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

Université Mohammed Seddik Ben Yahia-Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique

Projet de fin d'études pour l'obtention du Diplôme de
Master en Télécommunications

Option : Systèmes des Télécommunications

Thème :

Détection et classification des pathologies de la voix par SVM

Présenté par :

M^elle Latifa ZERIOUEL

M^elle Mounia BENHABILES

Encadré par :

Dr. Chaâbane BOUBAKIR

Promotion : 2020

Remerciements

Nous remercions Dieu le tout puissant, qui nous a donné la force et la patience pour accomplir ce travail.

Nous adressons nos vifs remerciements à notre encadreur Mr BOUBAKIR Chaâbane pour son aide, ses conseils, ses contributions, ses orientations précieuses et ses compétences.

Nous remercions vivement les membres du jury qui nous ont fait l'honneur d'accepter de juger notre travail.

N'oublions pas, bien sûr, nos professeurs de tout le cycle universitaire, de leur présenter nos meilleurs vœux.

Et enfin, nous tenons à remercier tous ceux qui ont contribué de près ou de loin, par leurs encouragements, leurs conseils et leurs soutiens à mener bien ce travail.

Latifa & Mounia

Dédicaces

Je dédie ce modeste travail

À mes chers Parents ma mère et mon père

Pour leur patience, leur soutien et leur encouragement.

À mes grands-parents que je souhaite une bonne santé.

À mon cher frère et mes chères sœurs.

À ma chère nièce « Loujaine ».

À mes deux grandes familles, maternelle et paternelle.

À tous mes proches, mes amis, mes collègues de classe.

À ma chère binôme « Mounia » et à toute sa famille.

À notre promoteur Mr BOUBAKIR Chaâbane.

À tous nos profs de la spécialité.

Et enfin, à tous ceux et celle que j'aime dans ma vie.

Latifa

Dédicaces

Je dédie ce modeste travail à ma chère maman et à la mémoire de mon papa.

À ma petite sœur, mon grand frère.

À tous mes oncles, tantes, cousins, cousines et tous les membres des deux familles, maternelle et paternelle.

À tous mes proches, mes amis, mes collègues de classe.

À toute la famille ZERIQUEL, plus particulièrement à « Latifa » qui est une chère amie avant d'être mon binôme et mon collègue.

À notre promoteur Mr BOUBAKIR Chaâbane.

À tous nos profs de la spécialité.

Et enfin, à tous ceux qui m'ont aidé dans mon parcoure.

Mounia

Sommaire

Chapitre I : Généralités sur les voix pathologiques

Chapitre II : Méthodes d'évaluation et de classification de la parole pathologique

Chapitre III : Méthodes de classification de la parole pathologique

Chapitre IV : Implémentations, résultats et interprétations de la classification

Liste des figures

Liste des tableaux

Liste des abréviations

ADC	Analog to Digital Converter
APQ	Amplitude Perturbation Quotient
AUC	Area Under Curve
DFT	Discret Fourier Transform
FFT	Fast Fourier Transform
FN	Faux Négatif
F_0	Fréquence fondamentale
FP	Faux Positif
GRBASI	Grade, Roughness, Breathiness, Asthenia, Strain, Instability
HNR	Harmonic to Noise Ratio
KKT	Karush Kuhn et Tucker
LR	Likelihood Ratio
MAJ	Mean Absolute Jitter
MAS	Mean Absolute Shimmer
MEEI	Massachusetts Eye and Ear Infirmary
MFCC	Mel Frequency Cepstral Coefficients
NHR	Noise Harmonic Ratio
ORL	Oto-Rhino-Laryngologie
PPQ	Pitch Perturbation Quotient
RAL	Reconnaissance Automatique du Locuteur
RAP	Relative Average Perturbation

RBF	Radial basis function
ROC	Receiver Operating Characteristic
SE	Sensibilité
SP	Spécificité
SVD	Saarbrücken Voice Database
SVM	Support Vector Machines
TAP	Traitement Automatique de Parole
TBC	Taux de Bonne Classification
TCD	Transformée en Cosinus Discrète
TFN	Taux de faux négatifs
TFP	Taux de faux positifs
TMP	Temps Maximum de Phonation
VN	Vrai Négatif
VP	Vrai Positif

Introduction générale

La voix est un phénomène multidimensionnel composé d'un certain nombre d'éléments qui contribuent à sa qualité globale et à son intelligibilité. Une altération de la production de la parole est fréquemment représentée par une dysphonie ou une voix dysphonique. La dysphonie est une qualité perceptive de la voix qui indique que certains changements négatifs se sont produits dans les organes de phonation. Le terme dysphonie signifie littéralement voix anormale / difficile / altérée / voix pathologique.

De nos jours, les troubles de la voix augmentent considérablement en raison du mode de vie moderne. La relation entre la pathologie de la voix et les caractéristiques de la voix acoustique a été cliniquement établie et confirmée à la fois quantitativement et subjectivement par des experts de la parole. Les principales méthodes utilisées par la communauté médicale pour évaluer le système de production de la parole et diagnostiquer les pathologies sont soit des méthodes directes qui nécessitent une inspection directe des cordes vocales et provoquent une gêne pour le patient, soit des méthodes subjectives dans lesquelles la qualité de la voix est évaluée directement par un médecin expert.

La détection et l'évaluation de l'état d'un patient est l'étape la plus cruciale et importante dans le diagnostic d'une pathologie. Pendant la consultation de nombreux problèmes peuvent être rencontrés par le médecin ou par le malade qui compliquent l'évaluation et conduisent à des mauvais résultats de diagnostic tel que : la difficulté de la prise de décision par la méthode de l'évaluation subjective, absences ou inefficacité des moyens de diagnostic, complexité des cas (maladies rares chroniques et urgentes) et problème de suivi du patient (le coût des soins, incapacité de déplacement, mauvaise qualité de traitement, …). Pour réduire le coût du diagnostic et aider les médecins à diagnostiquer avec précision les troubles vocaux, il y a eu une croissance récente de l'utilisation des techniques de traitement du signal vocal et de l'analyse des données pour détecter et diagnostiquer avec précision les individus.

En raison de sa nature non invasive, l'évaluation automatique de la pathologie vocale est fortement considérée comme un outil de dépistage primaire ou un outil d'aide pour le clinicien. Un système d'évaluation automatique peut discriminer entre les échantillons normaux et pathologiques et de classer les pathologies de la voix. Le processus de différenciation entre les sujets normaux et pathologiques est un problème à deux classes appelé détection de pathologie. En revanche, la discrimination entre les différents types de pathologies est un problème multi-classes appelé classification des pathologies [1].

La détection automatique et la classification des pathologies est un domaine d'actualité et toujours exploré par la communauté des chercheurs [1-4]. Une large gamme de paramètres acoustiques a été utilisée pour la détection de pathologie à savoir le pitch, le jitter, le shimmer, le rapport harmonique sur bruit (HNR : Harmonics to Noise Ratio), l'énergie de bruit normalisée (NNE : Normalized Noise Energy), les coefficients cepstraux (MFCC : Mel-Frequency Cepstral Coefficients), etc. Dans le domaine de la détection automatique de la pathologie vocale, divers classificateurs ont été proposés tels que le perceptron multi-couches, le modèle de mélange gaussien, le réseau neuronal probabiliste, l'analyse discriminante linéaire, le classificateur de k plus proche voisin (KNN : K-Nearest Neighborhood), les machines à support de vecteurs (SVM : Support Vector Machine), etc.

L'objectif de notre mémoire est l'étude, l'implémentation, le choix des meilleurs paramètres acoustiques et les mesures de performances de la détection et la classification des pathologies de la voix par SVM.

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 (pitch, MFCC).

Dans le deuxième chapitre, nous aborderons l'évaluation subjective et objective de la voix, les différents indices de perturbation du pitch et de l'intensité, le rapport harmonique sur bruit, ainsi que la présentation des résultats d'extraction de ces indices.

Le troisième chapitre sera réservé aux différentes méthodes de la classification des données, suivi par une présentation détaillée de la méthode de classification SVM, ses équations, le choix de ses paramètres et ses variantes.

Dans le quatrième 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.

I.1 Introduction :

L'importance de la communication orale dans la société humaine a poussé à la recherche scientifique sur la voix et sa nature, et sur le fonctionnement vocal. L'utilisation de la voix dans différents domaines fonctionnelles (enseignement, musique, ...) conduit dans la plus part du temps à des troubles vocaux, et avec le développement du domaine médical qui a pu relier les changements et les dysphonies de la voix avec certaines troubles et pathologies dans le corps humain, tous cela dans le but de pouvoir arriver à traiter ses problèmes en connaissant leur source à partir de la voix, c'est pour cette raison que nous allons parler de la voix, des différentes pathologies de la voix, en les classifiant selon leur source et leur cause.

I.2 L'essentiel sur le mécanisme de production de la parole :

Les sons de la parole se produisent via le mouvement de nombreux muscles et organes de phonation. La coordination de ces organes est contrôlée par le système nerveux central.

Les organes de l'appareil phonatoire humain peuvent être classés en trois groupes principaux qui sont : les poumons et la trachée, le larynx et le conduit vocal [5]. La figure (I.1) montre une représentation simplifiée de l'appareil phonatoire humain.

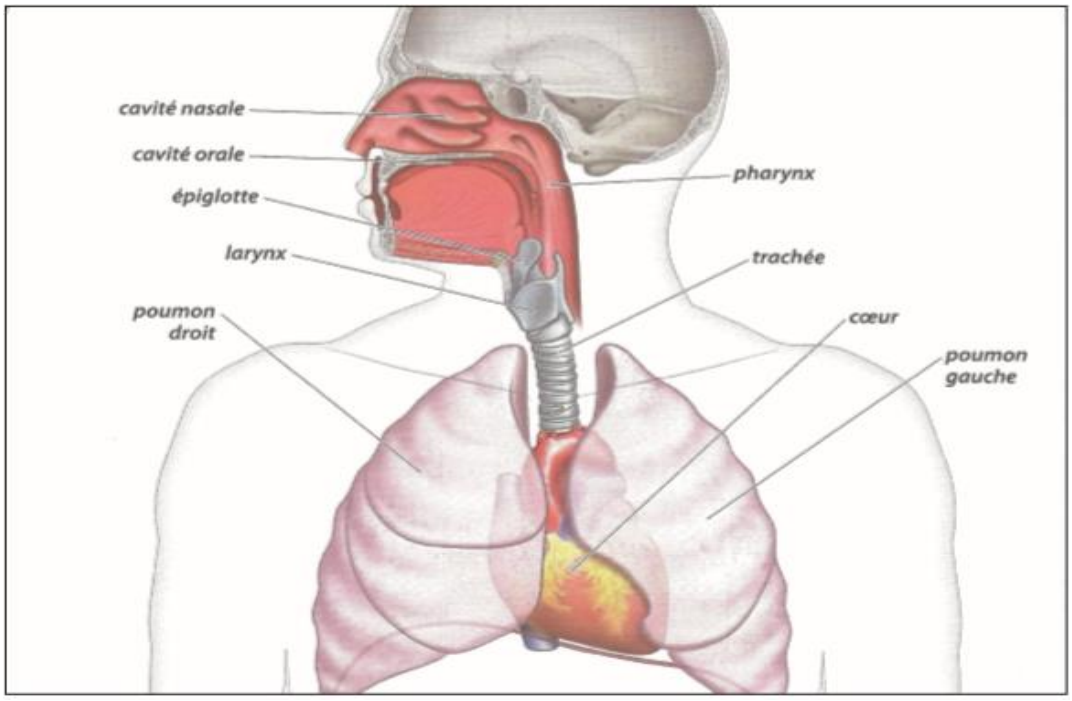

Figure I.1 : Anatomie de l'appareil phonatoire humain [6].

L'appareil phonatoire repose sur la partie haute de l'appareil respiratoire. Il est composé de trois systèmes [6] :

- **La soufflerie pulmonaire :** elle fournit l'énergie nécessaire à la production sonore. La production de tous les sons s'appuie sur un flux d'air qui provient des poumons et qui est nécessaire à la vibration des cordes vocales.

- **Le vibrateur :** le larynx, qui contient les cordes vocales, produit le son laryngé primaire.
- **Les résonateurs :** 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.

Comme le montre la figure (I.2), le larynx est constitué d'une armature cartilagineuse et d'un ensemble de muscles assurant sa fermeture ou son ouverture. Les plus volumineux de ces muscles sont représentés par les cordes vocales, qui fonctionnent comme une valve ou un sphincter. La mobilité des cordes vocales est sous le contrôle de plusieurs nerfs.

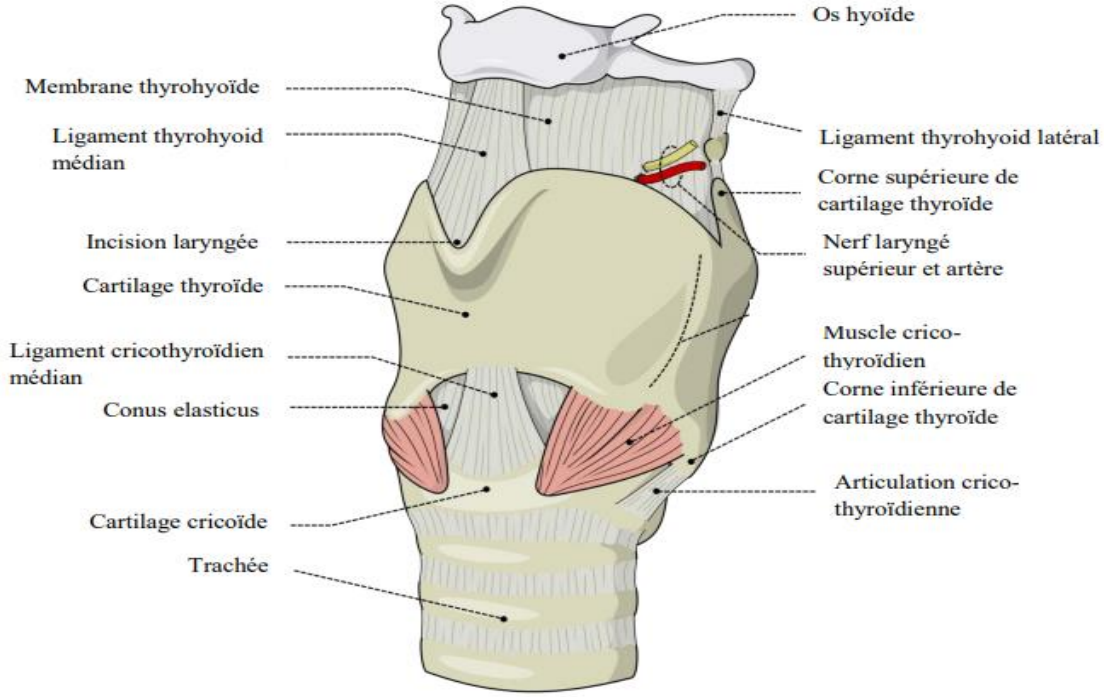

Figure I.2 : Anatomie du larynx [5].

La figure (I.3) montre l'aspect normal des cordes vocales lors de la respiration (cordes vocales ouvertes) et lors de la phonation (les cordes vocales se rapprochent et s'accolent).

Figure I.3 : Cordes vocales normales.

I.3 Les dysphonies :

La dysphonie est un trouble de la voix qui peut concerner son intensité, sa hauteur et son timbre. Plus précisément les changements dans ses paramètres ou toutes difficultés de fonctionnement du système phonatoire, cela peut être durable ou chronique, et due aux différents lésions, malformations, ou inflammations des organes du système phonatoire ou d'autre organe qui ont une certaine influence du loin ou du prés sur les cordes vocales ou le larynx, tous cela peut causer une dysphonie (modification dans les paramètres de la voix), ou une aphonie (la perte définitive de la voix ou pouvoir seulement chuchoter), on distingue :

I.3.1 Les dysphonies d'origines morphologiques :

Elles englobent les changements anatomiques de la glotte provoqués par les lésions des cordes vocales, comme les nodules, les polypes, les kystes, etc.

- **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'ampoule bilatéraux 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).

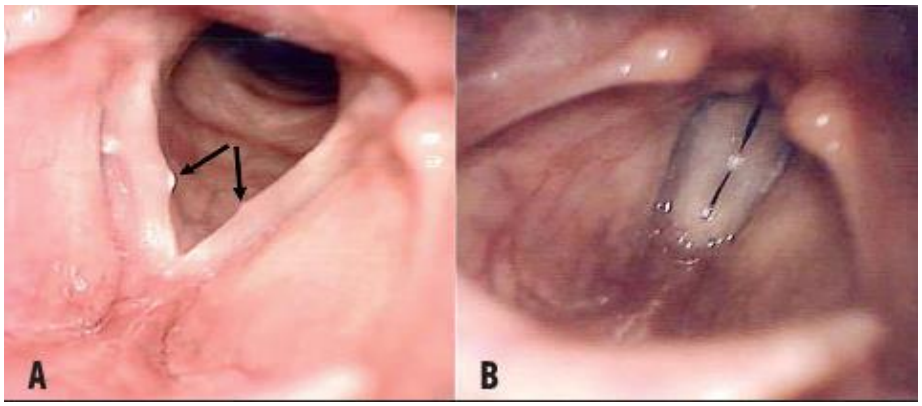

Figure I.4 : Nodules sur les cordes vocales.

- **Polypes :**

C'est une forme de tumeur non cancérigène (bénigne), de petite dimension, unilatéral et la plupart du temps d'une forme arrondie. Il est due au forçage de la voix, de la toux ou du tabagisme. Il apparait souvent chez les hommes que chez les femmes. Il peut rendre la voix râpeuse, enrouée, poussée et forcée en intensité. Les deux photos de la figure (I.5) présentent le cas d'un polype sur une corde vocale droite, en position d'ouverture des cordes vocales (A) et (B) en position de fermeture.

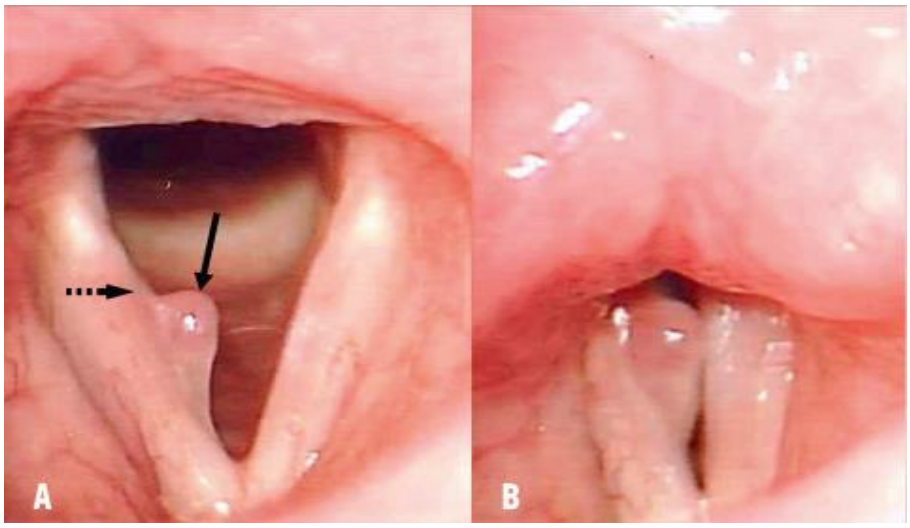

Figure I.5 : Polype de la corde vocale droite.

- **Kyste :**

C'est un relief intracordal (développé dans les cordes vocales) et couvre la muqueuse. Il gène la fermeture des cordes vocales donc les vibrations, ce qui provoque une voix enrouée, voilée et fatigable. La gêne des cordes vocales est souvent variable dans le temps, selon la quantité ou la présence ou l'absence du contenu de kyste, parce que des fois le kyste est troué ce qui lui permet de vider son contenu, donc plus il est vidé plus les cordes vocales peuvent s'accoler (figure I.4).

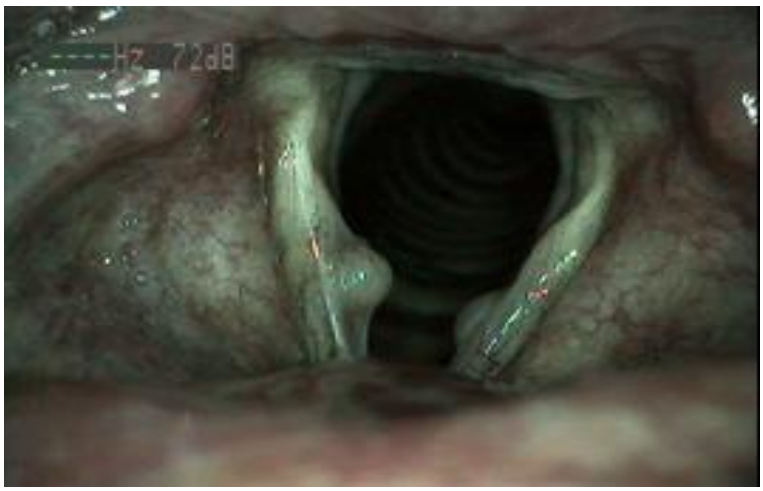

Figure I.6 : Exemple de kyste sur les cordes vocales.

- **Œdème de Reinke :**

Est une laryngite chronique œdémateuse, les cordes sont soufflées et lourdes (figure I.7), ce qui rend la voix rauque et très grave, et d'une faible puissance. Elle est connue plus souvent chez les femmes que chez les hommes (inflammation aigue d'origine viral des cordes vocales).

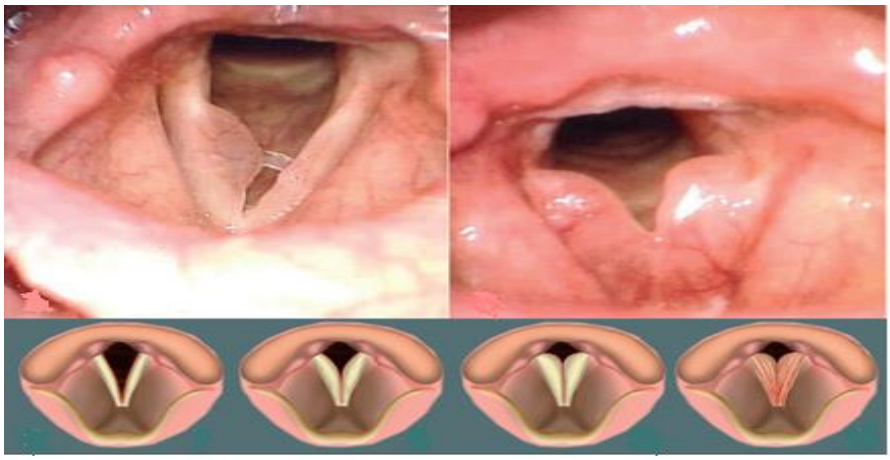

Figure I.7: Œdème de Reinke.

- **Tumeur cancérigène :**

Elle est présentée par des changements tissulaires dans le larynx ou les cordes vocales ainsi par des lésions irrégulières et une couleur blanchâtre ou rougeâtre (figure I.8), elle provoque en général des douleurs, une dysphonie persistante (voix enrouée, d'une faible intensité et voilée), ou elle peut parfois atteindre une aphonie.

Figure I.8 : Différence entre un larynx normal et un cancéreux.

- **Traumatismes chirurgicaux :**

Les plus importants changements anatomiques du larynx sont provoqués par des traumatismes chirurgicaux à la suite de l'ablation d'un cancer, d'un polype, ou n'importe quelle lésion au niveau de larynx. Ainsi que d'autres interventions sur le système phonatoire et des organes associés à l'appareil phonatoire. La voix est très dégradée, quoique fonction de la technique chirurgicale, avec parfois des désonorisations. Elle est grave, de faible intensité, mais intelligible sauf dans le bruit. Le timbre est très rauque.

I.3.2 Les dysphonies d'origines neurologiques :

Dans le cas d'un examen laryngoscopique où le résultat ne révèle aucun changement structurel ou organique des cordes vocales et de larynx, et il n'y a aucune lésion mais la voix comporte des troubles dysphoniques il s'agit de dysphonie neurologique, ils sont souvent

consécutifs aux problèmes cérébrales (tumeur, cavernome, Alzheimer, etc.), ou aux problèmes dans le système nerveux (des atteintes des nerfs commandant le système phonatoire). Les dysphonies d'origines neurologiques les plus citées sont :

- **L'hypotonie et l'hypertonie :**

Les dysphonies dysfonctionnelles se manifestent au niveau de l'appareil phonatoire soit par un excès de fonction, on parle alors d'hypertonie, soit par une insuffisance de fonction, on parle alors de l'hypotonie.

L'estimation du tonus musculaire fait classiquement partie de l'examen neurologique. Il est défini comme la résistance qui mobilise passivement les membres ou un segment de membre où les organes du patient, cas de résistance excessive par rapport à une situation normale, on parle d'hypertonie et dans le cas contraire, d'hypotonie. Il s'agit d'une notion subjective, car difficilement quantifiable. Les modifications du tonus musculaire et les signes qui y sont associés permettent toutefois d'orienter le diagnostic clinique vers une entité nosologique précise.

- **Tremblements :**

Les tremblements sont des mouvements involontaires, rythmiques, oscillatoires réciproques de groupes musculaires antagonistes, impliquant généralement les mains, la tête, le visage, les cordes vocales, le tronc ou des jambes, qui manifeste par conséquence des changements de la hauteur, d'intensité et le timbre de la voix.

- **Le parkinson :**

Est une maladie dégénérative qui touche les cellules nerveuses du Locus Niger, permet de perdre l'initiative motrice et une diminution de mouvement corporelle. Les troubles de la paroles dans la maladie de parkinson associent une dysphonie (altération des qualités acoustiques de la voix) et une dysarthrie (trouble de l'articulation de la parole).
Cette dysphonie dysarthrie touche les différents effecteurs de la chaîne parlée : la respiration, la phonation, les résonances (nasalisation), l'articulation et la prosodie (mélodie, intensité, durée).

- **Les dysphonies Spasmodiques** :

C'est une atteinte dystonique des muscle abducteur et adducteur du larynx (adduction : fermeture excessive de la glotte en phonation, abduction : prédominance des spasmes respiratoire, ouverture excessive de la glotte en phonation) ; causés par les problèmes dans la commande nerveuse, qui gêne l'excitation (l'énervation) des cordes vocales**,** ou la mobilité de larynx, elle est très rare et apparait plus particulièrement chez les femmes que chez les hommes.

Dans la dysphonie spasmodique, la voix est forcée, hachée, et le temps maximum de phonation est baissé (des coupures de la voix), elle peut devenir une aphonie (des chuchotements non audibles).

- **Paralysie laryngée :**

C'est un dysfonctionnement des cordes vocales présenté par une immobilité (paralyse) unilatérale (seulement une corde qui est paralysée) ou bilatérale (les cordes sont paralysées), dans le cas bilatéral les cordes sont souvent fixées à des positions variables soit en position médiane, intermédiaire ou latérale, et la personne souffre d'une aphonie totale. Alors au cas d'une paralysie unilatérale la voix est essoufflée, nasonnée ou bitonale associée à une dysphonie importante.

I.4 Paramètres de la voix :

Correspond aux différents caractéristiques et valeurs allouées à une voix pour pouvoir la décrire et la mesurer, elles se représentent par [7,8] :

I.4.1 Paramètres de description de la voix :

I.4.1.1 La hauteur :

Elle correspond à la fréquence d'ouverture et de fermeture des cordes vocales, présente le nombre de cycle de vibration (ouverture et fermeture) des cordes par seconde.

La voix peut être grave ça veut dire d'une fréquence basse, medium ou aigue ; elle varie selon la taille de larynx, plus les cordes sont épaisses plus la voix est grave ce qui est vue plus fréquemment chez les hommes ; et plus les cordes sont fines plus la voix est aigue ce qui est apparu chez les femmes et encore plus chez les enfants.

I.4.1.2 L'intensité :

Dans le langage courant cela correspond à parler fort ou doucement. Elle présente l'amplitude de la puissance de signal acoustique mesurée en décibel (dB) ; elle est réglée par la compression de l'aire envoyé par les poumons.

Lorsque l'intensité de la voix est faible on parle d'une hypotonie, et lorsque l'intensité est forte on parle d'une hypertonie.

I.4.1.3 Le timbre :

Représente l'identité de la voix d'une personne, il dépend des fréquences contenues dans un son qui se superposent, donc de la manière dont les cordes vocales s'accoles entre eux, ainsi que la mobilité du voile du palais, de la langue et des lèvres, et enfin la qualité de la muqueuse qui tapisse la cavité de résonance.

I.4.1.4 La durée :

Est l'intervalle de temps de la tenue des sons voisés, dépend souvent de la pression de l'aire expirée des poumons, avant de détendre. On peut la présenter par la période de prononciation d'une voyelle (un son voisé) sans arrêt, jusqu'à la prochaine aspiration de l'air.

I.4.1.5 Les formants :

Représentent le conduit vocal, possède 4 ou 5 fréquences de résonnance de résonateur, l'organisation de ses fréquences qui permet de distinguer les différentes voyelles entre elles, plus l'amplitude des formant est élevée plus la discrimination des voyelles est plus nette. Tout changement de conduit peut changer la fréquence de formant. Les valeurs de ces fréquences sont entre 500 Hz à 3500 Hz.

I.4.2 Paramètres acoustiques utilisés :

L'analyse du signal de la parole d'un patient est aujourd'hui une technique très utilisée pour la détection et la classification des pathologies de la voix, ces pathologies peuvent être identifiées par l'analyse de plusieurs paramètres de signaux acoustiques. Seulement les paramètres qui seront utilisés dans nos simulation seront présentés.

I.4.2.1 MFCC (Mel-scaled Frequency Cepstral Coefficients):

Les Coefficients spectraux de fréquence à l'échelle de Mel sont les paramètres les plus répandus dans le traitement de la parole, utilisés pour la reconnaissance automatique de la voix, du locuteur et des langues, ainsi que la détection et la classification des pathologies de la voix.

Le principe de calcul des MFCC est issu des recherches psycho-acoustiques sur la perception des différentes bandes de fréquences par l'oreille humaine. L'intérêt principal de ces coefficients est d'extraire des informations pertinentes en nombre limité en s'appuyant à la fois sur la production (théorie Cepstrale) et sur la perception de la parole (échelle des Mels). La figure (I.9) montre les phases d'obtention des coefficients MFCC :

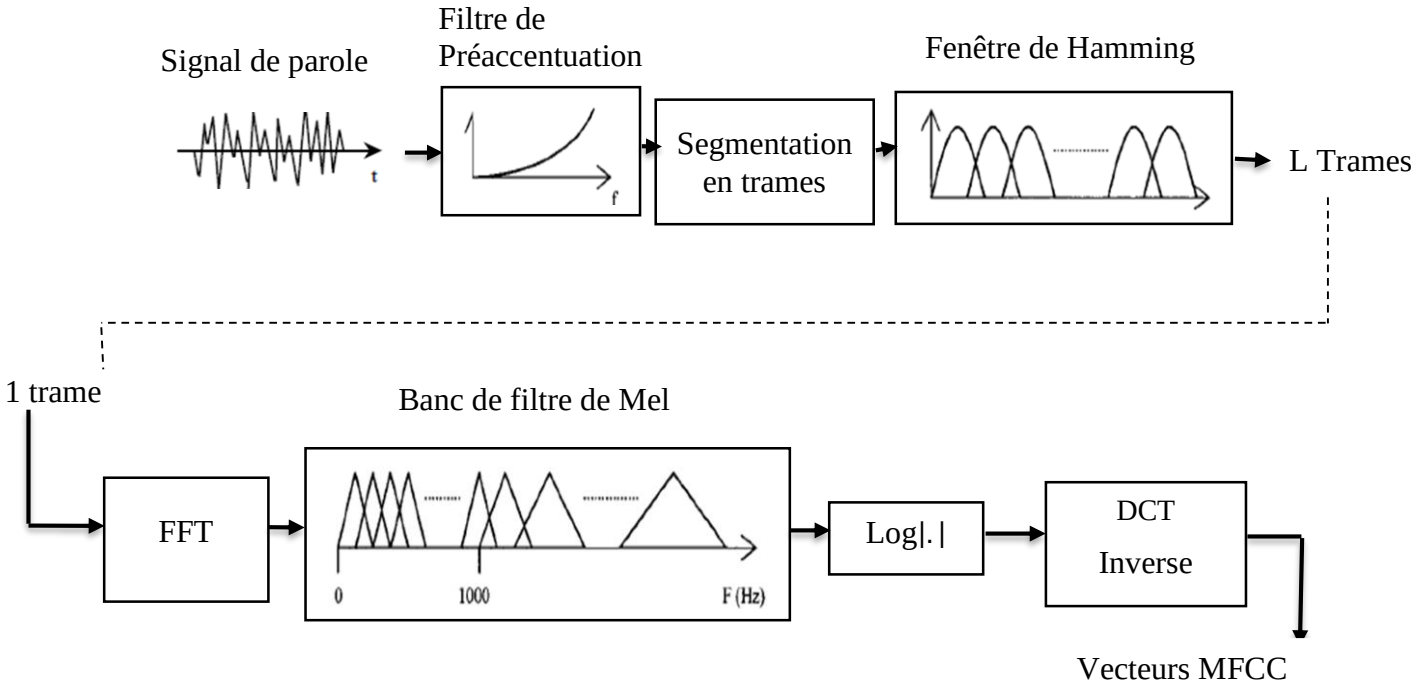

Figure I.9 : Phases d'obtention des coefficients MFCC à partir d'un signal de la parole.

Plusieurs étapes sont nécessaires pour transformer un fichier audio en Cepstre MFCC. Dans chaque étape nous donnerons la fonction réalisée et les approches mathématiques utilisés comme suite [9 - 11] :

- **Préaccentuation :**

Dans ce processus, nous mettons l'accent sur les fréquences les plus élevées ; cela augmentera l'énergie du signal aux fréquences élevées. Cette étape consiste à faire passer le signal dans un filtre numérique à réponse impulsionnelle finie (RIF) de premier ordre donné comme suit :

$$H(\mathrm{z}) = 1 - \alpha \mathrm{z}^{-1} \qquad \text{Avec } 0.9 \leq \alpha \leq 1 \tag{I.1}$$

Ainsi, le signal préaccentué s_1 dans la sortie du bloc de préaccentuation est donnée par:

$$s_1(n) = s(n) - \alpha s(n-1) \tag{I.2}$$

Où s(n) est le signal d'entrée.

- **Segmentation en trames :**

Les signaux vocaux sont des signaux non stationnaires. Cependant, sur un court intervalle, ils peuvent être considérés comme quasi stationnaires. C'est pourquoi il est utile de diviser le signal de parole en segments successifs stationnaires.

Dans cette étape de segmentation, le signal préaccentué est ainsi découpé en trames de N échantillons de parole. En général N est fixé de telle manière à ce que chaque trame corresponde à environ 20 à 30 ms de parole. Deux trames successives sont chevauchées de M échantillons selon le pourcentage de chevauchement.

- **Le Fenêtrage :**

La segmentation du signal en trames produit des discontinuités aux frontières des trames, donc le but du fenêtrage est de réduire l'effet résultant du processus de segmentation.

Le fenêtrage dans le domaine temporel est une multiplication point par point entre le segment et la fonction fenêtre. Selon le théorème de convolution, ceci correspond à une convolution du spectre court terme avec la réponse d'amplitude de la fonction fenêtre.

En général, une fonction fenêtre appropriée diminue aux bords de segment de sorte que l'effet des discontinuités est diminué.

La fonction fenêtre la plus utilisée dans le traitement de la parole est la fenêtre de Hamming, car elle entraîne un minimum de distorsion spectrale du signal de parole, par rapport aux autres fenêtres. Elle est définie par l'équation ci-dessous :

$$h(n) = \begin{cases} 0.54 - 0.46cos\left(2\pi\frac{n}{N}\right), 0 \leq n \leq N-1 \\ 0, autrement \end{cases} \quad (I.3)$$

Alors, l'application de la fenêtre de Hamming au signal s_1 nous fournit le signal fenêtré suivant :

$$s_2(n) = s_1(n) \times h(n) \quad (I.4)$$

- **La Transformée de Fourier Rapide (FFT) :**

Dans cette étape, nous sommes amenés à appliquer la transformée de Fourier rapide (TFR ou en anglais FFT : Fast Fourier Transform) sur les trames obtenues précédemment afin de les transposer dans le domaine fréquentiel.

La FFT est un algorithme de calcul de la transformée de Fourier discrète (TFD ou en anglais DFT : Discrete Fourier Transform). Ainsi, pour le temps de calcul de l'algorithme rapide peut être 100 fois plus petit que le calcul utilisant la formule de définition de la TFD. La figure suivante illustre la densité spectrale de puissance d'une portion de la voyelle /i/.

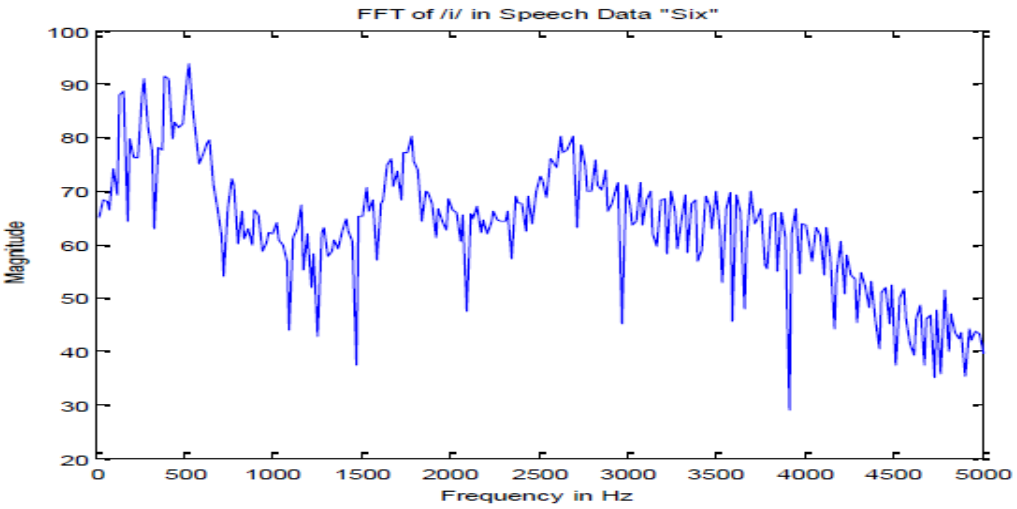

Figure I.10 : Exemple de FFT [11] .

- **Banc de filtre à échelle de Mel :**

Comme le signal de parole contient plus d'informations dans la gamme des basses fréquences, il faut un plus grand nombre de filtre dans cette gamme, ce qui incite à utiliser un banc de filtre.

Le banc de filtre Mel est appliqué dans le domaine fréquentielle avant le logarithme et la TCD. Le but du banc Mel est de simuler les filtres des bandes critiques du mécanisme d'audition. Les filtres sont également espacés sur une échelle Mel, et habituellement ils ont une forme triangulaire (figure I.11).

Dans l'échelle de mesure Mel, la correspondance est approximativement linéaire sur les fréquences au-dessous de 1*kHz* et logarithmique sur les fréquences supérieures à celle-ci.

L'échelle de Mel peut être approximée par la formule de conversion suivante :

$$Mel(f) = 2595log_{10}\left(1 + \frac{f}{700}\right) = 1125ln\left(1 + \frac{f}{700}\right) \quad \text{(I.5)}$$

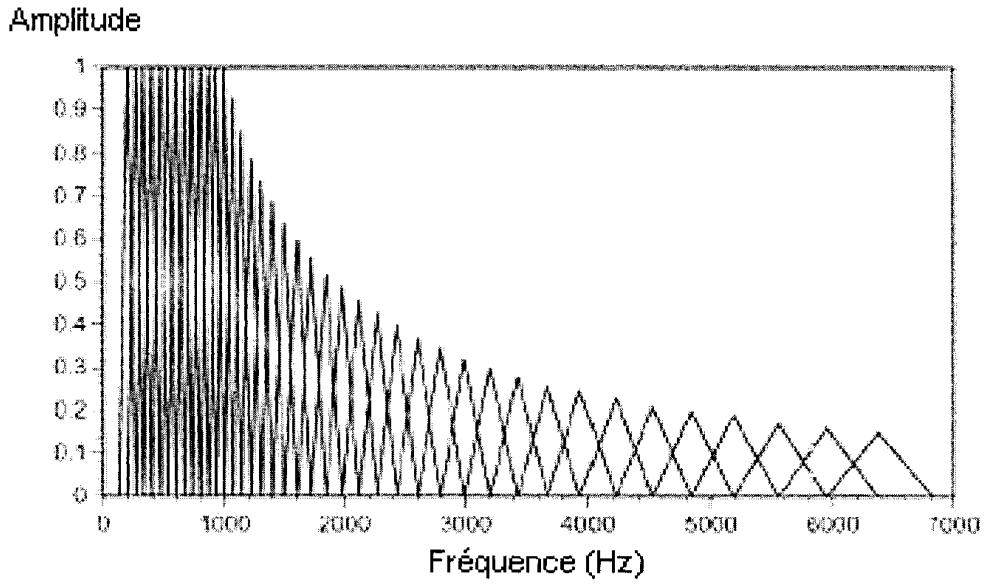

Figure I.11 : Approximation de l'échelle de Mel : Banc de filtres triangulaires.

- **Calcul de logarithme :**

Les sorties du filtre triangulaire sont comprimées en utilisant le logarithme. Le passage dans le domaine log-spectral permet de déconvoluer le signal et de le compresser. Par conséquent, cette étape convertit simplement la multiplication de l'amplitude de la transformée de Fourier en addition.

- **La Transformée en Cosinus Discrète (TCD)** :

La transformée en cosinus discrète (TCD ou en anglais DCT : Discrete Cosine Transform) inverse, appliquée au logarithme du vecteur d'énergie du signal obtenu en sortie du banc de filtres de Mel.

Nous utilisons ce processus pour changer le spectre Mel en domaine temporel. Ce faisant, nous avons obtenu les coefficients de cepstre de la fréquence Mel (MFCC).

$$MFCC(i) = \sum_{j=1}^{N_f} S(j)\, cos\left(i\left(j - \frac{1}{2}\right)\frac{\pi}{N_f}\right), i = 1,2,\dots,M \quad \text{(I.6)}$$

Où N_f indique le nombre de filtres Mel utilisés, M est le nombre de coefficients MFCC et S(j) est le logarithme de l'énergie obtenue avec le filtre triangulaire d'indice «*j*».

Généralement, seuls les 13 premiers coefficients MFCC (de MFCC (0) à MFCC (12)) sont retenus. Le premier coefficient, MFCC (0), représente l'énergie moyenne dans la trame de la parole ; MFCC (1) reflète la balance d'énergie entre les basses et les hautes fréquences.

I.4.2.2 Le pitch :

La mélodie de la voix résulte de la vibration des cordes vocales et se traduit par la fréquence de vibration laryngienne (fréquence fondamentale). Le pitch représente la fréquence fondamentale perçu par l'oreille, il existe au cours de l'émission des sons voisées lors de la prononciation des voyelles et quelque consonnes (b, d, g, ...) [12]. Il est très utile dans le codage de la parole, le discernement de la voix féminine ou masculine, la séparation des différentes phrases et c'est un paramètre sur lequel on peut distinguer une voix pathologique et une voix saine, etc. La plage de variation moyenne de cette fréquence varie d'un locuteur à l'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.

Plusieurs travaux de recherches, mémoires et thèses ont été consacrés à l'estimation de la fréquence du pitch où les détails des différentes méthodes ont été présentés. Dans le cadre de notre travail, nous allons utiliser le logiciel PRAAT [13] pour l'extraction du pitch et ses différentes mesures. Par ailleurs, le schéma global d'estimation du pitch contient trois phases :
Phase de prétraitement réservée à la préparation de signal issu d'un microphone, choisir la durée des trames d'analyses et du recouvrement afin de moins compromettre la condition de stationnarité.
Phase de traitement réservée à l'extraction de F0 par plusieurs algorithmes appartenant à l'une des trois catégories temporelles, fréquentielles et statistiques.
Phase de post-traitement qui a pour rôle de diminuer les erreurs d'estimation qui sont :

- Erreur de voisement : la présence de F0 dans des zones non voisées et son absence dans des zones voisées.
- Erreur grossière : F0 correspond à une harmonique ou sous harmonique.
- Erreur fine : la valeur trouvée est située à plus ou moins de 10 % de la valeur réel.

I.5 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. De plus, 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 mesures de perturbations du pitch. Les définitions et les différentes étapes d'obtention de ces paramètres ont été détaillé à la fin de ce chapitre, les autres seront abordés au chapitre suivant.

II.1 Introduction :

Les différents troubles de la voix (altération du timbre, souffle, raucité, instabilité, voix bitonale, essoufflement, diminution de l'intensité, …) aident à distinguer la présence ou l'absence d'une voix pathologique et à révéler les différents dysfonctionnements de la production de la voix ou les lésions organiques dont il peut souffrir un patient, le diagnostic peut être réalisé par les patients eux-mêmes ou par un phoniatre ou par un ORL (Oto-Rhino Laryngologie), on parle dans ceci d'une évaluation subjective; il y a aussi l'évaluation objective qui est basée sur l'analyse des paramètres ou des mesures physiques du son obtenu à l'aide des microphones et des différents capteurs, ainsi que les différents logiciels utilisés dans l'analyse des paramètres acoustiques.

II.2 Évaluation des troubles de la voix :

L'évaluation de la voix est une forme d'analyse et de consultation phoniatrique. Elle est basée d'une part sur une évaluation perceptive, à l'oreille, de la qualité de la voix (évaluation subjective), et d'autre part, sur une analyse instrumentale basée sur des mesures et des paramètres acoustiques et aérodynamiques du son en utilisant des capteurs, des applications et des techniques de traitement de signal (évaluation objective).

II.2.1 Évaluation subjective :

Basée sur l'opinion d'un expert (phoniatre, médecin) ou sur l'auto-évaluation, elle peut être réalisée sur trois échelles : les données de l'anamnèse, l'échelle d'auto-évaluation et l'échelle d'évaluation perceptuelle [14].

Les données de l'anamnèse :

Sur l'échelle de l'anamnèse, l'examen est basé sur un ensemble de données sur le patient pris et recueillie par un phoniatre ou par lui-même pour détecter le problème vocal dans son contexte. Ces données portent sur :

- L'histoire médical du patient : problèmes vocaux déjà rencontrés par le patient, les troubles médicaux qu'il a déjà vécus (trouble respiratoire, de la voix, neurologiques et hormonaux), intervention passée (chirurgie abdominale, thoracique cervicale, faciale, cérébrale, …) et traitements ou soins médicaux consécutives à ces troubles.

- Contexte sociale et professionnelle : (situation sociale du patient, utilisation quotidienne de sa voix, condition de travail, …).

- Les facteurs déclenchants et favorisants une dysphonie : qu'on peut les trouver au sein de pathologies médicales, où liées à un mode de vie particulier.

- L'avis du patient ainsi que ses contraintes : le ressenti du patient face à sa voix et ces plaintes (douleur, difficulté de respiration, diffèrent changement de la voix, …).

Échelle d'auto-évaluation

Il existe plusieurs questionnaires d'auto-évaluation par le patient de la qualité vocale comme le « Voice Outcome Survey », « Voice-Related Quality of Life Measure », « Voice Handicap Index », etc. L'échelle la plus facile, fiable et la plus utilisée est celle du VHI qui a l'avantage de quantifier l'impact d'une variété de troubles vocaux sur la qualité de vie. Elle enrichit donc l'anamnèse du patient ; tout en prenant compte des difficultés vocales rencontrées dans la vie quotidienne du patient.

Échelle d'évaluation perceptuelle :

Basée principalement sur l'écoute de la voix par l'oreille, c'est la méthode la plus courante dans les cliniques médicales, elle consiste à faire juger la voix du patient par un auditeur qui a la capacité de bien évaluer la qualité de la voix et sa fiabilité, on parle des experts du domaine (phoniatres, orthophonistes), qui ont le rôle de fournir un grade de dysphonie sur une échelle.

L'échelle d'évaluation perceptive subjective, la plus largement utilisée au monde est l'échelle GRBAS, proposée par la « Japanese Society of Logopedics and Phoniatrics » et développée par Hirano (1981). Elle consiste à faire lire un paragraphe aux patients ensuite soumis à des diverses juges expérimentées, en lui donnant une note entre 0 et 3 pour chacun des critères ou des paramètres suivants :

- G : grade (Grade), cela correspond à une évaluation générale et globale de la qualité de la voix (0 : voix normale, 1 : dysphonie légère, 2 : dysphonie moyenne, 3 : dysphonie sévère).
- R : raucité (Roughness), il s'agit d'évaluer la raucité de la voix et toutes les altérations du timbre (éraillure, craquement), ainsi que la régularité des vibrations des plis vocaux.
- B : le souffle (Breathiness), c'est la composante du souffle dans la voix et elle est liée directement à la présence d'une fuite d'aire lors de la phonation.
- A : la sensation de faiblesse (Asthenia), qui traduit un manque de puissance dans la voix liée à une intensité faible où au manque d'harmoniques aigues.
- S : la sensation de forçage (Strain), on observe ici le forçage vocal, l'hypertonie, en évaluant la sensation d'un effort important et d'une tension musculaire excessive lors de la production vocale.

II.2.2 Évaluation objective et indices utilisés :

L'analyse objective est la deuxième méthodologie proposée comme une alternative à l'évaluation perceptive pour pallier ses inconvénients et faiblesses [15], l'évaluation objective utilise les techniques de traitement automatique de la parole (TAP), elle permet de porter un

regard objectif sur les troubles et la qualité de la voix grâce à des mesures acoustiques et aérodynamiques pour extraire des indices acoustiques pertinents, permettant de déterminer les caractéristiques de la voix pour renseigner sur l'état du larynx du locuteur. Les cliniciens préfèrent des indices acoustiques qui sont corrélés avec les caractéristiques perceptuelles. Cependant, il serait vain de rechercher un accord acoustique-perceptuel lorsque l'évaluation perceptive n'est pas suffisamment fiable.

Les études de caractérisation des troubles de la voix sont consacrées au développement de méthodes d'analyse dédiées à l'estimation des dyspériodicités vocales dans le signal de parole. Les indices les plus citées et les plus utilisées seront présentées dans les paragraphes suivants.

II.2.2.1 Indices de perturbations du pitch :

- **Jitter :**

Est un paramètre acoustique très important qui permet la détection des micro perturbations de la voix due à la fluctuation à court terme correspondant à la variation de la période fondamentale pendant un cycle glottique. Le Jitter doit sa définition et sa naissance à Lieberman (1963) qui appuya le fait que lors d'une voyelle tenue [5], les perturbations laryngées sont en rapport non pas avec les variations lentes de l'amplitude et de la fréquence, mais plutôt avec les variations rapides de celles-ci. Sur le plan auditif, il correspond à une variation de la hauteur de la voix. Il apparaît que le Jitter serait le paramètre le plus corrélé à la raucité et au souffle. La figure (II.1) montre la variation du pitch d'un cycle à un autre dans un signal de parole.

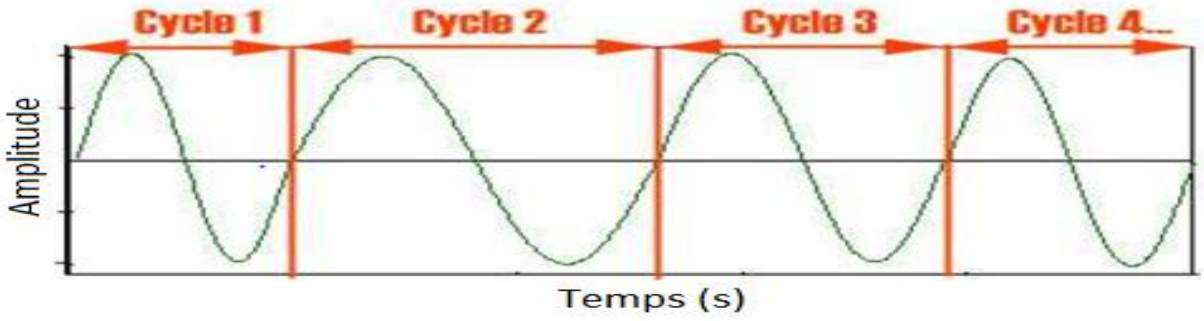

Figure II.1: Variations de fréquence (Jitter) [16].

Il existe plusieurs représentations du Jitter :

- Le jitter local absolu (Jitta) ou le MAJ (Mean Absolute Jitter) est la variation de cycle à cycle de la période fondamentale, c'est-à-dire la différence moyenne absolue entre des périodes consécutives, exprimée en [17] :

$$Jitta = MAJ = \frac{1}{N-1}\sum_{i=2}^{N}|T_i - T_{i-1}| \qquad \text{(II.1)}$$

Où T_i est la période fondamentale de la iéme trame avec $T_i=1/F_i$, N est le nombre de trames.

- Le jitter local relatif ou jitter (%) est la différence absolue moyenne entre des périodes consécutives, divisée par la période moyenne. Elle est exprimée en pourcentage par [18] :

$$Jitter(\%) = \frac{\frac{1}{N-1}\sum_{i=2}^{N}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \quad \text{(II.2)}$$

Selon le manuel de PRRAT le seuil pathologique de Jitter est : S_0=1.04 %

- Le jitter RAP (Relative Average Perturbation) est définie comme la moyenne relative de la perturbation, la différence absolue moyenne entre une période et la moyenne de celle-ci et de ses deux voisins, divisée par la période moyenne [18].

$$Jitter\ RAP\ (\%) = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|T_i - \frac{1}{3}(T_{i-1}+T_i+T_{i+1})\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} * 100 \quad \text{(II.3)}$$

- Le Jitter (PPQ5**)** est le Quotient de Perturbation Périodique à cinq points, calculé comme la différence absolue moyenne entre une période et la moyenne de celle-ci et de ses quatre voisines les plus proches, divisée par la période moyenne [18].

$$PPQ5 = \frac{\frac{1}{N-4}\sum_{i=3}^{N-2}\left|T_i - \frac{1}{5}(T_{i-2}+T_{i-1}+T_i+T_{i+1}+T_{i+2})\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} * 100 \quad \text{(II.4)}$$

II.2.2.2 Indices de perturbations d'amplitude :

- **Shimmer :**

C'est le paramètre de détection des fluctuations d'amplitude à court terme, il évalue l'instabilité de l'amplitude du signal de parole, d'un cycle glottique à l'autre. La figure suivante illustre les variations de l'amplitude qui seront quantifiées par plusieurs mesures de Shimmer :

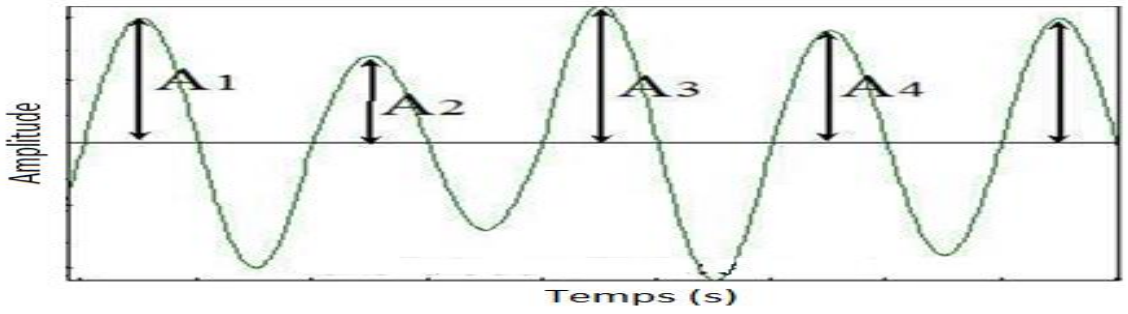

Figure II.2: Variations d'amplitude (Shimmer) [16].

- Le shimmer (dB) est exprimé comme la variabilité de l'amplitude crête à crête en décibels, c'est-à-dire le logarithme absolu moyen en base 10 de la différence entre les amplitudes de deux périodes consécutives, multiplié par 20 [17] :

$$Shimmer(dB) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20log\left(\frac{A_{i+1}}{A_i}\right)\right| \quad \text{(II.5)}$$

Où Ai sont les données d'amplitude crête à crête extraites et N est le nombre de périodes de fréquence fondamentale extraites.

- Le shimmer (relatif) est défini comme la différence absolue moyenne entre les amplitudes de deux périodes consécutives, divisée par l'amplitude moyenne, exprimée en pourcentage [17].

$$Shimmer(relatif) = \frac{\frac{1}{N-1}\sum_{i=2}^{N}|A_i - A_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \quad \text{(II.6)}$$

Le shimmer (APQ 3) est le Quotient de Perturbation d'Amplitude en trois points, la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de ses voisines, divisée par l'amplitude moyenne [18].

$$APQ3\ (\%) = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|A_i - \frac{1}{3}(A_{i-1}+A_i+A_{i+1})\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} * 100 \quad \text{(II.7)}$$

- Le shimmer (APQ5) est défini comme le Quotient de Perturbation d'Amplitude en cinq points, la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de cette période et de ses quatre périodes voisines les plus proches, divisée par l'amplitude moyenne [18].

$$APQ5\ (\%) = \frac{\frac{1}{N-4}\sum_{i=3}^{N-2}\left|A_i - \frac{1}{5}(A_{i-2}+A_{i-1}+A_i+A_{i+1}+A_{i+2})\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} * 100 \quad \text{(II.8)}$$

- Le shimmer (APQ11) est exprimé par le quotient de Perturbation d'amplitude à 11 points, la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de celle-ci et de ses dix périodes voisines les plus proches, divisée par l'amplitude moyenne [17].

$$APQ11\ (\%) = \frac{\frac{1}{N-10}\sum_{i=6}^{N-5}\left|A_i - \frac{1}{11}(A_{i-5}+\cdots+A_{i-1}+A_i+A_{i+1}+\cdots+A_{i+5})\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} * 100 \quad \text{(II.9)}$$

II.2.2.3 Rapport harmonique sur bruit :

- **HNR (Harmonic to Noise Ratio) :**

Le passage de l'air par les cordes vocales lors d'une phonation présente des interruptions quasi périodiques, ce qui explique le phénomène harmonique du signal vocal, alors que le passage turbulent et continue de l'air à travers la glotte correspond donc à du bruit, il est plus tôt connu par « Les inter harmoniques » qui sont différents des harmoniques, ils en résultent lorsque le fonctionnement du larynx se détériore. L'énergie des harmoniques est alors progressivement remplacée par du bruit sur le spectrogramme.

En grande partie le calcul du rapport (harmonie/bruit) permet d'enlever une portion du bruit du signal vocal où son timbre est pauvre d'harmonie [14].

Le seuil du rapport HNR d'une voix pathologique donnée par PRRAT est inférieur à 20 dB pour le phonème /a/, il est exprimé par la relation suivante :

$$HNR = 10 * log10\left(\frac{\mathrm{ACv(T)}}{\mathrm{ACv(0)-ACv(T)}}\right) \quad \text{(II.10)}$$

Où $\mathrm{ACv(0)}$ est le coefficient d'autocorrélation à l'origine qui est aussi l'énergie du signal, et $\mathrm{ACv(T)}$ est la composante d'autocorrélation correspondante à la période du pitch. La différence $\mathrm{ACv(0)} - \mathrm{ACv(T)}$ peut être supposée comme l'énergie du bruit.

- **NHR (Noise to Harmonic Ratio) :**

Le NHR est le rapport moyen de l'énergie des composantes inharmoniques, il évalue la présence de bruit dans le signal analysé (le rapport entre l'énergie non harmonique de l'intervalle 1500 Hz à 4500 Hz avec l'énergie harmonique de 70 Hz à 4500 Hz), la valeur de NHR tend vers zéro pour une voix normale, elle est autour de 0.1 pour une voix pathologique et elle augmente avec la sévérité de la dysphonie.

II.3 Exemple d'illustration :

Pour obtenir certaines caractéristiques représentatives de la parole, il était nécessaire de mettre en œuvre des fonctions d'extraction des paramètres acoustiques pour construire la matrice des caractéristiques de chacun des fichiers normaux ou pathologiques qui seront utilisés dans la classification. Dans notre travail, nous avons utilisé plusieurs paramètres comme les coefficients MFCC, les paramètres de variation du pitch (Jitter, shimmer) et les mesures (HNR, NHR).

II.3.1 Conditions d'implémentation :

La majorité de nos simulations et tests ont été faites sous MATLAB version 2015 (Matlab R2015b) qui est un environnement de calcul technique conçu pour le calcul numérique et la visualisation à haute performance. Ses dernières versions contiennent plusieurs fonctions et outils dédiés à la classification des signaux.

De plus, nous avons utilisés Praat qui est l'un des plus importants logiciels téléchargeables gratuitement sur internet en analyse acoustique de la parole et la manipulation de sons [13]. Ce logiciel, assez complet, donne accès à une analyse fine de la parole (analyse spectrale, analyse des formants, analyse de l'intensité, …), à des chiffres précis de mesures acoustiques (jitter, shimmer), à la création de graphiques et à des explorations statistiques.
La figure suivante illustre l'interface graphique du logiciel PRAAT où la forme d'onde d'un fichier de parole est présentée en haut, suivie par son spectrogramme en niveau de gris. Sur ce

dernier, on montre aussi les variations des valeurs du pitch (en bleu) et les différentes valeurs des cinq formants (F1, F2, …, F5) en rouge.

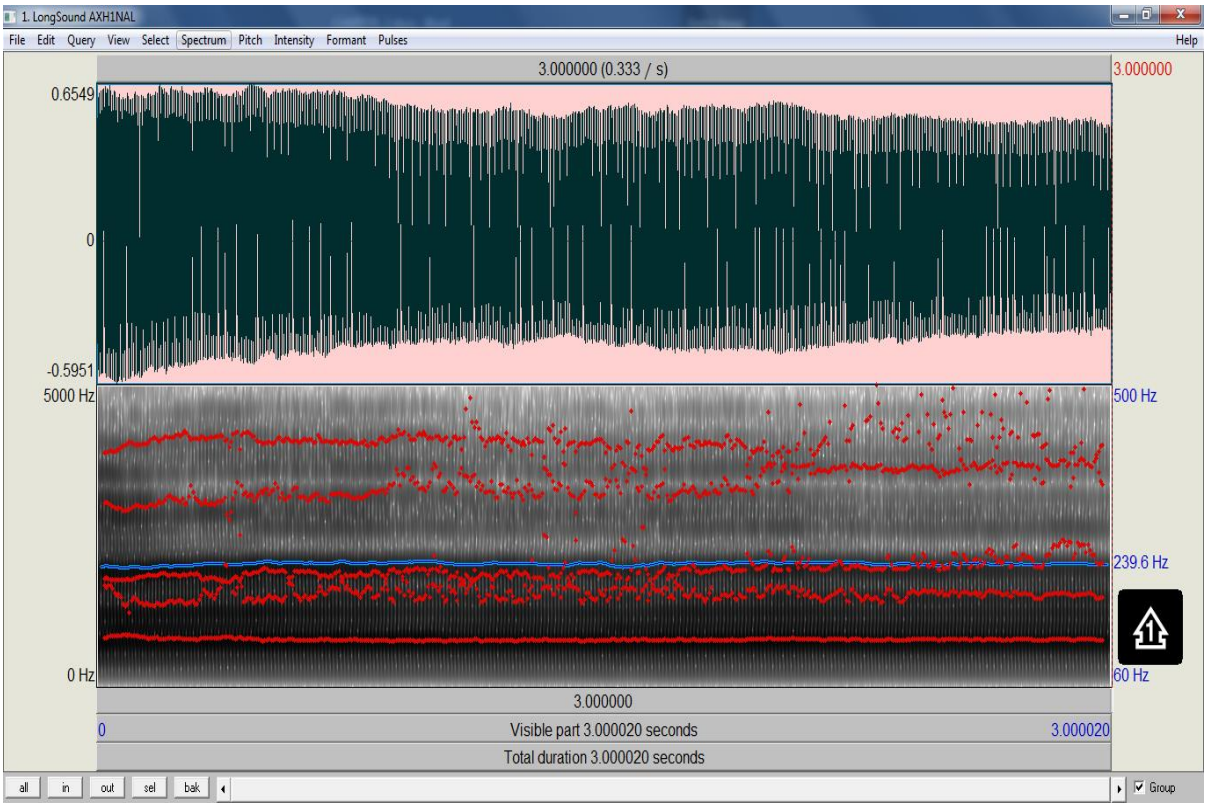

Figure II.3: Interface graphique du logiciel PRAAT.

II.3.2 Exemple d'extraction des paramètres MFCC :

Pour l'extraction des coefficients MFCC sous MATLAB, les paramètres suivants sont utilisés :

Une durée de trame d'analyse de 25 ms, un décalage de 10 ms et une fenêtre de pondération de Hamming ;

Un coefficient alpha = 0.97 de préaccentuation (preemphasing) ;

Un nombre des coefficients MFCC = 12 ;

Un nombre M = 20 de banc de filtres à échelle de Mel.

Ainsi, pour chaque fichier de parole échantillonné à une fréquence d'échantillonnage (fs) nous obtenons 12 coefficients MFCC pour chaque trame.

La figure suivante montre la forme d'onde d'un exemple de fichier de parole utilisé (voyelle /ah/ normal), l'énergie dans les bancs de filtres Log (mel) et le spectrogramme des coefficients MFCC du signal.

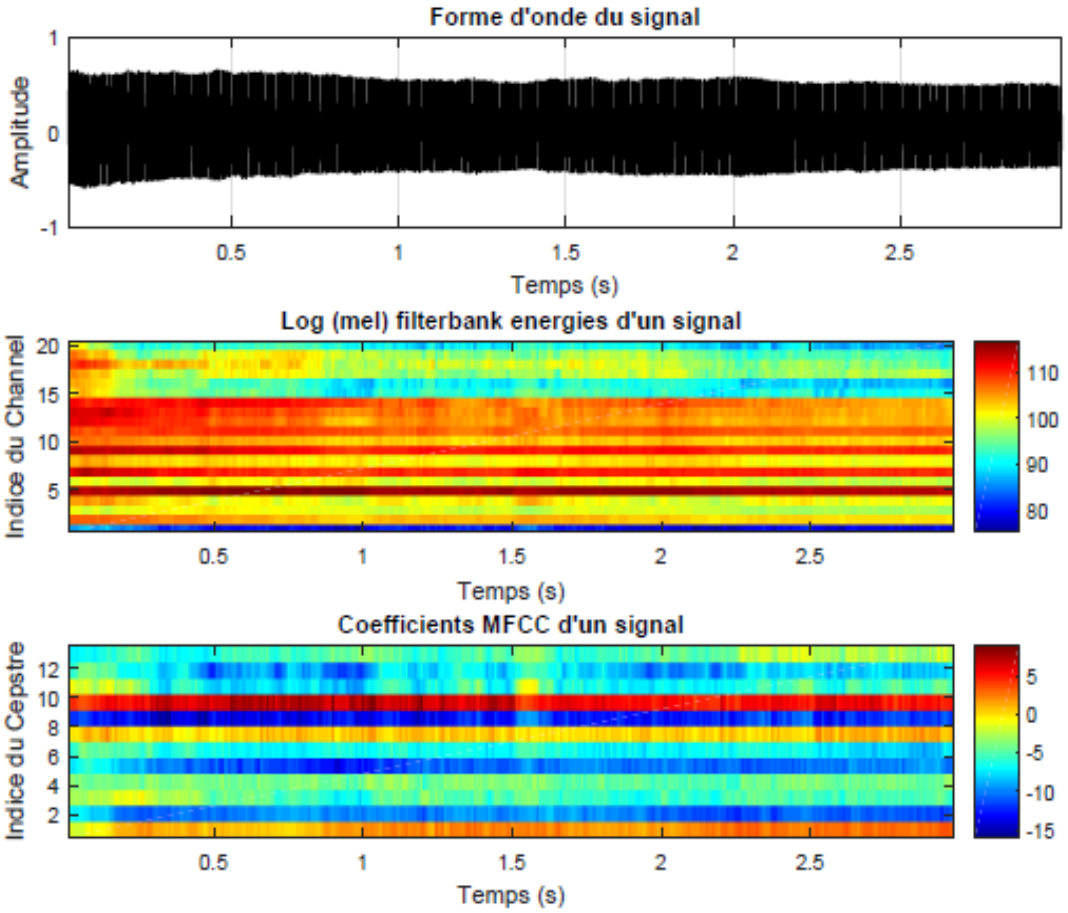

Figure II.4 : Forme d'onde, énergies Log (mel) et coefficients MFCC d'un signal.

II.3.3 Exemple d'extraction des paramètres de variation du pitch :

Pour l'extraction d'indices acoustiques des signaux vocaux nous avons utilisé le logiciel PRAAT. Nous pouvons donc réaliser de nombreuses mesures acoustiques les plus courantes en pathologie vocale, comme les perturbations de la fréquence (Jitter local, Jitter absolue, Jitter ppq3, …), les perturbations d'amplitude (Shimmer local, Shimmer (dB), Shimmer apq3, …) et les mesures de bruit (HNR, NHR).

Le help de PRAAT nous a permis la prise en main du logiciel et à titre d'illustration nous avons sélectionné six fichiers de parole (un normal et cinq pathologiques) dont nous présenterons ses différentes mesures dans le tableau (II.1).

Les seuils normal / pathologique selon le logiciel PRAAT sont : jitter local (1.04%), jitter rap (0.68%), shimmer local (3.81%), shimmer apq3 (3.070%). Nous remarquons que dans le cas d'une voix normale que les valeurs obtenues pour le jitter local (=0.218%), le jitter rap (=0.118%), le shimmer local (=1.018%) et le shimmer apq3 (=0.551%) sont toutes en dessous des seuils de pathologie, et une valeur élevée du HNR (=27.133 dB).

Paramètres	Jitter Local (%)	Jitter Rap (%)	Shimmer Local (%)	Shimmer APQ3 (%)	HNR (dB)
Voix Normale	0.218	0.118	1.018	0.551	27.133
Paralysis	9.168	4.094	18.773	9.442	-2.619
Vocal fold polyp	1.204	0.672	8.191	4.485	6.713
Vocal fold nodules	0.656	0.370	8.831	4.928	12.948
Keratosis	6.741	3.102	17.975	10.013	1.222
Adductor spasmodic dysphonia	8.935	4.661	22.296	8.310	-0.291

Tableau II.1 : Exemple de valeurs des indices acoustiques.

Nous remarquons aussi que pour les trois fichiers sélectionnés pour les trois types de pathologies : Paralysis, Keratosis, Adductor spasmodic dysphonia, prononcés par trois locuteurs différents, les valeurs obtenues sont au-dessus du seuil de pathologie et avec des valeurs un petit peu élevées, ce qui correspond à des types de pathologies avec sévérités élevées. Dans ce cas, les valeurs du HNR sont faibles et décroit avec la sévérité de la pathologie.

Dans le cas de la pathologie (vocal fold polyp), les mesures obtenues du fichier sélectionné sont supérieures aux seuils spécifiés sauf le jitter rap (=0.672% < seuil = 0.68%) et la valeur du HNR (=6.713 dB) qui est moyenne. Le degré de sévérité de ce type de pathologie est inférieur à celui des trois autres pathologies citées en haut.

La pathologie la moins sévère est celle du vocal fold nodules où nous observons seulement des valeurs élevées du shimmer local (8.831%) et le shimmer apq3 (4.928%), les autres valeurs du jitter sont inférieures aux seuils.

Nous sélectionnons seulement deux fichiers pour la représentation graphique (cas normal et un cas pathologique). Les mesures des différentes valeurs du pitch ont été obtenues par PRAAT avec la méthode de covariance et sauvegardées dans un fichier de données. Comme la représentation graphique de MATLAB est meilleure que celle de PRAAT, nous avons élaboré un script MATLAB pour la représentation de la forme d'onde, le spectrogramme et les variations des valeurs du pitch pour les deux signaux (figures II.5 et II.6). Nous remarquons des variations dans l'allure des valeurs du pitch du signal pathologique par rapport aux variations faibles du signal normal. De plus, les spectrogrammes montrent l'influence sur toutes les fréquences avec des degrés de contamination différents.

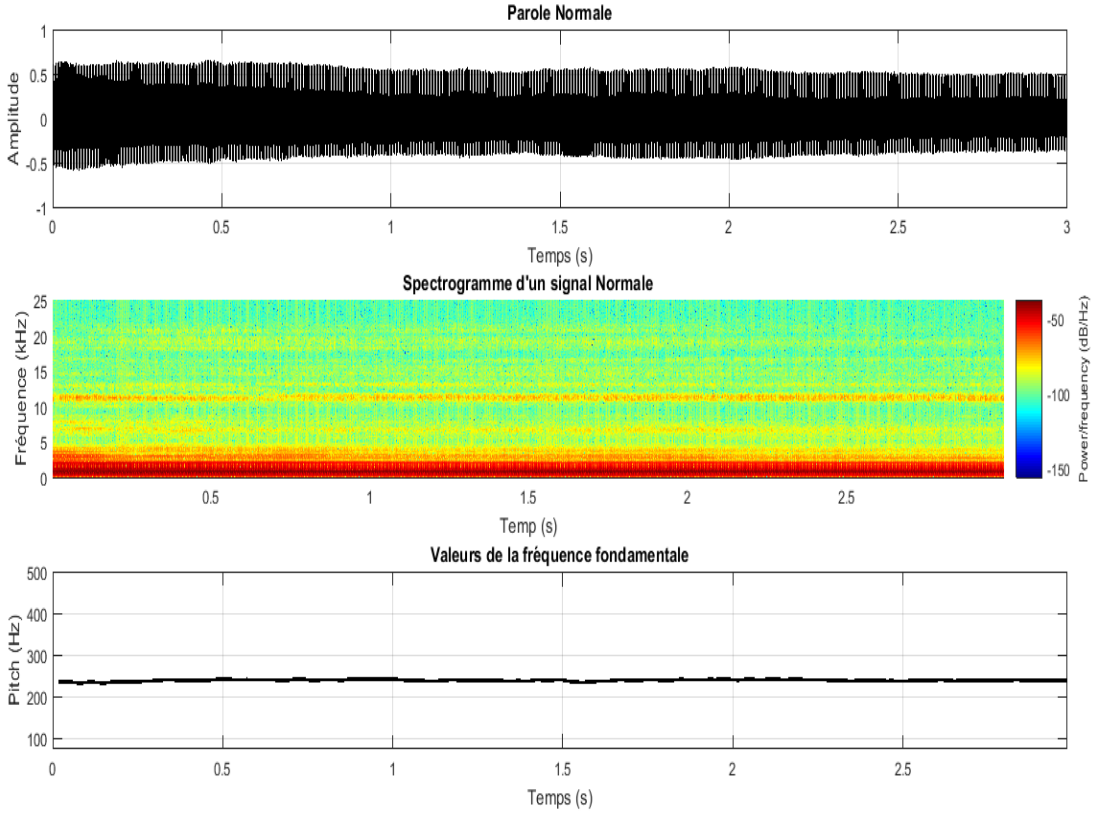

Figure II.5 : Forme d'onde, spectrogramme et valeurs du pitch d'un signal normal.

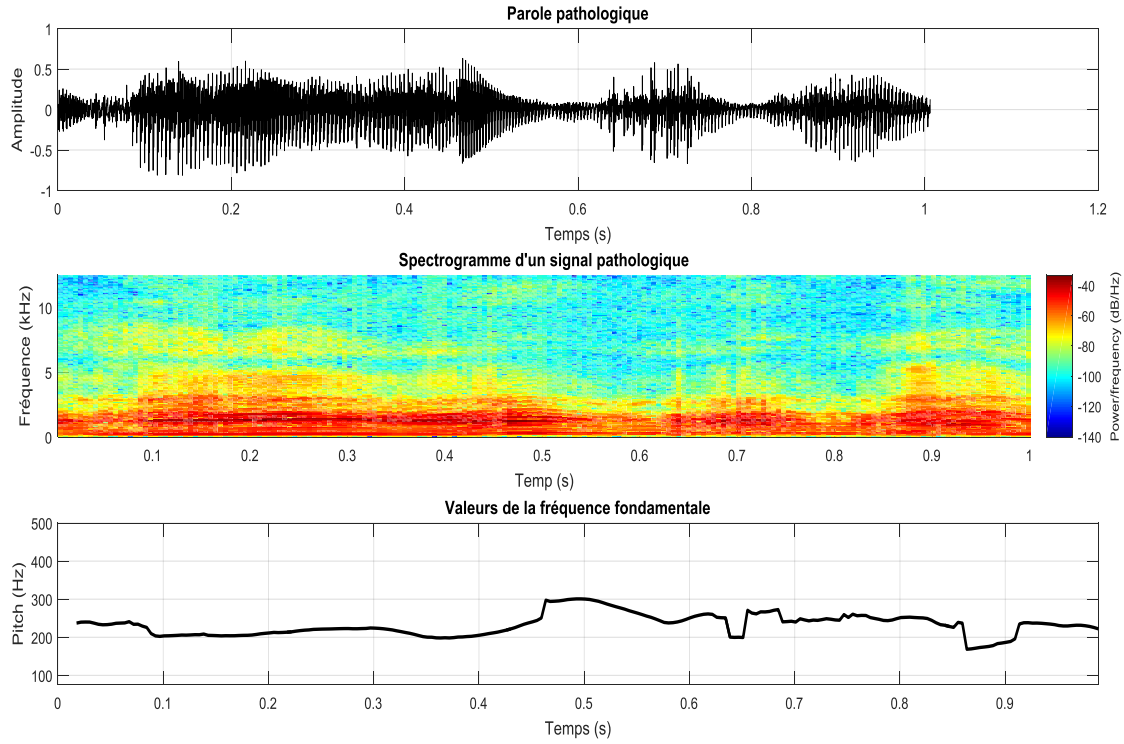

Figure II.6 : Forme d'onde, spectrogramme et valeurs du pitch d'un signal pathologique.

II.4 Conclusion :

L'évaluation de la qualité de la voix comprend une évaluation perceptive qui évalue le timbre de la voix sur des différentes échelles à l'oreille et une évaluation instrumentale par la mesure de paramètres acoustiques et aérodynamiques. Les méthodes acoustiques ont gagné en efficacité et en simplicité d'utilisation et d'interprétation. La mesure des différents paramètres vocaux par l'outil informatique a permis d'objectiver et visualiser un certain nombre de marqueurs pathologiques corrélés aux données perceptives et à l'auto-évaluation.

Comme dans le cas de la reconnaissance de la parole et l'identification du locuteur, les paramètres MFCC sont utilisés dans la majorité des travaux dans la littérature sur la classification des pathologies. Ces paramètres seront aussi les paramètres de base dans notre travail de classification, en ajoutant d'autres paramètres acoustiques très utiles dans l'analyse des signaux pathologiques à savoir le Jitter et ses variantes, le Shimmer et ses variantes et le HNR. Ces différents paramètres cités constitueront les paramètres issus du premier bloc d'un système de classification, nommé bloc d'extraction des paramètres. Le chapitre suivant sera consacré aux autres blocs de la classification.

III.1 Introduction :

Les nouvelles technologies dans l'apprentissage et les méthodes de classification ont aussi un rôle important sur l'évolution et le développement des méthodes et moyens de diagnostic des troubles vocaux, elles permettent une évaluation objective, rapide, facile et accessible de la voix. Parmi plusieurs techniques d'apprentissage automatique existantes dans la littérature, la classification par SVM (Support Vector Machines) a été largement utilisée dans le traitement du signal vocal. Avant de détailler le principe et les fondements mathématiques du SVM, des définitions de la classification d'ordre général et les types de classification seront présentés au début. Nous présenterons aussi une méthode de classification de référence, celle des K plus proches voisins afin de réaliser une étude comparative des performances avec le SVM.

III.2 Définition et type de classification :

Les méthodes de classification permettent de grouper des objets (observations ou individus) dans des classes (clusters) de manière à ce que les objets appartenant à la même classe sont plus similaires entre eux qu'aux objets appartenant aux autres classes. Il s'agit donc d'opérer des regroupements en classes homogènes d'un ensemble d'individus. On distingue deux branches de la classification automatique, la classification supervisée et non supervisée.

III.2.1 Classification non supervisée :

La classification non supervisée ou "Clustering" ou encore « Cluster analysis » est une méthode mathématique d'analyse de données. Pour faciliter l'étude d'une population d'effectif important (animaux, plantes, malades, gènes, ...), on regroupe les individus qui la forment en plusieurs classes de telle sorte que les individus d'une même classe soient les plus semblables possibles et que les classes soient les plus distinctes possibles les unes des autres.

Regrouper des éléments entre eux facilite mieux l'interprétation d'une grande quantité des données. Ainsi, les objectifs de la classification sont de regrouper les individus décrits par un ensemble de variables, ou regrouper les variables observées sur des individus et d'interpréter ces regroupements par une synthèse des résultats. L'intérêt de regrouper les individus est ici de les classer en conservant leur caractère multidimensionnel, et non pas seulement à partir d'une seule variable. Si les variables sont nombreuses il peut être intéressant de les regrouper afin de réduire leur nombre pour une interprétation plus facile [19].

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.

Il existe une très large famille de méthodes dédiées à la classification non supervisée. Parmi ces méthodes les plus utilisées, citons deux types d'approches : la classification hiérarchique et les centres mobiles (k-means).

III.2.2 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, 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 [20]. Parmi les méthodes supervisées on cite : les k-plus proches voisins, les arbres de décision, les réseaux de neurones, les machines à support de vecteurs (SVM) et les classificateur de Bayes.

Le schéma présenté sur la figure (III.1) illustre la structure générale d'une chaîne de classification statistique supervisée.

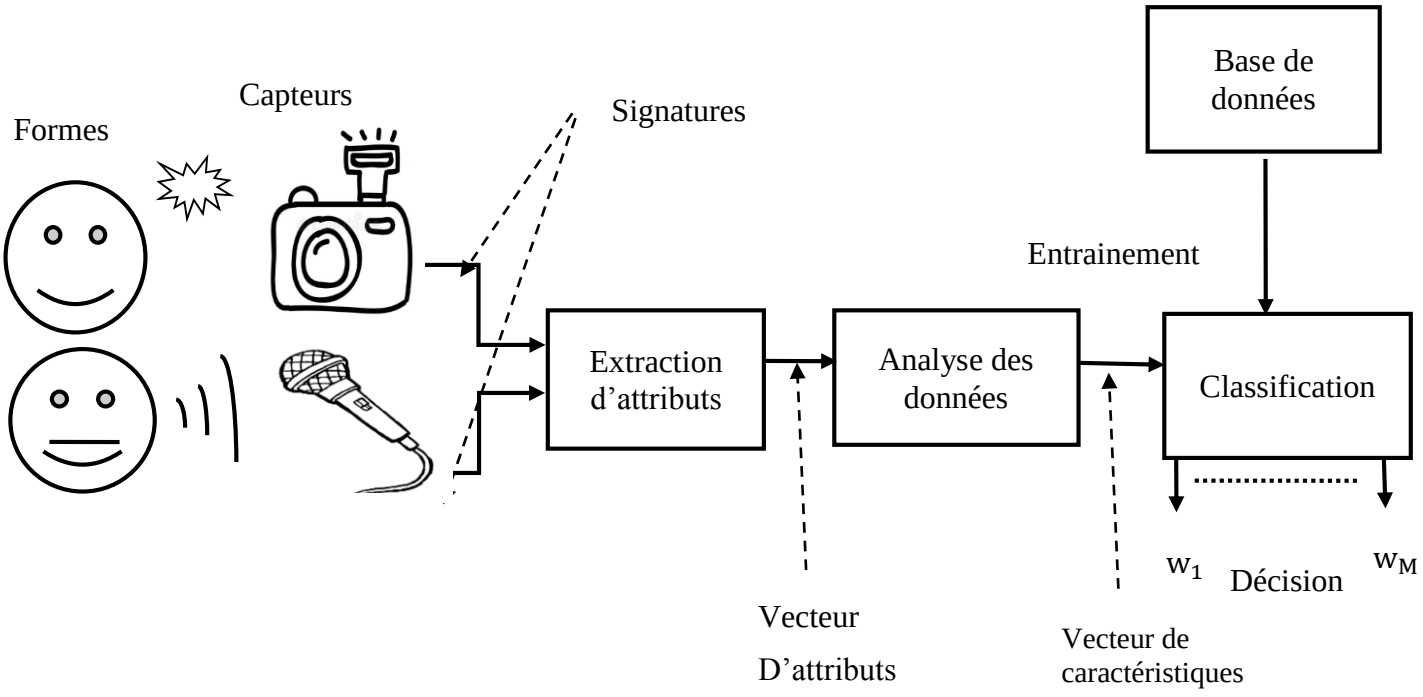

Figure III.1 : Schéma général d'une chaîne de classification statistique supervisée.

Ce processus, qui a pour but de classifier toute nouvelle forme inconnue à partir de sa signature, comprend les étapes suivantes [21] :

- L'extraction d'un vecteur d'attributs (ou primitives), à partir des signatures enregistrées, qui met en évidence l'information discriminante liée à la structure particulière de chaque classe.

- L'analyse des données qui sert généralement à réduire la dimension de l'espace initial tout en gardant les propriétés discriminantes des attributs extraits.
- La classification qui est l'étape décisionnelle et qui consiste à associer une étiquette à la forme dont la signature est présentée en entrée.

III.3 Classification par SVM :

Les machines à vecteurs de support ou séparateurs à vaste marge sont des techniques d'apprentissage supervisées destinées à résoudre des problèmes de classification binaire. Ils ont été introduits par V. Vapnik en 1995 [22] dans son livre, mais leur première apparition était en 1992 après était publiées par Boser, Guyon et Vapnik dans un article [23]. Le succès de cette méthode revient à la solidité des bases théoriques qui la soutiennent, elle permet d'élaborer des divers problèmes dont la classification par SVM est particulièrement adaptée à des problèmes de très haute dimension.

III.3.1 Définition des SVM :

Le SVM appartient à la catégorie des classificateurs linéaires (qui utilisent une séparation linéaire des données), et qui dispose de sa méthode à lui pour trouver la frontière entre les catégories.

Pour que le SVM puisse trouver cette frontière, il est nécessaire de lui donner des données d'entraînement. En l'occurrence, on donne au SVM un ensemble de points, dont on sait déjà la nature ou à quelle catégorie ou classe revient ces points. A partir de ces données, le SVM va estimer l'emplacement le plus proche de la frontière : c'est la période d'entraînement, nécessaire à tout algorithme d'apprentissage automatique [24].

Une fois la phase d'entraînement terminée, le SVM a ainsi trouvé, à partir de données d'entraînement, l'emplacement supposé de la frontière. En quelque sorte, il a « appris » l'emplacement de la frontière grâce aux données d'entraînement. Le SVM est maintenant capable de prédire à quelle catégorie appartient une entrée qu'il n'avait jamais vue avant et sans intervention, c'est là tout l'intérêt de l'apprentissage automatique.

III.3.2 Principe de base de la méthode de classification SVM :

Le but principal de SVM est de bien choisir la frontière entre les deux catégories, car quand on a un ensemble de points il existe une infinité de lignes qui peuvent discriminer ces deux classes de points, alors le problème est laquelle de ces lignes droites on doit choisir (figure III.2).

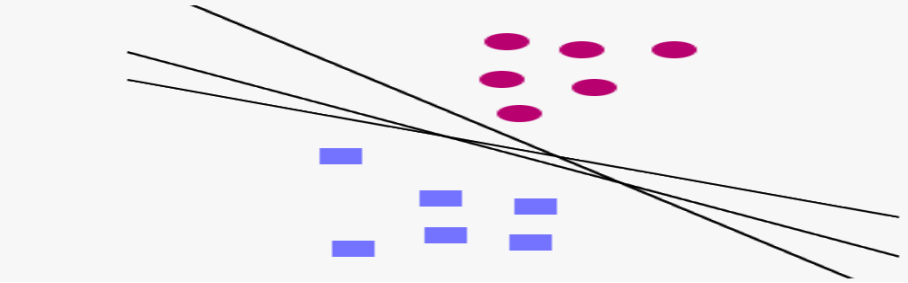

Figure III.2 : Les différentes frontières possibles.

Intuitivement, on se dit que si on nous donne un nouveau point, très proche de la première catégorie, alors ce point a de fortes chances d'appartenir à cette catégorie lui aussi. Inversement, plus un point est près de la deuxième classe, plus il a de chances d'appartenir lui-même à cette classe. Pour cette raison, un SVM va placer la frontière aussi loin que possible des deux classes de point [25].

Comme on le voit dans la figure, c'est bien la frontière la plus éloignée de tous les points d'entraînement qui est optimale, on dit qu'elle a la meilleure capacité de généralisation. Ainsi, le but d'un SVM est de trouver cette frontière optimale, en maximisant la distance entre les points d'entraînement et la frontière, pour objectif de procurer plus de sécurité lorsque 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 qui suit, 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 classée [26], comme l'illustre la figure suivante :

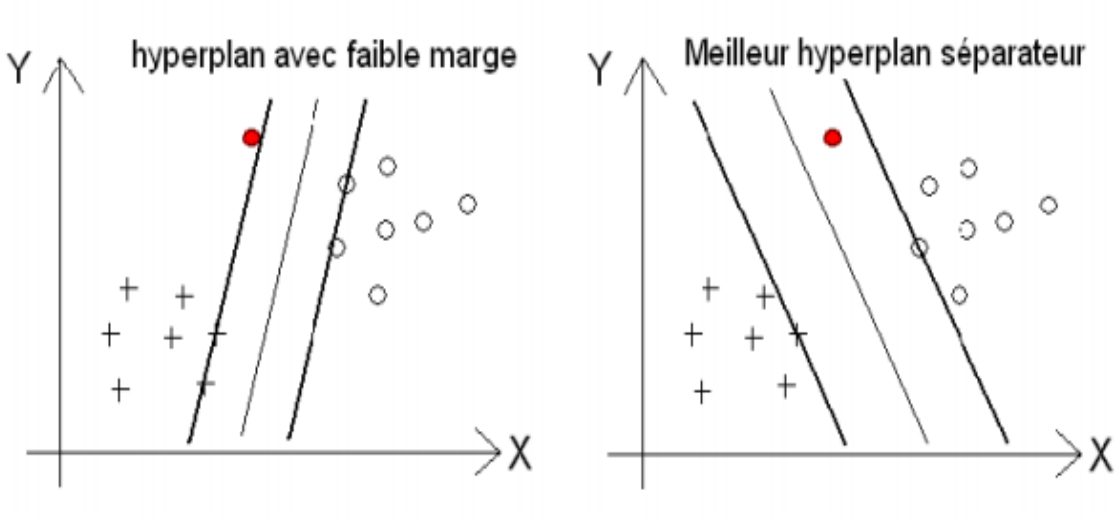

Figure III.3 : Différentes frontières entre les vecteurs de support.

Les points d'entraînement les plus proches de la frontière sont appelés vecteurs supports, et c'est d'eux que les SVM tirent leur nom : SVM signifie Support Vector Machine, ou Machines à Vecteurs de Support en français. Support, parce que ce sont ces points qui « supportent » la frontière. Les vecteurs de support sont indiqués sur la figure (III.4).

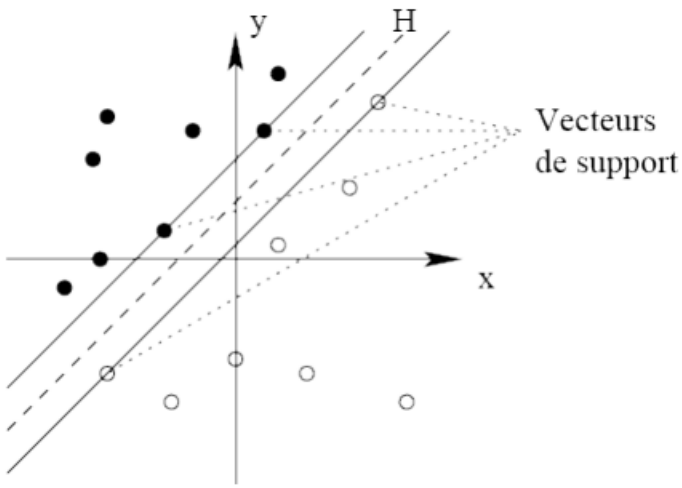

Figure III.4 : Hyperplan optimal (H) et illustration des vecteurs de support [26].

La frontière ou la ligne droite (H) qui sépare les deux catégories représente l'hyperplan séparateur. De façon plus générale que dans les exemples donnés précédemment, les SVM ne se bornent pas à séparer des points dans le plan. Ils peuvent en fait séparer des points dans un espace de dimension quelconque.

Par exemple, si on cherche à classer une voix alors qu'on connaît ses paramètres, le nombre de paramètres sur lequel on travaille représente les dimensions de l'espace. Un autre exemple est celui de la reconnaissance d'image : une image contient des millions de pixels. Il est ainsi courant de travailler dans des espaces de très grandes dimensions.

Fondamentalement, un SVM cherchera simplement à trouver un hyperplan qui sépare les deux catégories de notre problème.

III.3.3 Formalisme d'un SVM :

Dans un espace vectoriel de dimension finie *N*, un hyperplan est un sous-espace vectoriel de dimension $N-1$. Ainsi, dans un espace de dimension 2 un hyperplan sera une droite, dans un espace de dimension 3 un hyperplan sera un plan, etc.

Soit un espace vectoriel *E* de dimension *N*. L'équation caractéristique d'un hyperplan est de la forme :

$$w_1 x_1 + w_2 x_2 + \cdots + w_N = 0 \qquad \text{(III.1)}$$

où $w_1, \ldots, w_N$ sont des scalaires.

Par définition, tout vecteur $x = (x_1, \dots, x_N) \in E$, vérifiant l'équation (III.1) appartient à l'hyperplan. De plus, un hyperplan sépare complètement l'espace vectoriel en deux parties distinctes. Ainsi, une ligne droite sépare le plan en deux régions, il est donc possible de diviser notre espace vectoriel en deux catégories distinctes. Comme nous pouvons le constater, un hyperplan vectoriel passe toujours par 0. C'est pour cette raison qu'on utilisera un hyperplan affine, qui n'a pas quant à lui l'obligation de passer par l'origine.

Ainsi, si l'on se place dans R^N, pendant son entraînement le SVM calculera un hyperplan vectoriel d'équation (III.1) et un scalaire (un nombre réel) *b*. C'est ce scalaire *b* qui va nous permettre de travailler avec un hyperplan affine, comme nous allons le voir.

Le vecteur $w = (w_1, \dots, w_N)$ est appelé vecteur de poids, le scalaire *b* est appelé biais. Une fois l'entraînement terminé, pour classer une nouvelle entrée $x = (a_1, \dots, a_N)$, le SVM regardera le signe de :

$$h(x) = w_1 a_1 + \cdots + w_N a_N + b = \sum_{i=1}^{N} w_i \cdot a_i + b = w^T \cdot x + b \qquad \text{(III.2)}$$

Si $h(x)$ est positif ou nul, alors *x* est d'un côté de l'hyperplan affine et appartient à la première catégorie, sinon *x* est de l'autre côté de l'hyperplan, et donc appartient à la seconde catégorie.
En résumé, on souhaite savoir, pour un point *x*, s'il se trouve d'un côté ou de l'autre de l'hyperplan. La fonction *h* nous permet de répondre à cette question, grâce à la classification suivante :

$$h(x) \geq 0 \Longrightarrow x \in catégorie\ 1 \qquad \text{(III.3)}$$

$$h(x) < 0 \Longrightarrow x \in catégorie\ 2 \qquad \text{(III.4)}$$

Ainsi, étant donné un hyperplan de vecteur de poids *w*, et de biais *b*, nous pouvons calculer si un point x_k appartient à telle ou telle catégorie, grâce au signe de $h(x_k)$.
En particulier, supposons que l'on assigne à tout point x_k un label l_k qui vaut 1 si x_k appartient à la première catégorie, et -1 si x_k appartient à la seconde catégorie. Alors, si le SVM est correctement entraîné, on a toujours :

$$l_k . h(x_k) \geq 0 \qquad \text{(III.5)}$$

C'est-à-dire :

$$l_k (w^T \cdot x_k + b) \geq 0 \qquad \text{(III.6)}$$

Le but d'un SVM, lors de l'entraînement, est donc de trouver un vecteur de poids *w* et un biais *b* tels que, pour tout x_k de label l_k appartenant aux données d'entraînement, l'équation (III.6) est satisfaite. Autrement dit, de trouver un hyperplan séparateur entre les deux catégories.

III.3.4 Calcul de la marge :

Si l'on prend un point $x_k \in R^N$, on peut prouver que sa distance à l'hyperplan de vecteur support *w* et de biais *b* est donnée par :

$$\frac{l_k(w^T \cdot x_k + b)}{\|w\|} \tag{III.7}$$

Où $\|w\|$ désigne la norme euclidienne de *w*. La marge d'un hyperplan de paramètres (w, b) par rapport à un ensemble de points (x_k) est donc :

$$min \frac{l_k(w^T \cdot x_k + b)}{\|w\|} \tag{III.8}$$

On veut trouver l'hyperplan de support *w* et de biais *b* qui permettent de maximiser cette marge, donc on cherche un unique hyperplan dont les paramètres (w, b) sont données par la formule suivante :

$$\arg \max_{w,b} \min_k \left\{ \frac{l_k(w^T \cdot x_k + b)}{(\|w\|)} \right\} \tag{III.9}$$

Même si on peut montrer que l'hyperplan optimal est unique, il existe plusieurs couples (w, b) qui décrivent ce même hyperplan. On décide de ne considérer que l'unique paramétrage (w, b) tel que les vecteurs support x_s vérifient :

$$(w^T \cdot x_s + b) = 1 \tag{III.10}$$

Par conséquent $\forall k, l_k(w^\top \cdot x_k + b) \geq 1$, et l'égalité est atteinte si x_k est un vecteur support. Autrement dit, cette normalisation sur *w* et *b* permet de garantir que la marge :

$$min \frac{l_k(w^T \cdot x_k + b)}{\|w\|} = \frac{1}{||w||} \tag{III.11}$$

On se retrouve donc avec le problème suivant :

$$\begin{cases} maximiser \ \frac{1}{||w||} \\ sous\ les\ contraintes \ \ \forall k, \ l_k(w^T \cdot x_k + b) \geq 1 \end{cases} \tag{III.12}$$

Que l'on peut reformuler de la façon suivante :

$$\begin{cases} minimiser \ ||w|| \\ sous\ les\ contraintes \ \ \forall k, \ l_k(w^\top \cdot x_k + b) \geq 1 \end{cases} \tag{III.13}$$

Que, pour des raisons pratiques, on reformule à nouveau :

$$\begin{cases} minimiser \ \frac{(||w||)^2}{2} \\ sous\ les\ contraintes \ \ \forall k, \ l_k(w^T \cdot x_k + b) \geq 1 \end{cases} \tag{III.14}$$

Ce genre de problème est appelé problème d'optimisation quadratique, et il existe de nombreuses méthodes pour le résoudre. Dans le cas présent, on utilise la méthode des multiplicateurs de Lagrange [25].

Méthode des multiplicateurs de Lagrange :

Lagrange est une méthode mathématique utilisée pour résoudre des problèmes d'optimisation quadratique sous contraintes, ça signifie que lorsque nous voulons optimiser (trouver les min ou max extrêmes) une fonction de base $f\ (x_1,\ \dots\ , x_N)\ :\ R^N\ \rightarrow\ R$, vous pouviez simplement utiliser le test de la dérivée seconde, en plus de cette fonction, vous avez également une contrainte $g\ (x_1,\ \dots\ ,\ x_N) = 0$. Donc, nous essayons d'optimiser f tout en contraignant f avec g. Vous pouvez considérer une contrainte comme une frontière.
Par exemple, disant qu'on est dans une pièce et nous voulons connaitre la distance la plus élevée pour laquelle nous pouvons lancer une balle, alors nous somme conditionné par le plafond, nous ne pouvons pas lancer la balle plus haut que le plafond. Alors, la contrainte est le plafond. Le cœur de Lagrange est l'équation suivante :

$$\nabla f(x) = \lambda\, \nabla g(x) \tag{III.15}$$

Cela dit que le gradient de f est égal à un scalaire multiplier par le gradient de g, rappeler que :

$$g(x) = 0 \tag{III.16}$$

Souvent et surtout dans le contexte SVM, les deux équations (II.14) et (II.15) sont combinées en une seule équation appelée le Lagrangien :

$$L(x, \lambda) = f(x) - \lambda\, g(x) \tag{III.17}$$

On utilise cette équation et on cherche le point où :

$$\nabla L\ (x, \lambda) = 0 \tag{III.18}$$

Lorsque on a plusieurs contraintes, la relation (III.17) devient :

$$L(x, \lambda) = f(x) - \sum_i \lambda_i g_i(x) \tag{III.19}$$

Application de Lagrange en SVM :

Il s'agit de faire rentrer les contraintes dans la fonction objective et de pondérer chacune d'entre elles par une variable duale

$$L(w, b, \lambda) = \frac{1}{2}\left|\left|w\right|\right|^2 - \sum_{i=1}^{N} \lambda_i \{l_i(w^T x_i + b) - 1\} \tag{III.20}$$

Les variables duales intervenant dans le Lagrangien sont appelées multiplicateurs de Lagrange. Notons que L doit être minimisé par rapport aux variables primales w et b et maximisé par rapport aux variables duales λ_i. Le point (minimal par rapport à une variable, maximal par rapport à l'autre) doit donc satisfaire les conditions nécessaires de stationnarité qui correspondent aux conditions Karush Kuhn et Tucker (KKT), nous trouvons [24] :

$$\frac{\partial L(w,b,\lambda)}{\partial w} = 0 \quad \text{(III.21)}$$

$$\frac{\partial L(w,b,\lambda)}{\partial b} = 0 \quad \text{(III.22)}$$

Ce qui nous permet d'obtenir :

$$w = \sum_{i=1}^{N} \lambda_i l_i x_i \quad \text{(III.23)}$$

$$\sum_{i=1}^{N} \lambda_i l_i = 0 \quad \text{(III.24)}$$

Cette résolution donnera une valeur optimale pour *w*, mais rien pour *b*. Pour retrouver *b*, il suffit de se rappeler que pour les vecteurs support $l_s(w^T \cdot x_s + b) = 1$. On en déduit donc que *b* est tel que :

$$\min l_k \, (w^T \cdot x_k + b) = 1 \quad \text{(III.25)}$$

III.3.5 Cas de SVM non linéaire :

Dans la plupart des problèmes réels, les données ne sont pas linéairement séparables, il est donc nécessaire de contourner ce problème (difficile de séparer n'importe quel jeu de données par un simple hyperplan). Si par exemple les données des deux classes se chevauchent sévèrement, aucun hyperplan séparateur ne sera satisfaisant, il est nécessaire de projeter les points d'apprentissage dans un espace de dimension plus élevée.

On applique aux vecteurs d'entrée x une fonction de transformation non-linéaire φ , qu'on appelle fonction noyau [26].
On définit l'opération de redescription des points x de E vers E′ par l'opération :

$$\varphi : E \rightarrow E' \quad \text{(III.26)}$$

$$x \rightarrow \varphi(x) \quad \text{(III.27)}$$

Dans ce nouvel espace E', on va tenter d'entraîner le SVM, comme nous l'aurions fait dans l'espace E. Si les données y sont linéairement séparables, c'est gagné ! Par la suite, si l'on veut classer x, il suffira de classer $\varphi(x)$: on obtient ainsi un SVM fonctionnel [25].

L'astuce du noyau pour simplifier les calculs :

Quand on pose le problème d'optimisation quadratique dans l'espace E', on s'aperçoit que les seules apparitions de φ sont de la forme $\varphi(x_i)^T \cdot \varphi(x_j)$. De même dans l'expression de la fonction de classification h'. Par conséquent, il n'y a pas besoin de connaître expressément E', ni même φ : il suffit de connaître toutes les valeurs $\varphi(x_i)^T \cdot \varphi(x_j)$, qui ne dépendent donc que des x_i.
On appelle donc fonction noyau, la fonction $K: E \times E \rightarrow R$ définie de la façon suivante :

$$K(x, x') = \varphi(x)^T \cdot \varphi(x') \quad \text{(III.28)}$$

A ce moment, le calcul de l'hyperplan séparateur dans E' ne nécessite ni la connaissance de E', ni de φ, mais seulement de K.

Noyau symétrique semi-défini positif :

Une fonction φ est dite symétrique si et seulement si :

$$\forall x, y \quad \varphi(x, y) = \varphi(y, x) \quad \text{(III.29)}$$

(Dans le cas où φ est à deux variables).

Une fonction symétrique est dite semi-définie positive si et seulement si, pour tous ensemble fini $(x_1, \dots, x_n)$, et pour tous $réels(c_1, \dots, c_n)$:

$$\sum c_i c_j K(x_i, x_j) \geq 0 \quad \text{(III.30)}$$

Ainsi, il est possible d'utiliser n'importe quelle fonction noyau afin de réaliser une redescription dans un espace de dimension supérieure. La fonction noyau étant définie sur l'espace de description *E* (et non sur l'espace de redescription *E'*, de plus grande dimension), les calculs sont beaucoup plus rapides.

Voici une liste non exhaustive de noyaux couramment utilisés [25].

- Le noyau polynomial,

$$K(x, x') = (\alpha x^T \cdot x' + \lambda)^d \quad \text{(III.31)}$$

- Le noyau Gaussien,

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad \text{(III.32)}$$

- Le noyau Laplacien,

$$K(x, x') = \exp\left(-\frac{\|x-x'\|}{\sigma}\right) \quad \text{(III.33)}$$

Le noyau rationnel,

$$K(x, x') = 1 - \frac{\|x-x'\|^2}{\|x-x'\|^2+\sigma} \quad \text{(III.34)}$$

III.4 SVM pour le cas multi-classes :

La plupart des problèmes ne se contentent pas de deux classes de données. Il existe plusieurs méthodes pour faire la classification multi-classes. Citons les plus utilisées :

La première méthode, celle que nous utilisons dans notre application, est une méthode dite Un-contre-Un (One-Versus-One). Au lieu d'apprendre N fonctions de décisions, ici chaque classe est discriminée d'une autre (figure III.5). En créant des « voteurs » : chaque voteur V_{ij} détermine si mon entrée x a plus de chances d'appartenir à la catégorie i ou à la catégorie j. Ainsi, un voteur V_{ij} est un SVM qui s'entraîne sur les données de catégorie i et j uniquement. L'inconvénient de cette méthode est que le nombre de voteurs est proportionnel au carré du nombre de catégories, d'où un temps de calcul de plus en plus élevé.

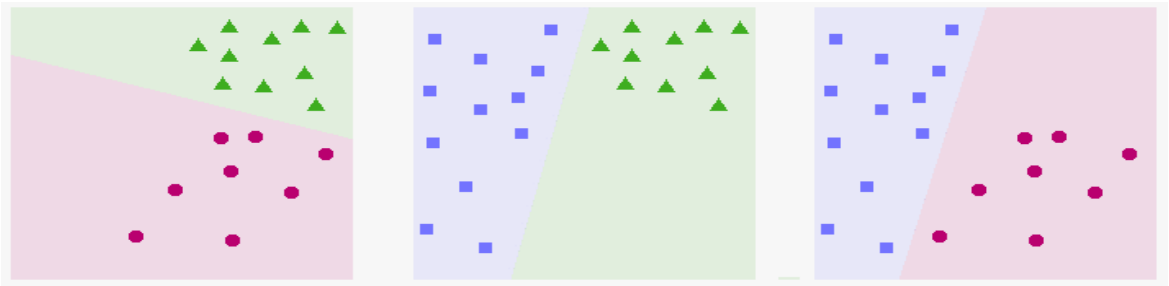

Figure III.5 : Discrimination par un contre un.

Pour classer une entrée on retournera tout simplement la catégorie qui aura remporté le plus de duels.

La deuxième méthode est appelée Un-Contre-Tous (One-Versus-All). C'est une approche étendant la notion de marge aux cas multi-classes. Cette formulation intéressante permet de poser un problème d'optimisation unique. Le problème fait intervenir N fonctions de décision. En créant deux catégories : la catégorie 1, qui contient toutes les entrées d'une classe précise (spécialisés dans cette catégorie), et la catégorie 2, qui contient toutes les autres entrées des classes qui restent (figure III.6). Et on fait de même pour les SVM spécialisés dans les autres catégories.

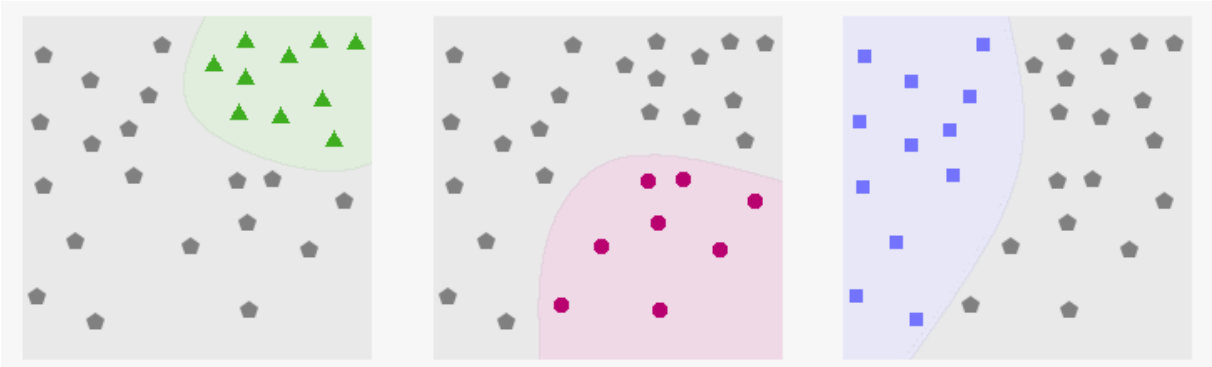

Figure III.6 : Discrimination par un contre tous.

Pour classer une nouvelle entrée, on regarde à quelle catégorie la nouvelle entrée est le plus probable d'appartenir [24,25].

Remarque :

Il arrive que plusieurs SVM aient un résultat positif. Dans ce cas-là, on prend celui qui est le plus certain de son résultat. De même, quand tous les résultats sont négatifs, on prend alors la catégorie du SVM pour lequel l'entrée est le plus près possible de la frontière [25].

III.5 Classifieur K plus proches voisins (KPPV) :

La méthode de référence des méthodes non-paramétriques de classification est celle des K plus proches voisins, connue sous le nom K-NN (K-Nearest Neighbors). C'est une méthode d'apprentissage supervisée qui raisonne avec le principe sous-jacent : "dis-moi qui sont tes amis, je te dirais qui tu es".

L'algorithme KNN est l'un des plus simples de tous les algorithmes d'apprentissage automatique, il est basé sur l'apprentissage paresseux (lazy learning). En d'autres termes, il n'y a pas de phase d'entraînement explicite ou très minime. Cela signifie que la phase d'entraînement est assez rapide [27].

La méthode KNN suppose que les données se trouvent dans un espace de caractéristiques. Cela signifie que les points de données sont dans un espace métrique. Les données peuvent être des scalaires ou même des vecteurs multidimensionnels.

Cette méthode est utilisée pour la classification et la régression. Dans les deux cas, l'entrée se compose des k données d'entraînement les plus proches dans l'espace de caractéristiques.

Pour trouver la classe d'un nouveau cas, Cet algorithme se base sur le principe suivant : il cherche les k plus proches voisins de ce nouveau cas, ensuite, il choisit parmi les candidats trouvés le résultat le plus proche et le plus fréquent.

Pour affecter un nouvel individu à une classe, l'algorithme cherche les k plus proches voisins parmi les individus déjà classés. Ainsi, l'individu est affecté à la classe qui contient le plus d'individus parmi les candidats trouvés.

Principe de l'algorithme :

Soit un ensemble E contenant n données labellisées : $E = \{(y_i, \vec{x}_i)\}$ avec i compris entre 1 et n, où y_i correspond à la classe (le label) de la donnée *i* et où le vecteur $\vec{x}_i$ de dimension *p* représente les variables prédictrices de la donnée i.

$$\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \quad \text{(III.35)}$$

Soit une donnée *u* qui n'appartient pas à *E* et qui ne possède pas de label (*u* est uniquement caractérisée par un vecteur $\vec{x}_u$ de dimension p).
Soit une fonction d qui renvoie la distance entre la donnée u et une donnée quelconque appartenant à *E*, et *K* un entier inférieur ou égal à n [28]. La figure (III.7) représente le fonctionnement de l'algorithme 'KNN'.

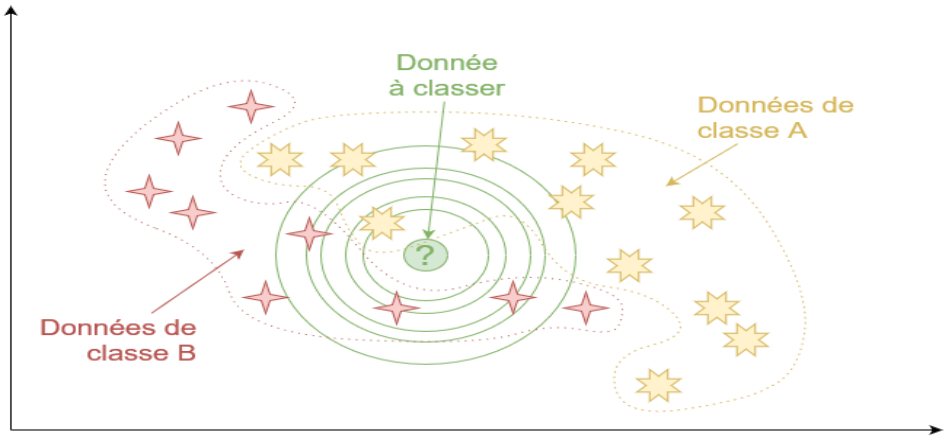

Figure III.7 : Fonctionnement de l'algorithme 'KNN'.

Pour appliquer cette méthode, les étapes à suivre sont les suivantes :

- On calcule les distances entre la donnée u et chaque donnée appartenant à E à l'aide de la fonction d.
- On retient les k données du jeu de données E les plus proches de u.
- On attribue à u la classe qui est la plus fréquente parmi les k données les plus proches.

Il est possible d'utiliser différents types de distance parmi lesquels on trouve :

La distance Euclidienne : $d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ (III.36)

La distance de Minkowsky : $d(x,y) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p}$ (III.37)

La distance de Manhattan $d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ (III.38)

Où : x, y sont des vecteurs et p un paramètre.

Avantages :

1. L'algorithme KNN est robuste envers des données bruitées.
2. La méthode des k plus proches voisins est efficace si les données sont larges et incomplètes.
3. Cette méthode est l'une des plus simples de tous les algorithmes d'apprentissage automatique.

Inconvénients :

1.Le besoin de déterminer la valeur du nombre des plus proches voisins (le paramètre k).
2.Le temps de prédiction est très long puisqu'on doit calculer la distance de tous les exemples.

III.6 Conclusion :

Nous avons présenté dans ce chapitre les notions essentielles sur la classification en général et le SVM en particulier. 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 diagnostiques médicales et ce même sur des ensembles de données de très grandes dimensions. Nous allons vérifier l'efficacité de ces méthodes dans le cas de la détection et la classification des pathologies par nos implémentations des différentes variantes, les résultats et les discussions seront l'objet du chapitre quatre.

IV.1 Introduction :

Les systèmes de détection automatique de la pathologie de la voix et les systèmes de classification étudiés et présentés dans les chapitres précédents ont été implémenté et testé sur des fichiers de parole et des bases de données. Les conditions des différentes implémentations, les mesures de performances, les résultats et des discussions seront l'objet de ce chapitre.

IV.2 Bases des données utilisées en classification des voix pathologiques :

IV.2.1 Base de données (SVD) :

Une des bases de données citée dans les travaux de recherches sur la classification des voix pathologiques est la SVD (Saarbrücken Voice Database), publiée en ligne par l'institut de phonétique de l'université de Saarland, gratuitement téléchargeable sur internet et contient différents enregistrements de la voix de plusieurs patients (2041 fichiers), ces enregistrements sont pris dans des conditions précises, dans un cas normal ou un cas où le patient soufre d'une pathologie. Elle contient 1354 voix pathologiques prononcées par 627 locuteurs et 727 locutrices, souffrant de 71 différentes pathologies. Les 687 fichiers restants sont ceux prononcés par des patients sains (259 locuteurs et 428 locutrices) [3].

Cette base contient des fichiers audio de durée allant de 1 à 3 seconds, correspondant aux voyelles / a /, / i / et / u / prononcées avec des intonations normales, basses, hautes et basses-hautes-basses, et à des fichiers correspondant à une phrase en Allemand "Guten Morgen, wie geht es Ihnen ?" signifie ("Bonjour Comment allez-vous ?"). Tous les enregistrements sont échantillonnés à 50 kHz et leur résolution est de 16 bits.

IV.2.2 Base de données (MEEI) :

Massachusetts Eye and Ear Infirmary (MEEI), est une base de données commercialisée et disponible sur le marché. Elle est utilisée depuis des année, distribuée par « Kay Elemtrics », pour la comparaison des voix pathologiques et saines ainsi que la classification des pathologies de la parole. Elle est constituée d'un catalogue d'enregistrement des voix de 12s (environ 1400 voix) qui sont de deux types : une voyelle tenue /ah/ et une phrase bien spécifique « Le passage de l'arc-en-ciel ».

Ces enregistrements ont été pris avec un microphone à condensateur dans un environnement contrôlé, par 700 locuteurs au total 53 sains et 657 locuteurs pathologiques avec différentes pathologies ; ils sont échantillonnés à Fs = 50 kHz ou Fs = 25 kHz. Toutes les informations cliniques des personnes sont archivées dans la base de données et ont été réalisées au sein de MEEI Voice and Speech Laboratory ; ces informations d'identification du patient sont utilisées pour le diagnostic [29].

IV.2.3 Base de données utilisée :

Dans notre projet, nous avons sélectionnés seulement un sous-ensemble de 156 fichiers de sons de la base de données MEEI. Ces enregistrements se repartirent en six groupes, un

correspond à des voix des personnes sains (voix normales) et les autres cinq groupes correspondent à des voix des personnes qui souffrent de différentes maladies mentionnées dans le tableau ci-dessous. Chaque groupe contient un nombre précis des enregistrements des différentes personnes qui souffrent de la même maladie. Tous ces fichiers vocaux sont échantillonnés à la même fréquence d'échantillonnage (25 kHz), les signaux aves des fréquences d'échantillonnages de 50 kHz et de 10 kHz seront ré-échantillonnés à la fréquence de 25 kHz. De plus, la quantification utilise 16 bits.

Sujets	**Type de pathologie**	**Nombre de fichiers**	**Total**
Pathologique	Paralysis	20	103
	Vocal fold polyp	20	
	Vocal fold nodules	19	
	Keratosis	26	
	Adductor spasmodic dysphonia	18	
Normal	-	53	53

Tableau IV.1 : Distribution des fichiers normaux et pathologiques utilisés.

IV.2.4 Bases d'apprentissage et de test :

On travaille en général lors de la mise en œuvre d'un algorithme de classification sur des différents ensembles de données, pour arriver à créer un modèle mathématique, à partir des données collectées. La partie des données sur laquelle on élabore un modèle et les équations associées à ce dernier est appelée base de données d'apprentissage. Alors que la partie utilisée pour quantifier les performances de ce modèle et appelée base de test.

Plus généralement, lors de l'évolution de tout algorithme d'exploitation de données on divise les données aléatoirement en deux parties, la plus grande partie est pour l'apprentissage et la plus petite est pour le test, d'une façon pour que les deux bases soient d'une part représentative de l'ensemble de données globales.

Autrement dit, les caractéristiques d'ensemble de données d'apprentissage et de test seront similaires, pour minimiser les effets d'écart de donnée et mieux comprendre les caractéristiques du modèle. Et d'autre part que l'ensemble de données soit suffisamment volumineux pour produire des résultats statistiquement significatifs.

IV.2.5 Validation croisée :

La validation croisée désigne le processus qui permet de tester la précision prédictive d'un modèle dans un échantillon test (parfois aussi appelé échantillon de validation croisée) par rapport à la précision prédictive de l'échantillon d'apprentissage à partir duquel le modèle a été développé.

Diverses techniques ont été développées pour réaliser des validations croisées avec de petits échantillons en construisant des échantillons test et des échantillons d'apprentissage partiellement (mais pas complètement) indépendants [30-31].

La première méthode est très simple, il suffit de diviser l'échantillon de taille n en échantillon d'apprentissage (> 60 % de l'échantillon) et échantillon de test. Le modèle est bâti sur l'échantillon d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant l'erreur quadratique moyenne.

Dans la seconde, on divise k fois l'échantillon, puis on sélectionne un des k échantillons comme ensemble de validation et les (k-1) autres échantillons constitueront l'ensemble d'apprentissage. On calcule comme dans la première méthode l'erreur quadratique moyenne. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les (k-1) échantillons qui n'ont pas encore été utilisé pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. La moyenne des k erreurs quadratiques moyennes est enfin calculée pour estimer l'erreur de prédiction.

La troisième méthode est un cas particulier de la deuxième méthode où k=n, c'est-à-dire que l'on apprend sur (n-1) observations puis on valide le modèle sur la nième observation et l'on répète cette opération n fois.

La figure ci-dessous présente les différentes étapes de la méthode de la validation croisée la plus courante :

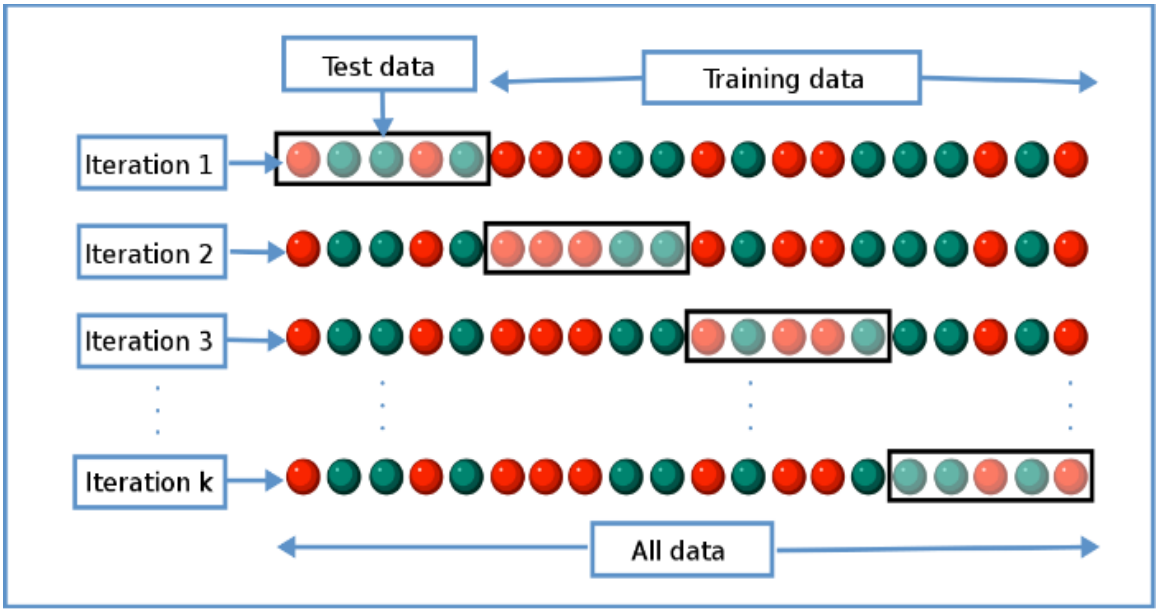

Figure IV.1 : Etapes de la méthode de la validation croisée.

IV.3 Mesures des performances :

Évaluer les performances d'un système de classification est un enjeu de grande importance car ces performances peuvent être utilisées pour l'apprentissage en tant que tel ou pour optimiser les valeurs des hyper-paramètres du classifieur.
Afin de mesurer les performances d'un détecteur des pathologies de la voix ou celles de la classification du type de pathologie, plusieurs indices doivent être pris en considération tels que la sensibilité, la spécificité, la précision et l'efficacité. Ces mesures sont utilisées dans le cas de notre travail et dans le cas de la classification en général.

IV.3.1 Matrice de confusion :

La performance 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 a 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 ensuite comptés et les totaux sont affichés dans la matrice [32].

	Décision Positive	Décision Négative	
Etiquette positive	Vrais Positifs (VP)	Faux Négatifs(FN)	Pos
Etiquette négative	Faux Positifs (FP)	Vrais Négatifs (VN)	Neg
Total (T)	Ppos	Pneg	N

Tableau IV.2 : 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 : ("sensitivity" en anglais), 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).

$$sensibilité = \frac{VP}{VP+FN} \quad \text{(IV.1)}$$

La spécificité : (″specificity" en anglais) 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 de faux positifs**.**

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

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

Taux de faux négatifs (TFN) : proportion de positifs détectes 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.

$$précision = \frac{VP}{VP+FP} \quad \text{(IV.3)}$$

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.

$$TBC = \frac{VP+VN}{VP+FN+VN+FP} \quad \text{(IV.4)}$$

Le TBC est généralement appelé efficacité.

IV.3.2 Courbe ROC et mesure AUC :

Les courbes ROC (Receiver Operating Characteristic) ont ainsi été inventées au cours de la seconde guerre mondiale par des ingénieurs électriciens et des ingénieurs radar pour la mise au point de moyens efficaces de détection des avions japonais et pour améliorer la séparation entre les signaux de radar et le bruit de fond [33].

Cette représentation a été largement étendue au domaine médical afin de discriminer entre la population des patients atteins de trouble vocal et des patient sains.

La courbe ROC est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test, calculée pour toutes les valeurs seuils possibles. Elle est la courbe qui va résumer les performances du test. Elle a pour objectif de déterminer la valeur seuil optimale de la sensibilité et de la spécificité du test. Nous présentons en figure (IV.2) un exemple d'une courbe ROC. On a en abscisse le taux de faux positifs et en ordonnée le taux de faux négatifs.

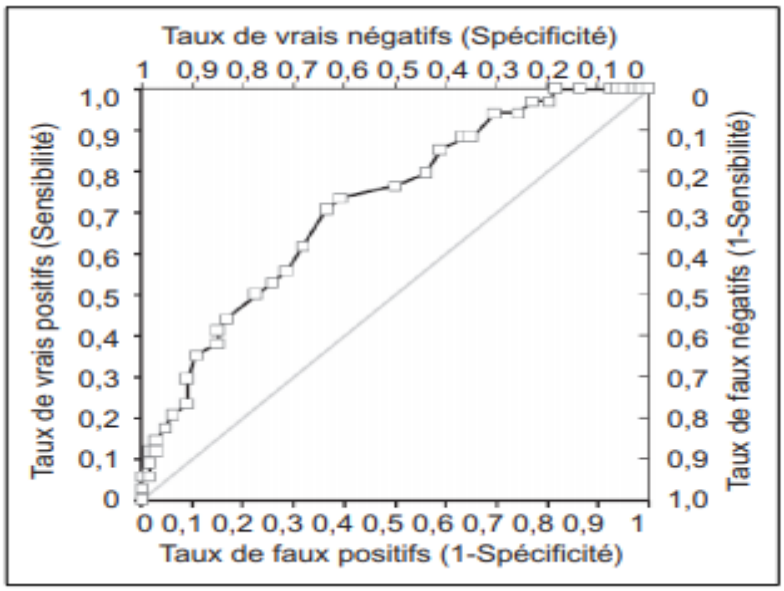

Figure IV.2 : Exemple de courbe ROC représentant la relation existant entre la sensibilité et la spécificité d'un test, calculée pour toutes les valeurs seuils possibles [34].

Aire sous la courbe (AUC) : L'AUC (Area Under Curve) est une mesure globale de la performance du test parmi les plus utilisées. C'est la mesure de l'aire de la surface située sous le tracé du ROC. Formellement, cette valeur correspond à l'intégral de cette fonction. Elle varie entre 0.5 dans le cas d'un test non informatif à 1 dans le cas d'une performance parfaite.

Ainsi, une AUC de 0,50 signifie que le test est mauvais et qu'il ne fait pas mieux que la chance pour classer les individus. Plus l'aire sous la courbe est élevée, plus le test est performant.

IV.4 Conditions et paramètres de simulation :

En général, la conception d'un système de détection de la pathologie de la voix doit passer par deux phases. La phase d'apprentissage (Training) et la phase de test (Prediction), comme le montre la figure suivante. Les 156 fichiers audio de la base de données utilisée sont répartis en groupe comme dans le tableau (IV.1) et seront utilisés pour l'apprentissage et pour le test. Le pourcentage des fichiers utilisés dans chaque phase dépend des techniques choisies de la validation croisée.

Une fois les fichiers de parole de l'étape d'apprentissage sont connus, nous passons à l'étape d'extraction des paramètres (Feature extraction). Une matrice des coefficients MFCC constituée de 12 colonnes (MFCC1, MFCC2, …, MFCC12) et un nombre de lignes égale au nombre de fichiers sélectionnés. Les différentes étapes du calcul des coefficients MFCC d'un seul fichier ont été présentées dans le premier chapitre et illustrées par un exemple dans le deuxième chapitre. Nous avons élaboré un programme MATLAB qui nous a permis d'avoir automatiquement les différentes matrices MFCC des sous-répertoires (normal, paralysis, adductor spasmodic dysphonia, keratosis, vocal fold nodules, vocal fold polyp) de la base d'apprentissage. Pour chaque fichier, la valeur de chaque coefficient MFCC (i) est la valeur moyenne des différentes valeurs du même coefficient dans les différentes trames du fichier.

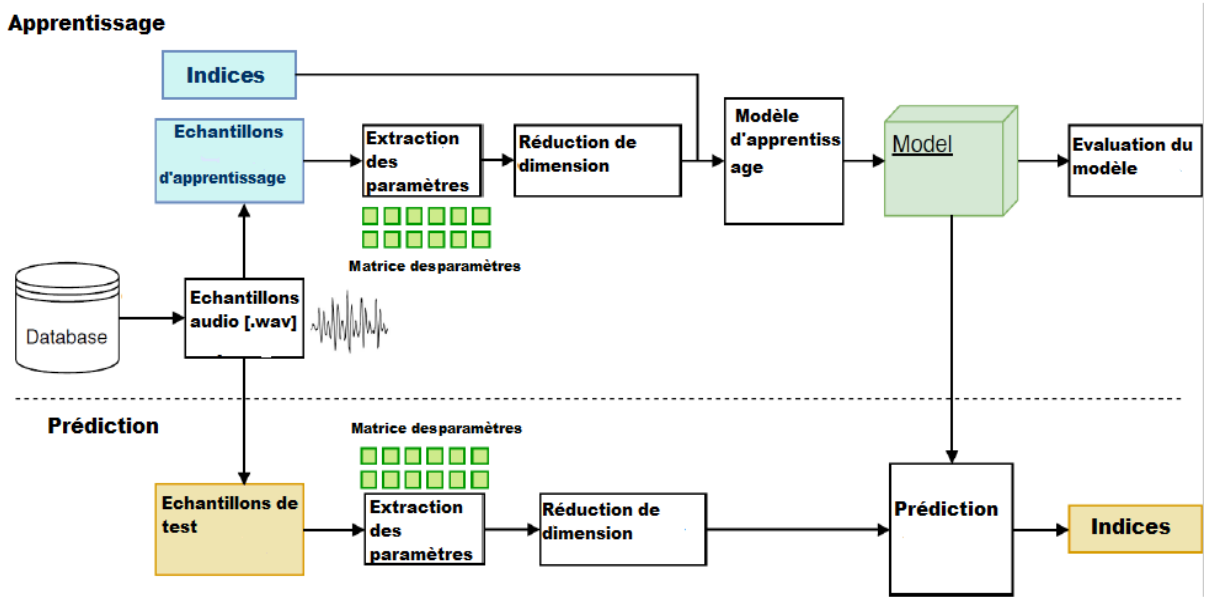

Figure IV.3 : Eléments d'un système de détection des pathologies de la parole.

En plus de la matrice des MFCC, nous avons introduit une deuxième matrice des paramètres de variations du pitch, de l'intensité et du HNR. Un fichier script programmé sous l'environnement PRAAT nous a permis d'obtenir les valeurs du jitter local, jitter rap, shimmer local, shimmer apq3 et le HNR de chaque fichier de parole automatiquement. Ainsi, pour chaque sous-répertoire de la base de données d'apprentissage nous aurons une matrice de ces nouveaux paramètres avec 5 colonnes et le même nombre de lignes que celui du nombre de fichiers utilisés. Enfin, notre matrice des paramètres extraits sera constituée des deux matrices précédentes avec 17 colonnes et un nombre de lignes égal au nombre des fichiers.

L'étape de réduction de dimensionnalité (Dimentionality reduction) n'a pas été implémentée à cause du nombre moins élevé des paramètres sélectionnés et des fichiers choisis. Par ailleurs, chaque fichier traité de la base de données aura un label (étiquette) soit normal soit pathologique. Dans le deuxième cas, le type de pathologie devra être mentionné aussi.

Après l'étape d'apprentissage, nous obtenons des modèles SVM et KNN pour des signaux normaux, des signaux pathologiques et pour les signaux de chaque type de pathologie.

Les performances de ces modèles peuvent être obtenues par les mesures de performances citées en haut et durant l'étape de test (prédiction), en choisissant un ensemble de fichiers test selon la technique de validation croisée sélectionnée, suivi par l'étape d'extraction des paramètres et éventuellement l'étape de réduction de dimensionnalité si elle est utilisée, enfin la prédiction du label à la sortie du système en fonction du modèle SVM ou KNN utilisé. La matrice de confusion peut donc être calculée à partir des labels réels de l'ensemble de test et les labels à la sortie du système de détection. Ce qui facilite le calcul des mesures de performances.

IV.5 Résultats et interprétations de la détection des pathologies :

IV.5.1 Résultats de la validation simple :

Commençons par une validation simple en divisant l'ensemble des échantillons en deux parties aléatoires, une partie pour l'apprentissage (80%) et une partie pour le test (20%), nous aurons ainsi 125 fichiers pour l'apprentissage et 31 fichiers pour le test.

La méthode de classification KNN est utilisée avec un nombre k=3, 5 et 10 des plus proches voisins utilisés par l'algorithme de classification, les deux types de distances : la distance euclidienne et celle de Manhattan.

La méthode de classification SVM est utilisée avec ses paramètres de base par défaut tout en changeant le type de noyau soit linéaire, polynomial soit gaussien.

Les résultats de nos implémentations présentés dans le tableau (IV.3) sont exprimés en termes d'efficacité, sensibilité et spécificité pour chaque variante.

Méthode		**Efficacité (%)**	**Sensibilité (%)**	**Spécificité (%)**
KNN (K=3)	Euclidienne	87.50	95.45	70
	Manhattan	87.50	100	60
KNN (K=5)	Euclidienne	93.75	100	80
	Manhattan	87.50	100	60
KNN (K=10)	Euclidienne	87.50	100	60
	Manhattan	87.50	95.45	70
SVM (Noyau Linéaire)		**96.88**	100	**90**
SVM (Noyau Polynomial)		84.38	100	50
SVM (Noyau Gaussien)		90.63	100	70

Tableau IV.3 : Mesures de performance de la détection de la pathologie avec KNN et SVM, cas d'une validation simple.

Nous remarquons que la méthode de classification SVM avec un noyau linéaire assure les meilleures performances en terme d'efficacité (96.88 %) et de spécificité (90 %). En ce qui concerne la sensibilité, elle est presque la même pour toutes les variantes. Les variantes du KNN donnent des mesures de performance moins par rapport à celles des variantes du SVM. Néanmoins, la variante du KNN avec k=5 et une distance euclidienne présente les meilleurs résultats par rapport aux autres variantes de la méthode de classification KNN.

IV.5.2 Résultats de la validation croisée :

Nous nous intéresserons maintenant aux résultats de la détection voix normale / voix pathologique obtenus dans le cas d'une validation croisée d'un classifieur SVM seulement. Nous avons mené une série de tests dans lesquels nous avons utilisé la validation croisée où l'ensemble des échantillons est divisé en (kv) parties (k-fold cross-validation) comme expliqué

au début du chapitre. Les mesures de performances de la détection normale/pathologique obtenues dans ce cas sont présentées dans le tableau ci-dessous :

Détection Normale / Pathologique		**Sensibilité (%)**	**Spécificité (%)**	**Efficacité (%)**	**AUC (%)**
kv = 5	Noyau linéaire	**94.34**	**93.20**	**93.59**	66.09
	Noyau gaussien	90.57	94.17	92.95	66.24
	Noyau polynomial	83.02	93.20	89.74	65.84
kv = 8	Noyau linéaire	90.57	94.17	92.95	66.24
	Noyau gaussien	92.45	93.20	92.95	66.24
	Noyau polynomial	84.91	93.20	90.38	65.92
kv = 10	Noyau linéaire	**94.34**	**93.20**	**93.59**	66.06
	Noyau gaussien	90.57	**95.15**	93.59	**66.36**
	Noyau polynomial	83.02	93.20	89.74	65.85

Tableau IV.4: Mesures de performance pour la détection voix normale / voix pathologique par validation croisée.

Dans le tableau (IV.4), chaque colonne indique la valeur la plus élevée en gras de la sensibilité, la spécificité, l'efficacité et l'AUC pour des valeurs de kv = 5, 8, 10 pour les trois variantes du classificateur SVM (noyau linéaire, gaussien et polynomial). Comme nous pouvons le voir, la valeur de l'AUC est autour de 66 % et la valeur maximale obtenue est de 66.36 % dans le cas d'un noyau gaussien avec kv = 10. Une autre constatation est que le classificateur SVM avec un noyau linéaire assure les meilleures performances en terme d'efficacité (93.59 %), de spécificité (93.20 %) et de sensibilité (94.34 %), et ce pour les deux valeurs de la validation croisée kv = 5 et kv = 10.

La figure (IV.4) présente les courbes ROC du classificateur SVM avec une validation croisée kv = 10 pour les trois types de noyaux : linéaire, gaussien et polynomial. Les ROC démontrent que les trois noyaux peuvent être utilisés pour la détection avec une capacité de discrimination élevée dans le cas d'un noyau linéaire et gaussien par rapport au noyau polynomial.

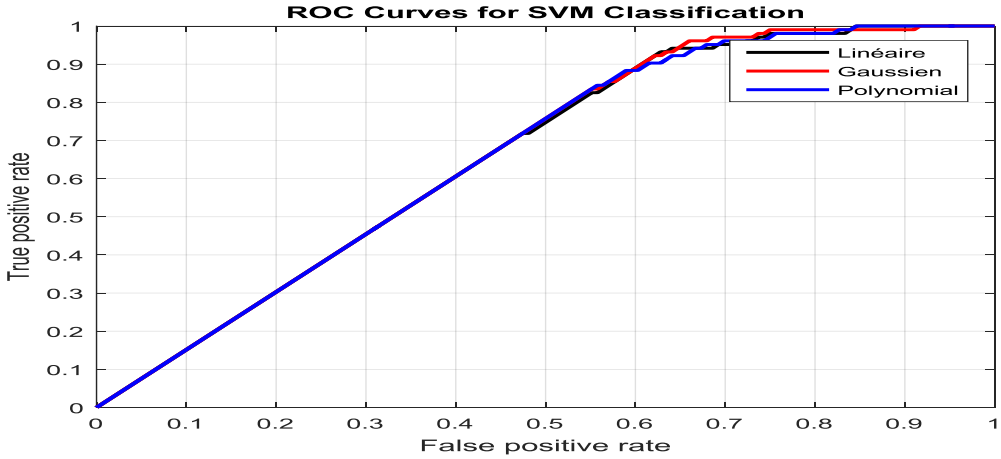

Figure IV.4 : Courbes ROC pour différents noyaux du SVM.

IV.5.3 Résultats en fonction des paramètres choisis :

Nous avons étudié aussi l'influence des paramètres choisis sur les performances de notre système de détection voix normale/ voix pathologique. Nous avons sélectionné la variante du classificateur SVM avec noyau linéaire et 10 kfold pour la validation croisée. L'utilisation de tous les 17 paramètres améliore toutes les mesures de performances, comme le montre le tableau suivant. Par contre l'utilisation des 12 paramètres MFCC seuls assure des performances moins par rapport à l'utilisation mixte des paramètres. De même, lors de l'utilisation des 5 différents paramètres du Jitter, Shimmer et du HNR sans les paramètres MFCC, le classificateur a des performances acceptables, mais moins par rapport aux deux variantes précédentes. Alors, la première variante a les plus hautes précisions car elle combine les avantages de tous les paramètres.

Paramètres utilisés	**Sensibilité (%)**	**Spécificité (%)**	**Efficacité (%)**	**AUC (%)**
Tous les paramètres	**94.34**	**93.20**	**93.59**	**66.06**
MFCC seulement	88.68	87.38	87.82	64.16
Jitter, Shimmer, HNR	81.13	87.38	85.26	64.90

Tableau IV.5: Mesures de performance pour la détection voix normale / voix pathologique selon les paramètres utilisés.

IV.6 Résultats et interprétations de la classification des voix pathologiques :

Dans notre projet, un système à deux niveaux a été proposé pour détecter le type de pathologie vocale. Au premier niveau, le système identifie si la voix donnée est normale ou non. Pour une voix normale, aucune classification supplémentaire n'est nécessaire. Si ce n'est pas normale, elle est reconnue comme voix pathologique puis on passe au deuxième niveau qui vise à identifier le type particulier de pathologie. La figure (IV.5) illustre les étapes impliquées dans le système proposé.

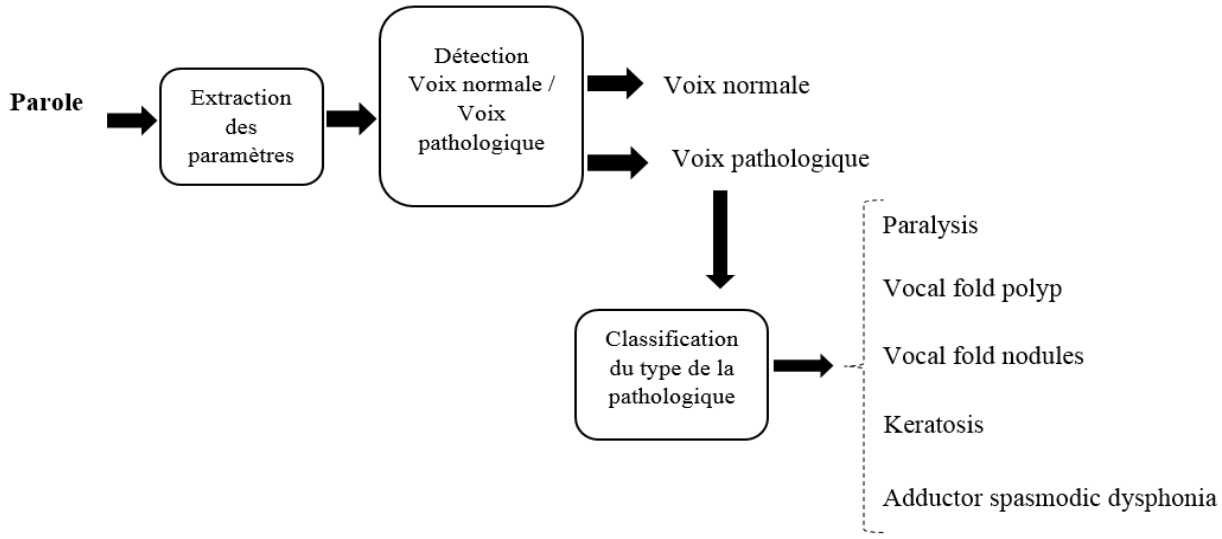

Figure IV.5 : Système proposé pour la détection et la classification des pathologies.

Les expériences, tests et résultats du premier niveau ont été présentés et interprétés dans la section précédente. Cette partie sera donc consacrée seulement au deuxième niveau réservé à la classification du type de la pathologie. Le tableau (IV.6) résume les résultats de la classification des pathologies par paire (classification binaire entre deux pathologies à chaque fois) appelée aussi classification multi-classes Un-contre-Un (One-Versus-One). Nous remarquons que les taux de classification Un-contre-Un sont acceptables avec une nette supériorité de la classification Polyp vs Nodules avec une efficacité de 82.05 % et un AUC 68.49 %.

Classification par paire complète	**Sensibilité (%)**	**Spécificité (%)**	**Efficacité (%)**	**AUC (%)**
Paralysis vs Adductor	60	44.44	52.63	52.92
Paralysis vs Keratosis	55	73.08	65.22	59.62
Paralysis vs Polyp	55	55	55	53.63
Paralysis vs Nodules	75	78.95	76.92	65.79
Adductor vs Keratosis	84.62	66.67	77.27	61.71
Adductor vs Polyp	55	61.11	57.89	54.21
Adductor vs Nodules	68.42	66.67	67.57	63.94
Keratosis vs Polyp	20	73.08	50	53.91
Keratosis vs Nodules	63.16	84.62	75.56	63.57
Polyp vs Nodules	80	84.21	82.05	68.49

Tableau IV.6 : Mesures des performances de la classification multi-classes Un-contre-Un.

Ce dernier tableau montre les mesures de performances du classificateur SVM multi-classes Un-Contre-Tous (One-Versus-All).

Type de pathologie	**Sensibilité (%)**	**Spécificité (%)**	**Efficacité (%)**	**AUC (%)**
Paralysis vs All	15	96.39	80.58	52.28
Keratosis vs All	90.91	15.38	71.84	62.64
Polyp vs All	98.80	0	79.60	50.95
Nodules vs All	88.10	5.26	72.82	60.68
Adductor spasmodic vs All	100	0	**82.52**	61.81

Tableau IV.7 : Mesures des performances de la classification multi-classes Un-contre-Tous.

Comme résumé du résultat de la classification, la meilleure efficacité est obtenue pour la classification de la pathologie Adductor Spasmodic, puis celle du Paralysis, après ça Vocal fold polyp et Vocal fold nodules. Celle du Keratosis présente la plus faible valeur par rapport aux autres.

IV.7 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 où nous avons étudié l'influence des divers paramètres et choix sur les mesures de performances.

Conclusion générale

L'objectif de notre travail était la détection et la classification des pathologies de la voix par la méthode de classification SVM et l'évaluation de l'impact des paramètres sélectionnés du signal vocal sur cette identification. Plusieurs variantes du classificateur SVM et quelques variantes du classificateur KNN ont été utilisés pour comparer leurs performances.

Le 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. Le système proposé et étudié emploie des mesures non invasives, peu coûteuses et entièrement automatisées des caractéristiques du signal vocal.

À partir des résultats obtenus, nous avons remarqué que les paramètres MFCC, Jitter, Shimmer et le HNR sont plus pertinents pour discriminer les pathologies des cordes vocales, lorsqu'ils sont évalués individuellement et surtout lorsqu'ils sont combinés.

Nous avons montré que les classifieurs SVM et KNN peuvent détecter efficacement les voix pathologiques et normales. Les résultats de la validation simple confirment que la méthode de classification SVM avec un noyau linéaire assure les meilleures performances avec une efficacité de 96.88 %. Pour ce qui est des résultats de la détection obtenus dans le cas d'une validation croisée d'un classifieur SVM seulement. La valeur maximale de l'AUC obtenue est de 66.36 % dans le cas d'un noyau gaussien avec kv = 10, le classificateur SVM avec un noyau linéaire assure les meilleures performances : une efficacité = 93.59 %, une spécificité =93.20 % et une sensibilité = 94.34 %, et ce pour les deux valeurs de la validation croisée kv = 5 et kv = 10.

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. Pour les SVM à Un-Contre-Tous (One-Versus-All) et à Un-contre-Un (One-Versus-One), nous avons prouvé que les performances de classification des SVM sont très bonnes.

Dans les travaux futurs, pour améliorer le taux de classification obtenu, nous proposons d'améliorer la phase de classification par l'utilisation d'un système hybride en combinant plusieurs techniques d'apprentissage automatique, d'augmenter le nombre de paramètres utilisés dans la phase d'extraction des paramètres, d'étudier l'influence de la sélection de certains paramètres sur les performances du système, d'optimiser certaines valeurs et coefficients du système, d'étendre cette étude sur d'autres bases de données et de développer un système de diagnostic en ligne.

Bibliographie

[1] 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.

[2] Xiang Wang et al., "Discrimination between pathological and normal voices using GMM-SVM approach," *Journal of Voice*, vol. 25, n°1, 2011.

[3] Laura Verde et al., "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16246-16255, 2018.

[4] Zuzana Dankovičová et al., "Machine Learning Approch to dysphonia detection," *Applied Sciences*, vol. 8, n°10, c.1927, 2018.

[5] Mounir Boudjerda, "Analyse du signal de parole pour l'évaluation automatique des voix pathologiques," Thèse de Doctorat, Université Mohammed Seddik BENYAHIA-Jijel, Algérie, 2018.

[6] Marie Daumet, "Élaboration de profils types en fonction de pathologies vocales à partir de critères d'analyse objectifs, par le logiciel Vocalab," Mémoire du Certificat de Capacité d'Orthophonie, Université de Nice Sophia Antipolis, Nice, France, 2015.

[7]http://tpe-sur-la-voix.mozello.fr/ii-le-mecanisme-de-production-du-son/2-les-parametres-de-lavoix/?fbclid=IwAR2YVwCW7FQ6muzsGfFVpGL9uJqkjYCNmLD0w9dIRb71zPjHKVQCI3u6uKg

[8]http://voixdenseignant.canalblog.com/archives/2009/11/12/15777106.html?fbclid=IwAR0JQdufvd7FNMt4R0cQiH67O2HBzpOKgOYE6QQkGQJthanGde8w8rF5i7U

[9] Jorge Martinez, Hector Perez and Enrique Escamilla, "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) techniques," in 22nd International Conference on Electronics Communications and Computing, United States, February 2012, pp. 248-251.

[10] Filipe Velho, "La reconnaissance du locuteur à l'aide de la transformée en ondelettes continue," Mémoire de la Maitrise en Génie Electrique, Ecole de Technologie Supérieure, Université du Québec, Montréal, Canada, 8 Mars 2006.

[11] Rishiraj Mukherjee, "Speaker Recognition Using Shifted MFCC," Master Thesis, University of South Florida, January 2012.

[12] Fathi Brioua, "Estimation du pitch d'un signal de parole," Mémoire de Master II, Université Mohammed Seddik BENYAHIA-Jijel, Algérie, 2011.

[13] www.praat.org

[14]https://www.institut-numerique.org/chapitre-3-levaluation-subjective-et-objective-de-la-voix-5194afdf80cc9

[15] Gilles Pouchoulin, " Approche statistique pour l'analyse objective et la caractérisation de la voix dysphonique," Thèse de Doctorat, Université d'Avignon et des Pays de Vaucluse, France, 2008.

[16] Alain Ghio, " Bilan instrumental de la dysphonie," *Rééducation orthophonique*, Ortho édition, pp. 9-29, 2013.

[17] M. Farrús, J. Hernando and P. Ejarque, " Jitter and Shimmer measurements for speaker recognition," in INTERSPEECH, Antwerp, Belgium, August 27-31, 2007, pp. 778-781.

[18] J. Paulo Teixeiraa and A. Gonçalves, " Accuracy of jitter and shimmer measurements," *in* Procedia Technology, International Conférence on Project Management / HCIST 2014 - International Conférence on Health and Social Care Information Systems and Technologies, Portugal, vol.16, pp. 1190-1199, 2014.

[19] S. Chebbout, "La classification automatique," Chapitre du cours : "Reconnaissance des Formes," pour étudiants Master 1, Vision Artificielle, Université Badji Mokhtar-Annaba.

[20] Fatma Karem, Mounir Dhibi et Arnaud Martin, " Combinaison de classification supervisée et non-supervisée par la théorie des fonctions de croyance," *Revue des Nouvelles Technologies de l'Information*, vol. RNTI-E-23, pp. 53-64, 2012.

[21] André Quinquis, *Le traitement du signal sous Matlab*, Hermes Science, Lavoisier, Paris, France, 446p, 2007.

[22] V.Vapnik, *The nature of statistical learning theory*, Springer-Verlag, New-York, 1995.

[23] Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik, "A training algorithm for optimal margin classifiers," in Fith Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania, USA, 27-29 July 1992, pp. 144-152.

[24] http://dspace.univ-tlemcen.dz/bitstream/112/322/11/ChapitreII.pdf

[25] https://zestedesavoir.com/tutoriels/1760/un-peu-de-machine-learning-avec-les-svm/

[26] Gilles Lebrun, "Sélection de modèles pour la classification supervisée avec des SVM (Séparateurs à Vaste Marge). Application en traitement et analyse d'images," Thèse de Doctorat, Université de Caen Basse-Normandie, France, 2006.

[27] Labiad Al, "Sélection des mots clés basée sur la classification et l'extraction des règles d'association," Mémoire de la Maitrise en Mathématiques et Informatique appliquée, Université du Québec, Montréal, Canada, Juin 2017.

[28] https://info.blaisepascal.fr/nsi-les-k-plus-proches-voisins

[29] Kay Elemetrices Corp, Disorderead Voice Database Model 4337, Verl. 03, Massachusetts Eye and Ear Infirmary Voice and Speech Lab, 2002.

[30]https://fr.qaz.wiki/wiki/Crossvalidation_(statistics)?fbclid=IwAR3deVY3STj5CTc4fwUMqclPYswqNmRxlwaXtVtXKdhAr4oQlk_rxJo0hzc

[31]https://docs.microsoft.com/fr-fr/analysis-services/data-mining/training-and-testing-data-sets?view=asallproductsallversions&fbclid=IwAR0bTlyqlaJliUkiPSvk6ymlknaPFtaoFWUnJ6xsjPKqU7y5JlOtu7ItdJg

[32] 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.

[33] https://www.xlstat.com/fr/solutions/fonctionnalites/courbes-roc

[34] D. Bertrand et al., " Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture," *L'Année psychologique*, vol. 110, pp. 299-320, 2010. (https://www.cairn.info/revue-l-annee-psychologique1-2010-2-page-299.htm)