

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



Département d'Electronique  
Laboratoire Signal & Communications

## Thèse de Doctorat

Présentée par :  
**Mr Mohamed DAHMANI**

Master/Ingénieur d'Etat en Électronique, ENP Alger

Pour l'obtention du titre de  
**Docteur Troisième Cycle (D/LMD) en Électronique**

Thème

# *Analyse Paramétrique du Signal Glottique en Vue de la Classification Automatique des Pathologies des Cordes Vocales*

Soutenue publiquement le 10/07/2021 devant le jury composé de :

Présidente	: Mme Latifa HAMAMI	Professeur	ENP
Directrice de thèse	: Mme Mhania GUERTI	Professeur	ENP
Examineurs	: Mme Malika KEDIR	Professeur	USTHB
	M. Salim CHITROUB	Professeur	USTHB
	M. Chérif LARBES	Professeur	ENP

ENP 2021



**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**  
**Ecole Nationale Polytechnique**



Département d'Electronique  
Laboratoire Signal & Communications

**Thèse de Doctorat**

Présentée par :  
**Mr Mohamed DAHMANI**

Master/Ingénieur d'Etat en Électronique, ENP Alger

Pour l'obtention du titre de  
**Docteur Troisième Cycle (D/LMD) en Électronique**

Thème

***Analyse Paramétrique du Signal Glottique en Vue de  
la Classification Automatique des Pathologies des  
Cordes Vocales***

Soutenue publiquement le 10/07/2021 devant le jury composé de :

Présidente	: Mme Latifa HAMAMI	Professeur	ENP
Directrice de thèse	: Mme Mhania GUERTI	Professeur	ENP
Examineurs	: Mme Malika KEDIR	Professeur	USTHB
	M. Salim CHITROUB	Professeur	USTHB
	M. Chérif LARBES	Professeur	ENP

**ENP 2021**

في هذا العمل ، تم تقديم مناهج جديدة لتحسين التحليل والتمييز بين الأصوات المرضية باستخدام تقنيات معالجة الإشارات غير الخطية. أولاً ، يتم استكشاف المعلومات المقدمة من مخطط التكرار (Recurrence Plot : RP) والتحليل الكمي للتكرار (Recurrence Quantification Analysis : RQA) ، الخاصة بشكل موجة الإشارة الغلوتينية (glottal signal) ، لتشخيص اضطرابات الصوت البشري وتصنيفها. لهذا الغرض ، تم أولاً استخلاص شكل موجة الإشارة الغلوتينية من الإشارة الصوتية باستخدام خوارزمية التصفية العكسية التكرارية التكيفية (Iterative Adaptive Inverse Filtering). تم تطبيق RP و RQA على موجة الإشارة الغلوتينية لتقييم الحركة الديناميكية للحبال الصوتية. تم استخدام ثمانية (08) معلومات RQA كمجموعة ميزات لإدخال لخوارزمية (PSO-SVM Particle Swarm Optimization-Support Vector Machines) : للفصل بين الأصوات العادية والمرضية. ثانياً ، نقتراح نموذجاً معيارياً يعتمد على التحليل الكمي للتكرار المتقاطع (Cross Recurrence Quantification Analysis) لإظهار كيف يمكن تكييفه لمقارنة وتقييم وقياس التغيرات الصوتية بعد عملية التعليم الأورطوفوني. من خلال مقارنة التباين CRQA قبل وبعد التعليم مع الأصوات العادية ، تم تمييز مراحل الصوت المرضي بسهولة. من أجل التحقق من صحة الاختبارات ، تم اعتماد قاعدة بيانات ساربروكن الصوتية (Saarbrücken Voice Database) ، حيث تم اختيار إشارة الصوت [a:] للعينات التي تلفظت بها أربع مجموعات من الأشخاص : الأشخاص الذين يعانون من شلل في الحبال الصوتية ، والأشخاص الذين يعانون من أورام البوليبات ، والأشخاص الذين يعانون من تشنج والأشخاص العاديين كمرجع. لضمان صحة النتائج وقدرات التعميم ، تم استخدام نظام K-fold Cross-Validation. أظهرت النتائج التي تم الحصول عليها فعالية RQA المطبقة على الإشارة الغلوتينية كتقنية لاستخراج الميزات. في الواقع ، قدم PSO-SVM ، كطريقة تصنيف ، أداة فعالة لتشخيص وتصنيف الأصوات المرضية بدقة أين بلغت دقة تشخيص هاته الأمراض نسبة : 97.41%.

**الكلمات المفتاحية :** تشخيص ، تصنيف أمراض الحبال الصوتية ، مخطط التكرار ، التحليل الكمي للتكرار ، PSO-SVM ، الإشارة الغلوتينية.

## Abstract

In the present work, new approaches are presented to improve the analysis and the discrimination of pathological voices using nonlinear signal processing techniques. Firstly, the information provided by the Recurrence Plot (RP) and the Recurrence Quantification Analysis (RQA), referred to the glottal signal waveform, are explored for detection and classification of the human voice disorders. For this purpose, the glottal signal waveform was firstly extracted from the voice signal using an Iterative Adaptive Inverse Filtering (IAIF) algorithm. The RP and the RQA were applied to a glottal signal waveform to evaluate the dynamic process of Vocal Folds (VFs). Eight RQA parameters were used as an input features set of a hybrid Particle Swarm Optimization-Support Vector Machines (PSO-SVM) algorithm to segregate between normal and pathological voices. Secondly, we suggest a parametric model based on Cross Recurrence Quantification Analysis (CRQA) for showing how it can be adapted to compare, assess and quantify the articulation changes of Vocal Folds, after voice therapy. By comparing the CRQA measures variation before and after the reeducation to the normal voices, pathological voice stages were easily discriminated. For the tests validation, Saarbrücken Voice Database (SVD) was adopted, in which the vowel [a:] was selected for samples uttered by four groups of subjects: persons having suffered from vocal folds paralysis, persons having vocal folds polyps, persons having spasmodic dysphonia and normal persons considered as reference. To ensure the validity of the results and the generalization capabilities, the K-fold Cross-Validation scheme was employed. The obtained results show the effectiveness of RQA applied to the glottal signal as a feature extraction technique. Indeed, the PSO-SVM, as a classification method, presented an effective tool for the diagnosis and the classification of pathological voices with an accuracy of 97.41%.

**Key items :** Glottal signal, Vocal folds, pathological, Recurrence Plot, Recurrence Quantification Analysis, assesment, classification, PSO-SVM, Saarbrücken Voice Database.

## Résumé

Dans ce travail, de nouvelles approches sont présentées pour améliorer l'analyse et la discrimination des pathologies des Cordes Vocales (CVs) en utilisant des techniques de traitement du signal non linéaire. En premier lieu, les informations obtenues par le diagramme de récurrence : Recurrence Plot (RP) et l'Analyse Quantitative de Récurrence (RQA), faisant référence à la forme d'onde du signal glottique, sont explorées pour la détection et la classification des troubles de la voix humaine. Pour cela, le signal glottique a été extrait premièrement du signal vocal à l'aide d'un algorithme de Filtrage Inverse Adaptatif : Iterative Adaptive Inverse Filtering (IAIF). Les RP et RQA ont été appliqués au signal glottique pour obtenir huit (08) indicateurs, ou mesures. Ces derniers varient sensiblement avec la présence des pathologies au niveau des CVs. De ce fait, leurs valeurs moyennes sont utilisées comme un vecteur de caractéristiques d'entrée de d'un l'algorithme Particle Swarm Optimization-Support Vector Machines (PSO-SVM) pour séparer les voix normales et pathologiques. Deuxièmement, un modèle paramétrique basé sur l'Analyse Quantitative de Récurrence Croisée (CRQA) est proposé pour montrer comment elle peut être adaptée pour comparer, évaluer et quantifier les changements d'articulation des plis vocaux après une thérapie ou rééducation. En comparant la variation de ces paramètres avant et après la thérapie aux voix normales, les stades pathologiques ont été facilement discriminés. Pour la validation des tests, la Base de Données Saarbrücken Voice Database (SVD) a été adoptée, dans laquelle la voyelle longue [a:] a été sélectionnée pour les échantillons prononcés par quatre groupes de personnes : sujets souffrant de la paralysie des CVs ; ayant des polypes de CVs ; ayant une dysphonie spasmodique et possédant une voix normale, considérée comme référence. Pour garantir la validité des résultats, la méthode K-Folds Kross Validation, a été utilisée. Les résultats obtenus montrent l'efficacité du RQA appliqué au signal glottique comme technique d'extraction des caractéristiques. En effet, le PSO-SVM, en tant que méthode de classification, semble un outil efficace pour le diagnostic et la classification des pathologies citées avec un taux de 97.41%.

**Mots clés :** pathologies des cordes vocales, Saarbrücken Voice Database, signal glottique, Recurrence Plot, Analyse Quantitative de Récurrence, classification, évaluation, PSO-SVM.

## *Dédicaces*

---

*Je dédie ce modeste travail,*

*A mes très chers parents, pour leur soutien et la patience dont ils ont fait preuve durant toutes les années de ma formation ;*

*A ma grande famille ;*

*Ainsi qu'à tous les électroniciens de ma promotion;*

*A tous ceux qui m'aiment et qui me souhaitent le bonheur ;*

*Mohamed*

# *Remerciements*

---

*Avant tout, je remercie Allah, de m'avoir donné la force d'aller jusqu'au bout.*

*Je tiens à remercier Madame **Mhania GUERTI**, Professeur à l'Ecole Nationale Polytechnique, pour avoir accepté de diriger et de suivre ce travail avec disponibilité, patience et bienveillance.*

*Je tiens à exprimer ma profonde gratitude et mes remerciements les plus vifs à Madame **Latifa HAMAMI**, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, pour avoir fait l'honneur d'accepter de présider le jury de cette soutenance.*

*Mes vifs remerciements, vont à Madame **Malika KEDIR**, Professeurs à l'USTHB. Je remercie également Mrs **Chérif LARBES** et Mr **Salim CHITROUB**, Professeurs à l' Ecole Nationale Polytechnique d'Alger et à l'Université des Sciences et Technologies Houari Boumediene Alger, d'avoir bien voulu accepter d'examiner et d'évaluer ce travail.*

*Je tiens aussi à remercier tous les enseignants du département d'Electronique de l'ENP qui ont contribué à ma formation et ce depuis le tronc commun jusqu'à la dernière année de graduation.*

# Table des matières

Liste des Tableaux

Liste des Figures

Liste des Abréviations

**Introduction Générale** **14**

---

**Chapitre 1 : Généralités Sur La Production de La Parole** **18**

---

1. Introduction.....	19
2. Anatomie de l'appareil phonatoire humain .....	19
2.1. Partie sous glottique .....	19
2.2. Partie glottique .....	20
2.3. Partie supra-glottique .....	21
3. Mécanisme de production de la parole .....	21
3.1. Vibration des cordes vocales .....	21
3.2. Contrôle de fréquence .....	23
3.3. Contrôle de l'intensité vocale .....	23
3.4. Articulation .....	23
4. Voyelles et consonnes .....	24
4.1. Voyelles .....	24
4.2. Consonnes .....	24
5. Modélisation de l'appareil phonatoire .....	25
6. Modèle de la source glottique .....	27
7. Techniques d'extraction de l'onde glottique .....	30
7.1. Filtrage inverse en phase fermeture .....	30
7.2. Filtrage inverse itératif et / ou adaptatif .....	31
8. Conclusion .....	33

**Chapitre 2 : Analyse Paramétrique des Pathologies des Cordes vocales** **34**

---

1. Introduction.....	35
2. Pathologies des Cordes Vocales .....	35
2.1. Polypes des cordes vocales .....	36
2.2. Nodules des cordes vocales .....	36
2.3. Les granulomes .....	37
2.4. Paralysie des cordes vocales .....	37

2.5. Dysphonie spasmodique .....	38
2.6. Kystes de Corde Vocale .....	38
2.7. Cancer des Cordes Vocales .....	39
3. Moyens d'évaluation des pathologies vocales .....	39
3.1. Evaluation perceptive .....	39
3.2. Evaluation objective .....	41
3.2.1. Paramètres aérodynamique (physique) .....	41
3.2.1.1. Temps maximal de phonation .....	41
3.2.1.2. Débit d'air buccal .....	41
3.2.1.3. Pression sous glottique .....	41
3.2.2. Analyse acoustique de la voix .....	41
3.2.2.1. Intensité de la voix .....	41
3.2.2.2. Timbre .....	42
3.2.2.3. Fréquence fondamentale ( $F_0$ ) .....	43
3.2.2.4. Energie à court terme .....	43
3.2.2.5. Taux de Passage par Zéro (TPZ) .....	43
3.2.2.6. Jitter .....	44
3.2.2.7. Shimmer .....	44
3.2.2.8. Les Formants .....	44
3.2.3. Paramètres issus de l'analyse par banc de filtres .....	45
3.2.3.1. Coefficients Mel Fréquence CC .....	46
3.2.3.2. Coefficients PLP .....	47
3.3. Analyse paramétrique basé sur l'estimation du signal glottique .....	47
3.3.1. Paramètres glottiques (articulatoires) .....	47
3.3.1.1. Paramètres du signal glottique dans le domaine temporel .....	48
3.3.1.2. Paramètres du signal glottique dans le domaine fréquentiel .....	49
4. Le diagramme de récurrence .....	50
4.1. Espace de Phase.....	50
4.2. Diagramme de récurrence .....	51
4.3. Différentes structures dans le diagramme de récurrence RP .....	51
4.4. Analyse Quantitative du diagramme de récurrence (Recurrence Quantification Analysis) RQA .....	54
4.5. Les Mesures RQA .....	54
5. Evaluation de la voix pathologiques basée sur l'analyse quantitative de récurrence croisée appliquée au signal glottique .....	56
5.1. Tracé de récurrence croisé (Cross Recurrence Plots (CRP)) .....	57
5.2. Analyse quantitative de récurrence croisée .....	58
6. Conclusion .....	58



---

1. Introduction.....	60
2. Classifieur Machines à Vecteurs Supports .....	60
2.1. Classifieur SVM pour classification binaire .....	61
2.1.1. Cas de données linéairement séparables .....	61
2.1.2. Cas de données non linéairement séparables .....	63
2.2. Classifieurs SVM multi-classes .....	66
2.3. Avantages et inconvénients des SVM .....	67
2.3.1. Avantages des SVM .....	68
2.3.2. Inconvénients des SVM .....	68
3. Classifieur K Plus Proches Voisins .....	68
3.1. Principe des KNN .....	69
3.2. Avantages et inconvénient .....	70
4. Classifieur Naïve Bayes Network .....	70
4.1. Construction d'un Classifieur Naïve Bayes network .....	70
4.2. Avantages et inconvénients .....	71
5. Optimisation par les méthodes de métaheuristiques .....	72
5.1. Algorithme d'Optimisation par Essaim de Particules (PSO) .....	72
5.2. Principe de la technique PSO .....	73
5.3. Configuration de l'algorithme PSO .....	73
5.3.1. Taille de la population .....	73
5.3.2. Initialisation de la position et de la vitesse .....	73
5.3.3. Mise à jour de la position et de la vitesse .....	74
5.4. Influence des paramètres sur l'algorithme de PSO .....	75
5.4.1. Facteur d'inertie.....	75
5.4.2. Coefficients d'accélération $c_1$ et $c_2$ .....	76
5.4.3. Coefficient de constriction .....	76
5.4.4. Topologie de voisinage.....	76
5.5. Critère d'arrêt .....	77
6. Conclusion .....	78

---

1. Introduction.....	80
2. Base de Données .....	80
3. Analyse et classification des pathologies des cordes vocales en se basant sur les paramètres glottiques .....	81
3.1. Méthodologie .....	82
3.2. Analyse des paramètres articuloire glottique .....	83
3.3. Résultats de classification et discussion.....	84
4. Analyse Quantitative de Récurrence appliqué au signal glottique pour la classification des pathologies vocales .....	85
4.1. Méthodologie et formulation du problème .....	85
4.2. Technique RP appliquée au signal glottique .....	87
4.3. Application de l'analyse quantitative de récurrence (RQA) au signal glottique .....	89
4.4. Optimisation des paramètres de SVM par l'algorithme d'Optimisation PSO .....	91
4.5. Résultats de classification et discussions .....	95
5. Evaluation de la voix pathologiques basée sur l'analyse quantitative de récurrence croisée appliqué au signal glottique .....	95
5.1. Tracé de Récurrence Croisé (Cross Récurrences Plots : CRP) entre une voix normal et pathologique .....	95
5.2. Résultats et discussion .....	97
6. Conclusion .....	98

---

**Conclusions Générales et perspectives.....100**

**Références Bibliographiques .....** 104

## Listes des Tableaux

<b>Tableau</b>	<b>Titre</b>	<b>Page</b>
Tableau 2.1 :	Ségnification et définition des parametres de l'échelle de GRBAS.....	40
Tableau 2.2 :	Caractéristiques de diagramme de récurrence RP et Ses interprétations .....	53
Tableau 4.1 :	Détails sur la base de données utilisée .....	80
Tableau 4.2 :	Analyse statistique des paramètres glottique de chaque classe.....	83
Tableau 4.3 :	Résultats de Classifications KNN et SVM .....	84
Tableau 4.4 :	Rendement de la méthode proposée .....	93
Tableau 4.5 :	Resultats des differentes classifications.....	93
Tableau 4.6 :	Paramètres d'Analyse Quantitative de récurrence croisée entre une voix et patients avec différentes pathologies vocale savant et après une rééducation orthophonique (valeur moyenne $\pm$ la déviation standard.....	97

## Liste des figures

<b>Figure</b>	<b>Titre</b>	<b>Page</b>
Figure 1.1	: Anatomie de l'appareil phonatoire.....	21
Figure 1.2	: structure des cordes vocales.....	22
Figure 1.3	: Spectrogramme des fricatives non voisée [s]/voisée [h].....	23
Figure 1.4	: La forme temporelle (a) et spectrale (b) de la voyelle longue [a :].....	24
Figure 1.5	: Modélisation de l'appareil phonatoire Source-filtre.....	25
Figure 1.6a	: Signal de parole.....	26
Figure 1.6b	: Spectre de signale de parole.....	26
Figure 1.6c	: Autocorrélation du signal.....	27
Figure 1.6d	: Spectre du filtre conduit vocal.....	27
Figure 1.6e	: Forme du signal glottique.....	27
Figure 1.6f	: Forme spectre du signal glottique.....	27
Figure 1.7	: Schéma principal du signal glottique utilisé par la plupart des modèles.....	28
Figure 1.8	: Un filtre source linéaire invariant dans le temps modèle de production de la parole pour l'IAIF algorithme.....	31
Figure 1.9	: Diagramme en bloque de la méthode de filtrage inverse (IAIF).....	33
Figure 2.1	: Polype des cordes vocales.....	36
Figure 2.2	: Nodules des cordes vocales.....	37
Figure 2.3	: Granulome des cordes vocales.....	37
Figure 2.4	: Paralysie unilatérale des cordes vocales.....	38
Figure 2.5	: Kyste de corde vocale.....	38
Figure 2.6	: Formants d'un signal vocal.....	45
Figure 2.7	: Calcul des coefficients MFCC.....	46
Figure 2.8	: Processus de Markov discret.....	47
Figure 2.9	: Forme du signal glottique (a) :Voix Normal, (b):Nodules, (c):Paralysie des CVs.....	50
Figure 2.10	: Topologie caractéristique des diagrammes de récurrence (a) Homogène, (b) périodique, (c) dévié et (d) interrompu.....	52
Figure 3.1	: Séparation linéaire de deux classes par SVM.....	62
Figure 3.2	: Séparation non linéaire de deux classes par SVM.....	64
Figure 3.3	: Changement d'espace.....	65

Figure 3.4 : Principe un-contre-un pour $w=3$ .....	67
Figure 3.5 : Influence du nombre de voisins sur la classification principe KNN.....	69
Figure 3.6 : Minima global et local.....	72
Figure 3.7 : Principe du déplacement possible entre deux points.....	74
Figure 3.8 : Voisinages social et géographique.....	77
Figure 4.1 : Interface de l’Outil Aparat.....	82
Figure 4.2 : Méthodologie de classification.....	83
Figure 4.3 : boîtes à moustaches des variations des paramètres articulatoires en fonction de la présence des pathologies.....	84
Figure 4.4 : Organigramme de la méthode proposée.....	85
Figure 4.5 : (a) Signal glottique, (b) : structure RP correspondant au ; polypes.....	87
Figure 4.6 : (a) Signal glottique, (b) : structure RP correspondant à une voix normale.....	87
Figure 4.7 : (a) Signal glottique, (b) : structure RP correspondant à une Paralysie des cordes.....	87
Figure 4.8 : (a) Signal glottique, (b) : structure RP correspondant à une Paralysie des cordes.....	87
Figure 4.9 : Boîtes à moustaches (Boxplot) distribution des données des paramètres RQA de chaque type de voix (Normale, Polypes, Paralysie des CVs, dysphonie Spasmodique) ((a) : RR, (b) : DET, (c) : Vmax, (d) : TT, (e) : Lmax, (f) : ENTR.....	88
Figure 4.10 : 3-D scatter plot pour voir l’aspect discriminatif des parametres RQA.....	91
Figure 4.11 : Taux de classification de la méthode PSO-SVM et KNN.....	92
Figure 4.12 : Tracé de récurrence croisé entre une voix Normal et Polypes des CVs.....	96
Figure 4.13 : Tracé de récurrence croisé entre une voix Normal et dysphonie spasmodique.....	96
Figure 4.14 : Tracé de récurrence croisé entre une voix Normal et Paralysie des CVs.....	96
Figure 4.15 : Tracé de récurrence croisé entre une voix Normal et Nodules des CVs.....	96

<b>Acronyme</b>	<b>Signification</b>
<b>AQ</b>	Amplitude Quotient
<b>BD</b>	Bases de Données
<b>CIQ</b>	Closing Instance Quotient
<b>CPIF</b>	Closed Phase Inverse Filtering
<b>CRP</b>	Cross Recurrence Plots
<b>CRQA</b>	Analyse Quantitative de Récurrence Croisée
<b>CVs</b>	Cordes Vocales
<b>DET</b>	DETerminisme
<b>EGG</b>	ElectroGlottoGramme
<b>ENTR</b>	<i>ENTR</i> opie
<b>F<sub>0</sub></b>	Fréquence fondamentale
<b>GCI</b>	Glottal Closure Instances
<b>GOI</b>	Glottal Opening Instances
<b>HRF</b>	Facteur de Relation Harmonique
<b>HUPA</b>	Hospital Universitario Príncipe d'Asturias
<b>IA</b>	Intelligence Artificielle
<b>IAIF</b>	Iterative Adaptive Inverse Filtering
<b>Ka</b>	Phase d'ouverture glottique
<b>KFV</b>	<b>K</b> -Folds <b>K</b> ross Validation
<b>KNN</b>	<b>K</b> Nearest Neighbors
<b>Ko</b>	Phase de fermeture glottique
<b>L</b>	Longueur de ligne diagonale moyenne
<b>LAM</b>	<b>LAM</b> inarité
<b>LF</b>	<b>L</b> ijlencrants <b>F</b> ant
<b>L<sub>max</sub></b>	Longueur <b>max</b> imale des lignes diagonales
<b>LPC</b>	Linear Prediction Coding
<b>LOI</b>	Line Of Identity
<b>LPC</b>	Linear Prédicatif Coding
<b>MFCC</b>	Mel Frequency Cepstrum Coefficients
<b>MEEI</b>	Massachusetts Eye & Ear Infirmary

<b>NBN</b>	<b>Naïve Bayes Networks</b>
<b>ORL</b>	<b>Oto Rhyno Laryngologist</b>
<b>PLP</b>	<b>Perceptual Linear Predictive</b>
<b>PSO-SVM</b>	<b>Particle Swarm Optimization-Support Vector Machines</b>
<b>QAN</b>	<b>Quotient d'Amplitude Normalisé</b>
<b>QDA</b>	<b>Quadratique Discriminate Analysis</b>
<b>QoQ</b>	<b>Quotient d'Ouverture</b>
<b>QO</b>	<b>Quotient d'Ouverture</b>
<b>RP</b>	<b>Récurrents Plot</b>
<b>RQA</b>	<b>Recurrence Quantification Analysis</b>
<b>RR</b>	<b>Recurrence Rate</b>
<b>RTE</b>	<b><i>Recurrence Time Entropie</i></b>
<b>SVD</b>	<b>Saarbrucken Voice Database</b>
<b>SQ</b>	<b>Speed Quotient</b>
<b>TT</b>	<b><i>Trapping Time</i></b>
<b>TAP</b>	<b>Traitement Automatique de la Parole</b>

# *Introduction Générale*



Dans notre société, la communication est une faculté fondamentale de l'être humain et représente une clé pour la réussite sociale et professionnelle. Elle peut être réalisée sous différentes modalités : orale, écrite, gestuelle ... etc. La communication parlée reste malgré tout le centre de la communication humaine en permettant une communication simple et efficace pouvant contenir différents messages (informer, demander, exprimer un sentiment/avis,...etc.). Elle permet aussi grâce au riche vocabulaire et style de parole de véhiculer et traduire plusieurs nuances parfois difficiles à exprimer autrement [1]. Ce moyen de communication peut être affecté de certaines anomalies que l'on définit par pathologie du langage. Celle-ci doit être traitée pour pouvoir communiquer. Généralement ces anomalies touchent surtout l'appareil articulatoire par conséquent il nécessite différentes thérapies.

Les Cordes Vocales (CVs) présentent une partie très importante de l'appareil phonatoire humaine, permettant la production de sons voisés dans les langages de la parole. Par conséquent, la présence d'une pathologies provoque des changements importants dans le processus de la phonation. Généralement, les méthodes utilisées pour la détection des pathologies des CVs sont basées sur une évaluation perceptive effectuée par des experts d'écoute témoignant l'évolution de la voix. Cette évaluation peut prendre plusieurs formes, elle est basée sur différents critères perceptifs tels que l'intelligibilité, la compréhensibilité, le débit de parole, l'articulation, ... etc. Dans ce type d'analyse, un expert humain décidera, suite à l'écoute de la parole d'un patient, de la présence ou non d'une pathologie et jugera de son état d'avancement et de sa sévérité [2, 3]. Les méthodes les plus courantes de diagnostic des pathologies des CVs reposent sur l'implication des instruments utilisant les signaux de la source laryngée (laryngoscopiques, vidéolaryngoscopiques, électroglottographiques et stroboscopiques). Cependant, malgré leur importance dans l'observation morphologique de l'appareil phonatoire, ils sont mal adaptés à l'étude du fonctionnement laryngé dû à difficulté d'intubation pour un examen visuel pendant le processus de phonation. En plus, lors de l'utilisation de ces instruments, quelques locuteurs manifestent une action réflexe dans la cavité supra-glottique qui peut donner une fausse évaluation [3].

Dans ce contexte, il existe un besoin pour les chercheurs en traitement et reconnaissance de la voix pathologique de développer de nouveaux outils objectifs pour soutenir l'analyse clinique. De tels outils pourraient fournir une évaluation et détection automatique des perturbations de la production vocale en utilisant uniquement le signal vocal, sans avoir besoin de dispositifs spéciaux et en utilisant des connaissances extraites de la base de données existante. De ce fait l'évaluation et la détection automatique des pathologies vocales peuvent être considérées comme un dépistage primaire d'évaluation objective de la qualité fonctionnelle de la production de la voix.

Dans la littérature, plusieurs techniques de diagnostic et d'évaluation automatique de la voix pathologiques plus ou moins sophistiquées ont été proposées [4]. Ces techniques peuvent être appliquées directement au signal vocal tel que les paramètres acoustiques (Fréquence fondamentale, jitter, Shimmer...etc), soit sur une estimation adéquate du signal glottique. Ce signal obtenu juste après les cordes vocales et avant le conduit vocal. Il décrit le

mouvement vibratoire des cordes vocales, donc l'extraction de certains paramètres extraits du signal glottique peut aider au diagnostic des pathologies des cordes vocales mieux que celles extraites du signal vocale [5]. Cependant, les paramètres acoustiques ont fait l'objet de très nombreux travaux et ne présentent pas de méthodologies innovantes dans le domaine du diagnostiques de la voix pathologique. C'est pourquoi, durant nos travaux, nous nous sommes intéressés à une nouvelle méthode de traitement du signal.

Le diagramme de récurrence ou Récurrents Plot (RP) en anglais, est une technique graphique avancée pour l'analyse de données non linéaires basée sur la théorie du Chaos [6, 7], a été sujet à de nombreuses recherches dans le domaine médical [8, 9], compte tenu des signaux étudiés dans ce domaine (signaux cardiaques, neurologiques ou cérébraux) et qui présentent souvent des dynamiques de types transitoires, non-périodiques et intermittents [10]. Cependant, cette méthode est considérée comme une analyse qualitative, car elle n'offre pas de paramètres permettant de quantifier la dynamique du signal étudié. C'est pourquoi, une méthode complémentaire à la RP, appelée « Recurrence Quantification Analysis ou RQA, permettant de quantifier cette dynamique à travers de nombreux indicateurs, est introduite. Ces derniers sont utilisés comme des entrées à diverses méthodes de classification [11]. Dans le domaine du traitement automatique de la voix pathologique, l'analyse quantitative de récurrence appliquée au signal vocal a connu un rendement de plus de 90 % pour la discrimination entre les patients avec ou sans troubles laryngés [12]. Les paramètres RQA ont été proposés comme un vecteur d'entrée pour un classificateur Linear Discriminate Analysis (LDA) et Quadratique Discriminate Analysis (QDA). Les résultats montrent une discrimination importante dans la possibilité de distinguer les voix saines des voix pathologiques en vue de la classification des pathologies laryngales [13]. En analysant les résultats des travaux réalisés dans le contexte de notre sujet, nous pouvons conclure que la l'analyse quantitative de récurrence et l'analyse quantitative de récurrence croisé n'ont pas été étudiée d'une manière approfondie dans le domaine de la classification automatique de la voix pathologique. Cependant, l'application de RQA à un signal décrivant un phénomène physiologique bien spécifié comme le signal glottique (vibration des cordes vocales) présenterait une forte capacité de discrimination des pathologies liées à cet organe.

Notre travail se porte sur l'exploitation des informations fournies par le tracé du diagramme de Récurrence (Recurrence Plots) et l'analyse quantitative de récurrence pour la discrimination des pathologiques des cordes vocales. Les deux principaux objectifs de ce travail sont : proposer un système de diagnostic et de classification des voix pathologiques en appliquant des techniques d'Intelligence Artificielle (IA). Deuxièmement, nous proposons une méthode d'évaluation automatique pour estimer le degré de gravité des maladies de la voix avant et après une rééducation orthophonique.

En premier lieu, les informations obtenues à partir du tracé du diagramme de (RP) et l'Analyse Quantitative Récurrence (RQA), faisant référence à la forme d'onde du signal glottique, sont explorées pour la détection et la classification des troubles de la voix humaine. Pour cela, premièrement le signal glottique a été extrait premièrement du signal vocal à l'aide

d'un algorithme de filtrage inverse : Iterative Adaptive Inverse Filtering (IAIF). Le tracé RP et le RQA ont été appliqués au signal glottique pour évaluer le processus dynamique des Cordes Vocales (CVs). Huit (08) paramètres RQA ont été utilisés comme un vecteur de caractéristiques d'entrée d'un algorithme Particle Swarm Optimization-Support Vector Machines (PSO-SVM) afin séparer les voix normales et pathologiques.

Deuxièmement, nous proposons un modèle paramétrique basé sur l'Analyse par Quantification Réurrence Croisée (CRQA) pour montrer comment que cette méthode peut être adaptée pour comparer, évaluer et quantifier les changements d'articulation des plis vocaux après une thérapie ou une rééducation (orthophonique). Huit (08) mesures CRQA sont extraites pour évaluer le degré d'amélioration de la qualité vocale. En comparant la variation de ces paramètres avant et après la thérapie par apport aux voix normales, les stades pathologiques ont été facilement discriminés.

Pour la validation des tests, la Base de Données Saarbrücken Voice Database (SVD) a été adoptée, dans laquelle la voyelle longue [a:] a été sélectionnée pour les échantillons prononcés par quatre groupes de sujets : personnes souffrant de la paralysie des CVs, personnes ayant des polypes de CVs, personnes ayant la maladie de la dysphonie spasmodique et personnes possédant une voix normale considérée comme référence. Pour garantir la validité des résultats et les capacités de généralisation, la méthode K-Folds Cross Validation a été utilisée. Les résultats obtenus montrent l'efficacité du RQA appliqué au signal glottique comme technique d'extraction des caractéristiques. En effet, le PSO-SVM, en tant que méthode de classification, semble un outil efficace pour le diagnostic et la classification des voix pathologiques.

Cette thèse est organisée en quatre (04) chapitres :

Le premier, expose une présentation du principe de production de la parole et quelques généralités sur les caractéristiques physico-acoustiques de la parole. Nous avons donné un aperçu sur la modélisation et les techniques d'extraction de l'onde glottique ainsi que sur les paramètres articulatoires. Le deuxième chapitre consiste en une description des différentes pathologies abordées affectant les cordes vocales. Ensuite nous présentons les méthodes les plus importantes de diagnostic de la voix pathologiques basées sur les techniques de traitement de signal. Nous attacherons une importance particulière sur l'approche proposée. Le troisième chapitre est consacré à une description mathématique des techniques de classification et optimisation introduites dans cette thèse. Le chapitre quatre (04) détaillera les démarches expérimentales et les résultats de classification et évaluation des pathologies des cordes vocales. Finalement, ce travail se clôturera par des conclusions et des réflexions sur l'approche proposée, permettront la proposition de perspectives sur l'utilisation des outils de traitement automatique de la parole pour l'évaluation et la classification de la voix pathologique.

***Chapitre 1 :***  
***Généralités Sur la Production de la Parole***

## 1.1. Introduction

La parole est le moyen de transmettre nos messages, nos pensées, nos émotions et nos sentiments. Il est aussi l'image sur notre santé et elle peut aussi être porteuse d'informations sur notre situation sociale. D'autre part, la voix est le support acoustique de la parole, Une perturbation ou une lésion de la voix peut donner lieu à des conséquences négatives sur notre vie quotidienne ce qui nécessite une consultation chez un spécialiste de la voix.

La parole est produite suite à des phénomènes physiologiques propres à l'être humain qui se produisent en partant des poumons qui représentent la source d'énergie produisant la voix jusqu'au l'extérieur. De point de vue acoustique, les sons voisés de la parole sont produits par la vibration des cordes vocales qui peut être mesurée à l'aide d'un microphone et permet de représenter la parole sous la forme d'un signal acoustique qu'on peut analyser en utilisant des outils de traitement du signal.

Depuis quelques décennies, l'analyse de la voix et de parole pour la caractérisation et l'évaluation des troubles vocales a fait l'objet d'un grand nombre de travaux de recherche. L'objectif de l'analyse du signal acoustique est d'extraire des indices pertinents permettant de déterminer les caractéristiques de la voix afin de renseigner sur l'état des cordes vocales. Dans ce chapitre nous présentons le principe de production de la parole, commençant par une présentation de l'anatomie et la physiologie du système de production de la voix. Ensuite nous abordons la description des différents modèles de production de la parole et la modélisation de l'onde glottique. Cette dernière caractérise les vibrations des cordes vocales et donne une image concrète de son comportement sous la présence d'une pathologie.

## 1.2. Anatomie de l'appareil phonatoire humain

L'appareil phonatoire représente l'ensemble des organes qui permettent de produire les sons constituant la voix. L'évolution rapide de la médecine et les techniques d'investigation ont permis de mieux comprendre comment le son était généré avant d'être rayonné dans le milieu extérieur. Comme indique la figure 1.1, Les sons de la parole sont produits par la modification du courant d'air de l'expiration en provenance des deux poumons [14]. Ce souffle d'air se transforme, selon la configuration du conduit vocal, lors de son passage dans le larynx, le pharynx et les cavités buccale et nasale (Figure 1.1) [15][16]. L'appareil phonatoire peut être divisée en trois principales partie :

- La partie sous glottique comprend le diaphragme, poumons et la trachée.
- La partie glottique (larynx avec les cordes vocales).
- La partie supra glottique (pharynx et cavités buccale, nasale et Labiale).

### 1.2.1 Partie sous glottique

Comme indique la figure 1, Les poumons peuvent être considérés comme des sacs pour le stockage de l'air représentent la source d'énergie produisant la voix. Ils jouent un rôle majeur dans les échanges d'oxygène ( $O_2$ ) et de dioxyde de carbone ( $CO_2$ ) entre l'air et le sang pendant la respiration. Les poumons peuvent se gonfler ou se dégonfler en augmentant ou en

diminuant le volume de la cavité thoracique, grâce à l'action du diaphragme et les muscles thoraciques. La trachée est un conduit constitué de tissu fibreux et cartilagineux reliant le larynx en haut aux bronches principales gauche et droite en bas [15].

L'inspiration se caractérise par une extension de la cage thoracique associée à une dilatation du poumon, celui-ci étant solidaire des parois de la cage thoracique grâce aux deux feuillets de la plèvre. L'expiration se produit par un recul élastique passif de la cage thoracique et une relaxation du diaphragme, induisant une décroissance du volume de la cavité thoracique et des poumons. Cela correspond à une augmentation de la pression dans les poumons forçant l'air à sortir. Pendant la respiration Normale, le temps pris pour les deux phases inspiration et expiration forment un processus rythmique comprenant de 12 à 15 cycles par minute [16].

En cas de production de la parole, la situation est très différente. La phase d'inspiration est plus rapide que dans la respiration normale et plus profonde en termes de volume d'air inspiré. La phase d'expiration (au cours de laquelle la parole est produite) est par contre plus longue. Pendant cette phase, le flux d'air provenant des poumons se transforme en une série d'impulsions par constriction au niveau des cordes vocales, cette série d'impulsion sera modulée en passant par le conduit vocal produisant un son ou parole.

### 1.2.2 Partie glottique

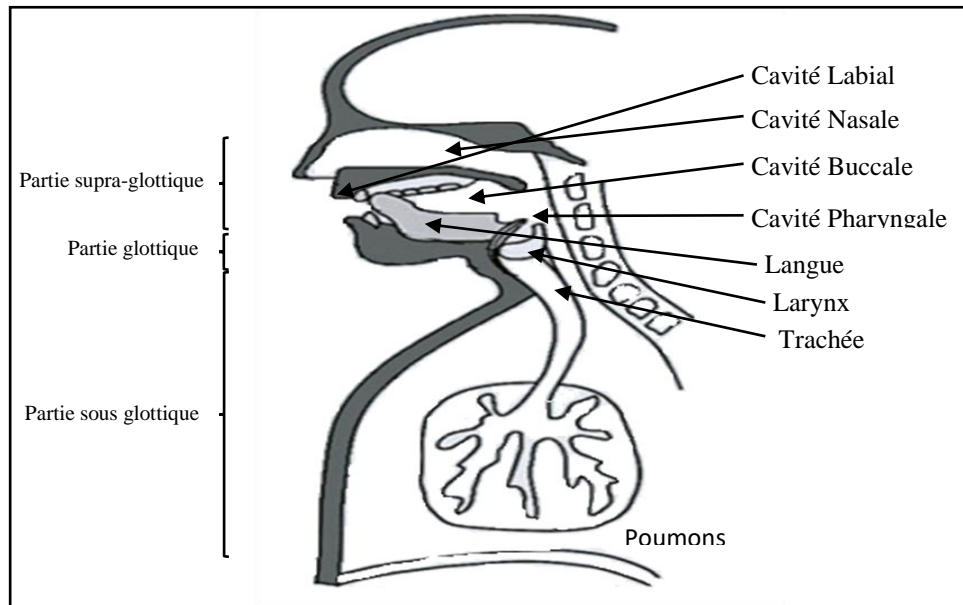
Elle contient le larynx, qui constitue la partie articulaire de l'appareil phonatoire, se situe à l'extrémité supérieure de la trachée et le bas de la langue. C'est un élément essentiel non seulement sert à la protection de la voix aérienne (la trachée) contre les aliments au cours de la déglutition mais aussi dans la production de la parole. Il contient les cordes vocales et des muscles modifiant l'ouverture et la fermeture des plis vocaux produisant les sons voisés. Il se compose de pièces cartilagineuses multiples, mobiles, entre lesquelles sont tendus des replis membraneux, les cordes vocales [16].

Les cartilages constituant le larynx s'articulent en bas à la trachée et se rattachent en haut, par l'intermédiaire de l'os hyoïde, au plancher buccal. Directement en rapport avec les téguments du cou (pomme d'Adam), le larynx s'ouvre en arrière dans le pharynx, juste avant l'œsophage. C'est donc l'un des éléments essentiels du carrefour aérodigestif où confluent voies respiratoires et voies digestives.

L'épiglotte est une lame fibro-cartilagineuse triangulaire située sur la ligne médiane en avant de l'orifice supérieur du larynx qu'elle ferme pendant les mouvements de déglutition. Son sommet s'attache à l'angle rentrant du cartilage thyroïde, au-dessus des cordes vocales, au moyen du ligament thyro-épiglottique. Sa face antérieure regarde la base de la langue, sa moitié supérieure est libre et présente des replis glosso-épiglottiques au nombre de trois, constitués par la muqueuse qui passe de la langue à l'épiglotte, entre ces replis se trouvent les fossettes glossoépiglottiques ou vallécules.

Les cordes vocales sont situées à environ 8mm du bord inférieur du cartilage thyroïde en dessous de l'épiglotte elles s'étendent de l'avant vers l'arrière, et forment une structure en V pointé vers l'avant comme indique la figure 1.2. À l'avant, les cordes vocales se fixent sur le

cartilage thyroïde du larynx. À l'arrière, les cordes vocales se fixent sur le cartilage aryténoïde, au niveau de l'apophyse vocale [17].



**Figure 1.1** : Anatomie de l'appareil phonatoire [17]

### 1.2.3 Partie supra-glottique

La partie supra-glottique (appelée aussi le conduit vocale) comprend les cavités : buccale, nasale, labiale et le pharynx. La cavité buccale est délimitée par les lèvres, le palais, la langue. (Marquant la jonction entre la bouche et l'oropharynx). Les cavités nasales sont le point d'arrivée de l'air dans le système respiratoire et sont responsables du réchauffement et de l'humidification de l'air. Le pharynx est un tube musculaire attaché à la base du crâne et continu avec l'œsophage. Il peut être divisé en trois parties : le nasopharynx (derrière les cavités nasales), l'oropharynx (derrière la bouche) et le laryngopharynx (derrière et autour de l'entrée laryngée) [18].

## 1.3 Mécanisme de production de la Parole

La production de la parole implique un mouvement coordonné d'un certain nombre de structures dans la tête, le cou et le thorax en d'autres termes, les parties anatomiques décrites ci-dessus. Trois étapes distinctes sont nécessaires à la production de la parole :

- la génération d'un flux d'air (expiration) provient des poumons
- la conversion de ce flux d'air dans une série de vibrations (la phonation)
- et la modification de ces vibrations (l'articulation)

### 1.3.1 Vibration des cordes vocales

La vibration des cordes vocales est un phénomène qui se produit pendant l'expiration d'air expulsé des poumons à travers la trachée. Si les cordes vocales sont maintenues fermées par l'action des muscles laryngés intrinsèques la pression sous-glottique augmente. Lorsque cette pression est suffisante pour surmonter la force musculaire, les cordes vocales sont forcées de

s'ouvrir un peu et une petite quantité d'air est libérée dans le conduit vocal, entraînant une diminution de la pression en dessous des plis vocaux. Comme l'action musculaire dépasse la force produite par force sous-glottique, les cordes vocales se ferment, sous l'action du phénomène de Bernoulli et par leur recul élastique naturel. Ce cycle se répète à une certaine fréquence qui représente la fréquence fondamentale, jusqu'à ce que les cordes vocales soient détendues ou qu'il n'y ait plus de flux d'air [19].

La structure des cordes vocales est plus ou moins compliquée. En effet, différentes parties des plis vocaux vibrent de différentes manières pendant le cycle de vibration. Ceci est en partie responsable de la présence d'harmoniques de la fréquence fondamentale dans le spectre du signal glottique. Le changement des propriétés physiques des cordes vocales comme par exemple la présence d'une pathologie provoque un changement dans les caractéristiques articulatoire agissant la qualité de la voix.

- **Son voisé**

Les sons voisés tels que les voyelles par exemple, sont produits par le passage de l'air des poumons à travers la trachée, qui met en vibration les cordes vocales. Pour les sons voisés, l'excitation possède un caractère périodique et des propriétés particulières dues à la forme de l'onde de débit glottique. Les sons voisés sont généralement quasi-périodiques. Ce type de sons représente la majorité du temps de phonation, et est caractérisé en général par une énergie élevée en basse fréquence avec environ un formant par kHz de bande passante, et dont seuls les trois ou quatre premiers contribuent de façon importante à l'information linguistique [20].

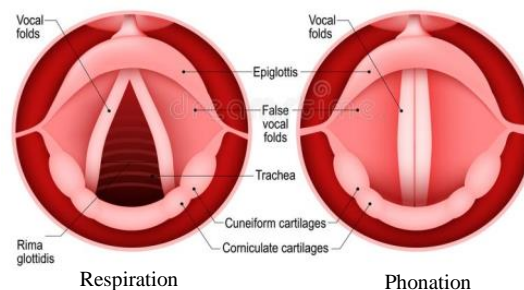
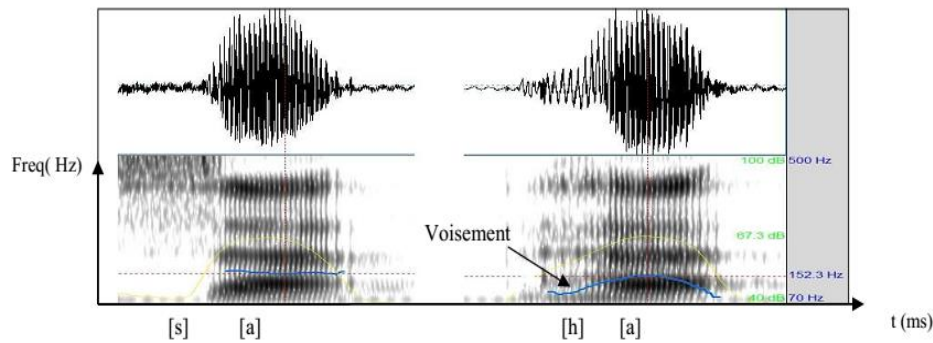


Figure 1.2 : structure des cordes vocales

- **Son non voisé**

Les sons non voisé n'utilise pas les cordes vocale par conséquent elles sont étendues permettant à l'air de circuler librement des poumons au conduit vocal modulant le son. Le signal correspondant présente une structure aperiodique. Le son non-voisé est souvent considéré comme un bruit blanc, L'énergie de ce type de sons est concentrée dans les hautes fréquences et correspond à du bruit.





**Figure 1.3** : Spectrogramme des fricatives [s] / [h] en contexte vocalique [a]

Comme le montre la figure 1.3, le spectrogramme des parties voisées du signal apparaissant sous forme de successions de pics spectraux denses en énergie sur lesquels on a superposé les courbes des trois premiers formants (F1, F2 et F3), dont les fréquences centrales ne sont pas forcément des multiples de la fréquence fondamentale. Par contre, le spectre d'un signal non voisé ne présente aucune structure particulière absence de bande d'énergie indiqué par une partie sourde correspond au son non voisé.

### 1.3.2 Contrôle de fréquence

La fréquence fondamentale varie selon l'âge, en raison de dimension l'anatomie de l'appareil phonatoire qui se diffère en longueur, d'épaisseur et de tension des cordes vocales. Des variations autour de cette fréquence peuvent être faites en ajustant les propriétés physiques des cordes vocales. En agissant sur la position des cartilages aryténoïdes cela modifie la longueur des cordes vocales, ce changement en longueur est accompagné d'un changement d'épaisseur et une croissance de tension. La tension peut également être modifiée indépendamment des autres variables physiques en contractant les muscles thyroaryténoïde. Sachant cela, une longueur croissante des cordes vocales s'accompagne d'une épaisseur décroissante et d'une tension interne croissante, ces trois effets induisant une fréquence fondamentale plus élevée.

### 1.3.3 Contrôle de l'intensité vocale

L'intensité de la voix peut être modifiée en agissant sur la pression sous-glottique, sur la portion du cycle de phonation pendant laquelle la glotte est ouverte et sur la forme de l'appareil vocal. Les deux premiers mécanismes sont étroitement liés au comportement des cordes vocales lors de la phonation. Cela induit une croissance de l'intensité de la voix et souvent liée à une augmentation de la fréquence fondamentale. Le troisième effet consiste à augmenter l'intensité vocale en ajustant la forme de l'appareil vocal pour que ses fréquences de résonance (les formants) correspondent aux harmoniques de la fréquence fondamentale.

### 1.3.4 Articulation

En fonction de l'état de l'appareil phonatoire les natures des sons produits peut être déterminée. Dans le cas de phonation les cordes vocales sont tendues et vibrante à cause du flux d'air. Le flux d'air est découpé en un train d'impulsions quasi périodique qui résonne

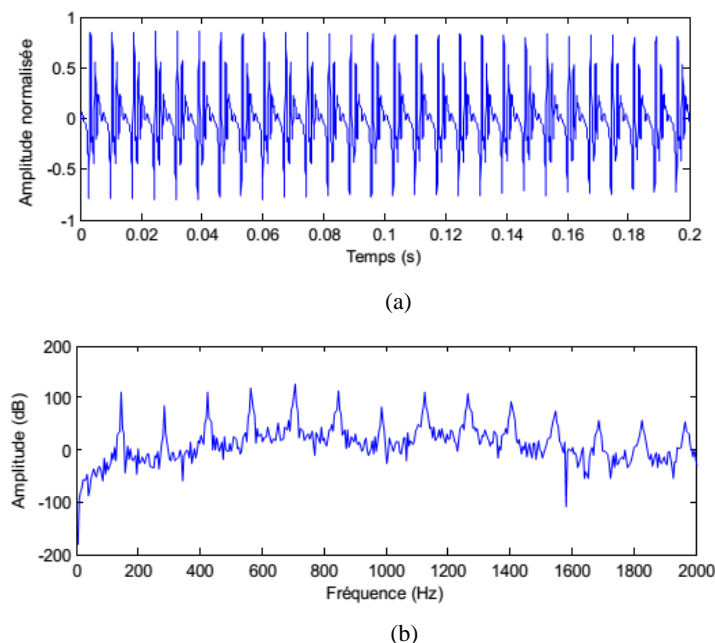
dans les différentes cavités. Physiquement, le train d'impulsion quasi périodique subit une modulation en fréquence, on obtient donc un son voisé. En passant par les différentes cavités changeant le spectre du signal source par l'effet du conduit vocal. Lorsque les cordes vocales sont relâchées, l'air passe librement au niveau du larynx sans les faire vibrer. On obtient alors un son non voisé.

## 1.4 Voyelles et consonnes

Le phonème représente la plus petite unité de parole : il s'agit soit d'une voyelle, soit d'une consonne. Le nombre de phonèmes est toujours limité, chaque phonème se différencie de l'autre par le mode et le lieu d'articulation. Par exemple : la langue française comprend 36 phonèmes et la langue arabe comporte 40 phonèmes (28 consonnes, 6 voyelles (3 longues et 3 courtes voyelles)).

### 1.4.1. Voyelles

Les voyelles sont produites lorsque le conduit vocal est ouvert et que le son est voisé. Elles sont orales ou nasales selon que la cavité nasale n'est pas ou est mise en parallèle à la cavité buccale. Ils sont classés selon : la nasalité, l'ouverture du conduit vocal, la position de la constriction du conduit vocal et l'arrondissement des lèvres. Les voyelles se différencient principalement les unes des autres par leur lieu d'articulation, leur ouverture, et leur nasalisation [21]. La figure 1.4 ci-dessous montre le signal et spectre de la voyelle longue [a:].



**Figure 1.4 :** La forme temporelle (a) et spectrale (b) de la voyelle longue [a:] prononcée par un locuteur masculin

### 1.4.2. Consonnes

Les consonnes sont produites lorsqu'un rétrécissement apparaît dans l'appareil phonatoire due au passage de l'air dans le conduit vocal d'où la naissance d'un bruit : une consonne est

produite. Dans le cas de la langue arabe elles se différencient en fonction de leur mode d'articulation, de leur lieu d'articulation, et de leur nasalisation. Comme pour les voyelles, d'autres critères de différenciation peuvent être nécessaires dans un contexte plus général : l'organe articulaire, la source sonore, l'intensité, l'aspiration, la palatalisation, et la direction du mouvement de l'air. Les consonnes peuvent être classées en trois sous classes différentes ayant des caractéristiques distinctes les fricatives, les occlusives et les sonantes [21]. De plus, si la prononciation d'une consonne s'accompagne d'une résonance du larynx, elle est considérée comme consonne sonore, ou sourde dans le cas contraire.

### 1.5 Modélisation de l'appareil phonatoire

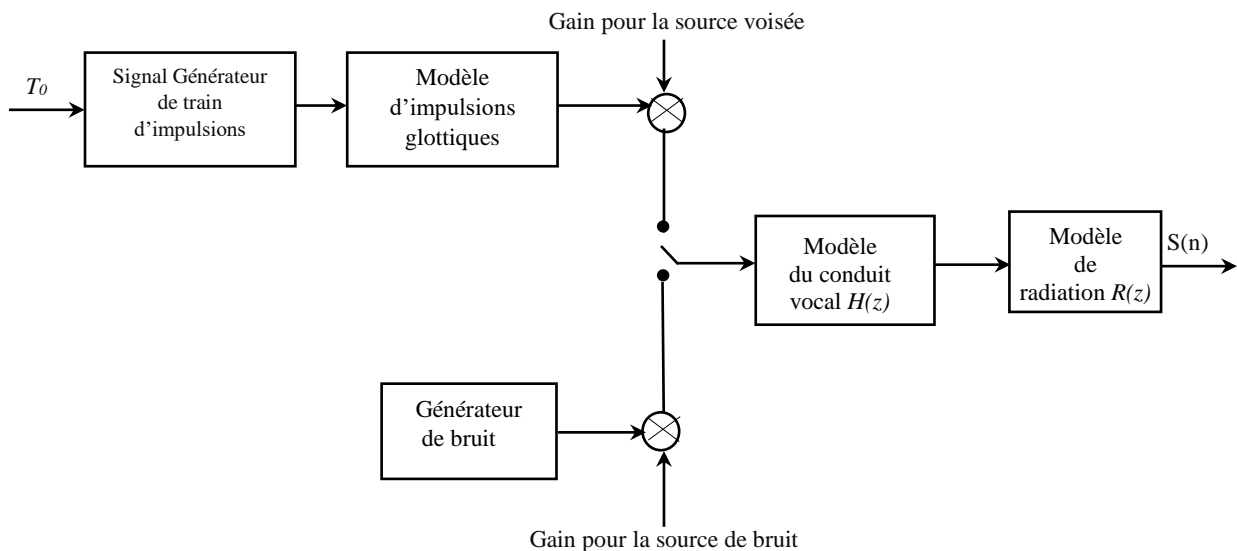
La modélisation du processus de production de la parole peut être représentée par le modèle source-filtre présenté dans la figure ci-dessous. Le signal de parole peut être modélisé comme la sortie d'un filtre linéaire variant dans le temps, qui simule les caractéristiques spectrales de la fonction de transfert du conduit vocal, excité par un signal source qui reflète l'activité des cordes vocales dans les zones voisées et le bruit de friction dans les zones non voisées [22], [23] Figure 1.5, avec :

$H(z)$  : filtre qui modélise le conduit vocal et les articulateurs ;

$R(z)$  : filtre qui simule le rayonnement aux lèvres ;

$G(z)$  : impulsion glottique ;

$T_0$  : la période correspond à la fréquence fondamentale de l'excitation ;



**Figure 1.5.** Modélisation de l'appareil phonatoire Source-filtre

La fonction de transfert exprimant le conduit vocal peut être donnée par la forme :

$$H(z) = \frac{H_0}{\prod_{k=1}^p (1 - p_k z^{-1})} \quad (1.1)$$

Où  $H_0$  : représente un gain et  $p_k, k = 1, \dots, p$  sont les pôles de la fonction de transfert.

La fonction de transfert modélisant le filtre de rayonnement aux lèvres est donnée par :

$$R(z) = 1 - z_0 z^{-1} \quad |z_0| < 1 \tag{1.2}$$

Dans le cas où la parole est non voisée, la sortie  $S(z)$  du modèle est définie par :

$$S(z) = E(z) H(z) R(z) \tag{1.3}$$

Où  $E(z)$  représente la transformée en  $z$  d'une réalisation du bruit blanc.

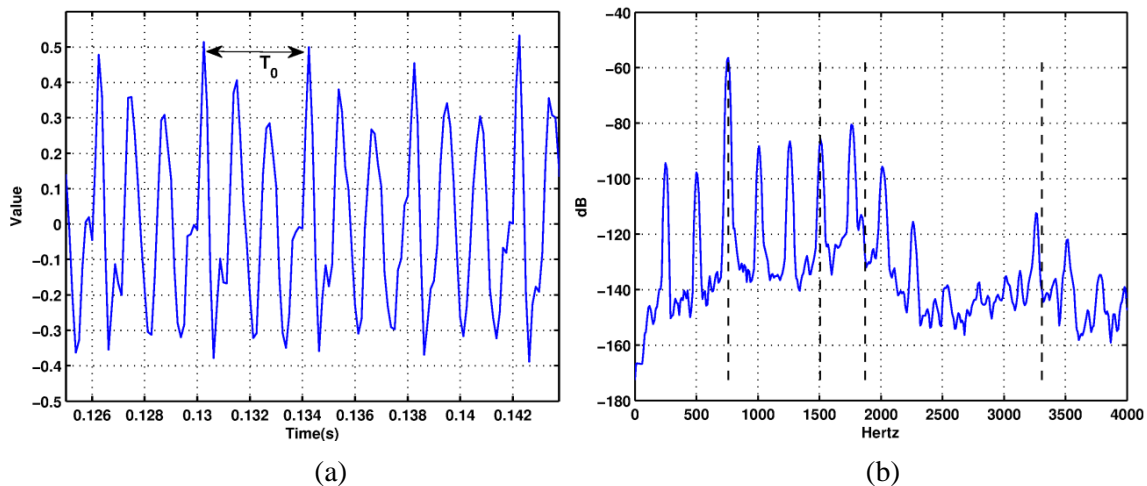
Dans le cas de la voix voisée, la sortie  $S(z)$  est définie comme suit :

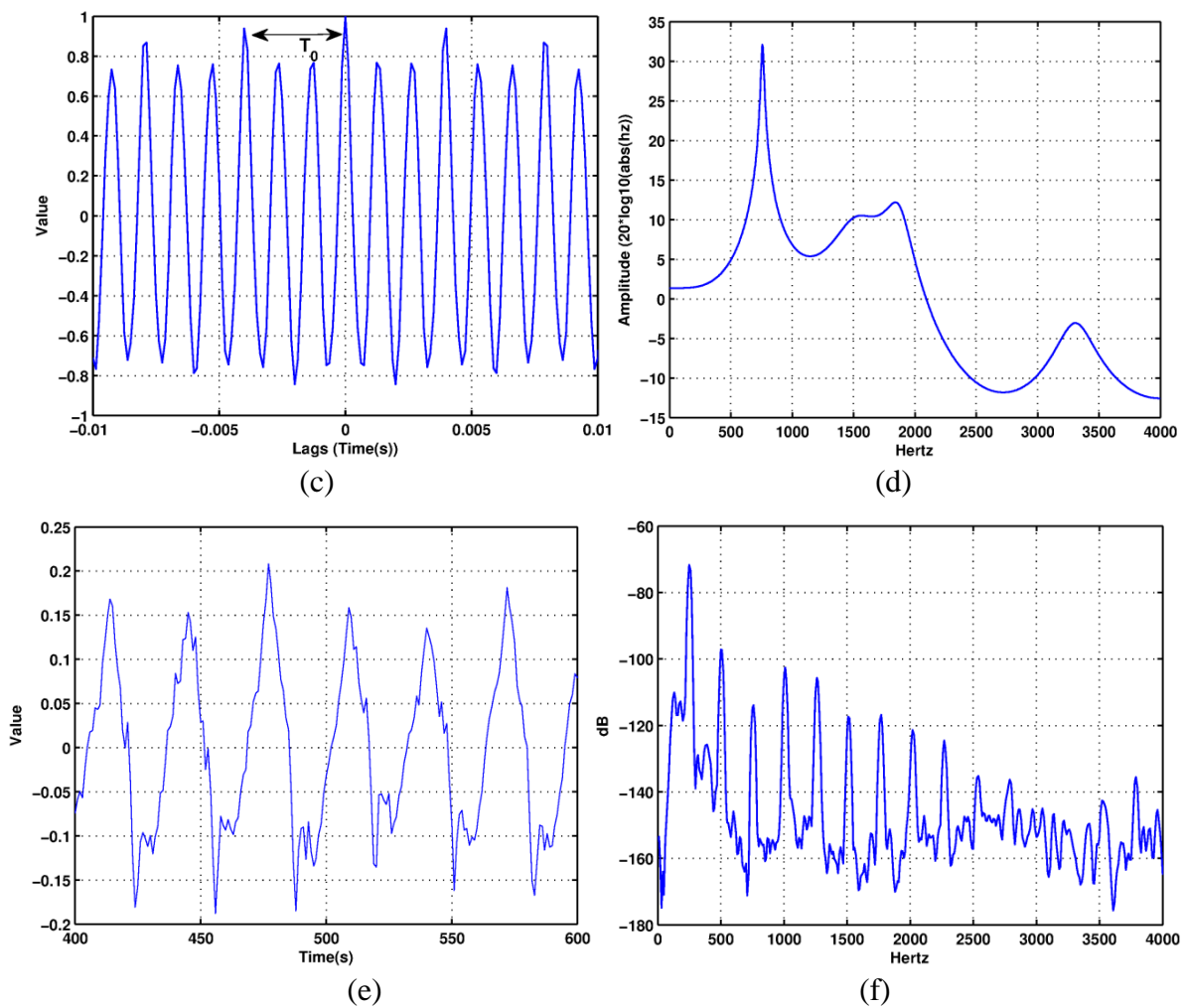
$$S(z) = E(z) G(z) H(z) R(z) \tag{1.4}$$

Où  $E(z)$  représente la transformée en  $z$  du train d'impulsions caractérisé par une période égale à la période correspond à la fréquence fondamentale  $T_0$  du signal de parole. La fonction de transfert du système globale  $Y(z)$  du modèle est définie par :

$$Y(z) = \begin{cases} H(z)R(z) & \text{Cas où les sons sont voisés} \\ G(z)H(z)R(z) & \text{Cas où les sons sont non voisés} \end{cases} \tag{1.5}$$

La fréquence fondamentale  $F_0$ , ou pitch, de la forme d'onde du signal de la parole  $s(n)$  peut être calculée à calculé en utilisant la méthode d'autocorrélation dont les résultats sont montrés à la figure 1.6. En se référant aux figures 6a, 6c, 6e, il est facile de démontrer comment la forme d'onde du source glottique et la forme d'onde acoustique de signal de parole ont la même période correspond à la fréquence fondamentale  $T_0$ , ce qui est compréhensible puisque la forme d'onde acoustique de sortie est finalement créée par la source glottique. La figure 6b montre le spectre de  $s(n)$  avec ses fréquences de formants estimées (lignes pointillées) en rapport avec le filtre  $H(z)$  du conduit vocal. Le spectre de  $H(z)$  est illustré à la figure 6d et présente des pics correspondant aux fréquences de formants illustrées à la figure 6b.





**Figure 1.6 :** Signaux vocaux, signal glottique et conduit vocal pendant la phonation : (a) : signal de parole  $S(n)$ , (b) : spectre de signal de parole, (c) : autocorrélation du signal, (d) spectre du filtre conduit vocal, (e) : forme du signal glottique, (f) : spectre du signal glottique.

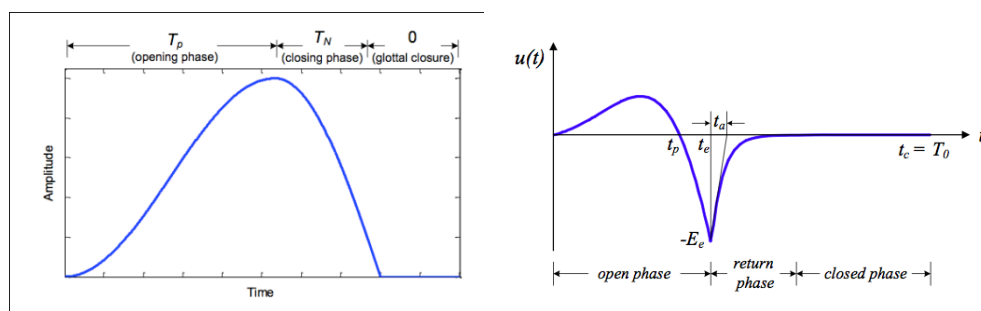
## 1.6 Modèle de la source glottique

Selon l'anatomie et la physiologie de la production de la parole, l'onde glottique représente l'écoulement du flux d'air qui traverse la glotte à travers le conduit vocal. Plusieurs recherches sont établies dans ce contexte dans l'objectif de modéliser la source glottique tel que : le modèle de Lijlencrants Fant (modèle LF) [24], le modèle de Klatt [25], le modèle de Rosenberg [26] et le modèle R++ [27]. D'autres modèles ont été proposés pour déterminer les paramètres biomécaniques des cordes vocales ou pour la synthèse des troubles de la voix [28].

Les différents modèles glottiques existants n'utilisent pas le même nombre de paramètres ou le même nom pour des paramètres similaires, ce qui rend difficile la compréhension des différences et des similitudes entre les modèles, tous partagent des caractéristiques communes et des expressions analytiques qui se différencient d'un modèle à un autre [29] Tel que :

- le flux glottal est toujours positif ou nul ;
- le signal glottique est un signal quasi-périodique ;
- le signal glottique est une fonction continue du temps ;
- Dans une seule période, le signal glottique a une forme de cloche: il commence à augmenter puis à diminuer et enfin nul ;
- le signal glottique est une fonction différentiable du temps, sauf dans certains instants comme l'instant de fermeture (GCI) et l'instant d'ouverture de la glotte (GOI) ;
- la phase d'ouverture de la glotte est plus longue que celle de fermeture ;

Malgré le fait que la plupart des modèles du signal glottique considèrent ces hypothèses, précédemment mentionné, des études ont montré que la majorité des individus présente une fuite glottale pendant la phase de fermeture, par conséquent, le signal glottique n'est pas complètement nul pendant cette phase. D'après les propriétés générales du signal glottique exprimées ci-dessus, les modèles existants utilisent principalement un ensemble d'instant temporels décrivant l'articulation des cordes vocales, comme le montre la figure 1.7.



**Figure 1.7** : Schéma principal du signal glottique utilisé par la plupart des modèles.

Où :

- $t_p$  : temps du maximum de l'impulsion. Ce maximum est appelé l'amplitude de la voix ;
- $t_e$  : temps du minimum de la dérivée du temps ;
- $t_c$  : la période correspond à la fréquence fondamentale ( $t_c = T_0 = 1/f_0$ ) ;
- $t_a$  : la durée de la phase de retour ;

Le modèle de l'onde glottique peut être représenté en deux phases qui sont la phase ouverte et la phase fermée comme la montre la figure 1.7. La phase ouverte est divisée elle-même en deux phases : la phase d'ouverture et la phase de fermeture. La phase d'ouverture est définie par une augmentation du débit glottique de l'état initial à l'instant  $T_p$  définissant l'instant d'ouverture. La phase de fermeture est décrite par la diminution de l'amplitude du signal glottique au moment  $T_n$  définissant l'instant de fermeture où la dérivée atteint son minimum négatif  $E_e$  ( $E_e$  : excitation maximale).

Pendant la phase de fermeture, le débit glottique retourne à l'état initial après une fermeture brutale  $T_a$  où la dérivée du débit glottique atteint 0 après une excitation maximale. En

pratique, il convient de mettre  $t_c = T_0$ , ce qui représente la période correspond à la fréquence fondamentale.

Plusieurs recherches sont entamées dans le contexte de modélisation de la source glottique qui se différencient principalement en nombre de paramètres décrivant le modèle. Le modèle proposé par Liljencrants and Fant [30] connu comme le modèle LF est l'un des modèles les plus utilisés pour la modélisation du signal glottique. C'est un modèle à cinq paramètres associé au long du cycle glottique : la période correspond à la fréquence fondamentale  $t_0$ , l'amplitude du minimum de la dérivée de la source glottique  $E_0$ , l'instant de fermeture où la dérivée atteint son minimum négatif  $t_e$ , l'instant du maximum de la source glottique  $t_p$  et Temps de retour à l'état initial  $t_a$ .

Le modèle LF est décrit par l'équation suivante :

$$g'_{LF}(t) = \begin{cases} E_0 e^{\alpha t} \sin(w_g t) & , 0 \leq t \leq t_e \\ -\frac{E_e}{\beta t_a} (e^{-\beta(t-t_a)} - e^{-\beta(t_c-t_a)}) & , t_e < t \leq t_c = T_0 \end{cases} \quad (1.6)$$

Où :

- $E_e$  : est l'amplitude du maximum négatif ;
- $\alpha = C\pi$ , ou C contrôle la sinusoïde en croissance exponentielle ;
- $w_g = 2\pi F_g$  représente la fréquence de la sinusoïde et  $F_g = \frac{1}{2t_p}$  ;
- $\beta$  est une constante de décroissance pour la phase de récupération de l'exponentielle ;

Les deux paramètres  $\alpha$  et  $\beta$  peuvent être calculés de l'équation (1.6) en imposant :

$$g'_{LF}(t_e) = E_e \quad (1.7)$$

Et le bilan énergétique :

$$\int_0^{T_0} g'_{LF}(t) dt = 0 \quad (1.8)$$

Le modèle LF est donc une fonction composée de deux parties. La première partie modélise la dérivée de l'écoulement de l'onde glottique à partir de l'instant d'ouverture glottique jusqu'à l'instant extrême du maximum négatif et correspond à la phase d'ouverture. Le deuxième segment caractérise la phase de fermeture. En plus de ces deux équations, le modèle est régi par le principe d'équilibre dans une phase d'ouverture-fermeture, ce qui signifie que l'intégrale de la fonction sur toute la période doit être égale à zéro.

Dans la figure 1.7, où la forme d'une onde glottique typique et sa dérivée est représentée selon le modèle LF, il est évident que les instants de synchronisation du modèle LF ont une correspondance avec le comportement de l'articulation des cordes vocales. Ce modèle semble être le modèle préféré par de nombreux chercheurs en raison de sa capacité à s'adapter aux variations naturelles de vibration des cordes vocales [31]. En outre, plusieurs études ont montré que le modèle LF était mieux que d'autres modèles surtout lorsque l'objectif est de modéliser le mécanisme de production de la parole naturelle. Cependant, son utilisation dans les synthétiseurs vocaux est limitée en raison de sa complexité de calcul, car elle implique la résolution d'une équation non linéaire.

### 1.7 Techniques d'extraction de l'onde glottique

L'estimation des paramètres articulatoire propres à la source glottique nécessite une estimation du signal glottique. Plusieurs méthodes ont été proposées pour l'estimation de la source glottique et du filtre du conduit vocal à partir du signal vocal. Cependant, ce problème n'est pas facile à résoudre car il est difficile de séparer efficacement la source glottique du conduit vocal. Les paramètres de la glotte peuvent également être mesurés à partir d'autres signaux obtenus lors du processus de production de la parole à l'aide des appareils médicaux plus ou moins sophistiqués, comme le signal EGG (**E**lectro-**G**lottal-**G**ramme) à l'aide d'un glottographe. Mais ces méthodes procèdent des inconvénients, ils sont très coûteuses en temps et en ressources humaines pour ces raisons en plus les dimensions des conduits vocales des malades diffèrent d'un patient à un autre ce qui amène une inaccessibilité ou difficulté de l'introduction de la sonde glottographique au lieu désiré. Pour cela l'estimation du signal glottique directement à partir du signal de parole peut être considérée comme un outil de dépistage primaire pour l'analyse du signal de la parole. Les sections suivantes donnent un aperçu des principales méthodes d'estimation du signal source signal glottique à partir du signal de parole.

La plupart des techniques d'estimation du signal glottique reposent sur un algorithme de filtrage inverse. Ces méthodes estiment d'abord un modèle paramétrique du conduit vocal, puis en supprimant la contribution du conduit vocal via un filtrage inverse permettent d'obtenir le signal glottique. Les méthodes de cette catégorie diffèrent par l'estimation du conduit vocal. Dans la section Filtrage Inverse en phase de fermeture, cette estimation est calculée pendant la phase de fermeture de la glotte, alors que dans l'autre technique, une procédure itérative et / ou adaptative est utilisée.

#### 1.7.1 Filtrage inverse en phase fermeture

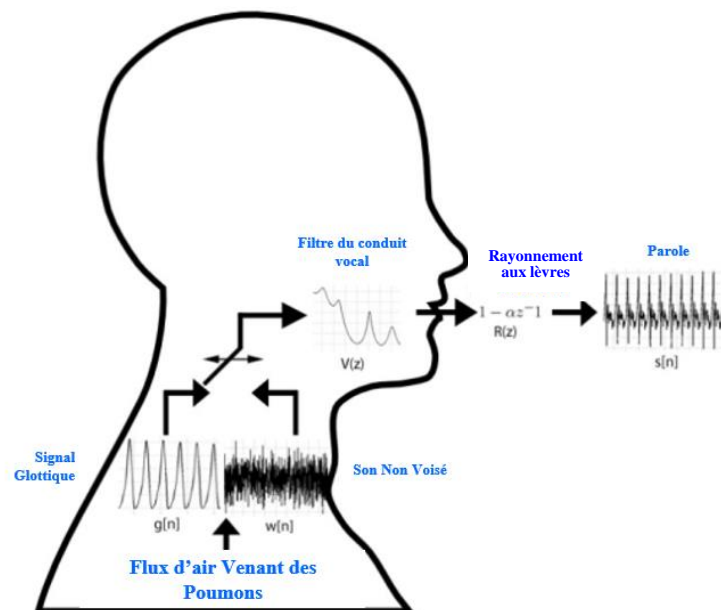
La phase fermeture fait référence à la période pendant laquelle la glotte est fermée. Pendant cette période, les effets des cavités sous-glottiques sont minimisés, ceci est mieux adapté pour estimer la fonction de transfert du conduit vocal. Par conséquent, les méthodes basées sur un filtrage inverse en phase fermée Closed Phase Inverse Filtering (CPIF) estiment un modèle paramétrique de l'enveloppe spectrale, calculé au cours de la durée estimée de la phase de fermeture [32]. Plusieurs approches ont été proposées dans la littérature pour résoudre ce



problème, les informations du signal électroglottographique sont utilisées pour identifier la période de fermeture de la glotte [33]. Une estimation basée sur une analyse formantique entre la phase d'ouverture et fermeture de la glotte a été proposée. Le principal inconvénient de ces techniques réside dans la difficulté d'obtenir une détermination précise de la phase fermée. Plusieurs approches ont été proposées dans la littérature pour résoudre ce problème. La robustesse de CPIF par rapport à la position de trame du signal de parole a été améliorée en imposant certaines contraintes. Outre ce problème de détermination précise de la phase fermée, il peut arriver que cette période soit si courte (pour les voix aiguës) qu'il ne reste pas assez d'échantillons pour une estimation fiable du filtre. Il a donc été proposé dans [34] une technique de codage linéaire prédictif LPC (Linear Predictive Coding) en cycle fermé à plusieurs cycles, dans laquelle un petit nombre de cycles glottiques voisins sont pris en compte afin de disposer de suffisamment de données pour une estimation précise du conduit vocal.

### 1.7.2 Filtrage inverse itératif et / ou adaptatif

Le filtrage inverse est la technique la plus utilisée pour extraire le signal glottique. L'équation 1.3 et le modèle Source-Filter de production de la parole de la figure 1.8 montrent que la forme d'onde de la parole acoustique peut être modélisée par une convolution de la source glottique avec la réponse impulsionnelle du filtre de conduit vocal et le rayonnement des lèvres [35]. La méthode IAIF utilise un processus en douze étapes de calcul du filtre du conduit vocal et de l'effet du rayonnement des lèvres par prédiction linéaire, filtrage inverse de la forme d'onde du signal de parole et intégration pour éliminer les effets du rayonnement sur les lèvres  $R(z)$  qui, dans cette méthode, est traité comme un différenciateur de premier ordre de l'air expulsé [36].



**Figure 1.8 :** Filtre source linéaire invariant dans le temps modèle de production de la parole pour algorithme l'IAIF

La source glottique peut ensuite être extraite par un filtrage inverse et annulation des effets du conduit vocal et du rayonnement illustrés dans l'équation I.4. Cependant, la précision des résultats dépend de la qualité de l'estimation du filtre du conduit vocal et dépend fortement de la qualité de la forme d'onde acoustique en entrée.

La méthode IAIF utilise l'analyse prédictive linéaire (LPC), pour modéliser le filtre du conduit vocal et les lèvres en tant que différentiateur. Elle réduit de façon itérative les effets du conduit vocal et des lèvres en filtrant de manière inverse la forme d'onde acoustique à différents ordres de LPC et en intégrant les résultats. Un système de prédiction d'ordre  $p$  linéaire est défini par l'équation :

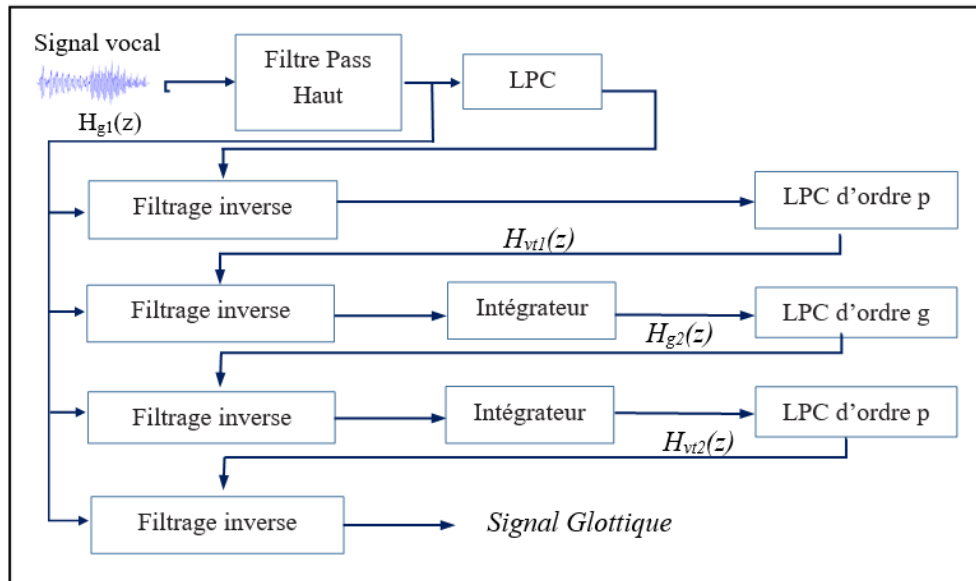
$$\hat{s}[n] = \sum_{k=1}^p \alpha_k s[n-k] \quad (1.9)$$

Où  $\hat{s}[n]$  est le signal de parole prévu et  $e[n]$  l'erreur est définie comme :

$$e[n] = s[n] - \hat{s}[n] = s[n] - \sum_{k=1}^p \alpha_k s[n-k] \quad (1.10)$$

Les coefficients  $p$  du système de prédiction sont choisis de manière à minimiser les erreurs de prédiction «  $n$  ». L'analyse LPC est utilisée dans les douze étapes de la méthode IAIF en raison de sa précision dans la modélisation du spectre de la parole. Elle peut être donc appliquée de manière itérative pour supprimer les effets du conduit vocal. La méthode IAIF est présentée par un schéma fonctionnel par les étapes suivantes :

- la forme d'onde du signal de la voix enregistrée est filtrée par un filtre passe-haut afin d'éliminer le bruit ambiant dont les basses fréquences ou les réverbérations pouvant être enregistrées par le microphone ;
- Le signal de parole filtré est analysé ensuite par un LPC de premier ordre ce qui fournit une  $H_{g1}(z)$  estimation préliminaire du flux glottique et effets de rayonnement aux lèvres.
- Filtrage inverse (inverse filtering) du signal vocal filtré passe-haut ;
- analyse par un LPC d'ordre  $p$ , indiquée par  $H_{vt1}(z)$ , est utilisée afin de réduire les effets du conduit vocal. L'ordre  $p$  de l'analyse LPC à la quatrième étape est généralement comprise entre 8 et 12 ;
- estimation de la forme d'onde de la parole avec annulation des effets du conduit vocal ;
- détermination d'une nouvelle estimation de l'onde glottique comme  $H_{g2}(z)$ , en utilisant une analyse LPC de l'ordre  $g$  qui est généralement entre 2 et 4 ;
- L'estimation finale des effets du conduit vocal est calculée par une autre analyse LPC d'ordre  $p$  à la dixième étape pour fournir  $H_{vt2}(z)$ . Ceci est utilisé pour filtrer inversement la forme d'onde acoustique filtrée passe-haut. En intégrant le résultat nous donne l'estimation finale de la forme du signal glottique,  $g[n]$  ;



**Figure 1.9 :** Diagramme en bloc de la méthode de Filtrage Inverse (IAIF).

La sortie de la septième étape présente l'estimation du signal glottique. La sortie de la septième étape est utilisée dans la huitième étape pour un filtrage inverse de la forme d'onde de la parole filtrée passe haut et annuler les effets de la contribution glottique, la sortie de la huitième étape étant intégrée dans la neuvième étape afin de réduire davantage les effets du rayonnement des lèvres. L'estimation finale des effets du conduit vocal est calculée par une autre analyse LPC d'ordre p à la dixième étape pour fournir  $H_{vt2}(z)$ . Ceci est utilisé pour filtrer inversement la forme d'onde acoustique filtrée passe-haut une dernière fois à l'étape onze en intégrant le résultat donne l'estimation finale de la forme du signal glottique,  $g[n]$ .

## 1.8 Conclusion

Ce premier chapitre a été consacré à la présentation de l'anatomie et la physiologie du système de production de la voix ainsi qu'à la description de des parties essentielles dans la production de la parole. Ensuite, nous avons présenté un modèle de production de la parole qui est le modèle source-filtre ainsi que le modèle LF de la source glottique ainsi que leurs détails mathématiques respectifs.

La description mathématique des méthodes les plus utilisées d'estimation de l'impulsion du signal glottique, qui comprenait différentes techniques de filtrage inverse. En outre, le paramétrage des impulsions glottales a été examiné, de même que l'évaluation des estimations du débit glottal.

Des résultats ont montré que la méthode de filtrage inverse IAIF semble de donner une estimation précise de l'onde glottique proche au signal réel. Ceci nous permet facilement d'évaluer la forme d'onde sous différents types de pathologies en analysant les quatre phases du cycle glottique. En général les procédures basées sur des instruments par exemple la glottographie, sont inconfortables pour le locuteur et interfèrent avec le comportement de la phonation normale, ce qui rend plus attrayant et motive le développement de techniques qui estiment l'impulsion glottique directement à partir du signal de parole.

*Chapitre 2 :*  
*Analyse Paramétrique des*  
*Pathologies des Cordes Vocales*

## 1 Introduction

La voix pathologique provient de certains troubles de la voix, qui se traduisent par une modification au niveau des paramètres acoustiques (altération objective) ou/et sonores (altération subjective) de la parole. Ce dysfonctionnement de peut être momentané ou durable. En général, il existe trois grandes catégories de pathologies vocales :

- **Les pathologies d'origines fonctionnelles** : mauvaise utilisation des organes de la phonation (conduit vocal), la cause est souvent liée à l'âge du patient (locuteur). On retrouve parfois une altération de la voix de cause psychologique comme par exemple, une dépression.
- **Les pathologies d'origines organiques** : laryngite aiguë, présence de lésion sur les cordes vocales, kystes, ... etc. Les principales causes de ces pathologies sont le forçage de la voix et les infections virales ou bactériennes du larynx.
- **Les pathologies d'origines cancéreuses** : l'ablation partielle ou totale du larynx est un acte chirurgical motivé par un cancer. La consommation d'alcool et l'usage du tabac en sont les principales causes.

L'élaboration d'un système automatique en vue d'une aide au diagnostic des troubles vocaux, nécessite une bonne connaissance de l'origine de ces troubles, leur localisation et dans la mesure du possible leur incidence sur la phonation.

Dans ce chapitre, nous présenterons de façon non exhaustive les différentes pathologies relatives aux cordes vocales dont nous avons entamés dans cette thèse. Ensuite les méthodes de diagnostic paramétrique et évaluation basées sur les techniques de traitement de signal. En vue de surmonter le manque du traitement automatique de la voix pathologique, nous attacherons une importance particulière dans ce présent chapitre à une toute nouvelle technique appelée l'Analyse Quantitative de Récurrence (Recurrence Quantification Analysis, ou RQA). Cette dernière permet de quantifier la dynamique du signal correspond aux vibrations des cordes vocales au long du processus de production de la voix.

## 2 Pathologies des Cordes Vocales

Les cordes vocales, jouant un rôle primordial dans la phonation, tout trouble l'affectant se répercute d'une façon ou d'une autre sur la qualité de la voix. La Production de la voix normale est basée sur sa qualité, son intensité, son débit. Une voix pathologique présente une altération d'un ou de plusieurs de ces paramètres [37]. Les pathologies des cordes vocales peuvent avoir différentes origines et différentes formes :

- troubles affectant la morphologie des plis vocaux,
- troubles affectant le système nerveux (périphérique ou central),
- troubles provenant d'autres phénomènes.

On peut classificateur les troubles de voix selon la régions pathologiques, c'est ce qui montre que la voix peut être altérée ou modifiée tout le long de sa production, Nous concentrons dans

notre étude sur les pathologies affectant les cordes vocales. Nous avons donné une attention particulière à deux cas pathologiques organique, à savoir les nodules et les polypes et deux cas neurologiques : la dysphonie spasmodique et la paralysie des cordes vocales, sur lesquelles nous avons fait une analyse du signal glottique et une classification automatique par rapport à la voix normale.

### 2.1 Polypes de cordes vocales

Souvent secondaire à un malmenage vocal, le polype peut être favorisé par une intoxication alcoolo-tabagique ou une exposition aux poussières. Le rôle de l'inflammation joue peu, il s'agit d'une tuméfaction bénigne de la corde vocale du plan glottique. Il est de forme lisse ou mamelonnée. Son implantation sessile ou pédiculée. Il est localisé au 1/3 moyen ou à la commissure antérieure. Il se présente sous une forme inflammatoire, œdémateuse ou hémorragique. Il est responsable d'une dysphonie : La voix est grave et bitonale, le diagnostic repose sur la laryngoscopie indirecte et la fibroscopie [38].

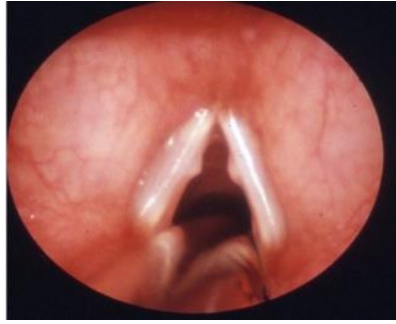


Figure 2.1 : Polype des cordes vocales [38]

### 2.2 Nodules des cordes vocales

Ils correspondent à un épaississement localisé, généralement bilatéral (des deux côtés) du rebord libre de la corde vocale. Cet épaississement se situe à la jonction du premier et du deuxième tiers des cordes vocales et empêche l'approximation de celles-ci durant la phonation. Les nodules sont la conséquence du choc répété des cordes vocales entre elles à cause d'une sur utilisation ou d'une mauvaise utilisation de la voix (p. ex. cris excessifs, vocalisations forcées, toux forte, éclaircissements répétés de la voix pour dégager des sécrétions, etc.). Certains facteurs peuvent aussi favoriser leur apparition : infections fréquentes des voies respiratoires supérieures (rhinite, sinusite), bronchites, asthme, allergies, exposition à la fumée de cigarette, reflux laryngopharyngé, présence de produits irritants dans l'atmosphère poussières, allergènes [39].

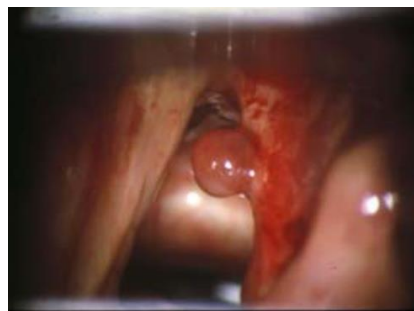
Caractéristiques acoustiques : La fréquence fondamentale est abaissée, les aigus sont diminués voire inexistant. Des dé-sonorisations en début et en fin de production sont fréquentes. L'intensité est instable. Le timbre est éraillé, rauque et soufflé.



**Figure 2.2 :** Nodules des cordes vocales [39]

### 2.3 Les granulomes

Sont des formations inflammatoires qui ne donnent vraiment de troubles dysphoniques que lorsqu'ils sont volumineux et gênent l'occlusion cordale. Le plus souvent ils se manifestent par une simple sensation de gêne pharyngée et sont découverts par l'examen laryngé systématique. Les granulomes sont unis ou bilatéraux, en général développés dans la partie postérieure de l'orifice glottique. Parfois, il n'y a pas de cause évidente, si ce n'est le surmenage vocal ou un reflux gastro-œsophagien qu'il faudra toujours soigneusement rechercher [40].



**Figure 2.3 :** Granulome des cordes vocales [40]

### 2.4 Paralysie des cordes vocales

Les facultés du nerf récurrent sont affaiblies unilatéralement, provoquant une immobilité d'un héli larynx. La corde vocale peut être paralysée en position latérale (provoquant une fermeture incomplète en phonation), intermédiaire, ou médiane (une dyspnée peut alors apparaître à l'effort). Caractéristiques acoustiques : Variation de la fréquence fondamentale : diminuée chez la femme et augmentée chez l'homme. La tessiture est diminuée, la voix est bitonale, inexistante lors des cris. L'intensité est réduite. Le caractère soufflé du timbre est le plus marquant. La voix est éteinte et détimbrée.



**Figure 2.4** : Paralysie unilatérale des cordes vocales [40].

## 2.5 Dysphonie spasmodique

C'est une atteinte dystonique des muscles du larynx qui se manifeste le plus souvent au moment de la phonation. Autant de femmes que d'hommes sont touchés, le plus souvent entre 45 et 60 ans. Les spasmes respiratoires et vocaux permettent de faire le diagnostic à la simple écoute de la voix. Il existe deux types de dysphonie spasmodique : en adduction (fermeture excessive de la glotte en phonation) et abduction (prédominance des spasmes respiratoires, ouverture excessive de la glotte en phonation). Caractéristiques acoustiques : Dans la dysphonie spasmodique en adduction, la voix est forcée, hachée, éraillée, ponctuée d'arrêts vocaux. Dans les formes les plus sévères, le sujet est presque aphone. Dans la dysphonie spasmodique en abduction (forme rare), la voix est chuchotée, à peine audible, avec des perturbations de la dynamique respiratoire.

## 2.6 Kystes des cordes vocales

Un kyste de corde vocale est une masse ferme de tissu, contenue dans une membrane, et il peut être situé près de la surface d'une corde vocale ou plus profond, près du ligament de corde vocale. Ces kystes sont moins courants que les nodules ou les polypes décrits ci-dessus. Ils peuvent être provoqués par l'utilisation tendue de la voix, par exemple, essai de converser en tant que normale pendant un accès de laryngite ou d'une infection des voies respiratoires supérieures. Habituellement, la chirurgie et le traitement de voix est recommandée pour des cas où le kyste entraîne la perturbation significative de voix.

Caractéristiques acoustiques : On note une aggravation de la hauteur, avec des aigus réduits voire inexistantes. L'intensité est forte avec des dé-sonorisations régulières. Le timbre est forcé, voilé, dur et éraillé.



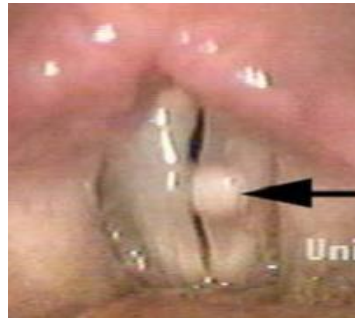


Figure 2.5 : Kyste de corde vocale [41]

## 2.7 Cancer des Cordes Vocales

Les cancers dans les zones glottiques se révèlent par une modification progressive de la voix qui devient couverte, voilée, rauque (dysphonie). Cette modification persiste et s'aggrave progressivement. Elle est parfois précédée d'épisodes transitoires de laryngite où complique une laryngite chronique ancienne, fréquente chez les fumeurs et/ou les personnes travaillant en atmosphère chaude et sèche, ou chargée de poussières. Les cancers des sub-glottiques siègent au niveau de l'épiglotte, par :

- un gêne ou une douleur d'un seul côté de la gorge ;
- une difficulté à avaler ;
- une sensation permanente de corps étranger ou d'angine d'un seul côté ;
- une douleur à une oreille ;
- l'apparition d'une boule dans le cou qui correspond à un ganglion.

## 3. Moyens d'évaluation des pathologies vocales

Pour mieux comprendre l'analyse du signal vocal, nous détaillons dans cette partie les différentes méthodes d'évaluation subjectives et objectives. Ce qui permettra aux médecins d'évaluer le pronostic et de suivre l'évolution de la maladie avant et après le traitement. La première approche est basée sur une évaluation perceptive à l'oreille de la voix et de la parole. La deuxième approche repose sur l'acquisition de mesures physiques en utilisant un capteur et l'application des techniques de traitement du signal sur les mesures extraites pour établir l'évaluation.

### 3.1 Evaluation Perceptive

L'analyse perceptive reste la première analyse réalisée par les cliniciens au moment de la consultation. C'est parce que la voix est altérée que les patients viennent consulter, c'est donc sur la qualité de la voix que doivent être jugés les résultats des thérapeutiques. Le principe d'une évaluation perceptive, est d'évaluer la voix uniquement à l'oreille, la quantité de dysphonie dans une voix. C'est la méthode d'évaluation vocale la plus couramment employée en clinique en premier lieu.

Pour simplifier la tâche des cliniciens, le classement des dysphonies sur une échelle perceptive se fait selon l'échelle GRBAS D'HIRANO [42] qui a décrit 5 paramètres qualitatifs à 4 niveaux (0=voix normale, 1=altération légère, 2=altération moyenne,

3=altération sévère). Il s'agit d'une échelle d'évaluation perceptive dont l'utilisation est largement recommandée par les spécialistes de ce domaine. Tableau 2.1.

**Tableau 2.1** Signification et définition des paramètres de l'échelle GRBAS [42]

<b>Lettre</b>	<b>Mot</b>	<b>Signification</b>	<b>Définition</b>
G	Grade	Grade globale de la dysphonie	cela correspond à une évaluation générale, globale de la qualité de la voix
R	Roughness	Raucité	il s'agit d'évaluer la raucité de la voix et toutes les altérations du timbre (ébrailure, craquement), ainsi que la régularité de la vibration des plis vocaux.
B	Breathiness	Caractère soufflé	c'est la composante du souffle dans la voix. Correspond à la production des voix dites voilées où soufflées. Cette caractéristique est liée directement à la présence d'une fuite d'aire lors de la phonation.
A	Asthenia	Asthénie	la sensation de faiblesse : qui traduit un manque de puissance dans la voix lié à une intensité faible où au manque d'harmoniques aigues.
S	Strain	Forçage	Effet d'un état hyperfonctionnel de phonation, 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.

L'une des limites les plus mentionnées dans la littérature est la variabilité inter et intra-individuelle dans la perception de la voix par un jury d'experts. Cette variabilité peut être influencée par le contexte, par l'état émotionnel ou l'attention de l'auditeur. Plusieurs travaux ont tenté d'étudier la fiabilité de l'analyse perceptive en étudiant le degré d'agrément entre les auditeurs [43, 44]. Les résultats mettent en évidence un désaccord important entre les auditeurs dans la perception d'une même voix. Néanmoins, un accord parfait et une analyse commune de la voix ne sont pas évidents au sein d'un groupe d'auditeurs. En pratique, l'analyse perceptive par un auditeur isolé ne peut pas être considérée comme un outil fiable de comparaison entre deux échantillons de voix et seul un jury d'experts peut être considéré

comme une méthode de référence. Il va sans dire qu'il s'agit d'une méthode extrêmement lourde et est difficile à proposer en routine.

### **3.2 Evaluation objective**

L'évaluation objective utilise des techniques de traitement du signal dans le but d'extraire des indices acoustiques pertinents pour déterminer les caractéristiques de la qualité vocale qui permettent de fournir des informations sur l'état des plis vocaux de l'orateur. Les cliniciens (Orthophonistes et médecins ORL) préfèrent une analyse acoustique paramétrique en corrélation avec les caractéristiques perceptives. Cependant, il serait utile de combiner un accord acoustico-perceptif lorsque l'évaluation perceptuelle n'est pas suffisamment fiable.

La caractérisation de la voix pathologique est consacrée au développement de méthodes d'analyse dédiées à l'estimation des perturbations dans le signal vocal. Il existe plusieurs méthodes d'évaluation objective de la voix pathologique, dans cette section, elles sont classées en deux catégories selon qu'elles utilisent directement le signal vocal et celles qui sont basées sur une estimation de la source glottique.

#### **3.2.1 Paramètres aérodynamique (physique)**

Nous citons : le temps maximal de phonation, le débit d'air buccal et la pression sous glottique.

**3.2.1.1 Le temps maximal de phonation :** C'est le temps de tenue maximale d'une émission vocalique après une inspiration profonde. Il représente un reflet tout à fait fidèle de la qualité de la fermeture glottique et, de ce fait, il est considéré par de nombreux auteurs comme un bon indicateur de la fonction vocale.

**3.2.1.2 Le débit d'air buccal :** le débit d'air buccal est un paramètre qui permet de mesurer le débit aérien lors de la phonation et de quantifier la fuite d'air glottique secondaire à la pathologie vocale. Ainsi, ce paramètre rend compte de la composante du souffle que l'on peut entendre dans le timbre des voix pathologiques et l'associe à une occlusion glottique insuffisante engendrant une fuite continue d'air pendant la phonation. Cette fuite d'air glottique se caractérise par une augmentation du débit [45]. On considère que les résultats sont pathologiques lorsqu'ils se situent au-delà de 40ml/s ou au-delà de 200 ml/s. [45].

**3.2.1.3 La pression sous glottique :** La Pression Sous Glottique est la pression à laquelle l'air arrive dans les poumons. Elle permet la mise en vibration des cordes vocales et ainsi la production du son. Elle augmente avec l'intensité de la voix et plus légèrement avec la hauteur tonale.

#### **3.2.2 Analyse acoustique de la voix**

Dans un système de reconnaissance automatique de la parole, l'analyse acoustique consiste à extraire du signal vocal un ensemble de paramètres pertinents dans le but de réduire la redondance du signal vocal pour une tâche de reconnaissance de la parole (voix pathologique dans notre cas). Ces paramètres doivent être discriminants en rendant les sons de base facilement séparables. Ils doivent être robustes au bruit. L'analyse du signal vocal

acoustique est indispensable à l'évaluation des variations de ses caractéristiques révélateurs d'éventuels changements de sa morphologie spectro-temporelle permettant la distinction entre un signal physiologique et un signal pathologique.

### 3.2.2.1 Intensité de la voix

L'intensité d'un son, appelée aussi volume, permet de distinguer un son fort d'un son faible. Elle correspond à l'amplitude de vibration des cordes vocales, qui dépend elle-même de la pression d'air sous glottique. Elle varie d'un individu à un autre, en fonction des habitudes vocales et des facteurs anatomie physiologiques. Elle peut être objectivement exprimée par le calcul des variations de pression d'air (exprimée en  $\text{watt/cm}^2$ ). Une échelle illustre les niveaux de pression sonore auxquels l'être humain est soumis. Un échèle graduée de 0 dB, seuil de perception de l'oreille humaine, à 120 dB considérée comme seuil de douleur. Une discussion normale a une valeur d'intensité d'environ 60 dB. Lorsque l'amplitude de l'onde est grande, l'intensité est grande et donc le son est plus fort. Une discussion forte peut aller jusqu'à 80 dB. Par contre, la parole chuchotée est à environ 10 dB d'intensité [46].

### 3.2.2.2 Le timbre

Chaque son est formé d'une fréquence fondamentale et de fréquences dites « harmoniques ». On dit qu'un son est riche, agréable à entendre, lorsqu'il contient de nombreuses harmoniques. Un son pauvre en harmoniques paraîtra terne à notre oreille. Les sons produits par l'appareil phonatoire sont des mélanges de sons graves et aigus. Ainsi un signal acoustique est la superposition d'un son fondamental et d'harmoniques dont les fréquences sont des multiples de la fréquence fondamentale. Ce sont celles-ci qui constituent le timbre du signal vocal acoustique et qui caractérisent chaque individu [47]. C'est en quelque sorte la couleur du son vocal à partir de laquelle, nous pouvons identifier une personne à une simple écoute de sa voix.

Le timbre vocal dépend de trois critères essentiels : l'accolement des cordes vocales, leur épaisseur et enfin les caractéristiques anatomiques des différentes cavités de résonance de l'appareil phonatoire. Les cavités de résonances contribuent également à la couleur de la voix, car en modifiant leurs volumes, nous obtenons telle ou telle voyelle. Le timbre vocal est l'un des paramètres acoustiques les plus caractéristiques des voix pathologiques. Selon que la voix est normale ou pathologique, nous distinguons les différents timbres :

- **clair** : caractéristique de la voix à l'état normale (voix naturelle), avec un accolement ferme et sans aucun dysfonctionnement des cordes vocales. C'est un timbre riche en harmoniques.
- **serré** : le patient en forçant sur son larynx, qui a une position trop haute, entraîne une contraction et une diminution du volume de résonateurs. La pression expiratoire est très importante et provoque un accolement brutal des cordes vocales. La voix présente beaucoup trop d'harmoniques aigus.
- **sombre** : la voix est exagérément grossie avec un larynx trop bas par manque de tonicité musculaire. Les résonateurs sont trop ouverts et sans tonicité. Les cordes

vocales ne s'accrochent pas suffisamment, entraînant des fuites d'air. Les cavités de résonance influent faiblement sur la voix et l'articulation devient floue.

- **érailé** : résultat de la superposition d'une vibration parasite irrégulière sur le son fondamental laryngé, provoqué par une lésion du bord libre d'une corde vocale. La vibration perd de sa souplesse et de sa régularité. La glotte n'est pas totalement fermée.
- **voilé** : les cordes vocales présentent un défaut de fermeture modéré entraînant une perte des harmoniques aigus. La voix est de faible intensité. Elle manque de netteté et clarté.
- **rauque** : la voix est grave, le signal de parole est apériodique. Les cordes vocales s'accrochent mal. Elles semblent rigides du fait d'une altération de leur capacité vibratoire. La voix est émise dans le bas pharynx. En conséquence, nous avons une sensation d'effort avec un son rugueux et dur.
- **soufflé** : il y a adjonction d'un bruit de souffle par défaut très important de fermeture glottique et la présence en excès d'harmoniques aigus.

### 3.2.2.3 Fréquence fondamentale ( $F_0$ )

La forme d'onde de la parole se compose de deux parties : une partie quasi périodique qui a tendance à se répéter après un bref intervalle de temps; cette période est appelée période correspond à la fréquence fondamentale et son inverse est la fréquence fondamentale. La deuxième partie est une partie semblable au bruit qui a une forme aléatoire. La fréquence fondamentale correspond aux vibrations des cordes vocales pour les sons vocaux de la parole.  $F_0$  dans une forme d'onde de parole naturelle varie généralement lentement avec le temps. Il peut être inférieur ou égal à 80 Hz pour les hommes adultes et supérieur à 300 Hz pour les enfants et certaines femmes adultes. La  $F_0$  est le principal paramètre dans un signal acoustique vocal de l'intonation et du stress dans la parole. Elle est essentielle dans les langages sonores pour l'identification des phonèmes.

### 3.2.2.4 Energie à court terme

Le signal à court terme est obtenu en utilisant une fonction de fenêtre et l'énergie est exprimée en tant que :

$$E_n = \sum_{m=n-N-1}^n [x(m)w(n-m)]^2 \quad (2.1)$$

Où  $w$  est la fonction de fenêtre et  $x(m)$  le signal. Il est démontré que l'énergie à court terme peut être utilisée pour distinguer les sons voisé/non-voisé : La valeur de  $E_n$  pour un son voisé est plus grande de celle pour un son non voisé.

### 3.2.2.5 Taux de passage par zéro

Le passage par zéro est réputé se produire lorsque des échantillons successifs présentent des signes algébriques différents. Le taux de passage par zéro est une mesure du nombre de fois où l'amplitude des signaux de parole passe par la valeur zéro dans un intervalle de temps

donné. Le taux auquel se produisent les passages par zéro est une simple mesure du contenu fréquentiel d'un signal. Le taux de passage à zéro est défini comme suit :

$$z(n) = \sum_{m=-\text{inf}}^{\text{inf}} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(m-n) \quad (2.2)$$

Où :

$$\text{sgn}[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases} \quad (2.3)$$

### 3.2.2.6 Jitter

C'est la variation de la fréquence fondamentale d'un cycle à un autre, c'est-à-dire la différence absolue entre les périodes consécutives, il est exprimé en :

$$Jitter = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}| \quad (2.4)$$

Où :  $T_i$  est les périodes extraites et  $N$  le nombre de périodes, comme illustré dans la figure ci-dessous.

### 3.2.2.7 Shimmer

Variabilité de l'amplitude crête à crête en décibels, c'est-à-dire le logarithme décimal absolu moyen en de la différence entre les amplitudes de périodes consécutives multipliées par 20.

$$Shimmer = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \text{Log} \left( \frac{A_{i+1}}{A_i} \right) \right| \quad (2.5)$$

$A_i$  sont les données d'amplitude crête à crête extraites et  $N$  le nombre de périodes.

### 3.2.2.8 Les Formants

On désigne par formant (acoustique) l'un des maxima d'énergie du spectre du signal sonore. Le spectre sonore représente la distribution en fréquences de l'énergie du signal vocal. L'analyse fréquentielle se fait par le biais d'un segment très court (appelé aussi fenêtre d'analyse ou trame), du signal vocal. Comme le spectre sonore, dont ils représentent l'un des aspects, les formants sont mesurés sur des segments du signal vocal très courts, et ils sont donc variables au cours du temps, dépendant du phonème prononcé et de la manière dont il est prononcé. On numérote les formants (qui sont donc les maxima du spectre sonore) en allant des basses fréquences vers les hautes fréquences. On les désigne par la notation physique  $F_i$  (mesurée en Hertz) en partant du premier formant  $F_1$  dans les fréquences basses. La notation  $F_0$  est réservée à la fréquence fondamentale, dont les variations au cours du temps constituent l'intonation du signal vocal.

Les deux premiers formants  $F_1$  et  $F_2$  sont des caractéristiques particulièrement adaptées à la description des voyelles, car les voyelles sont des sons relativement stationnaires. Chaque voyelle se caractérise ainsi par son timbre spécifique, déterminé en première approximation par  $F_1$  et  $F_2$ . La discipline phonétique a permis d'établir une corrélation entre les valeurs mesurées en Hertz de  $F_1$  et  $F_2$  et les articulations de l'appareil phonatoire nécessaires à la réalisation des voyelles. Ainsi  $F_1$  est corrélé avec l'aperture (ouverture de la bouche) et le  $F_2$  avec la position antérieure (valeur élevée de  $F_2$ ) et la position postérieure (valeur basse de  $F_2$ ) de la langue. Le troisième formant  $F_3$  est également une caractéristique intéressante, corrélée avec la configuration des lèvres pour les voyelles antérieures. La figure 2.6 présente un signal de la voyelle [a :] et les différents formants  $F_1$  et  $F_2...5$ .

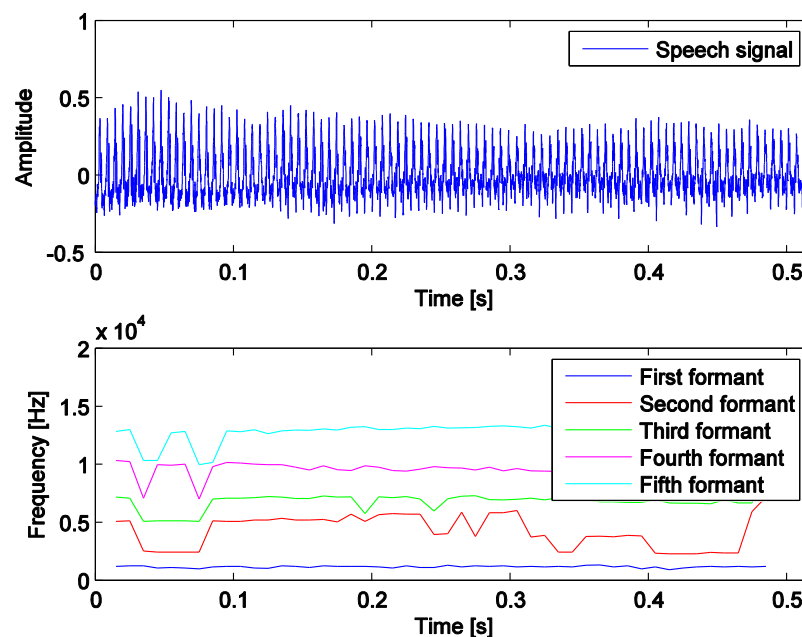


Figure 2.6 Formants d'un signal vocal

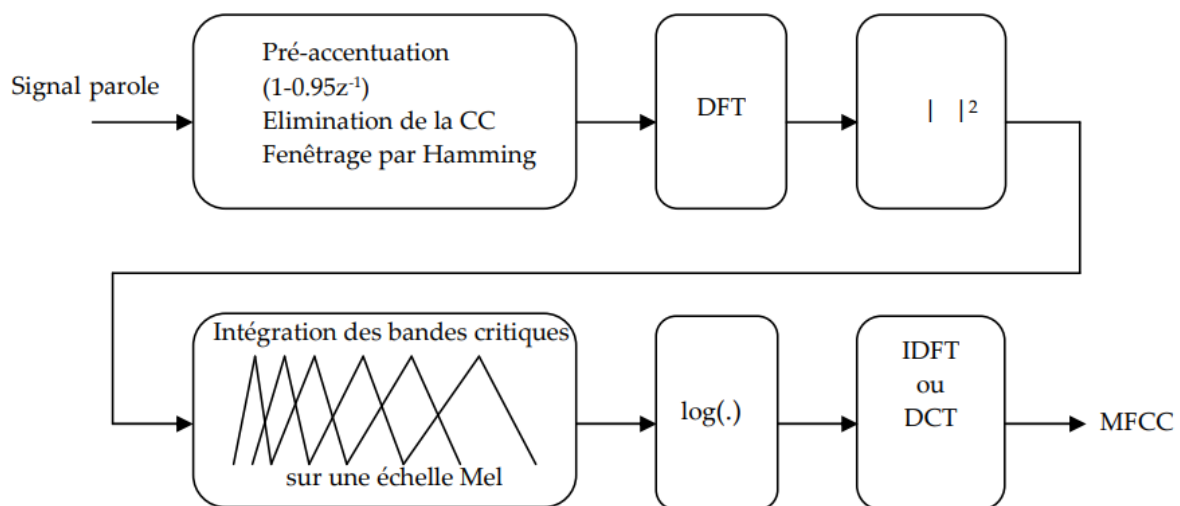
### 3.2.3 Paramètres issus de l'analyse par banc de filtres

L'analyse par banc de filtres est une technique qui a été initialement utilisée pour le codage du signal de parole (vocodateur à canaux). Dans ce cadre elle fait partie des techniques de codage par analyse-synthèse à partir du modèle source/filtre. Elle consiste à filtrer le signal par un ensemble de filtres passe-bande. L'énergie en sortie de chaque filtre est attribuée à sa fréquence centrale. Pour simuler le fonctionnement du système auditif humain, les fréquences centrales sont réparties uniformément sur une échelle perceptive. Plus la fréquence centrale du filtre est élevée, plus sa bande passante est large. Cela permet d'augmenter la résolution dans les basses fréquences, zone qui contient le plus d'information utile dans le signal de parole. Les échelles perceptives les plus utilisées sont le Mel ou le Bark. Du point de vue des performances des systèmes de reconnaissance, ces deux échelles sont quasiment identiques.

### 3.2.3.1 Coefficients MFCC (en Anglais « Mel Frequency Cepstrum Coefficients »)

Le traitement décrit dans le paragraphe précédent permet d'obtenir une estimation de l'enveloppe spectrale (densité spectrale lissée). Il est possible d'utiliser les sorties des bancs de filtres comme entrées dans le système de reconnaissance. Cependant, d'autres coefficients dérivés des sorties d'un banc de filtres, sont plus discriminant, plus robustes au bruit ambiant et moins corrélés entre eux. Il s'agit des coefficients cepstraux dérivés des sorties du banc de filtres répartis linéairement sur l'échelle Mel, ce sont les paramètres MFCC (Davis, 1980) [48]. Les étapes d'une analyse MFCC sont présentées dans la figure 2.7.

Une autre modification consiste à utiliser des paramètres (dits  $\Delta$ -cepstraux) qui caractérisent les variations des paramètres cepstraux dans les fenêtres proches de la fenêtre à court-terme courante (Furui, 1981a; Soong, 1988) [49]. Ces paramètres sont nettement moins efficaces que les paramètres cepstraux dans le cadre de la reconnaissance du locuteur lorsqu'ils sont utilisés seuls. Par contre, utilisés avec les paramètres cepstraux "instantanés" ils conduisent à une amélioration substantielle des performances. Les paramètres variationnels présentent en outre l'intérêt d'être insensibles à des variations linéaires du canal de transmission entre deux enregistrements. Toutefois, les paramètres variationnels ne sont réellement significatifs que dans les cas où il est possible de les comparer dans un même contexte, c'est à dire dans les applications où le texte prononcé par chaque locuteur est fixé (Soong, 1988) [50]. En mode indépendant du texte l'utilisation de paramètres,  $\Delta$ -cepstraux, en plus des paramètres cepstraux, ne semble pas améliorer sensiblement les performances de reconnaissance (et ce même dans les cas où le vocabulaire est fortement contraint comme dans (Tseng 1992)) [51].

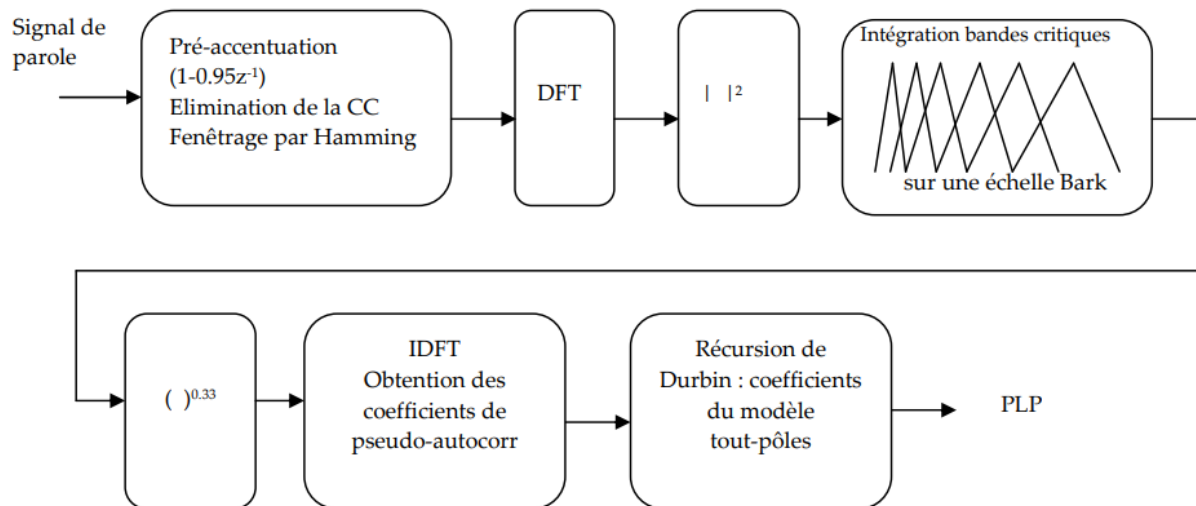


**Figure 2.7** Calcul des coefficients MFCC

Un autre point important est le fait que tous les coefficients cepstraux ne présentent pas des variations intra-locuteur de même ampleur : la variance des coefficients cepstraux décroît avec leur ordre. Pour tenir compte de cette propriété, on utilise généralement la distance dite de Mahalanobis, qui se réduit ici à une distance euclidienne pondérée du fait de la



décorrélation des coefficients cepstraux. Cette distance est beaucoup plus efficace dans le cadre de la reconnaissance du locuteur car elle permet de réduire l'influence des coefficients qui présentent une forte variabilité intra-locuteur (Soong, 1988) [50]. Compte tenu des propriétés du signal de parole, les poids à appliquer aux coefficients cepstraux pour le calcul de la distance pondérée peuvent être approximés par une progression linéaire en fonction de l'ordre des coefficients cepstraux (on obtient alors la distance dite root powers sum).



**Figure 2.8 :** Calcul des coefficients PLP

### 3.2.3.2 Coefficients PLP (en anglais « Perceptual Linear Predictive »)

La technique conventionnelle de la prédiction linéaire (LPC) est fondée sur un modèle tout-pôles autorégressif qu'on estime en utilisant une analyse par prédiction linéaire. Ce modèle constitue une approximation de l'enveloppe de la densité spectrale à court terme. Cette analyse est en accord avec les aspects de la production de la parole qu'avec les aspects de la perception de la parole. L'analyse PLP (Hermansky, 1990) [51] est une alternative à l'analyse LPC qui tient compte de trois aspects de la perception (Figure 2.8).

## 3.3 Analyse paramétrique basé sur l'estimation du signal glottique

L'analyse du signal de la source glottique a été utilisée dans plusieurs applications de traitement automatique de signal de la parole, et plus particulièrement pour la détection des troubles de la voix. En effet, les pathologies de la parole sont souvent associées à un dysfonctionnement des cordes vocales, par exemple, un nodule ou un polype. En plus, L'utilisation de la source glottique peut également être intéressante dans d'autres applications telles que la création d'un signal d'excitation artificielle basé sur la compréhension de la production de signal de la source glottique pour les patients ayant subi une laryngectomie totale, et la réparation de la voix dysphonique où la parole laryngée est re-synthétisée en utilisant un signal glottique synthétique [52].

### 3.3.1 Paramètres glottiques (articulatoires)

La production des signaux vocaux, en particulier ceux liée aux sons voisés, par exemple les voyelles, commencent par contraction-expansion de l'air des poumons, générant une différence de pression entre l'air dans les poumons et l'air près de la bouche. Le flux d'air traverse les cordes vocales, oscillant dans une fréquence appelée la fréquence fondamentale de la voix. Cette oscillation modifie le flux d'air provenant des poumons, le transformant en impulsions d'air. Le signal de pression formé par l'air les impulsions sont quasi-périodiques et s'appellent le signal glottique.

Le signal glottique est obtenu en effectuant un filtrage inverse sur le signal vocal, qui consiste à éliminer l'influence du conduit vocal et le rayonnement vocal provoqué par les lèvres, préservant les caractéristiques du signal. Le filtrage inverse l'algorithme utilisé ici est l'IAIF (Inverse Adaptatif Itératif Filtering). Il a été choisi en raison de sa haute performance et sa facilité de développement. Dans le contexte de notre travail nous avons utilisé une interface graphique implémenté sur l'environnement MATLAB (MathWorks), appelée Aparat, qui a été développé spécialement sur la base de la méthode IAIF pour obtenir le signal glottal et en extraire ses principales caractéristiques ou paramètres. Les paramètres qui seront utilisés peuvent être divisés en deux groupes : domaine temporel, domaine fréquentiel [53].

### 3.3.1.1 Paramètres du signal glottique dans le domaine temporel

Les paramètres qui peuvent être extraites du signal glottique obtenu par un filtrage inverse dans le domaine temporel sont décrits ci-dessous :

- **Quotient d'ouverture (QO)** : Ce paramètre mesure la durée de la phase ouverte à partir du moment où l'amplitude de l'impulsion glottique croise la ligne du marqueur à 50%. Il est défini comme le rapport entre la durée de la phase d'ouverture et la période glottique. Un faible OQ correspond à une intensité vocale plus élevée ;
- **Ko (Phase de fermeture glottique)** : correspond à l'intervalle de temps que prennent les cordes vocales entre la phase d'ouverture maximale et la phase de fermeture ;
- **Ka (Phase d'ouverture glottique)** : identifie le temps dans lequel les cordes vocales prennent de la phase totalement fermée à la phase d'ouverture ;
- **PP (distance entre les instants d'ouverture glottique maximal)**. Définit la distance entre les valeurs maximales de crête du signal glottique ;
- **CIQ (Quotient de fermeture)** est le rapport entre la phase de fermeture et la durée totale d'un cycle du signal glottique. Il est donné par l'équation : 2.6.

$$CIQ = \frac{K_o}{T} \quad (2.6)$$

- **SQ (Speed Quotient)** est le rapport entre la durée de la phase d'ouverture et la phase de fermeture du signal glottal. Equation 7.

$$SQ = \frac{K_a}{K_o} \quad (2.7)$$

- **AQ (Quotient d'Amplitude)** est le rapport entre l'amplitude du signal glottique (ouverture maximale) et la valeur minimale du signal glottal. Equation 8.

$$AQ = \frac{A_v}{D_{\min}} \quad (2.8)$$

- **QoQ (Quotient d'Ouverture pendant l'impulsion glottique)** est la relation entre le temps d'ouverture du signal glottique au moment exacte de l'oscillation et le temps de fermeture.
- **QAN (Quotient d'Amplitude Normalisé)** est calculé comme le rapport entre le quotient d'amplitude (AQ) et la durée totale d'une impulsion glottique. Equation 2.9.

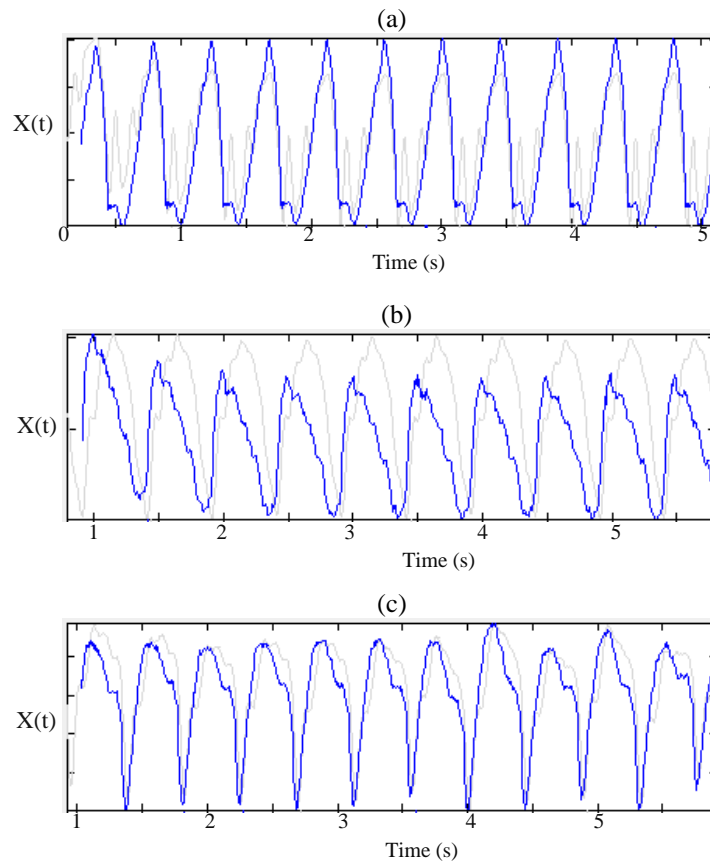
$$NAQ = \frac{AQ}{T} \quad (2.9)$$

### 3.3.1.2 Paramètres du signal glottique dans le domaine fréquentiel :

Deux paramètres dans le domaine fréquentiel sont étudiés dans le contexte de notre travail. Ces paramètres sont définis ci-dessous :

- **DH12 (Différence entre les harmoniques)** est la différence entre les valeurs de la première et de la deuxième harmonique du signal glottique ;
- **Facteur de Relation Harmonique (HRF)** relie le premier harmonique (H1) à la somme de l'énergie des autres harmoniques. Il a également été utilisé pour mesurer la qualité vocale ;

Le signal de pression formé par les impulsions d'air est quasi-périodique et il est appelé le signal glottique. Il révèle des informations importantes liées à de nombreuses applications dans le domaine du traitement de la parole comme les modèles de vibration des cordes vocales, l'identification de la voix pathologique, l'identification du locuteur et le traitement des empreintes vocales, ... etc. La figure ci-dessous décrit un signal glottique typique correspond à une voix normale (considérer comme référence) et les signaux glottiques correspondent à une voix pathologique (Nodules des cordes vocales et Paralysie unilatérale). Une analyse visuelle des trois signaux révèle une différence remarquable des trois signaux aux niveaux de la périodicité (perd de l'aspect quasiment périodique) ce qui induit une déformation de la forme d'onde en comparant à celle du sujet normale. La perte de la périodicité du signal glottique pathologique revient aux conditions d'articulation des plis vocaux. La présence d'une ou des lésions au niveau des cordes vocale perturbe les temps d'ouverture et fermeture des cordes vocales pendant la phonation, ceci est produit par la présence d'une pathologie.



**Figure 2.9** : Forme du signal glottique (a) : Voix Normale, (b) : Nodules, (c) : Paralyse unilatérale

#### 4 La technique du diagramme de récurrence (Recurrence Plots)

Cette partie décrit la description mathématique de la méthode dont nous avons l'utilisée dans le contexte de notre étude.

##### 4.1 Espace de Phase

Pour suivre l'évolution de l'état d'un système dynamique, préalablement décrit avec des lois physiques, on définira à un instant donné un point dans un repère. Ce point caractérise l'état du système dans l'espace à cet instant. Cet espace est appelé « l'espace des phases » de dimension  $N$ . Lorsque la variable d'évolution change de valeur (quand le temps s'écoule, par exemple), le point figurant à l'état du système décrit en général une courbe dans cet espace.

En pratique, il n'est pas possible de mesurer toutes les composantes d'un vecteur  $\vec{x}$  décrit dans un espace de phases à  $N$  dimensions. Ainsi, en 1981, Ruelle et Takens prouvent qu'il est possible de reconstruire l'attracteur d'une série temporelle. Dès lors, l'espace de phase est reconstruit en un « espace d'immersion » (ou espace de plongement) de dimensions  $m$  inférieur à  $N$  en utilisant les « vecteurs d'immersion » décrit comme suit :

$$\overrightarrow{x(t_i)} = [x(t_i), x(t_i - \tau), \dots, x(t_i - (m-1)\tau)]$$

où  $m$  représente la dimension d'immersion,  $\tau$  est le retard (ou le pas) et  $x(t_i)$  est la valeur de la série temporelle à l'instant  $t_i$ . Ce théorème permet donc de se renseigner sur le

comportement d'un système non linéaire, en l'étudiant sur un espace d'immersion reconstruit à partir d'une seule série temporelle, pourvu que les paramètres  $m$  et  $\tau$  de cette dernière soit connus. Pour l'estimation de ces deux paramètres, différentes méthodes sont adoptées [54].

#### 4.2 Diagramme de Récurrence

Pour observer les trajectoires dans l'espace des phases à dimension  $d$ , tout en conservant la dynamique du système étudié, Eckmann et al (1987) [54], les diagrammes de récurrence (**R**écurrences **P**lots ou **RP**) visant à révéler les récurrences dans les trajectoires de phases, c'est-à-dire les moments où une telle trajectoire visite la même zone de l'espace des phases. Supposons qu'on ait une trajectoire  $X$  d'un système dans son espace de phase. Ainsi, le RP correspondant est construit comme suit : chaque point de l'espace de phase  $X_i$  est testé s'il est proche ou non d'un autre point  $X_j$ . Ainsi, si la distance entre ces deux points est inférieure à un seuil  $\varepsilon$  prédéfini, les points sont dits récurrents et sont associés à un point noir. Dans le cas contraire, ils sont non récurrents et sont associés à un point blanc. Cela se traduit par une matrice  $N \times N$ , telle que :

$$R_{i,j} = \begin{cases} 1, & \|X_i - X_j\| \leq \varepsilon \\ 0, & \|X_i - X_j\| > \varepsilon \end{cases} \quad i, j = 1, \dots, N \quad (2.10)$$

Où  $N$  est le nombre d'états considérés et  $\|X_i - X_j\| \leq \varepsilon$  indique l'égalité avec une certaine erreur  $\varepsilon$ . Ainsi, l'équation (2.10) peut être réécrite sous sa forme matricielle :

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|X_i - X_j\|) \quad i, j = 1, \dots, N \quad (2.11)$$

Où  $N$  est le nombre de points  $X_i$  mesurés,  $\varepsilon$  est le seuil,  $\Theta(\cdot)$  est la fonction échelon ( $\Theta(x) = 0$  si  $x < 0$ , et  $\Theta(x) = 1$  autrement) et  $\|\cdot\|$  est une norme.

Pour un  $\varepsilon$  fixé (pour les états voisins de  $\varepsilon$ ), on introduit la notion suivante :

$$X_i \approx X_j \Leftrightarrow R_{i,j} = 1 \quad (2.12)$$

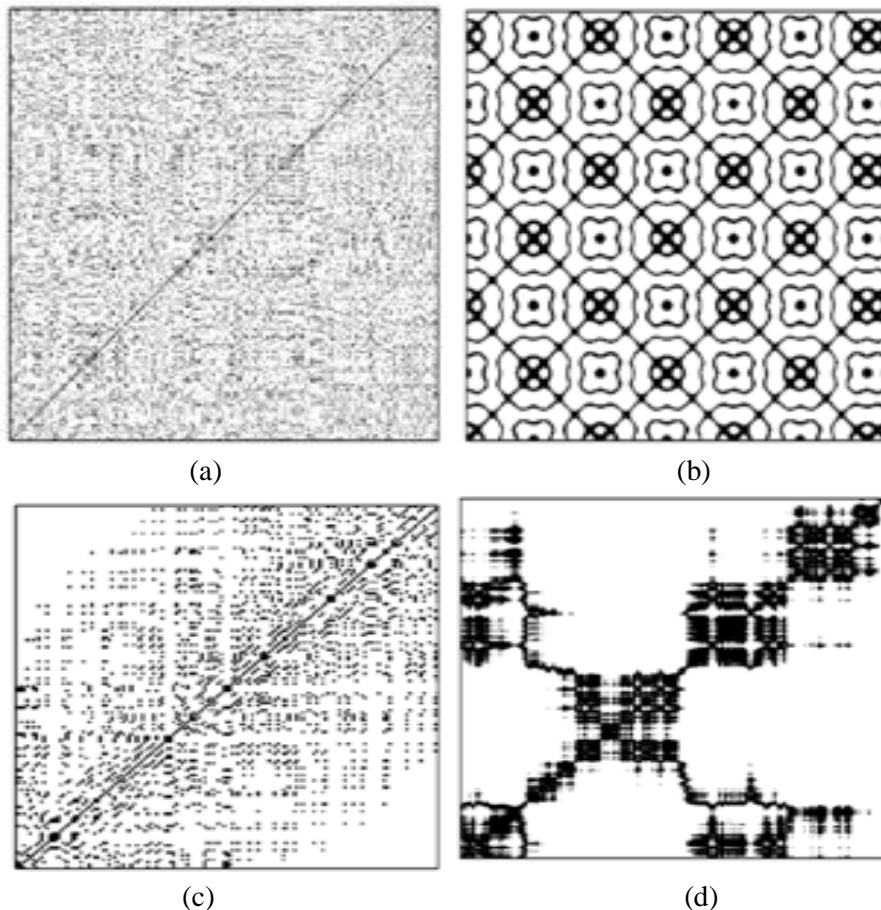
Puisque  $R_{i,j} = 1$  par définition, le diagramme de récurrence se caractérise par une ligne noire diagonale, appelé « la ligne d'identité » (Line Of Identity, LOI), avec un angle de  $\pi/4$ .

Le paramètre  $\varepsilon$  est crucial lors d'un diagramme de récurrence. Représente la distance seuil qui caractérise l'étendue du voisinage entre deux vecteurs et  $\Theta(x)$  fonction d'Heaviside. Ainsi, une attention spéciale doit lui être accordée. Si la valeur de  $\varepsilon$  est trop grande, la plupart des points seront voisins d'autres points, ce qui produira un diagramme condensé, et donc des points non pertinents. D'autre part, si la valeur de  $\varepsilon$  choisie est trop petite, on peut perdre la plupart des points récurrents, donc aucune information utile ne sera tirée. Par ailleurs, il n'existe pas de méthode standard pour choisir la valeur de  $\varepsilon$ . La plupart des chercheurs la choisissent comme étant égale à  $0.25 \sigma$ ,  $\sigma$  représente la déviation standard du signal étudié qui est le signal glottique dans notre cas [55].

### 4.3 Différentes structures dans le diagramme de récurrence RP

Comme déjà mentionné, l'objectif des diagrammes RP était de visualiser les trajectoires dans l'espace des phases, ce qui est particulièrement avantageux dans le cas de systèmes de grande dimension. Les structures RP fournissent des informations importantes sur l'évolution et les comportements temporels de ces trajectoires, car les modèles typiques des RP sont liés à un comportement spécifique du système. Les typologies des structures RP, peuvent être classés en : homogènes, périodiques ou quasi périodique, dévié et diagramme interrompu :

- Les diagrammes de récurrence homogènes sont typiques des systèmes stationnaires dans lesquels les temps de relaxation sont courts par rapport au temps écoulé par le RP. Un exemple d'un tel diagramme est celui d'une série temporelle aléatoire stationnaire ;
- Les diagrammes de récurrence des systèmes périodiques ou quasi-périodiques ont des structures récurrentes orientées d'une façon diagonales, périodiques ou quasi-périodiques (lignes diagonales, structures de forme en damier). La figure 5.b, montre le RP d'un système périodique avec deux fréquences harmoniques et avec un rapport de fréquence de quatre (deux et quatre lignes courtes se trouvent entre les lignes diagonales continues) ;



**Figure 2.10.** Topologie caractéristique des diagrammes de récurrence  
(a) Homogène, (b) périodique, (c) dévié et (d) interrompu

- Les diagrammes RP dévié : Une déviation est causée par des systèmes dont les paramètres varient lentement, c'est-à-dire des systèmes non stationnaires ;

- Diagramme RP interrompu : caractérisés par des changements brusques de la dynamique ainsi que des événements extrêmes (figure 5.d) Dans ces cas, les RP peuvent être utilisés pour trouver et évaluer des événements extrêmes et rares en notant la fréquence de leurs répétitions ;

Le RP est une structure en miroir dans tous les cas, c'est-à-dire que la courbe dans le quadrant inférieur droit est reproduite dans le quadrant supérieur gauche. La diagonale principale existe dans toutes les types de structures, ce qui correspond au fait que la distance entre le vecteur et la diagonale elle-même est nulle. Ainsi, le RP permet une analyse bidimensionnelle du comportement d'un système dynamique en permettant de vérifier le retour dans la trajectoire de l'espace des phases. Une forte densité de points et les segments blancs expriment dans notre cas la caractéristique non linéaire du signal étudié. Un « point récurrent » signifie que la distance entre deux vecteurs voisins dans l'espace de phase est inférieure ou égale à la valeur du seuil  $\varepsilon$ . Le point à la plus haute densité indique que la distance entre les vecteurs reconstruits dans l'espace de phase change légèrement. Il indique que les vecteurs incrustés sont très similaires aux vecteurs voisins dans  $m$ -dimensions d'espace de phase [56].

**Tableau 2.2.** Caractéristiques de diagramme de récurrence RP et Ses interprétations

Exemples	Signification
(1) Homogénéité	Il s'agit d'un processus stationnaire.
(2) Eclaircissement des coins haut gauche et bas droit	Données non stationnaires ; le processus contient un mouvement ou un décalage.
(3) Disruptions (bandes blanches)	Données non stationnaires ; certains état sont rares ; présence de transitions.
(4) Modèle périodique/ quasi-périodique	Processus cyclique ; la période correspond à la distance entre les lignes ; différentes distances entre les lignes diagonales révèlent un processus quasi-périodique.
(5) Points isolés	Forte fluctuation du processus ; si seuls les points isolés sont présents, le processus étudié est faiblement ou non corrélé.
(6) Points denses	La densité des points indique l'intensité du processus étudié. Si les points gagnent en densité, l'activité du processus étudié s'intensifie.
(7) Lignes diagonales (parallèles à la LOI)	L'évolution est similaire à d'autres périodes ; le processus peut être déterministe. Si ces lignes diagonales sont présentes à côté des points isolés, le processus peut être chaotique.
(8) Lignes diagonales (orthogonale à la LOI)	L'évolution est similaire à d'autres périodes mais en temps inverse ; parfois, cela indique une dimension d'immersion $m$ insuffisante.
(9) Lignes verticales et horizontals	Certains états présentent des changements peu nombreux ou lents ; indiquent des états laminaires.
(10) Lignes longues arquées	L'évolution est similaire à d'autres périodes mais à une vitesse différente ; la dynamique du système peut être variable.

Les points à faible densité représentent un dispersement évident dans le processus. Tandis que les espaces blancs indiquent un changement rapide de l'amplitude. Nous résumons les différentes significations des topologies des structures du diagramme de récurrence dans le tableau 2.2.

#### 4.4 Analyse Quantitative du diagramme de récurrence (Recurrence Quantification Analysis) RQA

L'interprétation visuelle des diagrammes de récurrence nécessite une certaine expérience pour évaluer le comportement d'un système. Cependant, les affichages graphiques avec une résolution insuffisante pour afficher les diagrammes RP présentent l'inconvénient que les utilisateurs sont obligés de intuitivement subjectivement et d'interpréter les motifs et les structures présentées dans le tracé de récurrence. Et différents observateurs voient les choses différemment. De ce fait, au début des années 1990, Zbilut et Webber ont introduit des définitions et des procédures pour quantifier les structures RP appelé « analyse quantitative de récurrence » ou (Recurrence Quantification Analysis, RQA) en anglais [57]. Cet outil est défini par un ensemble de paramètres de récurrence qui fonctionnaient comme des mesures de la complexité basées sur la densité des points récurrence et la structuration des lignes diagonales et verticales dans les tracés RP. Ceci permet de voir les transitions entre les états périodiques, laminaires et chaotiques lors d'un processus donné. Dans ce travail les paramètres d'analyse quantitative de récurrence seront utilisés pour mesurer la complexité et la non-linéarité du signal glottique.

#### 4.5 Les Mesures RQA

Le tracé de récurrence de chaque système dynamique a sa propre topologie [58]. Cette topologie peut être quantifiée par RQA. Les mesures de RQA peut être divisé en deux catégories : les mesures basées sur les lignes diagonales et celles basées sur les lignes verticales.

Nous introduisons l'histogramme :  $P(\varepsilon, l)$ , représentant le nombre total des lignes diagonales de longueur  $l$  dans un diagramme RP avec un seuil  $\varepsilon$  fixé :

$$P(\varepsilon, l) = \sum_{i,j=1}^N (1 - R_{i-1,j-1}(\varepsilon))(1 - R_{i+1,j+1}(\varepsilon)) \prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon) \quad (2.13)$$

Par analogie, le nombre de lignes verticales est calculé en remplaçant  $l$  par  $v$ , où  $v$  représente la longueur de la ligne verticale. Dans les prochaines équations, pour une raison de simplicité  $P(\varepsilon, l)$  sera remplacé par  $P(l)$  dans tout le texte ( $P(\varepsilon, l) = p(l)$ ). Nous omettrons de citer le symbole  $\varepsilon$  des mesures de RQA. Huit paramètres RQA du signal glottique sont étudiés.

- **Taux de Récurrence (Recurrence Rate (RR))** : RR mesure la densité des points sur la structure RP. Exprimé par le rapport entre tous les états récurrents (points noirs) de la structure RP et tous les états possibles. Par conséquent, cette métrique représente la



probabilité qu'un certain état se reproduise pour détecter les changements dans la dynamique du système. Le RR est définie par :

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j(i \neq j)}(\varepsilon) \quad (2.14)$$

- **DEterminisme (DET)** est représenté par la fraction entre le nombre de points de récurrence formant des structures diagonales (de longueur minimale  $l_{min}$ ) et tous les points  $R_{ij}$  formant la structure RP. Le déterminisme fournit une indication sur la prédictibilité du processus étudié, il est donné par l'équation :

$$DET = \frac{\sum_{l=l_{min}}^N l \times P(l)}{\sum_{l=1}^N l \times P(l)} \quad (2.15)$$

- **LAMinarité (LAM)** : de façon analogue à la définition du déterminisme défini par l'équation 2.15, la laminarité est le rapport entre les points de récurrence formant les structures verticales et l'ensemble des points de récurrence. Il décrit la quantité de points récurrents obtenus à partir des lignes verticales. Des faibles valeurs de laminarité indiquent que la structure RP comprend plus de points de récurrence uniques que de lignes verticales [59]. Donnée par l'expression :

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (2.16)$$

Où :  $P(v)$  est la distribution de fréquence des lignes verticales de longueur  $v$ , qui ont au moins une longueur de  $v_{min}$ .  $N$  est la longueur du signal.

- **Longueur de ligne diagonale moyenne** : La longueur de ligne diagonale moyenne est le temps moyen que deux segments de la trajectoire sont proches l'un de l'autre et peut être interprétée comme le temps de prédiction moyen. L'équation (2.17) exprime la formule de la longueur de ligne diagonale moyenne :

$$L = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (2.17)$$

- **Longueur maximale des lignes diagonales ( $L_{max}$ )** correspond à la plus longue ligne diagonale dans la structure RP. L'inverse de ( $L_{max}$ ) est lié à la divergence exponentielle. Il est défini par :

$$L_{max} = \max( l_i, i = 1, \dots, N_l ) \quad (2.18)$$

$$Div = \frac{1}{L_{\max}} \quad (2.19)$$

Dans le cas des systèmes déterministes, de longues diagonales sont produites, alors que pour des données aléatoires ou fortement fluctuantes, il s'agit de courtes diagonales.

- **ENTropie (ENTR)** : L'entropie se réfère à l'entropie de Shannon de la distribution des lignes diagonales dans la structure de récurrence RP : elle exprime la complexité de la structure RP par rapport aux lignes diagonales. Pour des signaux faiblement corrélés, la valeur **ENTR** est faible, indiquant ainsi une faible complexité. Elle est donnée par l'équation 2.20.

$$ENTR = - \sum_{l=l_{\min}}^N P(l) \ln P(l) \quad (2.20)$$

Où :  $P(l)$  représente la probabilité de trouver une ligne diagonale ait exactement la longueur  $l$ , et peut être estimée à partir de la distribution de fréquence  $p(l)$ , avec :

$$P(l) = \frac{p(l)}{\sum_{l=l_{\min}}^N p(l)} \quad (2.21)$$

- **Trapping Time (TT)** représente la valeur moyenne temporelle pendant lequel le système restera dans un seul et même état. Il est défini par la formule (2.22) :

$$TT = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=v_{\min}}^N P(v)} \quad (2.22)$$

Où  $P(v)$  est la distribution de fréquence des lignes verticales de longueur  $v$ , qui ont en moins une longueur  $v_{\min}$ .  $TT$  Estime la moyenne temporelle dans laquelle le système demeure dans un même état.

- **Récurrence Time Entropie (RTE)**. Le **RTE** représente la périodicité ou la répétitivité d'un signal. Il peut être appliqué aux systèmes dynamiques, ainsi qu'aux processus stochastiques. Il se calcule via l'équation (2.23) :

$$RTE = -(\ln T_{\max})^{-1} \sum_{t=1}^{T_{\max}} p(t) \cdot \ln p(t) \quad (2.23)$$

Où  $T_{\max}$  est la plus grande période du signal. Le **RTE** est très utilisé pour détecter les changements subtils dans les dynamiques des signaux stochastiques.

## 5. Evaluation de la voix pathologiques basée sur l'analyse quantitative de récurrence croisée appliqué au signal glottique

Cette partie consiste notre deuxième contribution ou nous avons appliqué l'analyse quantitative de récurrence croisé aux signaux glottiques d'un échantillon d'une voix normale et un échantillon pathologique du signal de la voyelle longue [a:].

### 5.1 Tracé de récurrence croisé (Cross Recurrence Plots (CRP)) :

Les tracés de récurrence croisé ou « Cross Recurrence Plots (CRP) est une extension du technique RP utilisés pour étudier l'évolution des dépendances et des similitudes entre deux systèmes différents en comparant leurs états. Elle repose sur le comportement de la récurrence de la trajectoire de phase des deux signaux [60]. Pour rendre le concept plus clair, en considérant deux signaux mesurés  $x(t)$  et  $y(t)$  avec des longueurs  $N_x$  et  $N_y$  respectivement et  $t$  comme intervalle d'échantillonnage. Dans le cas de notre contribution, les deux signaux  $x(t)$  et  $y(t)$  représentent la série temporelle du signal glottique d'un sujet de voix normal et un sujet de voix pathologique. La série de  $N$  échantillons est donnée comme suit :

$$NS(n) = [x(1), x(2), \dots, x(N_x)] \quad (2.24)$$

$$PS(n) = [y(1), y(2), \dots, y(N_y)] \quad (2.25)$$

Où : NS et PS sont les représentations du signal glottique : Normal & pathologique. Par l'espace de phase  $X, Y$  de dimension  $N$  et la méthode d'intégration de temporisation, les signaux glottaux peuvent être exprimés comme suit :

$$X_{NS}(t_i) = [x(t_i), x(t_i - \tau), \dots, x(t_i - (m-1)\tau)] \quad (2.26)$$

$$Y_{PS}(t_i) = [y(t_i), y(t_i - \tau), \dots, y(t_i - (m-1)\tau)] \quad (2.27)$$

Dans lesquelles,  $m$  est la dimension de l'espace de phase,  $\tau$  est un retard et  $i = 1; 2; \dots; N$  est l'indice d'échantillon. La comparaison du degré de similitude entre ces deux vecteurs conduit à la matrice de similitude des deux signaux correspond à la structure CRP définie comme :

$$d(i, j) = Sim(X_{NS}(i), Y_{PS}(j)) \quad (2.28)$$

Le structure CRP est ensuite obtenu par la comparaison de chaque coefficient de la matrice de similitude à un seuil défini par :

$$CRP_{i,j}^{X_{NS}, Y_{PS}} = \Theta(\varepsilon - \|X_i - Y_j\|) \quad i, j = 1, 2, \dots, N \quad (2.29)$$

Où,  $\theta$  représente la fonction Heaviside et  $\varepsilon$  est un seuil (threshold) choisi de manière appropriée.

Le but principal de la rééducation orthophonique est de récupérer les automatismes normaux du comportement de production vocale en diminuant l'effort et le stress vocal afin d'établir

une rétroaction auditive appropriée. De nos jours, les médecins ORL, les laryngologistes et les orthophonistes approuvent l'importance de la rééducation vocale et l'appliquent à plusieurs troubles de la voix. C'est le choix le plus souvent réalisé. Dans le contexte nos contributions dans ces travaux de thèse, nous appliquerons l'analyse par quantification récurrence croisé des signaux glottiques pour évaluer, analyser les pathologies des cordes vocales avant et après une rééducation orthophonique. Cette technique a présenté un bon outil servant à évaluer la voix pathologique dans différents stages de traitement.

## **5.2 Analyse quantitative de récurrence croisée**

L'analyse quantitative de récurrence croisée (**Cross Recurrence Quantification Analysis : CRQA**) est une extension de l'analyse quantitative de récurrence, un outil basé sur la description statistique de la distribution des lignes parallèles et des points dans le CRP. Les mesures de complexité définies à l'aide de la densité des points de récurrence et des structures de lignes diagonales dans les cartes CRP fournissent une description qualitative de la dynamique de la série temporelle du signal glottique étudiée [61]. Dans le cadre de notre étude, huit paramètres ont été suggérés considérés comme les paramètres les plus pertinents dans plusieurs classifications.

## **6. Conclusion**

Dans ce chapitre, nous avons présenté les techniques d'analyse paramétrique les plus utilisées dans le domaine de la reconnaissance automatique de la voix pathologique. Dans le but de diagnostiquer et analyser les pathologies des cordes vocales, l'analyse approfondie du signal glottique à travers la méthode RQA a été établie. Cette méthode a démontré une efficacité indubitable pour l'étude de la dynamique de variation de différents types de signaux tels que le signal ElectroCardioGramme. Huit indicateurs de la RQA ont été décrits, ses variations seront examiner dans chaque pathologies pour évaluer l'aspect discriminatif à travers les techniques de classification. De ce fait, Dans le chapitre suivant nous décrivons les techniques de classification utilisées dans notre système.

*Chapitre 3 :*  
*Techniques de Classification et*  
*d'Optimisation*

## 1 Introduction

Les techniques d'apprentissage pour la classification automatique sont une branche de l'intelligence artificielle (IA). Ils sont appliqués dans de nombreux domaines en particulier le domaine biomédicale en raison de leur capacité à traiter tous types de données. Dans le cadre de notre travail, des techniques de classification intelligentes ont été introduites en vue de l'identification de certains types de pathologies affectant les cordes vocales en se basant sur un modèle paramétrique déterminés par l'analyse quantitative de récurrence. Un algorithme d'optimisation a été développé aussi pour améliorer l'exactitude de la classification.

Ce chapitre est divisé en deux parties. Nous présentons en premier lieu les fondements théoriques des trois algorithmes de classification que nous avons utilisés, à savoir : le classifieur machine à vecteurs de supports (Support Vector Machines « SVM »), le classifieur « k plus proches voisins » (K Nearest Neighbors « KNN ») et la classification par réseaux Bayésien (Naïve Bayes Networks) en détaillant le processus de classification ainsi que les avantages et les inconvénients des différentes techniques. La deuxième partie est consacrée à la présentation de la technique d'optimisation par essaim particulaire (Particle Swarm Optimization « PSO ») dont nous allons introduire pour ajuster les paramètres du classifieur SVM.

## 2 Classifieur Machines à Vecteurs Supports

Les machines à vecteurs supports ou Support Vector Machines (SVM) en anglais, sont des techniques discriminantes dans la théorie de l'apprentissage statistique supervisé. Ces techniques sont destinées à résoudre des problèmes de discrimination et de régression traitant des données de grande dimension. Elles ont été proposées en 1995 par V. Vapnik dans son livre "The nature of statistical learning theory" [62]. Elles permettent d'aborder plusieurs problèmes divers et variés comme la régression, la classification, la fusion...etc. L'origine des machines à vecteurs de support (SVM) remonte à 1975, lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la dimension VC (Vapnik et Chervonenkis) pour caractériser la capacité d'une machine d'apprentissage. Mais leur dénomination par SVM n'est apparue qu'en 1995 avec Cortes et al. [63]. Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent. SVM est une méthode particulièrement bien adaptée pour traiter des données de très haute dimension. Elles ont été utilisées, entre autres, dans le diagnostic médical (analyse de séquences d'images IRM, détection d'arythmie cardiaque, évaluation des risques de cancer, ...), dans la reconnaissance de forme (traitement d'image et de la parole), etc. [64].

Les classifieurs est un algorithme qui a été développé initialement pour traiter le problème de la classification deux groupes (appelée également classification binaire supervisé) et très vite les SVM se sont étendus pour des problèmes de natures diverses multiples ou multi-classe. Les Classifieurs SVM génèrent un problème d'optimisation quadratique. Résoudre le problème d'optimisation quadratique conduit à trouver l'hyperplan

de séparation entre deux classes avec une marge maximale. Les données peuvent être linéairement ou non linéairement séparables.

## 2.1 Classifieurs SVM pour classification binaire

### 2.1.1 Cas de données linéairement séparables

Nous disposons donc d'un ensemble de données d'apprentissage  $T = \{x_i, y_i\}_i^m$ , où  $x_i$  est le vecteur d'entrée,  $y_i \in \{1, -1\}$  désigne l'étiquette de la classe de  $x_i$  et  $m$  le nombre total de données, On suppose qu'il existe un hyperplan qui sépare les points positifs et négatifs.

Dans un modèle linéaire, on a :

La fonction de décision  $f(x)$  de SVM peut être écrite comme suit [63] :

$$f(x_i) = w \cdot x_i + b \quad (3.1)$$

L'hyperplan séparateur (frontière de décision) a donc pour équation :  $w \cdot x_i + b = 0$

Où  $w$  est un vecteur orthogonal à l'hyperplan et  $b/\|w\|$  est la distance de l'hyperplan à l'origine de l'espace. La distance d'un point (exemple) est donnée par :

$$d(x_i) = |w \cdot x_i + b| / \|w\| \quad (3.2)$$

Notre objectif est la construction de l'Hyperplan Optimal (HO) qui sépare des données appartenant à deux classes différentes linéairement séparables.

Soit  $H : (w \cdot x_i) + b$  : l'hyperplan qui satisfait les conditions suivantes :

$$\begin{aligned} f(x_i) \geq 1 & \quad \text{si } y_i = 1 \Rightarrow x_i \text{ appartient à la classe } 1. \\ f(x_i) \geq -1 & \quad \text{si } y_i = -1 \Rightarrow x_i \text{ appartient à la classe } -1. \end{aligned}$$

Ces contraintes peuvent être fusionnées, comme suit :

$$y_i (<w \cdot x_i > + b) \geq 1 \quad \text{pour } i = 1, 2, \dots, m \quad (3.3)$$

Comme nous l'avons déjà mentionnée, un HO est un hyperplan qui maximise la marge  $M$  qui représente la plus petite distance entre les différentes données des deux classes et l'hyperplan. Maximiser la marge  $M$  est équivalent à maximiser la somme des distances des deux classes par rapport à l'hyperplan. Figure 3.1. Ainsi, la marge a l'expression mathématique suivante :  $M = 2 / \|w\|$ . Trouver l'hyperplan optimal revient donc à maximiser  $2 / \|w\|$ . Ce qui est équivalent à minimiser  $\|w\|^2 / 2$  sous la contrainte (3.3). Ceci est un problème de minimisation d'une fonction objective quadratique avec contraintes linéaires. L'hyperplan de séparation doit satisfaire les contraintes suivantes :

$$\begin{cases} y_i (\langle w \cdot x_i \rangle + b) \geq 1 & i = 1, 2, \dots, m \\ \min \frac{1}{2} \|w\|^2 \end{cases} \quad (3.4)$$

$$\|w\|^2 = w^T \cdot w \quad (3.5)$$

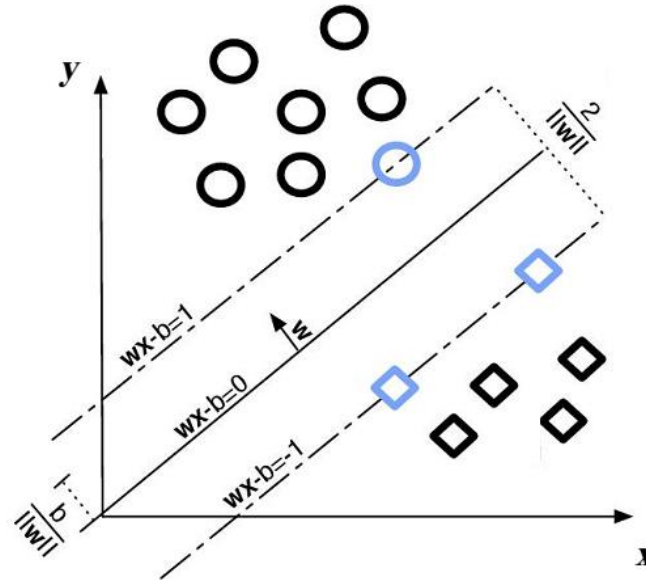


Figure 3.1. Séparation linéaire de deux classes par SVM

Le principe de Lagrange est utilisé Pour résoudre un problème d'optimisation sous contrainte pour la recherche de l'hyperplan optimale. Ceci permet de transformer l'équation (3.4) en :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{k=1}^m \alpha_k [y_k (\langle w \cdot x_k \rangle + b) - 1] \quad (3.6)$$

$\alpha_i$  sont les coefficients de Lagrange ( $\alpha_i > 0$ ).

Selon la condition d'optimalité :

$$\begin{cases} \frac{\partial L(w, b, \alpha)}{\partial w} = 0 \\ \frac{\partial L(w, b, \alpha)}{\partial b} = 0 \end{cases} \quad (3.7)$$

La résolution des équations (3.6) permet d'obtenir :



$$\begin{cases} w = \sum_{k=1}^m \alpha_k \cdot x_k \cdot y_k \\ \sum_{i=1}^m \alpha_i \cdot y_i = 0 \end{cases} \quad (3.8)$$

Un tel problème peut être résolu en utilisant la représentation double de Wolfe [65] :

$$\begin{cases} \text{maximiser } L(w, b, a) \\ \text{sujet à } w = \sum_{k=1}^m a_k \cdot x_k \cdot y_k \\ \sum_{k=1}^m a_k \cdot x_k = 0 \\ \forall k, a_k \geq 0 \end{cases} \quad (3.9)$$

Ce système d'équations est équivalent à :

$$\begin{cases} \text{maximiser } \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \langle x_i, x_j \rangle \\ \text{sous les contraintes, } a_i \geq 0, \text{ et } \sum_{i=1}^m a_i \cdot x_i = 0 \end{cases} \quad (3.10)$$

### 2.1.2 Cas de données non linéairement séparables

Dans le cas où les données sont non-linéairement séparables (Figure 3.2), l'hyperplan optimal est celui qui satisfait les conditions suivantes :

- La distance entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale.

Introduisons la marge d'erreur ( $\xi_i$ ) dans l'équation de l'hyperplan (3.10), comme suit :

$$\begin{cases} y_i (\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i & i = 1, 2, \dots, m \\ \forall i, 0 < \xi_i < 1 \end{cases} \quad (3.11)$$

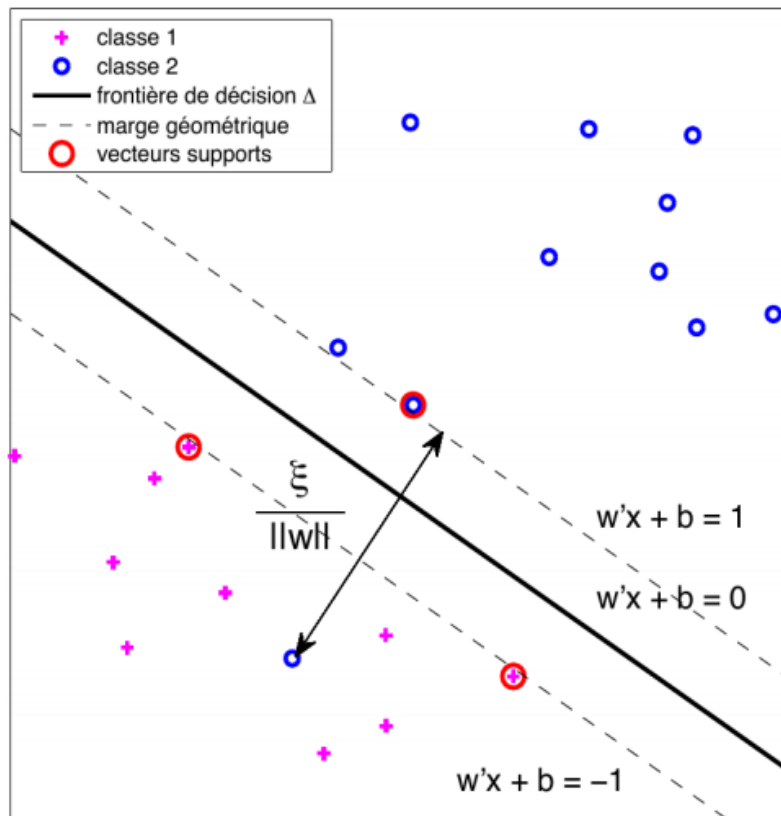


Figure 3.2. Séparation non linéaire de deux classes par SVM

Où « C » un coefficient de pénalisation du relâchement (le coût attribué à un échantillon mal classé). Le problème d'optimisation devient :

$$\begin{cases} \text{minimiser} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i \\ \text{sujet à} & y_i (w \cdot x_i + b) \geq 1 - \zeta_i \\ & \forall i, 0 < \zeta_i < 1 \quad i = 1, 2, \dots, m \end{cases} \quad (3.12)$$

Dans ce cas, dans le cas non linéairement séparable, la SVM peut ne pas trouver aucun hyperplan séparable. Une stratégie de prétraitement peut être utilisée pour simplifier la tâche. Il s'agit de changer l'espace original d' « attributs » en un nouvel espace appelé de « caractéristiques », ce qui implique de trouver une fonction  $\phi$  qui représente les points originaux dans un espace de dimension beaucoup plus grande ou il sera plus facile de trouver un séparateur linéaire, comme montré sur la figure 3.3.

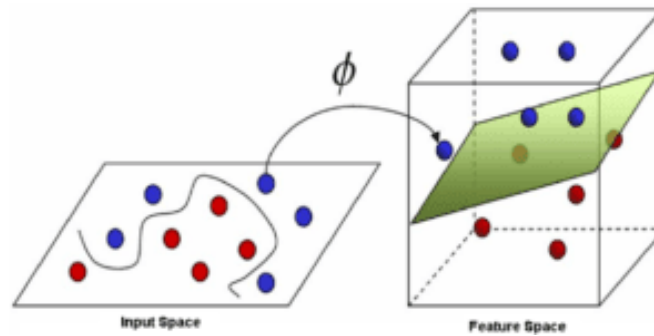


Figure 3.3. Changement d'espace

L'introduction de cette fonction a été par Boser et *al.* [66]. Le but de cette idée est de placer les observations dans un espace de Hilbert  $T$  de dimension plus élevée. L'espace  $T$  ainsi obtenu est appelé espace des caractéristiques ou espace transformé. En appliquant la transformation hermitienne sur le système d'équation (3.9), le problème d'optimisation devient :

$$\left\{ \begin{array}{l} \text{maximiser } \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \\ \text{sous les contraintes : } \forall i, 0 \leq \alpha_i \leq C, \text{ et } \sum_{i=1}^m \alpha_i x_i = 0 \end{array} \right. \quad (3.13)$$

Cependant la détermination de la fonction de transformation  $\phi$ , cette transformation pourrait être compromise si la dimension de l'espace de  $\phi(x)$  est très importante. Pour résoudre ce problème, Aizerman [67] a introduit une méthode dite "astuce de noyau" (en anglais : kernel trick). En effet, sous certaines hypothèses sur  $\phi$ , le produit scalaire  $\langle \phi(x_i), \phi(x_j) \rangle$  peut être calculé à l'aide d'une fonction symétrique  $k$  (noyau). Le noyau  $k$  doit satisfaire les conditions de Mercer [68] ( $k$  doit être symétrique et semi-définie positive).

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (3.14)$$

Il suffit donc de remplacer le produit scalaire par la fonction noyau non linéaire. De ce fait, les classifieurs SVM peuvent trouver un hyperplan séparateur linéaire. Le système (3.12) se transforme alors sous la forme suivante :

$$\left\{ \begin{array}{l} \text{maximiser } \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{sous les contraintes : } \forall i, 0 \leq \alpha_i \leq C, \text{ et } \sum_{i=1}^m \alpha_i x_i = 0 \end{array} \right. \quad (3.15)$$

Tel que "  $C$  " est le paramètre de marge, des faibles valeurs indique une bande de décision avec une grande marge et une petite marge pour les valeurs élevées. Le paramètre Gamma contrôle la distance d'influence des points uniques des vecteurs d'apprentissage.

Tel que, des valeurs faibles indiquent une grande similitude des radios, ce qui entraîne le regroupement de plus de points. Pour des valeurs élevées, les points seront très proches les uns des autres pour être considérés dans le même groupe.

En pratique, quelques familles de fonctions de noyau paramétrables sont couramment utilisées. Il revient simplement à l'utilisateur d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. Dans la littérature, on trouve les fonctions noyaux suivantes [71] :

- La fonction linéaire :  $K(x, y) = x \cdot y'$

- La fonction polynomiale :  $K(x, y) = (c + x \cdot y')^d$

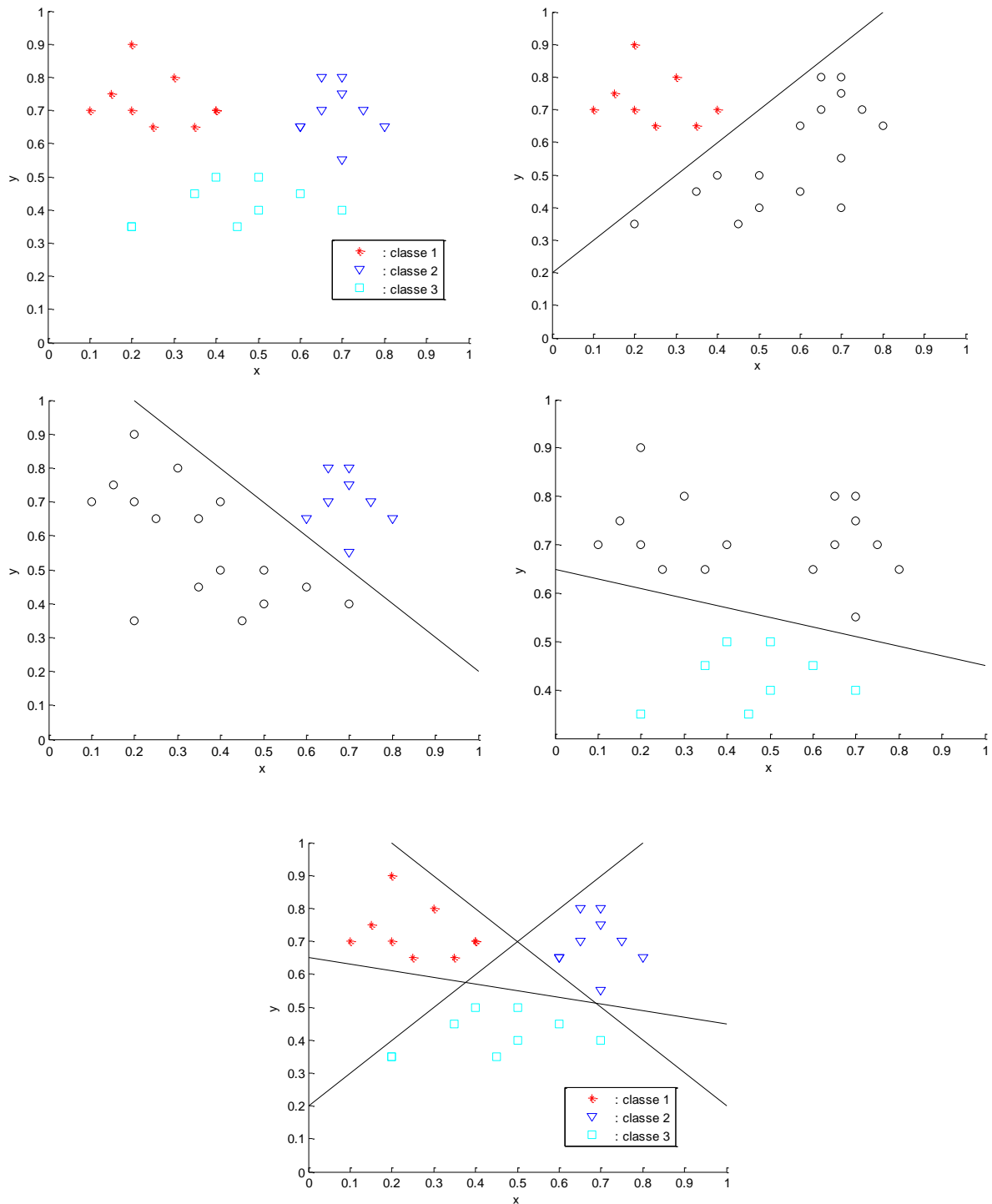
- La fonction à base radiale (Gaussienne) :  $K(x, y) = \text{Exp}\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$

- La fonction sigmoïde :  $K(x, y) = \tanh(\alpha_0(x \cdot y') + \beta_0)$

## 2.2 Classifieurs SVM multi-classes

La plupart des problèmes ne se contentent pas de deux classes de données. Il existe plusieurs méthodes pour faire la classification multiclassées. Citons : La première méthode est une méthode dite « Un-contre-Un » (one-against-one). Au lieu d'apprendre N fonctions de décisions, ici chaque classe est discriminée d'une autre. La deuxième méthode est appelé « Un-Contre-Tous » (one-against-all). C'est une approche étendant la notion de marge aux cas multiclassées. Cette formulation intéressante permet de poser un problème d'optimisation unique. Le problème fait intervenir N fonctions de décision. En comparant ces trois stratégies, il s'est avéré que "un-contre-un" produit une meilleure exactitude et exige moins de nombre de vecteurs de soutien, et par conséquent temps inférieur de calcul [69]. De ce fait, cette stratégie a été adoptée dans cette présente étude. Dans ce qui suit, nous présentons le processus d'apprentissage utilisant la stratégie un-contre-un.

La méthode un-contre-un a été initialement présentée par Knerr et al [70]. Cette technique a été associée aux SVM pour la première fois par Mayoraz et Alpaydin [71]. Cette stratégie construit  $w(w-1)/2$  classifieurs ( $w$  étant le nombre de classes) où chacun est formé par des données de deux classes. Donc il faut trouver  $w(w-1)/2$  fonctions de décisions. La solution du problème multi-classes est de le reformuler afin d'obtenir une classification binaire (en deux classes). Chaque classifieur sépare les données en deux catégories ; la première comporte une seule classe, les autres données sont fusionnées pour former la deuxième classe. La figure 3.4 illustre les différents scénarios qui en résultent, pour un problème de 3 classes. En fait, la stratégie un-contre-un représente un cas spécial des méthodes de décomposition proposée par Dietterich et al [72] pour résoudre le problème de multi-classes. La figure 3.4 montre comment trouver les fonctions de décisions étape par étape.



**Figure 3.4.** Principe un-contre-un pour  $w=3$

Pour un échantillon inconnu  $x_i$ , le classement est donné par le vote majoritaire ou par un graphe à cycle de décision (arbre de décision) si chaque classe reçoit le même nombre de votes.

### 2.3 Avantages et inconvénients des SVM

L'algorithme SVM offre une solution au problème sous forme d'une fonction convexe. Cette dernière contient, en fait, un seul optimum global, correspondant à la minimisation du risque structurel.

### 2.3.1 Avantage des SVM :

- Les SVM possèdent des fondements mathématiques solides.
- Les exemples de test sont comparés juste avec les supports vecteur et non pas  $\rightarrow$  avec tout les exemples d'apprentissage.
- Décision rapide. La classification d'un nouvel exemple consiste à voir le signe de la fonction de décision  $f(x)$ .

### 2.3.2 Inconvénients :

- Classification binaire d'où la nécessité d'utiliser l'approche un-contre-un.
- Grande quantité d'exemples en entrées implique un calcul matriciel important.
- Temps de calcul élevé lors d'une régularisation des paramètres de la fonction noyau.
- La sélection d'une mauvaise fonction noyau ou de mauvaises valeurs de ses paramètres peut conduire à un sur-apprentissage.

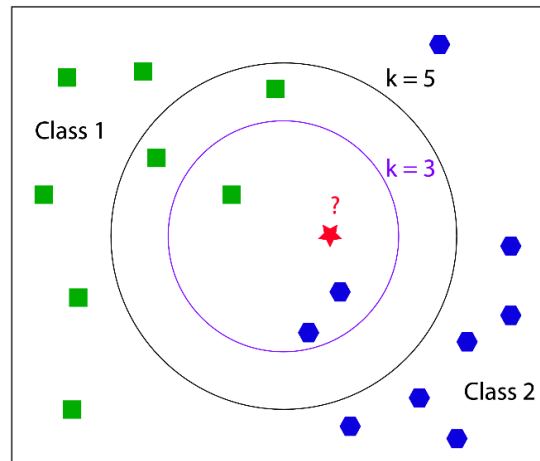
## 3 K-Nearest Neighbours (KNN)

Dans cette partie, nous abordons l'une des règles d'apprentissage les plus importantes pour les problèmes de classification, le classificateur k-plus proches voisins (k-Nearest Neighbours). Nous commençons par discuter brièvement de k-NN avec un exemple, nous donnons ensuite une formulation mathématique précise.

### 3.1 Principe

Le KNN est basé sur le calcul des distances entre les points d'échantillonnage aux voisins les plus proches de l'ensemble des points attribués [73]. La décision est basée sur le vote majoritaire des k voisins les plus proches.

Nous en illustrons un exemple dans la figure 3.5. Supposons que nous avons un ensemble de points dans un espace métrique  $\Omega$ , avec chaque point attribué une étiquette. Le choix de  $k=3$  (3-plus proche voisin) permet de classer la requête comme un hexagone bleu (les objets qui sont à l'intérieur du petit cercle en bleu). Quand même, la requête est classée comme un carré si l'on considère 5 – plus proches voisins (les objets qui sont à l'intérieur du grand cercle en ligne en noir). En effet, le choix du K (nombre de voisins) est un facteur essentiel dans le processus de classification.



**Figure 3.5.** Exemple de classification par kNN, la requête est un hexagone pour  $k=3$  et un carré pour  $k=5$

### Distances

La méthode des  $k$ -plus proches voisins (kNN) se base sur une comparaison directe entre le vecteur caractéristique de l'instance à classer et les vecteurs des instances de la base d'apprentissage. La comparaison consiste à un calcul de distances entre ces instances. Puis à l'instance à classer est assignée la classe majoritaire parmi les classes des  $k$  instances les plus proches. Il existe plusieurs distances employées par l'algorithme kNN pour comparer deux instances.

Notons par :

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$$

Le vecteur caractéristique de l'instance  $i$ , avec  $n$  le nombre de variables et par  $p$  et  $q$  deux échantillons à comparer.

$$\text{Distance Euclidienne} : D(Xp, Xq) = \sqrt{\sum_{i=1}^n (x_{pi} - x_{qi})^2}$$

$$\text{Distance de Manhattan} : D(Xp, Xq) = \sum_{i=1}^n |x_{pi} - x_{qi}|$$

$$\text{Distance de Tchebychev} : D(Xp, Xq) = \max_{i=1}^n |x_{pi} - x_{qi}|$$

### 3.2 Avantages et inconvénient du classifieur kNN

L'avantage de cette technique est qu'elle est facile à implémenter et adaptable à tout type de données, y compris celles qui sont plus ou moins complexes telles que des informations géographiques, le traitement d'image, parole... etc. Ainsi, elle est robuste au bruit. L'introduction de nouvelles données n'exige pas une reconstruction d'un modèle. Une fois les plus proches voisins exhibés. La classe est attribuée à un objet avec facilité et clarté.

Pour que le système de classification soit performant, il faut bien choisir le type de distance à impliquer ainsi que le nombre de voisins et mode dans lequel les réponses des

voisins sont combinées. Les résultats pourraient être de mauvaise qualité si le nombre d'attributs pertinents est faible relativement au nombre total d'attributs, car la proximité sur les attributs pertinents sera noyée par les distances sur les attributs non pertinents [74]. En l'absence d'apprentissage, tous les calculs se font dans la phase de classification, ce qui peut prendre énormément du temps.

Les principaux inconvénients de cette méthode sont le nombre d'opérations nécessaires pour classer une instance dans le cas d'une grande base d'apprentissage ainsi que sa sensibilité au bruit.

#### 4 Classifieur Naïve Bayes Network

Le classifieur Naïve Bayes Network est un algorithme de classification basé sur le théorème de Bayes. Il peut être utilisé pour résoudre des problèmes de diagnostics et prédiction. Les éléments de fonctionnalité d'une classe sont supposés être indépendants des autres valeurs d'attribut. L'attribution est de prédire la valeur des paramètres du modèle en termes de probabilité. En effet, la classification bayésienne est utilisée comme méthode d'apprentissage probabiliste [45].

##### 4.1 Construction d'un classifieur naïve Bayes network

La construction de classifieur est simple par l'appliquant de théorème de Bayes. Ceci concerne la probabilité conditionnelle et marginale de deux événements aléatoires. Supposant un ensemble de données  $D: \{(x_i, f(x_i)): i = 1, \dots, n\}$ , avec  $f(x_i) \in \{f_1, \dots, f_c\}$ .

$x_i$  : représente un échantillon avec ses variables caractéristiques  $x_i = (x_{i1}, \dots, x_{ip})$  ;

$f$  : représente une classe quelconque ;

$f(x_i)$  est la classe de l'échantillon  $x_i$  ;

$n$  dénote le nombre des échantillons et  $c$  le nombre de classes.

La décision du classifieur naïf de Bayes est définie par la probabilité postérieure  $P(f_k | x)$  donnée par l'équation suivante :

$$P(f_k | x_i) = \frac{P(x_i | f_k) P(f_k)}{P(x_i)} \quad k = 1, \dots, c. \quad (3.20)$$

$P(x_i | f_k)$  : la probabilité conditionnelle dont l'expression est :

$$P(x_i | f_k) = \prod_{k=1}^c P(x_i | f_k) \quad (3.21)$$

$P(f_k)$  : représente la probabilité antérieure de chaque classe qui peut être estimée selon deux possibilités :

$$P(f_k) = \frac{\text{le nombre des échantillons de la classe } k}{\text{le nombre total des échantillons}}$$



$P(f_k) = \frac{1}{c}$  C'est à dire que tous les classes possèdent la même probabilité antérieure. Pour les deux possibilités, on a :

$$\sum_{k=1}^c P(f_k) = 1 \quad (3.22)$$

$P(x_i)$  est la densité sans conditions qui normalise les postérieurs.

$$P(x_i) = \sum_{k=1}^c P(x_i | f_k) P(f_k) \quad (3.23)$$

La règle de décision consiste à choisir l'hypothèse la plus probable. De ce fait, il s'agit de la règle du maximum à posteriori ( $\max P(f_k | x)$ ) [75].

Si, on suppose, par exemple, que le calcul final de la probabilité postérieure, pour un problème de classification de trois classes  $w_1$ ,  $w_2$  et  $w_3$ , a donné l'inégalité suivante :  $P(f_2 | x) > P(f_3 | x) > P(f_1 | x)$ , l'échantillon  $x$  appartient, dans ces conditions, à la classe 3.

#### 4.2 Avantages et inconvénients

Nous citons ci-dessous quelques avantages du classificateur Naive Bayes :

- Il est simple et facile à mettre en œuvre.
- Il ne nécessite pas autant de données d'entraînement.
- Il gère à la fois les données continues et discrètes.
- Il est hautement évolutif avec le nombre de prédicteurs et de points de données.
- Il est rapide et peut être utilisé pour faire des prédictions en temps réel.
- Il n'est pas sensible aux fonctionnalités non pertinentes.

Etant donné sa simplicité, le classifieur bayésien exige relativement assez de données d'entraînement pour estimer les moyennes et les variances des différentes variables nécessaires à la classification. Or, son efficacité pourrait être compromise lors de la manipulation de bases de données vastes et/ou complexes. Par ailleurs, la classification bayésienne donne parfois de mauvais résultats si les données utilisées sont dépendantes entre elles, ou fortement corrélées.

#### 5. Optimisation par les méthodes de métaheuristiques

Lorsqu'une fonction possède plusieurs optimaux locaux, les techniques d'optimisation traditionnelles risquent de converger vers l'un de ces optimaux locaux. Les méthodes métaheuristiques (ou stochastiques) sont des algorithmes d'optimisation visant à résoudre des problèmes d'optimisation difficile (souvent issus des domaines de la recherche opérationnelle, de l'ingénierie ou de l'intelligence artificielle) pour lesquels on ne connaît pas de méthode classique plus efficace. Ces méthodes ne nécessitent ni le point de départ, ni la connaissance de la dérivée de la fonction objectif pour atteindre la solution optimale.

Le terme métaheuristique a été inventé par Fred Glover en 1986, lors de la conception de la recherche taboue [76]. Un problème d'optimisation se définit comme la recherche, parmi un ensemble de solutions possibles  $S$  (appelé aussi espace de décision ou espace de recherche), de la (ou des) solution(s)  $x^*$  qui rend (ent) minimale (ou maximale) une fonction mesurant la qualité de cette solution. Cette fonction est appelée fonction objectif ou fonction coût. Ces algorithmes manipulent une ou plusieurs solutions à la recherche de l'optimum [77]. La figure 3.6 illustre la différence entre un optimum globale et local ( $M_1$  optimum globale et  $M_2$  optimum local).

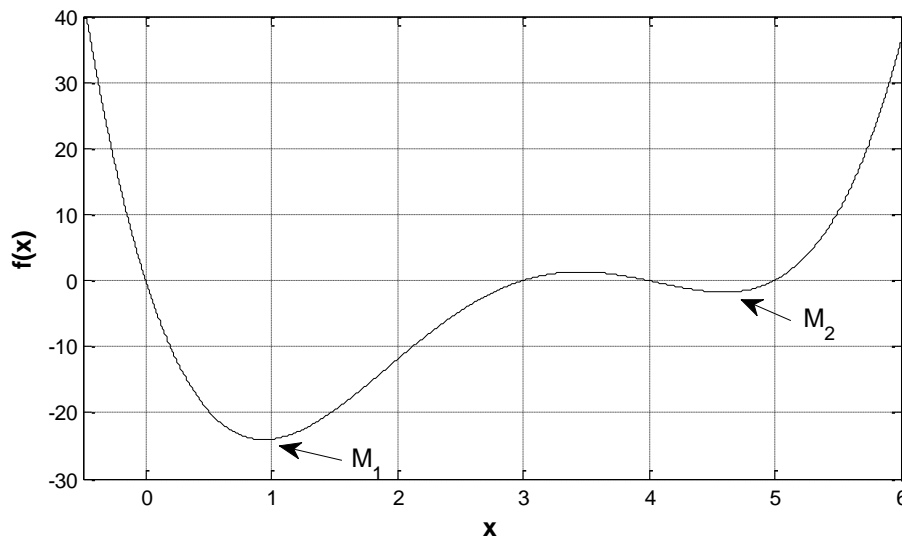


Figure 3.6. Minima global et local [77]

Les métaheuristiques sont inspirées souvent des processus naturels qui relèvent de la physique (algorithme du recuit simulé, par exemple), de la biologie (algorithmes génétiques) ou encore de l'éthologie (algorithmes de colonies de fourmis, optimisation par essaim particulaire ... etc.).

### 5.1 Algorithme d'Optimisation par Essaim de Particules (PSO)

L'intelligence en essaim peut être définie en tant que n'importe quelle tentative de conception des algorithmes ou des outils de résolution de problèmes distribués dont le comportement global émerge de l'interaction sociale entre les individus voisins. Elle a été initialement introduite en 1995 par Kennedy et Eberhart. L'intelligence en essaim est la simulation de l'interaction sociale entre des individus. Elle était destinée à simuler la migration et la collecte d'un troupeau d'oiseaux ou de poissons lorsqu'ils recherchent de la nourriture [78]. Chaque oiseau rappelle ses expériences de vol et partage les distances les plus courtes entre l'aliment et lui-même. Ce comportement social, appelé intelligence d'essaims, guide tous les oiseaux afin de déterminer la distance la plus courte à la nourriture.

## 5.2 Principe de la technique PSO

Un essaim est un ensemble de particules positionnées dans l'espace de définition de la fonction objectif. Le principe de l'algorithme consiste à trouver la solution optimale suite aux déplacements de ces particules dans l'espace de définition. Chaque particule est localisée par ces trois attributs, en l'occurrence, sa position, sa vitesse et son voisinage.

L'algorithme PSO a subi plusieurs améliorations. L'objectif est d'améliorer la qualité de la solution en réduisant au maximum le temps de calcul. Yann Cooren a proposé un nouveau mode d'initialisation de particules ; la distribution doit être régulière le long de l'espace de recherche. Cette nouvelle idée d'initialisation assure une couverture plus uniforme de particules. Une nouvelle stratégie hybride de déplacement de particules avec un algorithme d'estimation de distribution a été mise en œuvre pour garder la diversité au sein de l'essaim, tout au long du traitement [79].

Mendes et al ont proposé une nouvelle topologie dite FIPS (Fully Informed Partical Swarm). En outre de la meilleure voisine et la meilleure expérience propre à la particule, cette technique utilise également une partie des informations de chaque voisine [80].

Clerc et Kennedy ont proposé un model adaptatif de PSO qui utilise un nouveau paramètre " $\chi$ " appelé le coefficient de constriction. Cette idée sera éclairée dans la section suivante [81].

## 5.3 Configuration de l'algorithme PSO

L'algorithme PSO est configuré en trois étapes :

### 5.3.1 Taille de la population

Etant donné l'absence de règles relatives au choix du nombre de particules allouées à la résolution du problème, nous l'avons déterminé en fonction de deux paramètres, à savoir la taille de l'espace de recherche et le rapport entre les capacités de calcul de la machine et le temps maximum de recherche (ou temps d'exécution de l'algorithme). Ce dernier dépend du nombre de résolutions de la fonction objectif, correspondant au nombre de particules multiplié par le nombre d'itérations. De ce fait, pour réduire le nombre de résolutions de la fonction objectif, il faut réduire la taille de l'essaim ou le nombre d'itérations. Mais un trop petit nombre d'essaims risque de ne pas converger vers l'optimum global.

### 5.3.2 Initialisation de la position et la vitesse

Il faut définir l'espace de recherche des particules de la fonction objectif pour passer à l'étape de génération de la population. Les particules sont placées au hasard (aléatoirement) selon une distribution uniforme dans l'espace de recherche. Dans un espace de recherche de dimension  $d$ , la particule  $i$  de l'essaim est modélisée par son vecteur de position  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ , et par son vecteur vitesse  $V_i = (V_{i1}, V_{i2}, \dots, V_{id})$ .

### 5.3.3 Mise à jour de position et vitesse

La qualité de la position de la particule est déterminée par la valeur de la fonction objective en cette position (en ce point). Cette particule garde en mémoire la meilleure position par laquelle elle est déjà passée, que l'on note  $p_{best} = (p_{i1}, \dots, p_{i2}, \dots, p_{id})$ . Par ailleurs, la meilleure position atteinte par ses particules voisines est notée  $g_{best} = (g_{i1}, \dots, g_{i2}, \dots, g_{id})$ .

À chaque itération, les particules mettent à jour leurs positions et vitesses de déplacement en tenant compte non seulement de leurs meilleures positions, mais aussi celles de son voisinage. La figure 3.7 illustre le déplacement d'une particule.

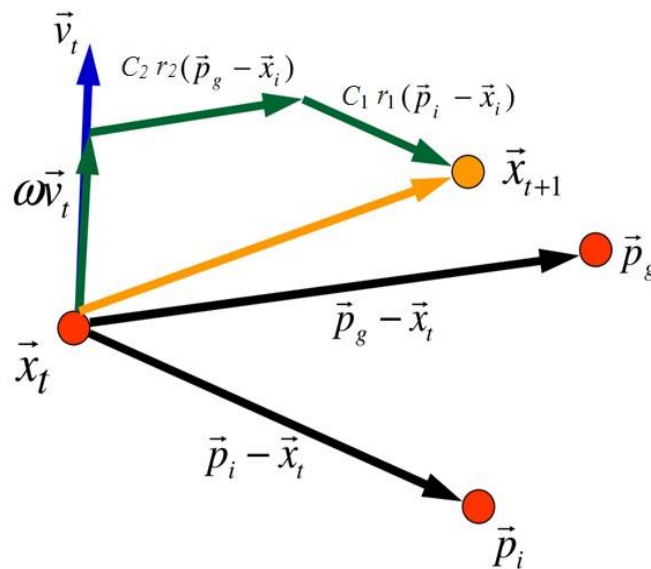


Figure 3.7. Principe du déplacement possible entre deux points [82]

La nouvelle vitesse est calculée à partir de la formule suivante [82] :

$$V_i(k+1) = w V_i(k) + c_1 r_1(k) (p_{best}(k) - X_i(k)) - c_2 r_2(k) (g_{best}(k) - X_i(k)) \quad (3.15)$$

Cette nouvelle vitesse permet la détermination de la position suivante comme suit :

$$X_i(k+1) = X_i(k) + V_i(k+1) \quad (3.16)$$

avec :  $X_i(k), X_i(k+1)$  : la position de la particule  $P_i$  à l'itération  $k$  et  $k+1$ , respectivement ;

$V_i(k), V_i(k+1)$  : les vitesses de la particule  $P_i$  à l'itération  $k$  et  $k+1$  respectivement ;

$P_{best}(k+1)$  : la meilleure position obtenue par la particule  $P_i$  à l'itération  $k+1$  ;

$g_{best}(k+1)$  : la meilleure position obtenue par l'essaim à l'itération  $k+1$  ;

$c_1, c_2$  : des constantes représentant les coefficients d'accélération ;

$r_1, r_2$  : des nombres aléatoires tirés de l'intervalle  $[0,1]$  ;

$w(k)$  : le poids inertiel.

Le pseudo-code de l'algorithme de PSO est donné comme suit :

---

***Pseudo code de l'algorithme d'optimisation par essaim de particules***

---

*Initialisation de la taille de population*

*Initialisation aléatoire des positions et des vitesses de chaque particule*

*pour chaque particule  $i$ ,  $p_i = x_i$*

***tant que*** le critère d'arrêt n'est pas atteint ***faire***

***pour***  $i = 1$  à  $N$  ***faire***

***Déplacement de particule***

$$\begin{cases} V_i(k+1) = W V_i(k) + c_1 r_1(k) (p_{best}(k) - p_i(k)) - c_2 r_2(k) (g_{best}(k) - p_i(k)) \\ x_i(k+1) = x_i(k) + V_i(k+1) \end{cases}$$

***Evaluation des positions***

***si***  $f(x_i)$  est meilleur que  $f(p_i)$

***Alors***  $p_i = x_i$

***fin si***

***si***  $f(g_{best})$  est meilleur que  $f(p_{best})$

***Alors***  $g_{best} = p_{best}$

***fin si***

***fin pour***

***fin tant que***

---

## 5.4 Influence des paramètres sur l'algorithme de PSO

Les paramètres affectant sur la convergence de l'algorithme de PSO vers l'optimum globale dans l'espace de recherche, consistent en :

### 5.4.1 Facteur d'inertie

Le facteur d'inertie  $w(k)$  est introduit par Shi et Eberhart. La force de vitesse de la particule est équilibrée grâce au facteur d'inertie  $w(k)$ . Shi et Eberhart ont trouvé qu'une meilleure convergence est obtenue pour  $w(k) \in [0,8 ; 1,2]$ . Selon,  $w(k)$  peut être évalué, dans l'intervalle  $[0,9 ; 0,4]$ , par la formule suivante :

$$w(k) = w_{min} + (w_{max} - w_{min}) \left( \frac{k}{max_{iter}} \right) \quad (3.17)$$

$w_{max}$  : Valeur maximale de  $w$  (= 0,9) ;

$w_{min}$  : Valeur minimale de  $w$  (= 0,4) ;

$iter$  : Itération courante ;

$max_{iter}$  : Nombre maximum d'itérations à effectuer.

D'après la formule précédente, le facteur d'inertie  $w(k)$  diminue au cours de l'exécution. En effet, au début des itérations, les particules établiront de grands déplacements. Ceci permettra d'étudier une large zone de l'espace de recherche. Quand le nombre d'itérations est augmenté, l'algorithme pourrait converger. A cet effet, il serait préférable de ralentir le déplacement, et donc diminuer le facteur d'inertie  $w(k)$ , de la particule pour ne pas rater l'optimum global [81].

#### 5.4.2 Coefficients d'accélération $c_1$ et $c_2$

$c_1$  et  $c_2$  sont deux paramètres constantes. Ils influent sur la direction que va prendre la particule à chaque itération vers la meilleure position (globale ou locale). Ces constantes sont comprise entre 0 et 2 (c'est-à-dire  $0 < c_1$  et  $c_2 \leq 2$ ). Les deux nombres  $r_1(k)$  et  $r_2(k)$  sont générés aléatoirement, à chaque itération, dans l'intervalle  $[0 ; 1]$  par une distribution uniforme [83]. Par voie de conséquence, les deux quantités  $r_1(k)*c_1$  et  $r_2(k)*c_2$  sont des variables au cours des itérations. Leurs valeurs sont comprises entre 0 et 2 pour chaque itération.

#### 5.4.3 Coefficient de constriction

L'équilibrage de l'algorithme pendant le processus de recherche est assuré par la combinaison des trois paramètres  $\omega$ ,  $c_1$  et  $c_2$ . Clerc et Kennedy [79] prouvent que la convergence peut avoir lieu en remplaçant les trois paramètres ( $\omega$ ,  $c_1$  et  $c_2$ ) par le coefficient de constriction  $\chi$ . L'algorithme ainsi obtenu est connu sous le nom de canonical PSO. Ce dernier ne nécessite pas la connaissance de la valeur de la vitesse maximale. La nouvelle équation de la mise à jour de la vitesse est telle que :

$$V_i(k+1) = \chi (V_i(k) + \phi_1(k)(p_{best}(k) - X_i(k)) - \phi_2(k)(g_{best}(k) - X_i(k))) \quad (3.18)$$

$$\text{Avec :} \quad \chi = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}} \quad (3.19)$$

où :  $\phi = \phi_1 + \phi_2$ ,  $\phi > 4$ .

Les valeurs optimales de  $\phi_1$  et  $\phi_2$  déterminées par Clerc et Kennedy sont :  $\phi_1 = \phi_2 = 2,05$  soit  $\phi = 4,1$ , ce qui donne un coefficient de constriction  $\chi = 0,7298844$ .

#### 5.4.4 Topologie de voisinage

Le réseau de communication entre les particules a pour rôle l'orientation des particules durant le processus de déplacement. Le type de communication a une influence sur les performances de l'algorithme. La communication entre les particules voisines peut être faite par deux types de structures : géométrique et social. La figure 3.8 montre la différence entre un voisinage géographique et un voisinage social.

- **Voisinage géographique** : le nombre de voisins doit être défini à l'avance. À chaque itération, la distance entre la particule "i" et les autres particules de l'essaim doivent recalculer à partir de la distance entre elles, prédéfinie dans l'espace de recherche ; cette distance change à chaque itération. Cette topologie dynamique nécessite un temps de calcul très élevé.
- **Voisinage social** : Au début de l'algorithme, les voisinages sont prédéfinis et ne sont pas modifiés par la suite. Le but d'utiliser ce type de voisinage est d'améliorer la rapidité de convergence de l'algorithme. A l'étape de convergence, un voisinage social tend à devenir un voisinage géographique.

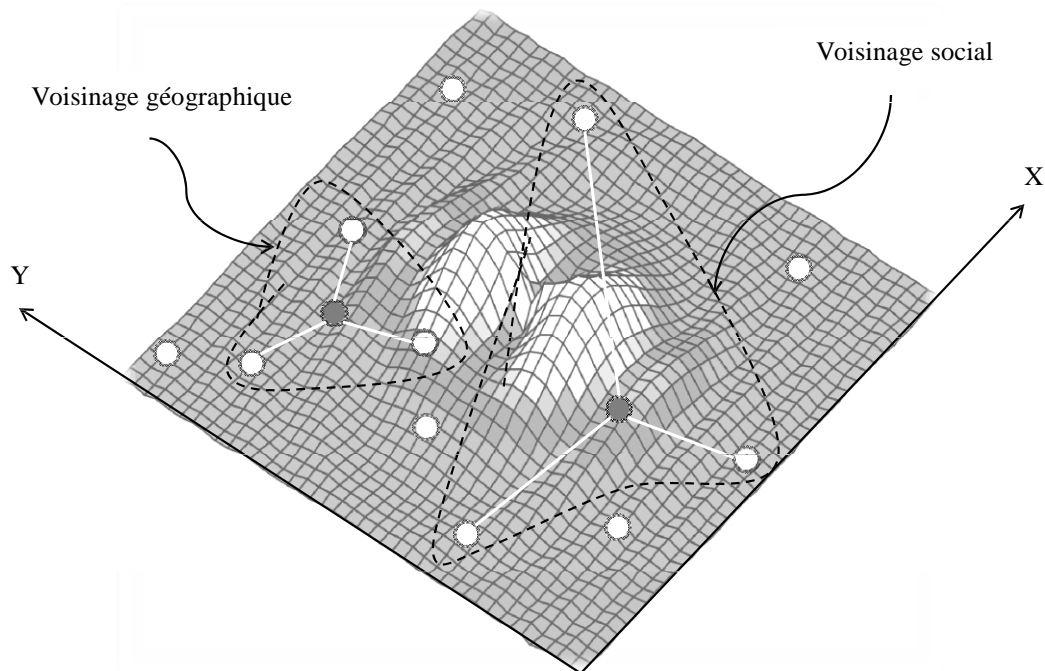


Figure 3.8. Voisinages social et géographique [84]

## 5.5 Critère d'arrêt

En général, les critères d'arrêt qui sont sélectionnés consistent en un critère :

- Statique : le nombre maximum d'itérations ;
- Critère dynamique : la stagnation de la vitesse ;

Etant donné que nous n'avons pas eu un problème de stagnation de vitesse dans les algorithmes que nous avons élaborés, nous nous sommes basées sur le premier critère d'arrêt [85].

## **6 Conclusion**

Dans ce chapitre, nous avons présenté les fondements théoriques essentiels pour comprendre le principe des méthodes d'apprentissage supervisé (SVM, KNN et Naïve Bayésienne). Nous verrons dans le quatrième chapitre l'utilisation de ces trois algorithmes aux problèmes de diagnostic des pathologies des cordes vocales. Nous verrons aussi comment adapter l'Optimisation par Essaims de Particules à l'algorithme SVM pour le réglage des paramètres de ce dernier, afin d'obtenir un meilleur diagnostic de ces pathologies en se basant sur l'analyse quantitative de récurrence appliquée au signal glottique.



---

***Chapitre 4 :***  
***Méthodologie et Résultats des***  
***Classifications des pathologies des***  
***cordes vocales***

## 1. Introduction

Ces dernières années, l'application des techniques de traitement de signal pour l'identification automatique de la voix pathologique a reçu une attention particulière dans les recherches en Traitement Automatique de la Parole (TAP).

Dans ce chapitre, nous allons implémenter les différentes techniques décrites aux chapitres précédents et évaluer les performances des paramètres du signal glottique. Dans un premier temps nous allons exposer la méthodologie et les résultats de classification des pathologies des cordes vocales en se basant sur les paramètres articulatoires de l'onde glottique. En deuxième lieu nous exposons une toute nouvelle technique appelée l'Analyse Quantitative de Récurrence (Recurrence Quantification Analysis, ou RQA). Cette dernière permet de quantifier la dynamique du signal glottique au long du processus de production de la voyelle longue [a:]. La RQA est alors introduite pour analyser les structures qui apparaissent sur les diagrammes de récurrence (Recurrence Plots) et pour extraire différentes caractéristiques du signal glottique en vue d'étudier l'état des cordes et diagnostiquer ses performances. Huit indicateurs RQA seront calculés sous les différentes pathologies étudiées. Les valeurs moyennes de ces huit indicateurs sont utilisés comme entrées à trois méthodes de classification (K-plus proches voisins, Naïve Bayes and Support Vector Machines (SVM) et Particle Swarm Optimisation (PSO-SVM).

Nous employons en dernier lieu une étude préliminaire servant à une évaluation (avant et après une rééducation orthophonique) paramétrique des pathologies des cordes vocales étudiées, appliquant l'analyse quantitative de récurrence croisée (Cross Recurrence Quantification Analysis : CRQA).

## 2. Base de Données

La disponibilité de données est une question cruciale dans le domaine du traitement et d'analyse de la voix pathologique. En effet, il n'est pas aisé de concevoir une base de données contenant des évaluations médicales et plusieurs enregistrements vocaux pour chaque patient, les contraintes sont liées, d'une part, à la disponibilité et l'approbation des malades et des organismes médicaux et d'autre part, aux conditions délicates d'enregistrement. De telles données existent seulement pour quelques langues, majoritairement en Anglais, en Espagnol, en Coréen et en Allemand (Baghai - Ravary et Beet, 2013). Plusieurs Bases de Données (BD) ont été utilisées dans la littérature pour la tâche de classification des voix saines et pathologiques. Nous pouvons les classer en données publiques et privées. Les bases de données publiques peuvent être librement ou commercialement distribuées pour la communauté scientifique. Nous pouvons affirmer que les données publiques sont plus utilisées, car elles permettent des comparaisons rapides. Des exemples de BD publiques sont : Massachusetts Eye & Ear Infirmary (MEEI) [86], Saarbrücken Voice Database (SVD) [87], MAPACI [88].

Saarbrücken Voice Database sur laquelle nous avons évalué notre travail est une BD Allemande qui a été récemment mise gratuitement à disposition en ligne. C'est une collection des enregistrements vocaux de plus de 2000 personnes, où une session est définie comme une collection :

- Des enregistrements des voyelles [a, i, u] produits à la hauteur normale, haut, bas et haut-bas ;
- Des enregistrements de la phrase "Guten Morgen, wie geht es Ihnen ?" ("Bonjour Comment allez-vous?") ;
- **ElectroGlottoGram signal (EGG) ;**

La taille des fichiers qui contiennent les voyelles soutenues est comprise entre 1 et 3 secondes. Tous les enregistrements sont échantillonnés à 50 kHz et leur résolution est de 16 bits. Soixante et onze (71) différentes pathologies sont contenues, y compris à la fois fonctionnelle et organique. Un nombre de 393 échantillons ont été sélectionnés correspondant à des voix normales et pathologiques. Le nombre d'échantillons correspondant à la voix normale a été choisi de façon à être adéquat aux nombres des échantillons pathologiques.

Trois types de pathologies affectant les cordes vocales ont été sélectionnés, étant données que les cordes vocales sont assimilées comme un résonateur permettant la production de sons voisés. De plus, certaines pathologies telles que la paralysie des cordes vocales et la dysphonie spasmodique ne présentent aucune lésion organique sur les cordes vocales, donc son diagnostic à l'aide des outils instrumentaux comme l'endoscopie peut conduire à un diagnostic erroné. **Le tableau 4.1** illustre le détail de la BD utilisée.

**Tableau 4.1** : Détails sur la Base de Données utilisée

Voix	Sujets		Marge d'Age		Moyenne d'Age		Ecart types (SD)	
	Hommes	Femmes	Hommes	Femmes	Hommes	Femmes	Hommes	Femmes
<b>Normal</b>	71	62	25-45	20-50	35	35	6.06	7.37
<b>Polypes</b>	45	30	26-50	25-45	38	35	7.61	5.48
<b>Paralysie des CVs</b>	57	40	25-40	28-38	32.5	29	4.80	4.97
<b>Dysphonie Spasmodique</b>	44	44	23-43	24-44	33	34	5.69	6.08

### 3. Analyse et classification des pathologies des cordes vocales en se basant sur les paramètres glottiques

Nous proposons dans un premier temps une analyse paramétrique basée sur les paramètres articulatoires correspondant aux vibrations des cordes vocales. Nous nous sommes concentrés sur 11 caractéristiques extraites de l'onde glottique, elle-même. Cependant, afin

de concevoir un système acceptable, en termes de précision, de coût de calcul et de performance, nous avons suggéré ces caractéristiques en sélectionnant les plus pertinentes pour discriminer les meilleures données possibles. Toutes les caractéristiques ont été calculées pour chaque échantillon de voix extrait de la base données SVD.

### 3.1. Méthodologie

Notre objectif est alors d'étudier la précision des paramètres articuloire du signal glottique, pour la classification des pathologies : Polypes et Paralysie des CVs, où la voix normale est considérée comme référence. Ces caractéristiques ont été transmises à classificateurs : KNN, classification bayésienne et les SVM, en utilisant la Base de Données allemande Saarbrecker Voice Database (SVD). L'extraction et la sélection des caractéristiques ont une importance particulière lors de l'analyse du signal vocal. En fait, les paramètres optimaux doivent contenir les informations nécessaires qui permettent la bonne discrimination des classes normales et pathologiques. Afin d'estimer la source glottique, de nombreuses approches ont été développées dans la littérature dans l'objectif est de supprimer la contribution du conduit vocal du signal vocal. Dans cette étude, nous avons utilisé l'algorithme populaire IAIF (Iterative Adaptive Inverse Filtering) en raison de ses performances élevées et de sa facilité de développement. Nous avons utilisé un toolbox (interface graphique) développé sous l'environnement MATLAB (MathWorks) appelé Aparat (figure 4.1) qui a été développée pour l'extraction du signal glottique par la méthode IAIF. Cet outil nous a permet d'extraire les paramètres articuloires de l'onde glottique de chaque échantillon pathologique et normal [89].

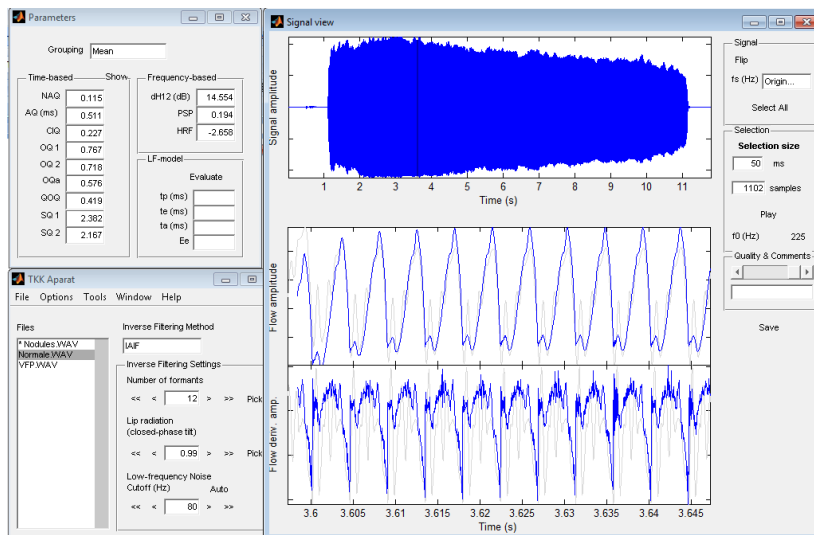


Figure 4.1 : Interface de l'Outil Aparat

Ces paramètres sont utilisés comme vecteur d'entrée d'entités pour les classificateurs cités précédemment. Figure 4.2.

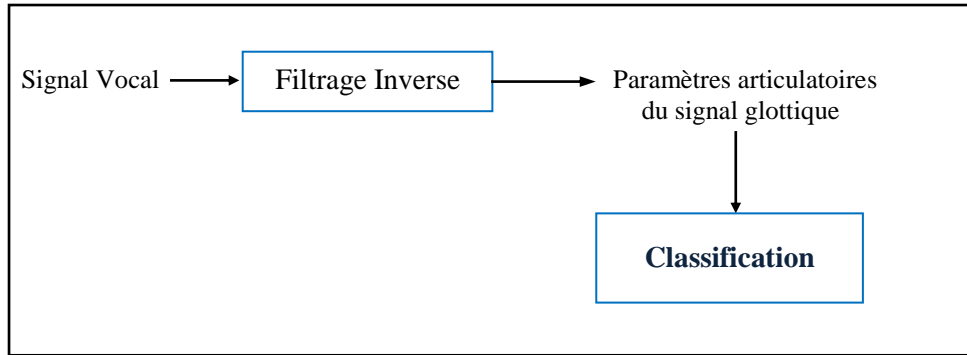


Figure 4.2 : Méthodologie de classification

### 3.2. Analyse des paramètres articulatoires glottique

Le tableau 4.2 montre que les mesures statistiques des paramètres articulatoires de chaque classe. Cette représentation nous permet de déterminer le comportement de chaque paramètre dans chaque classe. Ka et Ko leurs valeurs moyennes sont presque égales aux voix avec polypes et paralysie mais ses valeurs sont inférieures par rapport aux voix normales, nous pouvons conclure que les instants d'ouverture et de fermeture diminue dans le cas de présence de ces pathologies. Pour le groupe de paramètres OQ, CIQ, AQ, NAQ, OQA, QoQ, qui appartiennent aux paramètres du domaine temporel, ses valeurs d'écart-type (standard deviation) sont plus élevées pour les voix normales par rapport aux voix avec polypes et paralysie. Dans les paramètres du signal glottique dans le domaine temporel, l'écart type pour les voix avec paralysie et polypes sont très similaires et suivent les mêmes tendances. Les paramètres dans le domaine fréquentiel distinguent mieux les voix avec paralysie les voix avec polypes.

Tableau 4.2. Analyse statistique des paramètres glottiques de chaque classe

	Polypes				Paralysie				Normal			
	Moyenne	Dev Std	Min	Max	Moyenne	Dev Std	Min	Max	Moyenne	Dev Std	Min	Max
<b>KO</b>	0,00181	0,0098	0,00020	0,00450	0,001797	0,001504	0,00010	0,00720	0,002589	0,001332	0,0004	0,0074
<b>Ka</b>	0,00343	0,00083	0,00040	0,00500	0,00279	0,00185	0,00030	0,00990	0,00307	0,00133	0,0012	0,006
<b>CIQ</b>	0,3221	0,1502	0,0376	0,7223	0,24801	0,12573	0,0321	0,6072	0,39724	0,18526	0,0697	0,7607
<b>QA</b>	0,7613	0,42297	0,2555	2,0716	0,78916	0,50021	0,2179	1,995	0,99111	0,39366	0,2286	2,3553
<b>NAQ</b>	0,1212	0,07816	0,0472	0,3307	0,10521	0,05276	0,0162	0,2228	0,15842	0,06963	0,0374	0,4092
<b>QO</b>	0,32558	0,15168	0,1258	0,6189	0,26952	0,11744	0,0716	0,561	0,41325	0,13794	0,0628	0,7244
<b>QoQ</b>	0,3832	0,25283	0,0659	0,7872	0,28758	0,18497	0,0342	0,7514	0,4509	0,21067	0,0826	0,7261
<b>SQ</b>	1,7898	0,77705	0,4074	3,433	2,02462	1,07392	0,4432	7,9025	1,57107	1,0766	0,3146	4,0084
<b>DH12</b>	6,2828	5,82817	-22,72	13,2562	6,55555	6,69789	-7,2049	34,5436	5,32254	2,92909	-0,3915	11,0567
<b>HRF</b>	4,23594	6,87789	-3,0292	27,4274	6,19521	6,91366	-2,6794	26,7979	2,88172	2,44622	-2,9348	8,12

Pour améliorer le taux de classification de notre algorithme et analyser l'aspect discrimination des paramètres articulatoire de l'onde glottique, une analyse des variations des paramètres les plus pertinents a été établie en utilisant les boites à moustaches (boxplots) sous l'environnement Matlab. Nous avons représenté certaines variations des paramètres du signal glottique pour voir leurs comportements sur chaque trouble de voix

par rapport à la voix normale. Le quotient de fermeture des cordes vocales, présente une large plage dans le cas des voix normales par rapports celle dans les voix pathologiques. Ceci implique que l'intensité vocale diminue par la présence des lésions au niveau des cordes vocales figure 4.3a. Le quotient d'ouverture présente des valeurs plus importantes dans le cas des patients souffrant des nodules des cordes vocales ceci est due au non fermeture complète des cordes vocales lors de la vibration à cause de la présence des lésion nodulaire au niveau des cordes vocales figure 43b. Dans le cas du quotient de vitesse nous remarquons que la plus grande plage de données est enregistrée dans le cas de la dysphonie spasmodique, ce qui explique une fermeture excessive des cordes vocales lors de la phonation figure 4.3c. La différence entre la première et la deuxième harmonique présente des faibles plages de données dans toutes les pathologies, avec une petite variation ddans le cas de la voix normale, figure 4.3d.

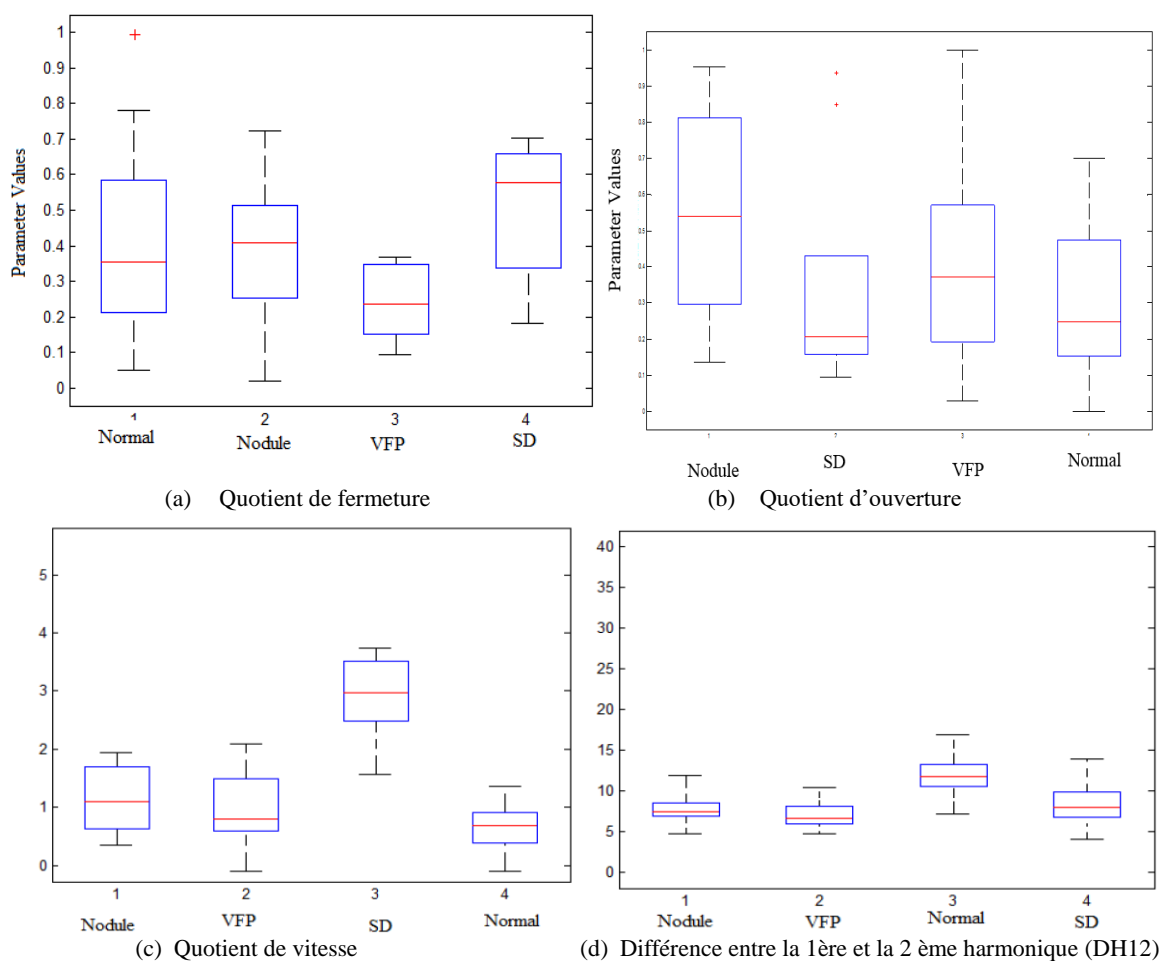


Figure 4.3 : boîtes à moustaches des variations des paramètres articulatoires en fonction de la présence des pathologies.

### 3.3. Résultats de classification et discussions

Notre expérience a été réalisée en utilisant la Base de Données SVD où nous avons sélectionné la voyelle longue [a:] de duré 02 secondes, la fréquence l'échantillonnage est de 50 kHz.

Dans cette configuration, un vecteur comprenant 11 paramètres extraits du signal glottique considéré comme des caractéristiques d'entrée des classificateurs. Nous avons considéré deux groupes de données, le premier correspond à la phase d'apprentissage et le deuxième correspond à la phase de reconnaissance. 120 échantillons de la voyelle longue [a:] normales et pathologiques sont sélectionnées, 75% de l'ensemble de données est utilisé pour la phase d'apprentissage et 25% est réservé à la phase de test. L'évaluation de la capacité de discrimination des paramètres glottiques a été réalisée en termes de rendement ou (accuracy) (ACC%) donné par :

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Où :

- TP représente le nombre d'échantillons pathologiques correctement classée ;
- FP indique le nombre d'échantillons pathologies énoncés classés comme normaux ;
- TN désigne le nombre d'échantillons normaux correctement classés ;
- FN est le nombre d'échantillons normaux classés comme pathologiques ;

Le tableau 4.3 récapitule les différents résultats obtenus lors de l'application des algorithmes SVM, KNN et naïve Bayes utilisant les paramètres glottiques. D'après le tableau, l'algorithme SVM offre le meilleur taux de réussite comparant aux KNN et naïve Bayes avec un rendement de 92.58%. Par ailleurs, l'algorithme KNN, et les paramètres du signal glottique sont avérés être un outil efficace, présentant un bon compromis pour la classification des troubles des cordes vocales sélectionnés. Le taux de classification était d'environ 87,67 %, présente de très bons résultats également. Quant à l'algorithme naïve Bayes, donne des résultats acceptables.

**Tableau 4.3 : Résultats de Classifications KNN et SVM**

Voix	Normal	Polypes	Paralysie
reconnaissance	30	30	30
test	10	10	10
Accuracy (SVM)	<b>92.58%</b>		
Accuracy (KNN)	<b>87.67%</b>		
Accuracy (Naïve Bayes)	<b>86.24%</b>		

#### 4. Analyse Quantitative de Récurrence appliquée au signal glottique pour la classification des pathologies des cordes vocales

Cette partie détaillera les démarches du système de classification proposé.

##### 4.1. Méthodologie et formulation du problème

L'approche proposée consiste en une méthode d'extraction des caractéristiques extraites par l'Analyse Quantitative du diagramme de récurrence **RP** (**R**ecurrence **P**lots), appliquée au signal glottique et une méthode de classification **Particle Swarm**

**Optimization-Support Vector Machines (PSO-SVM)** pour distinguer les différents types de pathologies affectant les cordes vocales. À cet effet, le signal glottique a été d'abord, extrait des échantillons enregistrés du signal vocal de la voyelle longue [a:] en utilisant un algorithme de filtrage inverse adaptatif itératif IAIF (Itérative Adaptive Inverse Filtering). L'analyse quantitative de récurrence RQA de la structure RP a été appliquée ensuite au signal glottique en vue d'extraire huit (08) paramètres RQA qui ont été utilisés comme un vecteur de caractéristiques d'entrée du classificateur. Ce dernier a été réalisé en utilisant l'algorithme PSO-SVM. Dans cet algorithme hybride, les paramètres SVM ( $C, \lambda$ ) ont été optimisés à l'aide de l'algorithme PSO, fournissant une implémentation simple et une technique facile à comprendre.

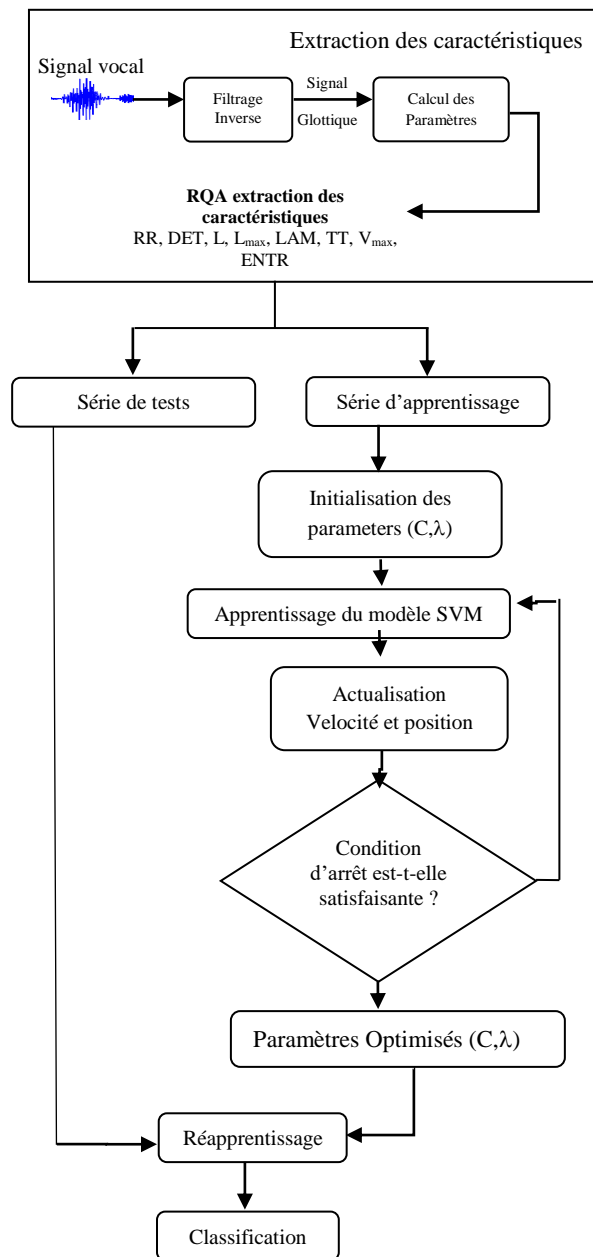


Figure 4.4. Organigramme de la méthode proposée



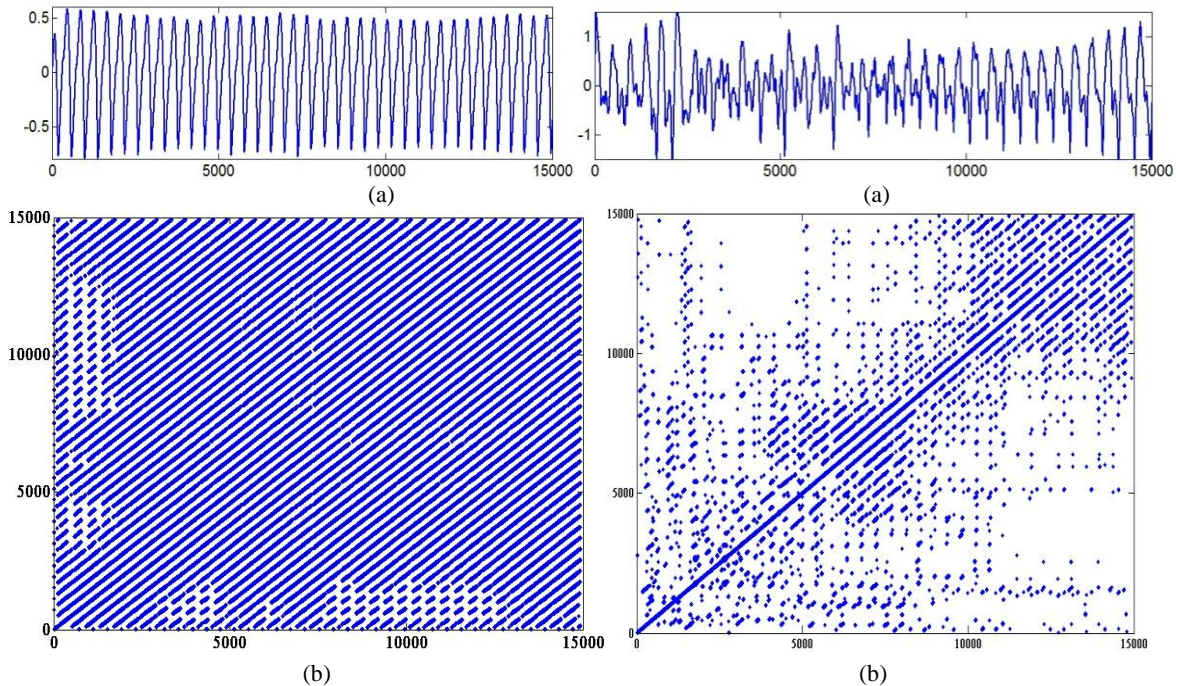
La méthode «validation croisée de K fois» ou «K- folds cross validation » a été utilisée pour évaluer les capacités de discrimination, où  $K = 10$ . Dans la validation croisée 10-folds cross validation, l'ensemble de données est divisé au hasard en 10 sous-ensembles différents de  $N$  échantillons. Dans notre cas, nous avons un ensemble de 393 échantillons correspondant à des sujets normaux et pathologiques ( $N = 40$ ). 10 répétitions ont été utilisées pour estimer les performances de classification. Où, nous avons formé un model d'apprentissage-machine sur  $K-1$  sous ensemble, puis nous évaluons le model par le sous ensemble restant (phase de test). Ce processus est répété  $k$  fois, avec un sous-ensemble différent réservé à l'évaluation. Les performances sont calculées en vérifiant les résultats obtenus à partir de chaque ensemble de données. Des méthodes de classification SVM et **KNN** (**K**-Nearest-Nighbours) ont été utilisées pour montrer l'efficacité de la méthode proposée. PSO-SVM a fourni une meilleure précision du taux de classification des cordes vocales. La figure 4.4 montre l'organigramme de l'approche proposée.

#### **4.2. Technique RP appliquée au signal glottique**

Les figures ci-dessous présentent des exemples qui comprennent quatre signaux glottiques extraits des signaux vocaux correspondants à la voyelle longue [a:], ainsi leur structure de tracé de récurrence correspondant RP. Chacune des figures concerne un type spécifique d'un signal glottique lié aux personnes ayant : une paralysie, une dysphonie spasmodique, Nodules et polypes au niveau des cordes vocales. Ces exemples illustrent les structures RP qui peuvent se produire sur chaque type de signal glottique. Ils ne représentent pas nécessairement la totalité de la Base de Données, la RP peut voir ses structures modifiées en fonction des caractéristiques et la gravité de la pathologie laryngée. Une trame du signal glottique 600 ms (15000 échantillons) obtenu à partir du signal de la voyelle [a:] par un filtrage inverse IAIF. Cette voyelle a été choisie parce qu'elle est couramment utilisée dans les évaluations des pathologies du larynx. C'est également une voyelle orale, moyenne, ouverte et non arrondie, ce qui permet une position plus neutre et intermédiaire du conduit vocal. La structure de récurrence tracée RP correspondante est aussi représentée sur la même figure. Notons que l'absence de pathologies au niveau des cordes vocales reflète des structures diagonales bien définies, car la régularité d'une voix saine pour une voyelle longue est approximativement un processus quasi-périodique.

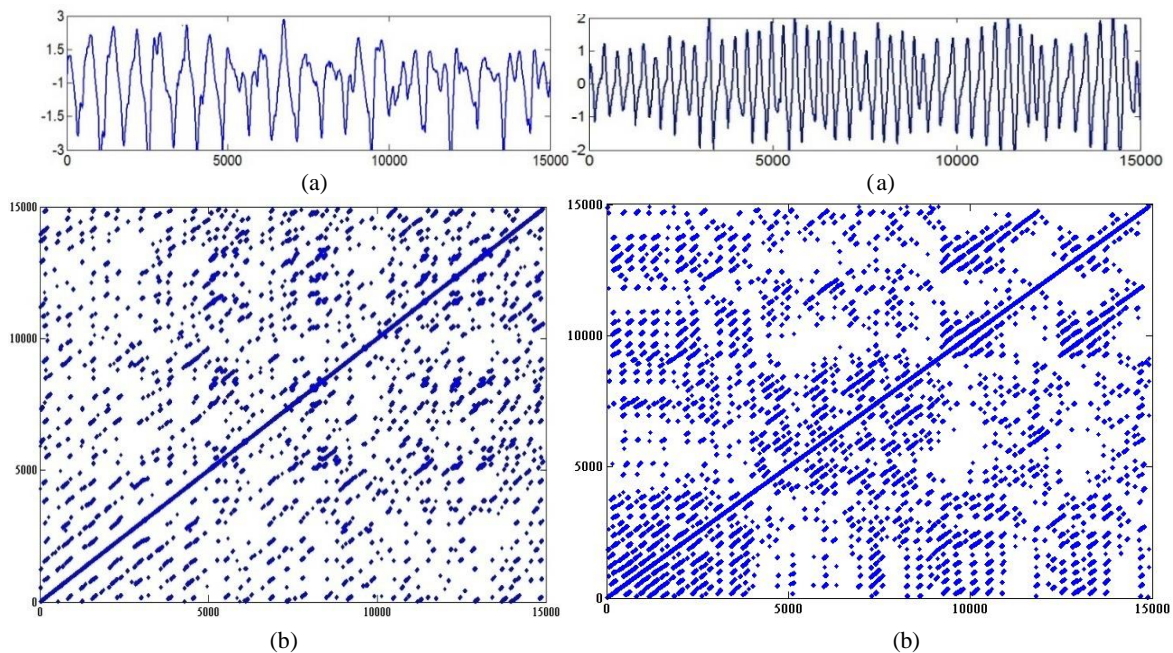
La formation de structures RP, qui reflète la complexité du système dynamique, diffère d'une pathologie à une autre. À titre de premier exemple, la figure 4.5, montre le signal glottique (15000 échantillons) et la structure RP pour un sujet normal, ne présentait, historiquement, aucune pathologie. La structure RP correspondante à une voix normale composée de points concentrés regroupés en lignes diagonales parallèles avec une homogénéité de la densité des points noirs et blancs figure 4.4. Cette configuration caractérise la périodicité et la faible variation d'amplitude du signal glottique. Cependant, la concentration des points récurrents et l'emplacement des bandes vides sur les cartes RP

dépendent du signal glottique des voix sélectionnées. Hors que, la périodicité est perdue dans les trois autres figures, ce qui indique la présence d'une pathologie.



**Figure 4.5 :** (a) Signal glottique, (b) : structure RP correspondant à une voix normale

**Figure 4.6 :** (a) Signal glottique, (b) : structure RP correspondant au ; polypes de cordes vocales



**Figure 4.7 :** (a) Signal glottique, (b) : structure RP correspondant à une Paralysie des cordes vocales

**Figure 4.8 :** (a) Signal glottique, (b) : structure RP à une correspondant dysphonique spasmodique

Pour les polypes des cordes vocales, les cartes RP montrent une faible densité des points de récurrence accompagnée d'un grand nombre de points isolés. La concentration de ces points augmente dans le cas de paralysie des cordes vocales. En comparant avec la structure RP correspondante aux autres pathologies, elle est moins irrégulière que les autres ce qui

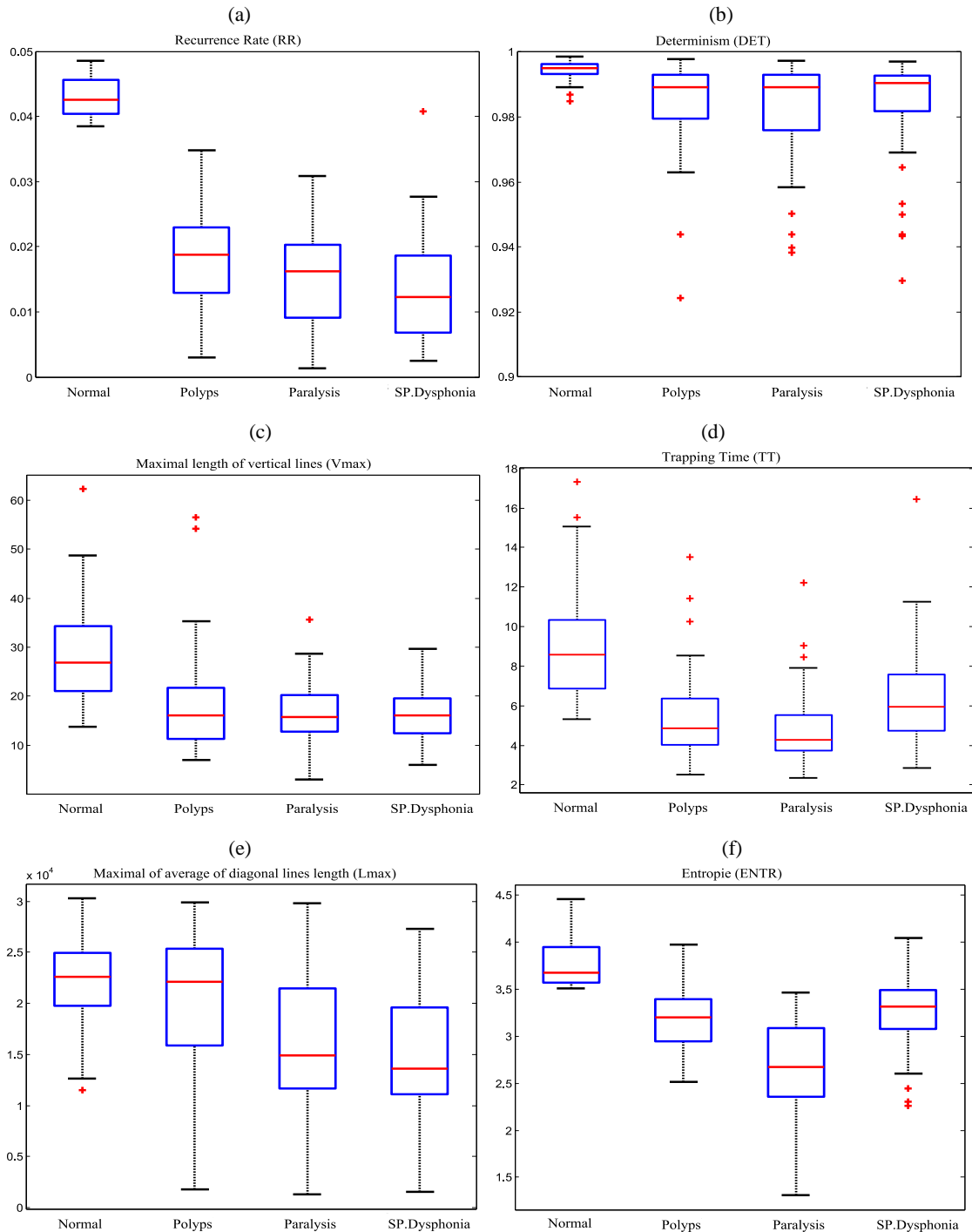
conduit que la forme d'onde du signal glottique est similaire à celle de la voix normale (Figure 4.6). La concentration de ces points augmente dans le cas de la paralysie des cordes vocales par rapport aux polypes et à la voix normale (Figure 4.7). Les bandes blanches indiquent l'absence de la voix ou une perte vocale produite par une fermeture excessive de la glotte. Cette constatation a été bien notée pour la maladie de dysphonie spasmodique. (Figure 4.8).

En général, il est possible d'extraire des informations concernant la présence de maladie dans la voix, à partir des tracés de récurrence appliqués au signal glottal. Cependant, l'analyse de la topologie graphique de la carte RP peut être assez difficile. À cet effet, une analyse de quantification de récurrence (RQR) a été utilisée à l'étape suivante de l'étude pour décrire quantitativement les caractéristiques du signal suivant une analyse paramétrique des structures graphiques de chaque pathologie.

### 4.3. Application de l'analyse quantitative de récurrence (RQA) au signal glottique

Dans le but d'évaluer le pouvoir discriminatif des paramètres RQA des pathologies des cordes vocales et voir l'ordre de la grandeur de chaque paramètre, une analyse paramétrique a été effectuée en se basant sur les boxplots ou boîtes à moustaches, où l'analyse quantitative de récurrence a été appliquée sur la structure RP du signal glottique. Dans les figures ci-dessous, six paramètres sont représentés selon le degré de pertinence en fonction de la pathologie traitée. Cette représentation (les boxplots) est un outil graphique disponible sur MATLAB, permettant une analyse visuelle et une comparaison des ensembles de caractéristiques de données pour faciliter la prise de décision. La valeur médiane de l'ensemble de données d'entités est représentée par une ligne horizontale rouge dans le boxplot. Les 25 et 75% des centiles sont illustrés par une ligne horizontale bleue, représentant le premier et le troisième quartile de la boîte à moustaches. Les moustaches représentent la valeur minimale et maximale dans la série de fonctions. Les croix rouges représentent l'ensemble de données résiduelles et les valeurs aberrantes actuelles. Le taux de récurrence RR montre une différence claire entre les médianes des voix normales et pathologiques avec une valeur autour de 0,043. Dans le cas de la dysphonie spasmodique, la valeur la plus basse est d'environ 0,012. Les polypes et la paralysie des cordes vocales (avec respectivement 0,015 et 0,018,) montrent des valeurs inférieures à celles des personnes normophoniques. Cela signifie que la tendance des points de récurrence dans la structure RP diminue par la présence de pathologies (Figure 4.9a). Sur la figure 4.9b, le déterminisme (DET) montre des valeurs médianes approximatives mais les valeurs maximales et minimales des moustaches présentent une différence remarquable et peuvent donc être utilisées comme indicateurs discriminants. La longueur maximale des lignes verticales  $V_{max}$ , présente des valeurs médianes qui sont proches entre eux à travers les trois types de pathologie et les sujets normophoniques. Le  $V_{max}$  Apporte des informations sur le comportement chaotique complexe dans chaque pathologie. Les valeurs les plus élevées

comprises entre 70 et 80, sont respectivement dans le cas de la voix normale et de la dysphonie spasmodique (Figure 4.9c).



**Figure 4.9 :** Boîtes à moustaches (Boxplot) distribution des données des paramètres RQA de chaque type de voix (Normale, Polypes, Paralysie des CVs, dysphonie Spasmodique) ((a) : RR, (b) : DET, (c) : Vmax, (d) : TT, (e) : Lmax, (f) : ENTR

Le Training Time (TT) possède à la plus petite distribution de données (min = 2, max = 8) et la valeur médiane la plus basse est de 4, dans le cas de la maladie de paralysie des CVs. La distribution de données la plus large (min = 5, max = 15) et la valeur médiane la plus élevée proche de 9 est enregistrée dans le cas des voix normales. Notons que des valeurs

élevées de TT indiquent l'absence ou de légers changements du signal glottal à travers le temps. La dysphonie spasmodique et les polypes présentent de faibles plages de données du paramètre TT avec respectivement des valeurs médianes 6 et 5. Dans la Figure 4.9d, le  $L_{max}$ , représente le temps moyen pendant lequel deux segments de la trajectoire sur la structure RP sont proches les uns des autres. Il montre une grande variation dans le cas de la paralysie des polypes et des FV et il varie entre  $(0,2 \text{ et } 3) * 10^4$  par rapport à la voix normale et des polypes des cordes vocales (Figure 4.9e). L'entropie est un paramètre indicateur important utilisé pour mesurer la complexité d'un signal donné et le comportement de périodicité (signal glottique). La figure 4.9f montre que la voix normale possède la distribution maximale de données (3,5 et 4,5) et une valeur médiane de 3,75. Cela indique la périodicité de l'onde glottique. Cependant, les valeurs médianes correspondent aux autres pathologies sont inférieures à la valeur de la voix normale, variant entre 2,75 et 3,25 ce qui indique la perte de périodicité due à la présence de pathologies. Les valeurs médianes et minimales les plus faibles sont respectivement de 2,25 et 1,15, dans le cas de la paralysie des cordes vocales. Par conséquent, ce paramètre présente un appui important pour distinguer les voix saines des voix pathologiques.

#### 4.4. Optimisation des paramètres de SVM par l'algorithme d'Optimisation par Essaim de Particules (PSO)

Afin de maximiser le taux de précision de diagnostic, les paramètres de marge ( $C$ ) et de conditionnement ( $\lambda$ ) du modèle SVM ont été optimisés par l'algorithme PSO. Cette technique, faisant partie de la métaheuristique, a été initialement introduite en 1995 par Kennedy et Eberhart [79]. Elle était destinée à simuler la migration et la collecte d'un troupeau d'oiseaux ou de poissons lorsqu'ils recherchent de la nourriture. Chaque oiseau rappelle ses expériences de vol et partage les distances les plus courtes entre l'aliment et lui-même. Ce comportement social, appelé intelligence d'essaims, guide tous les oiseaux afin de déterminer la distance la plus courte à la nourriture [80]. Les différentes étapes pour l'optimisation des paramètres SVM utilisant l'algorithme PSO sont décrites comme suit :

1. **Initialisation** des positions et des vitesses aléatoires dans l'espace de recherche de dimensions  $d$ . La  $i^{\text{ème}}$  particule possède une position  $P_i = (P_{i1}, P_{i2}, P_{i3} \dots P_{id})$  et une vitesse (déplacement)  $V_i = (V_{i1}, V_{i2} \dots V_{id})$ .  $P_i$  et  $V_i$  sont des matrices de  $N * d$ . Notons que chaque position et déplacement sont représentés par des couples (paramètre de marge, paramètre de conditionnement).
2. **Apprentissage** : L'algorithme SVM est exécuté pour chaque particule en utilisant les paramètres de marge et de conditionnement à partir la position actuelle de celle-ci. Cette opération est répétée  $N$  fois, soit pour toutes les particules de la population. Pour cela, la méthode 10-folds cross validation a été utilisée pour effectuer le processus d'apprentissage.
3. **Test** : L'algorithme SVM est testé sur le reste de la Base de Données. La fonction objectif représentant le taux de réussite (ou d'exactitude) de diagnostic de chaque particule est calculée selon l'équation (4.1). Notons que le taux de réussite correspond au nombre d'échantillons classés avec succès sur le nombre d'échantillons total formant l'ensemble de test. Il est donné par :

$$f_i = \frac{y_i - y_{fi}}{y_i} \times 100 \quad (4.1)$$

$y_t$  est nombre de données de test et  $y_{\bar{t}}$  le nombre des échantillons mal classés.

4. **Mise à jour** de la vitesse et la position de chaque particule conformément aux équations (4.2) et (4.3) pour obtenir  $\mathbf{P}_{\text{best}}$  et  $\mathbf{g}_{\text{best}}$  suivant

$$\mathbf{V}_i^d(t+1) = \mathbf{W} \mathbf{V}_i^d(t) + c_1 r_1(t)(p_{\text{best}}^d(t) - \mathbf{p}_i^d(t)) - \quad (4.2)$$

$$c_2 r_2(t)(\mathbf{g}_{\text{best}}^d(t) - \mathbf{p}_i^d(t))$$

$$\mathbf{p}_i^d(t+1) = \mathbf{p}_i^d(t) + \mathbf{V}_i^d(t+1) \quad (4.3)$$

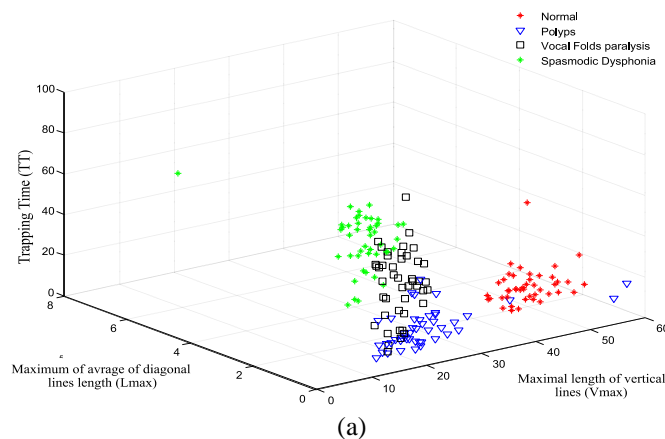
$\mathbf{W}$  est le poids d'inertie,  $c_1$  et  $c_2$  sont deux constantes positives, appelées coefficients d'accélération, et  $r_1(t)$  et  $r_2(t)$  sont des variables aléatoires entre  $\mathbf{0}$  et  $\mathbf{1}$ .

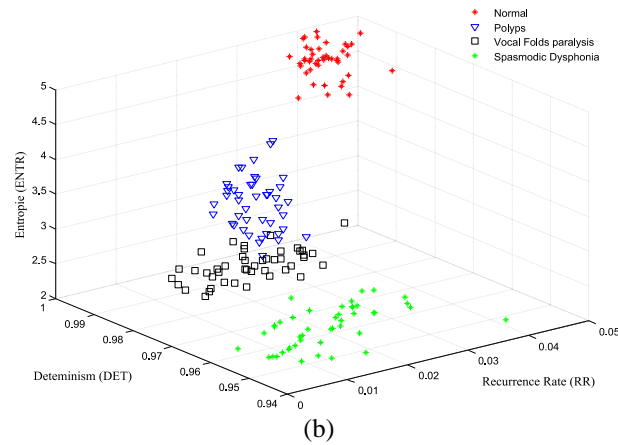
5. **Critère d'arrêt** : Les étapes 2 à 4 sont répétées jusqu'à atteindre le nombre maximal d'itérations fixé à l'avance

6. **Exploitation** : Les paramètres optimaux de SVM ainsi obtenus sont utilisés pour la classification.

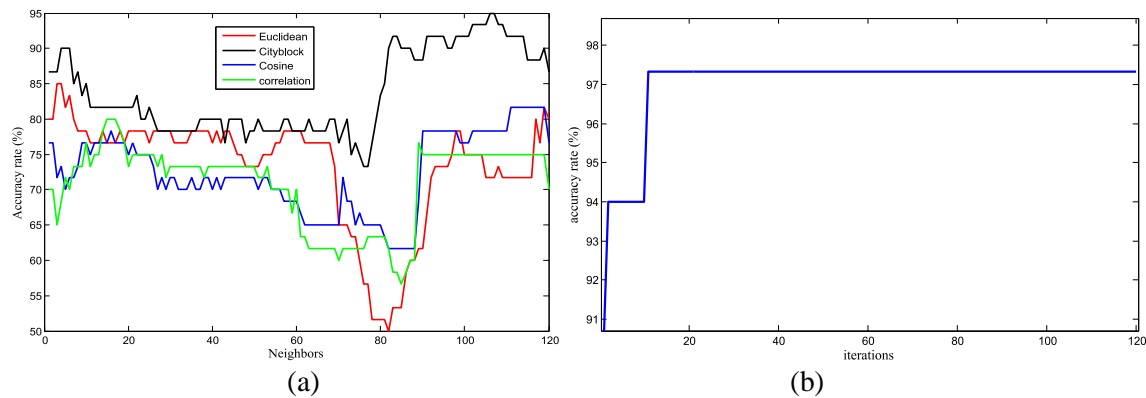
#### 4.5. Résultats de classification et discussions

Dans le but d'évaluer l'aspect séparabilité des ensembles de paramètres quantitatif de récurrence et la distribution des classes des différentes pathologies étudiées, la figure 4.10 illustre les tracés de dispersion tridimensionnels (3D scatter plots) des mesures RQA selon différents types de voix pathologiques (Paralysie, Polypes des CVs et la dysphonie spasmodique). Cette figure montre également la présence de nombreux points chevauchant parmi les paramètres de récurrence. Les échantillons sont visuellement arrondis en quatre groupes distincts liés aux trois pathologies et le groupe dont la voix normale considérée comme référence. Ces représentations expriment l'aspect discriminant des paramètres RQA. Une représentation suggère que les paramètres RQA du signal glottique sont directement liés aux états des cordes vocales normales ou pathologiques. En conséquence, ces constatations démontrent la capacité des indicateurs RQA à différencier les cordes vocales normales et pathologiques.





**Figure 4.10 :** 3-D scatter plot pour voir l'aspect discriminatif des paramètres RQA  
 (a) : Vmax, TT, Lmax. (b) : RR, DET and ENTR



**Figure 4.11 :** Taux de classification de la méthode PSO-SVM et KNN pour différents types de distances  
 (a) : KNN, (b) : PSO-SVM

Il est important de se rappeler qu'en général, SVM s'est vu offrir les meilleures précisions dans différents problèmes de classification et a supplanté d'autres approches de classification. Le KNN ne nécessite aucune phase d'apprentissage mais, il attribue la classe la plus courante correspondant au nombre maximal de  $k$  voisins les plus proches. Pour cela, les SVMs ordinaires et  $k$ -Nearest Neighbours (KNN) sont introduites pour évaluer et évaluer la robustesse de la méthode de prédiction proposée. Dans le cas de l'algorithme de classification KNN, plusieurs types de distance ont été utilisés à savoir : la distance "Euclidienne", "Cityblock", corrélation et "Cosine". Par ailleurs, le nombre de voisins  $k$  a été varié de 1 à 120. La valeur maintenue est celle permettant d'obtenir un meilleur taux de réussite. L'objectif de la figure 4.11a est de sélectionner le meilleur type de distance KNN et le nombre de voisin optimal qui donne le taux de reconnaissance maximal. La figure montre les performances du classificateur KNN. Un taux de 95 % a été obtenu aussi bien pour la distance "Cityblock" avec un nombre de voisins  $k = 106$ . Un taux de classification pour la distance "Euclidienne" de 85 % avec un  $k = 6$ , contre 70% pour  $k = 20$  obtenu lors de l'utilisation de la distance "Cosine". Dans notre classification utilisant l'algorithme Pentagone de Duval-KNN, nous avons choisi la distance "Euclidienne" avec  $k = 7$ .

Le tableau 4.4 illustre également la précision obtenue par SVM. Un taux de décision et de classification, correct est de 96,21% a été obtenu en utilisant 70% des données pour la phase d'apprentissage et les 30% restants pour la phase de reconnaissance, en utilisant la fonction de noyau polynômial. D'après la figure 4.10b, On peut remarquer que la méthode proposée (PSO-SVM) surpasse le classificateur KNN et Les SVMs en précision de détection basée sur les paramètres RQA. Le taux de classification atteint est de 97,41%, ce qui correspond aux paramètres C et  $\lambda$ , 78,085 et 8,333, respectivement. Par conséquent, la combinaison de RQA du signal glottal comme technique d'extraction de caractéristiques et de l'algorithme PSO-SVM utilisant la base de données SVD a présentée de bonnes performances.

**Tableau 4.4 :** Rendement de la méthode proposée.

Méthode	KNN	SVM	SVM-PSO
Taux de Classification (%)	95	96.21	<b>97.41</b>
Neighbors, Distance(KNN)	106/Cityblock	/	/
Parameters Optimisés (C, $\lambda$ )	/	/	<b>(78.085416,8.333069)</b>

Dans le tableau 4.5, un ensemble de 12 échantillons comprenant 3 échantillons de chaque maladie : dysphonie spasmodique, Paralysie et polypes des CVs et 3 échantillons normophoniques ont été considérés. Nous avons affiché les valeurs des huit paramètres RQA. Les 12 échantillons ont été classés avec succès à l'aide de PSO-SVM. Cependant, les KNN et SVM traditionnels offrent respectivement deux et trois erreurs de classement sur 12.

**Tableau 4.5 :** Résultats des différentes classifications

Echantillon	RR	DET	L	L <sub>max</sub>	LAM	TT	V <sub>max</sub>	ENTR	Classe assignee	Classe obtenue		
										KNN	SVM	PSO-SVM
1	0.0418	0.9962	0.9983	22383	8.7505	39.5522	28	4.7770	Normal	VFP*	Normal	Normal
2	0.0077	0.9500	0.9755	13367	5.5239	16.9547	21	2.9094	Dysphonia Spasmodique	Dysphonia Spasmodique	Dysphonie Spasmodique	Dysphonie Spasmodique
3	0.0091	0.9707	0.9245	25319	2.9518	12.9291	12	2.7167	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs
4	0.0126	0.9906	0.9827	24889	6.4516	22.7472	24	3.3405	Polypes	Polypes	Polypes	Polypes
5	0.0187	0.9697	0.9002	15732	4.0673	14.2030	8	2.7194	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs
6	0.0486	0.9976	0.9971	24897	12.3168	48.7418	36	4.4125	Normal	Normal	Normal	Normal
7	0.0095	0.9515	0.9622	13491	5.7926	16.1749	26	2.8147	Dysphonie Spasmodique	Paralysie Des CVs*	Normal*	Dysphonie Spasmodique
8	0.0091	0.9707	0.7705	25319	2.9518	27.9291	19	3.7167	Polypes des CVs	Polypes	Polypes	Polypes
9	0.0439	0.9926	0.9982	17383	6.6825	41.2136	26	4.5951	Normal	Normal	Normal	Normal
10	0.0165	0.9716	0.8965	10020	8.6874	18.3325	10	2.7054	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs	Paralysie des CVs
11	0.0100	0.9476	0.9900	16674	5.0394	14.4865	28	3.1647	Dysphonie Spasmodique	Normal*	Normal*	Dysphonie Spasmodique
12	0.0168	0.9901	0.9048	18288	4.8305	26.8312	15	3.3926	Polypes	Polypes	Polypes	Polypes



Même si le classificateur KNN a fourni trois erreurs de classification sur 12, il semble que ce soient des méthodes efficaces, sélectionnées pour évaluer les performances de la méthode PSO-SVM. Les paramètres RR et ENTR semblent des paramètres très pertinents, comme le montre le tableau, ces deux paramètres ont présenté la plus grande contribution à la classification. En fait, pour obtenir de meilleurs résultats de classification, il est nécessaire de concevoir un algorithme afin que ses caractéristiques correspondent de manière appropriée au choix de la technique d'extraction des caractéristiques et la méthode de classification. \* correspond à une erreur de classement.

## **5. Evaluation des pathologies des cordes vocales basée sur l'analyse quantitative de récurrence croisée appliquée au signal glottique**

Cette partie consiste une autre contribution, où nous avons appliqué l'analyse quantitative de récurrence croisée aux signaux glottique d'un échantillon d'une voix normale et d'une autre d'une voix pathologique du signal de la voyelle longue [a :]

### **5.1. Tracé de Récurrence Croisé (Cross Récurrences Plots : CRP) entre une voix normal et pathologique**

Les figures ci-dessous montrent le tracé de récurrence croisée CRP obtenues en utilisant la série temporelle d'un signal glottique d'une voix normale et celui d'un sujet présentant une pathologie. Quatre pathologies sont présentées dont deux organiques et deux fonctionnelles. Les comparaisons sont prises pendant une phase de rééducation orthophonique déterminée, dans le but de voir son effet sur la qualité vocale en la comparant à la voix saine.

Une inspection visuelle du tracé CRP dans le cas des polypes des CVs indique l'indépendance des modèles de récurrence entre les deux signaux. L'absence de lignes diagonales parallèles exprime la perte de périodicité causée par la présence de polypes dans les cordes vocales. La présence de lignes verticales signifie que le comportement de récurrence du signal n'est pas stable dans les différentes étapes de la série temporelle. L'homogénéité de la distribution ponctuelle exprime la stabilité des deux séries temporelles. Figure 4.12. Dans le cas de la paralysie des cordes vocales, il est évident que la structure du CRP correspondait à la voix normale suite à une évolution brutale des signaux glottiques présente une disparition des lignes diagonales, ce qui est référencée à la perturbation de la fréquence fondamentale. Dans cette situation, certains segments verticaux de signaux apparaissent dans la même région d'espace de phase dans certaines séries temporelles (Figure 4.13).

La structure du CRP entre le signal glottique d'un sujet normal et celui d'un patient présentant une dysphonie spasmodique montre une disparition des lignes diagonales, qui indiquent que les deux signaux ont une dynamique d'espace de phase similaire. Les distributions des points de récurrence formant les lignes le long des lignes verticales sont espacées par des bandes blanches, ce qui fournit une information sur le processus de

production de la voix correspondant à une dysphonie spasmodique (Figure 4.13). Les bandes blanches indiquent un souffle, l'absence de voix ou une perte vocale produite par une fermeture excessive de la glotte. Dans la figure 4.15, qui illustre le tracé de récurrence croisée entre le signal glottique normal et celui d'un patient présentant des nodules des CVs, nous avons utilisé un échantillon de voyelle dans un état avancé de thérapie vocale afin d'observer les changements du signal glottique pathologique en le comparant à une voix normophonique. Il est clair à travers la figure, l'occurrence de la ligne diagonale principale, qui indique le début de similitudes entre les deux signaux. On remarque également des lignes diagonales secondaires discontinues. Celui-ci signifie l'ajustement de la fréquence fondamentale ce qui exprime l'effet de la rééducation.

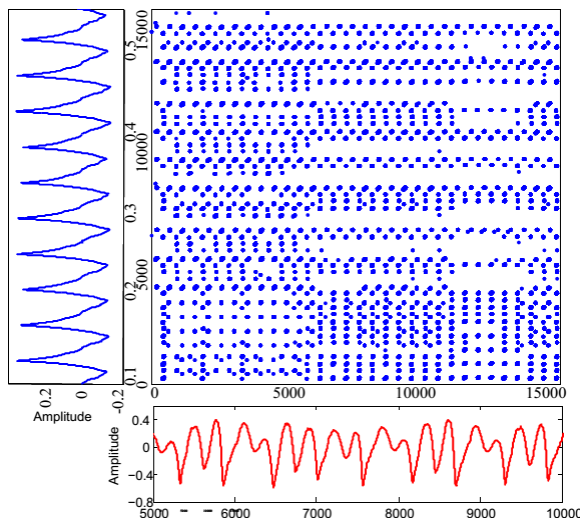


Figure 4.12 : Tracé de Récurrence Croisé entre une voix Normale et Polypes des CVs

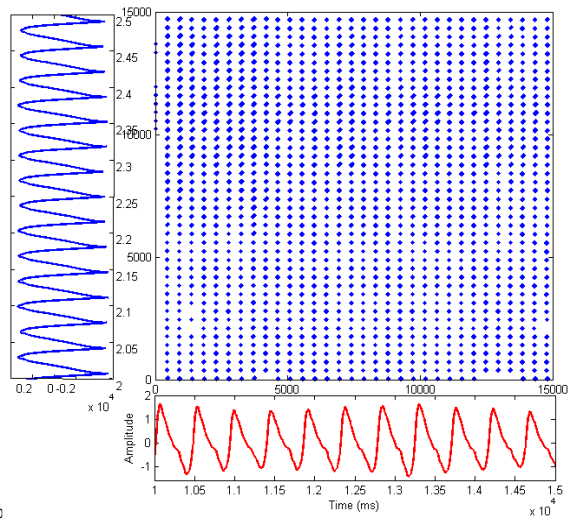


Figure 4.13 : Tracé de Récurrence Croisé entre une voix Normale et dysphonie spasmodique

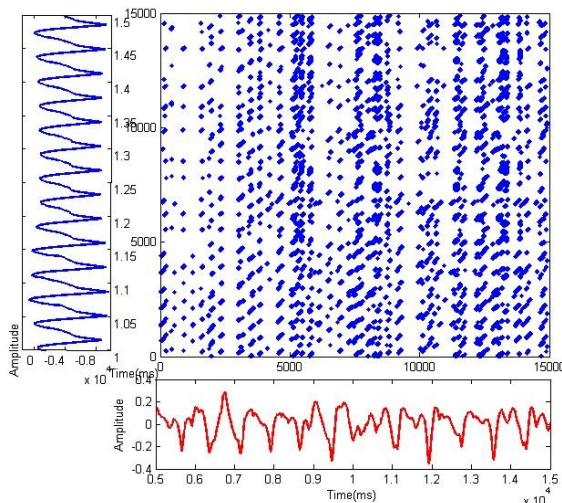


Figure 4.14 : Tracé de Récurrence Croisé entre une voix Normale et Paralysie des CVs

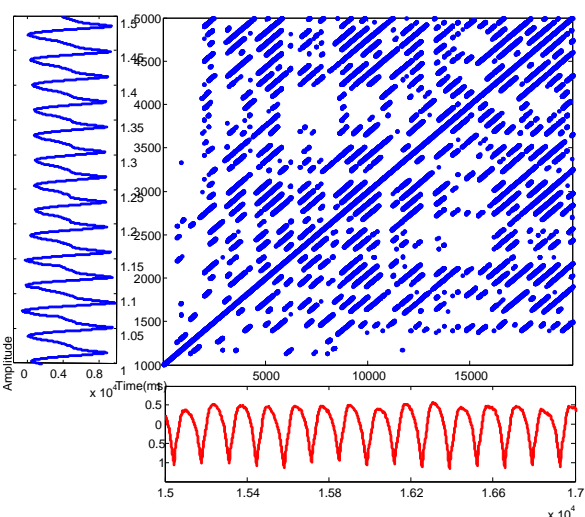


Figure 4.15 : Tracé de Récurrence Croisé entre une voix Normale et Nodules des CVs

## 5.2 Résultats et discussions

Le but principal de la rééducation orthophonique est de récupérer les automatismes normaux du comportement de production vocale en diminuant l'effort et le stress vocal afin d'établir une rétroaction auditive appropriée. De nos jours, les médecins en ORL (Oto Rhino Laryngologiste), les laryngologistes et les orthophonistes approuvent l'importance de la rééducation vocale et l'appliquent à plusieurs troubles de la voix. C'est le choix le plus souvent réalisé. Cette option ne compromet pas une intervention chirurgicale postérieure et demande moins de temps d'arrêt de travail. Elle est aussi moins traumatisante. La rééducation vocale consiste en différentes étapes de traitement, assistées par ordinateur, qui intègrent l'hygiène vocale, la relaxation, l'identification et l'élimination des abus vocaux, l'entraînement vocal et respiratoire classique appliquant des tentatives en transférant des comportements vocaux nouvellement appris à des situations de la vie quotidienne. La durée du traitement est très variable et associée au type de traitement et cas : de dix séances consécutives à plus de deux ans.

L'objectif de cette partie parie du travail est de prendre en charge les patients dysphoniques pendant la phase de rééducation et avant toute intervention chirurgicale, en aidant des spécialistes dans leur diagnostic et leur prise de décision. Cette étude portait sur l'évolution vocale de soixante (60) patients adultes (hommes et femmes) présentant une symptomatologie dysfonctionnelle (paralysie des CVs :  $n = 13$ , dysphonie spasmodique :  $n = 13$ ) ou une dysphonie organique (nodules des CVs :  $n = 18$ , Polypes des CVs :  $n = 16$ ) enregistrées en milieu hospitaliers algériens.  $n$  représente le nombre de patients (hommes et femmes).

En appliquant l'analyse quantitative de récurrence croisée sur la forme d'onde du signal glottique pour les différents groupes de patients avant et après la rééducation et en comparant ces paramètres au groupe témoin (voix normale), l'effet de la thérapie vocale est clairement observé grâce à une analyse paramétrique.

Les résultats obtenus sont regroupés dans le tableau 4.6. Le tableau montre que toutes les valeurs ont été améliorées par rapport à celles mesurées avant l'opération d'orthophonie en fonction du degré de perturbation de la voix. Avant la thérapie vocale, Nous notons que les valeurs moyennes les plus faibles des paramètres CRR, DET et L sont enregistrées respectivement dans le cas des pathologies : Nodule, Dysphonie spasmodique et Paralysie FV. Les valeurs les plus faibles de Liminalité croisée,  $L_{max}$  et TT sont observées dans le cas de la paralysie des cordes vocales et de la dysphonie spasmodique. Pour ENTR, nous remarquons des changements significatifs avant et après la thérapie vocale dans tous les cas de pathologies. Après l'opération de thérapie vocale, l'amélioration de la voix se traduit soit par une augmentation soit par une diminution de la valeur du paramètre en fonction de la valeur normative propre au groupe témoin (sujets homophoniques). Le CRR et l'ENTR croisé, ont été améliorés dans les quatre groupes de patients et une différence statistique a été principalement observée dans les polypes et nodules de CVs.

**Tableau 4.6 :** Paramètres d'Analyse Quantitative de récurrence croisée entre une voix et patients avec différentes pathologies vocale avant et après une rééducation orthophonique (valeur moyenne  $\pm$  la déviation standard)

Groupe	rééducation	CRR	DET	L	LAM
Polypes des Cordes vocales	Avant	0.0126 $\pm$ 0.0092	0.9906 $\pm$ 0.0095	0.9727 $\pm$ 0.0254	6.4511 $\pm$ 3.7755
	Après	0.0339 $\pm$ 0.0127	0.9985 $\pm$ 0.0025	1.225 $\pm$ 0.2101	9.3104 $\pm$ 2.2512
Paralysie des cordes vocales	Avant	0.0165 $\pm$ 0.0064	0.9716 $\pm$ 0.0125	0.9245 $\pm$ 0.1125	4.0673 $\pm$ 1.2265
	Après	0.0318 $\pm$ 0.0154	0.9926 $\pm$ 0.0254	0.9993 $\pm$ 0.0214	11.2366 $\pm$ 3.1522
dysphonique Spasmodique	Avant	0.0183 $\pm$ 0.0055	0.9476 $\pm$ 0.0135	0.9622 $\pm$ 0.1540	5.5239 $\pm$ 2.7418
	Après	0.0330 $\pm$ 0.0122	0.9976 $\pm$ 0.0022	1.4332 $\pm$ 0.2015	9.2254 $\pm$ 2.1421
Nodules des cordes vocales	Avant	0.0100 $\pm$ 0.0052	0.9912 $\pm$ 0.0192	0.9694 $\pm$ 0.0120	7.1654 $\pm$ 0.1655
	Après	0.0303 $\pm$ 0.0021	0.9965 $\pm$ 0.0092	1.0226 $\pm$ 0.1445	8.2600 $\pm$ 1.2205
Tous les patients	Avant	0.0740 $\pm$ 0.0065	0.9752 $\pm$ 0.0136	0.9572 $\pm$ 0.1003	5.8019 $\pm$ 1.9773
	Après	0.0322 $\pm$ 0.0106	0.9963 $\pm$ 0.0098	1.1700 $\pm$ 0.1925	9.5081 $\pm$ 2.1915
		Lmax	TT	Vmax	RNTR
Polypes des Cordes vocales	Avant	24889 $\pm$ 5225	25.7472 $\pm$ 6.0221	24 $\pm$ 06	3.3926 $\pm$ 1.5488
	Après	25055 $\pm$ 10225	27.4155 $\pm$ 9.4877	30 $\pm$ 05	4.0215 $\pm$ 0.7658
Paralysie des cordes vocales	Avant	15732 $\pm$ 2665	18.3325 $\pm$ 5.3325	18 $\pm$ 02	2.7144 $\pm$ 0.9894
	Après	17544 $\pm$ 2548	23.1447 $\pm$ 11.2254	21 $\pm$ 08	3.9755 $\pm$ 1.2247
Spasmodie dysphonique	Avant	13141 $\pm$ 2900	21.2548 $\pm$ 8.2554	21 $\pm$ 04	3.1636 $\pm$ 1.0788
	Après	18362 $\pm$ 0955	23.0995 $\pm$ 16.6658	28 $\pm$ 05	4.1250 $\pm$ 1.9220
Nodules des cordes vocales	Avant	19224 $\pm$ 1540	27.2258 $\pm$ 2.6598	22 $\pm$ 15	3.6602 $\pm$ 1.2450
	Après	19100 $\pm$ 0145	29.3695 $\pm$ 7.1654	27 $\pm$ 10	3.9544 $\pm$ 1.2356
Tous les patients	Avant	18246 $\pm$ 3082	21.1400 $\pm$ 5.5674	21 $\pm$ 06	3.9827 $\pm$ 1.2155
	Après	18665 $\pm$ 3468	27.2573 $\pm$ 11.1360	26 $\pm$ 07	4.0191 $\pm$ 1.2870

## 6. Conclusion

Ce chapitre nous a permis de présenter trois contributions concernant l'évaluation et la classification automatiques des pathologies des cordes vocales. Les approches proposées sont basées sur l'exploitation des caractéristiques du signal glottiques en appliquant différentes techniques d'extraction de paramètres décrivant le comportement des cordes vocales dans la présence ou absence d'une anomalie.

La première contribution consiste en une approche d'analyse et de classification des pathologies vocales, basée sur les paramètres articulatoires (Instants d'ouverture et fermetures des cordes vocales) en se référant à la forme de l'onde glottique.

En deuxième lieu, une analyse approfondie du signal glottique basée sur le diagramme de récurrence et l'analyse quantitative de récurrence RQA a été établie. Cette méthode a démontré une efficacité indubitable pour l'étude de la dynamique des cordes vocales. Huit indicateurs de la RQA ont été examinés à travers différents pathologies. Les valeurs moyennes des indicateurs de la RQA sont utilisées comme entrées à trois différentes méthodes de classification. Les résultats obtenus annoncent que la combinaison de la RQA, comme méthode d'extraction des caractéristiques, ainsi que de la PSO-SVM ou KNN, comme méthodes de classification, et de choix optimal assurant de bonnes performances de classification.

Une dernière étude préliminaire portant sur l'évaluation des pathologies des cordes vocales avant et après une rééducation orthophonique en utilisant l'analyse quantitative de récurrence croisée appliquée au signal glottique localisation proposée. Les performances de cette méthode d'analyse ont été évaluées en calculant les paramètres RQA croisés en les comparants à la voix saine. Les résultats obtenus confirment la capacité de cette technique à caractériser le suivi de la qualité vocale pendant les séances de rééducations orthophoniques.

*Conclusions Générales et  
Perspectives*

Le travail réalisé dans le cadre de cette thèse a été consacré au développement de nouvelles méthodes dédiées à la classification et l'évaluation automatique de la voix pathologiques. Pour atteindre cet objectif nous avons considéré quatre (04) types de voix pathologiques dont deux (02) fonctionnelles (dysphonie spasmodique et paralysies des cordes vocales) qui ne présentent aucun signe organique aux niveaux des cordes vocales, ce qui rend le processus de diagnostic difficile. Les deux (02) autres pathologies organiques (Nodules et Polypes des cordes vocales) causé par l'apparition de certaines lésions sur les cordes vocales.

Dans une première étape, les informations obtenues par le tracé du diagramme de Récurrence Plot (RP) et l'Analyse Quantitative de Récurrence (RQA), faisant référence à la forme d'onde du signal glottique, sont explorées pour la détection et la classification des troubles de la voix humaine. Pour cela, le signal glottique a été extrait premièrement du signal vocal à l'aide d'un algorithme de filtrage inverse : Iterative Adaptive Inverse Filtering (IAIF). Le RP et le RQA ont été appliqués au signal glottique pour évaluer le processus dynamique des Cordes Vocales(CVs). Huit (08) paramètres RQA ont été utilisés comme un vecteur de caractéristiques d'entrée d'un algorithme Particle Swarm Optimization-Support Vector Machines (PSO-SVM) pour séparer les voix normales et pathologiques. De telles études offrent ainsi aux spécialistes de la voix ou de la parole, des données objectives qui permettent d'estimer le degré de perturbation et d'apporter les solutions nécessaires pour y remédier. Dans une deuxième étape, nous proposons un modèle paramétrique basé sur l'Analyse par Quantification Récurrence Croisée (CRQA) pour montrer comment elle peut être adaptée pour comparer, évaluer et quantifier les changements d'articulation des plis vocaux après une rééducation orthophonique.

Ce travail a pour objectif la caractérisation des voix pathologiques en vue de leur exploitation en réhabilitation de la parole, la conduite de diagnostics automatiques et l'établissement de systèmes expert permettant de caractériser de façon fiable les anomalies vocales en milieu hospitalier algérien. Soulignons qu'en Algérie, peu de travaux ont été réalisés dans ce domaine.

Les résultats ont atteint une précision d'environ 97,41% en taux de classification du technique PSO-SVM, alors que 96,21% et 95,00% dans les classificateurs SVM et KNN respectivement. Ces résultats ont démontré clairement l'efficacité de l'approche proposée pour le diagnostic et la classification des pathologies des cordes vocales. L' ENTR et le RR ont présenté une forte capacité discriminative des pathologies étudiées. Les mesures Cross-RQA ont offert un excellent outil pour évaluer les troubles de la voix avant et après la rééducation vocale. La solution proposée permet également la discrimination des différentes étapes lors du protocole de rééducation.

En perspectives la méthodologie proposée a été appliquée à la Base de Données SVD et une petite base de données que nous avons élaboré locale avec des maladies spécifiques en raison de sa forte incidence. Cependant, il pourrait être étendu à d'autres maladies, qui feront l'objet de futures investigations avec de nouvelles expériences sur un plus grand nombre de

patients. Pour cela, les auteurs travaillent en collaboration avec le personnel médical du service ORL et orthophonistes dans les hôpitaux algériens pour collecter un grand nombre de sujets de différents âges et maladies. Cela nous permet non seulement de créer une nouvelle base de données de voix pathologiques, mais également d'appliquer nos méthodes proposées à l'examen des cliniciens et des orthophonistes en milieux hospitaliers.

Enfin, il nous paraît indispensable de proposer une nouvelle version de cette approche plus ergonomique et dont la portabilité lui permet d'être utilisée par les professionnels dans la pratique clinique. En effet, et malgré les différentes solutions proposées dans la littérature, peu d'outils d'évaluation et diagnostic automatique de la voix pathologique sont utilisés dans la pratique clinique. Le développement d'une solution facile à utiliser pourrait permettre à la fois une meilleure évaluation des performances des approches proposées et la constitution de larges corpus nécessaires pour l'amélioration de ces approches.



## *Références Bibliographiques*

- [1] I. Laaridh “Évaluation de la parole dysarthrique : Apport du traitement automatique de la parole face à l’expertise humaine,” Thèse de Doctorat, Université d’Avignon et des Pays de Vaucluse, 2016.
- [2] G. S. Do Amaral Catani, R. Hamerschmidt, A. Moreira, j. Timi, G. R. Wiemes, J. Ido, E. Macedo, “Subjective and objective analyses of voice improvement after phonosurgery in professional voice users,” *Med Probl Perform Art.* Vol 31, no.1, pp.18-24, 2016.
- [3] J. Hillenbrand and R. A. Houde, “Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech”, *J. Speech Lang. Hear. R.*, vol. 39, no. 2, pp. 311–321, 1996.
- [4] M. Dahmani, M. Guerti, “Vocal folds pathologies classification using naive bayes networks,” *IEEE 6th International Conference on Systems and Control (ICSC)*, pp. 426-432, Batna Algeria, 2017.
- [5] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, z. Ali, K. H. Malki, T. A. Mesallam, M. F. Ibrahim, “Voice pathology detection and classification using auto-correlation and entropy features indifferent frequency regions,” *IEEE Access*, Vol. 6, pp. 6961-6974, 2018.
- [6] M. Dahmani and M. Guerti, "Recurrence Quantification Analysis of Glottal Signal as non Linear Tool for Pathological Voice Assessment and Classification," *The International Arab Journal of Information Technology*, Vol. 17, No. 6, pp. 857–866, Nov. 2020.
- [7] J.P. Eckmann, S. O. Kamphorst, D. Ruelle, “Recurrence plots of dynamical systems,” *EPL (Europhysics Letters)*, Vol. 4, no.9, pp. 973-987, 1987.
- [8] J. Webber, C. L. Zbilut, “Recurrence quantification analysis of nonlinear dynamical systems, *Proceedings of the National Academy of Sciences*,” Vol. 12, pp. 26-94, 2005.
- [9] L. Tang and J. Liang, “C-C Method to Phase Space Reconstruction Based on Multivariate Time Series”, *2nd International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 438 – 441, Harbin, 25-28 July, 2011.
- [10] A. Chaou, A. Mekhaldi, M. Tegar, “Recurrence quantification analysis as a novel LC feature extraction technique for the classification of pollution severity on HV insulator model,” *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 22, no. 6, pp. 3376-3384, 2015.
- [11] U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu, P. C. A. Ang, “Application of recurrence quantification analysis for the automated identification of epileptic EEG signals,” *International journal of neural systems*, Vol. 21, no.03, pp.199-211, 2011.
- [12] C. de A. Washington, F. M. DE Assis, B. G. Aguiar, “Pathological Voice Classification Based on Recurrence Quantification Measures. ” 2012.
- [13] C. D. A. Washington, F. M. Assis, B. G. A. Neto, S. C. Costa, & V. J. D. Vieira, “Pathological voice assessment by recurrence quantification analysis,” In *2012 ISSNIP biosignals and biorobotics conference BRC* (pp. 1-6), IEEE, (2012, January).

- [14] A. Giovanni, M. Ouaknine, R. Garrel, “Physiologie de la phonation.” Encyclopédie médico-chirurgicale (Elsevier SAS, Paris, tous droits réservés), Oto-rhino-laryngologie, pp 20-632, 2003.
- [15] S. Fraj, “ Synthèse des voix pathologiques” thèse de doctorat, Université Libre de Bruxelles, 2010.
- [16] H. Gray, “Anatomy of the human body,” Philadelphia: Lea & Febiger ; Bartleby.com, 2000. [www.bartleby.com/107/](http://www.bartleby.com/107/).
- [17] L. Delebecque, “ Étude, analyse et modélisation physique de la production de la parole avec applications aux troubles liés à une surdité profonde”, Thèse de doctorat, laboratoire GIPSA-Lab, Université de Grenoble Alpes, 2015.
- [18] D. Chevalier, F. Dubrulle, B. Vilette, “Anatomie descriptive, endoscopique et radiologique du larynx. ”, EMC Oto-rhino-laryngologie, 2001.
- [19] L. Bailly, “Interaction entre cordes vocales et bandes ventriculaires en phonation : exploration in-vivo, modélisation physique, validation in-vitro.” Thèse de doctorat en acoustique, Ecole doctorale de l’université du Maine, 2009.
- [20] F. Alipour, D. A. Berry and I. R. Titze, “A finite-element model of vocal-fold vibration,” J. Acous. Soc. Am., vol. 108, no. 6, pp. 3003–3012, Dec. 2000.
- [21] L. Anna, S. REYT, “Etude comparative de la production des voyelles entre locuteurs sains et dysphoniques : Effet du contexte de réalisation et de la composante de souffle,” Université Claude Bernard Lyon 1, 2008.
- [22] T. Dubuisson “Glottal Source Estimation and Automatic Detection of Dysphonic Speakers,” PhD thesis University of Mons, Faculty of Engineering, TCTS Lab, 2011.
- [23] J. H. Deller, J. G. Proakis and J. H. L. Hansen, “Discrete-time processing of speech signals,” Prentice-Hall, 1993.
- [24] G. Fant, J. Liljencrants and Q. Lin, “A four-parameter model of glottal flow,” STLQPSR, vol. 4, pp. 1–13, 1985.
- [25] D. Klatt and L. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” J. Acous. Soc. Am., vol. 87, no. 2, pp. 820–857, Feb. 1990.
- [26] A. E. Rosenberg, “Effect of glottal pulse shape on the quality of natural vowels,” J. Acous. Soc. Am., vol. 49, no. 2, pp. 583–590, 1971.
- [27] R. Veldhuis, “A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation,” J. Acous. Soc. Am., vol. 103, no. 1, pp. 566–571, Jan. 1990.
- [28] P. G. Vilda, R. Martínez, F. Díaz, C. Lázaro, A. Alvarez, V. Rodellar, V. Nieto, “Estimation of vocal cord biomechanical parameters by non linear inverse filtering of voice,” In : Proc. of the 3<sup>rd</sup> Int. Conf. on Non-Linear Speech Processing (NOLISP’05), pp. 174–183, Barcelona, Spain, 19-22 April 2005.
- [29] S. Fraj, “ Synthèse des voix pathologiques,” Thèse de Doctorat, Université Libre de Bruxelles, 2010.

- [30] R. Veldhuis, "A computationally efficient alternative for the liljencrants-fant model and its perceptual evaluation," *Journal of Acoustics. Society* Vol 103, pp 566–571, 1998.
- [31] M. Boudjerda, "Analyse du signal de parole pour l'évaluation automatique des voix pathologiques," Thèse de Doctorat, Université Mohammed Seddik Benyahia – Jijel, 2018.
- [32] D. Y. Wong, J. Markel, J. A. Gray, "Least squares glottal inverse filtering from acoustic speech waveform," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol 27, pp 350–355, 1979.
- [33] T. Dubuisson, T. Drugman, T. Dutoit, "On the Mutual Information of Glottal Source Estimation Techniques for the Automatic Detection of Speech Pathologies," In *Proceedings of Maveba* Vol 09, 2009.
- [34] E. Moore, M. Clements, "Algorithm for Automatic Glottal Waveform Estimation without the Precise Glottal Closure Information," In *Proceedings of ICASSP 2004*, vol 14, pp 492–501, 2004.
- [35] P. Alku, E. Vilkman, U. K. Laine, "Analysis of glottal waveform in different phonation types using the new IAIF method," *Proceeding of International Congress of Phonetic Sciences (ICPhS)*, Aix-en-Provence, France 1991.
- [36] P. Alku, "Glottal Wave Analysis With Pitch Synchronous Iterative Adaptive Inverse Filtering," *Speech Communication*, Vol. 11, 3, Pp. 109-118.1992.
- [37] A. Kacha, "Analyse de la parole continue en vue de la caractérisation des troubles de la voix," Thèse de Doctorat, Université Libre de Bruxelles, 2006.
- [38] <http://www.orl-parc.com/Activite-ORL-Medicale-Pathologie-Voix-Cordes-Vocales>
- [39] [http://www.edimark.fr/phototheque/galerie\\_detail.php?id\\_galerie=998](http://www.edimark.fr/phototheque/galerie_detail.php?id_galerie=998)
- [40] S. D. Courbiere médecin ORL, F. Resuel, "la voix : la corde vocal et sa pathologie," *Collège international de médecine et chirurgie de l'hôpital américain de paris*, ISBN 2-905238-23-2.
- [41] A. Hajjij "la voix : la corde vocal et sa pathologie Place de la rééducation orthophonique dans la prise en charge des dysphonies ; étude perspective à propos de 42 cas," Thèse de doctorat, faculté de médecine et pharmacie, Université Sidi Mohammed Ben Abdellah, Fes, 2011.
- [42] M. S. De Bodt, F. L. Wuyts, P. H. Van de Heyning and C. Croux, "Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual ratings of voice quality", *J. Voice*, vol. 11, no. 1, pp. 74–80, 1997.
- [43] G. 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, 2008.
- [44] P. Dejonckere, C. Obbens, G. De Moor and G. Wieneke, "Perceptual evaluation of dysphonia : reliability and relevance," *Clin Linguist. Phonet*, vol. 45, pp. 76–83, 1993.
- [45] A. Reyt "la voix : la corde vocal et sa pathologie. Cancers du larynx : Diagnostique, principes de traitement," Professeur à la faculté de médecine de Grenoble, Mars 2005.

- [46] C. Finck, “ L’évaluation fonctionnelle des dysphonies d’origine laryngé,” Service d’Oto-rhino-laryngologie, CHU Sart Tilman, Université de Liège, Liège, Belgium.
- [47] S. Duvvuru, M. Erickson, “The Effect of Timbre, Pitch, and Vibrato on Vocal Pitch-Matching Accuracy,” *Journal of Voice*, Vol 30(3), 378- e1, 2016.
- [48] S. P. Davis & P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, pp. 357-366, Aug. 1980.
- [49] S. Furui, “Cepstral analysis technique for automatic speaker verification,” *IEEE Transactions on Signal Processing*] 29(2), 254–272. 1981.
- [50] B. L. Tseng & al, “Continuous probabilistic acoustic map for speaker recognition,” *Proceedings of the IEEE, ICASSP*, vol. 2, pp. 161-164, 1992.
- [51] F. K. Soong, A. E. Rosenberg, “On the use of instantaneous and transitional spectral information in speaker recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 6, pp. 871-879, 1988.
- [52] M. Dahmani, M. Guerti, “Automatic Classification of Vocal Folds Pathologies based on Glottal Signal Parameters using the K-Nearest Neighbors Method,” *International Conference on Health Science and Medical Technologies (ICHSMT)*, Tlemcen, Algeria 2018.
- [53] M. Dahmani, M. Guerti, “Glottal signal parameters as features set for neurological voice disorders diagnosis using K-Nearest Neighbors (KNN),” *IEEE 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pp 1-5, Algiers, Algeria 2018.
- [54] J. P. Eckmann, S.O. Kamphorst and D. Ruelle, “Recurrence plots of dynamical systems”, *Europhysics Letters* 4, Vol. 9, pp. 973-977, 1987.
- [55] J. Belaire Franch, D. Contreras, L. Tordera - Lledó, “Assessing nonlinear structures in real exchange rate using recurrence plot strategies”, *Physica D: Nonlinear Phenomena*, No. 171, Vol. 4, pp. 249-264. 2002
- [56] N. Marwan, M. C. Romano and M. Thiel, “Recurrence Plots for the Analysis of Complex Systems”, *Physics Reports*, Vol. 438, pp.237-329, 2007.
- [57] Jr. C. L. Webber and J. P. Zbilut, “Recurrence quantification analysis of nonlinear dynamical systems”, *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pp. 26-94. 2005.
- [58] H. Yang, “Multiscale recurrence quantification analysis of spatial cardiac vector cardiogram signals,” *IEEE Transactions on Biomedical Engineering*, Vol. 58, no. 2, pp. 339-347, 2011.
- [59] O. Afsar, U. Tirnakli, N. Marwan, “Recurrence Quantification Analysis at work : Quasi-periodicity based interpretation of gait force profiles for patients with Parkinson disease.” *Scientific reports*, 8(1), 9102.2018.
- [60] M. Dahmani and M. Guerti, "Cross-Recurrence Plots and Quantification of Glottal Signal for Pathological Voice Assessment,” *Traitment du signal*, Vol. 37, No. 2, pp. 307-317, 2020.

- [61] O. Le Bot, C. Gervaise, J. I. Mars, "Time-difference-of-arrival estimation based on cross recurrence plots, with application to underwater acoustic signals," In *Recurrence Plots and Their Quantifications : Expanding Horizons*, pp. 265-288, Springer, Cham. (2016).
- [62] V. Vapnik, "Statistical learning theory," John Wiley and Sons, New York, 1998.
- [63] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learn.*, Vol. 20, pp. 273-297, 1995.
- [64] J. Louradour, "Noyaux de séquences pour la vérification du locuteur par Machines à Vecteurs de Support», thèse de Doctorat de l'Université Toulouse III - Paul Sabatier, 2007.
- [65] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 2nd ed., Academic Press, Elsevier, Ch.3, 2003.
- [66] B.E. Boser, I.M. Guyon, V.N. Vapnik, "A training algorithm for optimal margin classifiers". In *COLT'92*, pages 144-152, 1992.
- [67] A. Aizerman, E.M. Braverman, L.I. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning", *Automation and Remote Control*, Vol. 25, pp. 821-837, 1964.
- [68] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London*, 209:415-446, 1909.
- [69] D. Kharroubi, "Etude de techniques de classement : Machines à vecteurs supports pour la vérification automatique du locuteur, " Thèse de Doctorat Ecole Nationale Supérieure des Télécommunications ParisTech (2002).
- [70] S. Knerr, L. Personnaz, G. Dreyfus. "Single-layer learning revisited: A stepwise procedure for building and training a neural network," In F. Fogelman-Soulié and J. Héroult, editors, *Neuro-computing : Algorithms, Architectures and Applications*, volume F68 of NATO ASI Series, pages 415-50. Springer-Verlag, 1990.
- [71] E. Mayoraz and E. Alpaydm. "Support vector machines for multi-class classification, " Technical Report 98-06, IDIAP, 1998.
- [72] T.G. Dietterich and G. Bakiri. "Solving multiclass learning problems via error correcting output codes," *Journal of Artificial Intelligence Research*, 2 : 263-286, 1995.
- [73] T. M. Cover, P. E. Hart, "Nearest neighbor pattern classification", *IEEE Trans. Inf. Theory*, Vol. 13, No. 1, pp. 21-27, 1967.
- [74] S. Hatko, "k-Nearest Neighbour Classification of Datasets with a Family of Distances," Mémoire de master en sciences Mathématiques, Université de Ottawa, Canada 2015.
- [75] K. S. Bhakre, A. Bang "Emotion Recognition on The Basis of Audio Signal Using Naive Bayes Classifier" 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, India.
- [76] M. Moreira, E. Mayoraz, "Improved pairwise coupling classification with correcting classifiers," *European Conference on Machine Learning*, Chemnitz, Germany, pp. 160-171, April 21-24, 1998.

- [77] S. Theodoridis, K. Koutroumbas, "Pattern Recognition," 2nd ed, Academic Press, Elsevier, Ch.3, 2003.
- [78] J. Kennedy, R. C. Eberhart, Y. Shi, "Swarm Intelligence," San Francisco : Morgan Kaufmann Publishers, 2001.
- [79] J. Kennedy and R.C. Eberhart, "Particle Swarm Optimization," Proceedings of the IEEE International Conference on Neural Networks, vol. 4. IEEE Press, 1995.
- [80] R. Mendes, J. Kennedy, and J. Neves. "The Fully Informed Particle Swarm: Simpler, Maybe Better". *IEEE Trans. Evolutionary Computation*, Vol. 8, No. 3, pp. 204–210, 2004.
- [81] Y. Shi, R. Eberhart "A modified particle swarm optimizer," Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence, IEEE International Conference, Anchorage, AK, pp. 69 - 73 4-9 May 1998.
- [82] J. Kennedy and R. Eberhart, "Particle swarm optimization," IEEE Int'l. Conf. Neural Networks, Perth, pp. 1942-1948, 1995.
- [83] Y. Cooren, "Perfectionnement d'un algorithme adaptatif d'Optimisation par Essaim Particulaire, Applications en génie médical et en électronique," thèse de Doctorat, Université Paris-Est Créteil, 2008.
- [84] Abbas El Dor : "Perfectionnement des algorithmes d'optimisation par essaim particulaire : applications en segmentation d'images et en électronique ", thèse de doctorat, Université Paris-Est, 2012.
- [85] K.E. Parsopoulos and M.N. Vrahatis, "Particle Swarm Optimization and Intelligence: Advances and Applications", New York, Information Science Reference; 2010.
- [86] Massachusetts. Eye and Ear Infirmary, "Voice disorders database," Kay Elemetrics Corp, 1994.
- [87] P. Manfred and J. B. William, "Saarbruecken Voice Database," [Online]. Available : [http://www.stimmdatenbank.coli.uni-saarland.de/help\\_en.php4](http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4).
- [88] MAPACI, "Voice Disorder Database," Available: <http://www.Mapaci.com/index-ingles.php>
- [89] Matti Airas, "TKK Aparat : An environment for voice inverse filtering and parameterization. Logopedics Phoniatrics Vocology, " 33(1), pp. 49-64, 2008.