ondelette continue
TOD/DWT	Transformée par ondelette discrète
SVM	Machines à vecteurs de supports (Support Vector Machines)
FFT	Transformation de Fourier rapide (Fast Fourier Transform)
LPC	Coefficients de prédiction linéaire (Linear Predictive Coding)
SOS	Statistique d'ordre supérieur
RNA	Réseau de neurones artificiels
EMD	Décomposition en modes empiriques (Empirical Mode Decomposition)
HOS	Statistiques d'ordre supérieur (Higher Order Statistics)
MFCC	Coefficients cepstraux de fréquence mel (Mel-Frequency Cepstral Coefficients)
HNR	Rapport harmonique à bruit (Harmonic-to-Noise Ratio)
KNN	K-plus proches voisins (K-Nearest Neighbors)
VN	Vrai négatif
FN	Faux négatif
VP	Vrai positif
FP	Faux positif

Introduction Générale

La voix humaine, en tant qu'outil fondamental de communication, revêt une complexité et une richesse qui transcendent les barrières linguistiques et culturelles. Elle est capable de véhiculer une vaste palette d'émotions, de concepts et de subtilités. Au-delà de son rôle de simple vecteur d'informations, la voix reflète notre identité et notre état de santé. Malheureusement, de nombreuses personnes dans le monde sont confrontées à des problèmes vocaux qui entravent leur capacité à s'exprimer et à interagir efficacement. Ces troubles, parfois sous-estimés, ont un impact significatif sur la qualité de vie des individus. Notre mémoire s'engage dans cette réalité en se consacrant à la classification des voix pathologiques.

- Problématique :

La détection précoce et précise des pathologies vocales est cruciale pour une intervention médicale appropriée. Traditionnellement, cette détection repose sur l'expertise clinique des professionnels de la santé, qui évaluent subjectivement les altérations vocales. Cependant, cette approche présente des limites en termes d'objectivité, de reproductibilité et de précision. L'avènement des technologies d'apprentissage automatique offre une opportunité sans précédent de développer des méthodes de classification des voix pathologiques plus objectives et plus précises. Ainsi, la question centrale qui guide notre recherche est la suivante : Comment les avancées dans le domaine de l'apprentissage automatique peuvent-elles être utilisées pour améliorer la détection et la classification des voix pathologiques ?

- Méthodologie :

Notre mémoire est construite autour d'une méthodologie rigoureuse qui repose sur une compréhension approfondie du signal vocal, l'extraction de caractéristiques vocales pertinentes, et l'application de modèles de classification performants. Les chapitres suivants détailleront cette méthodologie, qui constitue le socle de notre travail de recherche.

- Objectifs :

L'objectif majeur de notre mémoire est de développer une approche précise et objective pour la classification des voix pathologiques. En exploitant les avancées technologiques en matière d'apprentissage automatique, nous aspirons à fournir un outil de diagnostic plus fiable et reproductible pour les professionnels de la santé. Notre ambition est d'améliorer la qualité de vie des personnes souffrant de troubles vocaux en permettant une détection précoce et une intervention ciblée.

- Contributions attendues :

L'objectif central de notre mémoire est d'améliorer considérablement les capacités de diagnostic et de traitement des troubles vocaux. Nous visons à ouvrir de nouvelles perspectives pour les professionnels de la santé en leur fournissant des outils plus précis et objectifs pour évaluer et gérer ces affections. Parallèlement, nous cherchons à sensibiliser à l'importance vitale de la santé vocale et à souligner le besoin pressant d'investir dans des méthodes de diagnostic plus accessibles et précises. Notre engagement se concentre spécifiquement sur l'efficacité de la classification des voix pathologiques en tant qu'approche prometteuse pour améliorer la qualité des soins et la qualité de vie des individus concernés.

- Plan de la mémoire :

Notre travail se divise en trois chapitres principaux, chacun contribuant à répondre à la problématique posée :

Le premier chapitre s'attache à établir les bases fondamentales du signal vocal. Il explore l'anatomie de la production de la parole, les caractéristiques acoustiques des sons et les troubles vocaux courants. Ce chapitre jettera les fondements de notre compréhension de la voix humaine et de ses altérations.

Le deuxième chapitre introduit notre méthodologie de classification des voix pathologiques. Nous détaillerons l'extraction de caractéristiques vocales à l'aide de la transformation en ondelettes et l'application du modèle de classification SVM. Ce chapitre présentera notre approche méthodologique et ses principes sous-jacents.

Le troisième chapitre sera consacré à la présentation des résultats de notre recherche. Nous analyserons les performances de notre modèle de classification en utilisant la matrice de confusion comme outil clé. De plus, nous situerons nos résultats dans le contexte de la recherche existante en les comparant à d'autres travaux similaires.

En conclusion, notre mémoire s'inscrit dans une démarche visant à améliorer la détection et la classification des voix pathologiques. Nous espérons que notre travail contribuera

Chapitre 1 :
Généralités sur le
signal vocal

1. Introduction

La production de la parole est une action motrice complexe impliquant un grand nombre de muscles, de variables physiologiques et de contrôle neuronal, impliquant différentes aires corticales et sous-corticales. [1]

En tant que moyen de communication essentiel pour les êtres humains, le signal vocal est essentiel à notre quotidien. La voix humaine est un instrument complexe et riche en informations capable de transmettre une variété d'émotions, de pensées et de nuances subtiles.

Dans ce chapitre introductif, nous explorerons les fondamentaux du signal vocal, en nous penchant sur divers aspects tels que l'anatomie de la production de la parole, la nature variée des sons sonores et non sonores produits, ainsi que les troubles qui peuvent affecter la qualité et la clarté de la voix. De plus, nous aborderons les paramètres essentiels qui permettent de caractériser et de classifier les voix pathologiques, jetant ainsi les bases de notre exploration approfondie dans le domaine de la classification des voix pathologiques.

2. Système de production de la parole

La parole est l'un des moyens de communication les plus importants, elle offre un moyen d'exprimer des instincts, des sentiments et des pensées. Les sons de la parole se produisent par le mouvement de nombreux muscles et organes de phonation, en quelque sorte est la collaboration entre ces organes. [2]

La phonation est un acte volontaire, résultat de transformation de souffle d'air expirée le long de la conduite vocale. Après expiration, l'air des poumons passe par deux phases, la phase de modification en vibration au niveau de la partie glottique (Le larynx), et la phase de filtration au niveau de la partie supra-glottique. [3]

2.1 Partie sub-glottique (l'appareil respiratoire)

L'appareil respiratoire est la source d'énergie de la voix, la partie sub-glottique comprend trois organes : le diaphragme, les poumons et la trachée (Figure 1.1). [3][4]

2.1.1 Le diaphragme

Le diaphragme, qui se trouve sous les poumons, sert à séparer l'estomac de l'abdomen. Il est crucial pour l'acte de respirer et d'exhaler l'air. Il sert de moteur principal. La pression négative est créée dans le thorax lorsque le diaphragme se contracte, permettant à l'air d'entrer dans le système respiratoire à travers le tracé. Le processus est alors inversé lorsque le diaphragme se libère, fournissant l'air au résonateur.

2.1.2 Les poumons

Les principaux organes respiratoires du corps humain sont les poumons, qui se trouvent de chaque côté du médiastin. En fournissant l'air nécessaire à la soufflerie pour faire vibrer les cordes vocales, ils jouent un rôle essentiel dans la production de la parole. Les poumons produisent ainsi un flux d'air contrôlé, qui est essentiel pour produire des sons vocaux pendant la parole.

2.1.3 La trachée

Juste en dessous du larynx se trouve un tube cartilagineux appelé trachée. Elle se divise en deux bronches principales, avec une pour chaque poumon. Elle sert de passage d'air entre les poumons et le larynx, qui contient les cordes vocales. Lorsque nous parlons ou produisons des sons, l'air des poumons passe à travers la trachée et arrive au larynx.

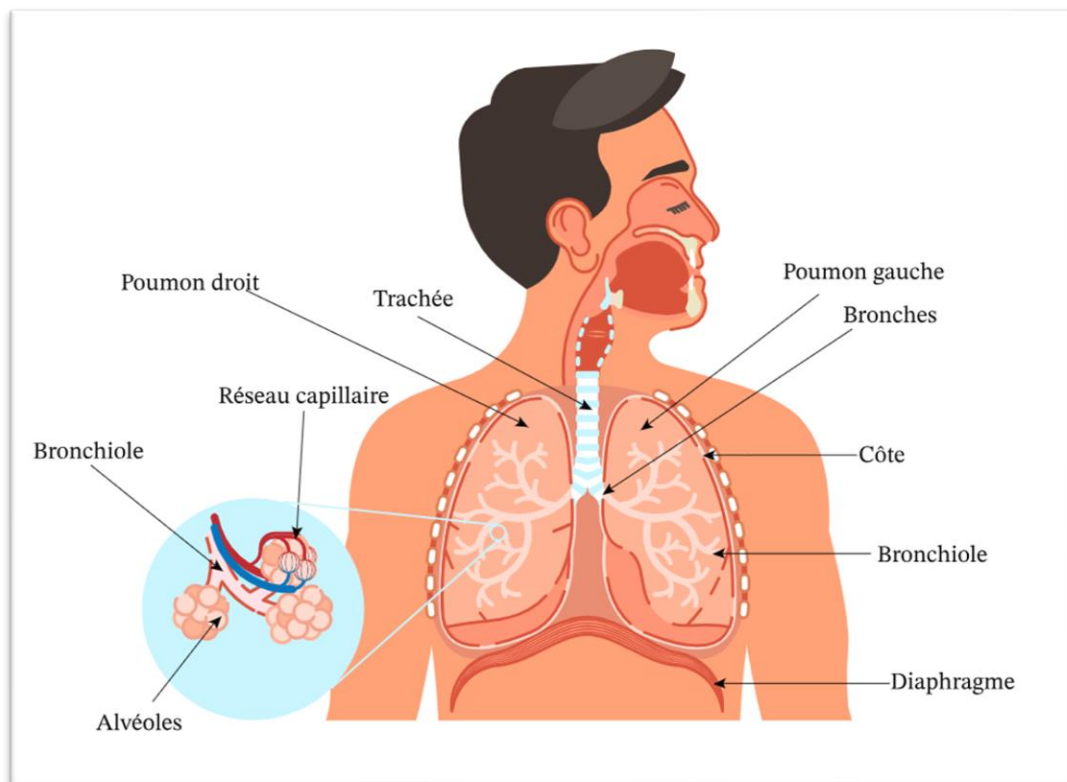


Figure 1.1 : Schéma de l'appareil respiratoire.

2.2 Partie glottique (Le larynx)

Structure cartilagineuse comprend les cartilages : trois cartilages impaires majeurs (cricoïde, thyroïde et épiglote) et trois paires de cartilages (aryténoïde, cornicule et cunéiforme), qui sont des muscles et des articulations qui assurent la mobilité de ces parties. L'organe essentiel pour transformer l'air expiré par les poumons en sons voisés ou non voisés sont les cordes vocales en nombre de deux.

Le larynx aussi joue un rôle de protection des voix respiratoires, il empêche le passage des aliments par l'épiglotte qui se ferme lors de déglutition (Figure 1.2). [4][5]

2.2.1 Les cartilages du larynx

2.2.1.1 Cartilages impairs

Les trois cartilages impairs sont le cartilage épiglottes, thyroïde et cricoïde.

2.2.1.1.1 Cartilage thyroïdien

Le cartilage thyroïdien est le plus grand des cartilages laryngés. Il constitue les segments antérieures et latéraux du larynx et n'a pas de segment postérieur. La proéminence laryngée, également connue sous le nom de "pomme d'Adam", est une projection antérieure en forme de V formée par les larges moitiés plates droite et gauche (lamines) du cartilage qui fusionnent antérieurement sur la ligne médiane.

2.2.1.1.2 Cartilage cricoïde

Directement sous le cartilage thyroïdien, le cartilage cricoïde est un cartilage hyalin beaucoup plus petit en forme d'anneau. Il constitue la face inférieure du larynx et est connecté à la trachée par la face inférieure du larynx. Il s'agit du seul anneau de cartilage complet qui entoure les voies respiratoires.

2.2.1.1.3 L'épiglotte

L'épiglotte est une plaque en forme de feuille de cartilage élastique qui marque l'entrée du larynx. Sa « tige » est attachée à l'arrière de l'aspect antérieur du cartilage thyroïdien. Pendant l'ingestion, l'épiglotte s'aplatit et se déplace à l'arrière pour fermer le larynx et prévenir l'aspiration.

2.2.1.2 Cartilages pairs

Il existe trois cartilages : l'aryténoïde, le corniculaire et le cunéiforme. Ils sont situés bilatéralement dans le larynx.

2.2.1.2.1 Les cartilages aryténoïdes

Les cartilages aryténoïdes sont des structures en forme de pyramide qui se situent sur le cartilage cricoïde. Ils se composent d'un sommet, d'une base, de trois côtés et de deux processus, et fournissent un point d'attachement pour diverses structures clés dans le larynx.

2.2.1.2.2 Les cartilages cornéliens

Les cartilages corniculés sont des structures cartilagineuses mineures. Ils articulent avec les apex des cartilages aryténoïdes.

2.2.1.2.3 Les cartilages cunéiformes

Les cartilages cunéiformes sont situés à l'intérieur des plis aryépiglottiques. Ils n'ont pas d'attachement direct, mais agissent pour renforcer les pattes.

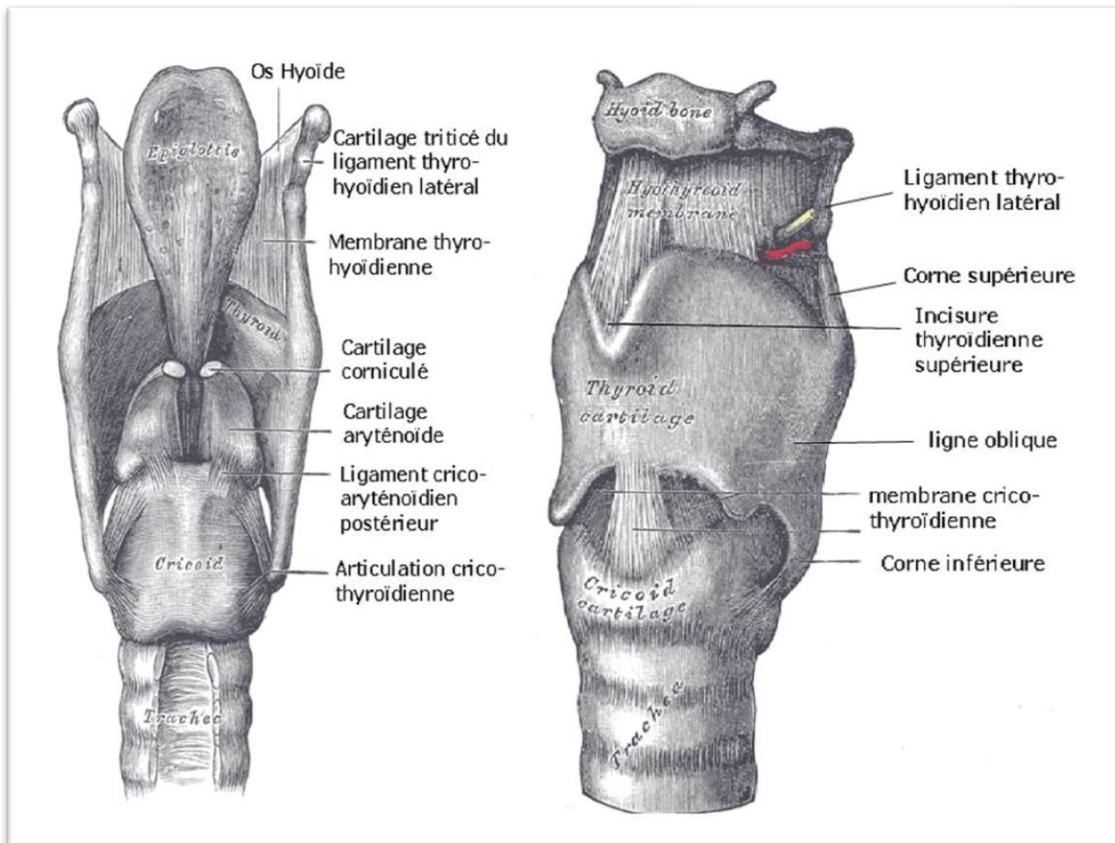


Figure 1.2 : Anatomie du Larynx.

2.2.2 Les muscles du larynx

Les muscles du larynx jouent un rôle essentiel dans la production de la parole et la phonation. Voici quelques points importants à connaître :

- Les muscles intrinsèques du larynx sont responsables de la tension et de la position des cordes vocales. Ils permettent d'ajuster la hauteur et la qualité du son produit lors de la parole.
- Les muscles extrinsèques du larynx sont attachés à d'autres structures adjacentes et contribuent au mouvement global du larynx. Par exemple, les muscles éleveurs aident à élever le larynx lors de la déglutition pour protéger les voies respiratoires.
- Les muscles adducteurs du larynx rapprochent les cordes vocales pour permettre la phonation. Ils sont activés lorsque nous parlons, chantons ou produisons des sons.
- Les muscles abducteurs du larynx écartent les cordes vocales, permettant ainsi le passage de l'air sans produire de son. Cela se produit pendant la respiration normale lorsque nous n'émettons pas de sons.

- Les muscles du larynx fonctionnent en coordination avec d'autres parties du système vocal, comme les muscles respiratoires et les muscles articulaires, pour produire un discours clair et intelligible.

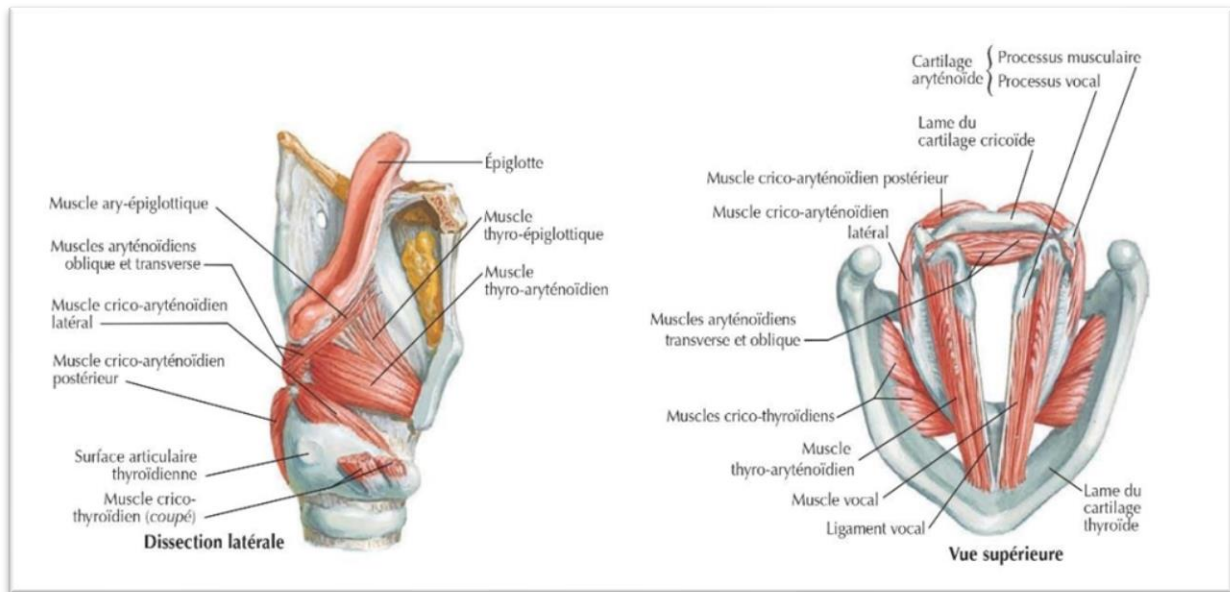


Figure 1.3 : Schéma des muscles du Larynx.

2.2.3 Les cordes vocales

Les cordes vocales, également appelées plis vocaux, sont des structures essentielles du système vocal. Elles sont situées dans le larynx. Elles sont composées principalement de tissu conjonctif élastique et de muscles. Chaque corde vocale est formée d'une fine couche de muqueuse qui la recouvre et d'un noyau de ligament vocal. Elles sont suspendues entre les cartilages aryénoïdes à l'arrière du larynx et sont disposées de manière parallèle.

Lorsque l'air expiré des poumons traverse les cordes vocales, elles se ferment partiellement, créant une vibration. Cela produit un son fondamental qui est ensuite modulé par les cavités supérieures du tractus vocal pour produire des sons spécifiques à la parole. Les muscles laryngés aident à contrôler la tension des cordes vocales pour produire une variété de sons (Figure 1.4).

Lors de la respiration ou de la phonation, ils contrôlent le passage d'air et prennent la forme d'une glotte.

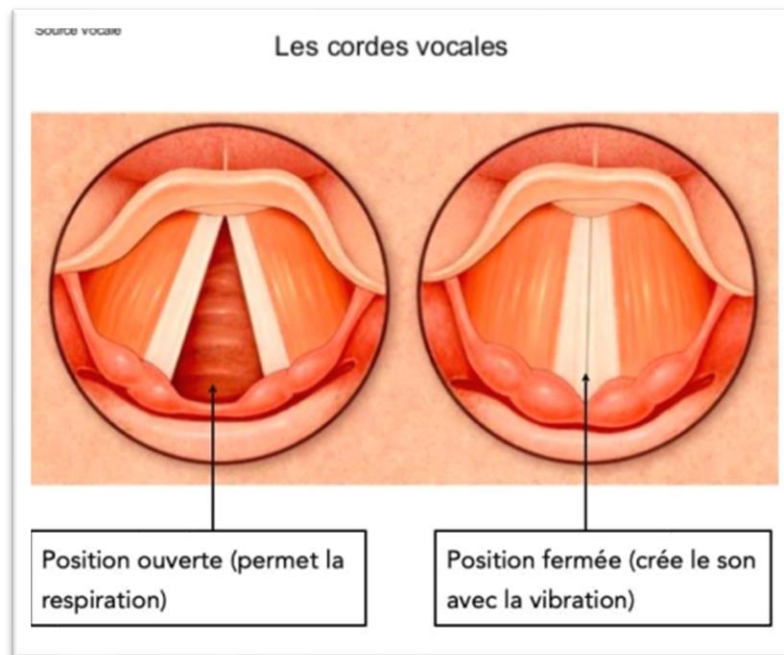


Figure 1.4 : Les cordes vocales.

2.3 Partie supra-glottique

Les cavités supra-laryngées font référence aux structures situées au-dessus du larynx, comprenant la cavité buccale, le pharynx et les cavités nasales et sinusales. Elles jouent un rôle essentiel dans la production du timbre vocal en modulant le son initial généré par les cordes vocales. Le canal buccal, formé par les cavités buccales et pharyngées, présente une géométrie et un volume qui varient considérablement, principalement en raison de la grande mobilité requise pour le langage. [5][6]

3. La production de la parole

La phonation correspond à l'ensemble des mécanismes permettant l'apparition et l'ajustement d'une vibration au niveau du bord des cordes vocales. [7]

Le processus de production de la parole est un mécanisme très complexe, qui repose sur l'interaction entre le système nerveux et le système physiologique. La parole commence par l'activité neuronale. Le système nerveux joue un rôle important, il est donc responsable de la génération et du décodage des signaux de parole. Après les pensées et la volonté de parler, le cerveau guide les opérations liées à l'activation des organes vocaux. La fonction de ces organes est de nature physiologique. [8][9]

Ce processus nécessite la mise en ordre de plusieurs éléments physiologiques et anatomiques pour produire une variété de sons qui constituent la base de la langue parlée. Les sons vocaux et non vocaux, deux catégories distinctes qui contribuent à la richesse et à la diversité de la communication humaine, sont au cœur de cette entreprise.

3.1 Sons vocaux

Les sons vocaux, également appelés sons sonores, sont créés lorsque les cordes vocales vibrent en réponse à l'air expiré par les poumons. Les ondes sonores produites par ces vibrations des cordes vocales produisent des sons distinctifs. La parole et la phonation dépendent des sons vocaux. Les sons vocaux ont les caractéristiques suivantes :

- Pendant la création de sons vocaux, les cordes vocales vibrent, produisant une série d'impulsions acoustiques régulières.
- Les phonèmes (unités de son) qui composent les mots et les langues parlées sont créés par les sons vocaux. Les sons vocaux comprennent les voyelles et les consonnes vocales telles que "m", "n" et "l".
- Amplitude acoustique : Les sons vocaux ont généralement un spectre de fréquence plus complexe et une amplitude acoustique plus élevée interaction humaine.

3.2 Sons non vocaux

Les sons non vocaux, également appelés sons non sonores, sont générés sans que les cordes vocales ne vibrent. D'autres parties de l'appareil vocal, comme la bouche ou le nez, les produisent en modifiant la forme de ces cavités pour produire des résonances particulières. Les sons non vocaux ont les caractéristiques suivantes :

- Les cordes vocales non vibrantes ne vibrent pas lorsqu'on produit des sons non vocaux.
- Sons consonantiques : Les consonnes non vocales (telles que "p", "t", "k") sont souvent créées à l'aide de sons non vocaux, qui sont caractérisés par des bruits d'explosion ou de turbulence.
- Amplitude acoustique réduite : Les sons non vocaux ont généralement un spectre de fréquence moins complexe et une amplitude acoustique plus faible que les sons vocaux.

La Figure 1.5 montre la différence entre un signal sonore et non sonore.

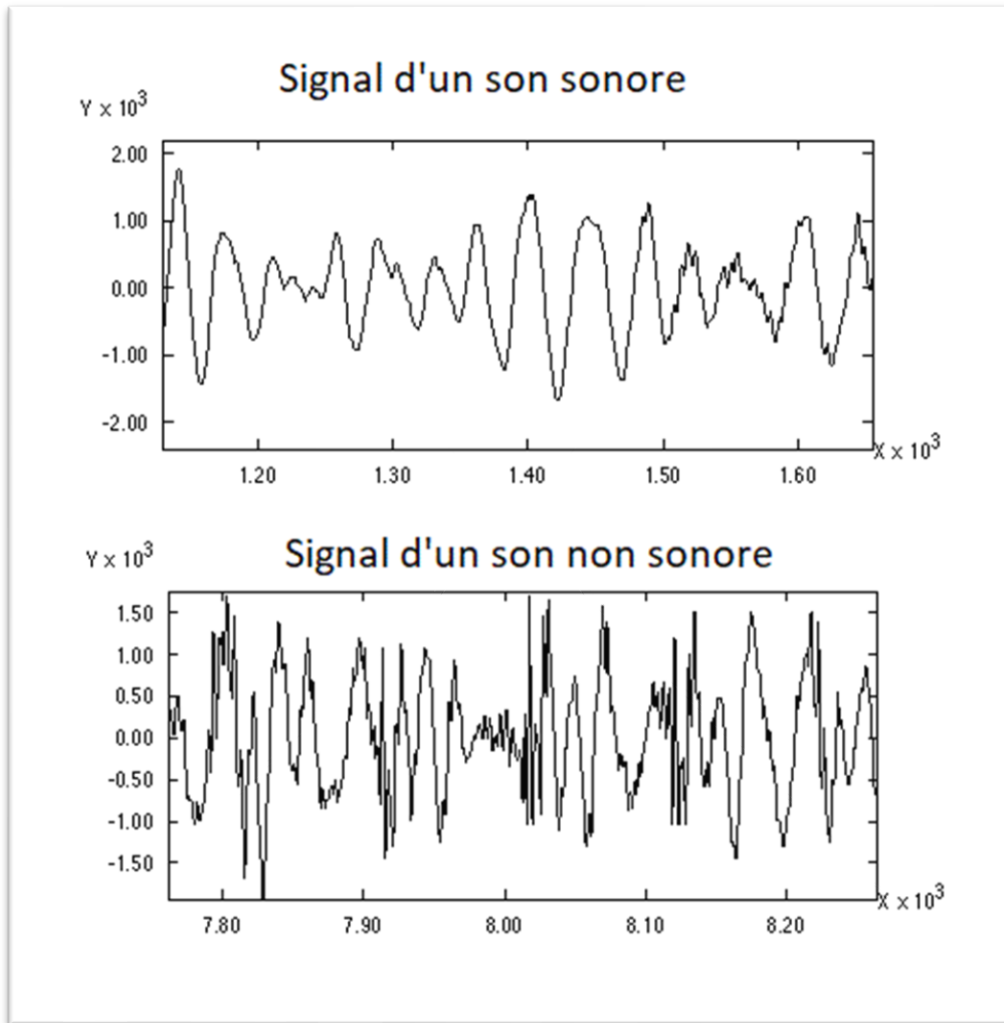


Figure 1.5 : signal sonore et non sonore

4. Troubles de la voix

4.1 Définitions

Les troubles de la voix, également connus sous le nom de dysphonie, font référence à toute altération de la production ou de la qualité de la voix. La respiration, la phonation (production de sons) ou la résonance peuvent être affectées par ces troubles. Ils peuvent être provoqués par une variété de choses, telles que des problèmes anatomiques, fonctionnels, organiques ou psychologiques.

Les troubles de la voix peuvent être organiques, fonctionnels ou combinés. Ils peuvent affecter n'importe qui, mais ils sont plus fréquents chez les personnes qui utilisent fréquemment leur voix, comme les enseignants, les agents de centre d'appel, les chanteurs et les fumeurs. Ils peuvent également être causés par des infections respiratoires, des événements stressants, des conditions à long terme qui affectent la respiration (comme l'asthme ou les maladies pulmonaires chroniques), des accidents vasculaires cérébraux ou des maladies neurologiques comme la maladie de Parkinson.

De temps en temps, les enfants peuvent présenter des perturbations vocales, que ce soit en raison de conditions innées ou de leur manière d'exprimer leur voix. Symptômes généraux

L'adulte ou l'enfant présente un ou plusieurs de ces symptômes : [10]

- Sensation de fatigue vocale,
- Besoin de faire des efforts pour parler,
- Manque de souffle,
- Sensation de boule dans la gorge,
- Douleurs,
- Présence de sécrétions,
- Raclement fréquent de la gorge,
- Toute autre sensation inconfortable.

4.2 Les pathologies

4.2.1 Polypes

Les polypes des cordes vocales sont fréquemment causés par un traumatisme vocal aigu, tel que crier lors d'un match de football, qui affectent souvent un seul cordon vocal. De nombreux facteurs, y compris le reflux gastrique, l'hypothyroïdie et la consommation chronique d'irritants (tels que la fumée de cigarette ou les fumées industrielles), peuvent conduire au développement de polypes. Ces facteurs entraînent souvent des polypes qui affectent les deux cordes vocales. Par rapport aux nœuds, les polypes sont généralement plus grands et plus en forme de pompe. Les adultes sont plus susceptibles de développer des polypes.

Les deux images de la Figure 1.6 montrent un polype situé sur une corde vocale, illustrant les positions d'ouverture et de fermeture des cordes vocales.

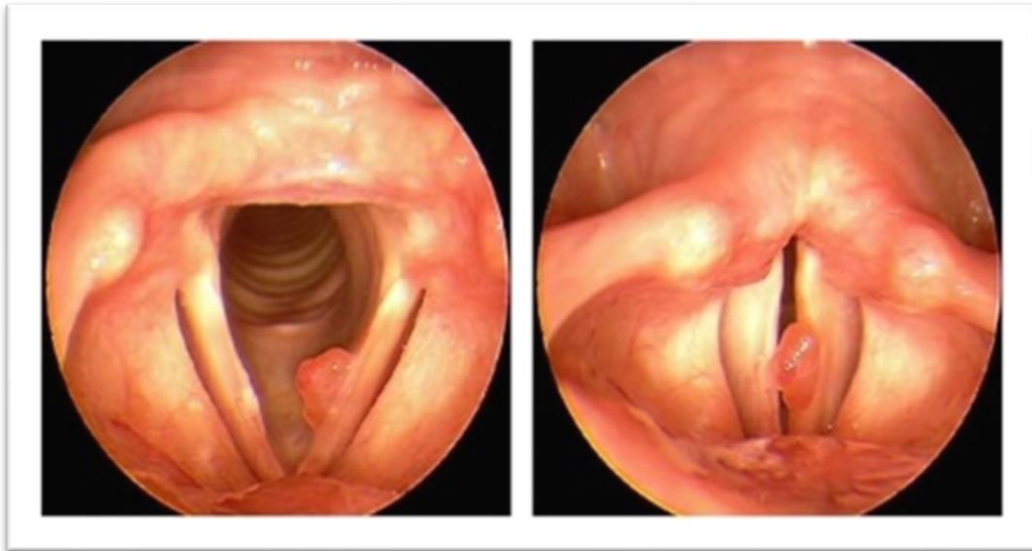


Figure 1.6 : Polype de la corde vocale

4.2.2 Nodules

Les nodules est l'une des causes potentielles de l'hyperkinèse dysphonique. Ils se rapportent à un épaississement localisé du bord libre du cordon vocal qui peut résulter d'une utilisation excessive ou d'un abus vocal, tels que des cris, des chants ou des conversations excessives. Cette opération est généralement bilatérale (se produit des deux côtés). Les cordes vocales peuvent ne pas se fermer complètement lorsque des nœuds sont présents, ce qui peut causer une voix étonnée ou une respiration.

La Figure 1.7 présente les cordes vocales avec des nodules, illustrant les positions d'ouverture (A) et de fermeture (B).

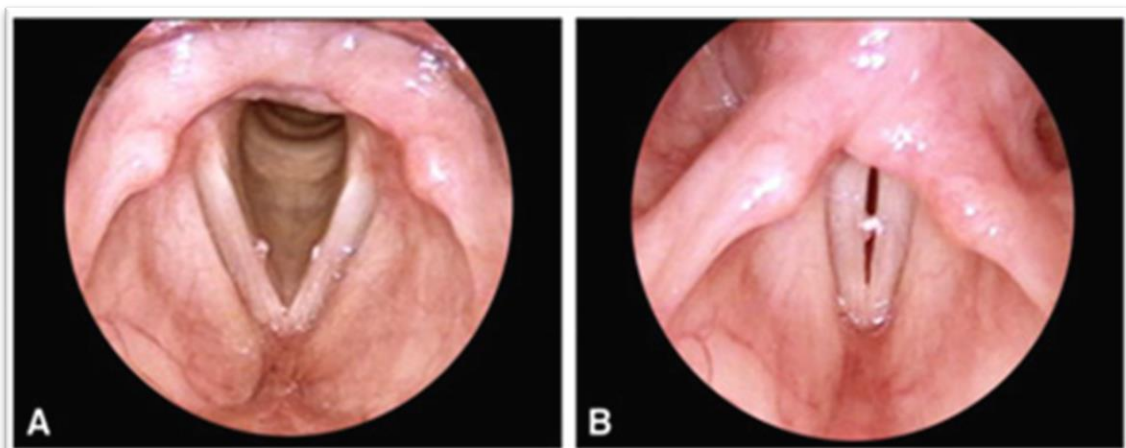


Figure 1.7 : Nodules sur la corde vocale

4.2.3 Kystes

Les kystes des cordes vocales apparaissent dans la couche superficielle à n'importe quelle partie de celles-ci. Au moment où le kyste croît, il exerce une pression sur l'épithélium et les ligaments des cordes vocales sans affectation des couches adjacentes. Les kystes induisent une augmentation de la masse et un durcissement de la couverture. Comme conséquences des kystes, la glotte ne se ferme pas complètement pendant la phonation [11].

Les kystes vocaux peuvent causer des symptômes similaires à ceux des nodules vocaux, tels qu'une voix enrouée, une altération de la qualité vocale, une fatigue vocale rapide et des difficultés à atteindre des notes élevées. Cependant, contrairement aux nodules, les kystes vocaux peuvent également provoquer une voix instable, une voix sifflante ou des fluctuations de la voix.

4.2.4 Tumeur cancérigène

Une tumeur vocale cancéreuse est une croissance maligne qui se développe dans les tissus des cordes vocales ou des structures voisines, comme le larynx. Le cancer des voies aérodigestives supérieures, qui est le type le plus courant, est généralement classé comme un carcinome squameux (Figure 1.8).

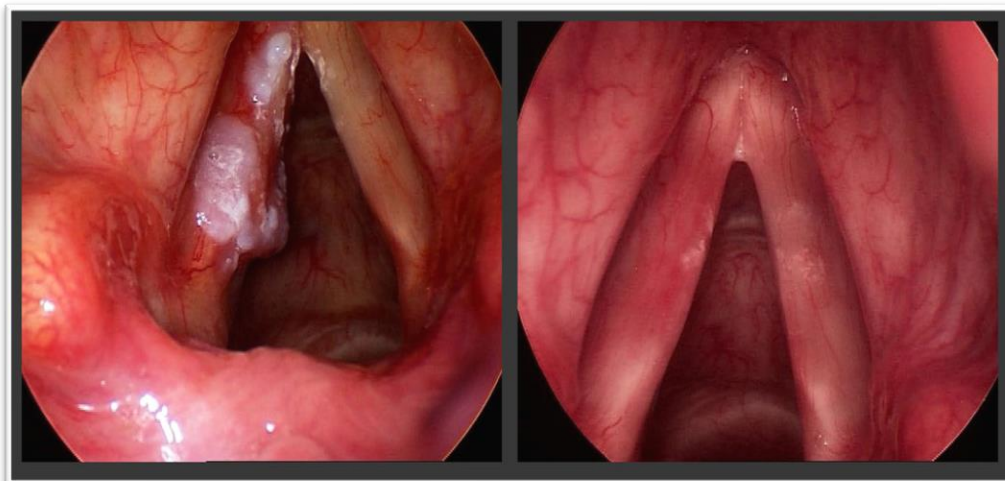


Figure 1.8 : Tumeur cancéreuse sur les cordes vocales

4.2.5 Reflux laryngite

Lorsque l'acide de l'estomac remonte dans le larynx, il y a une inflammation de la muqueuse laryngée appelée reflux laryngite. Le dysfonctionnement du sphincter, qui sépare l'estomac de l'œsophage, permet aux acides de remonter.

4.2.6 Œdème de Reinke

L'œdème de Reinke entraîne un épaissement anormal des cordes vocales, ce qui change considérablement la qualité et la tonalité de la voix. La voix de cette personne peut être grave, rauque, enrouée ou voilée. La voix peut également avoir une qualité fumée et sembler plus basse (Figure 1.9).

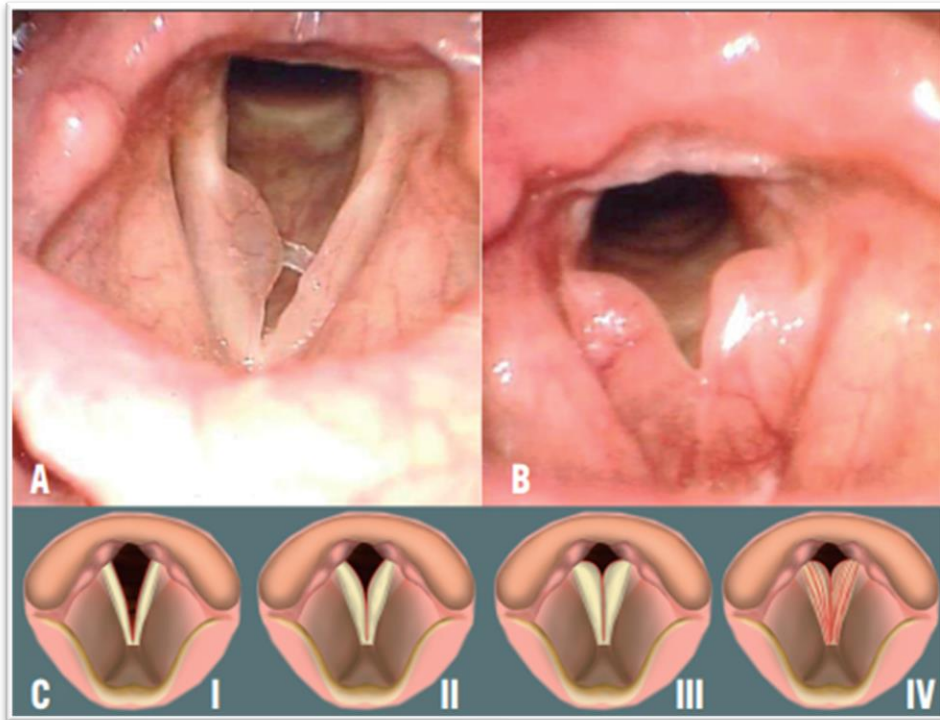


Figure 1.9 : Œdème de Reinke.

4.2.7 Paralysie des cordes vocales

La paralysie des cordes vocales se produit lorsque les muscles responsables du mouvement des cordes vocales ne fonctionnent pas correctement, ce qui entraîne une immobilité partielle ou totale d'une ou des deux cordes vocales. Cela peut modifier considérablement la voix, la respiration et la déglutition (Figure 1.10).

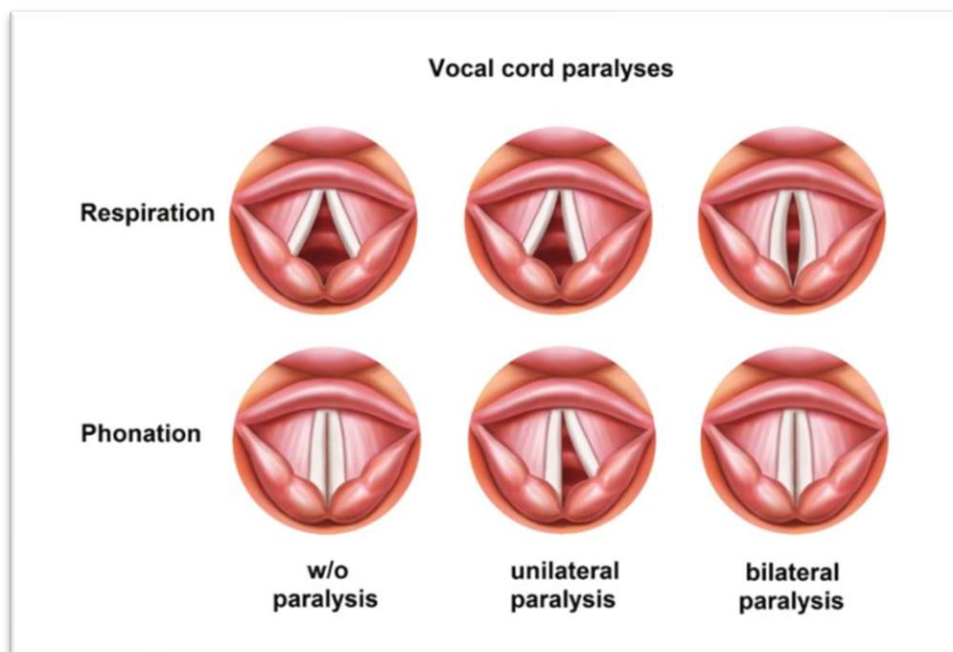


Figure 1.10 : Paralysie des cordes vocales.

4.2.8 Traumatismes chirurgicaux

Les dommages ou les blessures causés par une intervention chirurgicale au niveau des cordes vocales ou des structures voisines sont appelés traumatismes chirurgicaux vocaux. Des altérations de la fonction vocale et des symptômes tels qu'une voix enrouée, une voix faible, des difficultés à produire certains sons ou à atteindre des notes élevées, une fatigue vocale et des problèmes de déglutition peuvent être causés par ces traumatismes.

4.2.9 Hypercinésie dysphonique de la voix (hyperkinétique)

L'hypercinésie dysphonique de la voix, également appelée hyperkinétique, est un trouble de la voix caractérisé par une hyperactivité excessive des muscles vocaux et des structures environnantes, en résultent des mouvements vocaux involontaires, excessifs et inappropriés.

Les symptômes de l'hypercinésie dysphonique de la voix peuvent varier en fonction de la condition spécifique, mais ils incluent généralement une voix instable, saccadée, tendue ou tremblante, des interruptions de la voix, des difficultés à contrôler le volume et la tonalité et des variations incohérentes de la qualité vocale.

4.2.10 Hypokinésie dysphonique de la voix (hypokinétique)

L'hypokinésie dysphonique de la voix est un terme utilisé pour décrire un type de trouble de la voix qui résulte d'une hypokinésie, c'est-à-dire d'une diminution du mouvement et de l'activité musculaire des cordes vocales et des structures environnantes.

Les personnes atteintes de la maladie de Parkinson présentent une perte de contrôle et de coordination des muscles vocaux. Une voix faible, monotone et sans modulation sont des symptômes de l'hypokinésie dysphonique de la voix. De plus, les personnes atteintes peuvent avoir des difficultés à démarrer et à maintenir leur voix, ainsi qu'une perte de clarté et d'intelligibilité vocale.

5. Méthodes d'identification des troubles de la voix

Les troubles de la voix peuvent être identifiés de manière subjective et objective. Voici un aperçu des méthodes utilisées dans chaque approche :

5.1 Identification subjective

L'observation et l'évaluation de la voix par des professionnels de la santé tels que des phoniâtres, des orthophonistes ou des médecins spécialisés sont au cœur de cette méthode. En général, elle nécessite une évaluation clinique qui repose sur les interactions verbales avec le patient et une écoute attentive de sa voix.

Il existe plusieurs échelles d'évaluations subjectives, comme exemple, nous avons choisi de décrire le GRBAS D'Hirano (1981) : [12]

Le système GRBAS d'Hirano, créé par le Dr Hiroshi Hirano, un phoniatre japonais renommé, est une méthode d'évaluation subjective largement utilisée pour évaluer la qualité vocale d'une personne, en particulier pour identifier et quantifier les troubles de la voix. Le terme « GRBAS » représente les cinq paramètres évalués dans cette méthode :

G - Grade (Gravité) : Le degré de sévérité du trouble vocal, basé sur des échelles allant de 0 à 3 ou de 1 à 4. Ce paramètre évalue la gravité de la perturbation vocale globale, en tenant compte de facteurs tels que la rugosité, la rauqueur, l'enrouement, etc.

R - Roughness (Rugosité) : L'évaluation de l'irrégularité ou de la rugosité de la voix. Ce paramètre mesure la qualité de la voix en termes de douceur et de régularité tonale.

B - Breathiness (Souffle) : L'évaluation de la présence d'air audible ou de fuite d'air lors de la phonation. Cela peut être dû à une fermeture incomplète des cordes vocales.

A - Asthenia (Faiblesse) : L'évaluation de la force de la voix et de sa résistance. Ce paramètre évalue la puissance globale de la voix et son niveau de résistance à l'effort.

S - Strain (Effort) : L'évaluation de la tension vocale excessive et de l'effort nécessaire pour la production vocale. Ce paramètre mesure la pression sous-glottique nécessaire pour produire la voix.

Chaque paramètre est évalué sur une échelle discrète, généralement de 0 à 3, ou de 1 à 4, où un score de 0 représente un niveau normal et des scores plus élevés indiquent une perturbation vocale plus grave. Un profil vocal global peut être créé en combinant les scores attribués à chaque paramètre, ce qui peut aider à décrire et à quantifier les troubles vocaux d'une personne.

La communauté médicale et de la rééducation vocale utilise fréquemment le système GRBAS d'Hirano pour évaluer la qualité vocale, surveiller les progrès après des interventions thérapeutiques et guider les décisions de traitement. Cependant, il est important de garder à l'esprit que cette méthode est subjective et peut différer d'un évaluateur à l'autre.

5.2 Identification objective

La deuxième méthode suggérée comme une option à l'évaluation basée sur la perception pour compenser ses désavantages et lacunes est l'analyse objective. Les méthodes objectives comprennent la découverte de mesures acoustiques capables de quantifier la gravité de la pathologie des voix. Il s'agit principalement de techniques d'analyse de la voix acoustique.

L'analyse acoustique est une méthode non invasive d'exploration de la voix qui fournit des mesures quantitatives objectives sur un échantillon vocal. La hauteur (ou fréquence fondamentale (F0)), la sonie (ou intensité) et le timbre sont les éléments qui donnent à la voix sa couleur complète. D'autres paramètres, tels que le débit phonatoire et l'articulation, reflètent les perturbations de la parole.

6. Paramètres de la voix

Pour décrire et mesurer une voix, les paramètres de la voix correspondent aux différentes caractéristiques et valeurs accordées à celle-ci, voici quelques paramètres du signal vocale :

6.1 Fréquence Fondamentale (F0)

La fréquence à laquelle les cordes vocales vibrent pendant la production de la voix est connue sous le nom de fréquence fondamentale. Il est exprimé en Hertz (Hz). Une F0 élevée indique une voix aiguë et une F0 basse indique une voix grave. La dysphonie spasmodique, où la F0 peut être instable, est l'un des troubles qui peuvent entraîner des variations anormales de la F0.

La fréquence fondamentale (F0) varie en fonction du genre et de l'âge, voici des informations générales sur les variations de F0 pour différentes catégories :

Hommes : En moyenne, la F0 des voix masculines adultes se situe généralement entre environ 85 Hz et 180 Hz. Les voix masculines tendent à avoir des fréquences plus basses en raison de cordes vocales plus longues et plus épaisses.

Femmes : La F0 des voix féminines adultes a tendance à être plus élevée que celle des hommes. En moyenne, elle se situe généralement entre environ 165 Hz et 255 Hz. Les cordes vocales plus courtes et plus fines des femmes contribuent à ces fréquences plus élevées.

Enfants : Les enfants ont des cordes vocales encore plus petites et plus fines que les adultes. Par conséquent, la F0 des voix d'enfants est généralement plus élevée que celle des adultes. La F0 moyenne pour les enfants peut varier en fonction de leur âge, mais elle est souvent supérieure à 250 Hz.

6.2 Pitch

La perception subjective de la hauteur d'un son musical ou vocal est appelée pitch. Il représente le trait qui permet de faire la distinction entre les sons graves et les sons aigus. La fréquence fondamentale (F0) du signal sonore détermine principalement le pitch, une propriété auditive complexe, en général, plus la F0 est élevée, plus le pitch perçu sera haut, et vice versa.

6.3 Hauteur

Le signal vocal est fortement influencé par la hauteur de la voix. La fréquence fondamentale (F0) du signal vocal est généralement liée à la hauteur, qui représente le nombre de cycles de vibration des cordes vocales par unité de temps.

La hauteur d'une voix est liée directement à la F0 du signal vocal. Une F0 élevée indique une voix aiguë et une F0 basse indique une voix grave. Les variations de hauteur peuvent être utilisées pour évaluer la qualité de la voix et trouver des anomalies potentielles.

6.4 Intensité

L'intensité est une mesure de l'énergie acoustique du signal vocal et est liée à l'amplitude du son. Il est exprimé en décibels (dB). Les voix pathologiques peuvent avoir une intensité réduite ou excessive, ce qui peut indiquer des problèmes tels que des nodules vocaux ou un œdème des cordes vocales.

6.5 Timbre

Le timbre vocal, également connu sous le nom de qualité tonale ou de couleur vocale, est une caractéristique perçue de la voix humaine qui permet de différencier différentes sources sonores. La distribution spectrale des harmoniques et des formants dans le signal vocal le détermine. Les harmoniques, les formants, la résonance du tractus vocal, la vibration des cordes vocales et le filtrage des cavités buccale et nasale influencent le timbre.

6.6 Formants

Les pics de fréquence du spectre acoustique du signal vocal qui découlent de la résonance dans le tractus vocal sont appelés formants. Ils sont généralement classés par fréquence. Les conditions telles que la paralysie des cordes vocales peuvent provoquer des variations anormales des formants.

6.7 Harmoniques

Les harmoniques sont des composantes fréquentielles multiples de la fréquence fondamentale. L'analyse des harmoniques peut fournir des informations sur le timbre et la qualité de la voix. Des anomalies dans les harmoniques peuvent être associées à des troubles comme le choc phonique.

6.8 Jitter et Shimmer

Le jitter se réfère aux variations de la période fondamentale entre cycles vocaux consécutifs, reflétant l'instabilité de la F0. Le shimmer se rapporte aux variations d'amplitude entre cycles vocaux. Des valeurs anormalement élevées de jitter et shimmer peuvent indiquer des troubles vocaux tels que la dysphonie organique.

7. Conclusion

En conclusion, ce premier chapitre a jeté les bases de notre exploration sur la classification des voix pathologiques en examinant les généralités du signal vocal.

Nous avons découvert que la production de la parole humaine est un processus complexe qui nécessite une coordination précise de diverses structures anatomiques. Les sons générés pendant la parole peuvent être classés en sons sonores et non sonores, chacun ayant des caractéristiques acoustiques distinctes.

Nous avons également abordé les troubles de la voix, soulignant l'importance de comprendre les altérations qui peuvent affecter la qualité vocale.

Nous avons aussi mentionné les méthodes d'identification des troubles de la voix et les paramètres de la voix qui ont été identifiés comme des éléments clés pour l'analyse et la classification des voix pathologiques.

En comprenant ces concepts fondamentaux, nous serons mieux préparés à aborder les méthodes de classification des voix pathologiques présentées dans les chapitres suivants.

Chapitre 2 :
La méthodologie de la
classification des voix
pathologiques

1. Introduction

Ce chapitre se concentre sur la méthodologie utilisée pour classer les voix pathologiques et normales, un élément crucial pour le diagnostic vocal.

L'approche en plusieurs étapes débute par l'analyse des caractéristiques vocales, en utilisant des transformations telles que la transformation en ondelettes discrète, qui saisit les variations temporelles et fréquentielles, essentielles pour identifier les anomalies subtiles. Après l'extraction des caractéristiques, le chapitre se tourne vers la classification, en mettant en avant le système de classification SVM. L'attention se porte particulièrement sur l'approche linéaire du SVM, avec une exploration détaillée de ses principes et de son application à la classification des voix pathologiques.

En somme, ce chapitre présente une méthodologie complète intégrant des techniques de traitement du signal et de classification, offrant un cadre solide pour une identification précise des voix pathologiques et des diagnostics vocaux plus précis.

2. Transformée de Fourier (TF)

La technique mathématique essentielle pour analyser les signaux et les systèmes est la transformée de Fourier. Elle joue un rôle fondamental en décomposant une fonction périodique ou non périodique en une combinaison soigneusement ajustée de fonctions sinus et cosinus à des fréquences diverses. Cette décomposition permet une inspection minutieuse de la composition fréquentielle d'un signal, offrant ainsi une perspective dans le domaine fréquentiel qui complète son observation dans le domaine temporel.

La définition de la TF est donnée par la formule suivante :

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-2j\pi ft} dt$$

Où :

- $X(f)$ représente la transformée de Fourier du signal $x(t)$ à la fréquence f .
- $x(t)$ est le signal continu en fonction du temps t .
- f est la fréquence en hertz à laquelle la transformée de Fourier est calculée.

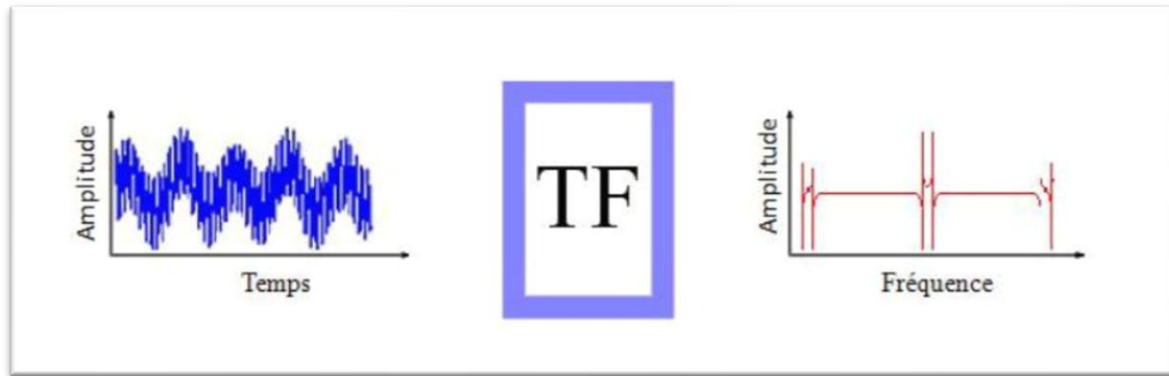


Figure 2.1 : Passage temporel vers fréquentiel par la TF

Pourtant, la méthode de Fourier présente des restrictions puisqu'elle requiert la prise en compte complète de l'évolution temporelle du signal. Même si l'information temporelle est encodée dans la transformation de Fourier, elle se retrouve dissimulée au sein des phases, ce qui complique son extraction. En conséquence, en l'absence de la capacité de fusionner ces deux approches, nous sommes contraints de choisir entre analyser un signal soit selon son évolution temporelle, soit selon ses caractéristiques fréquentielles, et pour résoudre ce problème, la transformée de Fourier à Court Terme (STFT) entre en jeu.

2.1 Transformée de Fourier à court terme (STFT)

La Transformée de Fourier à Court Terme (STFT) est une technique d'analyse qui combine les avantages de l'analyse fréquentielle de la transformée de Fourier standard avec une prise en compte des changements temporels dans un signal. Imaginons que nous souhaitons étudier la performance musicale d'un orchestre pendant un concert. Si nous n'utilisons que la Transformée de Fourier standard, nous obtenons une vue globale de l'ensemble des fréquences présentes dans tout le concert, mais nous ne saurons pas quand un instrument particulier joue ou quand des variations de tempo se produisent.

La STFT résout ce problème en découpant le signal en petits morceaux, comme si nous regardions le concert à travers une série de fenêtres courtes. À chaque instant, nous observons une fenêtre de temps restreinte, puis nous calculons la transformée de Fourier pour cette fenêtre. Cela équivaut à écouter l'orchestre pendant de courts intervalles et à analyser les variations des sons au fil du temps.

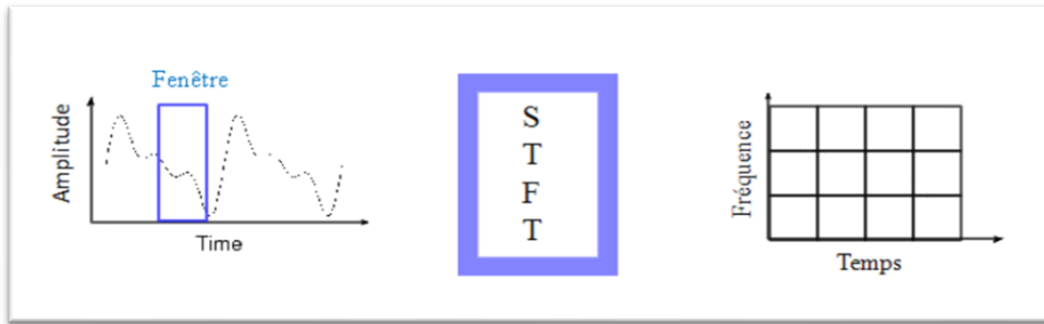


Figure 2.2 : Passage temporel vers temps-fréquence par STFT

La STFT est une technique d'analyse largement utilisée, mais elle a certaines limites lorsqu'il s'agit de caractériser des signaux complexes comme les voix pathologiques. L'un des inconvénients majeurs de la STFT est sa résolution temps-fréquence fixe, ce qui la limite à s'adapter aux variations de fréquence locales dans les signaux. Cela peut nuire à la détection et à la description des structures fluctuantes temporelles et fréquentielles, qui sont typiques des voix pathologiques. C'est là qu'intervient la transformée en ondelettes, une méthode alternative qui offre des avantages significatifs dans l'analyse des voix pathologiques.

3. Transformée en ondelettes

En utilisant des fonctions d'ondelettes étroitement concentrées dans le temps et en fréquence, la transformée en ondelettes permet une résolution temps-fréquence variable. Cette fonctionnalité permet de capturer efficacement les variations locales de fréquence et d'amplitude dans les signaux, ce qui est essentiel pour la caractérisation précise des voix pathologiques.

La différence principale entre la STFT et la transformée en ondelettes est que dans le cas de WT, la largeur de la fenêtre sera changée pendant la transformation et sera calculée pour chaque fréquence. [13]

La transformation en ondelettes consiste à décomposer un signal en une série de composantes d'ondelettes, chacune correspondant à une échelle et à une position donnée. Cette décomposition multi-échelle permet une représentation plus riche et adaptée aux signaux complexes. L'utilisation de la transformée en ondelettes dans le contexte de la classification des voix pathologiques peuvent permettre de différencier plus efficacement les caractéristiques distinctives entre les voix saines et pathologiques.

La fonction fondamentale employée dans la transformation en ondelettes pour réaliser cette décomposition est appelée ondelette mère. Cette ondelette mère peut être envisagée comme un filtre particulier qui subit des dilatations et des translations afin de saisir diverses fréquences et résolutions

présentes dans le signal initial. L'objectif consiste à extraire des informations significatives relatives aux différentes composantes fréquentielles du signal tout en considérant leur échelle temporelle.

La transformation en ondelette d'une fonction $f(t)$ en utilisant l'ondelette mère $\varphi(t)$ peut être exprimée mathématiquement comme suit :

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \varphi^* \left(\frac{t-b}{a} \right) dt$$

Où :

- $W_f(a, b)$ est le coefficient de la transformation en ondelette de $f(t)$ à une échelle a et une translation b .
- $\varphi^* \left(\frac{t-b}{a} \right)$ est la fonction conjuguée de l'ondelette mère $\varphi(t)$.
- a représente l'échelle ou l'étalement de l'ondelette.
- b est le facteur de translation qui déplace l'ondelette le long de l'axe des abscisses.

Cette formule capture comment l'ondelette mère est adaptée et décalée pour capturer différentes composantes du signal $f(t)$ à différentes échelles temporelles et a positions b .

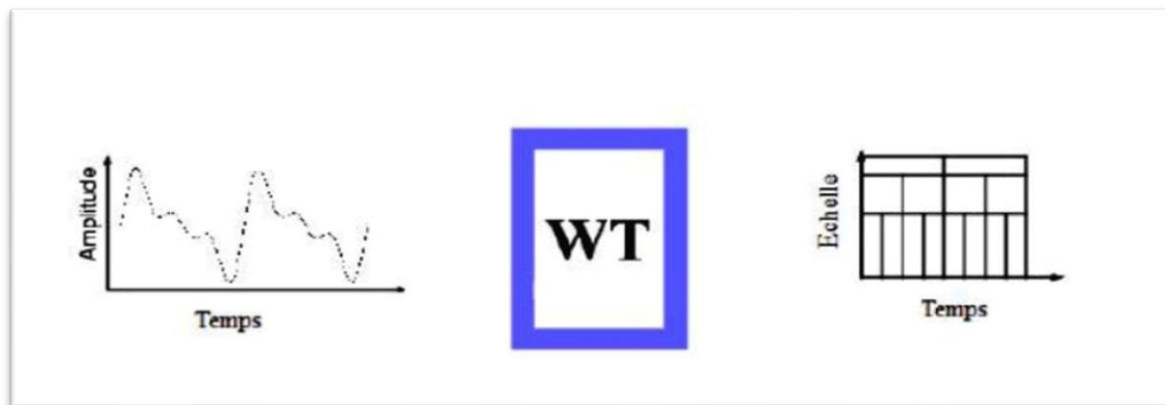


Figure 2.3 : Passage temporel vers temps-échelle par WT

Il y a deux types fondamentaux d'approches dans le domaine de la décomposition en ondelettes : les transformations d'ondelettes continues (TOC) et les transformations d'ondelettes discrètes (TOD).

Les ondelettes continues et discrètes sont des outils essentiels pour l'analyse de signaux à diverses échelles de fréquence. Les ondelettes discrètes fonctionnent mieux avec les signaux discrets, tandis que les ondelettes continues fonctionnent mieux avec les signaux continus. Dans les applications numériques où les signaux sont échantillonnés, la transformée en ondelette discrète est souvent préférée.

3.1 Familles d'ondelettes

Diverses familles d'ondelettes sont disponibles, chacune présentant des caractéristiques particulières de localisation dans le domaine temps-fréquence. Le Tableau 2.1 offre une liste indicative de ces familles d'ondelettes, tandis que des informations plus approfondies à leur sujet sont fournies dans la référence. [13]

Nom des familles d'ondelettes	Abréviations
Famille de Daubechies	<i>db</i>
Famille de Coiflets	<i>Coif</i>
Famille de Symlets	<i>Sym</i>
Famille de Haar	<i>Haar</i>
Famille de Meyer	<i>Meyr</i>
Famille de Morlet	<i>Morl</i>
Famille de Gabor	<i>Gaus</i>

Tableau 2.1 : Familles d'ondelettes

3.2 Choix de la meilleure famille

Le choix de la famille d'ondelettes la mieux adaptée à la classification des voix pathologiques repose sur divers facteurs, tels que les caractéristiques spécifiques que nous souhaitons extraire des signaux vocaux et les variations particulières introduites par les pathologies vocales. Chaque famille d'ondelettes possède ses propres caractéristiques et avantages distincts, et la sélection dépendra de l'objectif spécifique que nous poursuivons en matière de classification.

Dans le contexte de l'analyse des signaux vocaux pathologiques, certaines familles d'ondelettes ont gagné en popularité et sont fréquemment utilisées dans des applications de traitement de signal et de classification. Il convient de mentionner que ces familles ont démontré leur efficacité dans la détection et l'extraction de caractéristiques pertinentes pour identifier les anomalies vocales :

- **Daubechies (db)** : Les ondelettes de la famille Daubechies, telles que db4, sont largement utilisées en raison de leur propriété de compacité spectrale et de leur bonne représentation des bords et des transitions abruptes.

- **Symlet (sym)** : Les ondelettes Symlet sont similaires aux ondelettes de Daubechies mais avec un certain degré de symétrie, ce qui peut être utile pour certaines applications.
- **Coiflets (haar)** : Les ondelettes Coiflets sont également adaptées aux bords et aux discontinuités. L'ondelette de Haar (une ondelette Coiflet particulière) est souvent utilisée en raison de sa simplicité.

3.3 Transformée par ondelette continue (TOC)

Fondamentalement basée sur l'utilisation d'une ondelette mère $\varphi(t)$, une fonction de base modifiée en termes d'échelle et de position sur l'axe temporel, la transformée en ondelette continue calcule des coefficients d'ondelettes en multipliant le signal d'entrée par des versions dilatées et décalées de l'ondelette mère, suivies de l'intégration de ces produits sur la durée totale du signal. Ce processus est ensuite répété pour diverses échelles et positions, aboutissant à une représentation en coefficients d'ondelettes qui met en lumière les variations du signal à différentes résolutions.

En somme, l'ondelette continue est dilatée et translatée pour capturer différentes échelles et positions temporelles. Cela signifie qu'elle est étirée ou comprimée dans le temps et déplacée le long de l'axe temporel afin de détecter les variations du signal à différentes fréquences et à différents moments.

3.3.1 Décomposition de l'ondelette continue [14]

Cette décomposition porte sur l'analyse temps-échelle. Elle présente les résultats de cette analyse appliquée aux segments de la base de données mentionnée précédemment. Il est important de rappeler que nous avons utilisé l'ondelette de type Daubechies 'db4' avec une échelle de 128. [14]

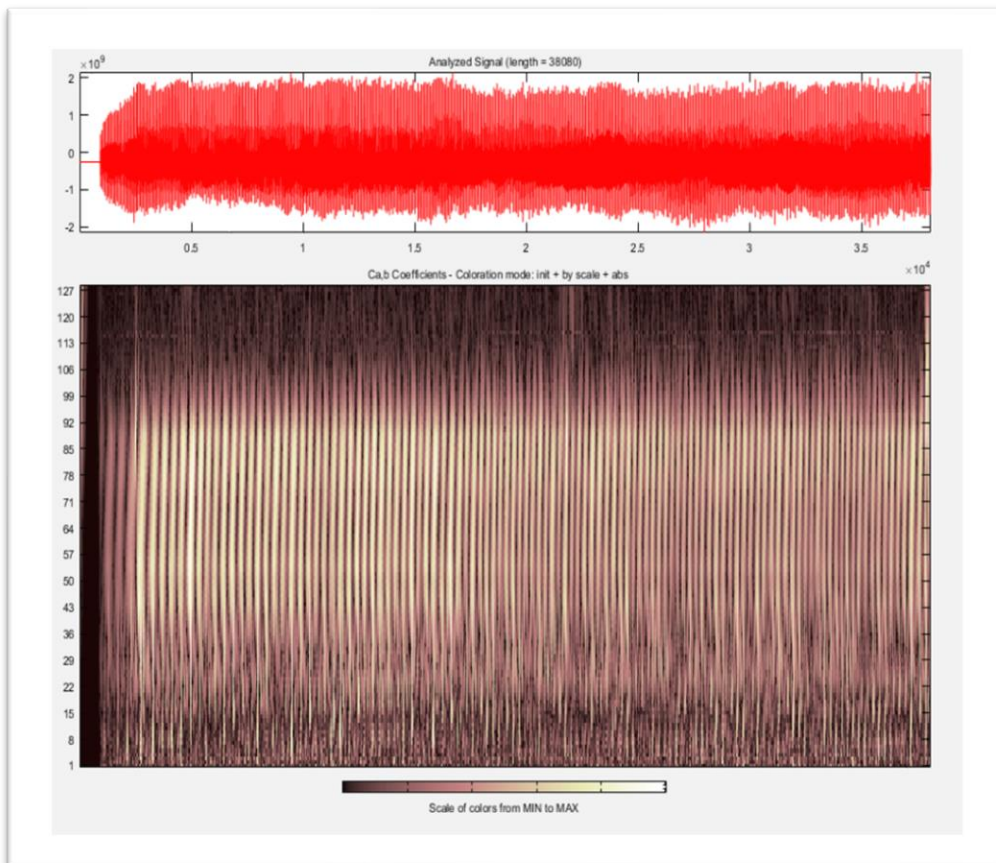


Figure 2.4 : Transformée en ondelettes continu pour une voix pathologique

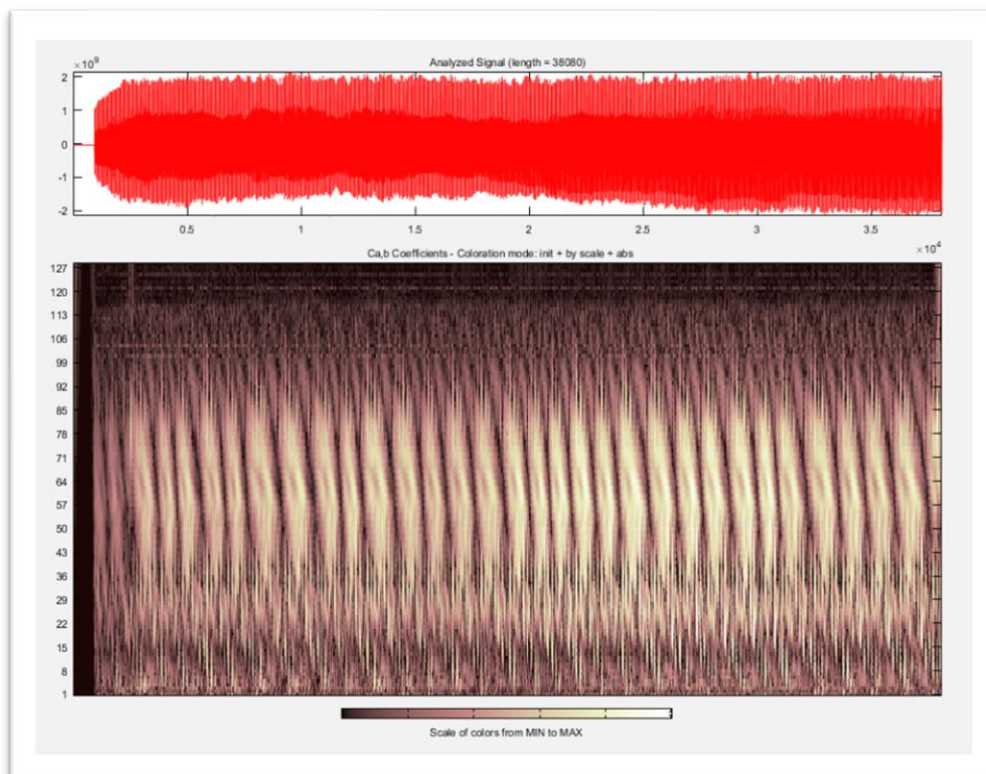


Figure 2.5 : Transformée en ondelettes continu pour une voix normale

Une comparaison entre la Figure 2.4 et la Figure 2.5 a mis en évidence des distinctions entre les signaux vocaux (pathologique et normal). Dans le cas de la voix normale, les coefficients d'ondelette présentent une répartition plus équilibrée à travers l'espace fréquentiel. Les variations des caractéristiques acoustiques se montrent généralement régulières, réparties de manière cohérente sur l'axe temporel et l'échelle 92.

En contraste, pour la voix pathologique, une déformation plus prononcée des coefficients d'ondelette a été constatée. Ces déformations se traduisent par des pics ou des valeurs inhabituellement élevées à certaines échelles (106) et positions temporelles. Cette observation reflète des anomalies et des variations atypiques dans les caractéristiques acoustiques des voix pathologiques.

3.4 Transformée par ondelette discrète

L'ondelette discrète, également connue sous le nom de transformée en ondelette discrète, représente une méthode d'analyse appliquée aux signaux discrets. Comparativement à la transformée d'ondelettes continue, l'approche centrale de la TOD réside dans sa capacité à décomposer un signal en différentes échelles d'ondelettes. Ce processus met en évidence les détails les plus subtils à certaines échelles tout en mettant en avant les caractéristiques globales à d'autres échelles. Contrairement aussi à l'ondelette continue, qui fonctionne avec des fonctions continues, la TOD a été spécialement développée pour traiter des signaux échantillonnés, tels que les signaux vocaux.

Les étapes pour effectuer une TOD simple sur un signal discret sont :

- **Sélection de l'ondelette de base :** Nous choisissons une ondelette de base qui servira de motif pour la transformation. Les ondelettes comme Haar, Daubechies, et bien d'autres sont couramment utilisées.
- **Étapes de décomposition :**
 - Étape 1 : Calcul des coefficients d'approximation et de détail :** La première étape consiste à calculer les coefficients d'approximation (décrivant les composantes à grande échelle) et les coefficients de détail (décrivant les détails fins) à partir du signal original.
 - Étape 2 : Sous-échantillonnage :** Les coefficients d'approximation calculés sont ensuite sous-échantillonnés en supprimant certains échantillons pour obtenir une résolution plus basse.
 - Étape 3 : Répétition :** Les étapes précédentes sont répétées sur les coefficients d'approximation obtenus à chaque étape pour obtenir des coefficients d'approximation et de détail à différentes échelles.

- **Répétition et pyramide d'échelle** : Les étapes de décomposition sont généralement répétées plusieurs fois pour obtenir une pyramide d'échelle de coefficients d'approximation et de détail à différentes résolutions (Figure 2.3).

La formule mathématique de la transformée en ondelette discrète (TOD) implique la convolution et le sous-échantillonnage de l'ondelette mère et du signal d'entrée. Voici comment cela peut être exprimé de manière générale :

$$DW(f, \varphi) = \sum_k a_{k,j} \cdot \varphi_{j,k}(t) + \sum_{j,k} d_{j,k} \cdot \phi_{j,k}(t)$$

Où :

- $f(t)$ est le signal d'entrée.
- $a_{k,j}$ sont les coefficients d'approximation à l'échelle j et la position k .
- $d_{j,k}$ sont les coefficients de détail à l'échelle j et la position k .
- $\varphi_{j,k}(t)$ est une version dilatée et translatée de l'ondelette mère $\varphi(t)$.
- $\phi_{j,k}(t)$ est une version dilatée et translatée de la fonction d'échelle $\phi(t)$, qui est liée à l'ondelette mère et est utilisée pour calculer les coefficients d'approximation.

Les coefficients $a_{k,j}$ et $d_{j,k}$ peuvent être calculés en utilisant des filtres de convolution et des opérations de sous-échantillonnage.

4. Machines à vecteurs de supports

Les SVM (Support Vector Machines) ont été inventées par les scientifiques Vladimir Vapnik et Alexey Chervonenkis dans les années 1960s. Elles sont devenues un outil populaire pour la classification discriminante. Un domaine passionnant de l'application récente des SVMs est dans le traitement de la parole. Ces modèles ont une stratégie de modélisation nettement différente dans la détection de troubles de la voix, par rapport à d'autres méthodes de classification reportées dans la littérature. [15]

4.1 Principe de fonctionnement

Les modèles d'apprentissage supervisés appelés SVM sont utilisés pour séparer des points de données en différents groupes (classes) en déterminant un hyperplan optimal dans un espace de caractéristiques de dimension supérieure. L'objectif est de découvrir un hyperplan dimensionnel qui permet de maximiser la distance entre les points de données de diverses classes.

Dans sa forme élémentaire, lorsque les deux classes sont linéairement séparables comme illustré dans la Figure 2.6 cette approche vise à identifier un hyperplan discriminant représenté par l'équation suivante :

$$w \cdot x + b = 0$$

Les éléments de cette équation sont définis comme suit :

- w : Vecteur des poids ($w_1, w_2, w_3, \dots, w_n$).
- x : Vecteur des attributs ($x_1, x_2, x_3, \dots, x_n$).
- b : Seuil du séparateur linéaire.

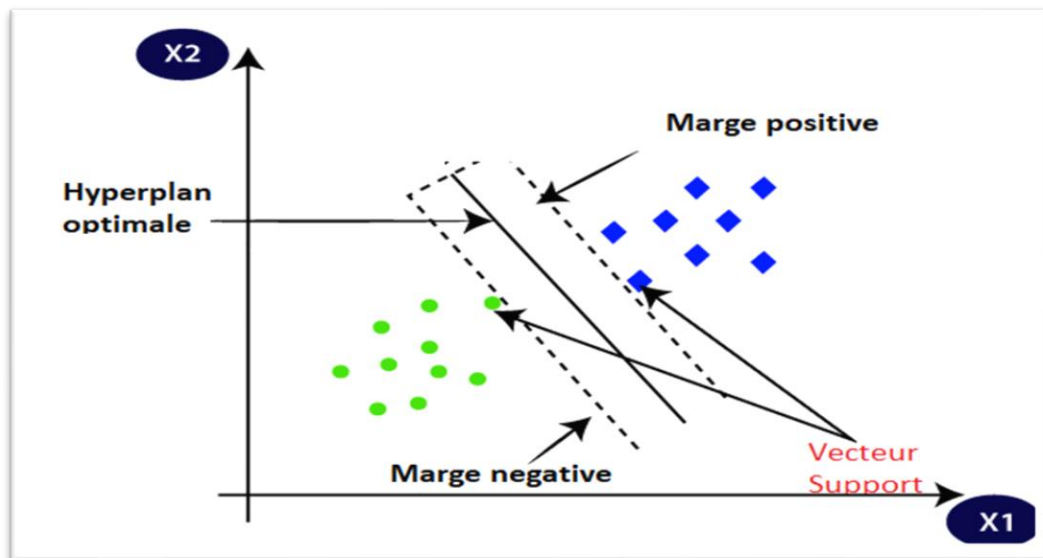


Figure 2.6 : Exemple de classification par SVM

- **Hyperplan** : L'hyperplan divise les données en différentes classes. La tâche consiste à trouver l'hyperplan idéal qui minimise les erreurs de classification tout en maximise la marge entre les points de données des différentes classes.
- **Marge** : La distance entre l'hyperplan de séparation et les vecteurs de support les plus proches est représentée par la marge d'une SVM, et l'optimisation vise à trouver l'hyperplan qui maximise cette distance afin d'obtenir une meilleure séparation entre les classes et une meilleure généralisation des prédictions.
- **Vecteur support** : Les points de données les plus proches de l'hyperplan de séparation entre les différentes classes sont appelés vecteurs de support. La position et l'orientation de l'hyperplan optimal sont déterminées par ces vecteurs.

4.2 Linéarité et non linéarité

4.2.1 SVM linéaire

Un classificateur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire. Dans la suite, nous supposons que nos exemples sont donnés dans un format vectoriel. Notre espace d'entrée x est composé de n composantes. Si les données sont linéairement séparables, alors il existe un hyperplan d'équation (Figure 2.7). [16]

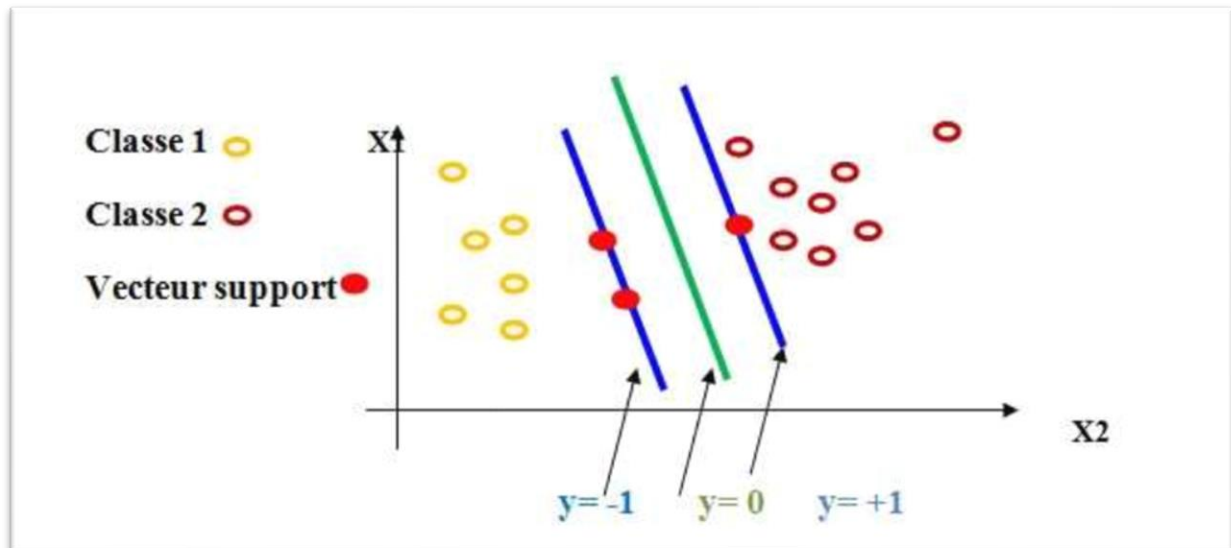


Figure 2.7 : Exemple graphique de données linéairement séparables

4.2.2 SVM non-linéaire

Un SVM non linéaire utilise des techniques pour traiter des données qui ne peuvent pas être séparées de manière linéaire dans l'espace d'origine. Pour cela, il projette les données dans un espace de dimension supérieure où elles peuvent potentiellement devenir linéairement séparables. Cela permet d'appliquer le concept de séparation linéaire dans cet espace étendu.

Mathématiquement, pour un SVM non linéaire avec une fonction de noyau K , l'hyperplan de séparation est représenté par :

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b$$

Où :

- $f(x)$ est la fonction de décision pour un nouvel échantillon x .
- n est le nombre total d'échantillons d'entraînement.
- α_i sont les coefficients obtenus lors de l'optimisation du SVM.
- y_i est l'étiquette de classe de l'échantillon d'entraînement x_i ($y_i = \pm 1$ pour les classes binaires).

- $K(x_i, x)$ est la fonction de noyau qui mesure la similarité entre les échantillons x_i et x dans un espace de caractéristiques transformé.
- b est le terme de biais.

Vapnik [17] a montré que toute fonction satisfaisant les conditions d'admissibilité peut être utilisée comme noyau, parmi les noyaux les plus utilisés pour la classification, nous citons : le noyau linéaire, le noyau polynomial et le noyau Gaussien.

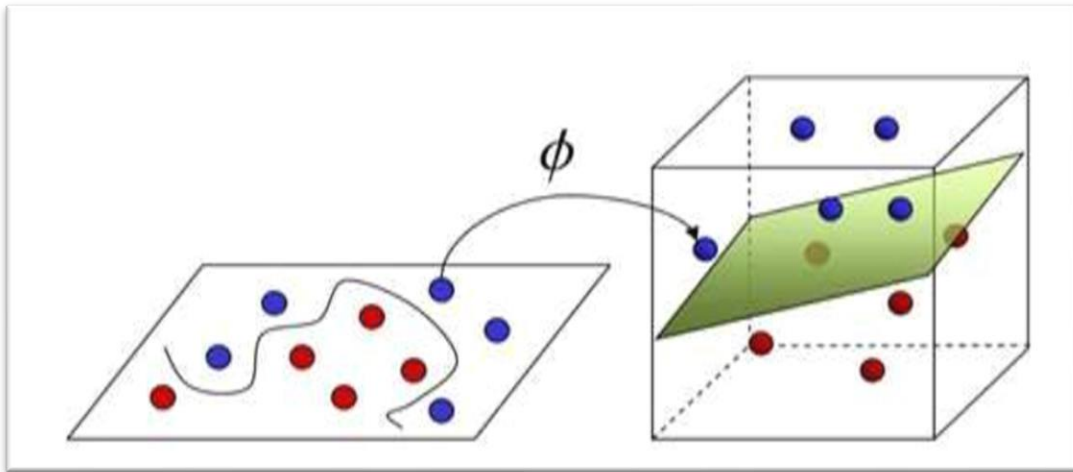


Figure 2.8: Exemple de transformation des données non linéaires dans un espace de dimension supérieure

5. Méthode proposée

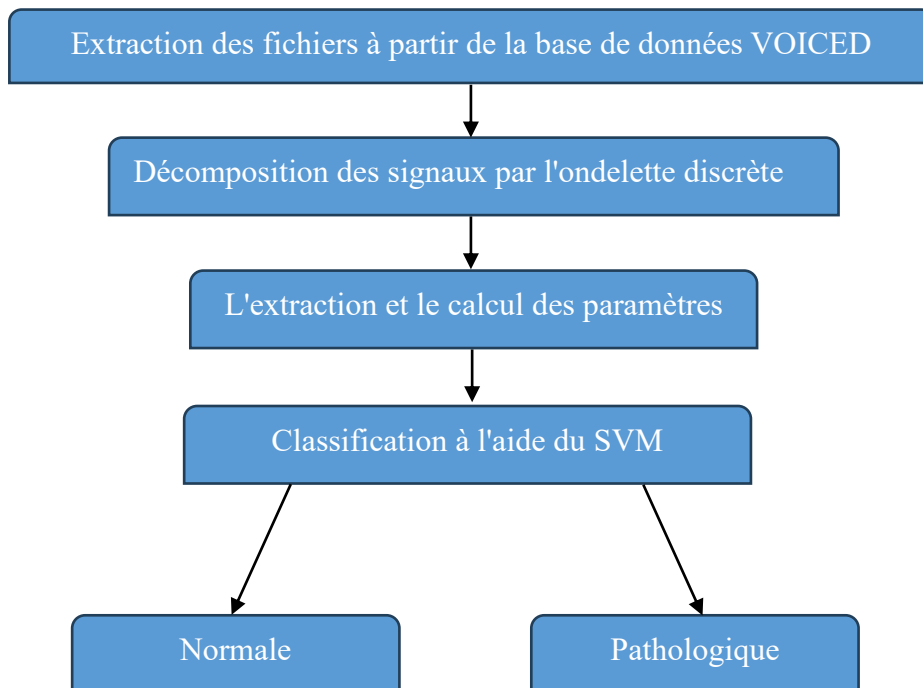


Figure 2.9 : Méthode Proposée de la classification des voix Pathologiques

5.1 Base de données

Pour la base de données, nous abordons notre travail avec la base de données VOICE ICar Federico II (VOICED), téléchargée depuis PhysioNet, en tant qu'élément essentiel de notre étude de classification des voix pathologiques. En travaillant avec VOICED, nous avons identifié sa pertinence significative pour notre recherche. Avec 208 enregistrements vocaux représentant à la fois des voix pathologiques et normales, cette base de données constitue un ensemble de données adéquat pour former et évaluer notre modèle de classification. Cette compréhension approfondie de la base de données VOICED jette les bases de notre méthodologie, que nous exposerons en détail dans le chapitre suivant.

5.2 Décomposition des signaux en ondelette

Notre approche repose sur l'application de la décomposition en ondelettes discrètes pour façonner le vecteur caractéristique de notre signal vocal. Cette décomposition est accomplie en se guidant par le choix spécifique et l'ordre des ondelettes employées dans le calcul des coefficients. Les tronçons du signal vocal sont fragmentés en quatre niveaux en utilisant l'ondelette de Daubechies 'db4', créant ainsi quatre détails et une approximation pour chaque signal. Voici les étapes :

- **Étape 1 : Préparation du Signal** : Nous prenons un signal unidimensionnel à analyser.
- **Étape 2 : Filtrage par le Filtre Passe-Bas (h)** : Nous appliquons un filtre passe-bas (h) au signal pour extraire les composantes à basses fréquences. Le résultat du filtrage est le signal d'approximation (cA), qui contient les caractéristiques globales du signal à cette échelle.
- **Étape 3 : Filtrage par le Filtre Passe-Haut (g)** : Nous appliquons un filtre passe-haut (g) au signal pour extraire les composantes à hautes fréquences. Le résultat du filtrage est le signal de détail (cD), qui contient les variations fines et les détails spécifiques du signal à cette échelle.
- **Étape 4 : Sous-Échantillonnage** : Nous réduisons la résolution temporelle du signal d'approximation (cA) et du signal de détail (cD) en conservant chaque deuxième échantillon. Ce sous-échantillonnage réduit la taille des signaux mais préserve les caractéristiques essentielles à différentes échelles.
- **Étape 5 : Répétition du Processus (Itérations)** : Nous répétons les étapes 2 à 4 pour le signal d'approximation (cA) obtenu dans l'étape précédente. Chaque itération divise le signal en nouvelles composantes d'approximation (cA) et de détail (cD) à une échelle plus fine.
- **Étape 6 : Répéter Jusqu'au Niveau Souhaité** : Nous répétons les étapes 2 à 5 pour un certain nombre prédéfini d'itérations (niveaux de décomposition) ou jusqu'à ce que la résolution temporelle souhaitée soit atteinte.

- Étape 7 : Coefficients d'Approximation et de Détail** : À la fin du processus, nous obtenons des coefficients d'approximation (cA) et de détail (cD) à différents niveaux de décomposition. Les coefficients d'approximation contiennent les caractéristiques globales du signal à différentes échelles. Les coefficients de détail contiennent les variations fines et les détails spécifiques du signal à différentes échelles.

Les exemples (Figure 2.10 et Figure 2.11) démontrent la décomposition en ondelettes jusqu'au **quatrième** niveau pour les signaux vocaux : normales et pathologiques.

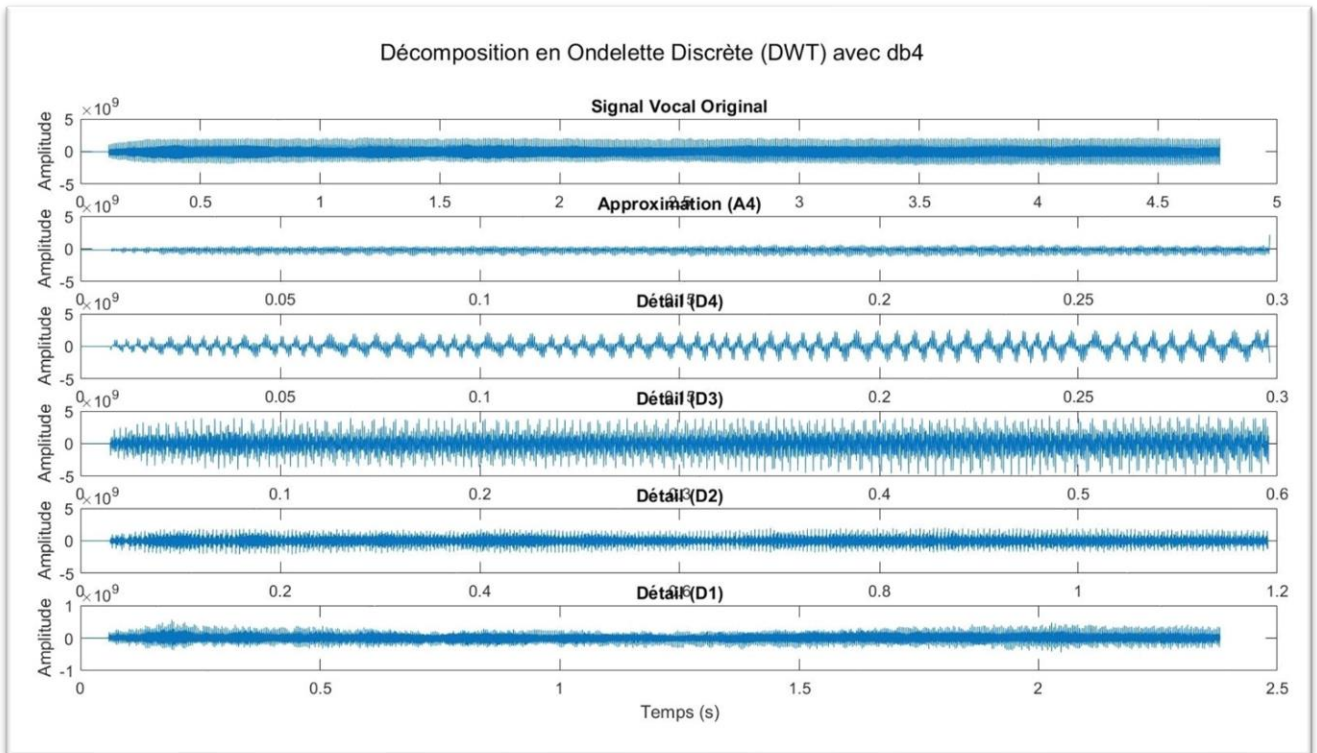


Figure 2.10 : Exemple de décomposition en ondelette discrète à 4 niveaux de voix normale

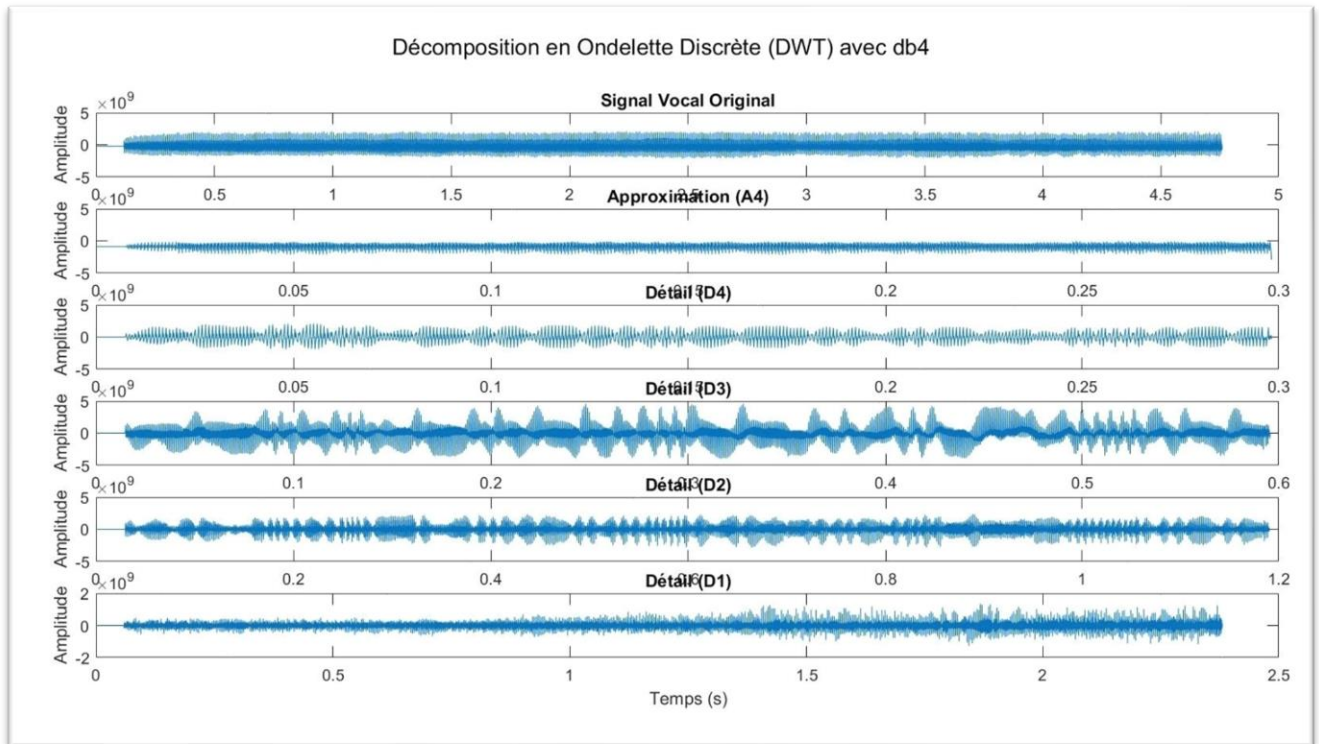


Figure 2.11 : Exemple de décomposition en ondelette discrète à 4 niveaux de voix pathologique

5.3 Extraction des paramètres

Après avoir achevé la transformation en ondelette, nous avons entrepris l'étape cruciale d'extraction des paramètres. Cette phase est essentielle pour saisir les caractéristiques distinctives des signaux vocaux, nécessaires à la distinction entre les voix pathologiques et normales. Nous avons calculé un ensemble de paramètres qui fournissent des informations sur différentes propriétés du signal vocal, contribuant ainsi à la construction d'un modèle de classification solide.

Avant d'extraire les paramètres, nous avons normalisé le signal vocal pour ramener toutes les valeurs du signal dans une plage spécifique, souvent entre -1 et 1. Cette étape est cruciale pour éviter que les valeurs extrêmes ne biaisent les calculs ultérieurs. La normalisation est effectuée comme suit :

$$\text{Signal normalise} = \frac{\text{Signal original}}{\max(\text{Signal original})}$$

Cette normalisation garantit que toutes les amplitudes du signal se situent entre -1 et 1, facilitant ainsi la comparaison et l'extraction de caractéristiques.

Paramètres de l'Approximation A4 :

- **Énergie** : L'énergie du signal est un indicateur de son intensité globale à cette échelle de fréquence. Elle est calculée en sommant les carrés des amplitudes du signal normalisé.

L'énergie est une caractéristique essentielle pour quantifier la force ou l'intensité du signal à cette échelle.

$$Energy = \sum_{i=1}^N (Signal_A4[i])^2$$

- **Moyenne** : La moyenne des amplitudes du signal reflète son niveau moyen et peut être liée à la sonorité globale du signal normalisé. Une valeur de moyenne significativement différente pourrait indiquer un changement dans le niveau du signal.

$$Moyenne = \frac{1}{N} \sum_{i=1}^N Signal_A4[i]$$

- **Écart Type** : L'écart type mesure la dispersion des amplitudes autour de la moyenne. Il offre des informations sur la variabilité du signal normalisé. Un écart-type élevé peut indiquer des variations importantes dans le signal, tandis qu'un écart-type faible peut indiquer une plus grande cohérence.

$$Ecart_{type} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Signal_A4[i]^2 - Moyenne)^2}$$

- **Variance** : La variance mesure la dispersion des valeurs du signal A4 autour de leur moyenne.

$$Variance = \frac{1}{N} \sum_{i=1}^N (Signal\ A4 - Moyenne)^2$$

- **Fréquence Moyenne** : La fréquence moyenne représente la fréquence centrale pondérée du signal d'approximation. Elle offre des informations sur la distribution des fréquences dans le signal normalisé.

Paramètres du Détail D4 :

Les étapes similaires ont été suivies pour extraire les paramètres du détail D4 :

- **Fréquence Dominante** : La fréquence dominante dans le signal est celle qui possède l'amplitude la plus élevée. Elle capture les variations fines à haute fréquence du signal normalisé.
- **Ratio des Bandes de Fréquences** : Le ratio des bandes de fréquences mesure la distribution d'énergie entre les bandes de fréquences basses et élevées du signal.
- **L'entropie spectrale** : L'entropie spectrale mesure la complexité du signal en termes de répartition de l'énergie sur différentes fréquences.

Les tableaux 2.2 et 2.3 présentent les paramètres calculés pour les signaux d'approximation A4 et de détail D4, tant pour une voix pathologique que pour une voix normale.

Paramètres à calculer	A4	Paramètres à calculer	D4
Énergie	125.1481	Fréquence Dominante	1835.9375
Moyenne	-0.0788		
Écart Type	0.2151	Ratio des Bandes de Fréquences	0.4799
Variance	0.0463		
Fréquence Moyenne	2650.3964	L'entropie spectrale	0.2617

Tableau 2.2 : Les paramètres d'un signal vocal normal

Paramètres à calculer	A4	Paramètres à calculer	D4
Énergie	560.2596	Fréquence Dominante	1101.5625
Moyenne	-0.4033		
Écart Type	0.2686	Ratio des Bandes de Fréquences	0.0213
Variance	0.0722		
Fréquence Moyenne	793.4092	L'entropie spectrale	0.1851

Tableau 2.3 : Les paramètres d'un signal vocal pathologique

Ces paramètres, calculés à partir des détails et de l'approximation normalisée, serviront de caractéristiques pour notre modèle de classification SVM, visant à différencier les voix pathologiques des voix normales.

5.4 Classification (normale ou pathologique)

Pour accomplir notre objectif de classification des voix en catégories pathologiques et normales, nous adoptons une approche basée sur une machine learning, en particulier en utilisant un SVM. Cette méthode s'est avérée efficace dans de nombreuses tâches de classification, y compris la nôtre. Nous avons extrait les caractéristiques pertinentes à partir des enregistrements vocaux de la base de données VOICE ICar Federico II (VOICED) que nous avons travaillée précédemment. Ces caractéristiques

seront utilisées comme entrées pour notre modèle SVM. En ajustant les paramètres du SVM, nous chercherons à obtenir la meilleure séparation possible, permettant ainsi de classifier efficacement de nouvelles voix. Notre choix d'un SVM repose sur sa simplicité et son efficacité pour des ensembles de données de taille modérée, comme celui que nous manipulons avec VOICED. Bien que les détails spécifiques de prétraitement, d'extraction de caractéristiques aient déjà été réalisés, nous allons détailler ces étapes dans les prochaines sections.

De plus, nous exposerons nos critères d'évaluation qui mesureront la performance de notre modèle. En capitalisant sur les propriétés discriminantes du SVM, notre objectif est de créer un outil de classification précis et fiable pour différencier entre les voix pathologiques et normales. Ces avancées ouvrent la voie à diverses applications potentielles, aussi bien dans le domaine médical que dans d'autres domaines connexes.

6. Conclusion

Ce chapitre a éclairé les concepts et les techniques fondamentaux qui sous-tendent notre projet de classification des voix pathologiques. En explorant la transformée en ondelettes, nous avons identifié la puissance de l'ondelette discrète pour notre travail. De plus, nous avons abordé en profondeur les SVM pour sa pertinence dans notre contexte. En mettant en évidence ces éléments, nous avons pu définir notre approche méthodologique qui combine judicieusement la base de données VOICE ICar Federico II (VOICED), l'extraction de paramètres à partir de l'ondelette discrètes, et la classification à l'aide du SVM. Ces choix et combinaisons nous dirigent vers la concrétisation de notre objectif de distinction précise entre les voix pathologiques et normales.

Dans le chapitre suivant, nous présenterons en détail les résultats que nous avons obtenus grâce à cette méthodologie.

Chapitre 3 :
Résultats et
Discussion

1. Introduction

Les progrès réalisés dans le domaine des technologies d'apprentissage et de classification ont notablement simplifié le développement de méthodes et d'outils de diagnostic pour les affections vocales.

L'objectif de ce chapitre est de présenter les résultats de notre étude sur la détection des pathologies vocales. Nous allons analyser les performances de notre modèle de classification SVM en mettant en évidence l'utilisation de la matrice de confusion comme outil d'évaluation clé. De plus, nous situerons nos résultats dans le contexte de la recherche existante en les comparant à d'autres travaux similaires. L'objectif ultime est de démontrer la pertinence et la valeur ajoutée de notre approche dans le domaine de la détection des pathologies vocales.

2. Travaux effectués

2.1 Identification Des Voix Pathologiques Par Réseaux De Neurones [18]

En 2009, Lotfi SALHI et son équipe ont présenté une méthode qui repose sur l'utilisation d'un réseau de neurones multicouches pour la détection des voix présentant des pathologies. Les données nécessaires pour cette étude ont été collectées conjointement à l'Université d'UCLA à Los Angeles, au sein du laboratoire G.E, ainsi qu'à l'Hôpital RABTA de Tunis.

L'ensemble de données est composé de 100 enregistrements audios provenant de différents locuteurs, dont 50 sont considérés comme vocalement normaux, tandis que les autres présentent des troubles vocaux ou neurologiques.

Les résultats mettent en évidence que le pitch (fréquence fondamentale) et les trois premiers formants se révèlent être les paramètres d'entrée les plus efficaces pour la distinction et l'identification des voix affectées par des pathologies, grâce à l'utilisation de réseaux de neurones. La Figure 3.1 montre le principe d'identification de la parole utilisé par Lotfi SALHI.

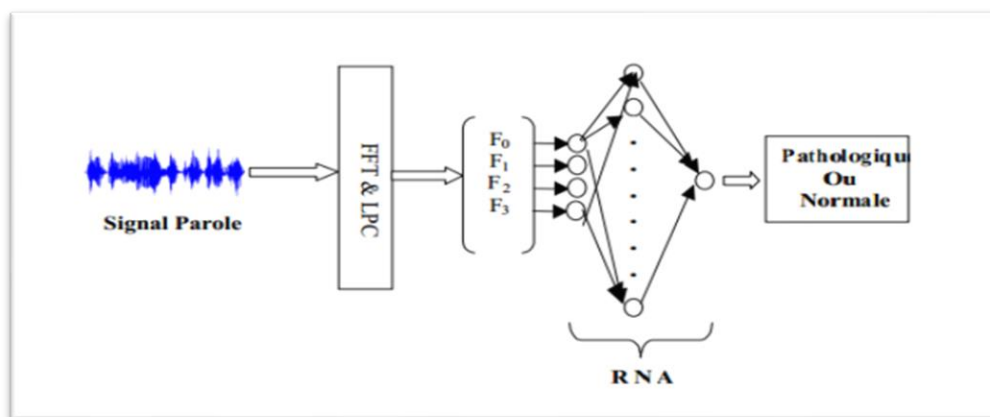


Figure 3.1 : Principe d'identification de la parole

2.2 Identification des troubles de la voix par l'analyse temps-fréquence et les machines à vecteurs de support [16]

En 2016, Nesrine BENAMMAR et Amina BOUTICHE ont mené cette étude à l'Université de M'HAMED BOUGARA de BOUMERDES, en utilisant deux bases de données internationales, à savoir la base MEEI et la base SVD.

Leur intérêt s'est porté sur les techniques d'analyse temps-fréquence, plus spécifiquement sur l'utilisation de la transformée d'ondelette continue (TOC). Ils ont élaboré une méthode d'extraction d'attributs qui repose sur la combinaison de la TOC avec des statistiques d'ordre supérieur. La classification est ensuite effectuée en utilisant des machines à vecteurs de support (SVM). Nous pouvons observer le processus d'extraction des attributs dans la Figure 3.2

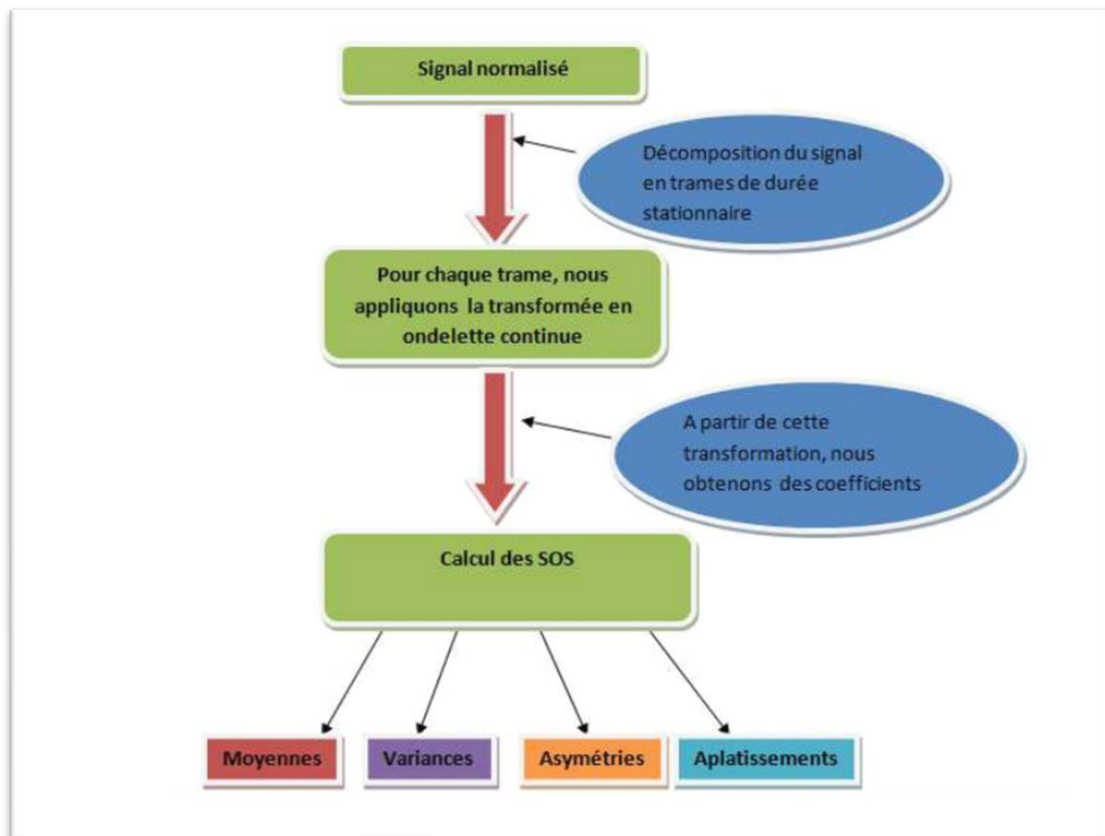


Figure 3.2 : Organigramme d'extraction d'attributs

2.3 Détection et classification des voix pathologiques utilisant l'analyse par ondelette discrète basée sur les caractéristiques statistiques d'ordre supérieures [19]

En 2019, I. Hammami et son équipe ont développé une méthode visant à identifier les voix présentant des pathologies en se basant sur des caractéristiques d'ordre supérieur (SOS) dérivées de l'analyse EMD-DWT. Leur recherche s'est appuyée sur la base de données vocales de Saarbrücken (SVD), qui

regroupe plus de 2000 fichiers audios d'enregistrements de voyelles soutenues (/a/, /i/ et /u/) à différentes hauteurs.

La méthodologie de cette étude se décompose en trois étapes majeures. Tout d'abord, les signaux vocaux subissent une décomposition en modes empiriques (EMD), suivie d'une transformation en ondelettes discrètes (DWT). À partir des coefficients issus de la DWT, plusieurs caractéristiques pertinentes sont extraites, notamment des caractéristiques d'ordre supérieur (HOS) telles que l'asymétrie, l'aplatissement et la variance, ainsi que d'autres caractéristiques comme la valeur moyenne, l'énergie et l'entropie. Ces caractéristiques sont ensuite utilisées pour former des vecteurs de données représentant les signaux vocaux. Enfin, un classifieur SVM est utilisé pour la détection et la classification des voix atteintes de pathologies.

La Figure 3.3 offre une illustration du processus de classification des voix pathologiques en utilisant l'analyse EMD-DWT.

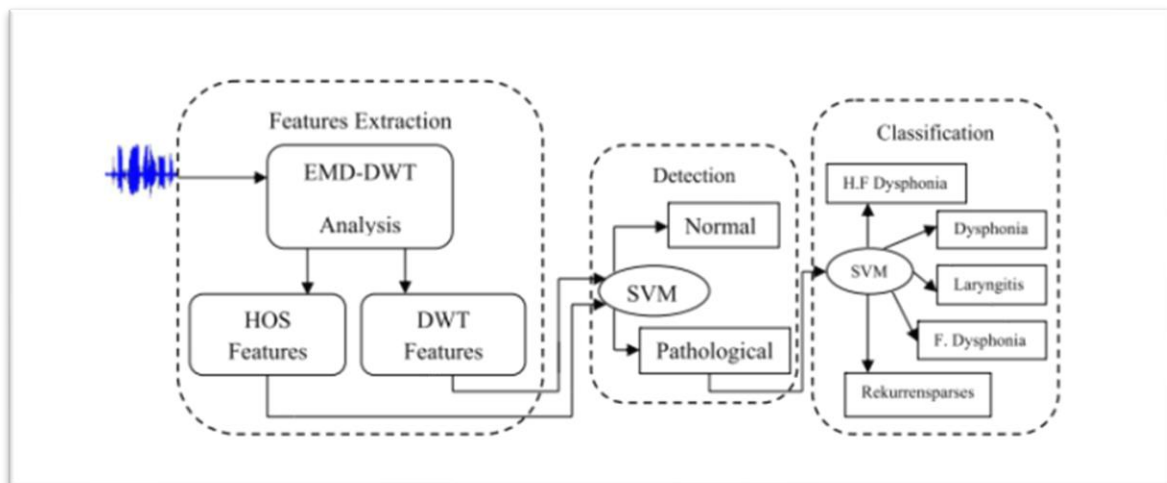


Figure 3.3 : Méthode de classification des voix pathologiques en utilisant l'analyse EMD-DWT

2.4 Développement d'un système de détection de la voix pathologique [20]

En 2021, G. MAROUA et K. RAHMA ont entrepris un projet visant à concevoir un système automatisé de détection des troubles vocaux à l'UNIVERSITÉ BADJI MOKHTAR - ANNABA.

Les expériences, évaluations et résultats présentés dans leur étude démontrent l'efficacité du classificateur SVM pour accomplir la détection et la classification des voix pathologiques. Les taux de précision obtenus grâce à la méthode SVM et l'utilisation des caractéristiques MFCC sont considérés comme satisfaisants pour la base de données SVD. En contraste, une précision totale a été atteinte avec la base de données HUPA, indiquant ainsi que les performances lors de l'utilisation de la base de données HUPA surpassent celles obtenues avec la base SVD.

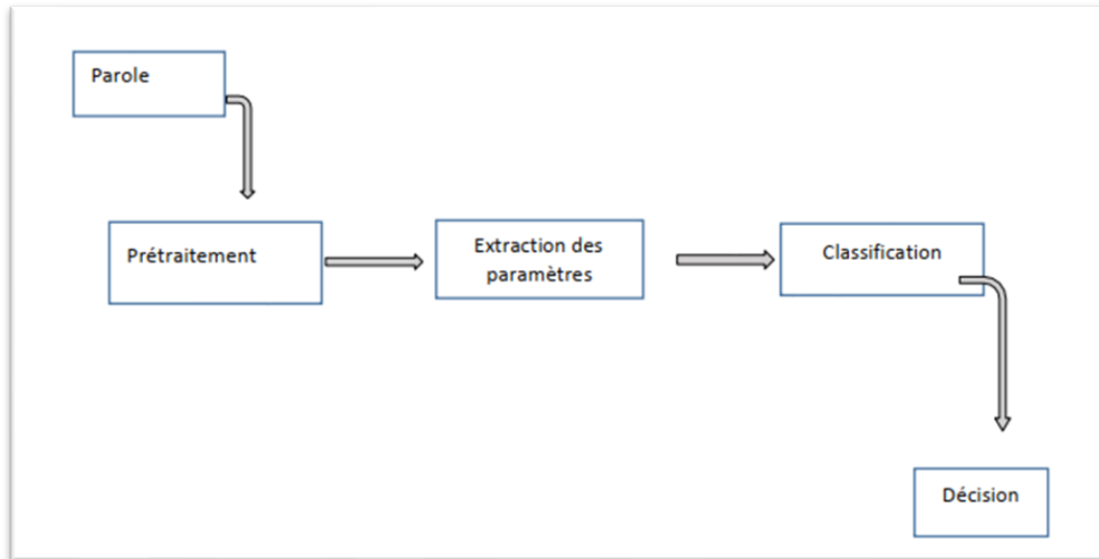


Figure 3.4 : Schéma de la reconnaissance vocale simplifiée

2.5 Identification des troubles de la voix à l'aide des techniques d'apprentissage automatique [21]

En 2021, M. BOUDJELLABA et N. BOUDJERIDA ont entrepris une étude visant à identifier les troubles vocaux à l'Université Mohammed Seddik Ben Yahia-Jijel.

L'objectif de leur travail était de mettre en place un système pour l'identification des troubles vocaux en utilisant des techniques d'apprentissage automatique. Ils ont proposé deux méthodes de classification, à savoir SVM et K-NN, afin d'évaluer leur efficacité dans la détection et la classification de diverses pathologies vocales. Ce système pourrait être employé par les orthophonistes pour effectuer une évaluation objective des voix de leurs patients, en se basant sur des mesures acoustiques et aérodynamiques telles que les MFCC, le Jitter, le Shimmer et le HNR.

À la suite de l'analyse des résultats obtenus, il a été observé que les techniques d'apprentissage automatique SVM et K-NN, évaluées à l'aide de la base de données MEEI, qui est fréquemment utilisée dans de nombreuses études de recherche similaires, se sont avérées efficaces pour la détection des voix pathologiques et normales, ainsi que pour la classification des différents types de pathologies.

3. Les étapes de la classification

3.1 Collecte de Données

La base de données "VOICE ICar Federico II (VOICED)", hébergée sur la plateforme PhysioNet, représente un jalon significatif dans l'étude des voix pathologiques.

Composée de 208 enregistrements vocaux minutieusement collectés, cette base de données offre une opportunité unique d'explorer les aspects acoustiques distinctifs des voix pathologiques par rapport aux voix normales.

En effet, elle contient 150 enregistrements de voix pathologiques, provenant de sujets présentant diverses conditions médicales affectant la voix, ainsi que 58 enregistrements de voix normales qui servent de groupe de contrôle (208 en total).

Elle est aussi stratifiée selon le genre, comprenant 135 enregistrements de voix pathologiques de femmes et 73 enregistrements de voix pathologiques d'hommes (208 en total)

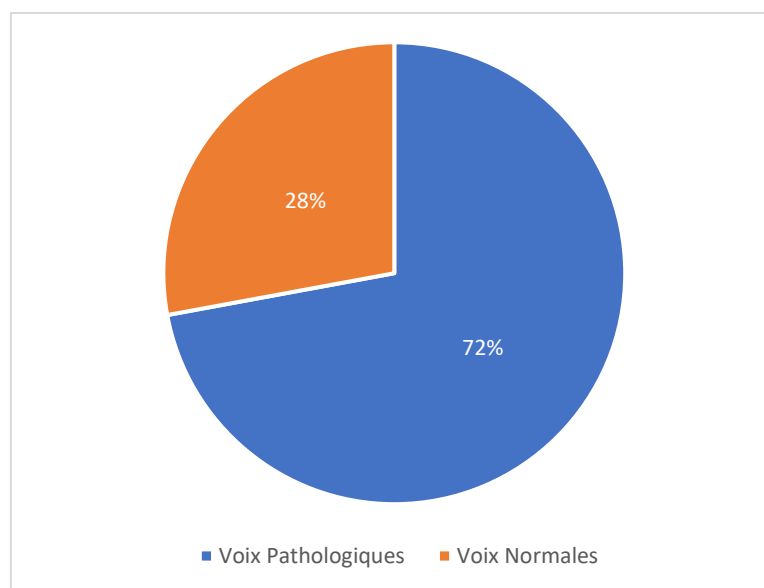


Figure 3.5 : Classification des données en pathologiques et normales

3.2 L'analyse des signaux et l'extraction de caractéristiques

Nous appliquons la transformation en ondelette discrète (DWT) à nos échantillons vocaux jusqu'au niveau souhaité, dans notre cas, jusqu'au niveau 4, comme détaillé dans le chapitre précédent de notre mémoire. La fréquence d'échantillonnage est égale à 8000 Hz.

Nous extrayons les 8 caractéristiques suivantes pour chaque échantillon vocal : l'énergie, l'écart type, la variance, la moyenne du signal et la fréquence moyenne du signal d'approximation, la fréquence dominante, le ratio des bandes de fréquences, et l'entropie spectrale du signal de détail.

3.3 Classification

La classification des voix pathologiques comprend deux étapes fondamentales : l'apprentissage initial, où un modèle est développé en se basant sur des données d'entraînement pour identifier les traits distinctifs des voix pathologiques, et une phase ultérieure de test, où ce modèle est testé sur de

nouvelles données pour évaluer son efficacité dans la classification des voix pathologiques. Ces deux phases sont essentielles pour assurer la précision et la fiabilité du système de classification.

3.3.1 Phase d'apprentissage

Dans cette phase, nous commençons par utiliser 70 % de notre ensemble de données vocales pour former notre modèle SVM. Cet ensemble de données d'apprentissage comprend des échantillons vocaux pathologiques et normaux, ainsi que les caractéristiques extraites de ces échantillons. L'objectif principal de cette phase est de permettre au modèle SVM d'apprendre à séparer de manière optimale les deux classes, à savoir les voix pathologiques et normales. Le modèle SVM ajuste ses paramètres en cherchant à maximiser la marge entre les deux classes tout en minimisant les erreurs de classification. En option, nous pouvons utiliser la validation croisée sur l'ensemble de données d'apprentissage pour optimiser les paramètres du modèle. Une fois le modèle SVM entraîné, nous évaluons sa performance sur l'ensemble de données d'apprentissage pour nous assurer qu'il a appris à discriminer efficacement les voix pathologiques et normales.

3.3.2 Phase de test

Dans cette phase, nous utilisons les 30 % restants de notre ensemble de données vocales, qui n'ont pas été utilisés lors de la phase d'apprentissage. Cet ensemble de données de test est crucial pour évaluer la capacité du modèle SVM à généraliser à de nouvelles données. Nous utilisons le modèle SVM formé précédemment pour évaluer les échantillons vocaux de l'ensemble de données de test. Le modèle classe chaque échantillon comme étant pathologique ou normal en fonction de la position de l'échantillon par rapport à l'hyperplan défini par le SVM. En analysant la matrice de confusion et en calculant la sensibilité, la spécificité et le taux de classification, nous obtenons une image complète de la performance de notre modèle SVM dans la classification des voix pathologiques, nous interprétons les résultats pour évaluer à quel point notre modèle SVM est capable de classer correctement les voix pathologiques par rapport aux voix normales sur de nouvelles données non vues lors de la formation.

Le tableau 3.1 donne un aperçu du nombre d'échantillons utilisés dans les ensembles d'apprentissage et de test.

Classe	Phase d'apprentissage	Phase de test	Total
Voix Pathologique	104	44	148
Voix Normal	41	17	58
Total	145	61	206

Tableau 3.1 : Le nombre d'exemples utilisés pour les bases d'apprentissage et de test

3.4 Matrice de confusion

La matrice de confusion est un outil fondamental en évaluation de modèle de classification, et elle est particulièrement utile dans le contexte de la classification des voix pathologiques. Elle nous permet de quantifier la performance de notre modèle en comparant ses prédictions avec les valeurs réelles des échantillons de voix.

Cette matrice trie tous les cas du modèle en catégories, en déterminant si la valeur prédite correspondait à la valeur réelle. Tous les cas dans chaque catégorie sont affichés dans la matrice. [22]

Elle comporte quatre entrées principales : TP, TN, FP et FN, organisées de la manière suivante :

	Prédiction Négative	Prédiction Positive
Etiquette Négative	Vrais Négatifs (VN)	Faux Positifs (FP)
Etiquette Positive	Faux Négative (FN)	Vrais Positifs (VP)

Tableau 3.2 : Matrice de Confusion

Où :

- **Etiquette Positive :** L'échantillon vocal est effectivement pathologique.
- **Etiquette Négative :** L'échantillon vocal est effectivement normal.
- **Prédiction Positive :** Le modèle prévoit que l'échantillon vocal est pathologique.
- **Prédiction Négative :** Le modèle prévoit que l'échantillon vocal est normal.

Vrais Positifs (TP) :

- TP représente le nombre d'échantillons qui appartiennent réellement à la classe positive (voix pathologiques) et qui ont été correctement classés comme tels par le modèle.
- Dans notre étude, TP signifie que le modèle a correctement identifié des cas de voix pathologiques.

Vrais Négatifs (TN) :

- TN représente le nombre d'échantillons qui appartiennent réellement à la classe négative (voix normales) et qui ont été correctement classés comme tels par le modèle.
- Dans notre étude, TN indique que le modèle a correctement identifié des cas de voix normales.

Faux Positifs (FP) :

- FP représente le nombre d'échantillons qui appartiennent réellement à la classe négative (voix normales), mais qui ont été incorrectement classés comme appartenant à la classe positive (voix pathologiques) par le modèle.
- FN se produit lorsque l'algorithme classe à tort un échantillon vocal sain comme étant pathologique.

Faux Négatifs (FN) :

- FN représente le nombre d'échantillons qui appartiennent réellement à la classe positive (voix pathologiques), mais qui ont été incorrectement classés comme appartenant à la classe négative (voix normales) par le modèle.
- FN se produit lorsque l'algorithme classe à tort un échantillon vocal pathologique comme étant sain.

À partir de la matrice de confusion, nous pouvons calculer plusieurs métriques d'évaluation essentielles pour évaluer la performance de notre modèle de classification des voix pathologiques.

Voici les principales métriques que nous pouvons calculer :

La sensibilité :

Exprimée en tant que taux de vrais positifs, la sensibilité mesure la capacité de notre modèle à détecter correctement les échantillons vocaux pathologiques. Elle se calcule comme suit :

$$\text{Sensibilité} = \frac{VP}{VP + FN} \times 100$$

La spécificité :

La spécificité mesure la capacité de notre modèle à distinguer correctement les échantillons vocaux normaux. Elle se calcule comme suit :

$$\text{spécificité} = \frac{VN}{VN + FP} \times 100$$

Taux de classification :

Le taux de classification mesure la précision globale du modèle. Il se calcule comme suit :

$$\text{Taux} = \frac{VN + VP}{VN + FP + VP + FN} \times 100$$

4. Résultats et discussion

4.1 Matrice de Confusion en SVM

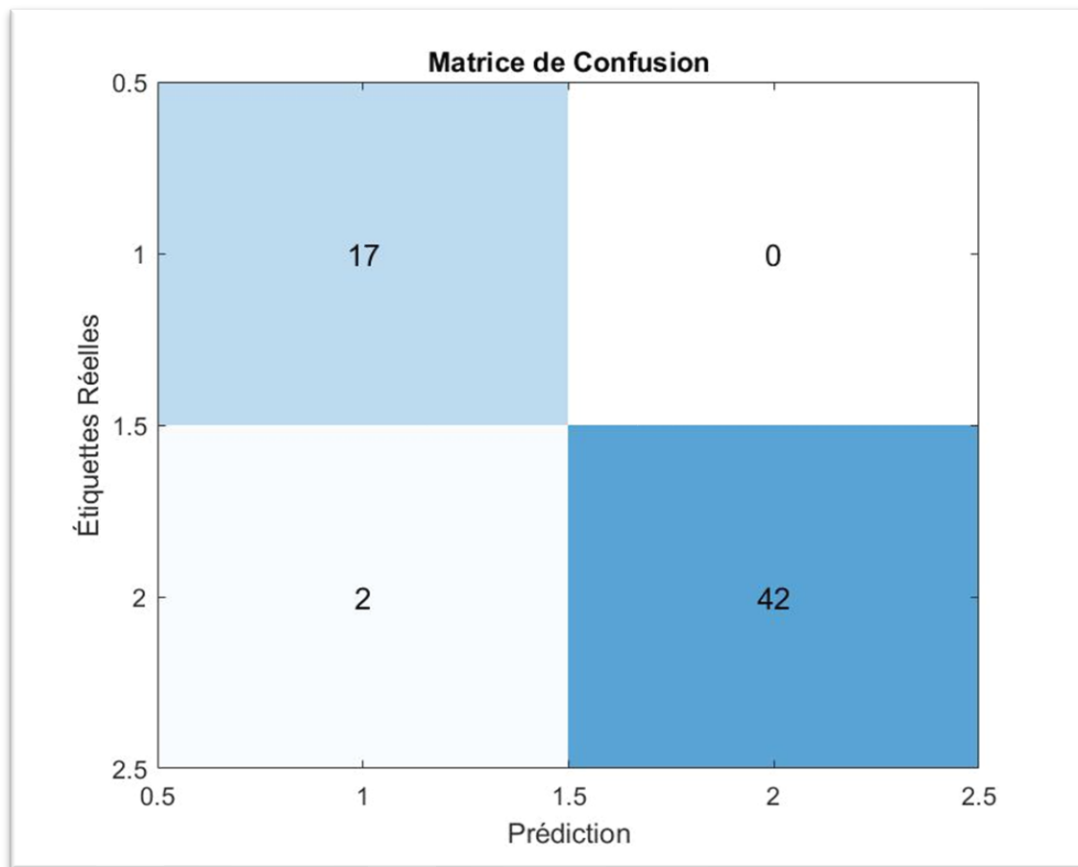


Figure 3.6 : Résultats de la matrice de confusion

- **Vrais Positifs (VP) : 42**

Ce chiffre (42) représente le nombre d'échantillons vocaux pathologiques qui ont été correctement identifiés comme pathologiques par notre modèle. En d'autres termes, ce sont les cas où notre modèle a fait une prédiction positive (pathologique) qui était correcte.

- **Faux Négatifs (FN) : 2**

Le nombre (2) de faux négatifs indique le nombre d'échantillons vocaux pathologiques qui ont été mal classés comme normaux par notre modèle. Ces échantillons étaient en réalité pathologiques, mais notre modèle les a mal interprétés.

- **Vrais Négatifs (VN) : 17**

Ce chiffre (17) représente le nombre d'échantillons vocaux normaux qui ont été correctement identifiés comme normaux par notre modèle. En d'autres termes, ce sont les cas où notre modèle a fait une prédiction négative (normal) qui était correcte.

- **Faux Positifs (FP) : 0**

Le nombre (0) de faux positifs indique que notre modèle n'a pas fait d'erreurs en classant à tort des échantillons vocaux normaux comme pathologiques. Il n'y a pas de cas où notre modèle a incorrectement prédit un échantillon vocal normal comme pathologique.

Les résultats de vrais négatifs (VN), vrais positifs (VP), faux négatifs (FN), faux positifs (FP), la sensibilité, la spécificité et le taux de classification sont présentés dans le tableau 3.3 :

Les entrées	Résultats
VN	17
FN	2
VP	42
FP	0
Sensibilité	95.45%
Spécificité	100%
Taux de classification	96.72%

Tableau 3.3 : Résultats de classification obtenus grâce au SVM

Sensibilité : 95.45% : Notre sensibilité signifie que notre modèle a correctement identifié près de 95.45% de tous les échantillons vocaux pathologiques dans notre ensemble de test. En d'autres termes, notre modèle a démontré une capacité élevée à détecter avec précision les cas pathologiques parmi les échantillons réellement pathologiques.

Spécificité : 100% : Notre spécificité indique que notre modèle a correctement identifié tous les échantillons vocaux normaux dans notre ensemble de test. Cela signifie que notre modèle n'a commis aucune erreur en identifiant les échantillons vocaux normaux, ce qui est très positif.

4.2 Taux de classification

Le taux de classification de représente la précision globale de notre modèle en termes de classification correcte. Il englobe à la fois les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs, offrant ainsi une vue d'ensemble de la performance globale de notre modèle. Notre modèle a correctement classé près de 96.72% de tous les échantillons, à la fois pathologiques et normaux, dans notre ensemble de test, ce qui démontre son efficacité globale.

4.3 Etude comparative des différents résultats des travaux réalisés :

Auteur	Année	Méthodologie employée	Taux de classification
Lotfi SALHI	2009	Variations du Pitch et LPC + RNA	85%
Nesrine BENAMMAR et Amina BOUTICHE	2016	TOC + SVM	95.17%
I. HAMMAMI	2019	TOD + SVM	93.1%
G. MAROUA et K. RAHMA	2021	FFT et MFCC + SVM	97%
M. BOUDJELLABA et N. BOUDJERIDA	2021	MFCC et HNR et variations du Pitch + SVM	95,70%
Notre travail	2023	TOD + SVM	96.72%

Tableau 3.4 : Etude comparative de notre travail avec des autres réalisés

5. Conclusion

En résumé, le chapitre 3 de notre mémoire se concentre sur la classification des voix pathologiques à l'aide d'un modèle SVM. Nos résultats sont très prometteurs, avec un taux de classification de 96.72% obtenu lors de l'évaluation de notre modèle sur la base de données VOICED. Ces performances confirment l'efficacité de notre approche en matière de détection des pathologies vocales.

De plus, notre travail s'intègre harmonieusement dans le contexte de la recherche existante sur la détection des affections vocales. Nous avons mené une comparaison minutieuse avec d'autres travaux similaires, mettant ainsi en évidence la pertinence de notre méthode et sa contribution significative à ce domaine.

En conclusion, ce chapitre souligne l'importance cruciale des avancées technologiques, notamment dans le domaine de l'apprentissage automatique tel que le SVM, pour l'amélioration du diagnostic des maladies vocales.

Conclusion Général

Ce mémoire a été dédié à la classification des voix pathologiques, une discipline cruciale pour la santé vocale. À travers une exploration approfondie des fondements du signal vocal, une méthodologie intégrant le traitement du signal et la classification, ainsi que la présentation de résultats prometteurs obtenus grâce à un modèle SVM, nous avons poursuivi l'objectif majeur d'améliorer le diagnostic et la prise en charge des troubles vocaux.

Nos investigations ont mis en lumière la complexité de la production de la parole humaine et la diversité des sons qu'elle engendre, qu'ils soient sonores ou non sonores. Nous avons souligné l'importance de comprendre les altérations vocales et présenté des méthodes d'identification des troubles vocaux, mettant en avant les paramètres de la voix en tant qu'éléments clés pour l'analyse et la classification des voix pathologiques.

La méthodologie que nous avons développée, en utilisant la base de données VOICED et le modèle SVM, s'est avérée efficace, comme en témoigne le taux de classification élevé de 96.72% que nous avons obtenu. Ces résultats prometteurs confirment la pertinence de notre approche dans la détection des pathologies vocales.

Toutefois, il est essentiel de noter que notre travail ne s'arrête pas ici. Des perspectives d'amélioration subsistent, notamment l'utilisation de bases de données plus riches et diversifiées pour affiner nos modèles. De plus, l'exploration de techniques avancées d'apprentissage automatique et l'intégration de l'intelligence artificielle pourraient ouvrir de nouvelles voies dans le domaine de la classification des voix pathologiques.

En fin de compte, ce mémoire représente une contribution significative à l'amélioration du diagnostic et de la prise en charge des troubles vocaux. Il souligne la nécessité d'accorder une attention accrue à la santé vocale et de continuer à investir dans des méthodes de diagnostic plus précises et accessibles. En mettant l'accent sur l'efficacité de la classification des voix pathologiques, nous ouvrons des perspectives porteuses pour les soins et la qualité de vie des individus touchés.

Nous espérons que ce travail stimulera davantage la recherche dans le domaine de la santé vocale et contribuera à offrir un avenir plus prometteur aux personnes souffrant de troubles vocaux.

Bibliographie

- [1] Kasuya.H, S.Ogawa, « AN acoustic analysis of pathological voice and its application to the evolution of laryngeal pathology », speech communication, 5-2. 1986.
- [2] F.Le Huche, A.Allali « La Voix Tome1 Anatomie et physiologie des organes de la voix et de la parole,4éme édition », 2010,204 pages.
- [3] Giovanni A., Yu P., Révis J., Guarella MD., Teston B., Ouaknine M. (2006) « Analyse objective des dysphonies avec l'appareillage EVA. Etat des lieux », Revue Oto-rhinolaryngologie Française, 90, p3 183-192.
- [4] GIOVANNI, A., HEIM, C., DEMOLIN, D., and TRIGLIA, J.M. (2000). « Estimate subglottic pressure in normal and dysphonic subjects » Otol. Rhinol. Laryngol, 109, 500-504.
- [5] <http://scgwww.epfl.ch/courses>
- [6] Castellengo(M), Ronbeau, Valette(C) « Study of the acoustical phenomena characteristic of the transition between chest voice and falsetto proceedings of SMCA références », Stockholm,1983, P 113-117.
- [7] <https://clemedicine.com/1-production-de-la-parole>
- [8] Beigi Homayoon, « Fundamentals of Speaker Recognition Springer », 2011.
- [9] Rudzicz F., « Articulatory Knowledge in the Recognition of Dysarthric Speech, IEEE Transactions on Audio, Speech, and Language Processing », vol. 19, No. 4, 2011.
- [10] Clinique d'orthophonie de Montréal, « Traitement des troubles de la voix », 2016.
- [11] S. Fraj, « Synthèse des voix pathologiques », Thèse de Doctorat, Université Libre de Bruxelles, 2010.
- [12] Institut Numérique., « L'évaluation subjective et objective de la voix », 16/05/2013.
- [13] T. ABID, « Analyse du signal ECG par les ondelettes », Annaba, Algérie, 2008.
- [14] Y. Karima & B. Chaima, « Classification des voix pathologiques », Université Aboubakr Belkaïd– Tlemcen, Algérie, 2023

- [15] J. I. Godino-Llorente, P. Gómez-Vilda, N. Sáenz-Lechón, M. Blanco-Velasco, F. CruzRoldán, Ferrer-Ballester and M. Angel, « Support vector machines applied to the detection of voice disorders," in *Nonlinear Analyses and Algorithms for Speech Processing* », 2005.
- [16] Nesrine BENAMMAR Et Amina BOUTICHE, « Identification des troubles de la voix par l'analyse temps-fréquence et les machines à vecteurs de support », projet de fin d'études, 2016.
- [17] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [18] Lotfi SALHI, Haytham BELGASSEM et Adnène CHERIF, « Identification Des Voix Pathologiques Par Réseaux De Neurones »,5th International Conference: Sciences of Electronic, March 22-26, 2009 – TUNISIA.
- [19] I. Hammami, L. Salhi, S. Labidi, « Voice Pathologies Classification and Detection Using EMD-DWT, Analysis Based on Higher Order Statistic Features », University of Tunis El Manar, Higher Institute of Medical Technologies of Tunis, Research Laboratory of Biophysics and Medical Technologies, 1006, Tunis, Tunisia.
- [20] G. MAROUA et K. RAHMA, « Développement d'un système de détection de la voix pathologique », l'UNIVERSITÉ BADJI MOKHTAR, ANNABA,2020
- [21] M. BOUDJELLABA et N. BOUDJERIDA, « Identification des troubles de la voix à l'aide des techniques d'apprentissage automatique », l'Université Mohammed Seddik Ben Yahia-Jijel, Algérie, 2021.
- [22] R. Abdelaziz et Z. Nab, « Système d'aide à la décision pour le diagnostic de la maladie de Parkinson à partir de la voix », Mémoire de Master II, Université de Blida, Algerie.