

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي



UNIVERSITÉ BADJI MOKHTAR - ANNABA  
BADJI MOKHTAR – ANNABA UNIVERSITY

جامعة باجي مختار – عنابة

Faculté : Sciences de L'ingénierat  
Département : Electronique  
Domaine : Sciences et Techniques  
Filière : Télécommunications  
Spécialité : Systèmes de Télécommunications

## Mémoire

Présenté en vue de l'obtention du Diplôme de Master

**Thème : Développement d'un système de détection de la voix pathologique**

Présenté par : *Ghenay Maroua*  
*Khelifi Rahma*

Encadrant : *Amara Fethi* Docteur UBMA

### Jury de Soutenance :

N. Kouadria	Docteur	UBMA	Président
Amara Fethi	Docteur	UBMA	Encadrant
M. Fezari	Professeur	UBMA	Examineur

Année Universitaire : 2020/2021

## ملخص

يتمثل العمل الذي نقدمه في هذه المذكرة في تطوير نظام يسمح بالكشف التلقائي عن الصوت المرضي ، وهما الأكثر شيوعاً في مجال التعرف على السماعات SVM ومصنف MFCCs يعتمد نظامنا على أكثر المعلمات الصوتية شيوعاً ، واكتشاف الصوت، لتحسين معدل التعرف على الصوت المرضي.

القرار النهائي هو ما إذا (Support Vector Machines -SVMs): يصنف هذا النظام الأصوات باستخدام آلات المتجهات الداعمة) بالإنجليزية Hospital Universitario " و "Saarbrücken Voice Database -SVD-" كان صوتاً سليماً أم مرضياً. تم استخدام قاعدتي بيانات ، المعدل (TBC) في الجزء التجريبي. مقاييس الأداء المستخدمة في هذه الدراسة هي: معدل التصنيف الجيد "Príncipe de Asturias.HUPA" ، الدقة، الخصوصية، الحساسية. سمحت لنا النتائج التي تم الحصول عليها بإعطاء (TFN) ، المعدل الخاطئ السلبي (TFP) الخاطئ الإيجابي استنتاجات مهمة حول قواعد بيانات الأصوات المرضية.

**كلمات المفتاح:** الكشف عن الأصوات المرضية، SVM, MFCC، تحديد الاضطرابات الصوتية.

## Abstract

The presented work consists in the development of a system that allows the automatic detection of pathological voice.

Our system is based on the most popular acoustic parameter MFCCs and SVM classifier, which are the most common in the field of speaker recognition and voice detection, to improve the recognition rate of pathological voice.

This system classifies voice sounds using Support Vector Machines (SVMs). The final decision is whether the voice is healthy or pathological. The databases "Saarbrücken Voice Database -SVD-" and "Hospital Universitario Príncipe de Asturias. HUPA" were used in the experimental part. The performance measures used in this study are: True Positive Rate (TPR), True Negative Rate (TNR), Accuracy, Specificity, Sensitivity. The results obtained allowed us to give important conclusions on the pathological voice databases.

**Key words:** Pathological voice, MFCC, SVM, voice disorder detection, pattern recognition.

# Résumé

Le travail que nous présentons dans ce mémoire consiste au développement d'un système qui permet la détection automatique de la voix pathologique.

Notre système est basé sur le paramètre acoustique le plus populaire MFCCs et le classificateur SVM, qui sont les plus courants dans le domaine de reconnaissance de locuteur et la détection de la voix, pour améliorer le taux de reconnaissance de la voix pathologique.

Ce système classe les sons de la voix en utilisant les machines à vecteurs de supports (En anglais : Support Vector Machines -SVMs). La décision finale consiste à indiquer s'il s'agit d'une voix saine ou pathologique. Les bases de données «Saarbrücken Voice Database -SVD-» et « Hospital Universitario Príncipe de Asturias.HUPA » ont été utilisées dans la partie expérimentale. Les mesures de performances utilisées dans cette étude sont : taux de vrais positifs (TVP), taux de vrais négatifs (TVN), La précision, La spécificité, La sensibilité. Les résultats obtenus nous ont permis de donner des conclusions importantes sur les bases de données des voix pathologiques.

**Mots clé :** Voix Pathologique, MFCC, SVM, La détection des troubles de la voix, La reconnaissance des formes.

# Dédicace

---

*Ma mère, la plus tendre, sincère et la plus pure au monde*

*A mon père, source d'amour ....*

*Sans vous, je n'aurais certainement pas fait mes études, Aucun hommage ne pourrait être à la hauteur de vos Sacrifices.*

*C'est le moment de plaisir mes frères et ma petite sœur et ma tante, pour tout l'encouragement et le soutien que j'ai trouvé auprès de lui.*

*A mes amies d'enfance Nour et Roumaïssa Pour tous moments d'encouragement et d'amour.*

*A mon binôme Maroua Sources de puissance ....*

---

***Rahma***

# *Dédicace*



*Je dédie cette thèse à ma petite famille que j'aime infiniment et inconditionnellement.*

*A mes chers parents,*

*Aucune dédicace ne saurait être assez éloquente pour exprimer ce que vous mérités car rien au monde ne vaut vos efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de vos sacrifices que vous avez consentis pour mon éducation et ma formation.*

*A mon cher frère **Mohamed Lamine**,*

*A ma chère grand sœur **Amel** et son époux et leur fils **Mohamed Djoud***

*A mes chères sœur **Hanen, Asma** et son époux*

*Aux mes amours, les enfants de ma sœur, **Noor El Yaqin, Mohamed Wael** et **Wasim**,*

*Pour tout l'amour et soutien que vous m'as offert.*

*A mon binôme,*

*Pour tous les moments inoubliables que j'ai passés avec vous.*

*A tous ceux que j'aime et qui m'aiment.*

*Je dédie ce travail espérant avoir répondu à leurs souhaits de me voir réussir.*



***Maroua***

# ***Remerciements***

*Nous tenons à remercier en premier lieu Dieu le tout puissant qui nous a dotés de toute la force nécessaire à l'aboutissement de ce mémoire.*

*Nos remerciements vont aussi à notre encadreur le Dr AMARA qui nous a fait bénéficier de ses nombreuses et riches connaissances durant toute la durée de ce travail.*

*Nous remercions les membres du jury ; qui nous honorent par la lecture de notre mémoire et de leur présence, le jour de notre soutenance.*

*Sans Oublier tous ceux qui ont contribué, de près ou de loin à mener à terme ce travail. Notamment nos enseignants qui nous ont instruits Tout au long de notre long parcours.*

# Liste des Tableaux

<b>TABLEAU III. 1 :</b> MATRICE DE CONFUSION.....	37
<b>TABLEAU III. 2 :</b> LA COMPOSITION SE BASE SVD.....	39
<b>TABLEAU III. 3 :</b> RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES SVD.....	40
<b>TABLEAU III. 4 :</b> RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES SVD APRES LDA .....	41
<b>TABLEAU III. 5 :</b> LA COMPOSITION DE LA BASE HUPA.....	43
<b>TABLEAU III. 6 :</b> RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES HUPA.....	43
<b>TABLEAU III. 7:</b> RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES HUPA APRES LDA .....	43
<b>TABLEAU III. 8:</b> MESURES DE PERFORMANCE DE LA DETECTION DE LA VOIX PATHOLOGIQUE AVEC MFCC ET SVM AVEC LDA (BASE HUPA ET SVD).....	45

# LISTE DES FIGURES

<b>FIGURE I. 1 : VUE D'ENSEMBLE DU L'APPAREIL VOCALE [20]</b> .....	5
<b>FIGURE I. 2 : LE LARYNX, D'APRES [25]</b> .....	7
<b>FIGURE I. 3 : LES PLIS VOCAUX. D'APRES [BRIN-HENRY ET AL. (2004), ENCART P.21] [4]</b> .....	7
<b>FIGURE I. 4 : NODULES SUR LES CORDES VOCALES</b> .....	13
<b>FIGURE I. 5 : EXEMPLE DE KYSTE SUR LES CORDES VOCALES</b> .....	14
<b>FIGUREII. 1 : SCHEMA DE LA RECONNAISSANCE VOCALE SIMPLIFIE</b> .....	17
<b>FIGUREII. 2 : LES ETAPES DE PRETRAITEMENT</b> .....	18
<b>FIGUREII. 3 : COMMUNICATION ENTRE DEUX PERSONNE</b> .....	19
<b>FIGUREII. 4 : REPRESENTATION TEMPORELLE D'UNE FENETRE DE HAMMING</b> .....	21
<b>FIGUREII. 5 : PROCESSUS DE CREATION DES COEFFICIENTS CEPSTRAUX</b> .....	22
<b>FIGUREII. 6 : CEPSTRUM D'UN VOIX NORMALE D'UN FEMME</b> .....	23
<b>FIGUREII. 7 : CEPSTRUM D'UNE VOIX NORMAL POUR L'HOMME</b> .....	24
<b>FIGUREII. 8 : L'HYPERPLAN H QUI SEPRE LES DEUX ENSEMBLES DE POINTS</b> .....	26
<b>FIGUREII. 9 : L'HYPERPLAN OPTIMAL, VECTEURS DE SUPPORT ET MARGE MAXIMALE</b> .....	27
<b>FIGUREII. 10 : MEILLEUR HYPERPLAN SEPARATEUR</b> .....	28
<b>FIGUREII. 11 : LES CAS LINEAIREMENT ET NON LINEAIREMENT SEPARABLE</b> .....	28
<b>FIGUREII. 12 : EXEMPLE GRAPHIQUE DES DONNEES LINEAIREMENT SEPARABLES</b> .....	29
<b>FIGUREII. 13 : ESPACE DE PROJECTION DES DONNEES NON LINEAIREMENT SEPARABLE</b> .....	31
<b>FIGUREIII. 1 : SIGNAL DE VOIX NORMALE</b> .....	34
<b>FIGUREIII. 2 : SIGNAL DE VOIX PATHOLOGIE</b> .....	34
<b>FIGUREIII. 3 : EXEMPLE DE COEFFICIENTS MFCC ET LEUR AMPLITUDE VOIX NORMALE D'UNE FEMME</b> .....	35
<b>FIGUREIII. 4 : EXEMPLE DE COEFFICIENTS MFCC ET LEUR AMPLITUDE VOIX NORMALE D'UNE FEMME</b> .....	36





# **LISTE DES ABREVIATIONS**

PLP Prédiction linéaire perceptive

LPC Linear predictive coefficients

MFCC Mel frequency cepstral coefficients

SVM support vectors machine

F0 Fréquence fondamentale

SVD Saarbrücken voice database

LPCC Linear predictive cepstral coefficients

FFT Fast fourier transform

IFFT Inverse fast fourier transform

HUPA Hospital universitario príncipe de Asturias

MEEI Massachusetts eye and ear infirmary

EGG Electroglottogramme

VN Vrai négatif

VP Vrai positif

FN Faux négatif

FP Faux positif

TFP Taux de faux positifs

TFN Taux de faux négatifs

TBC Taux de bonne classification

# Table des matières

Résumé

Dédicace

Remercîments

Liste des tableaux

Liste des figures

Les symboles

Table des matières

Introduction générale.....	1
Chapitre I Généralités sur les voix pathologiques.....	4
I. 1. Introduction.....	4
I. 2. Structures Anatomiques de l'appareil vocal.....	4
I. 2. 1. L'appareil respiratoire .....	5
I. 2. 2. L'appareil phonatoire .....	6
I. 2. 3. Les Cavités de résonance .....	8
I. 3. Production de la parole .....	9
I. 4. La fréquence fondamentale .....	9
I. 5. Traits acoustiques du signal de parole.....	10
I. 6. Propriétés spécifiques du signal vocal .....	11
I. 7. Troubles de la voix .....	12
I. 7. 1. Compléments sur les dysphonies .....	12
I. 8. Pathologies du larynx.....	13
I. 8. 1. Dysphonies organiques .....	13
I. 8. 2. Dysphonies fonctionnelles .....	14
I. 9. Conclusion .....	15
Chapitre II System de reconnaissance.....	16
II. 1. Introduction .....	16

II. 2. Le système d'identification .....	16
II. 3. Les bases de données de voix pathologiques .....	17
II. 4. Prétraitement du signal .....	18
II. 4. 1. Préalimentation du signal vocale .....	18
II. 4. 2. Sous échantillonnage .....	20
II. 4. 3. Fenêtrage du signal.....	20
II. 5. Extraction des paramètres .....	21
II. 5. 1. Le codage MFCC .....	21
II. 5. 2. Codage cepstrum .....	23
II. 6. Méthodes d'apprentissage des données acoustiques .....	24
II. 6. 1. Type de classification.....	24
II. 6. 2. Apprentissage de modèle SVM.....	25
II. 6. Conclusion.....	31
Chapitre III Expérimentations et résultats .....	32
III. 1. Introduction.....	32
III. 2. Méthodologie .....	32
III. 2. 1. Corpus de données.....	32
III. 2. 2. Prétraitement .....	34
III. 2. 3. Extraction des paramètres.....	34
III. 3. Résultats expérimentaux .....	36
III. 3. 1. Matrice de confusion .....	36
III. 3. 2. Classification binaire : Application a des bases de données standards .....	38
III. 4. Conclusion .....	48
Conclusion générale .....	49
Bibliographie	

## Introduction générale

La parole est d'une grande importance car la race humaine exprime verbalement ses sentiments, ses opinions, ses points de vue et ses concepts. La production de la parole comprend la prononciation, la voix et la fluidité [1, 2]

La production de la parole est un acte moteur complexe qui implique un grand nombre de muscles, de variables physiologiques et un contrôle neurologique impliquant différentes régions corticales et sous corticales [3]. « La voix humaine contient, elle aussi, un nombre étonnant de signaux : dès la première phrase au téléphone, on sait à qui l'on a affaire, son sexe, son âge, sa culture, son humeur agressive, abattue ou érotique, et même son niveau social ». Cette phrase, empruntée à Boris Cyrulnik, souligne que le but principal de la parole dans la communication est de transmettre un message linguistique. En plus de contenu linguistique, le discours contient également une multitude d'informations sur la langue et le dialecte, ainsi que sur le sexe, l'âge, l'humeur et la santé du locuteur.

D'après Brin-Henry et al. (2004)[4], la voix se définit, de façon générale, comme le « souffle sonorisé par le larynx, amplifié et modulé par les cavités de résonance sus-laryngées, ayant toutes les caractéristiques du son : hauteur ou fréquence ou registre (voix haute, aiguë, ou basse, grave), intensité (voix forte ou faible), timbre ou fourniture harmonique (voix sourde ou stridente, claire ou sombre), modulation (voix monocorde ou modulée), rythme, débit ou vitesse d'émission.». Nous utilisons notre voix en permanence dans la vie quotidienne, et nous pouvons jouer sur ses modalités dans les situations de communication pour tour à tour convaincre, attendrir, séduire, émouvoir, informer... Cependant, en raison de son utilisation constante, il est facile d'oublier la fragilité de cet outil. Les orthophonistes reçoivent de nombreuses plaintes concernant des changements dans leur voix, l'apparition d'un enrouement et une gêne dans la gorge. Il appartient donc à l'orthophoniste de construire un plan de traitement adapté à la plainte du patient. Cependant, la première étape d'évaluation est nécessaire. En utilisant les différentes méthodes d'évaluation vocale existantes, l'orthophoniste réalisera une évaluation vocale aussi précise que possible pour identifier les troubles vocaux et fonctionnels de la voix du patient, et les corrélés avec sa plainte.

Il y a eu plusieurs tentatives réussies dans le développement de systèmes qui peuvent analyser, classer et reconnaître les signaux vocaux. Le matériel et les logiciels qui ont été développés pour de telles tâches ont été appliqués dans divers domaines tels que les soins de santé, les secteurs gouvernementaux et l'agriculture. La reconnaissance du locuteur est la capacité d'un logiciel ou d'un

matériel à recevoir un signal vocal, à identifier le locuteur présent dans le signal vocal et à reconnaître le locuteur par la suite [5].

Les pathologies vocales sont dues à des infections, à des causes physiologiques et psychogènes et à une mauvaise utilisation de la voix qui prévaut dans des professions telles que les chanteurs, les enseignants et les représentants du service client [6], [7], et la détection automatique et la classification de la voix pathologique est un domaine d'actualité et toujours exploré par la communauté des chercheurs [8], [9], [10], [11]. La détection automatique de la pathologie vocale permet une évaluation objective et une intervention précoce pour le diagnostic. Un système de détection d'une pathologie de la parole typique se compose de trois étapes principales : la première étape est le traitement acoustique du signal vocal d'entrée, le signal vocal doit être traité pour éliminer le bruit avant que les caractéristiques vocales importantes puissent être extraites [12]. La deuxième étape est l'extraction de caractéristiques qui est la partie la plus pertinente de la reconnaissance du locuteur : Les méthodes d'extraction de caractéristiques produisent généralement un vecteur de propriété multidimensionnel pour chaque signal de parole [13]. Une grande variété d'options sont disponibles pour représenter de manière paramétrique le signal vocal pour le processus de reconnaissance, telles que la prédiction linéaire perceptive (PLP), le codage à prédiction linéaire (LPC) et les coefficients de fréquence Mel Cepstre (MFCC). Le MFCC est le plus connu et le plus courant [14, 15]. La troisième étape est le classificateur (i.e. décision normale versus pathologique). La présente étude se concentre sur la première phase [16].

L'objectif de notre mémoire est la discrimination entre la voix pathologique et la voix normale en utilisant un système de détection automatique. L'avantage principal de cette méthode est qu'elle est non invasive, rapide et non coûteuse. Notre système est basé sur le paramètre acoustique le plus populaire MFCCs et la classification des pathologies de la voix par SVM (Support Vectors Machine).

Le manuel est organisé comme suit :

Le premier chapitre sera consacré à la présentation des notions de base sur le mécanisme de production de la parole, les différents troubles vocaux (morphologiques ou neurologiques) et les paramètres de la voix qui seront utilisés.

Dans le deuxième chapitre, sera réservé aux le système d'identification automatisé ; une présentation détaillée de ses étapes et son mécanisme d'action ; la méthode de classification SVM, ses équations, le choix de ses paramètres et ses variantes.

Dans le troisième chapitre, nous présenterons les bases de données des fichiers de parole pathologique utilisées, les conditions d'implémentation, les mesures de performances à utiliser, les résultats de la détection et la classification obtenus et des discussions.

Enfin, nous terminerons ce document par une conclusion générale.

# **CHAPITRE I**

## **Généralités sur les voix pathologiques**



## 1. Introduction

La parole est l'une des formes les plus importantes de la communication humaine. En effet, La voix est considérée comme l'un des outils professionnels importants utilisés pour de multiples fonctions. Par conséquent, une utilisation intensive ou abusive de la voix peut provoquer des troubles de la voix et affecter la santé au travail, la vie sociale et les activités professionnelles en général.

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 [3].

Nous avons distingué trois systèmes qui contribuent à la production de la parole : le système sous-glottique (respiratoire), le système glottique (phonatoire) et le système supra-glottique (articulations).

En phonétique, le son audible de la parole est produit par la vibration des cordes vocales. Cette vibration peut être mesurée avec un microphone et permet d'exprimer la parole sous la forme d'un signal audio qui peut être analysé à l'aide d'outils de traitement du signal. Par conséquent, l'analyse du signal de parole peut donner une image de l'état du larynx (cordes vocales). Le but de l'analyse du signal audio est d'extraire des preuves pertinentes pour caractériser le son afin d'illustrer l'état du larynx du locuteur. Au cours des dernières décennies, l'analyse acoustique des signaux de parole pour caractériser et évaluer les troubles de la parole est devenue l'objet de nombreuses études [17], [18], [19].

Dans ce chapitre, nous présentons et expliquons la structure anatomique du système vocal et son mécanisme de génération. Ensuite, nous présentons les traits acoustiques et les propriétés du signal vocal. Enfin nous terminons le chapitre par un aperçu sur les troubles de la voix.

## I. 2. Structures anatomiques de l'appareil vocal

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.

L'appareil vocal s'étudie classiquement en trois parties [20] :

- La soufflerie ou appareil respiratoire : les poumons et la trachée.
- Le vibreur ou appareil phonatoire : le larynx et les cordes vocal (Partie Glottique).
- Les résonateurs : comprend le pharynx (niveau articulatoire), les cavités orales et nasales, les sinus.

La figure I.1 donne une vue d'ensemble de ces trois parties.

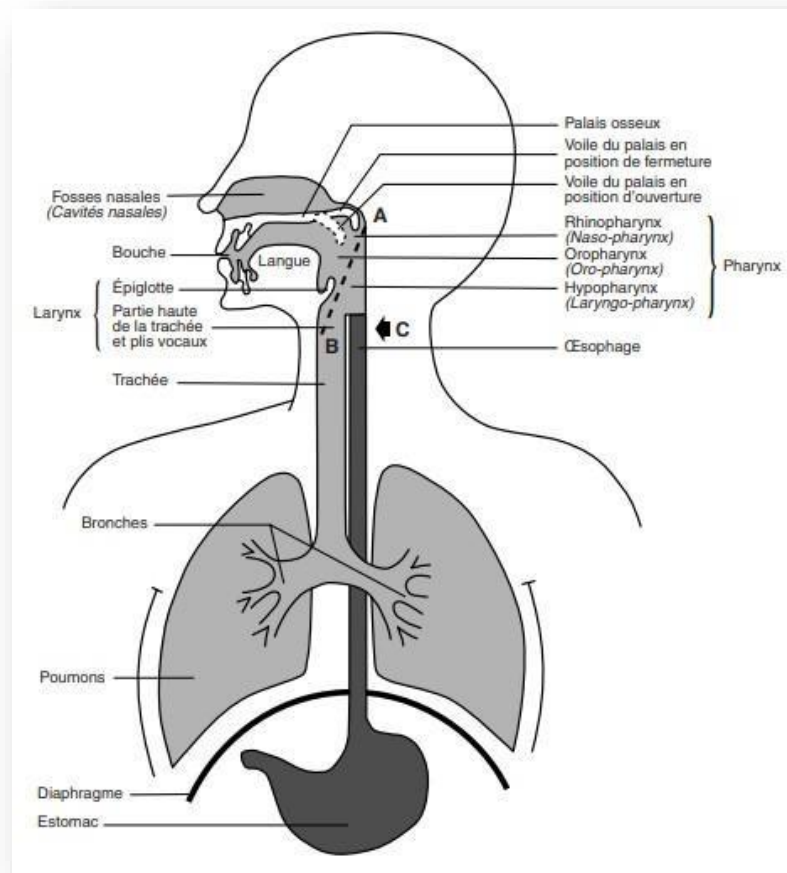


FIGURE I. 1 : VUE D'ENSEMBLE DU L'APPAREIL VOCALE [20]

### I. 2. 1. L'appareil respiratoire

Selon Le Huche et al. [20], le son peut être considéré comme une expiration sonore. Par conséquent, lorsque nous respirons au repos, les muscles inspiratoires remplissent les poumons puis se vident en relaxant ces muscles. L'expiration est alors passive. Cependant, lorsque nous parlons, l'expiration est active car les muscles d'expiration vident les poumons. Ensuite, nous disons "voix et respire".

Le système respiratoire regroupe l'ensemble des organes qui interviennent dans la respiration. Ce système comprend la trachée et les poumons. L'homme possède deux poumons, gauche et droit, deux organes thoraciques, séparés l'un de l'autre par le médiastin, médialement. Ils sont posés sur le diaphragme et protégés par la cage thoracique en avant, en dehors et en arrière, sauf au niveau de leur sommet, car ils dépassent le bord supérieur de la première côte, et montent même jusqu'au-dessus de la clavicule, à la base du cou, dans le creux supra claviculaire [21].

## I. 2. 2. L'appareil phonatoire

L'appareil phonatoire humain est formé de différentes parties qui peuvent nous sembler complexes. Cependant, il peut facilement être assimilé, et même souvent représenté comme un système composé simplement d'une source vibrante et d'un filtre (résultant du conduit vocal qui est formé d'une cavité résonante complexe) [22], [23], [24].

### a. Le larynx :

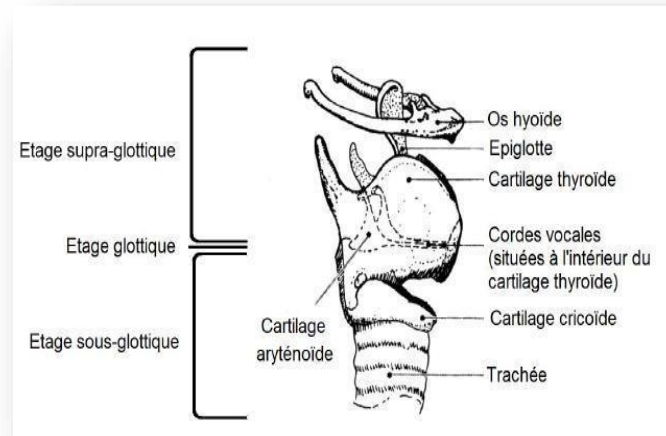
Le larynx représente le canal musculo-cartilagineux situé dans la partie médiane et antérieure du cou (**Burryat et Martin, 2005**).

Le larynx est un conduit cylindrique rétréci en son milieu, au niveau des plis vocaux, ayant une armature fibro-musculo-cartilagineuse recouverte de muqueuse. Il est situé en arrière du pharynx, à l'extrémité supérieure du tube trachéal et il est surplombé par la base de langue. C'est l'organe principal de la voix. Mais ce n'est pour lui qu'une fonction « secondaire » : sa fonction première est celle d'un sphincter permettant l'obturation de la trachée [20]

Le larynx est divisé en trois parties par les cordes vocales et les bandes ventriculaires comme la Figure I.2 montre :

- la région sus-glottique (ou supra-glottique).
- l'étage glottique (au niveau des plis vocaux).
- la région sous-glottique (ou infra-glottique)

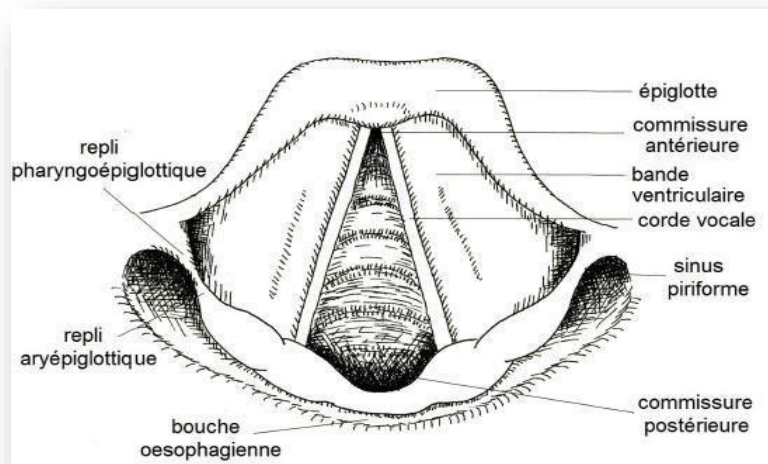
La figure I.2 donne une vue d'ensemble de ces trois étages



**FIGURE I. 2 : LE LARYNX, D'APRES [25]**

**b. Les plis vocaux :**

Les plis vocaux sont situés à l'intérieur du larynx, au niveau de la glotte, qui est l'espace laissé aux plis vocaux lorsqu'ils se séparent les uns des autres. Ce sont deux plis muqueux horizontaux sous les bandes ventriculaires. La figure I.3 montre une représentation simplifiée de Les plis vocaux humains [4].



**FIGURE I. 3 : LES PLIS VOCAUX [4]**

Les cordes vocales agissent comme une valve, bloquant le passage de l'air, grâce à leur position médiane, leur permettant d'offrir une certaine résistance au flux d'air.

### I. 2. 3. Les cavités de résonance

Ce sont la cavité buccale, les cavités naso-sinusiennes, et le pharynx, appelés cavités supra laryngées. Ils vont moduler le son laryngé initial et ainsi produire le timbre de la voix.

#### a. Le pharynx :

Le pharynx est un conduit musculo-membraneux qui s'étend verticalement depuis le larynx en bas jusqu'à la cavité buccale et aux fosses nasales en haut. Il est divisé en trois parties d'après Le Huche et al [20] :

- le rhinopharynx en haut qui va de la base du crâne au voile du palais,
- l'oropharynx au milieu qui va du voile du palais à l'épiglotte,
- l'hypo pharynx en bas, qui va de l'épiglotte à l'œsophage.

En ce qui concerne la phonation, le son de base émis par la vibration des cordes vocales est modulé par le filtre acoustique que constitue le conduit vocal, constitué par le pharynx, la cavité buccale, la cavité nasale et la cavité labiale. En effet, lorsque le voile du palais est relevé, l'air ne peut pas passer de l'oropharynx au rhinopharynx, les sons ne peuvent donc pas être transformés en voyelles ou en consonnes nasales. En revanche, si le voile du palais est baissé, l'oropharynx et le rhinopharynx communiquent ensemble, laissant passer l'air et permettant de les produire [26].

#### b. La cavité annexe (buccale et nasale) :

Au même titre que le pharynx, les cavités buccale et nasale (ou fosses nasales) sont des résonateurs très impliqués dans la production de la voix.

- Le Huche et al. (2001)[20] décrivent la bouche comme étant le résonateur principal (avec le pharynx) mais aussi l'articulateur principal. En effet, La cavité buccale est divisée en deux parties par l'interposition des arcades alvéolo-dentaires, qui permettent de moduler les sons de façon fine :
  - Le vestibule oral, en avant
  - La cavité orale propre, en arrière [27].
- Les cavités nasales ou fosses nasales sont deux espaces séparés par le septum nasal. Une résonance nasale est très caractéristique (nasillement). L'air passe par le nez lorsque le voile du palais (prolongement musculaire du palais osseux) est rabaisé : passage oro-nasal ouvert. Comme nous l'avons dit, la cavité nasale communique avec le pharynx par le rhinopharynx.

Cette communication est interrompue lorsque l'élévation du voile vient empêcher le passage de l'air [20], [28].

### **I. 3. Production de la parole**

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 [29] [30].

La parole est produite par l'excitation de la cavité résonante. Le système respiratoire fournit de l'énergie pour la production du son, qui est l'air. L'air fourni par les poumons est délivré à la source du résonateur à un débit clair à travers un appareil vocal [31]. Le son résultant peut être classé comme voisé ou non voisé, selon que l'air émis a fait vibrer les cordes vocales ou non. Quand l'air sous pression arrive au niveau des cordes vocales : si sont étirées, le flux d'air les fait vibrer, ce qui s'appelle sons voisés. Si les cordes vocales se relâchent, l'air passe librement à travers le larynx sans vibrer. Ensuite, nous obtenons une voix non voisée, qui est la deuxième position [26], [32], [33]. L'air, qu'il vibre ou non, continuera à traverser le canal sonore, puis se répandra dans l'atmosphère. Les différentes variantes possibles du corps de résonateur utilisé pour le traitement du son offrent une très large gamme de tonalités. Par conséquent, le conduit vocal est considéré comme un filtre pour différentes sources de parole, telles que les vibrations des cordes vocales ou les turbulences de l'air causées par la contraction du conduit vocal [33], [26], [34].

### **I. 4. La Fréquence fondamentale**

La fréquence fondamentale ( $F_0$ ) est la fréquence de vibration des cordes vocales. Dans le domaine temporel, c'est la période d'un signal voisé à un instant donné. Cette fréquence fondamentale nous donne un indice sur le pitch du signal vocal.  $F_0$  représente un indicateur émotionnel et est considérée comme étant la quantité à estimer par un algorithme de détection du pitch.

Pour le signal de parole, sa fréquence fondamentale n'est rien d'autre que la fréquence du cycle d'ouverture/fermeture des cordes vocales, déterminée par la tension des muscles qui contrôlent celles-ci.

La plage de variation moyenne de cette fréquence varie d'un locuteur à un autre en fonction de son âge et de son sexe. Elle s'étend approximativement de 80 à 200 Hz chez les hommes, de 150 à 450

Hz chez les femmes, et de 200 à 600 Hz chez les enfants [35]. La détection de F0 joue un rôle essentiel dans le domaine de traitement de la parole et doit être, si possible, pour les applications modernes, calculée en temps réel.

Tous les auteurs ont convenu que le sexe et l'âge jouent un rôle crucial dans la différence de voix, c'est-à-dire que jusqu'à l'âge de 35 ans, la fréquence continue de diminuer légèrement et lorsque l'âge de 55 ans, la fréquence moyenne tend à augmenter.

## I. 5. Traits Acoustiques Du Signal De Parole

Les traits acoustiques du signal de parole sont directement liés à sa production dans l'appareil phonatoire. Tout d'abord, nous avons l'énergie du son [36], celle-ci est liée à la pression de l'air en amont du larynx. Puis, pour les sons voisés, nous avons la fréquence fondamentale F0 [37] ; cette fréquence correspond à la fréquence du cycle d'ouverture/fermeture des cordes vocales. Enfin, nous avons le spectre du signal de parole [38] ; celui-ci résulte du filtrage dynamique du signal en provenance du larynx par le conduit vocal qui peut être considéré comme une succession de tubes ou de cavités acoustiques de sections diverses. Chacun de ces traits acoustiques est lui-même intimement lié à une autre grandeur perceptuelle, à savoir la hauteur, l'intensité, et le timbre.

- **La hauteur :**

La hauteur correspond à la fréquence de vibration des cordes vocales, mesurée en Hertz. Plus la fréquence est haute, c'est à dire plus les vibrations sont rapides, plus le son rendu est aigu. C'est l'équilibre entre la tension, la longueur et l'épaisseur des cordes vocales qui va déterminer cette fréquence de vibration.

La hauteur de la voix dépend de la longueur des cordes vocales (de 5 à 12mm chez l'enfant, de 14 à 18mm chez la femme et de 18 à 25mm chez l'homme (**Le Huche et Allali, 2010a**) [40]). L'âge et le sexe interviennent également dans la variation de la hauteur de la voix.

- **L'intensité :**

L'intensité de la voix traduit, quant à elle, l'amplitude des vibrations des cordes vocales et est principalement contrôlée par la variation de la pression sous-glottique.

En principe, plus la pression sous-glottique sera grande, plus l'intensité vocale le sera également. Du point de vue perceptif, plus l'intensité est importante, plus la voix sera considérée comme forte. Mesurée en décibel (dB), on considère à 10dB l'intensité d'une voix chuchotée, de 35 à 40 dB celle

d'une voix calme, de 50 à 70dB celle d'une conversation normale, de 80 à 90dB pour un cri. Il est à noter que la différence de mesures d'amplitude entre deux cycles successifs (à fréquence constante) est appelée le shimmer.

- **Le timbre :**

Le timbre est l'identité vocale d'un individu. Il dépend de la composition acoustique du son : le son glottique est composé d'un son fondamental et d'harmoniques. La qualité du timbre est directement liée à la richesse en harmoniques : plus la phase d'accolement des cordes vocales est longue et bonne, plus il y a d'harmoniques, et plus le timbre sera de qualité. Les harmoniques sont regroupées en ce que l'on appelle des formants. On distingue trois formants principaux :

- F1 : 500 – 1000 Hz
- F2 : 1200 – 2400 Hz
- F3 : 2400 – 3500 Hz.

Les deux premiers formants sont appelés formants vocaliques car ils caractérisent les voyelles. La position de la langue, des lèvres ainsi que l'ouverture de la mandibule sont des facteurs dont ils dépendent. A partir de F3, on parle de formants extra vocaliques. Plus il y a d'énergie au-delà de 1500 Hz (c'est à dire vers la fin des formants vocaliques) plus on considère que le timbre est riche, et plus la voix est audible.

Un beau timbre de voix résulte aussi de l'action des modulateurs sur le son source laryngé, modulateurs qui vont favoriser certains harmoniques au détriment d'autres. Le timbre est décrit avec un large panel d'adjectifs : on peut qualifier une voix de douce, chaude, colorée, pointue, rauque, nasonnée, voilée, éclatante, lumineuse, brouillée...

## **I. 6. Propriétés Spécifiques Du Signal Vocal**

### **1. Continuité :**

Une personne par nature parle un langage oral qui est une série continue de sons accompagnée de quelques arrêts temporaires de respiration (zones empoisonnées).

### **2. Variabilité**

La parole est une grande variété résultant d'un certain nombre de facteurs : la lumière (bruit ou échos), la distance, la direction de la voix et le sexe de l'orateur.



### 3. Le conduit vocal

Le tube vocal dans le larynx, qui est la source du son, est étiré avec une fréquence qui peut se développer très rapidement ; On peut aussi considérer le frottement qui se produit entre les cordes vocales comme une seconde source du tractus vocal.

### 4. Le codage

Le codage concerne les niveaux lexicaux, syntaxiques, sémantiques, morphologiques et phonétiques (phonèmes et leurs interactions) utilisés souvent pour assurer une meilleure qualité de la parole synthétique.

## I. 7. Troubles de la voix

Les pathologies du larynx ou anomalies laryngées, affectant les cordes vocales, sont à la base des troubles vocaux. Les laryngologues déterminent les troubles de la voix en rapportant des propriétés de la parole émise par les patients. Ces dysphonies ou troubles vocaux, qui désignent une altération de la voix d'origine laryngée, peuvent être décrites comme étant les conséquences audibles d'une malformation organique acquise ou congénitale des plis vocaux ou d'un contrôle déficient de la respiration ou de l'articulation laryngée. Les malformations organiques acquises peuvent être la conséquence d'un malmenage ou surmenage vocal. Des symptômes fréquents des troubles de la voix sont des irrégularités des cycles de parole. Les origines de ces irrégularités sont multiples [40].

### I. 7. 1. Compléments sur les dysphonies

En général, la dysphonie est assimilée à une voix ayant un timbre altéré. Plus précisément, le terme de dysphonie est associé à toute voix perçue comme pathologique c'est à dire dont l'altération vocale est perçue sur un ou plusieurs de ses caractères acoustiques comme le timbre, l'intensité et la hauteur tonale.

Cependant, il peut exister des voix « dysphoniques » sans perturbation acoustique et des voix « altérées » non pathologiques. C'est la raison pour laquelle [Le Huche & Allali (2001a) [41]] la définissent comme «un trouble momentané ou durable de la fonction vocale ressenti comme tel par le sujet lui-même ou son entourage. »

Nous conviendrons que la dysphonie est un trouble de la voix qui résulte d'une dysfonction de production et/ou d'une lésion organique, principalement liée aux cordes vocales. Dans les prochaines

sections, nous présenterons les dysphonies d'origines fonctionnelles puis les dysphonies d'origines organiques.

## I. 8. Pathologies du larynx

Il existe plusieurs types de pathologies qui peuvent affecter le larynx. Ces pathologies sont classées en deux catégories : organiques et fonctionnelles.

### I. 8. 1. Dysphonies organiques

Les causes principales des dysfonctionnements organiques sont les changements morphologiques de l'anatomie du larynx, surtout au niveau de la glotte. Parmi ces dysphonies, on trouve : les nodules, les kystes et les laryngites.

- **Nodule**

C'est une lésion bénigne, charnue ou entourée d'œdème de la muqueuse, souvent causé par les frottements des cordes vocales créent des boules sous forme d'ampoules bilatérales et symétriques. Lors d'une phonation les deux nodules se rencontrent et ne permet pas la fermeture des cordes vocales, ce qui donne une voix voilée et d'une faible intensité et des difficultés de parler et de tenir note longtemps. La figure (I.4) illustre deux photos des cordes vocales avec des nodules, en position d'ouverture (A) et de fermeture (B).

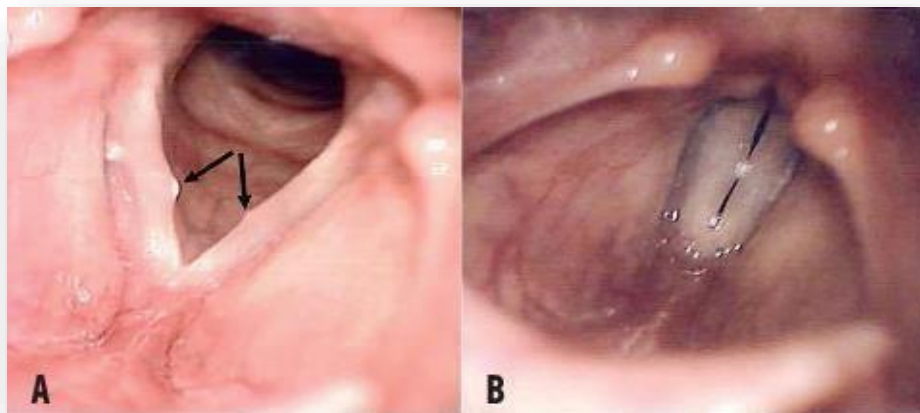
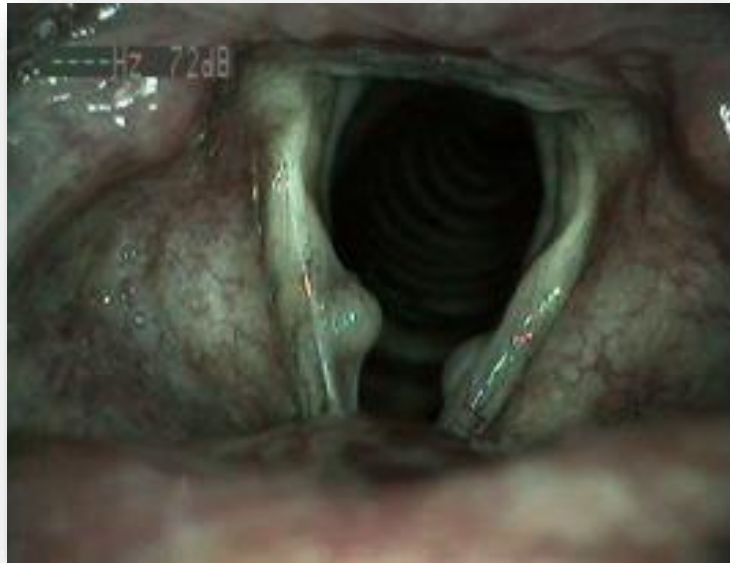


FIGURE I. 4 : NODULES SUR LES CORDES VOCALES.

- **Kyste**

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 [42].



**FIGURE I. 5 :** EXEMPLE DE KYSTE SUR LES CORDES VOCALES.

- **Laryngite**

Les laryngites sont des inflammation chroniques ou aiguës des cordes vocales. La cause principale de cette dysphonie est une infection virale et/ou bactérienne. Cette dysphonie augmente la masse et la dureté de la couverture ce qui entraîne une diminution de l'amplitude de vibration des cordes vocales [43]. La voix devient plus grave, rauque avec un timbre voilé, sourde et éraillée. Elle peut disparaître complètement [42].

### **I. 8. 2. Dysphonies fonctionnelles**

Les dysphonies fonctionnelles sont dues à une mauvaise utilisation du système vocal anatomiquement et physiologiquement intact. Parmi les dysphonies fonctionnelles, on trouve : la paralysie laryngée.

Dans les paralysies laryngées, les cordes vocales demeurent dans une position plus ou moins ouverte. La voix est soufflée et rauque avec une importante fuite d'air, entraînant un essoufflement en fin de phrase [43].

## **I. 9. Conclusion**

Un travail sur la classification des voix pathologiques nécessite la connaissance du mécanisme de production de la parole. Nous avons présenté seulement quelques notions élémentaires, la majorité des travaux et mémoires dans le domaine du traitement de la parole ont largement exposés ces notions. Concernant les différents types de dysphonies, leurs origines et ses sous classes une étude détaillée a été présenté.

Comme dans le cas de la reconnaissance de la parole et l'identification du locuteur, nous avons remarqué que la majorité des travaux sur la classification des voix pathologiques utilisent les paramètres MFCC et des machines à vecteurs de support. Dans le chapitre suivant nous présentant un système d'identification automatisé et une présentation détaillé du ses étapes ; et les définitions et les différentes étapes d'obtention de ces paramètres.

# **Chapitre II**

## **Systeme de reconnaissance**

## II. 1. Introduction

Comme ce que nous avons présenté dans le premier chapitre ; la voix est un signal de parole a été présenté comme un déplacement d'air transportant simultanément des informations linguistiques et émotionnelles, engendré par un ensemble de muscles et d'organes. A partir de ce point, et pour les besoins de notre étude, il sera considéré comme une forme d'onde contenant tout un jeu d'informations sur le message qu'elle transporte.

La reconnaissance automatique de la parole est la manière évoluée pour établir un dialogue artificiel « Homme-Machine », dans le but d'adapter une machine à un vocabulaire limité, qui traduit un besoin issu d'un locuteur.

Le domaine de cette application peut atteindre plusieurs domaines tels que : le pilotage d'avion, la composition du numéro téléphonique du correspondant, faire acquérir des informations à un PC, différentes aides à des handicapés...etc.

Ce domaine fait l'objectif des chercheurs depuis des années, par conséquent un bon nombre des méthodes sont incorporées, telles que les méthodes globales, analytiques, probabilistes et encore les méthodes connexionnistes qui sont adoptées depuis les années quarante.

Dans ce chapitre, nous introduisons d'abord le système d'identification automatisé ; Ensuite, nous expliquons toutes ses étapes et son mécanisme d'action. Enfin, nous terminons le chapitre par une explication détaillée du classifieurs SVM.

## II. 2. Le système d'identification

Les systèmes d'identification de la voix se concentrent sur le modèle de la parole d'un individu. Ces configurations de la parole sont constituées par une combinaison des facteurs comportementaux et physiologiques [44], parmi les méthodes utilisées dans le système d'identification « la reconnaissance vocale ».

La reconnaissance vocale est un sous-domaine interdisciplinaire de la linguistique informatique qui développe des méthodologies et des technologies permettant la reconnaissance et la traduction du langage parlé en texte par des ordinateurs [45]. Elle est programmée pour exécuter le système conformément aux instructions ; ce système sera nécessaire dans ce chapitre afin de pouvoir reconnaître la voix malade de la voix normale, pour nous avons cette décision nous avons tracé un algorithme qui suivie notre travail.

La reconnaissance vocale peut se rattacher à de nombreux plans de la science : traitement automatique des langues, linguistique, théorie des langages formels, théorie de l'information, traitement du signal, réseaux neuronaux, intelligence artificielle, etc [46].

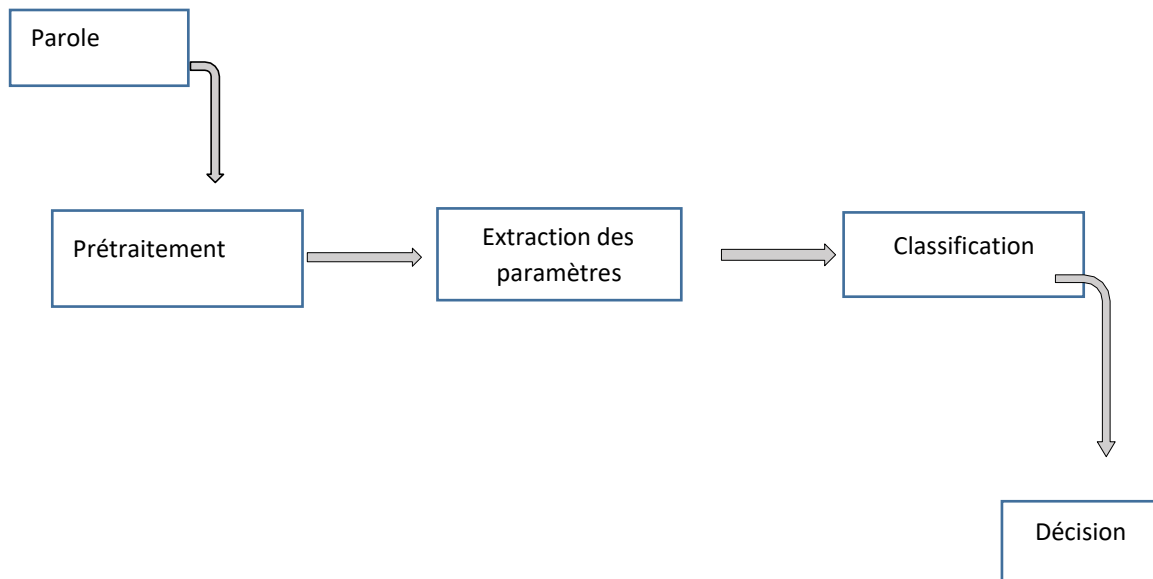


FIGURE II. 1 : SCHEMA DE LA RECONNAISSANCE VOCALE SIMPLIFIE

### II. 3. Les bases de données de voix pathologiques

La collecte des voix dépend de la disponibilité des patients et de leur consentement à participer au programme d'inscription, qui comprend un programme qui doit être suivi, comme la numérisation de la voix des patients, l'évolution de leur maladie, le degré de leur maladie sur leur corps, leur tranche d'âge et leur sexe, ce qui signifie que la combinaison dépend de la disponibilité d'un grand nombre de patients et de la différence d'âge et de sexe. L'enrichissement des données audio est une étape importante pour des travaux de recherche [47].

#### ○ SVD « Saarbrücken Voice Database » :

Cette base de données comprend :

- d'enregistrements de voix de plus de 2000 personnes. Elle est disponible gratuitement en ligne.
- La Fréquence d'échantillonnage des enregistrements est de 50 kHz. Environ 71 pathologies différentes sont incluses dans cette base de données [48].

## II. 4. Prétraitement du signal

C'est la première étape du processus du calcul des vecteurs acoustiques. L'objectif du prétraitement est de modifier le signal de parole,  $x(n)$ , pour qu'il soit plus convenable à l'étape de l'extraction de paramètres. Les opérations de prétraitement peuvent être vues dans la figure 2.5.

A ce stade, nous avons pu identifier les caractéristiques de notre son ; car nous ne pouvons pas utiliser le signal de parole directement car il contient beaucoup d'informations en plus de la répétition [49].



FIGURE II. 2 : LES ETAPES DE PRETRAITEMENT.

### II. 4. 1. Préaccentuation du signal vocale

C'est l'une des communications les plus importantes entre les humains et cela est dû à sa simplicité. Cette simplicité (pour les humains) contient un traitement très complexe par notre cerveau qui produit la parole, la perception et la compréhension, ce qui rend la parole difficile sur la machine et ses applications.

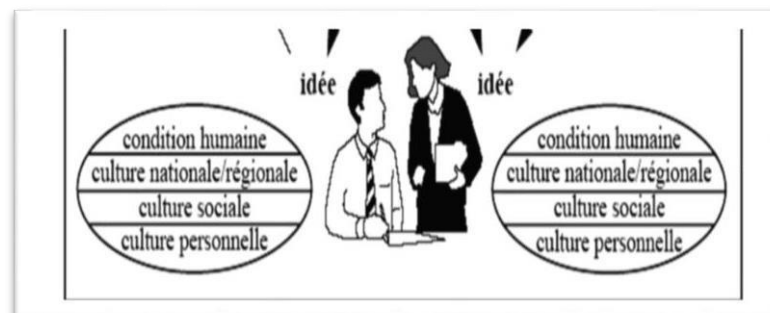


FIGURE II. 3 : COMMUNICATION ENTRE DEUX PERSONNE



Le signal de parole est traité par notre cerveau pour le filtrer de toute information inutile (bruit), en extraire des informations sur l'interlocuteur (voix, état mentale...), décoder le message en unités de base (phonèmes, syllabes...) qui seront réunies en mots choisis à partir d'un lexique [50] affecter des significations à ces mots (phrases) selon une sémantique et un contexte. Tout cela pour montrer le nombre colossal de connaissances nécessaires pour le traitement de la parole.

D'ailleurs on utilise plusieurs applications pour fait le traitement de la parole pour obtenir des informations complètes ; en effet le traitement de la parole contient de la préaccentuation ; elle rencontré dans le traitement pour effet d'accentuer la partie haute du spectre du signal ; la forte concentration énergétique des basses fréquences, observée dans la plus grande majorité des spectres de parole, est considérée comme une nuisance car, lors de l'analyse, elle minimise l'importance de l'énergie du signal aux moyennes et hautes fréquences. Ce problème est souvent partiellement résolu en passant le signal dans un filtre de transmittance  $1 - a \cdot z$  (un filtre de réponse impulsionnelle Fini). C'est le processus de préaccentuation. Concrètement, le signal pré accentué est donc donné par la formule II.1 (Kunt, 1984).

$$Y(n) = s(n) - a \cdot s(n - 1) \quad (\text{II. 1})$$

Où  $s(n)$  correspond au  $n^{\text{ème}}$  échantillon du signal d'entrée, et  $a$  est une constante dont la valeur est généralement fixée entre 0,9 et 1.

Ce procédé apparaisse dans la plupart des systèmes de traitement du signal, de nombreuses applications de reconnaissance vocale ne l'utilisent pas. La principale raison étant que, si le rehaussement de l'énergie des hautes fréquences ne modifie que très peu la structure des sons non voisés, dont les principales composantes spectrales sont situées dans cette zone, cela affecte en revanche très fortement la plupart des sons voisés, dont les caractéristiques majeures sont représentées par les premiers formants.

Vergin et O'Shaughnessy (1995) présentent même une méthode permettent de calculer un coefficient adapté à la catégorie du signal, pour accroître ainsi fortement la qualité de la reconnaissance ; Cependant, cela reste une méthode coûteuse en termes de nombre de comptes.

## II. 4. 2. Sous échantillonnage

Le sous-échantillonnage est le processus de réduction de la fréquence d'échantillonnage d'un signal. Ceci est habituellement fait pour réduire la taille des données et diminuer le temps de calcul. Le facteur de sous-échantillonnage est en général un nombre entier ou une fraction rationnelle supérieure à l'unité. Ce facteur multiplie le temps d'échantillonnage ou, de façon équivalente, qui divise la fréquence d'échantillonnage. Les équations suivantes illustrent un sous échantillonnage à M échantillons [51] [52].

## II. 4. 3. Fenêtrage du signal

L'utilisation d'une fenêtre de Hamming (sur 256 ou 512 points en général) avec recouvrement sur la moitié (128 ou 256 points) permet d'éviter la formation d'artefacts liés aux effets de bord durant la transformation du domaine temporel au domaine fréquentiel.

Fenêtre de Hamming (figure II .3) est une fenêtre "en cloche" qui s'écrit sous la forme :

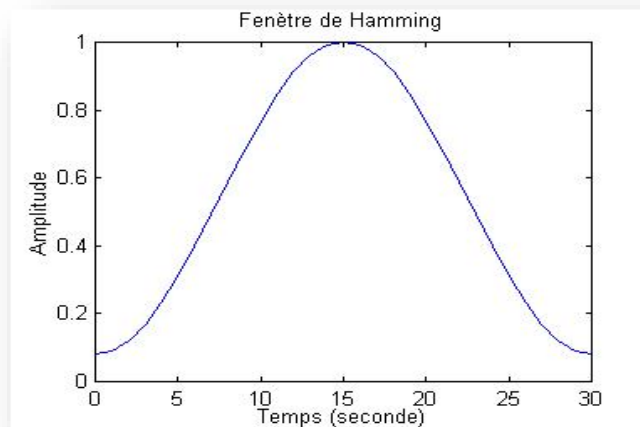
$$w_{\text{HAMMING}}(k) = \{0.54 + 0.46 \cdot \cos\left(\frac{2\pi \cdot k}{M-1}\right) \text{ si } k \in \left[-\frac{M}{2}, \frac{M}{2} - 1\right]\} \quad (\text{II. 2})$$

0 partout ailleurs

Et, l'échelle non linéaire Mel prend en compte des connaissances acoustiques sur la perception humaine.

Souvent une soustraction cepstral se fait sur les MFCC pour déconvoluer le bruit du canal (de la source d'enregistrement : micro, canal téléphonique...). On obtient alors un signal paramétré d'ébruité [53].

En effet, si on considère que le signal est convolé au canal et que les coefficients cepstraux ont une moyenne nulle alors pour ôter le bruit (causé par le canal), il suffit de soustraire à chaque coefficient cepstral la moyenne de tous les coefficients. La moyenne des coefficients du signal étant nulle, cela revient à enlever seulement le bruit (la « moyenne du bruit ») dû au canal.



**FIGURE II. 4 :** REPRESENTATION TEMPORELLE D'UNE FENETRE DE HAMMING.

L'inconvénient majeur de la représentation cepstrale réside dans son manque de lisibilité : il ne s'agit pas d'une représentation directement liée aux informations qu'un expert peut extraire de la lecture d'un sonagramme, ce qui complexifie l'interprétation des paramètres.

## II. 5. Extraction des paramètres

La parole est un signal très aléatoire, alors l'instabilité des paramètres classiques comme le pitch et les formants peuvent être utilisés comme paramètres distinctifs entre les deux types de voix pathologique et normale.

Les variations des paramètres d'un signal parole indiquent la présence d'un son pathologique. Depuis quelques années, le codage est utilisé dans plusieurs applications vocales car il participe au bon fonctionnement ; on trouve deux grandes familles de codage ; codage temporels comme LPC et LPCC et codage fréquentiels comme le MFCC et le cepstrum.

### II. 5. 1. Le codage MFCC

Les coefficients MFCC apparaissent très efficace pour caractériser ces changements. Dans ce qui suit nous présentons la théorie de base derrière ces coefficients. Cette partie est inspirée de la thèse.

Les MFCC sont utilisés en reconnaissance de parole et en identification du locuteur ou de la langue. Ils sont issus de l'hypothèse suivante, à savoir que le signal de parole est le résultat de la convolution entre un filtre (conduit vocal) et une excitation (cordes vocales) :

$$S_n = g_n * b_n \quad (\text{II. 3})$$

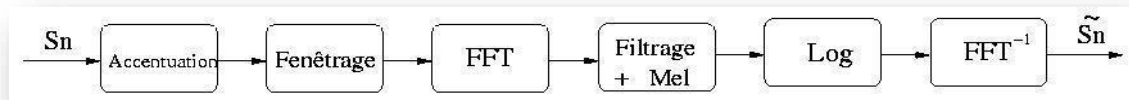
Avec :  $S_n$  le signal d'entrée,  $g_n$  la source et  $b_n$  le conduit.

Une transformation homographique permet de déconvoluer ce produit pour obtenir les « Mel Frequency Cepstral Coefficient » (MFCC). Ces MFCC permettent donc un dé convolution entre la source des sons produits (caractéristiques du locuteur) et le conduit oral (couplé ou non au conduit nasal) :

$$\tilde{S}_n = \tilde{g}_n + \tilde{b}_n \quad (\text{II. 4})$$

Ils ont la propriété d'être fortement dé corrélés.

Ces MFCC sont calculés à la suite d'une succession de transformations : un passage dans le domaine spectral par calcul de la FFT (Transformée de Fourier), puis une application du logarithme et enfin un retour au domaine temporel (nouveau calcul de FFT) (Figure II .4). Ils sont appelés MFCC car dans le domaine spectral, un changement d'échelle (utilisation de l'échelle perceptive Mel) est effectué.



**FIGURE II. 5** : PROCESSUS DE CREATION DES COEFFICIENTS CEPSTRAUX.

**FFT**: Transformée de Fourier (passage dans le domaine spectral).

**FFT<sup>-1</sup>** : Transformée de Fourier inverse (retour dans le domaine temporel).

Le calcul de la FFT se fait sur des fenêtres glissantes.

Remarquons qu'une accentuation des aigus est présente car les aigus sont toujours plus faibles que les graves.

## II. 5. 2. Codage cepstrum

Dans le traitement de la parole, la méthode Cepstral est utilisée pour déterminer qui parle, ou la nature de voix de la hauteur, la détection de voix est souvent effectuée dans le domaine Cepstral car le domaine Cepstral représente la fréquence dans le spectre de magnitude logarithmique d'un signal, le Cepstrum est formé en prenant la FFT (ou IFFT) du spectre de magnitude logarithmique d'un signal.

La raison d'utiliser la FFT ou l'IFFT de manière interchangeable est que l'un vous donnera simplement une version inversée de l'autre, de sorte que chacun est également valable pour le traitement que nous souhaitons effectuer. Le Cepstrum est une variante de décalage temporel. Par conséquent, nous ne peut pas simplement appliquer cette méthode à l'aveuglette. Au lieu de cela, nous devons aligner précisément vos fenêtres de domaine temporel de manière à ce qu'elles démarrent et s'arrêtent exactement sur un segment vocal exprimé.

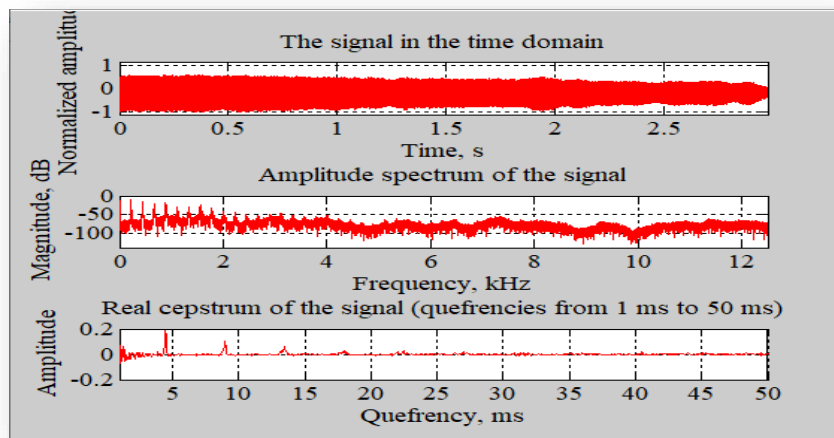


FIGURE II. 6 : CEPSTRUM D'UN VOIX NORMALE D'UN FEMME

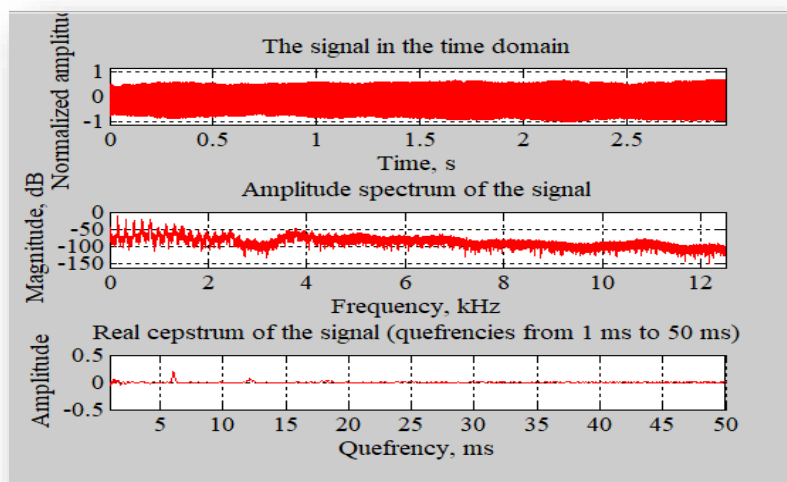


FIGURE II. 7 : CEPSTRUM D'UNE VOIX NORMAL POUR L'HOMME

## II. 6. Méthodes d'apprentissage des données acoustiques

### II. 6. 1. Type de classification

Les méthodes de classification des sons sont des méthodes sur lesquelles nous nous appuyons pour obtenir une bonne classification des sons, elles reposent sur le regroupement des objets en catégories afin que cette catégorie soit similaire entre elles, c'est-à-dire homogène [54].

L'idée de faire des assemblages en classes homogènes d'individus a formé deux branches de classification automatique, à savoir :

#### a. Classification non supervisée ou "Clustering» :

Cette méthode est considérée comme une méthode mathématique d'analyse de données, dans le but de faciliter l'étude d'une grande population (animaux, plantes, patients...), c'est-à-dire qu'elle facilite l'interprétation de grandes quantités de données [55].

Dans cette méthode, nous regroupons les individus qui la composent en plusieurs catégories aussi proches que possible les unes des autres.

Si les variables sont nombreuses il peut être intéressant de les regrouper afin de réduire leur nombre pour une interprétation plus facile [55]. D'une manière plus formelle et plus générale, la classification non supervisée consiste à créer une partition ou une décomposition de cet ensemble en groupes telle que : Critère 1 : les objets appartenant au même groupe se ressemblent. Critère 2 : les objets appartenant à deux groupes différents ne se ressemblent pas.

#### b. Classification supervisée :

C'est une méthode de classification basée sur l'apprentissage supervisé, on dispose d'un classifieur (ou classificateur) déjà entraîné sur une collection d'objets étiquetés (modèles) dite d'entraînement ou d'apprentissage avec un nombre de classes connues. Les données sont donc associées à des labels de ces classes. Alors l'objectif de la classification supervisée est d'apprendre à l'aide d'un modèle d'apprentissage des règles qui permettent de prédire la classe des nouvelles observations[54], ce qui permet de déterminer à partir d'un ensemble de descripteur qui caractérise les objets une fonction qui associe une classe et de pouvoir aussi affecter toute nouvelle observation à une classe parmi les classes disponibles.

On construit alors un modèle en vue de classer les nouvelles données. Une fois la phase d'apprentissage est réalisée, l'algorithme de classification est alors utilisé afin de déterminer la classification d'un ensemble d'individus tests composés d'un grand nombre d'échantillons [55] [52]. Parmi les méthodes supervisées on cite : les cas-plus proches voisins, les arbres de décision :

- les réseaux de neurones.
- les machines à support de vecteurs (SVM).
- classificateur de Bayes.

## II. 6. 2. Apprentissage de modèle SVM

Les machines à vecteurs de support sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de discrimination , c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent. SVM est une méthode de classification particulièrement bien adaptée pour traiter des données de très hautes dimensions telles que les textes, les images et la voix...etc. Dans ce qui suit nous présentons les aspects théoriques de la méthode SVM

Les fondations théoriques des SVM ont été explorées par Vapnik et ses collègues dans les années 70 avec le développement de la Théorie de Vapnik-Chervonenkis, et par Valiant et la théorie de l'apprentissage PAC [56].

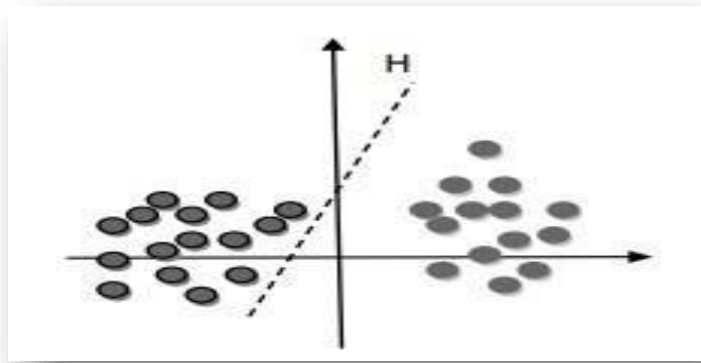
### a. Principe de fonctionnement :

Un SVM, comme un perceptron, trouve un séparateur linéaire entre les points de données de deux classes différentes. En général, il peut y avoir plusieurs séparateurs possibles entre les classes (en supposant le problème linéairement séparable) et qu'un perceptron n'a pas de préférence parmi celles-ci. Dans les SVMs, cependant, nous faisons un choix particulier parmi tous les séparateurs possibles : nous voulons celui avec la "marge" maximale [57].

### b. Notions de base :

❖ Hyperplan :

Dans le cas d'une classification binaire, on appelle hyperplan séparateur un hyperplan qui sépare les deux classes, en particulier il sépare leurs points d'apprentissage. Comme il n'est en général pas possible d'en trouver un, on se contentera donc de chercher un hyperplan discriminant qui est une approximation au sens d'un critère à fixer (maximiser la distance entre ces deux classes) [58].



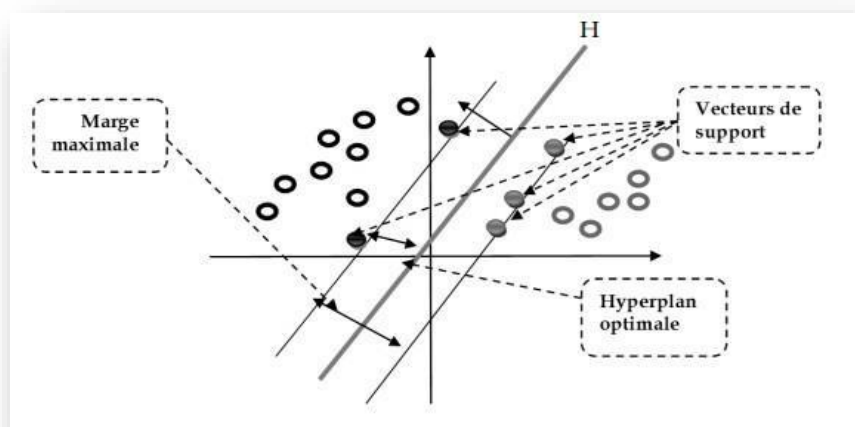
**FIGURE II. 8** : L'HYPERPLAN H QUI SEPRE LES DEUX ENSEMBLES DE POINTS.

❖ Vecteurs de support :

Pour une tâche de détermination de l'hyperplan séparable des SVM est d'utiliser seulement les points les plus proches (i.e. les points de la frontière entre les deux classes des données) parmi l'ensemble total d'apprentissage, ces points sont appelés vecteurs de support [56].

❖ Marge :

Il existe une infinité d'hyperplans capable de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs de support), cette distance est appelée la marge [59].



**FIGURE II. 9** : L'HYPERPLAN OPTIMAL, VECTEURS DE SUPPORT ET MARGE MAXIMALE.

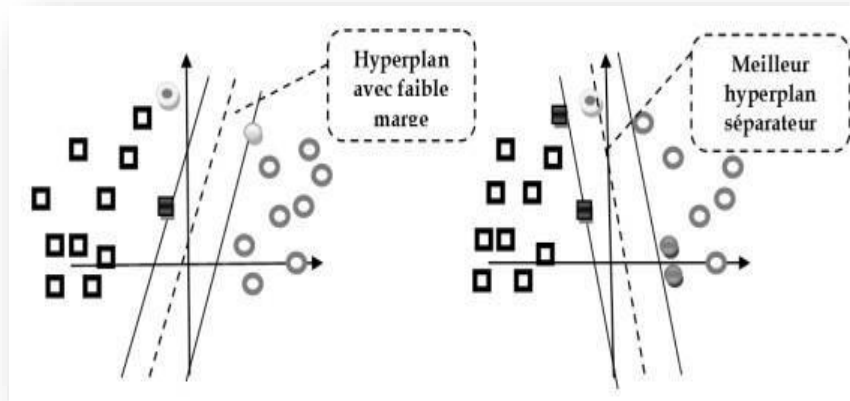


Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma figure III.3, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé [59].

La distance entre un point situé sur H1 et l'hyperplan H0 est donnée par :

$$\frac{|w \cdot x + b|}{\|w\|} = \frac{1}{\|w\|} \quad (\text{II. 6})$$

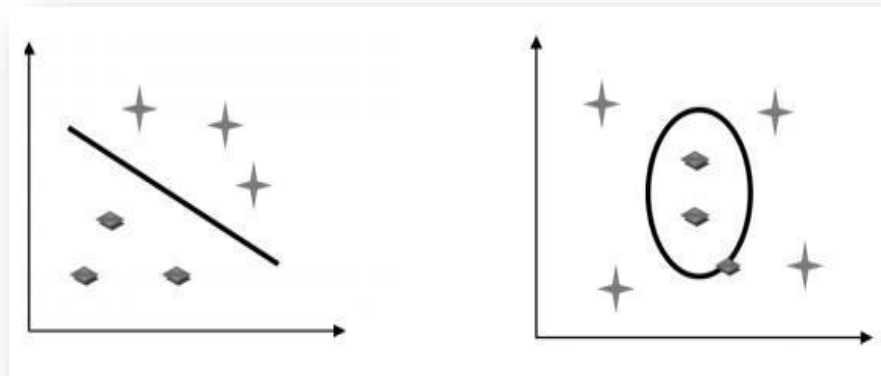
$$\text{la marge} : \frac{2}{\|w\|} \quad (\text{II. 7})$$



**FIGURE II. 10** : MEILLEUR HYPERPLAN SEPARATEUR.

❖ Linéarité et non-linéarité :

Parmi les modèles des SVM, on constate les cas linéairement séparables et les cas non linéairement séparables. Les premiers sont les plus simples des SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [60].



**FIGURE II. 11 :** LES CAS LINEAIREMENT ET NON LINEAIREMENT SEPARABLE.

**c. Fondements mathématiques :**

❖ Le cas linéairement séparable :

Si les données sont linéairement séparables, alors il existe un hyperplan d'équation

$$\langle W, X \rangle + b = 0 \tag{II. 8}$$

Tel que :

$$W \cdot X + b \geq +1 \quad \text{si} \quad Y_i = +1 \tag{II. 9}$$

Et

$$W \cdot X + b \leq -1 \quad \text{si} \quad Y_i = -1 \tag{II. 10}$$

On peut combiner ces deux inéquations en une seule :

$$Y_i (W \cdot X + b) \geq +1 \tag{II. 11}$$

La distance perpendiculaire de l'origine à l'hyperplan

$$H1 : W \cdot X + b = 1 \quad \text{et} \quad \frac{|1-b|}{\|w\|} \tag{II. 12}$$

De même pour :

$$H2 : W \cdot X + b = -1 \quad \text{et} \quad \frac{|1+b|}{\|w\|} \tag{II. 13}$$

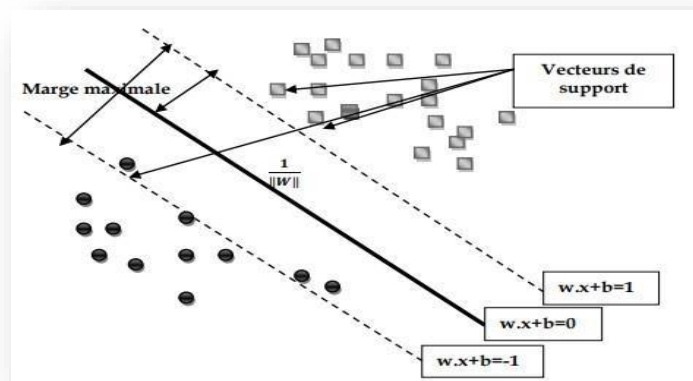


FIGURE II. 12 : EXEMPLE GRAPHIQUE DES DONNEES LINEAIREMENT SEPARABLES.

❖ Le cas linéairement non séparable :

Dans le cas non linéairement séparable, on introduit des variables d'écart

$$s_i = (i..N) \quad \text{avec} \quad s > 0 \quad \text{(II. 14)}$$

Dans les contraintes, qui deviennent :

$$W \cdot X + b \geq 1 - s_i \quad \text{et} \quad s_i \quad Y_i = +1 \quad \text{(II. 15)}$$

$$W \cdot X + b \leq -1 + s_i \quad \text{et} \quad s_i \quad Y_i = -1 \quad \text{(II. 16)}$$

Un moyen naturel de donner un coût aux erreurs est de remplacer la fonction à minimiser précédente par :

$$\frac{\|w\|^2}{2} + C (\sum s_i)^2 \quad \text{(II. 17)}$$

Dans le cas où les données sont non linéairement séparables, c'est-à-dire la surface séparatrice est non linéaire, on transpose le problème dans un autre espace  $F$  de dimension plus élevée pour rendre les nuages de points des deux classes linéairement séparables au moyen d'une transformation  $\Phi$  tel que :

$$\Phi : X \rightarrow \epsilon F \quad \text{(II. 18)}$$

$$\Phi^T(x_i) * \Phi(x_j) \quad \text{(II. 19)}$$

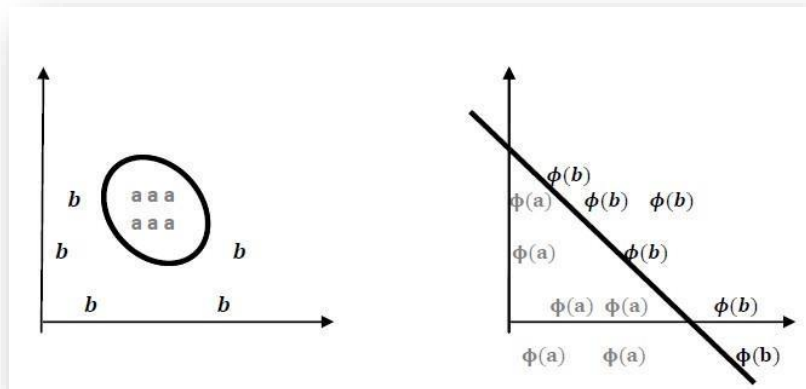
Cette dernière quantité peut être remplacée par une fonction de la forme  $K(x_i, y_j)$  (Les fonctions scalaires symétriques et définies positives, que l'on désigne souvent simplement par "noyaux", sont plus précisément des "noyaux de Mercer"), c'est ce qu'on appelle le noyau. Donc :

$$K(x_i, y_j) = \Phi^T(x_i) * \Phi(x_j) \quad (\text{II. 20})$$

Le lagrangien devient alors :

$$L(\mathbf{w}, \mathbf{b}, a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{j=1}^N a_i \cdot a_j \cdot y_i \cdot y_j \cdot K(x_i, y_j) \quad (\text{II. 21})$$

A ce stade, le problème se situe dans le choix de la transformation ou plus généralement à la fonction noyau K. Ils existent peu de noyaux régulièrement utilisés avec les SVM [61] [62] [63] [56].



**FIGURE II. 13** : ESPACE DE PROJECTION DES DONNEES NON LINEAIREMENT SEPARABLE.

❖ Les fonctions Noyau (Kernel) :

En pratique, l'astuce du noyau consiste à réécrire un algorithme où toutes les relations entre données d'entrée peuvent s'écrire sous forme de produits scalaires, en remplaçant ce produit scalaire par une fonction scalaire de deux variables (noyau) comme nous l'avons cité dans la partie précédente. L'astuce du noyau permet ainsi de généraliser un algorithme linéaire manipulant des vecteurs :

- pour traiter les vecteurs de façon non linéaire (parce que les données présentent des non linéarités qu'il est utile d'exploiter pour le problème visé).
- pour manipuler d'autres types d'objets que les vecteurs, c'est-à-dire prendre en compte des structures de données assez complexes et non homogènes (séquences, chaînes, graphes, ...etc.). Donc il suffit de trouver la bonne fonction noyau qui exprime le mieux la similarité entre les différentes structures. L'idée retenue dans SVM va dans un autre sens : on va tenter de trouver un mapping de l'espace d'entrée vers un autre espace (feature space) dans lequel les données sont linéairement séparables.

## **II. 6. Conclusion**

Dans ce chapitre, nous présentons l'architecture typique du détecteur de voix pathologique (système d'identification) ; Nous décrivons en détail les différentes étapes et les notions essentielles qui composent ce système.

Le SVM est une méthode de classification qui a montré de bonnes performances dans la résolution de problèmes variés, tels que le traitement d'image, la catégorisation de textes ou le diagnostic médical et ce même sur des ensembles de données de très grandes dimensions. Par conséquent, Nous terminons ce chapitre par une étude détaillée des SVM et de leurs standards. Nous allons vérifier l'efficacité de ces méthodes dans le cas de la détection et la classification des voix pathologiques par nous appliquant des paramètres MFCCs, les résultats et les discussions seront l'objet du chapitre trois

# **Chapitre III**

## **Expérimentations et résultats**

### III. 1. Introduction

Nous avons mentionné dans le chapitre précédent que les machines à vecteurs de support sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent.

Alors ; Ce chapitre présente l'utilisation des étapes du chapitre II relatives à la mise en œuvre d'un système de détection automatique d'une voix pathologique à partir d'un signal vocal. En fait, nos exemples sont basés sur des SVM.

L'objectif principal est d'exploiter les capacités discriminantes des SVM et aussi de tester l'efficacité du classificateur dans notre étude, nous utilisant deux bases de données différentes (SVD et HUPA).

### III. 2. Méthodologie

Nous avons déjà expliqué que le système d'identification vocale se réalise en deux phases c.-à-d. apprentissage et classification, alors dans cette partie nous visons à reproduire les résultats obtenus par les coefficients MFCC et le classificateur SVM.

Le premier point c'est l'extraction des MFCC, dont nous avons extrait un vecteur des 13 premiers coefficients, les dérivées premières ( $\Delta$ ) et deuxièmes ( $\Delta\Delta$ ) représentant la vitesse et l'accélération respectivement.

La méthode LDA pour la réduction de dimension est utilisée pour choisir les coefficients les plus informatifs.

Le deuxième point est l'apprentissage des SVM, Dans cette étape nous essayons de trouver la meilleure structure qui maximise une séparation optimale entre les deux classes . Le choix du noyau, il est souvent un problème pratique et critique lors de la mise en œuvre des SVM.

Nous avons utilisé un noyau RBF. Ce noyau se décrit par deux paramètres : ( $C$ ) est défini par le facteur de pénalité et  $\sigma$  est défini par la largeur du noyau.

### III. 2. 1. Corpus de données

Notre travail a été réalisé à l'aide de deux bases de données pour l'élaboration et l'évaluation de méthode proposé, ce qui assure la disponibilité d'une large quantité de données et une diversité suffisante dans les enregistrements : la base de données Hôpital Universitario Príncipe d'Asturias (HUPA) [64] [65] et la base de données sur les troubles de la parole de Saarbrücken (SVD) [66], [67] qui sont considérées comme les plus fiables et les plus standards. Consulter la base de données [68], [69], [70].

#### a. Base de données HUPA :

Cette base de données est enregistrée à l'Hospital Universitario Príncipe de Asturias à Alcalá de Henares, Madrid, Espagne [72] [65], L'ensemble de données contient la vocalisation continue de la voyelle / a / de 439 hispanophones adultes (239 personnes en bonne santé et 200 personnes malades). Initialement, les données étaient enregistrées à une fréquence d'échantillonnage de 50 kHz, puis sous-échantillonnées à 25 kHz. Les sons pathologiques comprennent une variété de maladies organiques, telles que les nodules, les polypes, l'œdème et le cancer. Plus de détails sur la base de données peuvent être trouvés dans [64] [65], [74].

Genre	Nature	Apprentissage	Test	Unité D'analyse
Femme	Normale	34	20	Voyelle "A"
	Pathologique	30	20	

TABLEAU III. 5 : LA COMPOSITION DE LA BASE HUPA

#### b. La base de données SVD :

Cette base de données a été enregistrée à Institut für Phonetik de l'Université de la Sarre et à la section de phoniatry de la Caritas Clinic St. Theresia à Sarrebruck, Allemagne [66], [67].

Les données comprennent des enregistrements de phonations soutenues des voyelles tenues /a/, /i/ et /u/ à différentes hauteurs de voix, ainsi qu'avec des hauteurs (normal, basse, haute) et la lecture de la phrase « Guten Morgen, wie geht es Ihnen ? ("Bonjour comment-allez-vous?"), déclinée sous sa forme audio et sous la forme d'un signal issu d'un Electro-GlottoGramme (EGG), renseignant sur la vibration des cordes vocales. Les données ont été enregistrées avec une fréquence d'échantillonnage de 50 kHz.



Nature	Apprentissage	Test	Unité D'analyse
Normale	25	10	Voyelle "A"
Pathologique	20	15	

TABLEAU III. 2 : CORPUS DE DONNEES SVD

Ces figures représentent des signaux de voix pathologique et de voix normale.

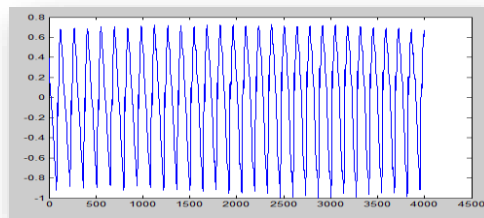


FIGURE III. 1 : SIGNAL DE VOIX NORMALE

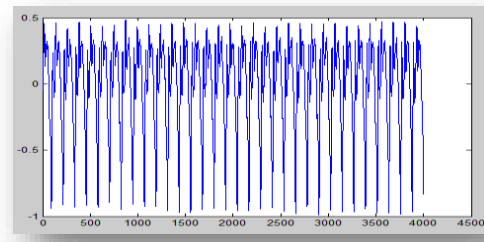


FIGURE III. 2 : SIGNAL DE VOIX PATHOLOGIE

### III. 2. 2. Prétraitement

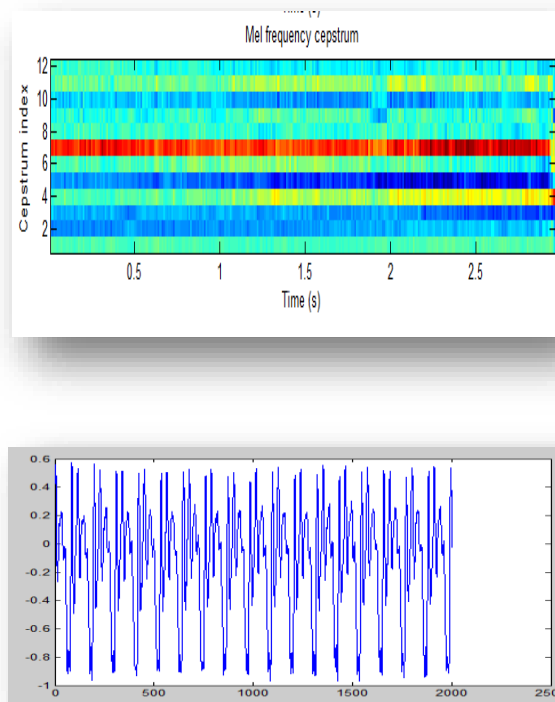
Cette phase est en général réservée à la préparation du signal, elle consiste à choisir la durée de la trame d'analyse et du recouvrement fin de moins de compromettre, le but de cette étape est : obtenir les caractéristiques acoustiques et les paramètres du signal de parole.

En fait des tests préliminaires effectués tant sur le système de reconnaissance de référence comme sur les systèmes hybrides, ont démontré que le signal de parole doit être normalisé entre -1 et 1 avant tout traitement, que ça soit pour la phase d'entraînement comme pour la phase de test. Il existe une différence de 10 à 20 points sur les taux de reconnaissance selon que l'on normalise le signal d'origine ou pas. Par conséquent, le module de prétraitement de tous nos systèmes, effectuera la normalisation des signaux avant toute autre opération.

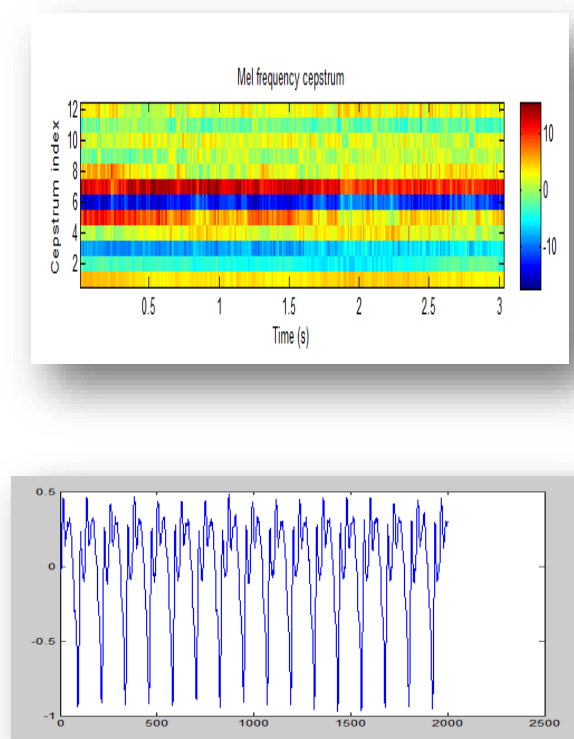
### III. 2. 3. Extraction des paramètres

L'extraction des vecteurs de caractéristiques se fait à l'aide de la technique des MFCC, Nous avons choisi de prendre (13, 26,39) coefficients incluant le coefficient d'énergie. L'analyse des signaux est effectuée à l'aide d'une fenêtre Hamming de 20 ms, ces fenêtres glissantes d'une façon régulière avec un chevauchement de 50 %. Les coefficients MFCC sont calculés à partir d'un banc de 24 filtres. Le vecteur de paramètres se compose de 39 coefficients incluant 12 premiers coefficients MFCC, l'énergie, leur dérivée première et deuxième Ces paramètres ont été extraits en utilisant la fonction melcepst (toolbox).

Une illustration des coefficients MFCC est montrée dans les figures :



**FIGURE III. 3 :** EXEMPLE DE COEFFICIENTS MFCC ET LEUR AMPLITUDE VOIX NORMALE D'UNE FEMME



**FIGURE III. 4 :** EXEMPLE DE COEFFICIENTS MFCC ET LEUR AMPLITUDE VOIX NORMALE D'UNE FEMME

### III. 3. Résultats expérimentaux

#### III. 3. 1. Matrice de confusion

Les performances de la classification est mieux décrite par un outil bien nommé appelé matrice de confusion. Une matrice de confusion de base est traditionnellement organisée sous la forme d'un tableau à deux dimensions. Les classes prédites sont disposées horizontalement en lignes et les classes réelles sont disposées verticalement en colonnes, bien que cet ordre soit parfois inversé.

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 [75].

	<b>Etiquette positive</b>	<b>Etiquette négative</b>
<b>décision positive</b>	Vrais Positifs (VP)	Faux Négatifs(FN)
<b>décision négative</b>	Faux Positifs (FP)	Vrais Négatifs (VN)

TABLEAU III. 1 : MATRICE DE CONFUSION

**Etiquette positive** : Patients malades.

**Etiquette négative** : Patients sains.

**Décision Positive** : Test était positif.

**Décision Négative** : Test négatif.

**Vrais Positifs (VP)** : individus malades réagissent positivement au test.

**Vrais Négatifs (VN)** : individus sains réagissent négativement au test.

**Faux Positifs (FP)** : individus sains réagissent positivement au test.

**Faux Négatifs (FN)** : individus malades réagissent négativement au test.

**La sensibilité** : est le taux de vrais positifs, correspond à la probabilité que le test soit positif sachant que le sujet est malade. Elle mesure donc la capacité d'un test à détecter les malades. Plus la sensibilité est proche de l'unité, moins il y a d'erreurs de détection des sujets malades (faux négatifs).

$$\text{La sensibilité} = \frac{VP}{VN+VP} \quad (\text{III. 1})$$

**La spécificité** : est le taux de vrais négatifs, correspond à la probabilité que le test soit négatif sachant que le sujet est sain. Elle mesure donc la capacité d'un test à détecter les individus sains. Plus la spécificité est proche de l'unité, moins il y a défaut positifs.

$$\text{La spécificité} = \frac{VN}{FP+VN} \quad (\text{III. 2})$$

**Taux de faux positifs (TFP)** : proportion de négatifs détectés comme des positifs par le test (1-Spécificité).

**Taux de faux négatifs (TFN)** : proportion de positifs détectés comme des négatifs par le test (1-Sensibilité).

**La précision** : est la proportion de prédictions correctes parmi les points que l'on a prédits positifs.

$$\text{La precision} = \frac{VP}{VP+FP} \quad (\text{III. 3})$$

**L'exactitude ou Taux de bonne classification (TBC) :** est définie comme la capacité du classificateur à sélectionner tous les cas qui doivent être sélectionnés et à rejeter tous les cas qui doivent être rejetés.

$$TBC = \frac{VP+VN}{VP+FN+VN+FP}$$

**. Résultats en utilisant la base SVD :**

D'après le test que nous avons fait, nous avons constaté l'effet des coefficients MFCC sur le taux de reconnaissance, on a obtenu les résultats qui représentent dans le tableau III.3 :

		Normale	Pathologique
<b>13</b> <b>Coef</b>	N	77%	23%
	P	25%	75%
<b>26</b> <b>Coef</b>	N	78%	22%
	P	25%	75%
<b>39</b> <b>Coef</b>	N	80%	20%
	P	20%	80%

**TABLEAU III. 3 : RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES SVD**

- Meilleur résultats est obtenu en utilisant les 39 coefficients dont il atteint les 80%,
  - Il y a une amélioration lorsqu'on utilise les dérivées premières et deuxièmes.
- **Les résultats après la réduction**

		Normale	Pathologique
<b>13</b> <b>Coef</b>	N	78%	22%
	P	20%	80%
<b>26</b> <b>Coef</b>	N	80%	20%
	P	20%	80%
<b>39</b> <b>Coef</b>	N	84%	16%
	P	14%	86%

**TABLEAU III. 4 : RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES SVD APRES**

- Nous remarquons que LDA permis d'améliorer le taux de reconnaissance, dont elle permet de choisir les coefficients les plus significatifs.

➤ Résultat avec la base HUPA :

Nous avons répété le test sur la base de donne HUPA et écrit les résultats dans le tableau (Tableau III.6)

		Normale	Pathologique
<b>13</b> <b>Coef</b>	N	85%	15%
	P	18%	82%
<b>26</b> <b>Coef</b>	N	90%	10%
	P	18%	82%
<b>39</b> <b>Coef</b>	N	90%	10%
	P	10%	90%

**TABLEAU III. 6** : RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES HUPA

- Les mêmes remarques sont valables pour la base de données HUPA, mais elle présente des taux plus élevés ont meilleur taux atteint les 90%.
- Les données (les voix pathologiques) de cette base sont triées selon l'échelle de GRBAS la chose qui a permet d'avoir des taux important par rapport à la base SVD.

- Les résultats après la réduction

		Normale	Pathologique
<b>13</b> <b>Coef</b>	N	90%	10%
	P	10%	90%
<b>26</b> <b>Coef</b>	N	92%	8%
	P	18%	90%
<b>39</b> <b>Coef</b>	N	98%	2%
	P	4%	96%

**TABEAU III. 7 :** RESULTAT DE CLASSIFICATION AVEC BASE DE DONNEES HUPA APRES  
LDA

- On remarque qu'à chaque fois lorsque en augmente le nombre de dérivé le taux de reconnaissance augmente aussi. Le taux de 39 coefficients atteints 98%.
- LDA améliore les résultats.

*a. Mesures de performance :*

		La spécificité	La sensibilité	La précision	TBC
<b>13</b> <b>Coef</b>	SVD	80%	49.36%	79.5%	79%
	HUPA	90%	50%	90%	90%
<b>26</b> <b>Coef</b>	SVD	80%	50%	80%	80%
	HUPA	83.33%	50.54%	83.63%	91%
<b>39</b> <b>Coef</b>	SVD	86%	49.41%	85.57%	85%
	HUPA	96%	50.51%	96.07%	97%

**TABEAU III. 8 :** MESURES DE PERFORMANCE DE LA DETECTION DE LA VOIX PATHOLOGIQUE



Ce que nous avons remarqué pour les voix de la base hupa (femme) le taux de spécificité atteint 96% lorsqu'on utilise 39 coefficients par 39, par contre 86% pour la base de données SVD.

Le taux de sensibilité presque le même pour les deux base.

D'après ces résultats on peut conclure que :

- La base de données HUPA et mieux que SVD
- Les coefficients MFCC et leurs deuxièmes dérivées donnent des informations plus pertinentes
- SVM et MFCC ont prouvé leur efficacité dans la détection des sons pathologiques.

### **III. 4. Conclusion**

Les simulations, tests et résultats présentés dans ce chapitre confirment l'efficacité du classificateur SVM pour la détection et la classification des voix pathologiques.

Les taux que nous obtenons par la méthode de SVM, MFCC sont considérés comme des taux acceptables pour la base SVD, par contre, nous avons obtenu un taux parfait, de là on déduit que les résultats en utilisant la base de données HUPA mieux que SVD.

## **Conclusion générale**

L'objectif de notre travail était la détection et la classification des pathologies de la voix par la méthode de classification SVM ; ce système proposé peut être utilisé comme un outil précieux par les chercheurs et les orthophonistes pour détecter si la voix est normale ou pathologique et également pour détecter un type spécifique de pathologie, cette méthode est peu coûteuse et entièrement automatisées des du signal vocal.

À partir des résultats obtenus, nous avons remarqué que le paramètres MFCC, est plus pertinent pour discriminer les pathologies des cordes vocales, cette utilisation de paramètre MFCC car la voix humain est de nature non linéaire et pour d'extraction de caractéristiques les plus fréquemment utilisée dans domaines de la reconnaissance vocale et des applications de vérification du locuteur.

Les résultats obtenus par le rapport de classificateur SVM de base sont parfait (97%) avec LDA, par contre 85% pour SVD avec LDA , on peut donc le classificateur permet d'améliorer les performances,sans augmenter la complexité du système.

Nous avons montré que le classificateur SVM peut détecter efficacement les voix pathologiques et normales

Il est conclu à partir des résultats expérimentaux que les SVM multi-classes offrent de bonnes performances pour la classification de ces données vocales en fonction des fonctionnalités utilisées ; à partir de ces résultats, nous avons prouvé que les performances de classification des SVM sont très bonnes.

A la fin on a connu que malgré les efforts et les travaux intensifs dans le domaine de la reconnaissance automatique de la parole, aucun système RAP n'est juge à 100%, mais au fur et mesure les auteurs essayant d'améliorer les scores pour meilleur résultats.

# **Bibliographie**

- [1] Hariharan M, Vijean V, Fook CY, Yaacob S. Speech stuttering assessment using sample entropy and Least Square Support vector machine. In: 8th International Colloquium on Signal Processing and its Applications (CSPA). 2012. pp. 240-245
- [2] Manjula GN, Kumar MS. Stuttered speech recognition for robotic control. International Journal of Engineering and Innovative Technology (IJEIT). 2014;3(12):174-177
- [3] Kasuya.H, S.Ogawa, “AN acoustic analysis of pathological voice and its application to the evolution of laryngeal pathology”, speech communication, 5-2. 1986.
- [4] BRIN F., COURRIER C., LEDERLE E., MASY V. (2004). Dictionnaire d'Orthophonie.Isbergues : OrthoEdition
- [5] Kurzekar PK, Deshmukh RR, Waghmare VB, Shrishrimal PP. A comparative study of feature extraction techniques for speech recognition system. International Journal of Innovative Research in Science, Engineering and Technology. 2014;3(12):18006-18016
- [6] A. Aronson, Clinical Voice Disorders; An Interdisciplinary Approach. Thieme Inc, 1985.
- [7] N. R. Williams, “Occupational groups at risk of voice disorders: a review of the literature,” Occupational Medicine, vol. 53, no. 7, pp. 456–460, 2003.
- [8] Malak Al Mojaly et al., “ Detection and classification of voice pathology using feature selection,” in The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 2014, pp. 571-577.
- [9] Xiang Wang et al., “Discrimination between pathological and normal voices using GMMSVM approach,” Journal of Voice, vol. 25, n°1, 2011.
- [10] Laura Verde et al., “Voice disorder identification by using machine learning techniques,” IEEE Access, vol. 6, pp. 16246-16255, 2018.
- [11] Zuzana Dankovičová et al., “Machine Learning Approach to dysphonia detection,” Applied
- [12] Shaneh M, Taheri A. Voice command recognition system based on MFCC and VQ algorithms. World academy of science. Engineering and Technology. 2009;57:534-538

- [13] Ravikumar KM, Rajagopal R, Nagaraj HC. An approach for objective assessment of stuttered speech using MFCC features. *ICGST International Journal on Digital Signal Processing, DSP*. 2009;9(1):19-24
- [14] Cornaz C, Hunkeler U, Velisavljevic V. An Automatic Speaker Recognition System. Switzerland:Lausanne;2003.Retrievedfrom:[http://read.pudn.com/downloads60/sourcecode/multimedia/audio/209082/asr\\_project.pdf](http://read.pudn.com/downloads60/sourcecode/multimedia/audio/209082/asr_project.pdf)
- [15] Kumar PP, Vardhan KSN, Krishna KSR. Performance evaluation of MLP for speech recognition in noisy environments using MFCC & wavelets. *International Journal of Computer Science & Communication (IJCSC)*. 2010;1(2):41-45
- [16] Ahmad AM, Ismail S, Samaon DF. Recurrent neural network with backpropagation through time for speech recognition. In: *IEEE International Symposium on Communications and Information Technology (ISCIT 2004)*. Vol. 1. Sapporo, Japan: IEEE; 2004. pp. 98-102
- [17] J. Hillenbrand and R. A. Houde, “Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech”, *J. Speech Lang. Hear. R.*, vol. 39, no. 2, pp. 311–321, 1996.
- [18] V. Parsa and D. G. Jamieson, “Identification of pathological voices using glottal noise measures”, *J. Speech, Language, Hearing Res.*, vol. 43, no. 2, pp. 469–485, Apr. 2000.
- [19] Y. Qi, R. E. Hillman and C. Milstein, “The estimation of signal-to-noise ratio in continuous speech of disordered voices”, *J. Acoust. Soc. Amer.*, vol. 105, no 4, pp. 2532–2535, 1999.
- [20].F.Le Huche, A.Allali *La Voix Tome1 Anatomie et physiologie des organes de la voix et de la parole*,4<sup>ème</sup> édition, 2010,204pages
- [21] H.Gray. *Anatomy of the human body*. Lea and Febiger, Philadelphia, United States of America, 1918. (Cited on pages 7, 8, 9 and 10.)
- [22] J. Trémolière, *Electronique Applications N°62*, « La synthèse de la parole », octobre 1988.
- [23] [www.r.batault.free.fr/probatoire](http://www.r.batault.free.fr/probatoire)
- [24] Rose R. C., course ECSE 570, *Automatic Speech Recognition*, McGill University, Win ter 2005.
- [25] Gillie-Guilbert C. (2001). « Et la voix s'est faite chair ... » : Naissance, essence, sens du geste vocal. *Cahiers d'ethnomusicologie*. 3.8.14 – 38.
- [26] [http://phoniatriestrasbourg.free.fr/Site\\_6/Resonateurs.html](http://phoniatriestrasbourg.free.fr/Site_6/Resonateurs.html)
- [27] [http://www.facmed-univ-oran.dz/ressources/fichiers\\_produits/fichier\\_produit\\_1961.pdf](http://www.facmed-univ-oran.dz/ressources/fichiers_produits/fichier_produit_1961.pdf)

- [28] [http://www.facmed-univ-oran.dz/ressources/fichiers\\_produits/fichier\\_produit\\_2612.pdf](http://www.facmed-univ-oran.dz/ressources/fichiers_produits/fichier_produit_2612.pdf)
- [29] Beigi Homayoon, *Fundamentals of Speaker Recognition* Springer, 2011.
- [30] Rudzicz F., *Articulatory Knowledge in the Recognition of Dysarthric Speech*, IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, No. 4, 2011.
- [31] Huang, X., Acero, A., & Hon, H.-W. (2001). *Spoken language processing: a guide to theory, algorithm, and system development*. Upper Saddle River, NJ : Prentice Hall PTR.
- [32] Lofqvist, Anders, Nancy A. McGarr and Kiyoshi Honda, *Laryngeal muscles and articulatory control*. The Journal of the Acoustical Society of America, 76(3):951-954, 1994.
- [33] T. Dubuisson, "Glottal source estimation and automatic detection of dysphonic speakers," PhD thesis, University of Mons, Belgium, 2012.
- [34] <http://www.claudegabriel.be/>
- [35] Boite, R., Boulard, H., Dutoit, T. and Flancq, J. (2000) "Traitement de la Parole", Presses Polytechniques Universitaires Romanes Lausanne.
- [36] Van Den Heuvel H., Rietveld T., Cranen B., *Methodological aspects of segment and speaker-related variability. A study of segmental durations in Dutch*. Journal of Phonetics, no 22, pp 389-406. 1994.
- [37] Atal, B. S. (1972). *Text-independent speaker recognition*. Paper presented at the Program of the 83rd meeting of the Acoustical Society of America. Abstracts only, 18- 27 April 1972, Buffalo, NY, USA.
- [38] Matsui T., Furui S., *Text-independent speaker recognition using vocal tract and pitch information*. In Proceedings ICSLP 90, pp 137-140, 1990.
- [39] Schoentgen, J. (2006) "Vocal cues of disordered voices: an overview", the Journal of the European Acoustics Association (EAA), Acta Acoustica United with Acoustica, Vol. 92, 667-680.
- [40] Titze, I.R. (2000) "Principles of voice production", The National Center of Voice and Speech, USA.
- [41] F. Le Huche & A. Allali, 2001a. *La Voix : Pathologie vocale d'origine fonctionnelle*, Volume 2.
- [42] S. Fraj, "Synthèse des voix pathologiques," Thèse de Doctorat, Université Libre de Bruxelles, 2010.

- [43] A. Kacha, “Analyse de la parole continue en vue de la caractérisation des troubles de la voix,” Thèse de Doctorat, Université Libre de Bruxelles, 2006.
- [44] aphaël Richard <https://www.24pm.com/117-definitions/502-reconnaissance-vocale>
- [45] technologies voix <https://www.biometrie-online.net/technologies/voix>
- [46] <https://www.techno-science.net/glossaire-definition/Reconnaissance-vocale.html>
- [47] Laura Verde et al., “Voice disorder identification by using machine learning techniques,” IEEE Access, vol. 6, pp. 16246-16255, 2018.
- [48] détection et classification des sons : application aux sons de la vie courante et à la parole ; Dans ISTRATE, Michel Vacher, Jean François SERIGNAT.
- [49] Fitch W.T., Raby D. (2001) « The descended larynx is not uniquely human », Proc Biol Sci, 268, 1669-75.
- [50] André Quinquis, Le traitement du signal sous Matlab, Hermes Science, Lavoisier, Paris, France, 446p, 2007.
- [51] thèse de doctorat de l’université de Talouse présentée par M. JEAN-ADRIEN VERNHES.
- [52] Théorie et Traitement du signal, tome 1 : Représentation des signaux et des systèmes - Cours et exercices corrigés Messaoud Benidir.
- [53].C. Mokbel, D. Jouvét, J. Monné « Blind Equalization using Adaptive Filtering for improving Speech Recognition over Telephone », Eurospeech ’95, Madrid, pp. 1987-1990.
- [54] Malak Al Mojaly et al., “Detection and classification of voice pathology using feature selection,” in The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 2014, pp. 571-577.
- [55] S. Chebbout, “La classification automatique,” Chapitre du cours : “Reconnaissance des Formes,” pour étudiants Master 1, Vision Artificielle, Université Badji Mokhtar-Annaba.
- [56] S. Khellat Kihel, A. Merzoug ; Reconnaissance automatique du diabète ; mémoire de fin d’étude USTO ; 2010.
- [57] A. Cornuejols, Une nouvelle méthode d’apprentissage : Les SVMs Séparateurs à vaste marge, Université de Paris-Sud, Orsay, 2002.
- [58] A. Statnikov, Automatic cancer diagnostic decision support system for gene expression domain, thèse de mastère, Université de Vanderbilt ; 2005.

- [59] B. Taachouche, O. Douak ; La reconnaissance automatique du locuteur en mode indépendant du texte en utilisant les méthodes à noyaux (Application du classifieur SVM) ; Mémoire de fin d'étude Ecole Militaire Polytechnique ; 2009.
- [60] N. Ayat, sélection de modèle automatique des machines à vecteurs de support : application à la reconnaissance d'images de chiffres manuscrits, Thèse de doctorat, Ecole De Technologie Supérieure, Québec, 2004.
- [61] J. Callut, Implémentation efficace des Support Vector Machines pour la classification, mémoire de maîtrise, Université De Bruxelles, 2003.
- [62] J. Kharroubi, Etude de Techniques de Classement Machines à Vecteurs Supports pour la Vérification Automatique du Locuteur, thèse de doctorat, Ecole nationale supérieure des télécommunications, 2002.
- [63] C. Bentaouza ; Classification des images médicales par les séparateurs à vaste marge ; Thèse de magister USTO ; 2007.
- [64] L. Moro-Velázquez, J. Gómez-García, J. I. Godino-Llorente, and G. Andrade-Miranda, "Modulation spectra morphological parameters: A new method to assess voice pathologies according to the grbas scale," in *BioMed research international*, 2015.
- [65] J. D. Arias-Londoo, J. I. Godino-Llorente, M. Markaki, and Y. Stylianou, "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices," *Logopedics Phoniatrics Vocology*, vol. 36, no. 2, pp. 60–69, 2011.
- [66] M. Pützer and W. J. Barry, "Instrumental dimensioning of normal and pathological phonation using acoustic measurements," *Clinical Linguistics & Phonetics*, vol. 22, no. 6, pp. 407–420, 2008.
- [67] ———, "Saarbrücken voice database, institute of phonetics, univ. of saarland," 2010, <http://www.stimmdatenbank.coli.uni-saarland.de/> (Last viewed April 20, 2019).
- [68] J. A. G. García, L. Moro-Velázquez, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part I: review of concepts and an insight to the state of the art," *Biomedical Signal Processing and Control*, vol. 51, pp. 181 – 199, 2019.
- [69] J. A. G. García, L. Moro-Velázquez, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part II: review of speaker recognition techniques

and study on the effects of different variability factors,” *Biomedical Signal Processing and Control*, vol. 48, pp. 128–143, 2019.

[70] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, “Voice pathology detection on the saarbrücken voice database with calibration and fusion of scores using multifocal toolkit,” in *Advances in Speech and Language Technologies for Iberian Languages*, 2012, pp. 99–109.

[71] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, “Automatic detection of pathological voices using complexity measures, noise parameters, and melcepstral coefficients,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 370–379, 2011.

[72] V. Parsa and D. G. Jamieson, “Identification of pathological voices using glottal noise measures,” *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 2, pp. 469–485, 2000.

[73] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Díaz-de-María, “Characterization of healthy and pathological voice through measures based on nonlinear dynamics,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1186–1195, 2009.

[74] J. I. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, I. Cobeta-Marco, R. González-Herranz, and C. Ramírez-Calvo, “Acoustic analysis of voice using wpcvox: a comparative study with multi-dimensional voice program,” *European Archives of Oto-Rhino-Laryngology*, vol. 265, no. 4, pp. 465–476, Apr 2008.

[75] 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, 2014.

[76] [M. El Emary ] M. Fezari M. El Emary and F. Amara. *Journal of Communications Technology and Electronics*. (Cited on page 71.)