REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOHAMED SEDDIK BEN YAHIA–JIJEL

N° d'ordre : …….
Série : ……..

THESE

Présentée à la Faculté des Sciences et de la Technologie
Département d'Electronique
En vue d'obtention du diplôme de Doctorat en Sciences
Spécialité : Electronique
Par :

BENHAMMOUD Redouane

Thème

Classification automatique de la parole continue basée sur des indices acoustiques en vue de la caractérisation des troubles de la voix

Soutenue le :……..……., devant le jury composé de :

M. BOUKABOU Abdelkrim	Professeur	U. Mohamed Seddik Ben Yahia-Jijel	Président
M. KACHA Abdellah	Professeur	U. Mohamed Seddik Ben Yahia-Jijel	Rapporteur
M. BRAHIMI Tahar	M.C.A	U. Mohamed Seddik BenYahia-Jijel	Examinateur
M. BENOUDJIT Nabil	Professeur	U. Mustapha Ben Boulaid-Batna 2	Examinateur
M. BOUGUEZEL Saad	Professeur	U. Ferhat Abbas-Sétif 1	Examinateur
M. DAAMOUCHE A/Hamid	Professeur	U. M'hamed Bougara-Boumerdes	Examinateur

Résumé

L'analyse du signal de parole est un moyen privilégié pour l'évaluation clinique de la qualité de la voix en vue d'un diagnostic et d'une documentation quantitative des pathologies du larynx. L'analyse acoustique du signal de parole permet une évaluation non invasive et peu coûteuse des troubles de la voix contribuant ainsi un diagnostic plus efficace, rapide et objectif.

Le travail de recherche présentée dans cette thèse est dédié à la classification automatique des voix pathologiques dans le but de fournir une méthode complémentaire d'évaluation clinique de la qualité de la voix permettant aux cliniciens le suivi des patients atteints de troubles de la voix. La méthode proposée utilise conjointement les séparateurs à vaste marge (SVM : support vector machine) et les modèles de Markov cachés (HMM : hidden Markov models) afin de profiter des avantages des deux méthodes simultanément. La classification par SVM est basée sur l'apprentissage automatique supervisé avec pour principe de séparer les données de différentes natures par un hyperplan alors que la classification par HMM basée sur l'apprentissage automatique non supervisé avec pour principe de suivre la règle de décision de Bayes.

Ces méthodes de classifications ont été appliquées sur un corpus contenant 251 fichiers composés de deux phrases concaténées suivies de la voyelle [a] produites par des sujets normophoniques et des sujets dysphoniques avec différents degrés de dysphonie. Les résultats de classification de la méthode hybride HMM-SVM sont comparés aux résultats obtenus par des méthodes de référence basées sur les SVMs et les HMMs séparément. Les résultats montrent que la combinaison de ces deux méthodes de classification améliore nettement le taux de classification.

Mots clés : classification des voix pathologiques, séparateurs à vaste marge, modèles de Markov cachés, indices acoustiques, modèle hybride HMM-SVM, parole continue.

Abstract

The analysis of the speech signal is a privileged means for the clinical evaluation of the quality of the voice with a view to diagnosis and quantitative documentation of pathologies of the larynx. The acoustic analysis of the speech signal allows a non-invasive and inexpensive assessment of voice disorders, thus contributing to a more efficient, rapid and objective diagnosis.

The research work presented in this thesis is dedicated to the automatic classification of pathological voices with the aim of providing a complementary method of clinical assessment of voice quality allowing clinicians to monitor patients with voice disorders. The proposed method jointly uses the support vector machines (SVMs) and hidden Markov models (HMM) in order to obtain advantages of both methods simultaneously. The SVM classification is based on supervised machine learning with the principle of separating data of different natures by a hyperplane whereas the HMM classification is based on unsupervised machine learning with the principle of following the Bayes decision rule.

These classification methods were applied to a corpus containing 251 files composed of two concatenated sentences followed by the vowel [a] produced by normophonic subjects and dysphonic subjects with different degrees of dysphonia. The classification results of the hybrid method are compared with the results obtained by reference methods based on SVM and HMM separately. The results show that the combination of these two classification methods significantly improves the classification rate.

Keywords : pathological voices classification, support vector machine, hidden Markov models, acoustic descriptors, hybrid HMM-SVM model, continuous speech.

ملخص

يعد تحليل إشارة الكلام وسيلة مميزة للتقييم العيادي لجودة الصوت بهدف التشخيص والتوثيق الكمي لأمراض الحنجرة. يسمح التحليل الصوتي لإشارة الكلام بتقييم غير جراحي وغير مكلف لاضطرابات الصوت، مما يساهم في تشخيص أكثر كفاءة وسرعة وموضوعية.

العمل البحثي المقدم في هذه الأطروحة مخصص للتصنيف التلقائي للأصوات المرضية بهدف توفير طريقة تكميلية للتقييم العيادي لجودة الصوت مما يسمح للأطباء بمتابعة المرضى الذين يعانون من اضطرابات الصوت. تستخدم الطريقة المقترحة كلاً من فواصل الهامش العريض (SVM) ونماذج ماركوف المخفية (HMM) من أجل الاستفادة من مزايا كلتا الطريقتين في وقت واحد. يعتمد التصنيف بواسطة SVM على التعلم الآلي الخاضع للإشراف مع مبدأ فصل البيانات ذات الطبيعة المختلفة عن طريق المستوى الفائق أما التصنيف بواسطة HMM يعتمد على التعلم الآلي غير الخاضع للإشراف مع مبدأ اتباع قاعدة قرار Bayes.

تم تطبيق طرق التصنيف هذه على مجموعة تحتوي على 251 ملفًا مكونًا من جملتين متسلسلتين متبوعين بحرف العلة [أ] الذي تم الحصول عليه من قبل الأشخاص الذين يعانون من خلل النطق مع درجات مختلفة. تتم مقارنة نتائج التصنيف بطريقة HMM-SVM الهجينة مع النتائج التي تم الحصول عليها بالطرق المرجعية بناءً على SVMs و HMMs بشكل منفصل. أظهرت النتائج أن الجمع بين هاتين الطريقتين يحسن بشكل كبير من معدل التصنيف.

الكلمات الرئيسية : تصنيف الأصوات المرضية، فواصل الهوامش العريضة، نماذج ماركوف المخفية، الواصفات الصوتية، نموذج HMM-SVM الهجين، الكلام المستمر.

Remerciements

Mes remerciements vont au premier lieu au bon Dieu pour la force, la patience et la volonté qu'il m'a donnée durant ces années d'études.

Ce travail a été effectué au laboratoire de Physique de Rayonnement et Applications de l'université de Jijel. Je tiens à remercier profondément le Professeur KACHA Abdellah pour m'avoir proposé un sujet intéressant, de l'avoir suivi, et de m'avoir aidé à le mener à terme, grâce à ces conseils précieux, ses interventions pertinentes et sa grande compétence dans le domaine de traitement du signal.

Mes remerciements s'adressent aussi à M. BOUKABOU Abdelkrim, Professeur à l'université de Jijel, de m'avoir honoré et accepté de présider le jury de ce travail. Je tiens aussi à remercier M. BRAHIMI Tahar, Maître de conférences classe A à l'université de Jijel, M. BENOUDJIT Nabil, Professeur à l'université de Batna 2, M. BOUGUEZEL Saad, Professeur à l'université de Sétif 1 et M. DAAMOUCHE Abdelhamid, Professeur à l'université de Boumerdes, pour avoir accepté d'examiner et de juger ce travail.

Finalement, je tiens à remercier vivement ma famille et spécialement mon père, ma mère, mes frères et sœurs, ma femme et mes filles adorées pour m'avoir aidé et encouragé durant ces années de préparation du Doctorat.

Table des matières

Résumé……..... i
Remerciements……..…… iv
Table des matières……..…….... v
Liste des figures……...….. x
Liste des tableaux……...… xiii
Liste des abréviations……...….. xiv

Introduction générale……..…... 1

Chapitre 1 : Evaluation des troubles de la voix

I.1. Introduction……..…... 5
I.2. La production de la parole……...……...……. 5
I.2.1. Anatomie de l'appareil phonatoire humain……..……. 6
I.2.1.1. La soufflerie……..…….. 6
I.2.1.2. Le larynx……..…….... 7
I.2.1.3. Les résonateurs……...……. 7
I.2.2. Physiologie du système de production de la parole……..…… 8
I.2.2.1. La vibration des cordes vocales……..…… 8
I.2.2.2. Articulations (résonateurs)……..…… 10
I.3. Modèle de la production de la parole…….……. 10
I.3.1. Modèle source-filtre……..…… 10
I.3.2. Modélisation de la source glottique……..…… 11
I.4. Pathologies du larynx……..…..... 14
I.4.1. Dysphonie organique……...…….. 14
I.4.2. Dysphonie dysfonctionnelle……..…… 15
I.4.3. Dysphonie neurologique……..…. 16
I.5. Examen clinique des patients atteint de troubles de la voix……...…. 16
I.5.1. Interrogatoire avec le patient……...….. 17
I.5.2. Examen physique du patient……..…... 17
I.5.3. Examen du comportement vocal……..... 19
I.6. Evaluation des voix dysphoniques……..…..... 19

I.6.1. Analyse perceptive……..…….. 20
I.6.1.1. Les échelles d'évaluation perceptive……..…… 21
I.6.1.2. Le jury d'écoute……..…… 22
I.6.2. Analyse objective et méthodes instrumentales……...…... 23
I.6.2.1. Méthodes instrumentales analytiques……..…... 24
I.6.2.2. Méthodes instrumentales automatiques……..…… 24
1.7. Classification automatique des voix pathologiques……...…. 25
I.8. Conclusion……...…… 26

Chapitre 2 : Etat de l'art de l'intelligence artificielle et des méthodes de classification

II.1. Introduction……...…. 28
II.2. Historique et application de l'intelligence artificielle……...…. 28
II.3. Apprentissage automatique et apprentissage profond……...…. 30
II.4. Apprentissage automatique (machine learning)……...….. 31
II.4.1. Apprentissage supervisé……...……..... 32
II.4.1.1. Classification……..…...… 32
i) Les séparateurs à vaste marge……...…… 33
ii) K plus proches voisins……..…... 33
iii) Bayes naïf……..……. 34
II.4.1.2. Régression……..... 34
i) Régression linéaire……...……. 35
ii) Arbres de décision……..……. 35
iii) Forets aléatoires……..…… 36
iv) Régression vectorielle de support……..……. 36
II.4.2. Apprentissage non supervisé……...……. 36
II.4.2.1. Partitionnement……..…… 36
i) Partitionnement en K moyennes……..……. 37
ii) Classification ascendante hiérarchique……..…….. 37
iii) les modèles de Markov cachés……..……..... 38
iv) Modèles de mélange gaussien……...…….. 38
II.4.2.2. Réduction de dimension……..... 39
i) Analyse en composantes principales……..…….. 40

ii) Sélection des fonctionnalités……..…. 40
iii) Analyse discriminante linéaire……..…. 41
II.4.3. Apprentissage par renforcement……..…… 41
II.5. L'apprentissage profond et les réseaux de neurones……..... 42
II.5.1. Eléments de base des Réseaux de neurones……...….. 43
II.5.1.1. Le neurone formel……...…... 43
II.5.1.2. Les fonctions d'activation……...…... 43
II.6. Conclusion……..…... 44

Chapitre 3 : Les machines à vecteurs de support

III.1. Introduction……..…… 46
III.2. Historique des SVMs……..……. 46
III.3. Notion d'apprentissage……..……... 47
III.4. Principe de fonctionnement général……..……... 48
III.4.1. Linéarité et non linéarité……..……... 49
III.4.1.1. Cas linéairement séparable……..……. 49
III.4.1.2. Cas non linéairement séparable……..….. 49
III.5. Fondements mathématiques……..…... 50
III.5.1. Principe……..……. 50
III.5.2. La forme primale du SVM……..…… 53
III.5.3. La forme duale du SVM……..……... 56
III.6. Noyaux……..……... 57
III.6.1. Principe du noyau……..……. 57
III.6.2. Exemples de noyaux……..…… 59
III.7. Séparateurs à vaste marge binaires…….…… 60
III.7.1. Séparateur à marge dure……..……... 61
III.7.2. Séparateurs à marge souple…….……... 65
III.8. Séparateurs à vaste marge multi-classes……..……... 67
III.8.1. Un contre tous (1 vs rest)…….……... 68
III.8.2. Un contre un (1vs1)……..…….. 70
III.9. Séparateur à vaste marge mono-classe (Novelty detection)…….…….. 71
III.10. Avantages et inconvénients des SVMs……..…… 76
III.10.1. Avantages des SVMs……..……. 76

III.10.2. Inconvénients des SVMs……..…….. 76
III.11. Conclusion……..…… 76

Chapitre 4 : Les modèles de Markov cachés

IV.1. Introduction……..…… 77
IV.2. Généralités sur les modèles de Markov cachés……..…….. 77
IV.2.1. Modèles de Markov observables……...…. 77
IV.2.2. Modèles de Markov cachés (HMM)……..…… 79
IV.3. Paramètres d'un HMM……...…….. 79
IV.4. Types de topologies des HMM……..…….. 81
IV.4.1. Le modèle ergodique……..…..... 81
IV.4.2. Le modèle gauche-droite……..…….. 81
IV.5. Génération d'une séquence d'état par un HMM……..…… 82
IV.6. Classification de séquences d'observation……..……. 82
IV.6.1. Les trois problèmes des HMM……..……. 82
IV.6.2 La classification bayésienne……...……. 83
IV.7. Evaluation de la probabilité d'observation……...……..... 84
IV.7.1. Evaluation directe……...……..... 84
IV.7.2. Evaluation par les fonctions avant-arrière (forward-backward)…..…….. 84
IV.8. Algorithme de Viterbi : calcul du chemin optimal……...…. 86
IV.9. Apprentissage des HMMs……..…... 88
IV.9.1. Formules de réestimation……..……. 89
IV.9.2. Algorithme de Baum-Welch……..…… 91
IV.10. Classification statistique……...…….. 92
IV.11. Fonctionnement du HMM……..…… 93
IV.12. Hypothèses simplificatrices pour la reconnaissance de la parole……...…… 93
IV.13. Reconnaissance de mots isolés……...…… 94
IV.14. Reconnaissance de parole continue……...……. 96
IV.15. Limitations des HMMs……...…..... 97
IV.16. Conclusion…….……..……... 97

Chapitre 5 : Classification automatique des voix pathologiques par HMM-SVM

V.1. Introduction……..……. 98
V.2. Corpus et évaluation perceptive……...…….. 98
V.3. Classification des voix pathologiques à base de HMM et SVM……..……. 100
V.3.1. Descripteurs acoustiques……..…... 101
V.3.1.1. Coefficients MFCCs (Mel-frequency cepstral coefficients)……..... 102
V.3.1.2. Coefficients LFPCs (log frequency power coefficients)……..... 103
V.3.1.3. Coefficients LPCCs (linear predictive cepstral coefficients)……..... 104
V.3.2. Classification par modèles de Markov cachés……...….. 104
V.3.3. Classification par séparateur à vaste marge……...….. 106
V.3.4. Classification hybride HMM-SVM……...……..… 107
V.3.4.1. Apprentissage……..….. 108
V.3.4.2. Classification……..... 109
V.4. Critères d'évaluation des performances……..... 111
V.5. Configuration expérimentale……..…... 111
V.6. Résultats et discussion……..……. 112
V.7. Conclusion……..…... 118

Conclusion générale……..... 120

Bibliographie……....…….... 122

Liste des figures

Figure I.1	Vue d'ensemble des organes de la voix et de la parole……..……..	6
Figure I.2	Coupe frontale du larynx……...	6
Figure I.3	Mouvements des cordes vocales au cours d'un cycle et à partir de la fermeture glottique……	10
Figure I.4	Modèle source-filtre de production de la parole……...	11
Figure I.5	Modèle de débit glottique (a) et sa dérivé (b)……...	13
Figure I.6	Miroir utilisé en laryngoscopie « indirecte »…….	18
Figure I.7	Laryngoscopie « directe »…….	18
Figure I.8	Les méthodes instrumentales pour l'évaluation objective de la voix dysphonique……..…….....	23
Figure II.1	Concept d'appartenance des différentes branches de l'intelligence artificielle……..…….....	30
Figure II.2	Principaux algorithmes d'apprentissage automatique……...	32
Figure II.3	Neurone formel……...……..	43
Figure II.4	Exemples de fonctions d'activation……..	44
Figure III.1	SVM avec ses différentes variantes……...	46
Figure III.2	Séparation de deux ensembles de points par un Hyperplan H……..	48
Figure III.3	Hyperplan optimal, marge et vecteurs de support…….	49
Figure III.4	Linéarité et non linéarité de séparation…….	50
Figure III.5	Transformation de l'espace de représentation et l'hyperplan séparateur dans le cas non linéairement séparable……	50
Figure III.6	Deux classes de données et les hyperplans les séparant……	54
Figure III.7	Non-adaptation de l'hyperplan aux problèmes réels de classification…….....	58
Figure III.8	Représentation idéale de la fonction de décision……..	58
Figure III.9	Séparateur à vaste marge binaire……...	60
Figure III.10	Séparateur binaire à marge souple…….	65
Figure III.11	Approche un-contre-tous avec des zones d'indécision…….	69
Figure III.12	Résolution des cas d'indécision dans la méthode 1vsR……	69
Figure III.13	Approche de classification un-contre-un……...	71
Figure III.14	Séparation des exemples d'une classe du reste de l'espace……..	72
Figure III.15	Séparateur à vaste marge mono-classe à marge maximale……...	73

Figure III.16	SVM mono-classe à base d'hypersphère……..	75
Figure IV.1	Dynamique d'un modèle de Markov caché……..	79
Figure IV.2	Le HMM à deux états……	80
Figure IV.3	Le HMM à trois états……..….	80
Figure IV.4	Modèle ergodique du HMM……..	81
Figure IV.5	HMM gauche-droite (ou Bakis) à quatre états……..	81
Figure IV.6	Treillis de Viterbis…….	87
Figure IV.7	Exemple de HMM à cinq états. Les états grisés sont émetteurs, les états blancs ne génèrent aucune observation…….	93
Figure V.1	Distribution de fréquence du grade moyen de dysphonie (G) pour le corpus de concaténation des deux phrases suivies de la voyelle [a]…….....	99
Figure V.2	Voyelles [a] produites par des locuteurs de chacune des trois classes…….....	100
Figure V.3	Concaténation des deux phrases produites par des locuteurs de chacune des trois classes……...	101
Figure V.4	Schéma bloc de l'algorithme d'extraction des coefficients MFCC..	103
Figure V.5	Banc de filtres à l'échelle de Mel……..	103
Figure V.6	Schéma bloc de l'algorithme d'extraction des coefficients LFPCs..	103
Figure V.7	Schéma bloc de l'algorithme d'extraction des coefficients LPCCs..	104
Figure V.8	Modèle HMM à 3 états de gauche à droite, chaque état peut émettre une ou plusieurs observations O_i selon une probabilité gouvernée par B……...…….	105
Figure V.9	Schéma illustrant le remplacement des probabilités de sortie des états cachés qi par les probabilités postérieures obtenues par le SVM……..	108
Figure V.10	Architecture générale du système hybride HMM-SVM……...	110
Figure V.11	Comparaison des performances de classification des voyelles [a]… en deux classes en termes de précision avec les MFCCs comme descripteurs acoustiques……	115
Figure V.12	Comparaison des performances de classification de la parole continue en deux classes en termes de précision avec les MFCCs comme descripteurs acoustiques……...	115

Figure V.13 Comparaison des performances de classification des voyelles [a] en trois classes en termes de précision avec les MFCCs comme descripteurs acoustiques…… 118
Figure V.14 Comparaison des performances de classification de la parole continue en trois classes en termes de précision avec les MFCCs comme descripteurs acoustiques……... 118

Liste des tableaux

Tableau I.1	Définition des paramètres de l'échelle GRBAS……...……	21
Tableau V.1	Répartition des voix selon le degré de dysphonie et le type de sexe pour la classification binaire……...….	101
Tableau V.2	Répartition des voix selon le degré de dysphonie et le type de sexe pour la classification en trois classes……...…	101
Tableau V.3	Temps de classification moyen des voyelles [a] en fonction du type de descripteurs et le nombre de classes……....…	112
Tableau V.4	Temps de classification moyen des concaténations de phrases suivies de la voyelle [a] en fonction du type de descripteurs et le nombre de classes……....….	112
Tableau V.5	Précision de classification du classificateur HMM-SVM en deux classes du corpus de la concaténation des phrases suivies de la voyelle [a] pour différentes valeurs de la longueur de trame et différents descripteurs acoustiques……..…….	113
Tableau V.6	Précision de classification du classificateur HMM-SVM en trois classes du corpus de parole continue pour différentes valeurs de la longueur de trame et différents descripteurs acoustiques……..…	113
Tableau V.7	Performances de classification en deux classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité en % pour le corpus de voyelles [a]……..…...	114
Tableau V.8	Performances de classification en deux classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité pour le corpus de parole continue……..…..…	114
Tableau V.9	Performances de classification en trois classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité pour le corpus de voyelles [a]……..……	117
Tableau V.10	Performances de classification en trois classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité pour le corpus de parole continue……..…..…	117

Liste des Abréviations

GRBAS	Grade, Rough, Breathy, Asthenic, Strained
HMM	hidden Markov model
SVM	Support Vector Machine
SVR	Support Vector Regression
RAL	Reconnaissance Automatique du Locuteur
RAP	Reconnaissance Automatique de la Parole
TAP	Traitement Automatique de la Parole
RNA	Réseau de neurones artificiels
MLP	Multi-layer perceptron
PCA	Principal Component Analysis
LDA	Linear discriminant analysis
RL	Reinforcement Learning
ML	Apprentissage automatique
IA	Intelligence artificielle
ML	Machine learning
DL	Deep Learning
ANN	Artificial Neural Network
1vsR	Un contre tous
1vs1	Un contre un
PAC	Probably approximately correct
OCR	Optical character recognition
HAC	Hierarchical Ascendant Clustering
UC	Unité centrale
LDA	Linear discriminant analysis
CNN	Convolutional Neural Network
GCI	Glottal Closure Instant

Introduction générale

Introduction générale

La voix est l'outil de communication le plus important dans la vie humaine. Pour certaines personnes telles que les enseignants, les chanteurs et les employés de centres d'appels, c'est l'outil sur lequel repose leur vie professionnelle. Plus ces personnes utilisent leur voix, plus elles seront affectés. Les troubles de la voix affectent aussi bien les enfants que les adultes. Chez l'enfant, la dysphonie est perçue principalement à partir de deux ans le plus souvent à la suite d'abus vocaux.

Les troubles de la voix peuvent être d'origine fonctionnelle ou organique. Les dysphonies fonctionnelles sont engendrées par une mauvaise utilisation du système vocal anatomiquement et physiologiquement intact et peuvent être corrigées au moyen d'une thérapie de la voix. Les pathologies laryngales organiques peuvent être corrigées au moyen d'une thérapie de la voix ou par une intervention chirurgicale. Dans ce cadre, le développement de méthodes efficaces pour l'évaluation clinique de la voix et le suivi de l'évolution de l'état du patient est d'une grande importance.

Dans le milieu clinique, l'évaluation de la qualité vocale se fait par jugement à l'oreille de la voix et de la parole afin de déterminer le degré du dysfonctionnement vocal. Or, de par sa subjectivité intrinsèque et sa grande variabilité (intra-auditeur et inter-auditeurs), le jugement perceptif reste une méthode controversée [1].

Il existe un lien étroit entre la qualité de la voix perçue et les paramètres du signal acoustique [2-5]. En effet, le dysfonctionnement des cordes vocales causé par les pathologies laryngales entraine une perturbation du signal acoustique. Pour mettre en évidence la relation entre la qualité de la voix et le signal acoustique, une analyse de corrélation est souvent utilisée. L'analyse du signal de parole offre donc un moyen privilégié pour l'évaluation clinique de la qualité de la voix en vue d'un diagnostic et d'une documentation quantitative des pathologies du larynx. Le système d'évaluation utilisant l'analyse acoustique présente plusieurs avantages. En plus de son coût peu élevé, il est simple à mettre en œuvre, il présente un caractère non invasif et documente quantitativement le degré d'enrouement perçu par le clinicien.

L'analyse du signal acoustique a pour objet d'extraire des indices pertinents en vue de déterminer les caractéristiques de la voix permettant de renseigner sur l'état du larynx du locuteur. Plusieurs indices acoustiques sont utilisés pour l'évaluation clinique de la voix. Un nombre de ces indices reflète la déviation du signal de parole voisée par rapport à la périodicité parfaite [6, 7]. Les causes de ces dyspériodicités sont diverses : vibrations non modales des cordes vocales, bruit de modulation comprenant les variations cycle à cycle de

la durée de cycle (jitter) et de l'amplitude (shimmer) dues aux perturbations externes, bruit additif dû à une turbulence excessive [8], etc.

Les méthodes d'analyse acoustique pour l'évaluation clinique de la voix utilisent souvent des fragments stationnaires de voyelles soutenues. La raison est que les voyelles sans attaques et déclins sont faciles à analyser parce que les hypothèses de stationnarité utilisées par les méthodes d'analyse sont valables pour beaucoup de locuteurs. En effet, les caractéristiques de la source vocale et du conduit vocal restent inchangées lors de la production des voyelles soutenue. La popularité de l'analyse des voyelles soutenues au détriment de la parole continue est due à la faisabilité technique et non pas à la pertinence clinique.

La parole continue est considérée comme plus informative que les voyelles soutenues dans le cadre de l'évaluation clinique de la voix. En effet, lors de la production de la parole continue, le larynx fonctionne dans des conditions non stationnaires parce que la vibration des cordes vocales doit commuter continuellement pour donner lieu à l'apparition ou l'extinction du voisement. De même, le larynx doit descendre et monter continuellement. La parole continue contient donc les caractéristiques dynamiques de la source de la voix et du conduit vocal telles que l'attaque et le déclin et les variations de la fréquence fondamentale et de l'amplitude [9].

Au cours des dernières années, beaucoup de travaux de recherche se sont focalisés sur la détection et la classification automatique des voix pathologiques en utilisant les méthodes de la reconnaissance automatique de formes ou les méthodes statistiques. Plusieurs méthodes utilisées dans le domaine de la reconnaissance automatique de la parole ou l'identification du locuteur ont été adaptées à la classification des voix pathologiques [10]. L'objectif est de trouver un classificateur performant en termes de précision de classification. Parmi les méthodes de classification utilisées, les résultats ont montré que les modèles de Markov cachés (HMM : hidden Marcov models) et les séparateurs à vaste marge (SVM : support vector machine) présentent les meilleures performances [11, 12].

La plupart des travaux de recherche sur la classification des voix pathologiques ont considéré le problème de classification binaire normale/pathologique des voix en utilisant des voyelles soutenues [13-19]. L'extraction des indices acoustiques à partir de la parole continue pour classification des voix pathologiques fait recours à des méthodes non stationnaires comme l'analyse conjointe temps-fréquence [20] ou le spectre d'amplitude [21]. Les résultats ont montré que ces classificateurs binaires permettent d'obtenir des précisions élevées [22-26] mais leur extension au cas de plusieurs classes reste un défi.

Dans le travail présenté dans cette thèse, on s'intéresse à la classification automatique des voix pathologiques en deux et trois classes en utilisant la parole continue. Comme alternative aux méthodes de classification à base de SVM et HMM, on propose un classificateur hybride HMM-SVM pour la classification des voix pathologiques en termes de degré de perturbation.

La thèse est organisée comme suit :

Le premier chapitre est consacré à l'évaluation et la classification des troubles de la voix. Un rappel sur l'anatomie et la physiologie de l'appareil vocal est présenté avant d'aborder les troubles de la voix, de la parole et du langage. Les dysphonies sont présentées selon leurs origines, fonctionnelles ou organiques. Les différentes méthodes d'évaluation et de classification des voix pathologiques sont aussi présentées dans ce chapitre.

Le deuxième chapitre présente l'état de l'art de l'intelligence artificielle. Cette science émergente est constituée de l'apprentissage automatique et l'apprentissage profond. Les classificateurs SVM et HMM utilisés dans cette thèse sont basés sur l'apprentissage automatique supervisé et non supervisé.

Le troisième chapitre se focalise sur les machines à vecteurs de support ou séparateurs à vaste marge (SVM : support vector machine), ce sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de discrimination, c'est à dire décider à quelle classe appartient un échantillon, ou de régression, c'est à dire prédire la valeur numérique d'une variable.

Le quatrième chapitre est dédié au fondement théorique des modèles de Markov cachés. Ces modèles sont un outil statistique permettant de modéliser les phénomènes stochastiques. Ils sont utilisés dans de nombreux domaines tels que la reconnaissance et la synthèse de la parole, la biologie, l'ordonnancement, l'indexation de document, la reconnaissance d'image, la prédiction de série temporelle, etc. ils sont aussi très bien adaptés pour la modélisation de la parole continue étant donné leur nature approprié à suivre les événements variables dans le temps. On présente les notations utiles et les principaux algorithmes qui constituent la théorie des HMMs.

Le cinquième chapitre présente les résultats expérimentaux obtenus en utilisant la méthode de classification hybride HMM-SVM et les modèles HMMs et SVMs séparément. On utilise trois types de descripteurs acoustiques comme entrée aux classificateurs et on considère la classification binaire normophonique/dysphonique ainsi que la classification en trois classes. Ces classificateurs sont appliqués à une base de données comprenant des voyelles soutenues et de la parole continue.

On termine par une conclusion générale dans laquelle on résume le travail réalisé, les remarques et les suggestions.

Chapitre 1 :

Evaluation des troubles de la voix

I.1. Introduction

Support privilégié de la communication parlée, la voix a comme rôle primordial de véhiculer nos pensées, de transmettre nos émotions et nos sentiments. Elle nous permet d'exprimer nos émotions tout en donnant des indications sur notre personnalité. C'est pourquoi elle est perçue comme le reflet de notre personnalité, le propre de notre identité. Elle est aussi une fenêtre sur notre santé, physique ou psychique, et peut aussi être porteuse d'informations sur notre situation sociale et culturelle. De par sa dimension esthétique, c'est souvent lorsqu'une altération ou une lésion de la voix apparaît que l'on prend réellement conscience de son importance dans notre vie quotidienne et que nous consultons des spécialistes tels que les ORL, les phoniatres et les orthophonistes.

Ce chapitre est consacré à l'évaluation et à la classification des troubles de la voix. On présente l'anatomie et la physiologie de l'appareil vocal avant d'aborder les troubles de la voix, de la parole et du langage. Les dysphonies, seront présentées selon leur origine fonctionnelle ou organique. L'évaluation des voix dysphoniques sera décrite avec ses différentes méthodes subjectives et objectives.

I.2. Production de la parole

La parole est produite par les mouvements de l'appareil phonatoire. Cet appareil se compose de trois parties distinctes [27], une source, un vibrateur et l'articulation (le résonateur) comme le montre les coupes schématiques des figures I.1 et I.2. La source d'énergie utilisée pour produire les sons est l'air contenu dans les poumons [28, 29]. Le flux d'air sous pression parvient à travers la trachée jusqu'au conduit vocal qui comprend le larynx avec les cordes vocales, des cavités résonantes plus ou moins déformables (pharynx, cavité buccale et fosses nasales, ces dernières pouvant être mises en communication grâce au voile du palais) et des organes d'articulation (langue, lèvres, mâchoires, etc.) permettant d'obturer le canal vocal ou d'en modifier la forme.

I.2.1. Anatomie de l'appareil phonatoire humain

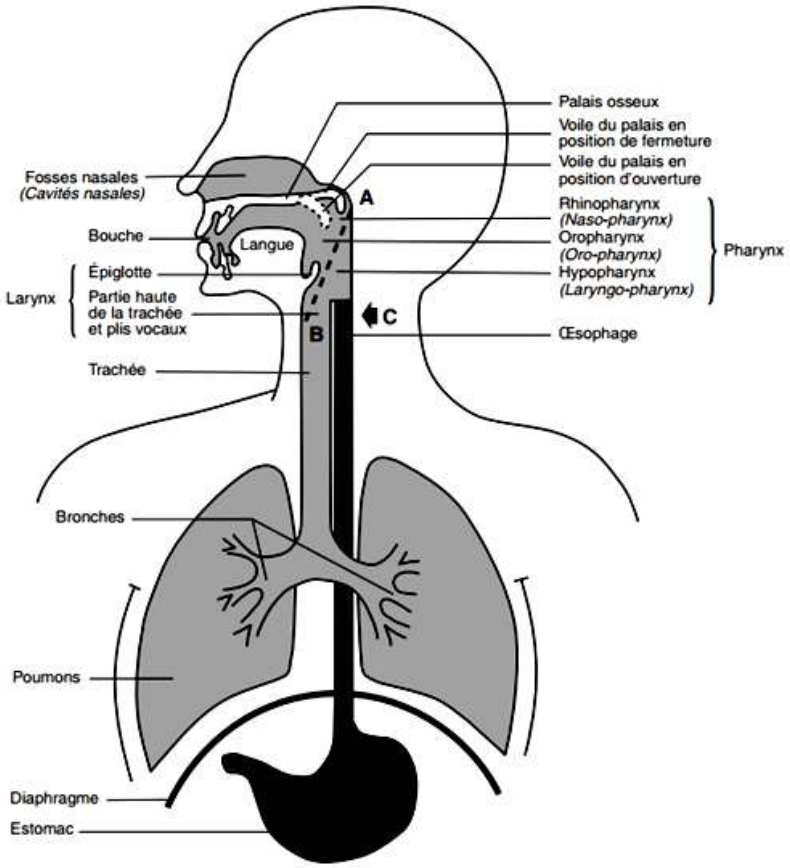


Figure I.1 : Vue d'ensemble des organes de la voix et de la parole [29].

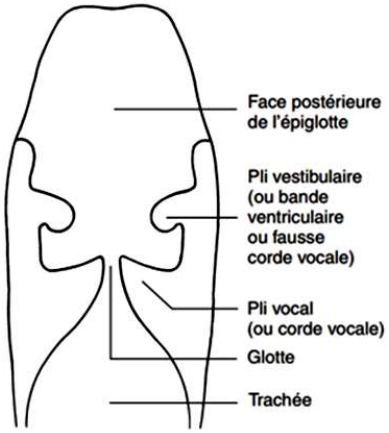


Figure I.2 : Coupe frontale du larynx [29].

Les sons de la parole sont produits par le mouvement de nombreux muscles et organes de phonation dont la coordination est assurée par le système nerveux central.

I.2.1.1. La soufflerie

La soufflerie fournit l'énergie nécessaire à la production sonore [30]. La production de tous les sons s'appuie sur un flux d'air qui provient des poumons et qui est nécessaire à la vibration des cordes vocales.

L'appareil respiratoire est constitué de la cage thoracique (qui contient les poumons, la trachée et les bronches) et des muscles respirateurs : diaphragme, muscles abdominaux, muscles intercostaux et muscles inspirateurs accessoires.

Lors de la respiration, l'abaissement du diaphragme crée une dépression intra-thoracique qui aspire l'air extérieur, remplissant ainsi les poumons [30]. Ceci assure une oxygénation suffisante pour une activité sans effort notable. Le relâchement du diaphragme ainsi que

l'élasticité du tissu pulmonaire permet d'expulser l'air dans un mouvement lui aussi passif en quasi-totalité. Lors de la phonation en revanche, l'augmentation des besoins en oxygène demande la mise en route des muscles inspirateurs accessoires, dont les plus importants sont les muscles intercostaux, qui permettent d'augmenter par l'horizontalisation des côtes le volume de la cage thoracique (et donc celui de l'air inspiré). Cette phase d'inspiration est courte, on parle « d'élan » phonatoire.

L'expiration qui suit est une expiration forcée et active qui met en jeu les muscles expirateurs accessoires, que sont les muscles abdominaux. Dans [29], les auteurs parlent de « souffle phonatoire » et en décrivent plusieurs types.

Pendant la phonation, le sujet exerce un contrôle musculaire très fin pour maintenir une pression d'air la plus constante possible dans la trachée.

I.2.1.2. Le larynx

Le larynx est un organe impair et médian [31], situé à la partie moyenne de la gaine viscérale du cou et qui occupe la partie supérieure de la région sous hyoïdienne. Il est formé par un squelette cartilagineux suspendu à l'os hyoïde. Les différents cartilages sont unis par des articulations et des membranes. L'ensemble est doublé d'un appareil musculaire et recouvert par une muqueuse de type respiratoire.

Le premier rôle du larynx est celui de sphincter : il permet d'obturer la trachée lors de la déglutition grâce à l'abaissement de l'épiglotte, ce qui protège les voies respiratoires. Il intervient également pendant la respiration, ainsi que lors d'un effort musculaire. Enfin, il est l'organe principal de la voix.

I.2.1.3. les résonateurs

Les « résonateurs » (le pavillon pharyngo-buccal et les cavités annexes) [32]

- **Le pharynx**

Le larynx débouche en haut dans le pharynx. Le pharynx, ou « carrefour aérodigestif », n'est rien d'autre que la cavité de l'arrière-bouche (ou arrière-gorge). C'est la cavité qui fait suite à la bouche en arrière de la langue. C'est une cavité musculaire capable de se rétrécir latéralement et d'arrière en avant (action des muscles constricteurs du pharynx). Le volume du pharynx est également susceptible de varier verticalement. Ces variations dépendent des mouvements d'élévation et d'abaissement du larynx, dus à l'action des muscles de l'appareil suspenseur du larynx. Ces mouvements interviennent de façon très importante dans l'articulation des voyelles.

- **La bouche**

Résonateur principal, au même titre que le pharynx, et articulateur principal. Les organes qui composent la bouche (langue, mâchoires, lèvre, etc.) interviennent dans la génération des bruits à partir desquels se sont individualisés les sons du langage.

- **Les fosses nasales et les sinus**

La cavité nasale communique avec le pharynx par le rhinopharynx. Cette communication est interrompue lorsque l'élévation du voile vient empêcher le passage de l'air.

Les sinus sont des cavités annexes des fosses nasales avec lesquelles elles communiquent par de petits orifices appelés ostium. Ces cavités sont remplies d'air.

Les sinus sont au nombre de quatre :

- sinus frontal, au-dessus de l'orbite.
- sinus maxillaire, entre la cavité orbitaire et l'os maxillaire supérieur (c'est-à-dire juste au-dessous de l'œil).
- sinus ethmoïdal, constituant la paroi creuse qui sépare l'œil de la fosse nasale.
- sinus sphénoïdal correspondant au plafond du rhinopharynx.

Si nous citons les sinus, c'est seulement à cause de leur importance en pathologie oto-rhino-laryngologique : sachant que les sinusites sont des affections fréquentes. En effet, les sinus ne jouent apparemment aucun rôle dans la phonation.

I.2.2. Physiologie du système de production de la parole

La production de la parole implique le mouvement coordonné d'un certain nombre de structures dans la tête [33], le cou et le thorax (en d'autres termes, les parties anatomiques décrites ci-dessus). Trois étapes distinctes sont nécessaires pour la production de la parole : la génération d'un courant d'air (expiration); la conversion de ce courant d'air en une série de vibrations (la phonation) et la modification de ces vibrations (l'articulation).

I.2.2.1. La vibration des cordes vocales

La vibration des cordes vocales peut s'expliquer par la théorie aérodynamique-myoélastique [34]. À l'expiration, l'air est expulsé des poumons par 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 vaincre la force musculaire, les cordes vocales sont obligées de s'ouvrir un peu et une petite quantité d'air est libérée dans le canal vocal, entraînant une diminution de la pression sous les cordes.

Comme l'action musculaire dépasse la force produite par la force sous-glottale, les cordes vocales se ferment, aidées par l'effet Bernoulli et par leur recul élastique naturel.
Ce cycle se répète à une certaine fréquence (la fréquence fondamentale) jusqu'à ce que les cordes vocales soient détendues ou lorsqu'il n'y a plus de circulation d'air. La figure I.3 montre un cycle de vibration d'une fermeture glottique à la suivante [35]. Comme le montre cette figure, la structure des cordes vocales n'est pas simple. En effet, différentes parties des plis vibrent de différentes manières au cours du cycle de vibration. Ceci est en partie responsable de la présence d'harmoniques de la fréquence fondamentale de vibration dans le spectre de l'excitation glottique. Différents effets vocaux peuvent être obtenus en agissant sur les propriétés physiques des cordes vocales.

- **Contrôle de la fréquence vocale :** la fréquence fondamentale habituelle varie en fonction de l'âge, en raison des différences anatomiques de longueur, d'épaisseur et de tension des cordes vocales. Des variations autour de cette fréquence peuvent être effectuées en ajustant les propriétés physiques des cordes vocales. Pour cela, la longueur est modifiée en agissant sur la position des cartilages aryténoïdes, cette variation de longueur s'accompagnant d'un changement d'épaisseur et d'une tension croissante. La tension peut également être modifiée indépendamment des autres variables physiques en contractant les muscles thyroaryténoïdes. Sachant cela, une longueur croissante des cordes vocales s'accompagne d'une épaisseur décroissante et d'une tension interne croissante, ces trois effets induisent une fréquence fondamentale plus élevée.

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

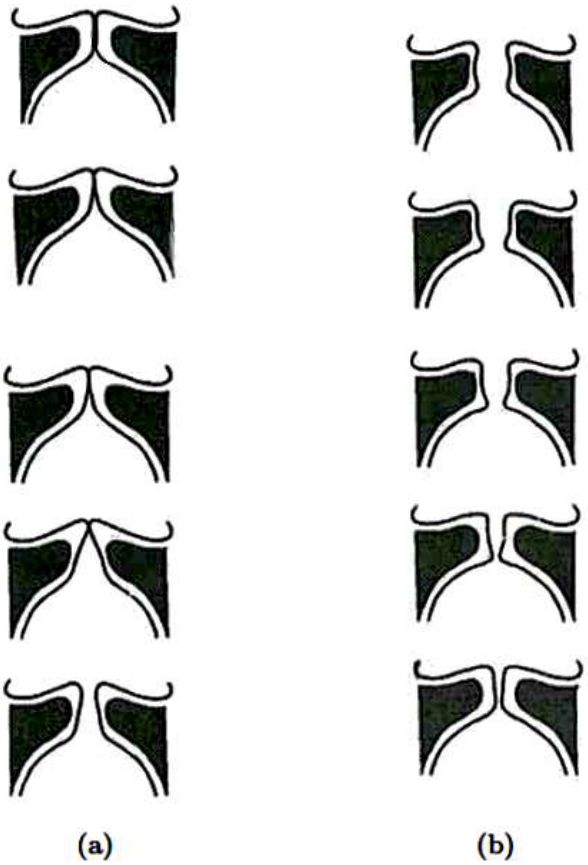


Figure I.3 : Mouvements des cordes vocales au cours d'un cycle et à partir de la fermeture glottique [35].

I.2.2.2. Articulations (résonateurs)

Le spectre du signal source doit être modifié pour produire des sons intelligibles. Ceci est réalisé par l'articulation, autrement dit les mouvements de la langue, des lèvres et de la mâchoire inférieure visant à modifier la taille et la forme des différentes parties du conduit vocal. Suivant la configuration de l'appareil vocal, le timbre de la voix est affecté par l'amplification ou l'amortissement de certaines harmoniques de la fréquence fondamentale des cordes vocales. L'onde de pression qui en résulte est émise vers l'extérieur par les lèvres.

I.3. Modèle de production de la parole

I.3.1. Modèle source-filtre

Basé sur le processus physique de production de la parole, le modèle source-filtre (Figure I.4) a été proposé par Gunnar Fant en 1960 [36] et est devenu largement utilisé comme modèle de production de la parole [33].

Ce modèle est composé de trois parties supposées linéairement indépendantes [37] :

- **Source :** elle représente l'impulsion glottique périodique générée par les cycles périodiques d'ouverture et de fermeture des cordes vocales dans le cas des sons voisés et le bruit dans le cas des sons non voisés.
- **Filtre du conduit vocal :** le conduit vocal est composé des cavités nasale, buccale et pharyngée. Il est modélisé comme un filtre linéaire tout pôle, dont les pôles correspondants aux fréquences de résonance du conduit vocal ont des effets, spectraux appelés formants.
- **Rayonnement labial :** lors du passage par l'ouverture des lèvres, les ondes stationnaires se déplaçant à travers le conduit vocal se propagent sous forme d'ondes sphériques vers l'extérieur. Le rayonnement aux lèvres est souvent modélisé comme l'effet d'une dérivation.

Suivant ce modèle, la parole est produite par la convolution des réponses impulsionnelles de ces trois éléments. Il est important de noter que ce modèle fait l'hypothèse qu'il n'y a pas d'effets de couplage entre le résonateur acoustique (le conduit vocal) et la source vibrante (les cordes vocales). Cela signifie par exemple que l'influence de la cavité supraglottique sur la cavité sous-glottique est négligée. Cela n'est pas toujours valable, particulièrement lorsque la parole est produite avec une période fondamentale élevée [38] ou lorsque le premier formant est très proche de la fréquence fondamentale [37]. Une discussion intéressante sur le couplage entre les différents éléments du modèle source-filtre peut être trouvée dans [39].

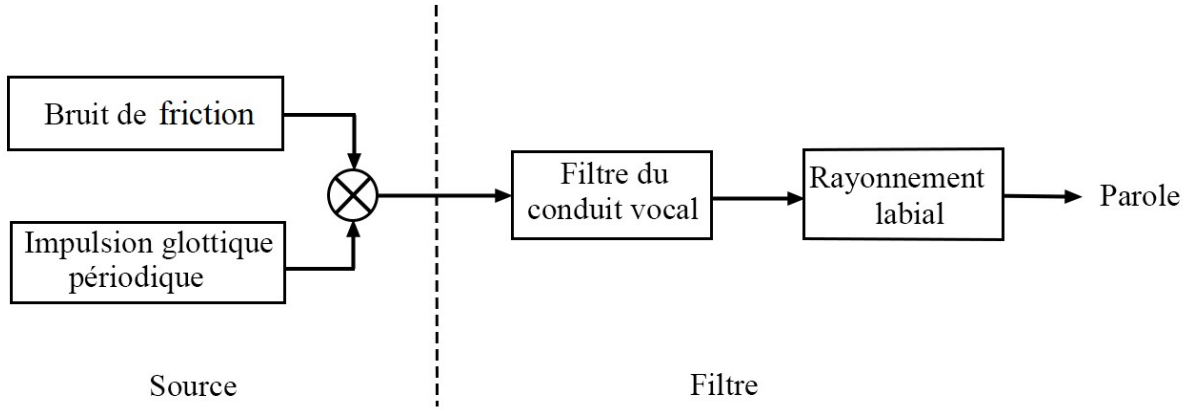


Figure I.4 : Modèle source-filtre de production de la parole [36].

I.3.2. Modélisation de la source glottique

L'objectif de la modélisation de la source glottique est la représentation de la forme d'onde glottique et/ou sa dérivée en utilisant certains paramètres bien définis [40], chacun de ces paramètres a une réalité physiologique ou physique. Dans la littérature, il existe

plusieurs modèles, qui reproduisent fidèlement le comportement de la forme de la source glottique.

Certains modèles visent à modéliser le débit d'air glottique comme le modèle de Lijlencrants-Fant (modèle LF) [41], le modèle de Klatt [42], le modèle de Rosenberg [43] et le modèle R++ [44]. D'autres modèles ont été proposés pour déterminer les paramètres biomécaniques des cordes vocales [45] ou pour la synthèse des troubles de la voix [46].

Généralement, la source glottique (qui modélise l'air glottique) a les caractéristiques temporelles suivantes [41, 47] :

- Le débit glottique est toujours positif ou nul.
- Le débit glottique est quasi-périodique.
- Au cours d'une période fondamentale, le débit glottique est en forme de cloche : il augmente, puis il diminue, jusqu'à ce qu'il devient nul.
- Le débit glottique est une fonction continue du temps.
- Le débit glottique est une fonction différentiable du temps, sauf dans certaines situations à l'instant de fermeture de la glotte GCI (en anglais, Glottal Closure Instant).

L'instant de fermeture de la glotte est l'instant d'excitation maximale due au flux d'air glottique pulsée. A cet instant, si la fermeture des cordes vocales est abrupte, la dérivée de la source glottique présente une discontinuité, tandis qu'elle présente une phase qui s'appelle la phase de retour si la fermeture est lisse.

La dérivée du débit glottique est souvent considérée à la place du débit glottique elle-même [47]. Cette considération est due au faite que lors de la modélisation de la production de la parole, la fonction de transfert de l'élément de rayonnement, liant le flux acoustique au niveau des lèvres au champ acoustique devant des lèvres, peut être considéré comme une dérivée d'une première approximation.

La dérivée de la source glottique pour tous les modèles a des caractéristiques communes qui sont les suivantes :

- La dérivée de la source glottique est quasi-périodique.
- Au cours d'une période fondamentale, la dérivé de la source glottique est positive (lorsque le débit glottique augmente), puis nul (au maximum du débit glottique), puis négative (lorsque le débit glottique décroit), puis nul (lorsque la source glottique est nulle).
- La dérivée du débit glottique est une fonction continue du temps, sauf dans certaines situations à l'instant de fermeture de la glotte.
- La dérivée du débit glottique est une fonction différentiable du temps, sauf dans certaines situations à l'instant de fermeture de la glotte.

Généralement, le modèle du débit glottique est décrit en termes de deux phases qui sont la phase ouverte et la phase fermée comme le montre la figure I.5 [47]. La phase ouverte elle-même est divisée en deux phases : la phase d'ouverture et la phase de fermeture. La phase d'ouverture est définie par l'augmentation du débit glottique de l'état initial à A_v son amplitude maximale à l'instant T_p définissant l'instant d'ouverture.

La phase de fermeture est décrite par la diminution de la source glottique de A_v à un point au moment T_e définissant l'instant de fermeture où la dérivée atteint son minimum négatif E (E: excitation maximale).

Dans 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 excitation maximale. En pratique, il convient de mettre $T_c = T_0$, soit la totalité de la période fondamentale [41].

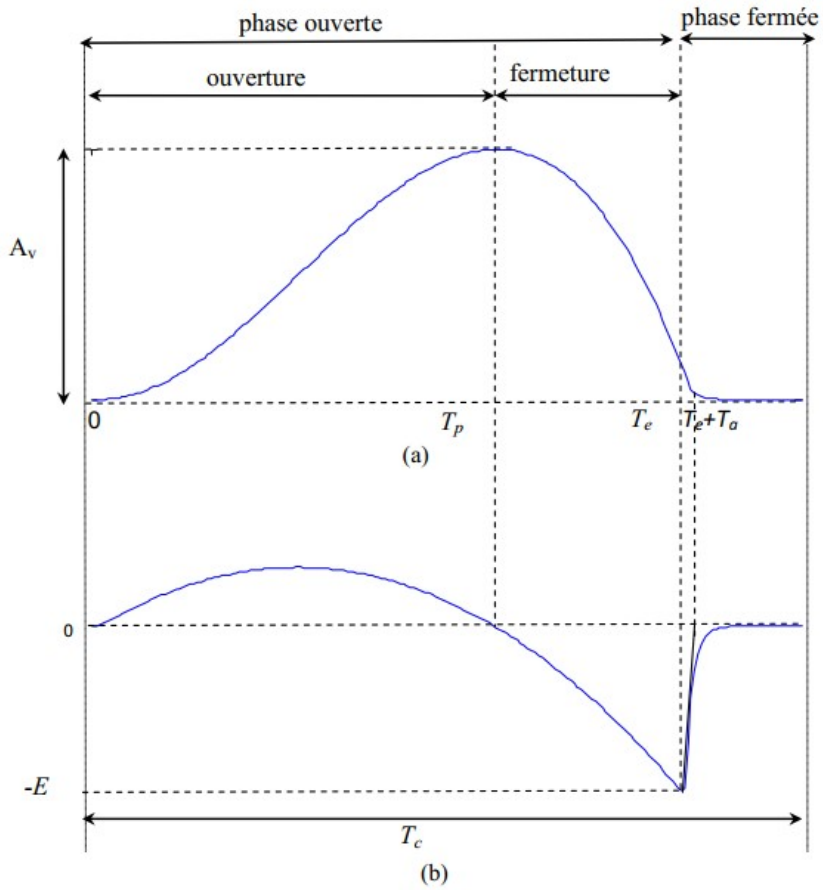


Figure I.5 : Modèle de débit glottique (a) et sa dérivé (b).

Dans la littérature, de nombreux modèles de la source glottique sont proposés. Le modèle proposé par Liljencrants-Fant connu comme le modèle LF [41] est l'un des modèles les plus couramment utilisés pour la synthèse et la modélisation de la parole [48, 49].

I.4. Pathologies du larynx

La dysphonie est définie comme la perturbation de la parole produite alors que le larynx est impliqué dans la phonation [50]. Il existe différents types de dysphonie [33]. Les dysphonies organiques sont causées par des modifications pathologiques des cordes vocales. Les dysphonies dysfonctionnelles sont caractérisées par des difficultés de phonation sans altération organique évidente des cordes vocales. Du fait de l'application de mécanismes de compensation par le patient, une dysphonie dysfonctionnelle peut induire une dysphonie organique. Enfin, les dysphonies neurologiques sont causées par des dommages neurologiques. Ces différents types de pathologies sont décrits dans [50-52].

I.4.1. Dysphonie organique

Ce type de trouble de la voix est essentiellement causé par des changements morphologiques de l'anatomie du larynx, en particulier au niveau de la glotte [52]. Une dysphonie organique très courante est la laryngite. Elle est causée par des agents infectieux (virus / bactéries) ou des séances de parole prolongées et consiste en une inflammation du larynx. Elle provoque une vibration asymétrique des cordes vocales et une augmentation de leur masse et de leur raideur. Les symptômes typiques sont l'enrouement et la douleur.

En se concentrant sur les cordes vocales, une autre pathologie courante est les polypes vocaux et les nodules, résultant souvent de l'abus ou de la sur-utilisation de la voix [50]. Les polypes et les nodules apparaissent à la partie des cordes vocales la plus sollicitée, à savoir le lieu de fermeture glottique.

Apparaissant sur l'une des cordes vocales, les polypes prennent la forme de masses charnues rouges ou blanchâtres, fixées aux cordes vocales ou liées à celles-ci par une tige. Contrairement aux polypes, les nodules peuvent apparaître symétriquement sur les deux plis sous forme de grumeaux fermes et blanchâtres. Les polypes et les nodules ont le même effet, en ce sens que la fermeture ne peut pas être complète en raison de la présence des grumeaux. Il en résulte une fuite d'air lors de la phonation (conduisant à la présence de bruit dans la parole produite) et une modification de la masse et de l'élasticité des cordes vocales (conduisant à l'enrouement).

La vibration des cordes vocales peut également être altérée par la présence d'un œdème. Par exemple, l'œdème de Reinke est situé à un ou deux cordes vocales et consiste en une accumulation de fluide dans la couche superficielle de la lamina propria (Membrane qui tapisse les cavités naturelles de l'organisme) du ou des plis. Bien qu'il n'empêche pas la fermeture glottique, l'œdème influence la vibration des plis en augmentant la masse du pli

vocal sur lequel il se trouve, induisant une diminution de la fréquence fondamentale et un enrouement de la voix produite. Un œdème plus important, l'œdème laryngé, peut également survenir en raison de réactions allergiques ou de médicaments. S'il n'est pas traité, il peut entraîner une obstruction des voies respiratoires, menaçant la vie de la personne. Ce cas est traité en urgence.

Dans les cas les plus sévères, un cancer peut survenir au niveau des cordes vocales, sous la forme d'un carcinome épidermoïde. Se manifestant par de gros morceaux blancs ou de grandes ulcérations, le cancer peut être enlevé chirurgicalement et la voix est préservée s'il est situé dans le tissu épithélial. Si le tissu sous-jacent est déjà envahi, il peut être nécessaire de réaliser une laryngectomie partielle ou totale (ablation permanente de la partie malade du larynx), souvent accompagnée d'une trachéotomie.

I.4.2. Dysphonie dysfonctionnelle

La dysphonie dysfonctionnelle se caractérise par une absence de lésion anatomique et peut être induite par des troubles respiratoires, neurologiques voire psychologiques. Elle est définie comme un trouble temporaire ou permanent de la fonction vocale, perçu comme tel par la personne ou son entourage [53].

Ce type de dysphonie est caractérisé par une hypotonie et une hypertonie des muscles laryngés et respiratoires [54]:

- **Hypotonie (hypofonction, hypokinésie) :** il s'agit d'une diminution pathologique du tonus musculaire. Dans ce cas, les cordes vocales ne sont pas suffisamment tendues provoquant donc une diminution à la fois de l'intensité vocale et de la fréquence fondamentale.

- **Hypertonie (hyperfonction, hyperkinésie) :** il s'agit d'une surestimation permanente du tonus musculaire. Ce trouble est la conséquence d'une sur-utilisation de la voix. Il induit une perturbation de la vibration des cordes vocales, dont le corollaire est une augmentation de la période fondamentale et une perte de ses harmoniques. D'autres effets sont les discontinuités dans les émissions vocales ou des hésitations au début de phonation.

Une autre dysphonie dysfonctionnelle est la dysphonie spasmodique. Elle se caractérise par des spasmes au niveau du larynx, dus à des troubles de la transmission d'informations entre les nerfs et les muscles. Les symptômes les plus courants sont de brèves coupures de la voix pendant la phonation, ainsi qu'une diminution de l'intensité et une dégradation désagréable du timbre. La voix présente souvent des tremblements.

I.4.3. Dysphonie neurologique

En cas de dysphonie neurologique périphérique, une ou les deux cordes vocales peuvent être paralysées [55]. En effet, tous les muscles laryngés intrinsèques (à l'exception du muscle cricothyroïdien) sont innervés par le nerf laryngé récurrent. La paralysie d'une ou de deux cordes vocales peut survenir en cas de paralysie de ce nerf. Deux types de paralysie sont possibles : la paralysie temporaire (certaines des unités motrices fonctionnent et la paralysie disparaît lorsque sa cause est résolue) et la paralysie permanente (toutes les unités motrices sont désactivées et la paralysie se termine lorsque les nerfs sont régénérés). Dans les deux cas, la paralysie induit une rigidité décroissante des muscles laryngés. Différents types de paralysie peuvent également être distingués suivant la position des cordes vocales.

Lorsque le système nerveux central est endommagé, une dysarthrie peut apparaître et consiste en des troubles de l'articulation et/ou de la phonation. Le rythme de la voix est modifié et des coupures de voix se produisent souvent. La dysarthrie peut être causée par des maladies neurologiques telles que la maladie de Parkinson ou la sclérose en plaques.

I.5. Examen clinique des patients atteint de troubles de la voix

L'examen clinique d'une personne atteinte de troubles de la voix ne se limite pas uniquement à l'examen endoscopique du larynx. La connaissance du comportement phonatoire du sujet dysphonique (par exemple, en situation de forçage vocal), l'importance de la voix dans son activité professionnelle et sociale, sa propre perception de la qualité de sa voix, la manière dont il vit sa phonation (sa souffrance, sa gêne, etc.) ou dont son entourage le perçoit, constituent un ensemble primordial et complémentaire de l'examen médical pour déterminer correctement le caractère pathologique ou non pathologique d'une voix [56].

En matière de pathologie vocale [53], les signes les plus importants à considérer ne relèvent pas directement de l'acoustique de la voix. C'est la gêne du sujet ou de son entourage qui est le fait primordial : sa gêne ou sa souffrance. C'est elle qu'il conviendra d'évaluer avec le plus de soin.

Dans cette section, nous présentons trois méthodes permettant d'établir un bilan clinique d'un patient dysphonique [50] : l'interrogatoire, l'examen physique et l'analyse du comportement vocal.

I.5.1. Interrogatoire avec le patient

L'examen clinique d'un sujet dysphonique commence par un interrogatoire durant lequel le patient expose l'historique de son anomalie actuelle, comment et depuis combien de temps les troubles de la voix sont apparus, les différents traitements entrepris ou explorations déjà faites (en terme de rééducation, de chirurgie, de cure thermale, etc.), les résultats observés, etc. L'ensemble de ces éléments vont permettre de constituer l'anamnèse c'est-à-dire « l'histoire de la maladie », d'établir un diagnostic complet permettant d'interpréter différemment les résultats d'examens complémentaires.

Il est important aussi que le patient dysphonique parle de son ressenti à propos de sa propre phonation, ce qu'il pense de sa voix (d'un point de vue esthétique, agréabilité, etc.) et de ses possibilités vocales (en terme de puissance, de fatigue, en milieu bruité ou téléphonique, etc.), sur ses capacités de contrôle, à chanter, etc. Il en est de même sur les sensations qu'il peut éprouver dans la région laryngée (comme picotement, irritation, douleur, etc.) ou respiratoire. Ainsi, même si ces éléments constituent des « signes subjectifs », ils pourront permettre au patient de prendre conscience de sa guérison dès leur disparition.

D'autres éléments pourront être recueillis par le clinicien afin de mieux comprendre la dysphonie du patient tels que sa santé générale, son genre de vie, son tempérament, l'importance de la voix dans sa profession, sa consommation d'alcool et de tabac, son exposition à la pollution aérienne et sonore, au surmenage, etc.

I.5.2. Examen physique du patient

A l'interrogatoire du patient, l'examen du larynx reste indispensable et nécessaire au médecin pour établir un diagnostic complet sur la pathologie. De façon générale, il comporte systématiquement la laryngoscopie et la laryngostroboscopie auxquelles est associé un examen ORL [57], permettant ainsi d'orienter le diagnostic étiologique. L'examen laryngoscopique dit « indirect » est réalisé à l'aide d'un miroir tel que celui présenté sur la figure I.6. Il permet d'une part, de vérifier la présence ou non d'une anomalie organique laryngée, d'inspecter l'aspect de la muqueuse laryngée et de contrôler l'existence d'un processus tumoral. Il permet d'autre part, d'évaluer, durant la phonation, le dysfonctionnement de la cinétique laryngée et d'étudier la mobilité des cordes vocales et des aryténoïdes durant les passages alternés entre le mode de phonation (glotte fermée) et respiratoire (glotte ouverte).

Figure I.6 : Miroir utilisé en laryngoscopie « indirecte » [50].

Figure I.7 : Laryngoscopie « directe » [50].

Depuis l'apparition de la fibre optique dans les années 1970, la fibroscopie par voie nasale (ou naso-fibroscopie) peut être utilisée en remplacement de l'examen laryngé pratiqué classiquement au miroir. Elle permet l'observation du larynx durant la phonation et le chant. La laryngoscopie au tube rigide (ou épipharyngoscopie) consiste à introduire l'endoscope dans la bouche afin de visualiser l'ensemble du pharyngolarynx avec une très grande précision. Offrant une meilleure définition de l'image laryngée, elle permet surtout l'examen en lumière stroboscopique. La laryngostroboscopie consiste à examiner les cordes vocales en phonation en utilisant la lumière stroboscopique. L'examen stroboscopique permet d'obtenir une image au ralenti de la vibration des cordes vocales. Devenu indispensable à l'étude de la fonction laryngée, il permet d'observer la qualité des vibrations glottiques, de révéler la présence éventuelle d'une lésion intra-cordale et d'explorer l'ondulation de la muqueuse des plis vocaux. En cas de dysphonie durable sans pathologie organique apparente, cet examen est obligatoire.

Connectée au fibroscope ou à un endoscope droit, la caméra à haute vitesse permet au médecin d'observer le phénomène extrêmement rapide de vibrations des plis vocaux, de visualiser la vidéo après examen du patient afin de compléter le dossier médical avec des photographies extraites de la vidéo. Il est à noter que les possibilités d'imagerie du larynx se sont considérablement améliorées grâce à l'apport du numérique.

L'examen clinique du larynx peut éventuellement être complété par un examen audiométrique. Cependant, en cas d'atteinte organique, le seul moyen d'obtenir un diagnostic de certitude reste la laryngoscopie dite « directe » (figure I.7) qui, réalisée au bloc opératoire sous neuro-analgésie, permet d'effectuer des prélèvements biopsiques.

Il est à noter que l'examen ORL ne serait être complet sans la palpation de la glande thyroïde et des aires ganglionnaires.

I.5.3. Examen du comportement vocal

L'appréciation de la voix durant l'examen interrogatoire du patient ne peut se limiter à la voix dite « conversationnelle ». L'observation du comportement vocal du patient dysphonique en diverses situations peut s'avérer d'un apport considérable pour approfondir la connaissance de la pathologie [58].

En effet, la dynamique de la voix dite « projetée », par exemple, se caractérise par un changement d'attitude et de posture (comme la verticalisation), par une modification du débit (ralentissement), de l'intensité et du mode respiratoire. Par rapport à la voix « conversationnelle », la dépense énergétique y est beaucoup plus importante. Ainsi en voix « conversationnelle », il consistera à relever les caractères acoustiques vocaux tels que l'intensité, la hauteur et le timbre. Il est aussi intéressant de noter la variabilité de la qualité vocale apparaissant dans le discours du patient (l'émotivité, les contextes favorisant les troubles pathologiques, etc.), ainsi que la présence ou non d'un comportement de forçage vocal, de troubles d'articulation et de débit de parole.

Cet examen étudie le comportement vocal dans les situations suivantes :

- Epreuve de « comptage projeté » pour mettre en évidence le forçage vocal.
- Voix « d'appel » pour faire découvrir au patient ses possibilités vocales.
- Voix « chantée » pour étudier la tessiture, la justesse, les altérations du timbre.
- Epreuve de « temps maximal de phonation » pour objectiver l'efficacité du souffle phonatoire.

De plus, faire écouter au patient sa propre voix enregistrée durant les différentes épreuves, peut lui permettre de mieux comprendre le mécanisme de ses difficultés et des troubles qu'il présente.

I.6. Evaluation des voix pathologiques

Le bilan vocal de personnes dysphoniques ne saurait être exhaustif si l'examen clinique n'était pas complété par une évaluation de la qualité vocale. Cette évaluation reste encore aujourd'hui un sujet sensible au centre de nombreuses études dans des domaines multidisciplinaires. Deux principales approches peuvent être considérées [59] : l'analyse perceptive et l'analyse objective.

L'analyse perceptive concerne le jugement « à l'oreille » de la voix et de la parole qui est l'une des méthodes les plus utilisées en milieu clinique. Elle permet de déterminer si une voix est pathologique ou non et de mesurer le degré du dysfonctionnement vocal. Or, de par sa subjectivité intrinsèque et sa grande variabilité (intra-auditeur et inter-auditeurs), le jugement perceptif reste une méthode controversée. A travers cette section, l'évaluation perceptive est présentée, de la définition d'une terminologie adéquate à l'élaboration d'échelles d'évaluation. Ces deux étapes ont été nécessaires pour la mise en œuvre de protocoles standardisés d'analyse perceptive de la voix permettant d'atteindre une fiabilité raisonnable.

Afin de pallier les « faiblesses » du jugement auditif, l'analyse objective est proposée comme complémentaire à l'analyse perceptive. Ces techniques dites instrumentales se subdivisent en deux catégories. La première catégorie repose sur l'acquisition de mesures physiques par le biais de capteurs et applique des techniques statistiques (analyse discriminante, corrélation, régression, etc.) sur les mesures extraites pour établir des résultats.

La deuxième catégorie repose sur les techniques utilisées en traitement automatique de la parole (TAP). Cette dernière a la particularité de s'appuyer uniquement sur une analyse acoustique du signal de parole permettant une facilité et simplicité d'utilisation. Ces deux techniques instrumentales, nommées respectivement « analytique » et « automatique »

I.6.1. Analyse perceptive

De par son essence perceptive, la voix est indissociable de l'oreille. Cela légitime le fait que la perception auditive soit le premier outil pour l'évaluation du trouble vocal [60], de résultats post-opératoires ou rééducatifs. Dans la pratique quotidienne, les cliniciens utilisent « l'écoute à l'oreille » pour décrire la qualité de la voix à l'aide d'une terminologie établie sur une impression auditive. Même si des règles ont été instaurées (notamment sur l'âge, le sexe ou la nature de la voix), l'évaluation perceptive n'est pas sujette à consensus en ce qui concerne la définition des termes perceptifs décrivant la qualité de la voix.

Selon les auditeurs, l'éraillement peut aussi bien désigner une voix rauque que pathologique. Il en est de même pour apprécier les nuances, parfois subtiles, entre deux termes très proches comme entre une voix éraillée et une voix rocailleuse.

I.6.1.1. Les échelles d'évaluation perceptive

Plusieurs méthodes d'évaluation perceptive ont été proposées dans la littérature [61, 62]. Ces dernières se proposent d'évaluer la qualité de la voix suivant des critères qualitatifs et quantitatifs [63, 64].

L'échelle GRBAS de Hirano [65, 66] est une échelle compacte et simple d'utilisation, composée de 5 paramètres décrits dans le tableau I.1.

Tableau I.1: Définition des paramètres de l'échelle GRBAS.

Initiale	Terme	Signification	Définition
G	Grade	Grade global de dysphonie	Impression globale du degré d'anormalité de la voix
R	Roughness	Raucité	Impression d'irrégularité des vibrations des cordes vocales qui correspond aux fluctuations irrégulières de F_o et/ou à l'amplitude du son glottique
B	Breathiness	Caractère soufflé	Impression d'une fuite d'air assez importante à travers les cordes vocales, relative à des turbulences
A	Asthenia	Asthénie	Manque de puissance de la voix relatif à une intensité faible du son et/ou un manque des harmoniques élevées
S	Strain	Forçage	Impression d'un état hyperfonctionnel de phonation relatif à une F_o anormalement haute (bruit dans les hautes fréquences et/ou richesse en harmoniques dans les hautes fréquences)

Actuellement, l'échelle GRBAS est la plus utilisée en pratique clinique quotidienne [64]. Elle permet d'évaluer la fonction laryngée d'un patient sur de la lecture ou de la parole spontanée, suivant 5 paramètres qualitatifs à quantifier selon 4 niveaux de sévérité de la dysphonie :

0 =voix normale	1 =altération légère	2 =altération moyenne	3 =altération sévère

Cette méthode d'évaluation ne prend pas en compte la fonction supralaryngée, la hauteur tonale et la sonie du patient. Il faut noter que les paramètres de grade global (G), de raucité (R) et de souffle (B) sont considérés comme étant les plus fiables et moins soumis à la variabilité d'un jury d'écoute que les paramètres d'asthénie (A) et de forçage (S) [67].

I.6.1.2. Le jury d'écoute

Pour un usage clinique quotidien, une échelle d'évaluation doit être simple, reproductible et pratique. Néanmoins, son utilisation n'est pas sans comporter une part d'imprécision. En effet, chaque spécialiste définit à l'usage (ou inconsciemment) ses propres critères subjectifs. La constitution de jury d'écoute tente de réduire cette imprécision avec une prise de décision par consensus lors de séances d'évaluation vocale. Constitué de plusieurs auditeurs, un jury d'écoute évalue à l'oreille et caractérise le degré de dysphonie d'une voix. Son but est d'obtenir un jugement fiable et reproductible sur la qualité vocale de chaque échantillon présenté.

En ce sens, la constitution du jury d'écoute doit prendre en compte plusieurs éléments tels que l'expérience des auditeurs en termes d'évaluation de la voix pathologique, leurs stratégies d'écoute ou la cohérence du groupe. En effet, un auditeur expérimenté peut ressentir des sensations kinesthésiques (ou physiques) à l'écoute d'une voix dysphonique [68], ce qui n'est pas forcément le cas pour un auditeur naïf. De même, sa capacité d'analyser une voix en discernant les mécanismes de production et le fonctionnement vocal est un élément important afin d'évaluer l'efficacité dans la communication. Enfin, l'origine socio-culturelle, la profession (phoniatre, médecin ORL, orthophoniste, etc.), le secteur d'activité (hôpital, cabinet médical, libéral, etc.), constituent des informations pertinentes pour obtenir une structure de groupe homogène et harmonieuse. Généralement, un jury est considéré comme fiable si ses évaluations sont reproductibles.

Plusieurs travaux ont porté sur la problématique du recrutement des auditeurs de jury d'écoute [69] sans vraiment confirmer le nombre optimal nécessaire pour constituer ce jury d'écoute. Des études contradictoires ont été publiées comme par exemple [70], pour ces auteurs, l'expérience des auditeurs peut influencer la fiabilité du jury d'écoute, ce qui est infirmé dans [66] en situation de «test-retest ».

D'une manière générale, il est admis que l'analyse perceptive doit être menée par plusieurs auditeurs experts et durant plusieurs sessions d'écoute pour qu'elle puisse atteindre une fiabilité raisonnable. La mise en place d'un tel protocole peut s'avérer finalement contraignante et difficilement applicable. En effet, les inconvénients majeurs de cette approche restent le coût humain non négligeable (réunions périodiques de plusieurs experts, durée des séances d'écoute, disponibilité simultanée des experts, etc.) et le manque de fiabilité dû à différents facteurs liés aux caractéristiques intrinsèques du jugement perceptif [71] :

- **Subjectivité** : le caractère « agréable/désagréable» est une sensation subjective par laquelle aucun objet n'est représenté. En particulier, le caractère esthétique ou dysharmonieux de la voix pathologique dépend du goût de chacun.

- **Variabilité intra-individuelle** : un même individu peut ressentir ou interpréter les objets différemment en fonction de son état psychologique (de son humeur, de son état d'esprit, de son stress, de sa disponibilité) ou à des moments différents. Cette variabilité se retrouve lors de l'analyse perceptive de dysphonie vocale, lorsqu'un auditeur du jury n'attribue pas la même note à un même stimulus d'une séance à l'autre.

- **Variabilité inter-individuelle** : elle est due essentiellement à la connaissance et/ou méconnaissance du phénomène dysphonique, à l'expérience des auditeurs. Cette variabilité est d'autant plus importante si le jury d'écoute est de culture et d'écoles cliniques différentes.

I.6.2. Analyse objective et méthodes instrumentales

L'analyse objective est la deuxième méthodologie proposée comme une alternative à l'évaluation perceptive pour pallier ses inconvénients et faiblesses précédemment décrits. Depuis plusieurs années, une multitude de travaux se sont consacrés à la détection et la classification des voix pathologiques au moyen d'analyses acoustiques et paramétriques, de méthodes statistiques et de techniques de modélisation issues des domaines de la reconnaissance des formes et du traitement automatique de la parole [72]. Les méthodes objectives instrumentales peuvent être divisées en deux catégories d'approches présentées sous les appellations « analytique » et « automatique » (figure I.8).

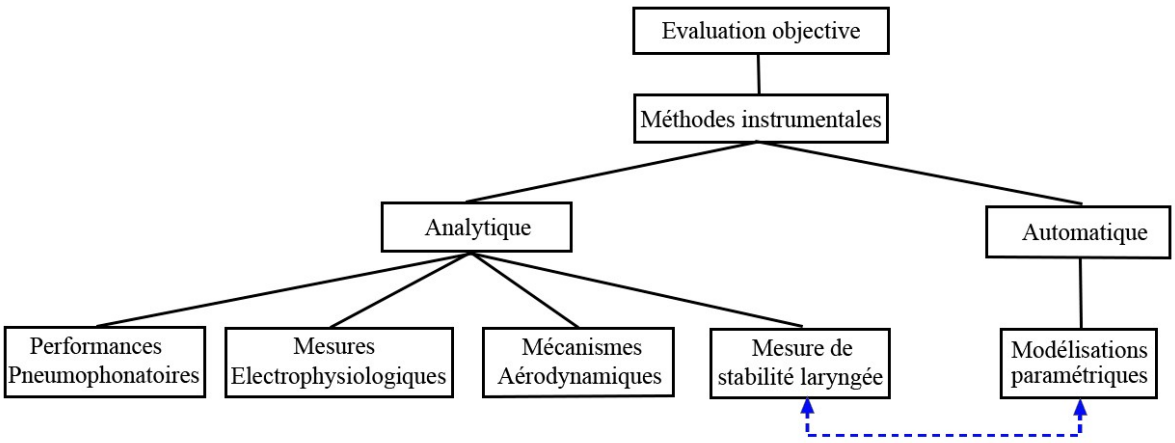


Figure I.8 : Les méthodes instrumentales pour l'évaluation objective de la voix dysphonique.

I.6.2.1. Méthodes instrumentales analytiques

De nombreux travaux [73] ont porté sur l'élaboration d'outils d'évaluation objective standardisés et applicables aux voix dysphoniques. Ces derniers reposent sur l'acquisition de mesures acoustiques, aérodynamiques et/ou électro-physiologiques par le biais de capteurs. Pour la plupart des auteurs, l'élaboration de méthodes objectives multiparamétriques se justifie par le caractère multidimensionnel de la production vocale [74]. Leur but est d'évaluer qualitativement et quantitativement les troubles vocaux à partir de mesures multiparamétriques. Dans ce contexte, les mesures sont associées au système qui applique sur ces dernières une technique statistique (analyse dyscriminante, étude de corrélation, régression linéaire, etc.) pour fournir une décision. Ces méthodes sont appelées « analytiques» dans le sens où les mesures estimées (acoustique, aérodynamique ou physiologique) apportent un ensemble restreint d'indices susceptibles d'être caractéristiques de dysfonctionnements vocaux. La dimension « multiparamétrique » de la méthode permet la constitution d'un ensemble d'indices facilitant ainsi la prise de décision. On peut distinguer différentes classes d'indices.

I.6.2.2. Méthodes instrumentales automatiques

Parallèlement, une deuxième approche d'évaluation objective a été proposée dans la littérature, les méthodes instrumentales dites « automatiques ». L'approche est basée sur une analyse automatique de la parole pour l'évaluation de la dysphonie. En ce sens, l'intérêt s'est plus particulièrement porté sur les techniques utilisées en traitement automatique de la parole (TAP) et à leur adaptation à l'évaluation et la classification des voix pathologiques.

La première étape des méthodes instrumentales automatique est la paramétrisation qui consiste à transformer le signal acoustique en une séquence de vecteurs de paramètres afin d'en obtenir une représentation simplifiée nécessaire avant les phases d'apprentissage et de test. Cette phase permet de réduire la redondance du signal de parole et d'en extraire les informations pertinentes. Différentes représentations paramétriques sont proposées. Les plus couramment utilisées sont issues de l'analyse en banc de filtres tels que les coefficients LFSC et MFSC (linear/Mel frequency spectral coefficients) et les coefficients cepstraux tels que les coefficients LFCC et MFCC (linear/Mel frequency cepstral coefficients).

Lorsque la dysphonie apparaît exclusivement dans la vibration glottique, des mesures acoustiques (caractéristiques de la stabilité laryngée comme F_0, jitter, shimmer, HNR)

peuvent être estimées sur des trames successives du signal de parole (à court ou moyen terme) afin d'adopter une représentation de séquences de vecteurs, nécessaire pour les phases de modélisation et de test [75].

En pratique, comparées aux méthodes instrumentales analytiques, les méthodes instrumentales automatiques ont l'avantage de la stabilité et la performance qui reposent sur :

1. La capacité à analyser la parole continue, bien plus significative que l'analyse des voyelles soutenues.
2. La capacité à traiter de grands corpus, permettant de mener des études à grande échelle et d'obtenir des informations statistiques significatives.
3. Une analyse acoustique, simple et automatique, permettant une utilisation clinique facile à caractère non invasif et à faible coût humain.

1.7. Classification automatique des voix pathologiques

La majorité des études dédiées à la classification des voix pathologique ont considéré le problème de classification binaire normale/pathologique en utilisant des voyelles soutenues. Le travail présenté dans [13] se focalise sur la classification des voix pathologiques en utilisant 30 indices acoustiques obtenus à partir de la voyelle soutenue [a]. La méthode comprend deux étapes. La première étape est la transformation de l'espace des indices et la réduction de la dimension des données en utilisant l'analyse en composantes principales. La deuxième étape est la classification à l'aide des machines à vecteurs de support (SVM). Les résultats obtenus montrent que les cas pathologiques peuvent être détectés avec une précision de classification atteignant 98.1 %.

Dans [14], les auteurs ont utilisé différentes bandes de fréquences pour l'extraction des indices acoustiques à partir de la voyelle soutenue [a] pour la classification binaire normale/pathologique des voix de trois bases de donnée. La précision de classification obtenue était entre 96.02 % et 99.54 % selon la base de données.

Bien que la précision de ces classificateurs soit élevée, ils utilisent beaucoup d'indices acoustiques ce qui limite leur capacité de généralisation. Les méthodes de classification proposées dans la littérature ont pour objectif d'atteindre des grandes valeurs de précision tout en utilisant un nombre limité d'indices acoustiques ayant une grande capacité de discrimination [15-19].

Le nombre d'études portant sur la classification des voix pathologiques en utilisant la parole continue reste faible par rapport au nombre de travaux utilisant les voyelles

soutenues. Dans [20], une représentation temps-fréquence a été proposée pour la classification binaire normale/pathologique des voix en utilisant la parole continue. Les expériences effectuées sur une base de données constituée de stimuli de parole continue produite par 51 locuteurs normophoniques et 161 locuteurs dysphoniques ont donné une précision de classification de 93.4 %.

Dans [21], les auteurs ont exploré les informations fournies par une représentation conjointe du signal acoustique et de la fréquence de modulation, appelée spectre de modulation, pour la détection et la discrimination des troubles de la voix à l'aide d'un classificateur SVM, l'approche proposée a permis d'atteindre une précision de classification de 95 %.

Dans [23], il a été proposé de décomposer le signal acoustiqueen en plusieurs bandes d'ondelettes pour l'extraction d'indices acoustiques et utiliser un algorithme génétique pour déterminer les paramètres de l'ondelette optimale. Une machine à vecteurs de support est utilisée comme classificateur des cas normaux et pathologiques à l'aide d'un système d'ondelettes à cinq bandes. La méthode a permis d'atteindre une précision de classification de 98.59 %.

Dans [24], un vecteur de caractéristiques basé sur la combinaison de la décomposition en paquets d'ondelettes (WPD : wavelet packet decomposition) et des coefficients cepstraux de fréquence Mel (MFCCs) a été proposé pour former un modèle hybride de réseau de neurones artificiels (ANN : artificial neural networks) et d'algorithme génétique pour la classification binaire normale/pathologique des voix, la précision de classification maximale était de 95.3%.

Le travail présenté dans [25] s'est concentré sur l'extraction de différentes bandes de fréquences en utilisant l'autocorrélation et l'entropie pour la détection et la classification des pathologies de la voix. La méthode a été testée sur plusieurs bases de données telles que « Arabic Voice Pathology Database » (AVPD), « Saarbrücken Voice Database » (SVD) et « Massachusetts Eye and Ear Infirmary » (MEEI), contenant des voyelles soutenues et des phrases produites par des locuteurs normophoniques et dysphoniques. La précision la plus élevée obtenue en utilisant un classificateur SVM était de 99.79 %.

I.8. Conclusion

L'évaluation perceptive de la qualité de la voix reste une composante essentielle pour la majorité des protocoles d'évaluation de la voix pathologique [76]. Malgré les inconvénients de l'évaluation perceptive comme la variabilité intra et inter juges, cette

méthode est toujours considérée à l'heure actuelle comme le standard de référence pour l'évaluation de la qualité vocale, face à laquelle les mesures instrumentales sont confrontées et indispensable pour suivre l'évolution d'un traitement médical, rééducatif ou post-opératoire.

Afin de pallier les faiblesses et inconvénients de l'évaluation perceptive, les méthodes instrumentales dites « analytiques » ont été proposées dans la littérature. Ces dernières sont basées sur l'acquisition de mesures acoustiques, aérodynamiques et/ou électrophysiologiques par le biais de capteurs. Les mesures estimées constituent un ensemble d'indices susceptibles de caractériser les dysfonctionnements vocaux directement liés à la dysphonie comme les fluctuations laryngées, la fuite glottique ou l'activité musculaire laryngée. Cependant, malgré des résultats très acceptables obtenus par cette approche « analytique », ceux-ci restent encore insuffisants pour les praticiens pour qu'elle puisse être considérée comme un outil d'évaluation clinique de la voix.

En outre les méthodes instrumentales dites automatiques ont plusieurs avantages. En effet, elles offrent la possibilité d'analyse de la parole continue et permettent une analyse acoustique facile et automatique ainsi qu'un usage journalier en milieu clinique.

Chapitre 2 :

Etat de l’art de l’intelligence artificielle et des méthodes de classification

II.1. Introduction

L'intelligence artificielle (IA) est une branche de l'informatique où on tente de comprendre le fonctionnement des entités intelligentes et on programme les ordinateurs pour imiter l'intelligence humaine afin d'améliorer et assister le quotidien de l'humanité ou optimiser le travail présentant une difficulté pour l'être humain. Cela réduit considérablement les efforts des humains et leur facilite beaucoup l'exécution de ces tâches.

L'intelligence artificielle est une technologie pour l'avenir de l'humanité qui vise à améliorer la vie des humains. Cependant, généraliser l'intelligence artificielle dans notre vie n'est pas possible en ce moment car il existe de nombreuses caractéristiques du cerveau humain qu'il n'est pas possible de décrire. Ce chapitre est dédié à l'état de l'art de l'intelligence artificielle et aux méthodes de classification automatique. On donne un aperçu sur l'historique de l'IA et on présente les différents types d'apprentissage et les méthodes de classification automatique.

II.2. Historique et applications de l'intelligence artificielle

On constate qu'il est possible d'identifier trois grandes périodes dans l'histoire de l'I.A [77-80] :

1. Les débuts des années quarante jusqu'à la fin des années cinquante, l'I.A naît en tant que science entourée d'espérances grandioses et de promesses impressionnantes.
2. Ensuite, le manque de résultats convaincants, finit par réorienter la recherche vers un discours plus modeste et moins éblouissant, cherchant des résultats concrets et des expérimentations réelles loin du sur optimisme.
3. Enfin, à partir des années quatre-vingt, et à la suite de ses propres progrès et aussi grâce aux nouvelles découvertes provenant de ses frontières interdisciplinaires, l'I.A reprend ses ambitions et ses espoirs.

En effet, dès les premiers travaux sur l'I.A, les chercheurs se sont inspirés de l'intelligence humaine. A partir des années quarante, plusieurs chercheurs ont travaillé sur la modélisation du cerveau humain et en particulier sur les neurones artificiels [81]. Ensuite, en 1950 une vision de l'intelligence artificielle a été énoncée par Alan Turing dans son article « les ordinateurs et l'intelligence » [82]. Après, c'est en 1955 dans les travaux de McCarthy [83] que le domaine de l'intelligence artificielle a vraiment pris naissance avec beaucoup d'ambitions. Par la suite, au cours des années 70, l'intelligence artificielle a fait l'objet d'un bref recul. Le manque de financement a limité la recherche en apprentissage

en profondeur et en intelligence artificielle. Cependant, certains ont poursuivi la recherche sans financement pendant ces années difficiles. En effet, ces dernières décennies, le domaine de l'intelligence artificielle a connu, un développement continu et accéléré. Les progrès considérables de l'I.A sont le résultat d'une part, du développement de nouvelles conceptions théoriques, modèles et techniques; et d'autre part des avancées technologiques en matière de vitesse de traitement des ordinateurs récents.

L'apprentissage est l'une des composantes principales de l'intelligence artificielle. En s'appuyant sur l'histoire de l'intelligence artificielle [84], on constate une forte relation entre l'I.A et l'apprentissage automatique. En effet, les premiers travaux sur la modélisation du cerveau humain, sont en réalité du domaine de l'apprentissage automatique. Ce dernier, considéré comme une approche de l'intelligence artificielle, connaît une progression intensive, surtout avec l'arrivée de nouvelles technologies.

L'intelligence artificielle est devenue très répandue dans une multitude de domaines et ne cesse de progresser de jour en jour. Parmi ces domaines on site les suivants :

- **L'intelligence artificielle du jeu vidéo :** l'IA joue un rôle crucial dans les jeux stratégiques ou les jeux de rôle où la machine peut penser à un grand nombre de réactions possibles des personnages.
- **Les systèmes experts :** certaines applications utilisent des machines et des logiciels permettant de raisonner et de proposer des conseils aux utilisateurs.
- **Les robots intelligents :** les robots sont des machines ou des appareils fonctionnant automatiquement ou par télécommande. Ils ont des capteurs pour détecter les données physiques du monde réel telles que la température, la chaleur, la lumière, les mouvements, le son, les chocs, la pression, etc. Ils sont capables d'apprendre de leurs erreurs et de s'adapter à un nouvel environnement afin d'accomplir les tâches confiées à un humain [85].
- **La reconnaissance et classification vocale :** certains systèmes intelligents sont capables d'entendre et de comprendre la voix humaine ou de faire de la classification vocale. Ces systèmes permettent d'analyser le langage en termes de phrases et de leurs significations ou leurs catégories pendant qu'un humain parle.
- **La reconnaissance de l'écriture manuscrite :** c'est l'un des domaines de recherche les plus commodes et les plus difficiles dans le cadre du traitement de l'image et de la reconnaissance des formes au cours des dernières années. Les applications dotées d'un système de reconnaissance de l'écriture manuscrite contribuent énormément à

l'avancement du processus d'automatisation et peuvent améliorer l'interface entre l'homme et la machine.

• **Les systèmes de vision :** ces systèmes interprètent et comprennent les entrées visuelles de l'ordinateur. Par exemple, les médecins utilisent un système expert clinique pour diagnostiquer le patient, un avion espion prend des photographies qui sont utilisées pour comprendre des informations topographiques, la police utilise un logiciel informatique capable d'identifier une personne criminelle grâce aux systèmes de la reconnaissance faciale.

II.3. Apprentissage automatique et apprentissage profond

Durant le développement continu et la vulgarisation des nouvelles technologies ces dernières années, des termes tels que intelligence artificielle, l'apprentissage automatique et l'apprentissage profond sont souvent confondus et sont définis d'une manière similaire, malgré que ces disciplines sont totalement différentes. Pour illustrer la relation entre ces termes, il est possible de schématiser la relation d'appartenance montrée dans la figure II.1.

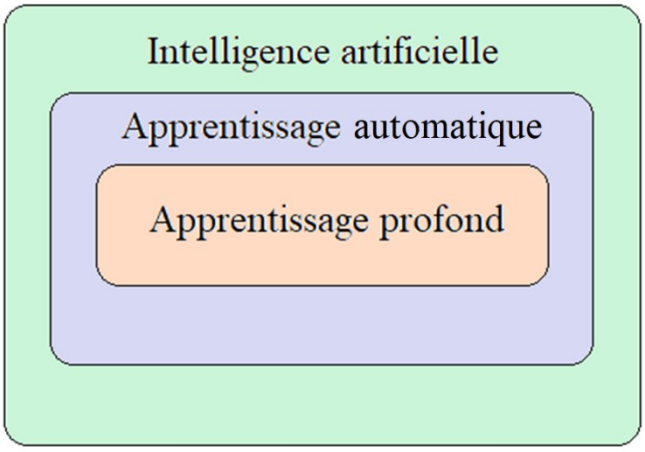


Figure II.1 : Concept d'appartenance des différentes branches de l'intelligence artificielle.

• **Intelligence artificielle (IA) :** est une technique qui permet aux ordinateurs d'imiter l'intelligence humaine.

• **Apprentissage automatique ou machine learning (ML) en anglais :** est un sous-ensemble de l'intelligence artificielle utilisant des techniques (telles que l'apprentissage profond), qui permettent aux machines de tirer des enseignements de leurs expériences pour améliorer la manière dont elles exécutent leurs tâches.

• **Apprentissage profond, aussi appelé Deep Learning (DL) en anglais :** est un sous-ensemble de l'apprentissage automatique, basé sur des réseaux neuronaux artificiels. Le processus d'apprentissage est qualifié de profond parce que la structure des réseaux

neuronaux artificiels se compose de plusieurs couches d'entrée, de sortie et ces couches sont masquées.

II.4. Apprentissage automatique (machine learning)

L'apprentissage automatique est une technique qui fait partie de l'intelligence artificielle pour former des modèles complexes [86]. Ces modèles peuvent faire fonctionner l'ordinateur ou le système de manière indépendante sans l'intervention humaine. En outre, l'apprentissage automatique utilise des données pour alimenter un modèle qui assure la relation entre l'entrée et la sortie. Quand la phase d'apprentissage est terminée, le programme peut prédire les valeurs ou les classes des données à tester. L'apprentissage automatique est aussi un domaine de recherche en informatique qui traite des méthodes d'identification et de mise en œuvre des systèmes et des modèles par lesquels un ordinateur peut apprendre sur la base des donnés en entrée. Cet apprentissage est un outil très efficace pour analyser, identifier et comprendre un ensemble de données très volumineux afin de permettre à un ordinateur de construire automatiquement des modèles complexes et à prendre des décisions aussi intelligentes que possible.

Le processus d'apprentissage automatique se compose de deux étapes de traitement des données. La première étape nécessite :

- **Un ensemble d'apprentissage** : il s'agit de la base de données utilisée pour former le modèle d'apprentissage automatique qui va être ensuite utilisé dans la phase test. La seconde étape nécessite :
- **Un ensemble de test** : ensemble des données qui seront utilisées uniquement pour évaluer les performances du modèle et qui n'ont en général pas de relation avec l'ensemble d'apprentissage.

La théorie de l'apprentissage automatique utilise des outils mathématiques provenant des lois des probabilités et de la théorie de l'information. Il y a essentiellement trois types d'apprentissage qui sont schématisés dans la figure II.2 : apprentissage supervisé, apprentissage non supervisé et apprentissage par renforcement [87].

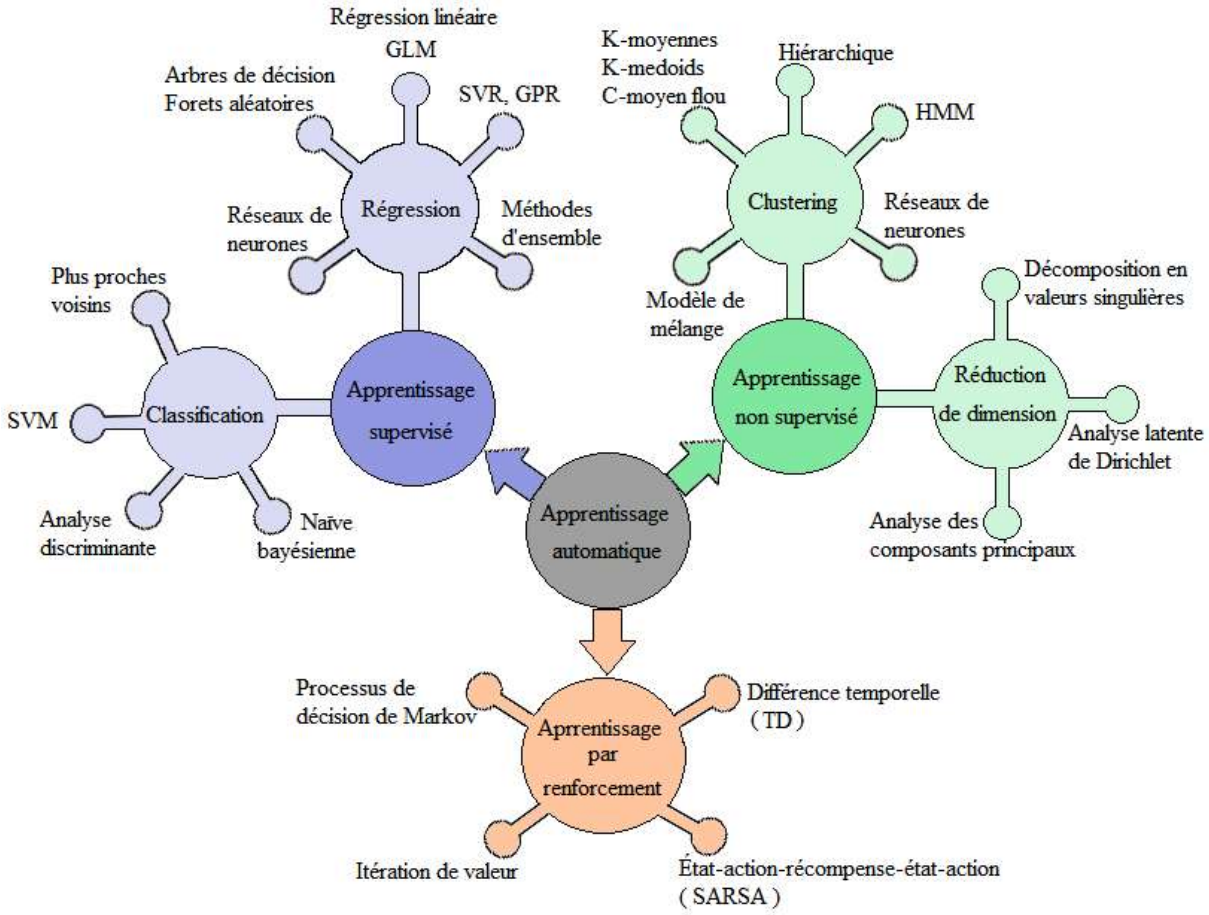


Figure II.2 : Principaux algorithmes d'apprentissage automatique [87].

II.4.1. Apprentissage supervisé

Dans l'apprentissage supervisé, on utilise des données étiquetées connues. Comme les données sont connues et étiquetés, l'apprentissage est donc supervisé [88]. Les données d'entrée passent par l'algorithme d'apprentissage automatique et sont utilisées pour former le modèle. Une fois que le modèle est bien formé sur la base des données étiquetées connues, il va tenter de déterminer le type de données dans la phase test. Les principales applications de l'apprentissage supervisé sont la classification et la régression.

II.4.1.1. Classification

Un classificateur est un mécanisme à n entrées [88] représentant les entités $e_1, e_2, \ldots, e_n$ calculées à partir d'un objet à classer, et une sortie. Un classificateur de classe S générera l'un des symboles $w_1, w_2, \ldots, w_s$ comme sortie, et l'utilisateur interprète cette sortie comme une décision concernant la classe de l'objet traité. Les symboles générés w_s sont les identificateurs de classes. Les algorithmes de classification tels que SVM, K plus proches voisins (K-nearest neighbor K-NN), Bayes Naïf sont des algorithmes pour classifier des

données. Les données peuvent être des mots, des couleurs, des sons, des pixels de ces images, etc.

i) Les séparateurs à vaste marge

Les séparateurs à vaste marge ou support vector machines (SVM) en anglais, sont un type d'algorithmes de classification [89-92], ils peuvent être utilisés pour résoudre des problèmes de discrimination. Autrement dit, ils peuvent décider à quelle classe appartient un échantillon. La technique des SVMs consiste à créer un hyperplan dans un espace de grande dimension afin de séparer les données à classifier, cette séparation de données est réalisée par l'hyperplan qui a la plus grande distance de séparation entre les points de données des classes les plus proches. Ces points de données sont appelés la marge. L'erreur de généralisation du classificateur SVM dépend de la taille de la marge [89]. L'algorithme d'apprentissage SVM construit un modèle sur la base d'une marge fonctionnelle qui crée un classificateur binaire.

Le modèle construit peut être utilisé pour la classification linéaire ou non linéaire. La classification non linéaire est effectuée à l'aide d'une fonction basée sur le noyau pour mapper l'entrée dans un espace d'entités de grande dimension [93]. Dans cet algorithme, la prise en charge des points de données de formation appartient chacune à deux classes dans le cas binaire avec comme objectif de décider de la classe d'un nouveau point de données.

ii) K plus proches voisins

Le principe de l'algorithme K plus proches voisins (K-nearest neighbor : K-NN) [94-96] est de trouver un nombre prédéfini d'échantillons d'apprentissage les plus proches en termes de distance par rapport à la nouvelle entité (entrée), où chaque point de données est caractérisé par un ensemble de variables. C'est à dire, chaque point est tracé dans un espace de grande dimension, où chaque axe de l'espace correspond à une variable individuelle. Autrement dit, l'algorithme K-plus proches voisins consiste à récupérer les entités voisines les plus proches par rapport à la nouvelle entité et à les affecter à une classe. L'algorithme K-NN est sans doute le plus simple de tous les algorithmes de classification supervisée, cet algorithme continue de fonctionner assez bien pour des données de grandes tailles en apprentissage. Il ne nécessite que le choix de K qui est le nombre de voisins à considérer lors de la classification. Le nombre K est généralement choisi comme la racine carrée de N qui est le nombre total de points dans l'ensemble de

données d'entraînement. Lorsqu'on a un nouveau point de données à tester, il faut trouver les K voisins les plus proches en termes de distance.

iii) Bayes naïf

Bayes naïf (Naive Bayes : NB) [97-98] est un algorithme d'apprentissage supervisé basé sur l'application du théorème de Bayes. Il s'appuie sur la règle de révision permettant de modifier une distribution a priori, en une distribution a posteriori tenant compte des données d'apprentissage. En fait, cette règle énonce que la probabilité a posteriori d'une hypothèse est égale à sa probabilité a priori multipliée par la vraisemblance des données étant donnée l'hypothèse. Plus formellement, la célèbre règle de Bayes des probabilités s'écrit :

$$P(y \mid x_1,...,x_n) = \frac{P(y)P(x_1,...,x_n \mid y)}{P(x_1,...,x_n)} \quad \text{(II.1)}$$

On dénote par $P(y)$ et $P(y|X)$, respectivement, la densité de probabilité a priori et la densité de probabilité a posteriori après prise en compte des données X, où $P(X|y)$ est la mesure de probabilité d'un ensemble d'apprentissage et $P(X)$ est la probabilité a priori de l'événement X, où $X=x_1, x_2, ..., x_n$.

Bien que l'algorithme Bayes naïf soit simple dans sa conception et son développement, l'utilisation de cet algorithme a donné des résultats très encourageants, et il est très utilisé dans le domaine de la classification.

II.4.1.2. Régression

La régression sert à trouver la relation d'une variable par rapport à une ou plusieurs autres. Dans l'apprentissage automatique, le but de la régression est d'estimer une valeur (numérique) de sortie à partir des valeurs d'un ensemble de caractéristiques en entrée. Par exemple, estimer le prix d'une maison en se basant sur sa surface, nombre des étages, son emplacement, etc. Donc, le problème revient à estimer une fonction de calcul en se basant sur des données d'entrainement :

$$\hat{y}(x) = f(x_1, x_2, ..., x_m) \quad \text{(II.2)}$$

Il existe plusieurs algorithmes pour la régression : régression linéaire, régression polynomiale, régression logistique, régression quantile, etc.

i) Régression linéaire

La régression linéaire ou (linear regression : LR) tente de modéliser la relation entre deux variables en ajustant une équation linéaire aux données observées, l'une des variables est considérée comme variable explicative et l'autre est considérée comme une variable dépendante [99]. Par exemple, relier les poids des individus à leurs hauteurs à l'aide d'un modèle de régression linéaire. Une droite de régression linéaire possède une équation de la forme $Y = aX + b$, où X est la variable explicative, Y est la variable dépendante, a est la pente de la droite et b est l'ordonnée à l'origine.

ii) Arbres de décision

L'apprentissage avec les arbres de décision (decision tree : DT), est une méthode classique en apprentissage automatique. Son but est de créer un modèle qui prédit la valeur d'une variable cible depuis la valeur de plusieurs variables d'entrée.

Une des variables d'entrée est sélectionnée à chaque nœud intérieur. Chaque arête vers un nœud-fils correspond à un ensemble de valeurs d'une variable d'entrée, de manière que l'ensemble des arêtes vers les nœuds-fils couvrent toutes les valeurs possibles de la variable d'entrée.

Chaque feuille (ou nœud terminal de l'arbre) [100] représente soit une valeur de la variable cible, soit une distribution de probabilité des diverses valeurs possibles de la variable cible. La combinaison des valeurs des variables d'entrée est représentée par le chemin de la racine jusqu'à la feuille.

L'arbre est en général construit en séparant l'ensemble des données en sous-ensembles en fonction de la valeur d'une caractéristique d'entrée. Ce processus est répété sur chaque sous-ensemble obtenu de manière récursive, il s'agit donc d'un partitionnement récursif.

L'ensemble d'apprentissage est généralement fourni sous la forme d'enregistrements du type [101] :

$$(x, Y) = (x_1, x_2, x_3, \ldots, x_k, Y) \tag{II.3}$$

La variable Y désigne la variable cible que l'on cherche à prédire, classer ou généraliser. Le vecteur x est constitué des variables d'entrée $x_1, x_2, x_3, \ldots$ etc., qui sont utilisées dans ce but. Chaque nœud interne de l'arbre correspond à un test effectué sur une des variables x_i.

iii) Forets aléatoires

Les forets aléatoires (random forest : RF), sont des méthodes qui permettent d'obtenir des modèles prédictifs pour la classification et la régression. La méthode met en œuvre des arbres de décision binaire, notamment des arbres CART (classification and regression trees) proposés par Breiman en 1984 [102].

L'idée générale derrière la méthode est la suivante : au lieu d'essayer d'obtenir une méthode optimisée en une fois, on génère plusieurs prédicteurs avant de mettre en commun leurs différentes prédictions.

iv) Régression vectorielle de support

La régression vectorielle de support (support vector regression : SVR), algorithme de classification des vecteurs de support peut être étendu pour résoudre les problèmes de régression [103, 104]. Comme son nom l'indique, le SVR est aussi un algorithme de régression, nous pouvons donc l'utiliser pour travailler sur des valeurs continues contrairement à la classification.

L'objectif principal du SVR est de trouver des modèles dont les prédictions ont un écart maximal par rapport aux valeurs de la fonction de régression pour les données d'apprentissage [104]. Bien que le SVR permet de définir un hyper plan qui est une ligne de séparation entre les classes de données, la ligne de séparation aidera à prédire la valeur continue ou la valeur cible.

II.4.2. Apprentissage non supervisé

Dans l'apprentissage non supervisé, les données sont inconnues et non étiquetées pour la phase d'apprentissage. L'algorithme doit découvrir lui-même la structure en fonction des données, dans ce cas on parle d'apprentissage non supervisé ou clustering. Ces données sont introduites dans l'algorithme d'apprentissage automatique et sont utilisées pour former le modèle. Le modèle formé essaie de rechercher le type de données dans la phase test.

Les principales branches utilisées pour l'apprentissage non supervisé sont le partitionnement (clustering) et la réduction de dimension.

II.4.2.1. Partitionnement

Le partitionnement (clustering) est une classification non supervisée des objets [105]. C'est une technique courante d'analyse de données, utilisée dans de nombreux domaines,

notamment l'apprentissage automatique, la reconnaissance de formes, l'analyse d'images et la bioinformatique. Les algorithmes de clustering partitionnent les données en un certain nombre de clusters (catégories, sous-ensembles ou groupes). C'est un processus de regroupement des objets similaires en différents groupes, ou plus précisément, le partitionnement d'un ensemble de données en sous-ensembles selon une mesure de distance définie (distance euclidienne par exemple).

i) Partitionnement en K moyennes

Généralement, le partitionnement en K moyennes (K-means clustering) est considéré comme l'une des techniques les plus efficaces de segmentation non supervisée [106, 107], elle partitionne un certain nombre de points en K groupes dans lesquels chaque point appartient au groupe qui a la moyenne la plus proche. Les étapes de l'algorithme de partitionnement en K moyennes sont les suivantes :

1) Calculer la valeur moyenne de chaque groupe.

2) Affecter chaque point au groupe le plus proche en calculant la distance de chaque point par rapport à la valeur moyenne du groupe correspondant.

L'algorithme de K-means est un processus itératif pour calculer la valeur moyenne de chaque groupe et pour calculer la distance de chaque point par rapport au groupe le plus proche. Ce processus itératif est répété jusqu'à ce qu'aucune réduction supplémentaire ne puisse être obtenue dans la somme des erreurs carrées dans chaque groupe.

ii) Classification ascendante hiérarchique

La classification ascendante hiérarchique (hierarchical ascendant clustering : HAC) [108], est l'une des méthodes de regroupement les plus populaires, en outre c'est une méthode de classification itérative dont le principe est simple :

1. On commence par calculer la dissimilarité entre les *N* objets.
2. Puis on regroupe les deux objets dont le regroupement minimise un critère d'agrégation donné, créant ainsi une classe comprenant ces deux objets.
3. On calcule ensuite la dissimilarité entre cette classe et les *N*-2 autres objets en utilisant le critère d'agrégation. Puis on regroupe les deux objets ou classes d'objets dont le regroupement minimise le critère d'agrégation.

On continue ainsi jusqu'à ce que tous les objets soient regroupés.

Ces regroupements successifs produisent un arbre binaire de classification (dendrogramme), dont la racine correspond à la classe regroupant l'ensemble des individus.

Ce dendrogramme représente une hiérarchie de partitions. On peut alors choisir une partition en tronquant l'arbre à un niveau donné qui dépend soit des contraintes de l'utilisateur (l'utilisateur sait combien de classes il veut obtenir), soit de critères plus objectifs.

La classification ascendante hiérarchique (CAH) est une méthode de classification qui présente les avantages suivants :

- On travaille à partir des dissimilarités entre les objets que l'on veut regrouper. On peut donc choisir un type de dissimilarité adapté au sujet étudié et à la nature des données.
- L'un des résultats est le dendrogramme, qui permet de visualiser le regroupement progressif des données. On peut alors se faire une idée d'un nombre adéquat de classes dans lesquelles les données peuvent être regroupées.

iii) Modèles de Markov cachés

Les modèles de Markov cachés (hidden Markov models : HMM) sont des modèles statistiques définis par un ensemble de paramètres qui sont appris sur un corpus de données d'apprentissage. Ils implémentent une densité de probabilité sur les données séquentielles. Il existe des états cachés ou non observés dans un modèle de Markov caché, contrairement à une chaîne de Markov standard où tous les états sont visibles pour l'observateur.

Le modèle de Markov caché a été développé par le mathématicien Baum et ses collègues dans les années 1960 [109]. Ce modèle tente de prédire l'état futur d'une variable en utilisant des probabilités basées sur l'état actuel et passé. La principale différence entre une chaîne de Markov et le modèle de Markov caché est que l'état de ce dernier n'est pas directement visible par un observateur, même si la sortie l'est.

Les modèles de Markov cachés sont utilisés pour les tâches d'apprentissage automatique et d'exploration de données. Certains d'entre eux comprennent la reconnaissance vocale, la reconnaissance de l'écriture manuscrite, le balisage d'une partie du discours et la bioinformatique.

iv) Modèles de mélange gaussien

Les modèles de mélange gaussien (Gaussian mixture models : GMM) font partie des approches de partitionnement (clustering) qui utilisent des modèles probabilistes pour distribuer les points dans différentes partitions (clusters) [110]. Comme leur nom l'indique, les GMMs sont un mélange de plusieurs distributions gaussiennes, ces modèles permettent de mesurer l'incertitude ou la probabilité qui indique à quel point de données est associé à

un groupe spécifique. En outre, les GMMs supposent qu'il existe un certain nombre de distributions gaussiennes, et chacune de ces distributions représente un cluster. Par conséquent, un modèle de mélange gaussien a tendance à regrouper les points de données appartenant à une seule distribution. En fait, un GMM peut être utilisé pour regrouper des données non étiquetées de la même manière que l'algorithme K-means, ce qui signifie qu'il associera chaque point à un seul cluster. Une limite à cette approche est qu'il n'y a pas de mesure d'incertitude ou de probabilité qui indique à quel point un point de données est associé à un cluster spécifique.

Un mélange gaussien est une fonction qui est composée de plusieurs gaussiennes, avec chaque gaussienne comprenant les paramètres suivants :

- Une moyenne μ qui définit son centre.
- Une covariance Σ qui définit sa largeur.
- Une probabilité de mélange π qui définit la taille de la fonction gaussienne.

II.4.2.2. Réduction de dimension

La réduction de dimensionnalité en apprentissage automatique consiste à réduire le nombre de variables dans les données d'apprentissage [111], afin d'obtenir plus d'efficacité en termes de résultats et de temps d'analyse.

En apprentissage automatique, les grandes dimensions nuisent à l'efficacité des systèmes d'apprentissage automatique. La production de résultats en trop grand nombre rend difficile d'associer ou de comparer ces quantités de données. En utilisant un espace de plus petite dimension, on obtient des algorithmes plus efficaces, ainsi qu'un ensemble de solutions plus réduit.

La réduction de dimensionnalité permet, en réduisant le nombre de variables dans les données d'apprentissage, de mieux visualiser les données obtenues pour effectuer des comparaisons et des analyses plus fiables. Son intérêt est également de réduire les coûts de calcul et de stockage des informations. La réduction de la dimension des données permet une économie de l'espace mémoire, une meilleure efficacité des algorithmes et un gain de temps parfois considérable. De même, elle permet une amélioration de l'apprentissage automatique, en construisant des modèles plus simples, où les variables inutiles ont été écartées.

i) Analyse en composantes principales

L'analyse en composantes principales (principal component analysis : PCA) [112] est une méthode de réduction de la dimensionnalité qui est souvent utilisée pour transformer un grand ensemble de variables en un ensemble plus petit qui contient encore la plupart des informations dans le grand ensemble [113].

L'idée est de transformer des variables corrélées en nouvelles variables non corrélée en projetant les données dans le sens de la variance croissante. Les variables avec la variance maximale seront choisies comme les composants principaux.

Le PCA est une technique de l'apprentissage automatique très utilisée dans le traitement des données scientifiques, elle intervient dans de nombreuses situations ou le jeu de données en question est compliqué et difficile à décortiquer. Elle est appliqué dans :

- L'étude des corrélations entre les variables.
- L'obtention de facteurs non corrélés.
- La visualisation des observations dans un espace ou une base à deux ou trois dimensions.

Le PCA est donc une méthode de réduction de la dimensionnalité qui permet non seulement d'éviter le sur-apprentissage, mais aussi de simplifier les données et de gagner en matière de temps et de puissance de calcul.

ii) Sélection des fonctionnalités

Les algorithmes de sélection des fonctionnalités (feature selection) [114] est une partie importante de l'apprentissage automatique. Elle désigne le processus visant à réduire les entrées à traiter et à analyser, ou à identifier les entrées les plus pertinentes en visant à extraire des informations ou des fonctionnalités utiles à partir de données existantes.

La sélection des fonctionnalités est essentielle à la création d'un modèle performant pour plusieurs raisons. L'une d'elles est d'optimiser le processus de modélisation en supprimant les colonnes inutiles pour générer un modèle qui permet de libérer davantage l'unité centrale et l'espace mémoire pendant le processus d'apprentissage.

Les algorithmes les plus utilisés dans la sélection des fonctionnalités sont les modèles de filtre [115] et du wrapper [116]. Le modèle de filtre évalue le score de chaque fonctionnalité selon certains critères, ensuite il sélectionne les caractéristiques qui ont le score le plus élevé. Le modèle wrapper utilise un algorithme de clustering pour évaluer la qualité des caractéristiques sélectionnées. Il commence en premier par trouver un sous-ensemble de caractéristiques, en second, il évalue la qualité du clustering en utilisant le

sous-ensemble sélectionné, enfin, il répète les deux étapes jusqu'à ce que le critère souhaité soit atteint.

En fait, la sélection des fonctionnalités permet à la fois d'éviter d'avoir trop de données qui présentent peu d'intérêt et de n'avoir pas assez de données utiles. L'objectif principal est d'identifier le nombre minimum de colonnes de la source de données qui sont importantes pour le modèle à créer.

iii) Analyse discriminante linéaire

En statistique, l'analyse discriminante linéaire (linear discriminant analysis : LDA) [117] fait partie des techniques d'analyse discriminante prédictive. Il s'agit de prédire l'appartenance d'une entité à une classe prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives. En fait, l'analyse discriminante linéaire peut être comparée aux méthodes supervisées développées en apprentissage automatique et à la régression logistique développée en statistique.

Comme son nom l'indique, les techniques de réduction de la dimensionnalité réduisent le nombre de dimensions ou de variables dans un ensemble de données tout en conservant autant d'informations que possible. Cette technique est utilisée lorsqu'il y a une ou plusieurs variables indépendantes. L'analyse discriminante linéaire effectue un test de différence multivarié entre les groupes. Elle est également utile pour déterminer le nombre minimum de dimensions nécessaires pour décrire ces différences. La méthode LDA est souvent utilisée comme technique de réduction de dimensionnalité dans l'étape de prétraitement pour les applications de classification de modèles et d'apprentissage automatique. L'objectif est de projeter un ensemble de données sur un espace de dimension inférieure avec une bonne séparabilité de classe afin de réduire les coûts de calcul.

II.4.3. Apprentissage par renforcement

L'apprentissage par renforcement (reinforcement learning : RL) est une approche de l'intelligence artificielle qui met l'accent sur l'apprentissage du système à travers ses interactions avec l'environnement [118-120]. A l'aide de cet apprentissage, le système adapte ses paramètres en fonction des commentaires reçus de l'environnement. Ensuite, ce système va fournir des commentaires sur les décisions prises. Par exemple, un système qui modélise un joueur d'échecs qui utilise le résultat des étapes précédentes pour améliorer ses performances est un système qui apprend par renforcement. La recherche actuelle sur l'apprentissage par renforcement est fortement inter-disciplinaire faisant intervenir des

chercheurs spécialisés dans les algorithmes génétiques, les réseaux de neurones, la psychologie et l'ingénierie du contrôle.

II.5. L'apprentissage profond et réseaux de neurones

L'apprentissage profond ou Deep Learning en Anglais, est un type d'apprentissage automatique (ML) et d'intelligence artificielle (IA) qui imite la façon dont les humains acquièrent certains types de connaissances. Le Deep Learning est un élément important de la science des données, qui comprend les statistiques et la modélisation prédictive. Il est extrêmement utile aux scientifiques chargés de collecter, d'analyser et d'interpréter de grandes quantités de données et rend ce processus plus rapide et plus facile.

Les réseaux de neurones artificiels (artificial neural network en Anglais) sont des systèmes informatiques logiciels et/ou matériels dont le fonctionnement imite celui des neurones du cerveau humain, il s'agit là d'une variété de technologie d'apprentissage profond (Deep Learning).

En règle générale, un réseau de neurones repose sur un grand nombre de processeurs opérant en parallèle et organisés en tiers. Le premier tiers reçoit les entrées d'informations brutes, un peu comme les nerfs optiques de l'être humain lorsqu'il traite des signaux visuels. Par la suite, chaque tiers reçoit les sorties d'informations du tiers précédent. On retrouve le même processus chez l'homme, lorsque les neurones reçoivent des signaux en provenance des neurones proches du nerf optique. Le dernier tiers, quant à lui, produit les résultats du système.

Les réseaux de neurones artificiels apprennent à effectuer de nouvelles tâches en analysant des exemples pour s'entraîner, ces exemples ont préalablement été étiquetés afin que l'ordinateur puisse être autonome pour ces décisions futures.

Les usages des RNA sont nombreux, et se multiplient au fil du temps. Parmi les exemples d'application, on peut citer la reconnaissance d'écriture manuscrite, la prévision des marchés financiers, ou encore la reconnaissance faciale, la prédiction météo, ils sont aussi utilisés dans les algorithmes de traitement du langage. Toutes ces tâches ont un point commun, elles sont complexes à modéliser et elles ne requièrent pas une solution unique et exacte, mais plutôt une estimation de la réponse en se basant sur l'apprentissage artificiel.

II.5.1. Eléments de base des Réseaux de neurones

II.5.1.1. Le neurone formel

L'unité de base du RNA est le neurone formel tel qu'illustré dans la figure II.3.

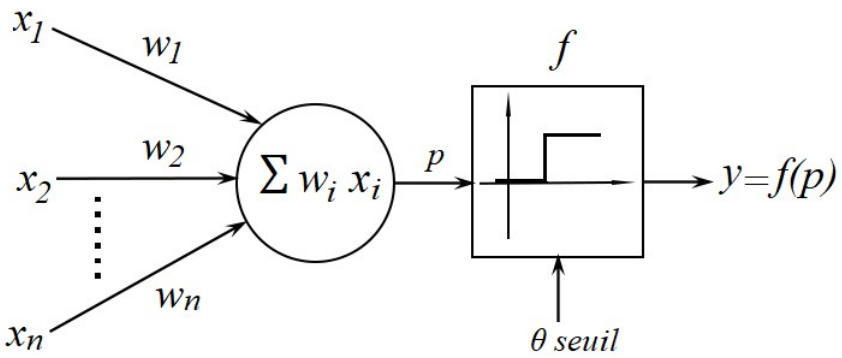


Figure II.3 : Neurone formel [81].

Le neurone formel reçoit un ensemble de signaux d'entrées x_i et fournit un signal de sortie y. Ces signaux sont pondérés par des poids w_i. Il est à noter que x_i et w_i sont représentés sous forme de vecteur. Chaque neurone formel calcule son potentiel $p = \sum w_i\, x_i$, ensuite le signal de sortie du neurone (ou activation) $y= f(p)$ est calculée en invoquant une fonction de transfert (une fonction seuil) f qui est appliquée sur le potentiel calculé.

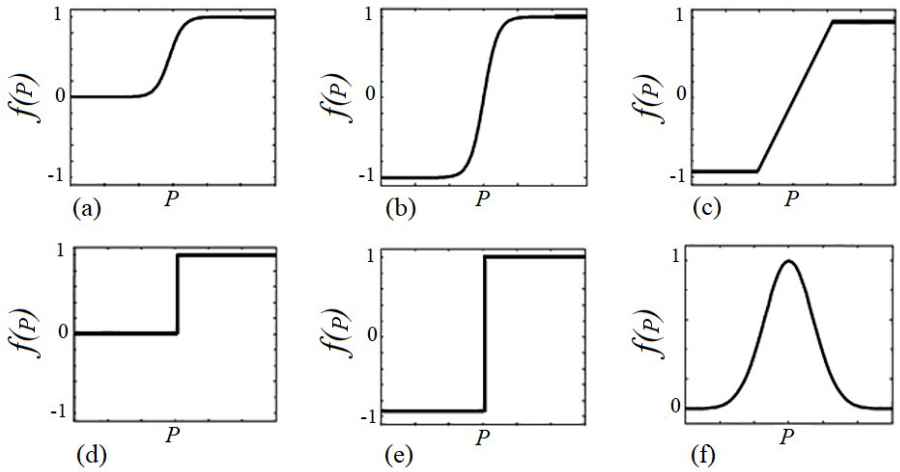
II.5.1.2. Les fonctions d'activation

La fonction d'activation peut avoir plusieurs formes différentes [121]. Dans leur modèle d'origine, McCulloch et Pitts [81] ont utilisé des fonctions d'activation à seuil. Les états des neurones avec de telles fonctions sont binaires et les ensembles de valeurs possible les plus couramment utilisées sont {-1, 1} ou {0, 1}. Les deux formes des fonctions à seuil sont :

$$Yseuil = \begin{cases} 1 \text{ si } P > \theta \\ 0 \quad \text{sinon} \end{cases} \tag{II.4}$$

Et

$$Yseuil = \begin{cases} +1 \text{ si } P > \theta \\ -1 \quad \text{sinon} \end{cases} \tag{II.5}$$



(a) Fonction seuil. (b) Fonction seuil bipolaire.
(c) Fonction RBF (d) Fonction sigmoïde.
(e) Fonction sigmoïde bipolaire. (f) Fonction linéaire saturée

Figure II.4 : Exemples de fonctions d'activation [121].

L'examen du comportement des cellules nerveuses par rapport aux signaux qui leur arrivent a donné lieu à l'utilisation d'un autre type très intéressant comme fonction d'activation : les fonctions sigmoïdales. Une fonction sigmoïde est donnée par :

$$f_{sigmoide}(P) = \frac{1}{1+\exp(-P)} \tag{II.6}$$

La figure II.4 donne quelques exemples des fonctions d'activation. Les fonctions RBF forment une catégorie particulière des réseaux de neurones : Les réseaux RBF.

II.6. Conclusion

L'intelligence artificielle évolue et progresse continuellement depuis les années quarante dans le domaine de l'informatique. Malgré qu'aujourd'hui l'IA ne soit pas totalement autonome c'est à dire qu'elle n'est pas capable de raisonner elle-même, cependant elle est tout de même capable de remplacer l'être humain dans plusieurs domaines tels que l'assistance et l'aide à la décision, le traitement et la reconnaissance de signaux complexes, la classification et l'identification des images, l'analyse et les prévisions, etc.

Ce chapitre a été consacré à l'étude des différentes branches de l'intelligence artificielle, notamment l'apprentissage automatique, l'apprentissage profond et les réseaux de neurones afin de recenser et décrire les différents algorithmes existant. Ces algorithmes basés sur l'apprentissage artificiel offre des concepts et des outils puissants et prometteurs

pour le développement de systèmes informatiques et pour la résolution de problèmes complexes.

L’étude de l’IA nous a offert une vision claire sur les concepts et les outils puissants utilisés dans ce domaine. L'apprentissage automatique n’a pas besoin d’un gros volume de données contrairement à l'apprentissage profond qui nécessite des grandes bases de données. D'un autre côté, l'apprentissage automatique a besoin d’un temps de calcul pour l'entrainement de quelques minutes tandis que l’apprentissage profond peut prendre de quelques heures à plusieurs jours. L’apprentissage profond a l’avantage de donner d’excellentes performances sur un grand ensemble de données, par contre l’apprentissage automatique est mieux adapté pour les petits et moyens ensembles de données.

Chapitre 3 :
Les machines à vecteurs de support

III.1. Introduction

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais, support vector machine : SVM), sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de discrimination, c'est à dire décider à quelle classe appartient un échantillon, ou de régression, c'est à dire prédire la valeur numérique d'une variable.
Les séparateurs à vastes marges reposent sur deux idées principales : (i) la notion de marge maximale de séparation des données des différentes classes et (ii) la notion de fonction noyau qui permet au SVM d'exploiter différentes façons de classifier autre que le cas linaire. Ces deux notions existaient déjà avant l'invention des SVMs. Ce chapitre est dédié à la théorie des SVMs. On présente les SVMs à deux classes et multi-classes ainsi que les cas linéaire et non linéaire.

III.2. Historique des SVMs

Les SVMs ont été appliqués à de très nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finance [122], etc.). Selon les données, les performances des machines à vecteurs de support sont du même ordre ou même supérieures, à celles d'un réseau de neurones ou d'un modèle de mélanges gaussiens [123, 124]. Les différents types existants des SVMs sont représentés dans la figure III.1 [125].

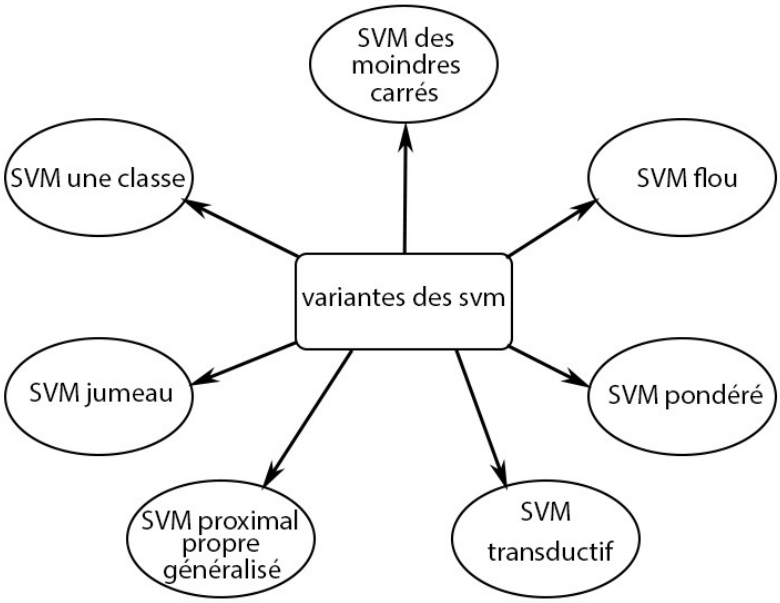


Figure III.1 : SVM avec ces différentes variantes [125].

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et Aleksander Lerner [126], ensuite en 1973 par Richard Duda et Peter Hart dans leur livre Pattern Classification [127].

Les fondements théoriques des SVMs ont été explorées par Vapnik et Chervonenkis en 1971 par le développement de la Théorie de VC [128] qui est une forme de théorie de l'apprentissage automatique, qui tente d'expliquer l'apprentissage d'un point de vue statistique, ensuite par les travaux de Leslie Valianten en 1984 [129] avec la théorie de l'apprentissage PAC (probably approximately correct) qui est un cadre théorique pour l'apprentissage automatique qui permet notamment d'évaluer la difficulté d'un problème dans le contexte de l'apprentissage supervisé.

La notion des fonctions noyaux n'est pas non plus nouvelle, le théorème de Mercer date de 1909 [130] et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizerman et ses collègues [131]. Ce n'est toutefois qu'en 1992 que ces idées seront bien exploitées et rassemblées par Boser et ses collègues dans leur article [132], qui sera le fondateur des séparateurs à vaste marge.

Il y a eu ensuite le concept des variables ressorts ajoutées dans la maximisation de la marge, afin de permettre la résolution de certaines limitations pratiques, qui a été introduit en 1995 par Cortes et Vapnik [133]. À partir de cette date qui correspond à la publication du livre de Vapnik [134], les SVMs gagnent en popularité et seront utilisés dans de nombreuses applications.

III.3. Notion d'apprentissage

Pour les SVMs, l'apprentissage est du type supervisé et se fait en présentant des exemples de données qui sont représentés par un ensemble de couples « descriptions d'objet / classe prédéfinie » dans le but de former une fonction qui correspond aux exemples présentés et qui prédit les sorties pour les entrées qui n'ont pas encore été présentés afin de prédire leur classe. Cette façon d'apprendre s'appelle apprentissage par induction.

Les SVMs engendrent par construction un hyperplan qui repose sur un sous-ensemble de données d'apprentissages appelées vecteurs de support. Ces vecteurs de support et leurs poids associés suffisent pour construire la frontière de décision. Le processus d'apprentissage des SVMs consiste donc en un problème d'optimisation visant à maximiser la marge de l'hyperplan séparateur dans l'espace de représentation (dépendant du réglage du noyau) avec comme variables d'optimisation les poids associés à chacun des exemples de la base d'apprentissage. Finalement, les SVMs sont entraînés via un algorithme d'optimisation des paramètres de l'hyperplan séparateur [134].

III.4. Principe de fonctionnement général

Pour deux classes d'exemples de données, le but des SVMs est de trouver un classificateur qui sépare les données et maximise la distance entre ces deux classes. Avec le SVM, ce classificateur est linéaire, quadratique ou gaussien selon le noyau utilisé. Ce classificateur se manifeste par un hyperplan de séparation [135]. Dans la figure III.2, on détermine un hyperplan qui sépare deux ensembles de points.

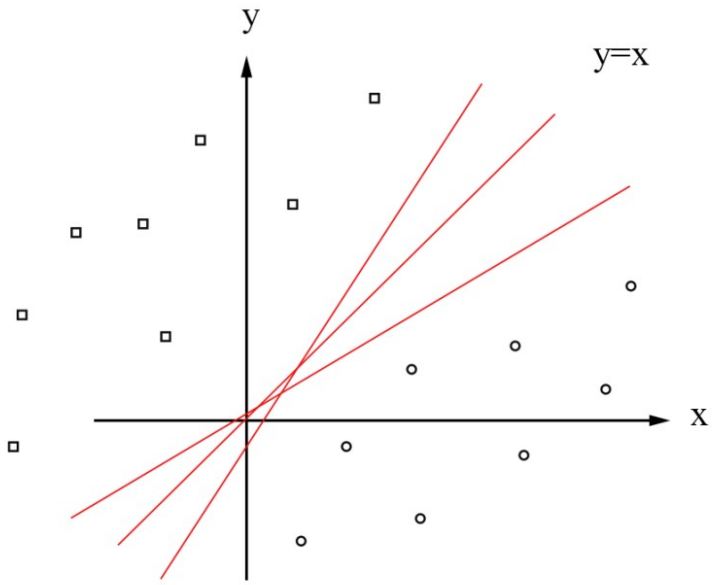


Figure III.2 : Séparation de deux ensembles de points par un Hyperplan H.

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

Il est évident qu'il existe une multitude d'hyperplans valables mais la propriété remarquable des SVMs est que l'hyperplan doit être optimal. On cherche donc, en plus parmi les hyperplans valables, celui qui passe « au milieu » des points des deux classes d'exemples. Cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale. On appelle cette distance « marge » entre l'hyperplan et les exemples.

L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge [136]. Dans la figure III.3, on détermine un hyperplan qui sépare les deux ensembles de points.

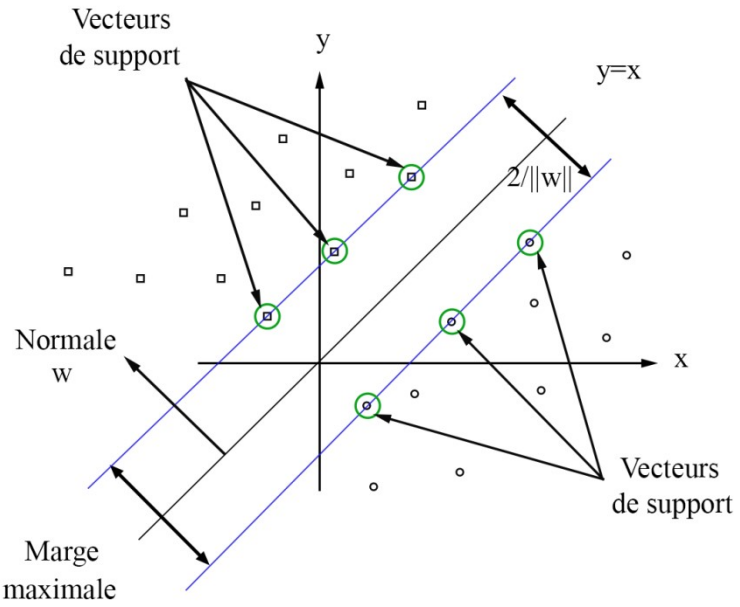


Figure III.3 : Hyperplan optimal, marge et vecteurs de support.

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsqu'on classe un nouvel exemple. De plus, si on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples.

III.4.1. Linéarité et non linéarité

Dans les SVMs, on distingue deux cas de séparabilité : linéairement séparable et non linéairement séparable.

III.4.1.1. Cas linéairement séparable

Dans un problème linéairement séparable, les SVMs trouvent facilement un séparateur (classificateur linéaire) qui maximise la marge.

III.4.1.2. Cas non linéairement séparable

Dans la plupart des problèmes réels, il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [137] (figure III.4).

Le principe consiste à projeter les données de l'espace d'entrée (appartenant à deux classes différentes) non linéairement séparables dans un espace de plus grande dimension appelé

« espace de caractéristiques ou espace de redescription » de façon à ce que les données deviennent linéairement séparables (figure III.5).

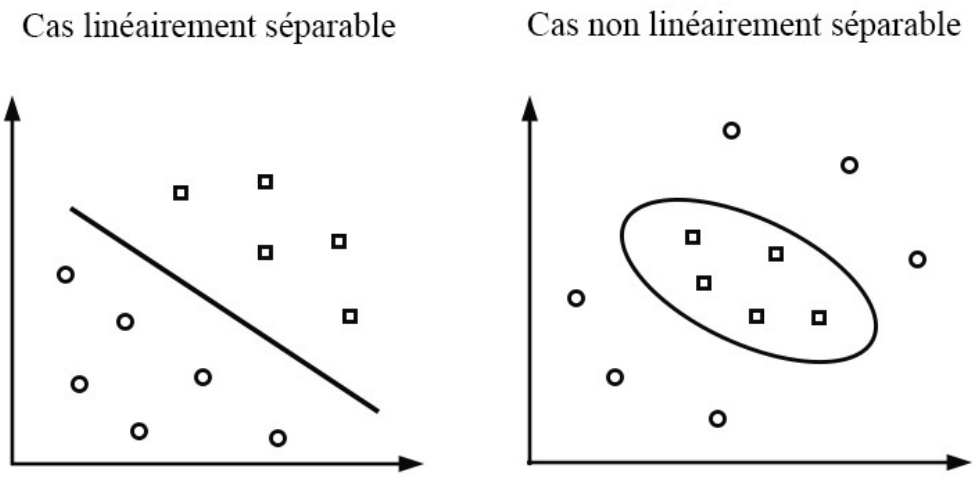


Figure III.4 : Linéarité et non linéarité de séparation.

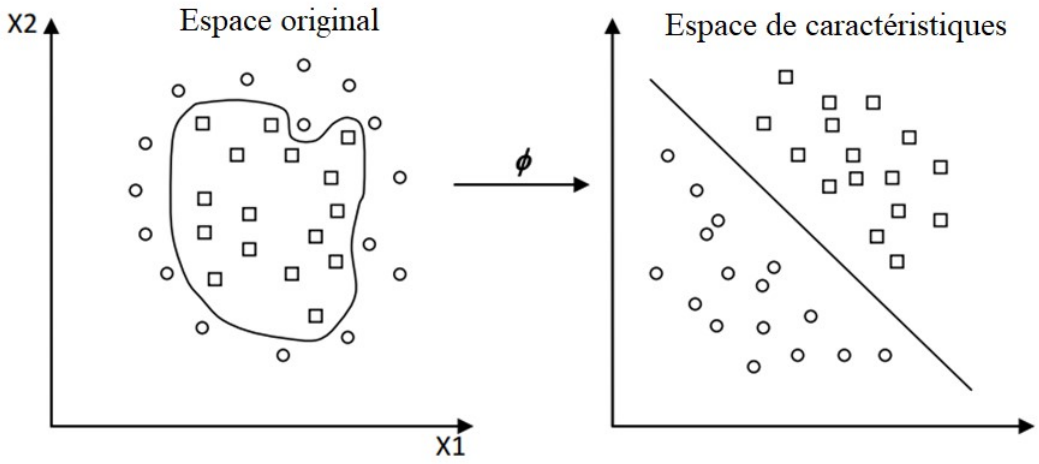


Figure III.5 : Transformation de l'espace de représentation et l'hyperplan séparateur dans le cas non linéairement séparable.

On a donc une transformation d'un problème de séparation non linéaire dans un espace de représentation en un problème de séparation linéaire dans un espace de description de plus grande dimension. Cette transformation est réalisée via une fonction noyau.

III.5. Fondements mathématiques

III.5.1. Principe

Pour bien comprendre les SVMs, il faut savoir qu'ils sont basés sur deux algorithmes clés, à savoir la combinaison d'un classificateur linéaire et d'un classificateur K plus proche voisin (K nearest neighbors : K-NN) [138]. Pour le classificateur linéaire c'est simple, une

fois formé, il devient un outil ou une machine qui peut être utilisé pour dire si catégoriser une donnée à l'une des classes. Cependant, un classificateur K-NN est également simple et peut séparer les données de manière non linéaire.
Un SVM est fondamentalement un classificateur linéaire binaire, principalement avec deux objectifs importants à atteindre :

- Maximiser la marge qui sépare les deux classes;
- N'utiliser que quelques données d'apprentissage (ou vecteurs supports) pour déterminer l'hyperplan qui sépare les deux classes.

Afin d'être capable de classer des données non linéairement séparables, il est nécessaire d'ajouter un noyau au SVM.
Une fois ces objectifs atteints, on construit véritablement une machine combinant les avantages du classificateur linéaire et du classificateur K-NN, tout en surmontant leurs inconvénients.

- **Classificateur linéaire**

Dans cette partie, on se concentre uniquement sur le problème de classification à deux classes, car tout problème de classification peut être étendu en une classification multi-catégories qui est basée sur l'un des algorithmes, *un contre tous* ou *un contre un*.
On suppose qu'il existe une frontière de décision entre chaque paire de classes, avec les paramètres de la fonction discriminante qui peuvent être estimés à l'aide d'échantillons d'apprentissage disponibles. Avec des données qui sont représentées sous la forme d'un vecteur de caractéristiques à n dimensions $x = (x_1, x_2, \ldots, x_n)$, alors la fonction discriminante linéaire est formulée comme suit [139] :

$$f(\mathrm{x}) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \tag{III.1}$$

Où x_i sont les variables et w_i sont les coefficients ou poids. Supposons $x_0 = 1$, et $f(x)$ peut être écrit comme :

$$f(\mathrm{x}) = \sum_{j=0}^{n} w_j x_j \tag{III.2}$$

Géométriquement, $f(\mathrm{x}) = 0$ est un hyperplan dans un espace de Hilbert à n dimensions $\Re^n$, sachant que $f(\mathrm{x}) = 0$ est la frontière de décision entre deux classes. Un échantillon de

données représenté par un vecteur de caractéristiques x est classé dans l'une des classes à l'aide du critère suivant :

$$\mathrm{x} \in \begin{cases} classe1 & si\ f(\mathrm{x}) > 0 \\ classe2 & si\ f(\mathrm{x}) < 0 \end{cases} \tag{III.3}$$

Il s'agit ensuite de trouver les paramètres ou l'ensemble des poids de $f(\mathrm{x})$: $w_0, w_1, w_2, \ldots. w_n$. Une manière classique de trouver les poids consiste à résoudre un ensemble d'équations linéaires à partir d'un ensemble d'échantillons d'apprentissage. Supposons que $d_i = 1$ représente la classe 1 et $d_i = -1$ représente la classe 2, et les $n+1$ échantillons d'apprentissage sont donnés par (x_i, d_i), où :

$$x_i = (x_{i1}, x_{i2}, ..., x_{in}), \quad i = 1, 2, ..., n+1 \tag{III.4}$$

Ensuite, en remplaçant (III.2) par chacune des données d'apprentissage x_i, l'ensemble des poids $w_i (i = 0, 1, \ldots, n)$ peut être résolu à l'aide des $n+1$ équations linéaires suivantes :

$$\sum_{j=0}^{n} w_j x_{ij} = d_i,\ i = 1, 2, ..., n+1 \tag{III.5}$$

L'équation (III.5) peut être écrite sous forme matricielle :

$$\mathrm{w}^T = \mathrm{X}^{-1}\mathrm{d}^T \tag{III.6}$$

Où w^T et d^T sont les transposées de w et d, et la solution de w est alors donnée par l'équation (III.6).

- **Classification du plus proche voisin**

L'algorithme du K voisin le plus proche ou K-NN [138] est un algorithme simple qui stocke tous les cas disponibles et classe les nouveaux cas en utilisant une fonction de décision (c'est-à-dire, une mesure de distance).

Étant donné un ensemble de données d'entraînement D et une mesure de distance $dist$:

- (x_i, y_i), $i = 1, 2, \ldots, N$.
- x_i est une donnée d'apprentissage dans $\Re^n$.
- y_i est la classe correspondante des données x_i, et $y_i \in \{c_j, j = 1, 2, \ldots, M\}$.
- dist $(x - x_i) = \|x - x_i\|$.

Une nouvelle donnée d'observation x est classée dans l'une des classes y_j à l'aide de l'algorithme suivant :

1. Saisissez les nouvelles données x.

2. Calculez la distance de x à tous les échantillons d'apprentissage x_i dans l'ensemble de données : $dist(x-x_i)$.

3. Trier $dist(x-x_i)$ ($i = 1, 2, \ldots, N$) par ordre croissant et classer tous les x_i en conséquence : $x_{r1}, x_{r2}, \ldots, x_{rk}, \ldots, x_{rN}$.

4. Pour une classification du plus proche voisin (NN), classez x en y_{r1}.

Pour une classification K-NN, classez x dans la classe majoritaire y_{rp} parmi les k premières données classées : $\{x_{r1}, x_{r2}, \ldots, x_{rk}\}$.

Contrairement à tous les autres classificateurs qui sont indépendants des données d'apprentissage d'origine une fois formés, le classificateur K-NN est lié aux données de formation. Cependant, l'un des principaux avantages d'un classificateur K-NN est qu'il peut classer des données qui ne sont pas linéairement séparables. C'est l'idée clé derrière les séparateurs à vaste marge (SVM) basée sur le noyau.
Pour formuler un séparateur à vaste marge, nous partirons de la forme primale d'un SVM, après la forme duale est introduite et enfin les SVMs basés sur les noyaux sont décrits en détail.

III.5.2. La forme primale du SVM

- **La marge entre deux classes**

Pour les SVMs, il faut chercher un hyperplan pour séparer les deux classes de données. Cependant, il existe un nombre infini d'hyperplans entre ces deux classes. L'hyperplan optimal ou le meilleur que nous percevons est celui séparant les deux classes avec la marge maximale [140] comme le montre la figure III.6. C'est l'hyperplan que nous trouvons.
Supposons que les deux sous-espaces correspondants aux deux classes de données sont, respectivement,

$$\begin{aligned} &< \mathrm{w} \mid \mathrm{x} > + b \geq 1 \quad pour \quad y_i = +1 \\ &< \mathrm{w} \mid \mathrm{x} > + b \leq -1 \quad pour \quad y_i = -1 \end{aligned} \tag{III.7}$$

Avec $< w \mid x >$: le produit scalaire entre w et x.

Les deux inégalités (III.7) peuvent être combinées en une seule :

$$y_i(< w \mid x > +b) - 1 \geq 0 \qquad \text{(III.8)}$$

Les frontières entre les deux sous-espaces sont respectivement, les hyperplans H_1 et H_2 :

$$\begin{aligned} H_1 &: \ < w \mid x > +b - 1 = 0 \\ H_2 &: \ < w \mid x > +b + 1 = 0 \end{aligned} \qquad \text{(III.9)}$$

Et l'hyperplan entre H_1 et H_2 est donné par H_0 :

$$H_0 : \ < w \mid \ x > +b = 0 \qquad \text{(III.10)}$$

Les deux classes de données et les trois hyperplans les séparant sont représentés sur la figure III.6.

Notre but n'est pas seulement de trouver H_0 mais aussi de maximiser la distance entre H_1 et H_2, qui est la marge entre les deux classes de données. Pour calculer la distance entre H_1 et H_2, on commence par exprimer l'équation de l'hyperplan séparateur comme suit :

$ax + by + c = 0$, où la constante c est appelée l'ordonnée à l'origine, et $|c|$ est associé à la distance de l'origine à la ligne.

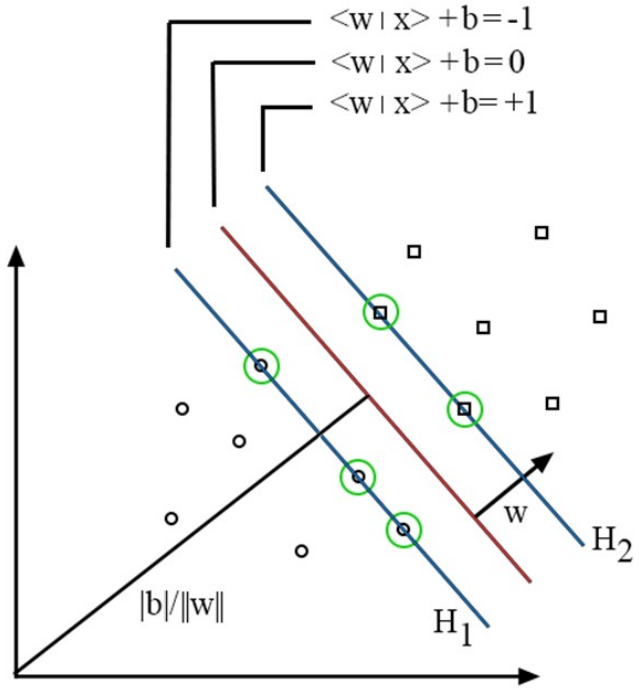


Figure III.6 : Deux classes de données et les hyperplans les séparant.

Ceci est également vrai dans l'espace de dimension supérieure. Par exemple, dans un espace 3D, un plan est donné par $Ax + By + Cz + D = 0$, et $|D|$ est associé à la distance de l'origine au plan, et ainsi de suite.

Par conséquent, la distance entre H_1 et H_2 est égale à la différence entre la distance de chacun d'eux à l'origine. Plus précisément, en se basant sur la théorie de la géométrie, la distance d'un point (x_0, y_0, z_0) à un plan dans l'espace 3D : $Ax + By + Cz + D = 0$, est donnée par :

$$\frac{\left|Ax_0 + By_0 + Cz_0 + D\right|}{\sqrt{A^2 + B^2 + C^2}} \tag{III.11}$$

Comme on peut également appliquer l'expression (III.11) aux dimensions supérieures, par conséquent la distance de H_0 à l'origine (0, 0, …, 0) dans un espace à n dimensions est donnée par :

$$\frac{|b|}{\|\mathrm{w}\|} \tag{III.12}$$

Où $\|\mathrm{w}\|$ est la longueur du vecteur w, et la distance entre H_1 et H_2 de l'origine (0, 0, …, 0) dans l'espace à n dimensions est donnée par les deux expressions suivantes :

$$\frac{|b+1|}{\|\mathrm{w}\|} \quad et \quad \frac{|b-1|}{\|\mathrm{w}\|} \tag{III.13}$$

Ainsi, en calculant la différence entre les deux expressions (III.13), la marge entre les deux hyperplans H_1 et H_2 s'obtient comme :

$$\frac{2}{\|\mathrm{w}\|} \tag{III.14}$$

Les points de données qui se trouvent sur H_1 et H_2 sont appelés vecteurs de support (marqués par des cercles sur la figure III.6), qui sont à la fois nécessaires et suffisants pour définir les hyperplans des frontières.

- **Maximisation de la marge**

Par conséquent, d'après l'équation (III.14), maximiser la distance entre les deux sous-espaces équivaut au problème d'optimisation suivant [141] :

$$Minimiser: f(\mathrm{w}) = \frac{1}{2}\|\mathrm{w}\|^2 = \frac{1}{2} < \mathrm{w} \mid \mathrm{w} > \quad \text{(III.15)}$$

$$\text{sous contraintes}: g_i(\mathrm{w}, b) = y_i(< \mathrm{w} \mid \mathrm{x}_i > + b) - 1 \geq 0, \quad i = 1, 2, ..., N \quad \text{(III.16)}$$

Ce qui nous mène à un problème d'optimisation. En outre, l'optimisation des équations (III.15) et (III.16) est équivalente à les combiner dans la fonction de Lagrange suivante et que l'on doit résoudre :

$$\nabla L = 0 \quad ou \quad \partial_{W,b} L = 0 :$$

$$L(\mathrm{w}, b, \alpha_i) = f(\mathrm{w}) - \sum_i \alpha_i g_i(\mathrm{w}, b) = \frac{1}{2}\mathrm{w}.\mathrm{w} - \sum_i \alpha_i y_i(\mathrm{w}.\mathrm{x}_i + b) + \sum_i \alpha_i \quad \text{(III.17)}$$

A partir de (III.17), on peut obtenir la forme primale du SVM :

$$\text{Minimiser}: L(\mathrm{w}, b, \alpha_i) = \frac{1}{2}\mathrm{w}.\mathrm{w} - \sum_i \alpha_i y_i(\mathrm{w}.\mathrm{x}_i + b) + \sum_i \alpha_i \quad \text{(III.18)}$$

$$\text{sous contraintes}: \alpha_i \geq 0, \; i = 1, 2, ..., N$$

III.5.3. La forme duale du SVM

La forme primale dans (III.18) permet de trouver les poids *w* et par conséquent un hyperplan qui sépare les deux classes de données avec le maximum de marge [142]. Cependant, comme nous devons optimiser deux ensembles de paramètres en même temps : *w* et α_i, c'est très couteux en termes de calcul comme approche d'optimisation. Avec la forme duale, nous allons la rendre plus efficace.

Puisque l'équation (III.18) est une fonction quadratique, le gradient ou les dérivées partielles de $L(\mathrm{w}, b, \alpha_i)$ sont égaux à 0.

Par conséquent, nous avons :

$$\nabla L(\mathrm{w}, b, \alpha_i) = \nabla f(\mathrm{w}) - \nabla[\sum_i \alpha_i g_i(\mathrm{w}, b)] = 0 \quad \text{(III.19)}$$

On déduit alors

$$\frac{\partial L}{\partial \mathrm{w}} = 0 \quad et \quad \frac{\partial L}{\partial b} = 0 \quad \text{(III.20)}$$

Cela mène à :

$$\mathrm{w} = \sum_i \alpha_i y_i \mathrm{x}_i \quad et \quad \sum_i \alpha_i y_i = 0 \tag{III.21}$$

La substitution de la forme primale (III.18) par (III.21) conduit à la forme duale suivante du SVM :

$$\text{Maximiser} : L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j < \mathrm{x}_i \,|\, \mathrm{x}_j > \tag{III.22}$$

$$\text{sous contraintes} : \ \alpha_i \geq 0 \ \ et \ \ \sum_i \alpha_i y_i = 0$$

En maximisant, L_D on obtient ces deux caractéristiques :

1. Faire apparaitre les vecteurs caractéristiques sur les frontières opposées des deux classes,
2. Ignorer les vecteurs caractéristiques éloignés des frontières des deux classes.

C'est exactement ce que nous voulons, car en termes de recherche des hyperplans séparant les deux classes avec une marge maximale, seuls les vecteurs sur ou proches des frontières entre les deux classes comptent le plus. Ces vecteurs caractéristiques sont appelés vecteurs de support et le classificateur défini par ces vecteurs de support est appelé une machine à vecteurs de support.

III.6. Noyaux

III.6.1. Principe du noyau

Même si un hyperplan est optimisé dans certain cas, il y aura toujours des exemples mal classés (figure III.7) [143]. Plutôt qu'une droite, la représentation idéale de la fonction de décision serait une représentation qui colle le mieux aux données d'entrainement (figure III.8). La détermination d'une telle fonction non linéaire est très difficile voire impossible. Pour cela les données sont amenées dans un espace où cette fonction devient linéaire (figure III.5), ce qui permet de garder les mêmes modèles de problèmes d'optimisation, utilisant les SVMs basés essentiellement sur le principe de séparation linéaire. Cette transformation d'espace est réalisée souvent à l'aide d'une fonction $F = \{\phi(x) \,|\, x \in X\}$ appelé fonction de correspondance (Mapping function) et le nouvel espace est appelé espace de caractéristiques (Feature space).

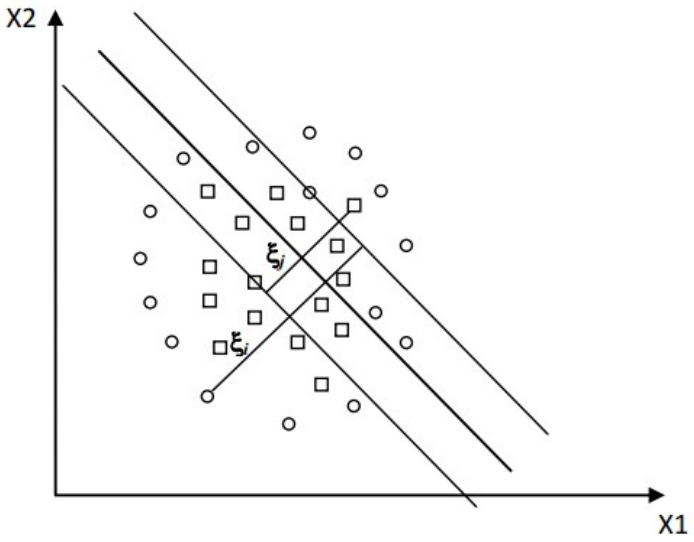


Figure III.7 : Non-adaptation de l'hyperplan aux problèmes réels de classification.

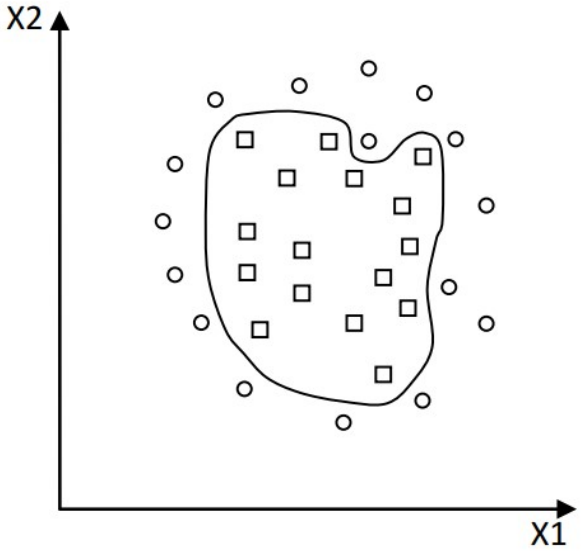


Figure III.8 : Représentation idéale de la fonction de décision.

Dans ce nouvel espace de caractéristiques, la fonction objective à optimiser devient :

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \phi(x_i) | \phi(x_j) \right\rangle \qquad \text{(III.23)}$$

Où $\left\langle \phi(x_i) | \phi(x_j) \right\rangle$ est le produit scalaire des deux images des vecteurs x_i et x_j dans le nouvel espace et dont le résultat est un scalaire.

Dans le calcul de l'optimum de la fonction (III.23), on utilise une fonction appelée noyau (Kernel). Au lieu de calculer $\varphi(x_i)$, $\varphi(x_j)$ et leur produit scalaire, on calcule plutôt une

fonction $K(x_i,x_j)$ qui représente à la fois les deux transformations (qui peuvent être inconnues) et leur produit scalaire. Cette fonction permet de surmonter le problème de détermination de la transformation φ et permet d'apprendre des relations non linéaires par des machines linéaires. La fonction $K(x_i,x_j)$ peut être vue comme une matrice $G[n, n]$ dite de Gram [144] qui représente les distances entre tous les exemples :

$$\begin{bmatrix} K(x_1,x_1) & \dots & K(x_1,x_n) \\ \dots & \dots & \dots \\ K(x_n,x_1) & \dots & K(x_n,x_n) \end{bmatrix}$$

Pour qu'une matrice G (fonction K) soit un noyau, il faut qu'elle respecte les conditions de Mercer [145], c'est-à-dire qu'elle doit être semi-définie positive (symétrique et n'a pas de valeurs propres négatives). La construction de tels noyaux est largement étudiée dans la littérature [146], néanmoins il existe certains noyaux qui sont très utilisés et qui sont considérés comme standards. Une fois le noyau choisi, la fonction objective (III.23) peut être calculée comme suit :

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{III.24}$$

Et la fonction de décision devient :

$$H(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x) + b \tag{III.25}$$

Où S représente l'ensemble des vecteurs supports.

III.6.2. Exemples de noyaux

- Noyau linéaire : Si les données sont linéairement séparables, on n'a pas besoin de changer d'espace, et le produit scalaire suffit pour définir la fonction de décision [147] :

$$K(x_i, x_j) = x_i^T x_j \tag{III.26}$$

- Noyau polynomial : Le noyau polynomial élève le produit scalaire à une puissance naturelle d :

$$K(x_i, x_j) = (x_i^T xj)^d \tag{III.27}$$

Si d = 1 le noyau devient linéaire. Le noyau polynomial dit non homogène $K(x_i, x_j) = (x_i^T x_j + C)^d$ est aussi utilisé.

- Noyau à fonctions de bases radiales : Les noyaux à fonctions de bases radiales (Radial Basis functions : RBF) sont des noyaux qui peuvent être écrits sous la forme : $K(x_i, x_j) = f(d(x_i, x_j))$ où d est une métrique sur X et f est une fonction dans l'espace $\Re^n$.

Un exemple de noyau RBF est le noyau Gaussien donné par :

$$K(x_i, x_j) = e^{(-\frac{\|x_i - x_j\|^2}{2\sigma^2})} \quad \text{(III.28)}$$

Où σ est un paramètre réel positif qui représente la largeur de bande du noyau.

III.7. Séparateurs à vaste marge binaires

Le cas le plus simple est celui où les données d'entrainement viennent uniquement de deux classes différentes (+1 ou -1), on parle alors de classification binaire. L'idée des SVMs est de rechercher un hyperplan (droite dans le cas de deux dimensions) qui sépare le mieux ces deux classes (figure III.9) [148].

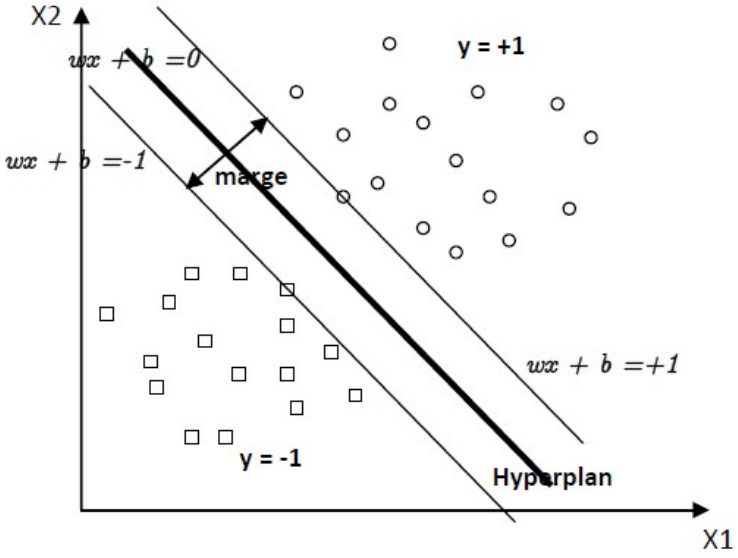


Figure III.9 : Séparateur à vaste marge binaire.

Si un tel hyperplan existe, c'est-à-dire si les données sont linéairement séparables, on parle d'une machine à vecteur de support à marge dure (hard margin).

III.7.1. Séparateur à marge dure

L'hyperplan séparateur est représenté par l'équation [149] :

$$H(x) = w^T x + b \tag{III.29}$$

Où *w* est un vecteur de *m* dimensions et *b* est une constante. La fonction de décision, pour un exemple *x*, peut être exprimée comme suit :

$$\begin{cases} Classe = 1 \quad si \;\; H(x) > 0 \\ Classe = -1 \; si \;\; H(x) < 0 \end{cases} \tag{III.30}$$

Puisque les deux classes sont linéairement séparables, il n'existe aucun exemple qui se situe sur l'hyperplan, c'est-à-dire qui satisfait $H(x) = 0$. Il convient alors d'utiliser la fonction de décisions suivante :

$$\begin{cases} Classe = 1 & si & H(x) > 1 \\ Classe = -1 & si & H(x) < -1 \end{cases} \tag{III.31}$$

Les valeurs +1 et -1 à droite des inégalités peuvent être des constantes quelconques $+a$ et $-a$, mais en divisant les deux parties des inégalités par a, on trouve les inégalités précédentes qui sont équivalentes à :

$$y_i(w^T x_i + b) \geq 1, \; i = 1...n \tag{III.32}$$

L'hyperplan $w^T x + b = 0$ représente un hyperplan séparateur des deux classes, et la distance entre cet hyperplan et l'exemple le plus proche s'appelle la marge (figure III.9). La région qui se trouve entre les deux hyperplans $w^T x + b = -1$ et $w^T x + b = +1$ est appelée la région de généralisation de la machine d'apprentissage. Plus cette région est importante, plus est la capacité de généralisation de la machine. La maximisation de cette région est l'objectif de la phase d'entrainement qui consiste, pour la méthode SVM, à rechercher l'hyperplan qui maximise la région de généralisation c'est-à-dire la marge. Un tel hyperplan est appelé "hyperplan de séparation optimale". En supposant que les données d'apprentissage ne contiennent pas des données bruitées (mal-étiquetées) et que les données de test suivent la même probabilité que celle des données d'entraînement, l'hyperplan de marge maximale va certainement maximiser la capacité de généralisation de la machine d'apprentissage.

La détermination de l'hyperplan optimal passe par la détermination de la distance euclidienne minimale entre l'hyperplan et l'exemple le plus proche des deux classes. Puisque le vecteur *w* est orthogonal sur l'hyperplan séparateur, la droite parallèle à *w* et reliant un exemple *x* à l'hyperplan est donnée par :

$$\frac{a\,w}{\|w\|} + x = 0 \tag{III.33}$$

Où *a* représente la distance entre *x* et l'hyperplan. La résolution de (III.33), donne :

$$a = -\frac{w^T x + b}{\|w\|} \tag{III.34}$$

La distance de tout exemple de l'hyperplan doit être supérieure ou égale à la marge δ :

$$\frac{y_i(w^T x_i + b)}{\|w\|} \geq \delta \tag{III.35}$$

Si une paire (*w, b*) est une solution alors (*aw, ab*) est une solution aussi où *a* est un scalaire. On impose alors la contrainte suivante :

$$\|w\|\,\delta \geq 1 \tag{III.36}$$

Pour trouver l'hyperplan séparateur qui maximise la marge, on doit déterminer, à partir des deux dernières inégalités, le vecteur *w* qui possède la norme euclidienne minimale et qui vérifie la contrainte de l'équation (III.32), de bonne classification des exemples d'entrainement. L'hyperplan séparateur optimal peut être obtenu en résolvant le problème :

$$\begin{cases} \text{Minimiser} & \frac{1}{2}\|w\|^2 \\ \text{sous contraintes} & \\ y_i(w^T x_i + b) \geq 1 & \forall i = 1...n \end{cases} \tag{III.37}$$

Remarquons que nous pouvons obtenir le même hyperplan même en supprimant toutes les données qui vérifient l'inégalité de la contrainte. Les données qui vérifient l'égalité de la contrainte s'appellent les vecteurs supports, et ce sont ces données seules qui contribuent à la détermination de l'hyperplan. Dans la figure III.9, les données qui se trouvent sur les deux droites +1 et -1 représentent les vecteurs de support.

Le problème de l'équation (III.37) est un problème de programmation quadratique avec contraintes linéaires. Dans ce problème, les variables sont *w* et *b*, c'est-à-dire que le nombre de variables est égal à *m* + 1. Généralement, le nombre de variables est important ce qui ne permet pas d'utiliser les techniques classiques de programmation quadratique. Dans ce cas, l'équation (III.37) est convertie en un problème dual équivalent et sans contraintes en introduisant les multiplicateurs de Lagrange :

$$Q(w,b,\alpha)=\frac{1}{2}w^T w-\sum_{i=1}^{n}\alpha_i\left\{y_i(w^T x_i+b)-1\right\} \tag{III.38}$$

Où les α_i sont les multiplicateurs non négatifs de Lagrange. L'optimum de la fonction objective *Q* peut être obtenu en la minimisant par rapport à *w* et *b* et en la maximisant par rapport aux α_i. À l'optimum de la fonction objective *Q*, ses dérivées par rapports aux variables *w* et *b* s'annulent ainsi que le produit des α_i aux contraintes (équation (III.39)) :

$$\begin{cases} \dfrac{\partial Q(w,b,\alpha)}{\partial w}=0 & (a)\\ \dfrac{\partial Q(w,b,\alpha)}{\partial b}=0 & (b)\\ \alpha_i\left\{y_i(w^T x_i+b)-1\right\}=0 & (c)\\ \alpha_i\geq 0 & (d) \end{cases} \tag{III.39}$$

De (III.39-a), on déduit :

$$\begin{cases} w=\sum_{i=1}^{n}\alpha_i y_i x_i\\ \sum_{i=1}^{n}\alpha_i y_i=0 \end{cases} \tag{III.40}$$

En remplaçant (III.40) dans (III.38), on obtient le problème dual à maximiser suivant :

$$\begin{cases} \text{Maximiser } Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j\\ \text{sous contrainte}\\ \sum_{i=1}^{n}\alpha_i y_i=0\\ \alpha_i\geq 0 \end{cases} \tag{III.41}$$

Si le problème de classification est linéairement séparable, une solution optimale pour les α_i existe. Les exemples ayant des $\alpha_i \neq 0$ représentent les vecteurs de support appartenant aux deux classes. La fonction de décision est donnée par :

$$H(x) = \sum_{S} \alpha_i y_i x^T x_i + b \tag{III.42}$$

Où S représente l'ensemble des vecteurs supports. Le paramètre b peut être calculé à partir de n'importe quel vecteur support par :

$$b = y_i - w^T x_i \tag{III.43}$$

D'un point de vue précision, on prend la moyenne de b pour tous les vecteurs de support :

$$b = \frac{1}{|S|} \sum_{i \in S} y_i - w^T x_i \tag{III.44}$$

La fonction de décision H peut être calculée, donc pour chaque nouvel exemple x par l'équation (III.42) et la décision peut être prise comme suit :

$$\begin{cases} x \in \text{classe} +1 & si \quad H(x) > 0 \\ x \in \text{classe} -1 & si \quad H(x) < 0 \\ x \text{ inclassifiable} & si \quad H(x) = 0 \end{cases} \tag{III.45}$$

La zone -1 $< H(x) <$ 1 est appelée la zone de généralisation.
Si on prend un exemple x_k de l'ensemble d'entrainement appartenant à la classe y_k et on calcule sa fonction de décision H(x_k), on peut se trouver dans l'un des cas suivants :

1. $y_k * \mathrm{H}(x_k) > 1$: dans ce cas l'exemple est bien classé et ne se situe pas dans la zone de la marge. Il ne représente pas un vecteur support.

2. $y_k * \mathrm{H}(x_k) = 1$: dans ce cas l'exemple est bien classé et se situe aux frontières de la zone de la marge. Il représente un vecteur support.

3. $0 < y_k * \mathrm{H}(x_k) < 1$: dans ce cas l'exemple est bien classé et se situe dans de la zone de la marge. Il ne représente pas un vecteur support.

4. $y_k * \mathrm{H}(x_k) < 0$: dans ce cas l'exemple se situe dans le mauvais côté, il est mal classé et ne représente pas un vecteur support.

III.7.2. Séparateurs à marge souple

En réalité, un hyperplan séparateur n'existe pas toujours, et même s'il existe, il ne représente pas généralement la meilleure solution pour la classification. En plus une erreur d'étiquetage dans les données d'entrainement (un exemple étiqueté +1 au lieu de -1 par exemple) affectera crucialement l'hyperplan [150].

Dans le cas où les données ne sont pas linéairement séparables, ou contiennent du bruit (valeurs aberrantes, données mal étiquetées), les contraintes de l'équation (III.32) ne peuvent être vérifiées, et il y a nécessité de les relaxer légèrement. Ceci peut être réalisé en admettant une certaine erreur de classification des données (figure III.10) ce qui est appelé SVM à marge souple (soft margin).

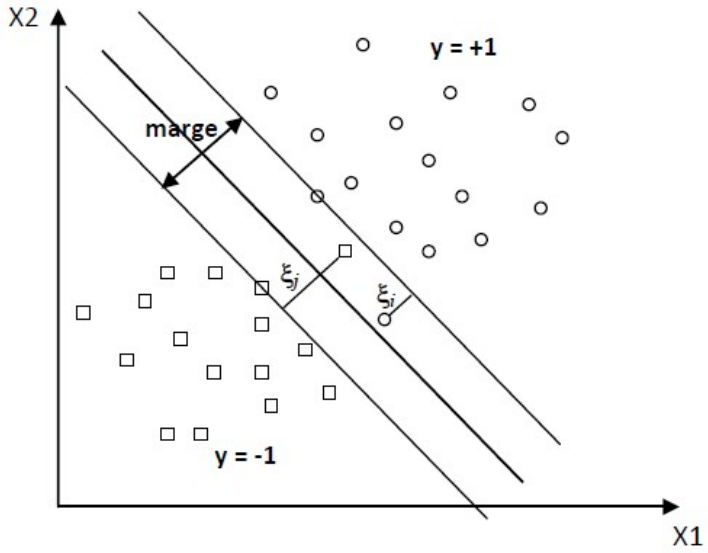


Figure III.10 : Séparateur binaire à marge souple.

On introduit alors sur les contraintes des variables ξ_i dites de relaxation pour obtenir la contrainte donnée par :

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad i = 1...n \tag{III.46}$$

Grâce aux variables de relaxation non négatives ξ_i, un hyperplan séparateur existera toujours.

Si $\xi_i < 1$, x_i ne respecte pas la marge mais reste bien classé, sinon x_i est mal classé par l'hyperplan. Dans ce cas, au lieu de rechercher uniquement un hyperplan séparateur qui maximise la marge, on recherche un hyperplan qui minimise aussi la somme des erreurs permises c'est-à-dire minimiser $Q(w) = \sum_{i=1}^{n} \xi_i$.

En combinant avec l'équation (III.37), on obtient le problème primal suivant :

$$\begin{cases} \textit{Minimiser} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ \text{sous contraintes} \\ y_i(w^T x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \tag{III.47}$$

Où C est un paramètre positif libre (mais fixe) qui représente une balance entre les deux termes de la fonction objective (la marge et les erreurs permises), c'est-à-dire entre la maximisation de la marge et la minimisation de l'erreur de classification. On obtient le problème dual en introduisant les multiplicateurs de Lagrange α_i et β_i :

$$Q(w,b,\alpha,\xi,\beta) = \frac{1}{2}w^T w + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i y_i(w^T x_i + b) - 1 + \xi_i - \sum_{i=1}^{n}\beta_i\xi_i \tag{III.48}$$

À la solution optimale, les dérivées par rapport aux variables w, b, α, β s'annulent ainsi que le produit des contraintes aux multiplicateurs. Les conditions suivantes sont alors vérifiées:

$$\begin{cases} \dfrac{\partial Q(w,b,\xi,\alpha,\beta)}{\partial w} = 0 & (a) \\ \dfrac{\partial Q(w,b,\xi,\alpha,\beta)}{\partial b} = 0 & (b) \\ \alpha_i\left\{y_i(w^T x_i + b) - 1 + \xi_i\right\} = 0 & (c) \\ \beta_i\xi_i = 0 & (d) \\ \alpha_i \geq 0;\ \beta_i \geq 0;\ \xi_i \geq 0 & (e) \end{cases} \tag{III.49}$$

On déduit que :

$$\begin{cases} w = \sum_{i=1}^{n}\alpha_i y_i x_i \\ \sum_{i=1}^{n}\alpha_i y_i = 0 \\ \alpha_i + \beta_i = 0 \end{cases} \tag{III.50}$$

En remplaçant (III.49) dans (III.48), on obtient le problème dual suivant :

$$\begin{cases} \textit{Maximiser} \quad Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{sous contraintes} \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{cases} \tag{III.51}$$

La seule différence avec le SVM à marge dure est que les α_i ne peuvent pas dépasser C, ils peuvent être dans l'un des trois cas suivants :

1. $\alpha_i = 0 \Rightarrow \beta_i = C \Rightarrow \xi_i = 0$: x_i est bien classé,

2. $0 < \alpha_i < C \Rightarrow \beta_i > 0 \Rightarrow \xi_i = 0 \Rightarrow y_i(w^T x_i + b) = 1$: x_i est un vecteur support et est appelé dans ce cas vecteur support non borné (unbounded),

3. $\alpha_i = C \Rightarrow \beta_i = 0 \Rightarrow \xi_i \ge 0 \Rightarrow y_i(w^T x_i + b) = 1 - \xi_i$: x_i est un vecteur support appelé dans ce cas vecteur support borné (bounded). Si $0 \le \xi_i < 1$, x_i est bien classé, sinon x_i est mal classé.

Ces conditions sur les α_i sont appelées les conditions de Karush-Kuhn-Tucker (KKT) [151, 152], elles sont très utilisées par les algorithmes d'optimisation pour rechercher les α_i optimums et par conséquent l'hyperplan optimal.

La fonction de décision est alors calculée de la même manière que dans le cas des SVMs à marge dure mais uniquement à base des vecteurs supports non bornés par :

$$H(x) = \sum_{i \in U} \alpha_i y_i x_i^T x + b \tag{III.52}$$

Pour les vecteurs supports non bornés, nous avons :

$$b = y_i - w^T x_i \tag{III.53}$$

Pour garantir une bonne précision, on prend la moyenne de b pour tous les vecteurs supports non bornés :

$$b = \frac{1}{|U|} \sum_{i \in U} y_i - w^T x_i \tag{III.54}$$

III.8. Séparateurs à vaste marge multi-classes

Les séparateurs à vaste marge sont dans leurs origines binaires. Cependant, les problèmes du monde réel sont dans la plupart des cas multi-classes, l'exemple le plus simple en est la

reconnaissance des caractères optiques. Dans de tels cas, on ne cherche pas à affecter un nouvel exemple à l'une des deux classes mais à l'une parmi plusieurs, c'est-à-dire que la décision n'est plus binaire et un seul hyperplan ne suffit plus.

Les méthodes des machines à vecteur support multi-classes, réduisent le problème multi-classes à une composition de plusieurs hyperplans bi-classes permettant de tracer les frontières de décision entre les différentes classes [153, 154]. Ces méthodes décomposent l'ensemble d'exemples en plusieurs sous-ensembles représentant chacun un problème de classification binaire. Pour chaque problème un hyperplan de séparation est déterminé par la méthode SVM binaire. On construit lors de la classification une hiérarchie des hyperplans binaires qui est parcourue de la racine jusqu'à une feuille pour décider de la classe d'un nouvel exemple. On trouve dans la littérature plusieurs méthodes de décomposition :

III.8.1. Un contre tous (one against all)

C'est la méthode la plus simple et la plus ancienne. Selon la formulation de Vapnik [155], elle consiste à déterminer pour chaque classe k un hyperplan $H_k(w_k,b_k)$ la séparant de toutes les autres classes. Cette classe k est considérée comme étant la classe positive (+1) et les autres classes comme étant la classe négative (-1), ce qui résulte, pour un problème à K classes, en K SVM binaires. Un hyperplan H_k est défini pour chaque classe k par la fonction de décision suivante :

$$\begin{aligned} H_k(x) &= signe(\langle w_k, x\rangle + b_k) \\ &= \begin{cases} +1 \;\; si \;\; f_k(x) > 0; \\ 0 \quad \text{sinon} \end{cases} \end{aligned} \tag{III.55}$$

La valeur retournée de l'hyperplan permet de savoir si x appartient à la classe k ou non. Dans le cas où il n'appartient pas à k ($H_k(x) = 0$), nous n'avons aucune information sur l'appartenance de x aux autres classes. Pour le savoir, on présente x à tous les hyperplans, ce qui donne la fonction de décision suivante :

$$k^* = \underbrace{Arg}_{(1\le k\le K)} \; Max(H_k(x)) \tag{III.56}$$

Si une seule valeur $H_k(x)$ est égale à 1 et toutes les autres sont égales à 0, on conclut que x appartient à la classe k. Le problème est que l'équation (III.56) peut être vérifiée pour plus

d'une classe, ce qui produit des régions d'ambiguïté, et l'exemple *x* est dit non classifiable. La figure (III.11) représente un cas de séparation de 3 classes.

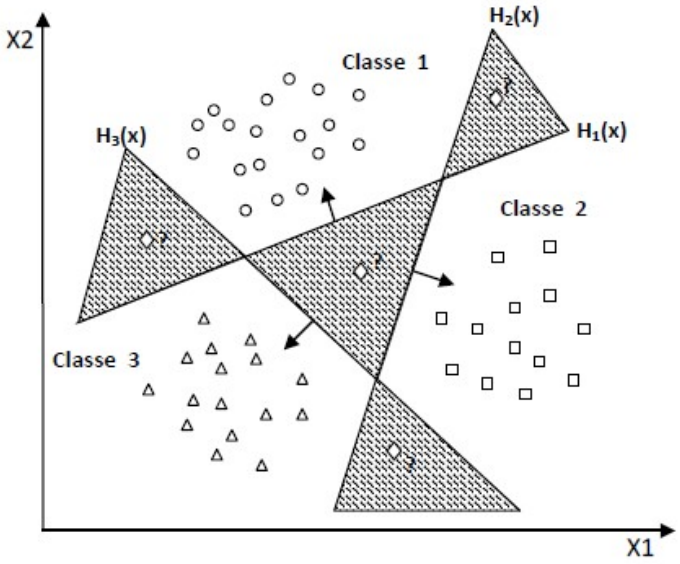


Figure III.11 : Approche un-contre-tous avec des zones d'indécision.

Pour surmonter cette situation, la méthode *un contre tous* utilise le principe du "gagnant prend tout" (winner takes all), comme c'est représenté dans la figure III.12 : la classe *k* retenue est celle qui maximise $f_k(x) = \langle wk, x \rangle + b_k$

$$k^* = \underbrace{Arg}_{(1 \le k \le K)} \; Max(\langle w_k, x \rangle + b_k) \qquad \text{(III.57)}$$

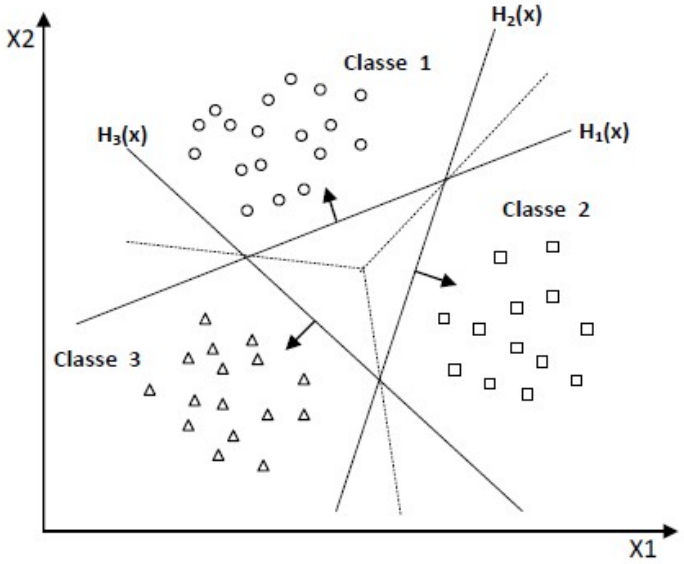


Figure III.12 : Résolution des cas d'indécision dans la méthode *un contre tous*.

La méthode *un contre tous* peut être utilisée pour découvrir même les cas de rejet où un exemple n'appartient à aucune des *K* classes. Pour cela, on prend les deux fonctions de décision les plus élevées, puis on calcule leur différence, si elle est supérieure à un certain seuil, l'exemple est rejeté.

Souvent, la méthode *un contre tous* est critiquée à cause de son asymétrie [156], puisque chaque hyperplan est entrainé sur un nombre d'exemples négatifs beaucoup plus important que le nombre d'exemples positifs. Par exemple dans le cas de reconnaissance de caractères optiques, le classificateur du caractère 'A' est entrainé sur des exemples positifs représentant 'A' et des exemples négatifs représentant tous les autres caractères. La méthode *un contre un* suivante est une méthode symétrique qui corrige ce problème.

III.8.2. Un contre un (one against one)

Cette méthode, appelée aussi "pairwise", revient à Knerr et ses co-auteurs [157] qui l'ont proposée pour les réseaux de neurones. Elle consiste à utiliser un classificateur pour chaque paire de classes. Au lieu d'apprendre *K* fonctions de décisions, la méthode *un contre un* discrimine chaque classe de chaque autre classe, ainsi $K(K - 1)/2$ fonctions de décisions sont apprises.

Pour chaque paire de classes (*k, s*), la méthode *un contre un* définit une fonction de décision binaire $h_{ks} : \Re \rightarrow \{-1, +1\}$. L'affectation d'un nouvel exemple se fait par liste de vote. On teste un exemple par le calcul de sa fonction de décision pour chaque hyperplan. Pour chaque test, on vote pour la classe à laquelle appartient l'exemple (classe gagnante). On définit pour le faire la fonction de décision binaire $H_{ks}(x)$ donnée par :

$$\begin{aligned} H_{ks}(x) &= \text{signe}(f_{ks}(x)) \\ &= \begin{cases} +1 & si \quad f_{ks}(x) > 0; \\ 0 & \text{sinon} \end{cases} \end{aligned} \tag{III.58}$$

Sur la base des $K(K - 1)/2$ fonctions de décision binaires, on définit *K* autres fonctions de décision :

$$H_k(x) = \sum_{s=1}^{m} H_{ks}(x) \tag{III.59}$$

Un nouvel exemple est affecté à la classe la plus votée. La règle de classification d'un nouvel exemple *x* est donnée par :

$$k^* = \underbrace{Arg}_{(1 \le k \le K)} \ (Max\, H_k(x)) \qquad \text{(III.60)}$$

Malheureusement, la fonction (III.60) peut être vérifiée pour plusieurs classes, ce qui produit des zones d'indécisions. La méthode de vote affecte dans ce cas, un exemple aléatoirement à l'une des classes les plus votées.

La Figure III.13 montre un exemple de classification de trois classes avec la zone d'indécision.

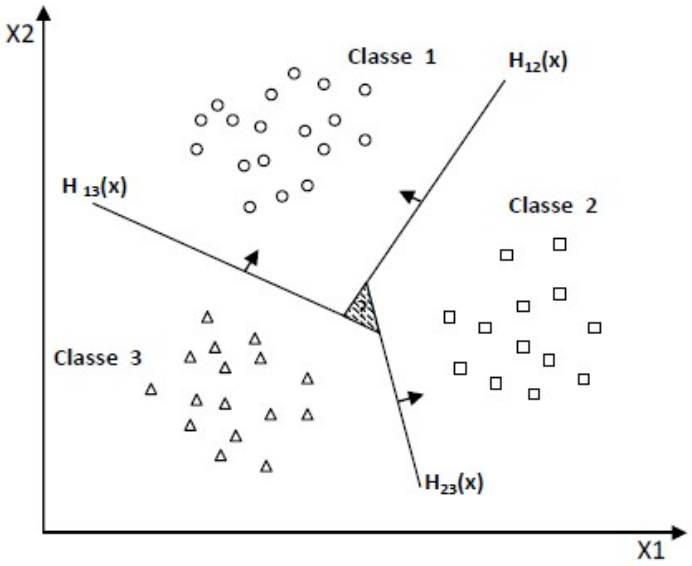


Figure III.13 : Approche de classification un-contre-un.

Bien que La méthode *un contre un* utilise pour l'entrainement, un nombre plus important d'hyperplans que la méthode *un contre tous*, elle est souvent plus rapide. Cela est dû, d'une part au nombre limité d'exemples utilisés pour entrainer chaque hyperplan, et d'autre part à la simplicité des problèmes à résoudre. En effet, chaque deux classes prises à part sont moins chevauchées que toutes les classes.

III.9. Séparateur à vaste marge mono-classe (Novelty detection)

Dans les machines à vecteurs de support binaires et multi-classes précédentes, nous avons toujours des exemples positifs et d'autres négatifs, c'est-à-dire des exemples et des contre-exemples. De telles informations ne sont pas disponibles dans tous les cas pratiques. Parfois, il est très coûteux, voire impossible, de trouver des contre-exemples qui représentent réellement la classe négative.

Prenons l'exemple de reconnaissance d'une catégorie particulière de pièces par un robot dans une usine, il est facile d'avoir des exemples suffisants de cette pièce, mais il est

difficile d'avoir des exemples de toutes les pièces différentes. Il est souhaitable, dans de tels cas, d'avoir un modèle de décision permettant de reconnaître autant d'exemples possibles de cette catégorie et de rejeter tous les autres. Ce problème est souvent appelé "Novelty detection" ou détection des nouveautés, puisque le modèle de décision connaît un ensemble d'exemples et détecte tous ce qui est nouveau (étranger ou outlier).

Pour la classification SVM mono-classe, il est supposé que seules les données de la classe cible sont disponibles. L'objectif est de trouver une frontière qui sépare les exemples de la classe cible du reste de l'espace, autrement dit, une frontière autour de la classe cible qui accepte autant d'exemples cibles que possible [158]. Cette frontière est représentée par une fonction de décision positive à l'intérieur de la classe et négative en dehors. La figure III.14 représente, en deux dimensions, un cas de séparation d'une classe de toute autre classe.

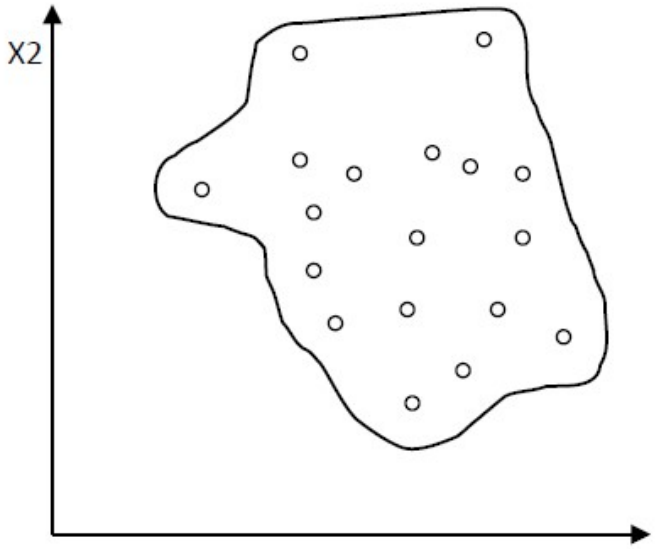


Figure III.14 : Séparation des exemples d'une classe du reste de l'espace.

Pour résoudre ce cas de problèmes, la technique SVM mono-classe utilise le même modèle binaire décrit dans la section III.6 avec une astuce en plus; l'origine de l'espace est considérée comme étant la seule instance de la classe négative. Le problème revient donc à trouver un hyperplan qui sépare les exemples de la classe cible de l'origine et qui maximise la marge entre les deux (figure III.15).

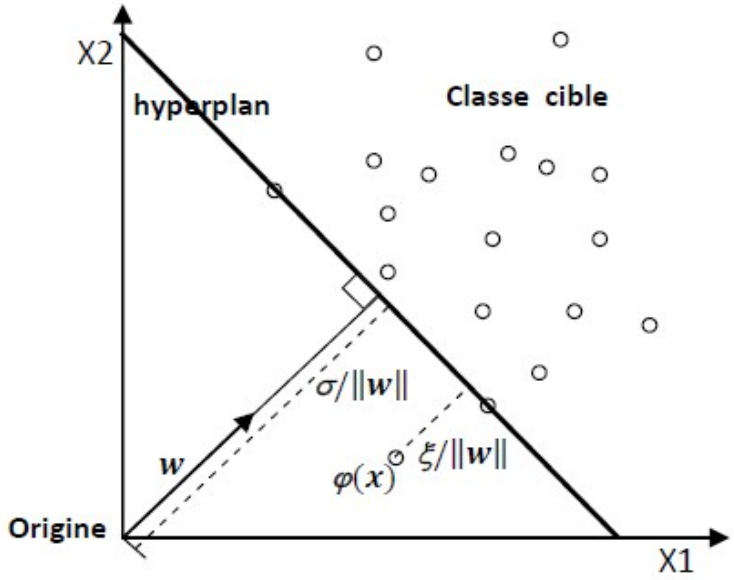


Figure III.15 : Séparateur à vaste marge mono-classe à marge maximale.

Le problème est modélisé par le problème primal de programmation quadratique de l'équation (III.61) dont l'objectif est de maximiser la marge et minimiser les erreurs de classification. La contrainte est la bonne classification des exemples d'entrainement.

$$\begin{cases} \min_{w,\xi,\rho} \dfrac{1}{2}\|w\|^2 + \dfrac{1}{\nu N}\sum_{i=1}^{l}\xi_i - \rho \\ \langle w, \phi(x_i)\rangle \geq \rho - \xi_i \\ \xi_i \geq 0 \quad i = 1,2,...,N \end{cases} \tag{III.61}$$

Où N est le nombre d'exemples de la classe cible, (w,ρ) sont les paramètres permettant de localiser l'hyperplan, ξ_i représentent les erreurs permises sur les exemples, pénalisées par le paramètre ν et ϕ est une transformation d'espace semblable à celle du cas binaire.

Une fois (w,ρ) déterminés, tout nouvel exemple pourra être classé par la fonction de décision suivante:

$$f(x) = \langle w, \phi(x_i)\rangle - \rho \tag{III.62}$$

x appartient à la classe cible si $f(x)$ est positive.

En fait, la résolution du problème de l'équation (III.61) est réalisée par l'introduction des multiplicateurs de Lagrange pour obtenir le problème dual suivant :

$$\begin{cases} \text{Minimiser}_{\alpha} \quad \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j K(x_i,x_j) \\ \text{sous contraintes} \\ \sum_{i=1}^{n}\alpha_i = 1 \\ 0 \le \alpha_i \le \frac{1}{\nu N} \end{cases} \tag{III.63}$$

Où K est un noyau qui représente la transformation d'espace ϕ.

Une fois les α_i déterminés ils peuvent être dans l'un des trois cas suivants :

- $\alpha_i = 0$: correspondent aux exemples bien classés, c'est-à-dire qui se situent au-dessus de l'hyperplan,
- $\alpha_i = \frac{1}{\nu N}$: correspondent aux exemples qui se situent à l'intérieur de la marge (au-dessous de l'hyperplan),
- $0 < \alpha_i < \frac{1}{\nu N}$: correspondent aux exemples vecteurs support qui se situent sur l'hyperplan.

La fonction de décision pour tout exemple x est donnée par :

$$f(x) = \sum_{i=1}^{l}\alpha_i K(x_i,x) - \rho \tag{III.64}$$

Où ρ peut être déterminé à partir d'un exemple x_i d'apprentissage dont $\alpha_i \neq 0$ par :

$$\rho = \sum_{j}\alpha_j K(x_j,x_i) \tag{III.65}$$

Cette méthode de résolution n'est pas la seule. Dans [159], David Tax et Robert Duin ont proposé de résoudre le problème par la recherche de l'hyper-sphère de rayon minimal qui englobe tous les exemples d'apprentissage (figure III.16).

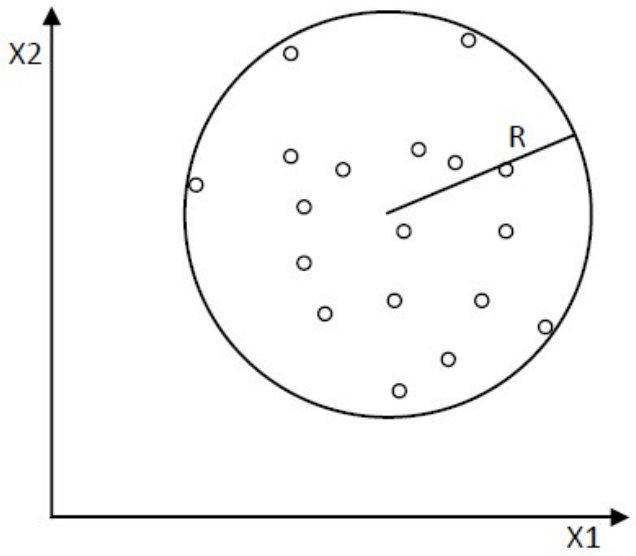


Figure III.16 : SVM mono-classe à base d'hypersphère.

Le problème revient à minimiser R tout en gardant les exemples dans l'hyper-sphère :

$$\begin{cases} \text{Minimiser}_{R,\xi} \quad \|R\|^2 + \dfrac{1}{\nu N}\sum_{i=1}^{l}\xi_i; \quad 0 < \nu \le 1 \\ \text{sous contraintes} \\ \qquad \|\phi(x_i) - c\|^2 \le R^2 + \xi_i \\ \qquad \xi_i \le 0 \quad i = 1, 2, ..., N \end{cases} \tag{III.66}$$

Où R est le rayon de l'hyper-sphère, c est le centre des exemples, ξ_i sont les erreurs permises pénalisées par le paramètre νet ϕ est une transformation d'espace. Le problème est résolu en introduisant les multiplicateurs de Lagrange α_i et en le transformant vers le problème dual donné par :

$$\begin{cases} \textit{Minimiser}_{\alpha} \quad \sum_{i,j}\alpha_i\alpha_j K(x_i, x_j) - \sum_i \alpha_i K(x_i, x_i) \\ \text{sous contraintes} \\ \qquad \sum_{i=1}^{n}\alpha_i = 1 \\ \qquad 0 \le \alpha_i \le \dfrac{1}{\nu N} \end{cases} \tag{III.67}$$

Après la détermination des α optimaux, la fonction de décision pourra être calculée pour un nouvel exemple x par la fonction :

$$f(x) = R^2 - \sum_{i,j}\alpha_i\alpha_j K(x_i, x_j) + 2\sum_i \alpha_i K(x_i, x) - K(x, x) \tag{III.68}$$

Cette méthode est moins précise que la précédente à cause de la forme sphérique de la frontière de la décision qui limite les capacités de généralisation de la machine apprise.

III.10. Avantages et inconvénients des SVMs

III.10.1. Avantages des SVMs

Les séparateurs à vaste marge présentent des avantages multiples dont on se contente de recenser quelques uns :

- Leur capacité à manipuler de grandes quantités de données,
- Elles ont une base théorique solide [155],
- Le faible nombre d'hyper paramètres utilisés par ces méthodes,
- Elles sont bien fondées théoriquement,
- Les résultats pertinents qu'on obtient avec les SVMs en pratique,
- Différentes fonctions noyaux peuvent être spécifiées [146].

III.10.2. Inconvénients des SVMs

Malgré leurs performances, les séparateurs à vaste marge présentent aussi des faiblesses, notamment celles-ci :

- Elles utilisent des fonctions mathématiques complexes pour la classification [146],
- Besoin d'une bonne fonction noyau pour avoir de bons résultats de classification dans le cas d'un séparateur non linéaire,
- Besoin d'un temps considérable pour l'apprentissage des données de grandes dimensions.

III.11. Conclusion

Dans ce chapitre, on a présenté le principe de fonctionnement général des séparateurs à vaste marge et leur fondement mathématique. On a considéré les deux cas des données linéairement séparable et non linéairement séparable. Ce dernier cas nécessite l'utilisation d'une marge souple et une fonction noyau pour trouver un hyperplan séparateur optimal.
Une synthèse des différentes approches des SVMs binaires et multi-classes a été présentée. Les SVMs multi-classes, utilisent deux techniques principales qui sont : un-contre-un et un-contre-tous. Les séparateurs à vaste marge sont des classificateurs très utilisés et très prometteurs vu leurs utilisations répondues dans de nombreux domaines tels que la bio-informatique, la recherche d'information, la vision par ordinateur, les finances, etc.

Chapitre 4 :
Les modèles de Markov cachés

IV.1. Introduction

Les modèles de Markov cachés (HMMs : hidden Markov models) sont un outil statistique permettant de modéliser les phénomènes stochastiques. Ces modèles sont utilisés dans de nombreux domaines [160] tels que la reconnaissance et la synthèse de la parole, la biologie, l'ordonnancement, l'indexation de documents, la reconnaissance d'image, la prédiction des séries temporelles, etc. Récemment, les modèles de Markov cachés ont même été utilisés pour l'ordonnancement des tâches [161] et les technologies de l'information [162,163].

Pour pouvoir utiliser les HMMs, il est nécessaire de connaitre les règles théoriques. Par conséquent, ce chapitre est consacré aux fondements mathématiques, notations utiles et principaux algorithmes qui constituent la théorie des modèles de Markov cachés comme les algorithmes avant, arrière, Baum-Welch et Viterbi.

IV.2. Généralités sur les modèles de Markov cachés

Les modèles de Markov cachés ont été introduits en 1966 par Baum et Petrie [164] sous le nom de « *probabilistic functions of Markov chains* ». En 1970, les termes « *modèles de Markov cachés* » ou « *chaînes de Markov cachées* » (*hidden Markov models* en anglais) ont été inventés par Neuwirth afin de remplacer l'appellation « *fonction probabiliste d'une chaine de Markov* ».

Les premiers travaux sur les HMMs pour la reconnaissance automatique de la parole ont été menés en parallèle en 1975 par le groupe IBM composé de Bahl et Jelinek [165] et par Baker au CMU [166, 167]. Ces travaux ont permis de découvrir les capacités des modèles de Markov cachés pour la reconnaissance de la parole.

Dans les années 1980, les modèles de Markov cachés incorporant les réseaux de neurones apparaissent [168]. Depuis lors, ces nouveaux modèles ont été très largement utilisés pour la reconnaissance de mots isolés [169-171], la reconnaissance de mots enchainés [172], la reconnaissance de la parole continue [173] et pour la localisation de mots dans une phrase [174, 175].

À partir des années 1990 les premières applications à la reconnaissance d'images [176] et de l'écriture apparaissent [177, 178].

IV.2.1. Modèles de Markov observables

Avant de décrire les HMMs proprement dits, nous présentons un modèle probabiliste plus simple pour l'observation de séquences : les modèles de Markov observables [179].

D'une manière générale, un processus ou modèle stochastique observable est un processus aléatoire qui peut changer d'état s_i, $i = 1,\ldots, n$ au hasard, aux instants $t = 1,2,\ldots, T$. Le résultat observé est la suite des états dans lesquels il est passé. On peut aussi dire de ce processus qu'il émet des séquences d'états $S = s_1, s_2,\ldots, s_T$. Chaque séquence est émise avec une probabilité $P(S) = P(s_1, s_2,\ldots, s_T)$. Pour calculer $P(S)$, il faut se donner la probabilité initiale $P(s_1)$ et les probabilités d'être dans un état s_t, connaissant l'évolution antérieure.

Un processus stochastique est markovien (ou *de Markov*) si son évolution est entièrement déterminée par une probabilité initiale et des probabilités de transitions entre états. Autrement dit, en notant $(q_t = s_i)$ le fait que l'état observé à l'instant t est s_i.

$$\forall t,\ \mathrm{P}(q_t = s_i \mid q_{t-1} = s_j,\ q_{t-2} = s_k \ldots) = \mathrm{P}(q_t = s_i \mid q_{t-1} = s_j) \tag{IV.1}$$

D'où :

$$\mathrm{P}(q_1 \ldots q_T) = \mathrm{P}(q_1) \times \mathrm{P}(q_2 \mid q_1) \times \ldots \times \mathrm{P}(q_T \mid q_{T-1}) \tag{IV.2}$$

Nous supposons pour simplifier, que les processus de Markov auxquels nous avons affaire sont stationnaires, c'est-à-dire que leurs probabilités de transition ne varient pas dans le temps. Cela autorise à définir une matrice de probabilité de transitions $A = [a_{ij}]$ telle que :

$$a_{ij} = \mathrm{P}(q_t = s_j \mid q_{t-1} = s_i),\ 1 \le i \le n,\ 1 \le j \le n \tag{IV.3}$$

Avec :

$$\forall i, j \quad a_{ij} \ge 0,\ \forall i \ \sum_{j=1}^{j=n} a_{ij} = 1 \tag{IV.4}$$

Nous désignons maintenant pour simplifier par modèle de Markov observable un processus stochastique observable, markovien et stationnaire. Dans un tel modèle, il y a un lien direct à tout instant entre l'état où se trouve le processus et l'observation réalisée à cet instant. C'est ce qui caractérise le fait que ce processus soit observable. Nous allons maintenant voir comment nous affranchir de cette contrainte en présentant d'autres processus stochastiques : les modèles de Markov cachés.

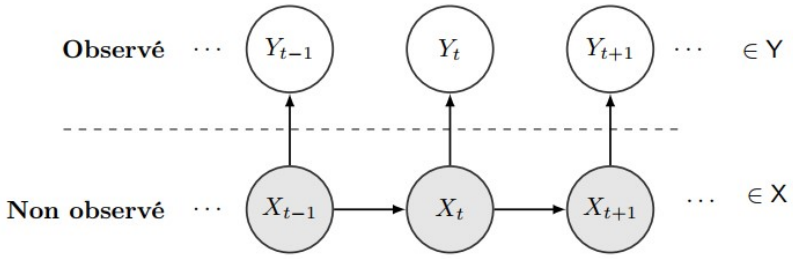


Figure IV.1 : Dynamique d'un modèle de Markov caché.

IV.2.2. Modèles de Markov cachés (HMM)

Le modèle de Markov caché généralise le modèle de Markov observable car il produit une séquence en utilisant deux suites de variables aléatoires; l'une cachée et l'autre observable (Figure IV.1) [170] :

• La suite cachée correspond à la suite des états $q_1, q_2, \ldots, q_T$, notée $Q(1 : T)$, où les q_i prennent leurs valeurs parmi l'ensemble des n états du modèle $\{s_1, s_2, \ldots, s_n\}$.

• La suite observable correspond à la séquence des d'observations $O_1, O_2, \ldots, O_T$, notée $O(1 : T)$, où les O_i sont des observations obtenues d'un total de M symboles observables $V = \{v_1, v_2, \ldots, v_M\}$.

Par conséquent, pour un HMM, un état n'est pas associé exclusivement à une observation donnée qu'il émettrait à coup sûr : chaque observation a désormais une certaine probabilité d'être émise par chaque état. En outre, ce ne sont pas les états qui sont observés, mais les observations qu'ils émettent.

Une conséquence importante est que l'on peut maintenant travailler avec un nombre de symboles observables M infinis. Une observation est alors émise avec une certaine densité de probabilité, correspondant à une distribution propre à chaque état.

En pratique, on cherche à construire des HMMs représentant des concepts dans l'espace de représentation des séquences. Pour des raisons de simplifications, les séquences construites sur des symboles observables $V = \{v_1, v_2, \ldots, v_M\}$ sont de taille M finie. Puisque la taille des symboles observables peut être infinie, ce qui signifie en pratique que chaque état peut émettre une variable continue ou un vecteur d'un espace euclidien de dimension d ($\Re^d$).

IV.3. Paramètres d'un HMM

Un HMM est noté $\Lambda = (A, B, \pi)$ et se définit par [170] :

- Ses états, en nombre n, qui composent l'ensemble $S = \{s_1, s_2, \ldots s_n\}$. L'état où se trouve le HMM à l'instant t est noté $q_t (q_t \in S)$, (Figure IV.2 et Figure IV.3).

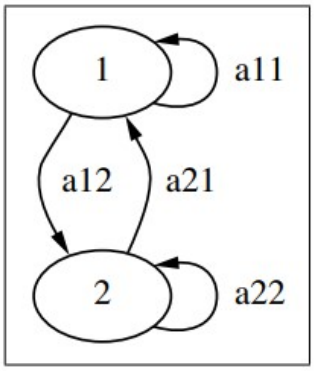


Figure IV.2 : Le HMM à deux états.

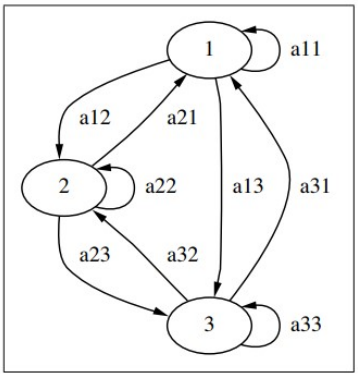


Figure IV.3 : Le HMM à trois états.

- M symboles observables dans chaque état. L'ensemble des observations possibles est noté $V=\{v_1, v_2, \ldots, v_M\}$. $O_t \in V$ est le symbole observé à l'instant t.

- Une matrice A de probabilités de transition entre les états : a_{ij} représente la probabilité que le modèle évolue de l'état i vers l'état j :

$$a_{ij} = A(i,j) = \mathrm{P}(q_{t+1} = sj \mid q_t = s_i), \quad \forall i, j \in [1 \ldots n], \ \forall t \in [1 \ldots T] \tag{IV.5}$$

Avec :

$$a_{ij} \geq 0, \quad \forall i, j \quad et: \quad \sum_{j=1}^{n} a_{ij} = 1 \tag{IV.6}$$

- Une matrice B de probabilités d'observation des symboles dans chacun des états du modèle : $b_j(k)$ représente la probabilité d'observer le symbole v_k alors que le modèle se trouve dans l'état j, soit :

$$b_j(k) = \mathrm{P}(O_t = v_k \mid q_t = s_j), 1 \leq j \leq n, \ 1 \leq k \leq M \tag{IV.7}$$

Avec :

$$b_j(k) \geq 0 \quad \forall j, k \quad et \quad \sum_{k=1}^{M} b_j(k) = 1 \tag{IV.8}$$

- Un vecteur π de probabilités initiales : $\pi = \{\pi_i\}_{i=1,2,\ldots,n}$. Pour tout état i, π_i est la probabilité que l'état de départ du HMM soit l'état i :

$$\pi_i = \mathrm{P}(q_1 = s_i) \ 1 \leq i \leq n \tag{IV.9}$$

Avec :

$$\pi \geq 0, \quad \forall i \quad et: \quad \sum_{i=1}^{n} \pi_i = 1 \tag{IV.10}$$

- Un ou plusieurs états finaux, sachant que le processus peut s'arrêter dans n'importe quel état, autrement dit que tout état peut être final.

IV.4. Types de topologies des HMMs

En pratique, on utilise deux types de modèles de Markov cachés, le modèle ergodique et le modèle gauche-droite [180].

IV.4.1. Le modèle ergodique

Il est sans contrainte, toutes les transitions d'un état vers un autre sont possibles. Comme il est montré dans la figure IV.4 :

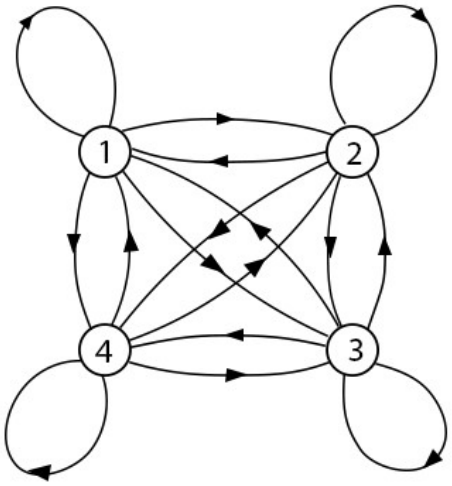


Figure IV.4 : Modèle ergodique du HMM.

IV.4.2. Le modèle gauche-droite

C'est un modèle contenant des contraintes résultantes de la mise à zéro de certaines valeurs a_{ij}. Dans le modèle le plus utilisé, celui de la figure IV.5, l'état i n'est relié par une transition de probabilité non nulle qu'à trois états : lui-même, l'état $i + 1$ et l'état $i + 2$, d'où le nom de modèle gauche-droite.

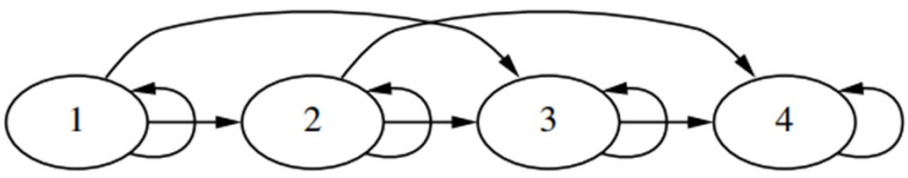


Figure IV.5 : HMM gauche-droite (ou Bakis) à quatre états.

IV.5. Génération d'une séquence d'état par un HMM

Un HMM peut être vu comme un processus permettant d'engendrer une séquence; inversement, on peut considérer une séquence comme une suite d'observations sur un HMM en fonctionnement [179].
En se plaçant du premier point de vue, la génération d'une séquence peut se décrire par l'algorithme 1 : c'est une procédure itérative gérée par des tirages aléatoires.

- **<u>Algorithme 1:</u> Génération d'une séquence par un HMM**

Début
$t \leftarrow 1$
Choisir l'état initial $q_1 = s_i$ **avec la probabilité** π_i
Tant que $t \leq T$ **faire**
Choisir l'observation $o_t = \upsilon_k$ **avec la probabilité** $b_i(k)$
Passer à l'état suivant $q_{t+1} = s_j$ **avec la probabilité** a_{ij}
$t \leftarrow t+1$
Fin tant que
Fin

Sachant qu'une séquence donnée peut en général être engendrée de plusieurs façons distinctes par un HMM.

IV.6. Classification de séquences d'observation

IV.6.1. Les trois problèmes des HMMs

Les définitions précédentes ne sont utilisables que si on sait calculer la probabilité qu'une séquence soit engendrée par un HMM et surtout si on sait entrainer un HMM à partir d'exemples.
On doit donc chercher des algorithmes pour résoudre les problèmes suivants [181] :

- *L'évaluation de la probabilité de l'observation d'une séquence* : Étant donné la séquence d'observations O et un HMM $\Lambda = (A, B, \pi)$, il faut évaluer la probabilité d'observation $P(O \mid \Lambda)$. Sachant que dans un problème de classification, on attribuera à une séquence la classe que modélise le HMM le plus probable étant donnée la séquence.

- *La recherche du chemin le plus probable* : Étant donnés la suite d'observations O et un HMM Λ, il est nécessaire de trouver une suite d'états $Q = q_1, q_2, ...,q_T$ qui maximise la probabilité d'observation de la séquence.

- *L'apprentissage* : On doit ajuster les paramètres (A, B, π) d'un HMM Λ pour maximiser $P(\mathbf{O} \mid \Lambda) = \prod_{O \in \mathbf{O}} P(O \mid \Lambda)$ à partir d'un ensemble O de séquences d'apprentissage.

IV.6.2. La classification bayésienne

Le principe est d'entrainer un HMM par classe à partir des exemples de cette classe. L'apprentissage d'un HMM s'effectue à partir d'un modèle initial; le HMM se modifie, mais en gardant jusqu'à sa convergence certaines caractéristiques du modèle initial (l'architecture reste la même) [182] :
- Le nombre d'états reste inchangé.
- Une transition de probabilité nulle entre deux états du modèle initial garde toujours une valeur nulle.

Le mieux est de prendre pour chaque classe un modèle initial ayant la même architecture : par exemple un modèle ergodique ou un modèle de Bakis (aussi appeler modèle gauche-droite). Pour chaque classe, le modèle initial peut simplement être pris avec le même nombre d'états.
Après C apprentissages indépendants, on dispose donc de C HMMs, que l'on peut noter $\Lambda_1, ... ,\Lambda_C$.

Étant donnée une séquence quelconque O, on a pour la classe de rang k :

$$P(\Lambda_k \mid O) = \frac{P(O \mid \Lambda_k).P(\Lambda_k)}{P(O)} \qquad \text{(IV.11)}$$

Le modèle qui doit être choisi par la règle bayésienne est celui qui maximise $P(\Lambda_k \mid O)$ (règle *MAP* : maximum a posteriori), ou si on suppose les classes équiprobables, celui qui maximise $P(O \mid \Lambda_k)$ (maximum de vraisemblance).
On doit donc être capable de calculer cette dernière valeur pour tout i. Cela nécessite un algorithme capable d'évaluer la probabilité qu'une séquence d'observation soit émise par un HMM.

IV.7. Evaluation de la probabilité d'observation

IV.7.1. Evaluation directe

Remarquons d'abord que la probabilité de la suite d'observations *O*, étant donné le modèle Λ, est égale à la somme sur toutes les suites d'états possibles *Q* des probabilités conjointes de *O* et de *Q* [183] :

$$P(O \mid \Lambda) = \sum_{Q} P(O, Q \mid \Lambda) = \sum_{Q} P(O \mid Q, \Lambda) P(Q \mid \Lambda) \tag{IV.12}$$

Or, on a les relations :

$$\begin{aligned} P(Q \mid \Lambda) &= \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \ldots a_{q_{T-1} q_T} \\ P(O \mid Q, \Lambda) &= b_{q_1}(O_1) b_{q_1}(O_1) \ldots b_{q_1}(O_1) \end{aligned} \tag{IV.13}$$

On déduit donc des formules précédentes, en réarrangeant les termes :

$$P(O \mid \Lambda) = \sum_{Q=q_1, q_2, \ldots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \ldots a_{q_{T-1} q_T} b_{q_T}(O_T) \tag{IV.14}$$

Cette formule directe nécessite d'énumérer toutes les suites d'états de longueur *T*, ce qui la rend coûteuse en temps de calcul.

IV.7.2. Evaluation par les fonctions avant-arrière (forward-backward)

Dans l'approche de Leonard Baum [184], on remarque que l'observation peut se faire en deux temps : d'abord, l'émission du début de l'observation $O(1 : t)$ en aboutissant à l'état q_i au temps *t*, puis, l'émission de la fin de l'observation $O(t + 1 : T)$ sachant que l'on part de q_i au temps *t*. Ceci posé, la probabilité de l'observation est donc égale à [185] :

$$P(O \mid \Lambda) = \sum_{i=1} \alpha_t(i) \beta_t(i) \tag{IV.15}$$

Où $\alpha_t(i)$ est la probabilité d'émettre le début $O(1 : t)$ et d'aboutir à q_i à l'instant *t*, et $\beta_t(i)$ est la probabilité d'émettre la fin $O(t + 1 : T)$ sachant que l'on part de q_i à l'instant *t*. Le calcul de α s'effectue avec *t* croissant tandis que le calcul de β est réalisé avec *t* décroissant, d'où l'appellation avant-arrière (*forward-backward*).

- **Calcul de α**

On a :

$$\alpha_t(i) = P(O_1 O_2 ... O_t, q_t = s_i \mid \Lambda) \quad \text{(IV.16)}$$

$\alpha_t(i)$ se calcule par l'algorithme 2, qui exprime que pour émettre le début de l'observation $O(1 : t + 1)$ et aboutir à l'état s_j au temps $t + 1$, on doit nécessairement être dans l'un des états s_i à l'instant t. Cette remarque permet d'exprimer $\alpha_{t+1}(j)$ en fonction des $\alpha_t(i)$ et d'utiliser un algorithme de programmation dynamique pour le calcul de tous les $\alpha_t(i)$ pour tout i, puis des $\alpha_{t+1}(i)$ pour tout i, etc.

- **Calcul de β**

De manière analogue, $\beta_t(i)$ se calcule par l'algorithme 3.

- **<u>Algorithme 2 :</u> Calcul de la fonction avant α**

Début

Pour $i = 1, n$ **faire** $\alpha_1(i) \leftarrow \pi_i b_i(O_1)$

$t \leftarrow 1$

Tant que $t < T$ **faire**

$j \leftarrow 1$

Tant que $j \leq n$ **faire**

$\alpha_{t+1}(j) \leftarrow \left(\sum_{i=1}^{n} \alpha_t(i) a_{ij} \right) b_j(O_{t+1})$

$j \leftarrow j + 1$

Fin tant que

$t \leftarrow t + 1$

Fin tant que

$P(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \alpha_T(i)$

Fin

- **<u>Algorithme 3 :</u> Calcul de la fonction arrière β**

Début

Pour *i = 1, n* **faire** $\beta_T(i) \leftarrow 1$

$t \leftarrow T - 1$

Tant que $t \geq T$ **faire**

$i \leftarrow 1$

Tant que $i \leq n$ **faire**

$\beta_t(i) \leftarrow \sum_{j=1}^{n} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$

$i \leftarrow i + 1$

Fin tant que

$t \leftarrow t - 1$

Fin tant que

$P(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \beta_1(i)$

Fin

Finalement, la probabilité d'observation d'une séquence est obtenue en prenant les valeurs de α et de β à un instant t quelconque : $P(O \mid \Lambda) = \sum_{i=1}^{n} \alpha_t(i) \beta_t(i)$. Cependant, on utilise le plus souvent les valeurs obtenues pour deux cas particuliers ($t = 0$) ou ($t = T$), ce qui donne:

$$P(O \mid \Lambda) = \sum_{i=1}^{n} \alpha_T(i) = \sum_{i=1}^{n} \pi_i \beta_0(i) \qquad \text{(IV.17)}$$

IV.8. Algorithme de Viterbi : calcul du chemin optimal

L'algorithme de Viterbi est apparu à la fin des années 60 dans [186], depuis cette époque, les applications de cet algorithme n'ont cessé de se développer. Il y a eu une utilisation importante de cet algorithme dans le domaine des télécommunications.

Il s'agit maintenant de déterminer le meilleur chemin correspondant à l'observation (Figure IV.6), c'est-à-dire de trouver dans le modèle Λ avec la meilleure suite d'états Q, qui maximise la quantité [187] :

$$\mathrm{P}(Q, O \mid \Lambda)$$

Pour trouver $Q = (q_1, q_2, \ldots, q_T)$ pour une séquence d'observations $O = (O_1, O_2, \ldots, O_T)$, on définit la variable intermédiaire $\delta_t(i)$ comme la probabilité du meilleur chemin amenant à l'état s_i à l'instant t, en étant guidé par les t premières observations :

$$\delta_t(i) = \underset{q_1,\ldots,q_{t-1}}{Max}\ P(q_1, q_2, \ldots, q_t = s_i, O_1, O_2, \ldots, O_t \mid \Lambda) \tag{IV.18}$$

Par récurrence, on calcule :

$$\delta_{t+1}(j) = [\underset{i}{Max}\ \delta_t(i) a_{ij}] b_j(O_{t+1}) \tag{IV.19}$$

On garde trace, lors du calcul, de la suite d'états qui donne le meilleur chemin amenant à l'état s_i à t dans la variable ψ sous la forme d'un tableau, comme montré dans la figure IV.6 :

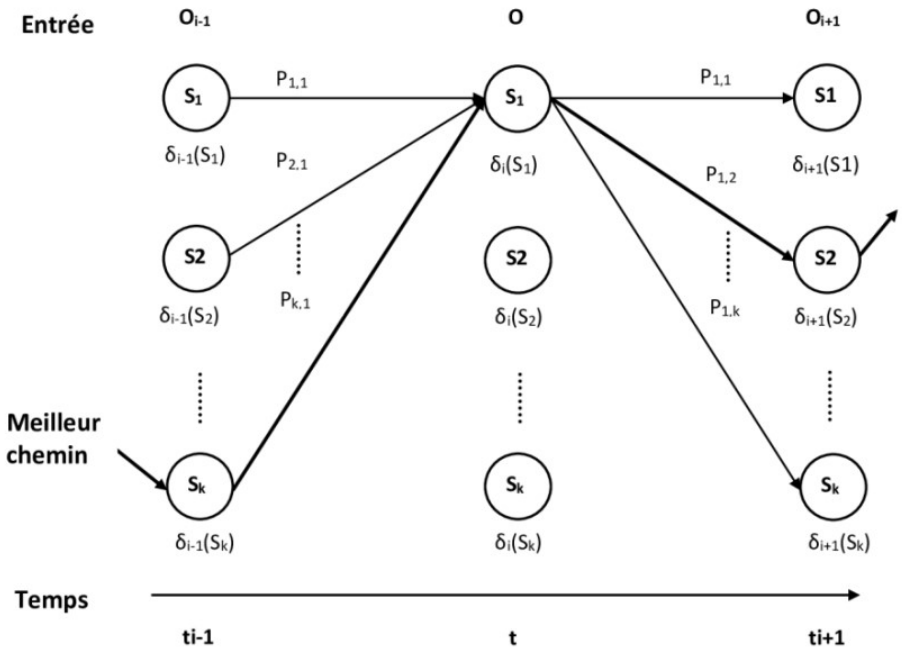


Figure IV.6 : Treillis de Viterbi.

On utilise une variante de la programmation dynamique, qui est l'algorithme de Viterbi (détailler dans l'algorithme 4) pour formaliser cette récurrence. Il fournit en sortie la valeur P*de la probabilité d'émission de la séquence par la meilleure suite d'états $(q_1^*, \ldots, q_T^*)$.

La fonction *Argmax* permet de mémoriser l'indice i, entre 1 et n, avec lequel on atteint le maximum des quantités $(\delta_{t-1}(i)a_{ij})$.

- **Algorithme 4 : Algorithme de Viterbi**

Début

pour $i = 1, n$ **faire**

$\delta_1(i) \leftarrow \pi_i b_i(O_1)$

$\psi_1(i) \leftarrow 0$

Fin pour

$t \leftarrow 2$

Tant que $t \leq T\text{-}1$ **faire**

$j \leftarrow 1$

Tant que $j \leq n$ **faire**

$\delta_t(j) \leftarrow Max_{1 \leq i \leq n} \left[\delta_{t-1}(i) a_{ij} \right] b_j(O_t)$

$\psi_t(j) \leftarrow ArgMax_{1 \leq i \leq n} \left[\delta_{t-1}(i) a_{ij} \right]$

$j \leftarrow j + 1$

Fin tant que

$t \leftarrow t + 1$

Fin tant que

$P^* \leftarrow \psi_{t+1}(q^*_{t+1})$

$t \leftarrow t - 1$

Fin tant que

Fin

IV.9. Apprentissage des HMMs

Supposons qu'on dispose d'un ensemble de séquences $\mathbf{O} = \{O_1, \ldots, O_m\}$, dont l'élément courant est noté O_k. Le but de l'apprentissage est de déterminer les paramètres d'un HMM d'architecture fixée $\Lambda = (A, B, \pi)$, qui maximisent la probabilité $P(O \mid \Lambda)$. Comme on suppose que les séquences d'apprentissages sont tirées indépendamment, on cherche donc à maximiser [188] :

$$P(\mathbf{O} \mid \Lambda) = \prod_{k=1}^{m} P(O_k \mid \Lambda) \qquad \text{(IV.20)}$$

L'idée est d'utiliser une procédure de réestimation qui affine le modèle progressivement selon les étapes suivantes :

- Choisir un ensemble initial Λ_0 de paramètres,

- Calculer Λ_1 à partir de Λ_0, puis Λ_2 à partir de Λ_1, etc.
- Répéter ce processus jusqu'à un critère de fin.

Pour chaque étape p d'apprentissage, on dispose de Λ_p et on cherche un Λ_{p+1} qui doit vérifier :

$$P(\mathbf{O} \mid \Lambda_{p+1}) \geq P(\mathbf{O} \mid \Lambda_p) \tag{IV.21}$$

Soit :

$$\prod_{k=1}^{m} P(O_k \mid \Lambda_{p+1}) \geq \prod_{k=1}^{m} P(O^k \mid \Lambda_p) \tag{IV.22}$$

Λ_{p+1} doit donc améliorer la probabilité d'émission des observations de l'ensemble d'apprentissage. La technique pour calculer Λ_{p+1} à partir de Λ_p consiste à utiliser l'algorithme dit *EM* (*expectation maximisation*). Pour cela, on effectue un comptage de l'utilisation des transitions A et des distributions B et π du modèle Λ_p quand il produit l'ensemble O. Si cet ensemble est assez important, ces fréquences fournissent de bonnes approximations *a posteriori* des distributions de probabilités A, B et π et sont utilisables alors comme paramètres du modèle Λ_{p+1} pour l'itération suivante.
La méthode d'apprentissage *EM* consiste donc dans ce cas à regarder comment se comporte le modèle défini par Λ_p sur O, à réestimer ses paramètres à partir des mesures prises sur O, puis à recommencer cette réestimation jusqu'à obtenir une convergence.

IV.9.1. Formules de réestimation

On définit $\xi_t^k(i, j)$ comme la probabilité, étant donnés une séquence d'observation O_k et un HMM Λ, que ce soit l'état s_i qui ait émis l'observation de rang t de O_k et l'état s_j qui ait émis celle de rang $t + 1$ (la lettre k en exposant fait référence à la séquence d'apprentissage concernée) [189].

Donc :

$$\xi_t^k(i, j) = \mathrm{P}(q_t = s_i, q_{t+1} = s_j \mid O_k, \Lambda) \tag{IV.23}$$

Ce qui se réécrit comme:

$$\xi_t^k(i, j) = \frac{P(q_t = s_i, q_{t+1} = s_j, O_k \mid \Lambda)}{P(O^k \mid \Lambda)} \tag{IV.24}$$

Par définition des fonctions avant-arrière, on en déduit :

$$\xi_t^k(i,j) = \frac{\alpha_t^k(i) a_{ij} b_j(O_{t+1}^k) \beta_{t+1}^k(j)}{P(O^k \mid \Lambda)} \tag{IV.25}$$

On définit aussi la quantité $\gamma_t^k(i)$ comme la probabilité que le symbole de rang t de la séquence d'observation O_k soit émise par l'état s_j.

$$\gamma_t^k(i) = \mathrm{P}(q_t = s_i \mid O_k, \Lambda) \tag{IV.26}$$

Soit :

$$\gamma_t^k(i) = \sum_{j=1}^{n} P(q_t = s_i,\ q_{t+1} = s_j \mid O^k, \Lambda) = \frac{\sum_{j=1}^{n} P(q_t = s_i,\ q_{t+1} = s_j, O^k \mid \Lambda)}{P(O^k \mid \Lambda)} \tag{IV.27}$$

On a la relation :

$$\gamma_t^k(i) = \sum_{j=1}^{n} \xi_t(i,j) \frac{\alpha_t^k(i) \beta_t^k(i)}{P(O^k \mid \Lambda)} \tag{IV.28}$$

Le nouveau modèle HMM se calcule à partir de l'ancien en réestimant π, A et B par comptage sur la base d'apprentissage. On mesure les fréquences :

$$\bar{a}_{ij} = \frac{\text{nombre de fois où la transition de } s_i \text{ à } s_j \text{ a été utilisée}}{\text{nombre de transitions effectuées à parir de } s_i}$$

$$\bar{b}_j(l) = \frac{\text{nombre de fois où le HMM s'est trouver dans l'état } s_j \text{ en observant } v_l}{\text{nombre de fois où le HMM s'est trouvé dans l'état } s_j} \tag{IV.29}$$

$$\bar{\pi}_i = \frac{\text{nombre de fois où le HMM s'est trouvé dans l'état } s_i \text{ en émettant le premier symbole d'une phrase}}{\text{nombre de fois où le HMM a émis le premier symbole d'une phrase}}$$

Compte tenu de ces définitions, on obtient:

$$\bar{\pi}_i = \frac{1}{m} \sum_{k=1}^{m} \gamma_1^k(i)$$

$$\bar{a}_{ij} = \frac{\sum_{k=1}^{m} \sum_{t=1}^{|O^k|-1} \xi_t^k(i,j)}{\sum_{k=1}^{N} \sum_{t=1}^{|O^k|-1} \gamma_t^k(i)} \qquad \text{(IV.30)}$$

$$\bar{b}_j(l) = \frac{\sum_{k=1}^{m} \sum_{t=1}^{|O^k|-1} \gamma_t^k(j)}{\sum_{k=1}^{m} \sum_{t=1}^{|O^k|-1} \gamma_t^k(j)} \quad \text{avec } O_t^k = v_l$$

Ces formules ont été établies par Leonard Baum [184], comme une application de la procédure *EM* à l'apprentissage des paramètres HMM. La suite des modèles construits par l'algorithme de Baum-Welch vérifie la relation cherchée :

$$P(\mathbf{O} \mid \Lambda_{p+1}) \geq P(\mathbf{O} \mid \Lambda_p) \qquad \text{(IV.31)}$$

IV.9.2. Algorithme de Baum-Welch

L'algorithme de Baum-Welch est décrit comme ci-dessous :

- **<u>Algorithme 5 :</u> Algorithme de Baum-Welch**

Début

Fixer les valeurs initiales (A, B, π)

Définir le HMM de départ comme $\Lambda_0 = (A, B, \pi)$

$p \leftarrow 0$

Tant que la convergence n'est pas atteinte faire

Obtenir le HMM Λ_p

Calculer pour ce modèle, sur l'ensemble d'apprentissage, les valeurs :

$\xi(i,j), \quad \gamma_t(i) \quad 1 \leq i, j \leq n \quad 1 \leq t \leq T-1$

Déduire $\bar{\pi}$, $\bar{A}$, $\bar{B}$ **en utilisant les formules de réestimation**

Le HMM courant est désormais défini par $\Lambda_{p+1} = (\bar{\pi}, \bar{A}, \bar{B})$

$p \leftarrow p+1$

Fin tant que

Fin

Il faut savoir aussi, qu'en pratique :

- Le choix du modèle initial influe sur les résultats; par exemple, si certaines valeurs de A et B sont égales à 0 au départ, elles le resteront jusqu'à la fin de l'apprentissage. Ceci permet en particulier de garder la structure dans les modèles gauches-droites.
- L'algorithme converge vers des valeurs de paramètres qui assurent un maximum local de $P(O \mid \Lambda)$. Il est donc important, si on veut être aussi près que possible du minimum global, de bien choisir la structure et l'initialisation.
- Le nombre d'itérations est fixé empiriquement. L'expérience prouve que, si le point précédent a été correctement traité, la stabilisation des paramètres ne correspond pas à un sur-apprentissage : il n'y a donc en général pas besoin de contrôler la convergence par un ensemble de validation. Mais cette possibilité est évidemment toujours à disposition.

IV.10. Classification statistique

Connaissant une observation x, nous devons connaitre la probabilité (ou densité de probabilité dans le cas continu) $p(\omega_k,x)$ pour que x appartienne à la classe ω_k, sachant qu'il existe K classes [182].

On peut montrer qu'un classifieur assigne x à la classe ω_i si :

$$p(\omega_i \mid x) = p(\omega_j \mid x), \ \forall j = 1...J, \ j \neq i \tag{IV.32}$$

Ce classifieur doit commettre un nombre minimum d'erreurs de classification. Cette stratégie optimale est une forme de la règle de décision bayésienne.

Il est en général impossible de calculer directement la probabilité $p(\omega_i,x)$. On peut cependant l'estimer à partir de données d'apprentissage. Malheureusement, l'algorithme d'apprentissage permet seulement d'estimer la probabilité d'observer le vecteur x connaissant sa classe d'appartenance ω_i : $p(x \mid \omega_i)$. La formule de Bayes [182] permet de calculer $p(\omega_i|x)$:

$$p(\omega_i \mid x) = \frac{p(x \mid \omega_i)\, p(\omega_i)}{p(x)} \tag{IV.33}$$

$p(\omega_i|x)$ est appelée probabilité *a posteriori*,

$p(x|\omega_i)$ est la probabilité d'observer x connaissant la classe ω_i, souvent appelée vraisemblance (likelihood en anglais),

$p(\omega_i)$ est la probabilité *a priori* de la classe ω_i,

$p(x)$ est la probabilité d'observation de x.

IV.11. Fonctionnement du HMM

Un HMM est un automate probabiliste contrôlé par deux processus stochastiques. Le premier débute sur l'état initial du HMM et se déplace ensuite d'état en état, en respectant les transitions autorisées par la topologie de cet automate. Le second génère une observation dans chaque état du HMM. Un exemple de l'HMM utilisé en reconnaissance automatique de la parole est représenté sur la figure IV.7 [180].

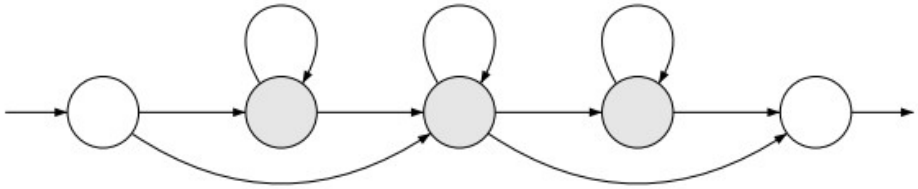


Figure IV.7 : Exemple de l'HMM à cinq états. Les états grisés sont émetteurs, les états blancs ne génèrent aucune observation.

IV.12. Hypothèses simplificatrices pour la reconnaissance de la parole

La mise en œuvre pratique d'un HMM en reconnaissance de la parole nécessite de poser un ensemble d'hypothèses simplificatrices résumées ci-dessous [181] :

a) Hypothèse de base

On suppose que le signal de parole est produit par une suite d'états, chaque état étant gouverné par une loi statistique. Chaque unité de parole (phone, diphone, triphone, mot) est associée à un modèle de Markov et la concaténation de tels modèles permet d'obtenir des mots ou des phrases.

b) Hypothèse n◦1

Les probabilités $p(\omega_k)$ de la formule (IV.33) peuvent être calculées séparément, c'est-à-dire sans données acoustiques. Habituellement, on utilise un modèle de langage appris sur un grand corpus de données.

c) Hypothèse n◦2

La probabilité que le modèle de Markov soit dans l'état *i* au temps *t* ne dépend que de l'état du modèle de Markov au temps *t-1* (modèle de Markov du premier ordre) :

$$P(q_t \mid q_{t-1}q_{t-2}\ldots q_0) \cong P(q_t \mid q_{t-1}) \qquad \text{(IV.34)}$$

d) Hypothèse n◦3

La chaîne de Markov est stationnaire :

$$P\left(q_t = j \mid q_{t-1} = i\right) \cong P(q_{t+v} = j \mid q_{t+v-1} = i), \quad \forall v \tag{IV.35}$$

e) Hypothèse n◦4

La probabilité qu'un vecteur soit émis au temps t ne dépend pas des vecteurs précédemment émis (indépendance des observations) :

$$P(o_t \mid q_0\, q_1 \ldots q_t,\, o_1\, o_2 \ldots o_{t-1}) \cong P(o_t \mid q_0\, q_1 \ldots q_t) \tag{IV.36}$$

f) Hypothèse n◦5

La probabilité qu'un vecteur soit émis au temps t ne dépend pas des états précédemment visités, mais uniquement de l'état courant :

$$P(o_t \mid q_t\, q_{t-1} \ldots q_0) \cong P(o_t \mid q_t) \tag{IV.37}$$

g) Hypothèse n◦6

La distribution des probabilités d'émission est approximée par un mélange de k lois normales (gaussiennes) de la forme :

$$b_j(o_t) = \sum_{i=1}^{k} \frac{c_i}{(2\pi)^d \left|\sum_i\right|} \exp\left(-\frac{1}{2}(o_t - \mu_i)'.\sum_i^{-1}.(o_t - \mu_i) \right) \tag{IV.38}$$

μ_i et Σ_i représentent la moyenne et la matrice de covariance de la $i^{\text{ème}}$ loi normale, c_i est la pondération de la loi i (avec $\sum_{i=1}^{k} c_i = 1$) , $|\Sigma_i|$ est le déterminant de la matrice Σ_i.

Ces hypothèses sont utilisées en pratique et les résultats obtenus prouvent la validité du modèle suivi.

IV.13. Reconnaissance de mots isolés

En reconnaissance de mots isolés en utilisant V modèles $M_1 \ldots M_V$, correspondant aux V mots du vocabulaire qui ont été préalablement appris, la reconnaissance revient à trouver la meilleure séquence d'états pouvant engendrer la séquence d'observations $O = (o_1, o_2, \ldots, o_T)$ correspondant à la prononciation d'un mot du vocabulaire. Il faut donc calculer $P(O|M_i)P(M_i)$ pour chacun des V modèles.

Une solution naïve revient à évaluer max $P(O|M, q)$ pour chaque séquence d'états possibles q, avec [169] :

$$P(O \mid M, q) = \pi_0 \prod_{i=1}^{T} a_{qt-1\,qt} b_{qt}(o_t) \qquad \text{(IV.39)}$$

Cette solution est inapplicable en général car le nombre de séquences d'états possibles est très grand, précisément de l'ordre de $T \times N^T$. Par exemple pour 5 états et une séquence de 100 observations, cela représenterait plus de 1072 séquences. L'algorithme de Viterbi, variante stochastique de l'algorithme de programmation dynamique, permet de trouver la séquence optimale d'états en effectuant beaucoup moins de calculs. Le principe revient à construire de façon itérative la meilleure séquence d'états à partir d'un tableau $T \times N$ contenant les valeurs δ. La vraisemblance $\delta_t(i)$ du meilleur chemin qui finit à l'état i au temps t peut être calculée par récurrence :

- Initialisation : $\delta_0(i) = \pi_i$ (par définition, à $t = 0$ la probabilité d'être à l'état i est égale à π_i).

- Récursion : pour se trouver à l'état i à l'instant t, le processus markovien se trouvait forcément dans l'un des états du modèle à l'instant $t-1$ et il a effectué la transition de cet état vers l'état i. En utilisant le principe d'optimalité de Bellman [179, 181], on en déduit que $\delta_t(i) = \underset{j}{Max}(\delta_{t-1}(i).a_{ij}).b_i(o_t)$, j étant l'un des états du modèle de Markov.

- Terminaison : $P = \underset{i}{Max}(\delta_T(i))$. La vraisemblance est obtenue en recherchant l'état qui maximise la valeur de δ à la dernière observation T.

Cet algorithme présente plusieurs avantages. Tout d'abord, sa complexité de calcul (nombre de séquences d'états possibles) est $T \times N^2$, donc linéaire en T à comparer avec les $T \times N^T$ de l'algorithme naïf [190, 191]. De plus, l'algorithme est « synchrone à la trame », c'est-à-dire que les calculs pour une trame de parole t ne nécessite que des données déjà calculées précédemment. Le traitement en temps réel est donc possible, sous réserve d'une puissance de calcul suffisante. Les calculs peuvent débuter dès que le locuteur commence à parler. Enfin, l'algorithme est très peu gourmand en mémoire puisque seules deux colonnes du tableau $T \times N$ sont nécessaires simultanément, celles correspondantes aux trames t et $t + 1$.

IV.14. Reconnaissance de la parole continue

La reconnaissance de la parole continue présente de nouvelles difficultés car on ne connaît pas le nombre de mots qui composent une phrase, ni les frontières de chaque mot. Une solution simpliste serait de reprendre pour la parole continue le même algorithme que pour les mots isolés en l'appliquant pour chacune des phrases autorisées par la grammaire de l'application considérée. Si cette solution convient pour des langages de commandes simples, elle est inapplicable pour des grands vocabulaires. Une modification de l'algorithme de reconnaissance des mots isolés est donc nécessaire [192].

Dès que le dernier état d'un des mots du vocabulaire est atteint, on peut commencer un autre mot, c'est-à-dire effectuer une transition vers le premier état d'un mot du vocabulaire.

L'algorithme de Viterbi comprend alors les phases suivantes :

- Initialisation : $\delta_0^k(i) = \pi_i^k$,k désignant l'un des mots du vocabulaire.

- Récursion :

État non initial du mot k : $\delta_t^k(i) = \underset{j}{Max}(\delta_{t-1}^k(i) * a_{ij}^k).b_i(o_t)$,

État initial du mot k : $\delta_t(i) = Max(\underset{j}{Max}(\delta_{t-1}^k(i) * a_{ij}^k)$,

$\underset{l}{Max}(\delta_{t-1}^l(\text{état final}(l)) * P(M_l \mid M_k))).b_i(o_t)$.

- Terminaison : $P = \underset{k}{Max}(\delta_T^k(i))$.

La connaissance de $P = \underset{k}{Max}(\delta_T^k(i))$ n'indique que le dernier état qui maximise la séquence, c'est-à-dire le dernier mot prononcé. Une structure supplémentaire doit être ajoutée afin de retrouver la séquence des modèles. Deux solutions peuvent être adoptées : La première revient à mémoriser dans le tableau des δ l'état qui avait contribué à calculer le maximum. Ainsi, à la fin de la reconnaissance (à la dernière trame T), il suffit de revenir en arrière à partir de l'état qui maximise $Max(\delta_T^k(i))$ pour retrouver la séquence optimale d'états et donc la séquence de mots. On obtient également les frontières temporelles pour chacun des mots reconnus.

La deuxième solution nécessite de mémoriser dans $\delta_t^k(i)$ la séquence de mots qui a permis d'arriver à l'état i du modèle k à l'instant t. Cette solution a l'avantage de ne

pas nécessiter de retour-arrière à la fin de la reconnaissance, mais elle nécessite plus de mémoire.

Un des avantages de l'algorithme est qu'il ne nécessite aucune hypothèse sur le nombre de mots qui constituent la séquence. A la fin de la reconnaissance, on obtient non seulement la séquence de mots prononcés mais aussi les instants de début et de fin de chaque mot reconnu.

IV.15. Limitations des HMMs

Le modèle HMM repose sur un ensemble d'hypothèses simplificatrices. Certaines d'entre elles sont des sources de limitation du modèle que les chercheurs ont tenté de pallier. Elles concernent en particulier [193] :

- La modélisation de la durée : celle-ci est implicite dans un HMM (par une distribution géométrique). L'adjonction d'une modélisation explicite des durées d'occupation des états du HMM a été proposée et expérimentée avec succès [194, 195].
- L'hypothèse d'indépendance conditionnelle des observations : cette hypothèse est nécessaire mais irréaliste. De nombreux travaux ont été menés pour compenser cette insuffisance. Une première solution, efficace quoique limitée et largement adoptée, revient à utiliser les dérivées premières et secondes des paramètres. La modélisation explicite de la corrélation entre les trames successives de parole a également été étudiée dans le cadre des HMMs conditionnellement gaussiens [196] et des HMMs segmentaux [197].

IV.16. Conclusion

Dans ce chapitre nous avons présenté les principaux fondements théoriques des modèles de Markov cachés considérés parmi les méthodes d'apprentissage non supervisé et sont caractérisé par leur nature statistique, avec une propriété de décision bayésienne.

Ils sont massivement utilisés notamment en reconnaissance de formes, en intelligence artificielle ou encore en traitement automatique du langage naturel. Dans ce travail de recherche, les HMMs sont utilisés pour la classification des voix pathologiques conjointement avec les SVMs afin de profiter des avantages des deux méthodes.

Les HMMs se sont montrés extrêmement utiles dans le domaine de la reconnaissance de la voix [198] et la classification de la parole [199], ils offrent plusieurs avantages, comme la stabilité des résultats en les utilisant sur différentes bases de données, ainsi que le bon taux de classification correcte comparés aux autres méthodes existantes.

Chapitre 5 :

Classification automatique des voix pathologiques par HMM-SVM

V.1. Introduction

Les méthodes d'évaluation des troubles de la voix basées sur l'analyse acoustique de la parole permettent d'évaluer quantitativement le degré d'enrouement perçu via des indices qui reflètent la sévérité du trouble. La plupart des méthodes d'évaluation clinique de la qualité de la voix supposent la stationnarité et la pseudo-périodicité du signal de parole; pour répondre à ces hypothèses, l'analyse est effectuée sur des fragments stables extraits de voyelles soutenues [200]; cependant l'analyse de la parole continue peut être plus informative que les voyelles soutenues.

Des méthodes de reconnaissance de formes ou des méthodes statistiques ont été proposées dans de nombreuses études comme une alternative pour la caractérisation et la classification des troubles de la voix. Cependant, la plupart de ces études ont porté sur la classification binaire normale/pathologique de la voix ou de la parole [201-203].

Les modèles hybrides basés sur les modèles de Marcov cachés (HMMs) et les séparateurs à vaste marge (SVMs) se sont avérés efficaces dans le cadre de la reconnaissance vocale. Dans [204], les auteurs ont suggéré la combinaison des HMMs et SVMs pour segmenter et étiqueter la forme d'onde de la parole en unités de phonèmes; Les HMMs génèrent la séquence des phonèmes et leurs frontières alors que les SVMs refont les frontières et corrigent les étiquettes; Les résultats ont montré que la reconnaissance vocale basée sur la segmentation HMM-SVM surpasse celle basée sur la segmentation d'apprentissage intégrée. Dans [205] une nouvelle méthode hybride de reconnaissance de la parole arabe a été présentée. Il a été conclu que le modèle hybride HMM-SVM est plus efficace que les HMM et le système hybride perceptron multicouches MLP avec HMM.

Dans ce chapitre, on propose un classificateur hybride HMM-SVM pour l'évaluation des voix pathologiques en termes de degré de perturbation; On considère la classification en deux classes et en trois classes.

V.2. Corpus et évaluation perceptive

Le corpus a été fourni par Dr. Youri Maryn [206] thérapeute de la parole à l'hôpital général de Sint-Jan AV, Bruges, Belgique. Il comprend des voyelles [a] soutenues et deux Phrases néerlandaises (« Papa en Marloes staan op het station. Ze wachten op de trein ») produites par 28 locuteurs normophoniques âgés de 19 à 40 ans et 223 locuteurs dysphoniques avec différents degrés de dysphonie âgés de 8 à 85 ans. Les diagnostics sont les suivantes :

nodules (42), dysphonie fonctionnelle (81), polype (11), œdème (29), parésie/paralysie (18), laryngite aigu (5), kyste (8) et autres (34).

Chaque locuteur a soutenu la voyelle [a] pendant au moins 5 secondes et a lu un texte phonétiquement équilibré à une tonalité et un volume confortables. Les stimuli de voyelles [a] ont été enregistrés pour 3 secondes seulement en ne gardant que la partie stable du milieu alors que les phrases n'ont subi aucune modification. A partir de ces stimuli, on forme deux corpus. Le premier corpus comprend des voyelles [a] de durée 3 secondes et le deuxième corpus comprend la concaténation des deux phrases suivies de la voyelle [a] de durée 3 secondes. Le deuxième corpus sera référé comme étant le corpus de parole continue. Les stimuli ont été enregistrés au moyen d'un microphone à condensateur AKG C420 à une fréquence d'échantillonnage de 44100 Hz et une résolution de 16 bits.

Le corpus a été évalué auditivement par cinq juges, qui sont des thérapeutes professionnels ayant au moins cinq ans d'expérience dans l'évaluation clinique de la qualité de la voix, chaque juge ayant assigné un score entre 0 et 3 du grade (G) de l'échelle GRBAS. Le score représente la qualité globale de la voix et est considérée avec la raucité (R) et le souffle (B) comme la plus fiable et la moins sujette à la variabilité d'un jury d'écoute par rapport à l'asthénie (A) et le forçage (S). Une voix modale se voit attribuer le score 0 et une voix très perturbée se voit attribuer le score 3. Le score global est obtenu comme étant la moyenne des scores attribués par les cinq juges. La figure V.1 montre la distribution des scores moyens attribués par les juges [206].

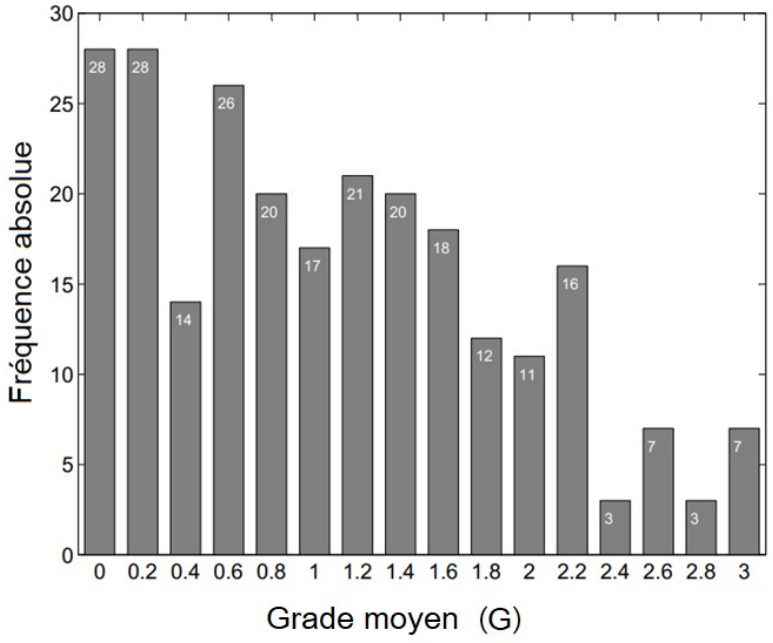


Figure V.1 : Distribution de fréquence du grade moyen de dysphonie (G) pour le corpus de concaténation des deux phrases suivies de la voyelle [a].

V.3. Classification des voix pathologiques à base de HMM et SVM

Les classificateurs SVM, HMM et HMM-SVM ont été appliqués aux corpus de voyelles [a] soutenues et de parole continue. Dans ce qui suit, on présente les résultats de classification en deux classes (classification binaire) et en trois classes.

La catégorisation des voix est basée sur les scores moyens assignés par les juges. Pour la classification en deux classes, la première classe contient 70 stimuli avec des scores moyens entre 0 et 0.4 (voix modales) et la deuxième classe contient 181 stimuli avec des scores moyens entre 0.6 et 3 (voix perturbées). Pour la classification en trois classes, la première classe contient 70 stimuli avec des scores moyens compris entre 0 et 0.4 (voix modales), la deuxième classe contient 104 stimuli avec des scores moyens entre 0.6 et 1.4 (voix modérément perturbées) et la troisième classe contient 77 stimuli, avec des scores moyens compris entre 1.6 et 3 (voix très perturbées). Comme exemple, les figures V.2 et V.3 montrent, respectivement, les formes d'onde des voyelles [a] soutenues et de la concaténation des deux phrases de chacune des trois classes. Les tableaux V.1 et V.2 [207] montrent la répartition par classe et par sexe des enregistrements du corpus pour la classification binaire et la classification en trois classes.

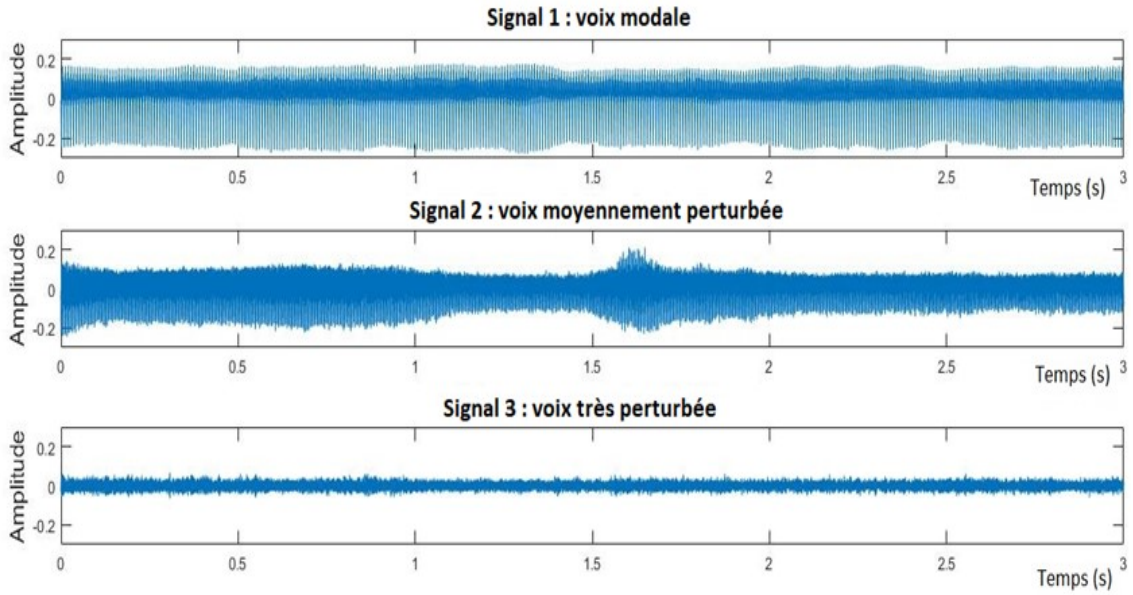


Figure V.2 : Voyelles [a] produites par des locuteurs de chacune des trois classes.

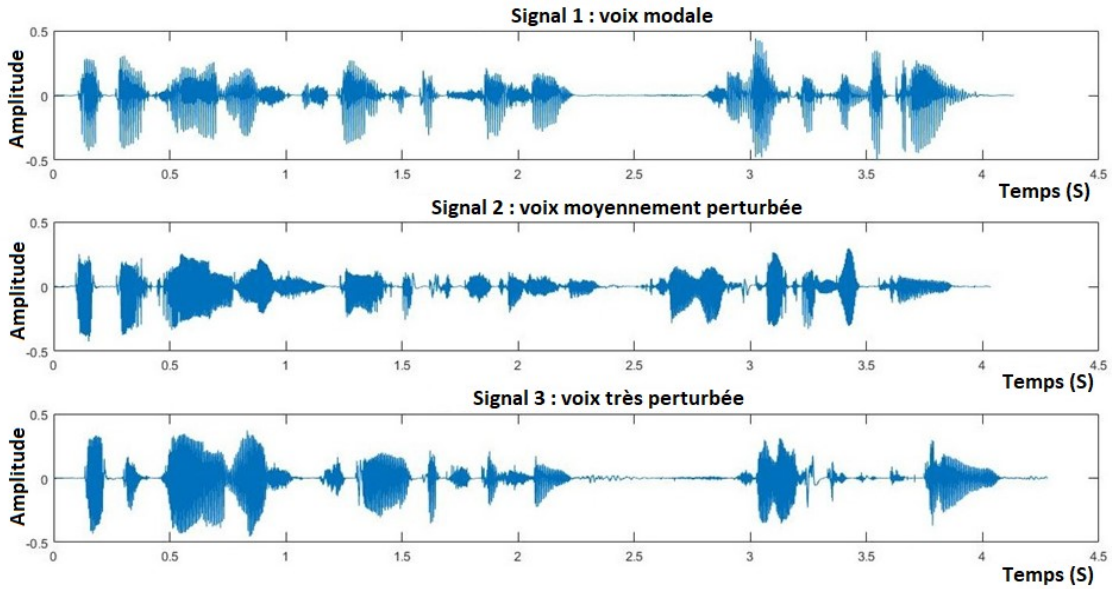


Figure V.3 : Concaténation des deux phrases produites par des locuteurs de chacune des trois classes.

Tableau V.1 : Répartition des voix selon le degré de dysphonie et le type de sexe pour la classification binaire.

Type de la classe	sexe		
	Masculins	Féminins	total
Voix modales	13	57	70
Voix perturbées	67	114	181

Tableau V.2 : Répartition des voix selon le degré de dysphonie et le type de sexe pour la classification en trois classes.

Type de la classe	sexe		
	Masculins	Féminins	total
Voix modales	13	57	70
Voix modérément perturbées	34	70	104
Voix très perturbées	33	44	77

V.3.1. Descripteurs acoustiques

Les indices acoustiques utilisés comme entrées des classificateurs sont les coefficients MFCC, LFPC et LPCC qui sont souvent utilisés dans la reconnaissance et la classification

de la parole et du locuteur. L'extraction des paramètres acoustiques a pour but de transformer le signal vocal en une séquence de vecteur de caractéristiques acoustiques qui préserve les caractéristiques acoustiques du locuteur.

V.3.1.1. Coefficients MFCCs (Mel-frequency cepstral coefficients)

Les coefficients MFCCs [208] peuvent être calculés selon les étapes suivantes (Figure V.4):

- Le signal de parole est divisé en trames sans chevauchement puis la fenêtre de Hamming est appliquée à chaque trame de 30ms.

- Le spectre de puissance est calculé en utilisant la transformation de Fourier discrète $X(k)$ pour chaque fenêtre.

$$X(k) = \sum_{n=0}^{N-1} x(n)\ e^{-j(2\pi k/N)n} \tag{V.1}$$

N : longueur de la trame.

- Le passage à l'échelle de Mel se fait en faisant passer le spectre de puissance par un banc de filtre Mel (Figure V.5) :

$$S(r) = \sum_{k} W_r(k).\left|X(k)\right|^2 \tag{V.2}$$

r = (1 ... M), M étant le nombre de bandes triangulaires, avec M = 20 filtres triangulaires espacées sur une échelle mel.

$W_r(k)$ est la fonction de pondération triangulaire associée à la k^{eme} bande.

Par conséquent, les bandes de fréquence deviennent plus larges lorsque la fréquence en Hz augmente. Pour obtenir la fréquence Mel, l'axe des fréquences du spectre de puissance est converti à l'aide de l'équation :

$$Mel(f) = 2595\, Log_{10}(1 + \frac{f}{700}) \tag{V.3}$$

- Les coefficients MFCC sont obtenus en prenant le logarithme de la sortie du banc de filtre puis en calculant la transformée en cosinus discrète (DCT : discret cosine transform).

$$MFCC(m) = \sum_{r=1}^{M} \log[S(r)].\cos\left[m.(r - \frac{1}{2}).\frac{\pi}{M} \right] \tag{V.4}$$

où m=(1 ... L); avec $L \leq M$; M étant le nombre de coefficients.

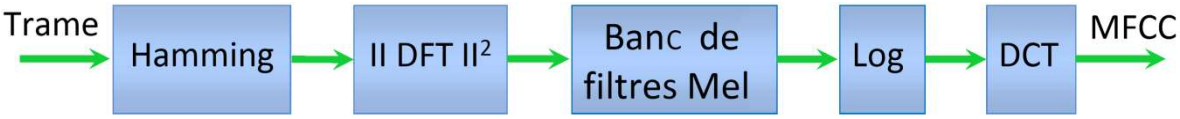


Figure V.4 : Schéma bloc de l'algorithme d'extraction des coefficients MFCC.

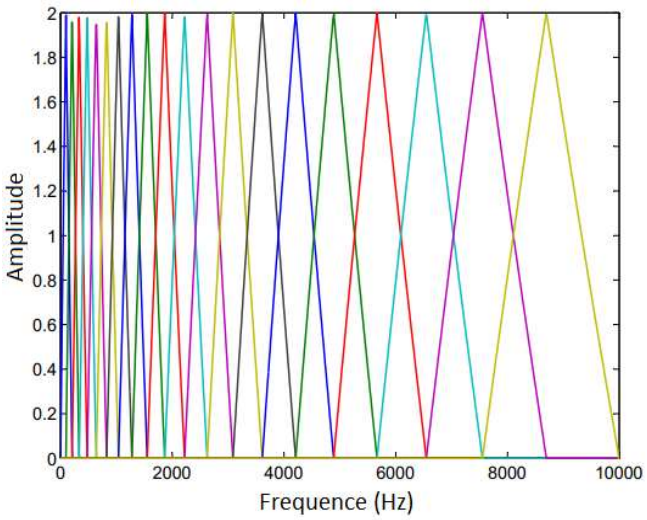


Figure V.5 : Banc de filtres à l'échelle de Mel.

V.3.1.2. Coefficients LFPCs (log frequency power coefficients)

Les Coefficients de puissance de fréquence logarithmique (LFPC) [209] sont conçus pour simuler les caractéristiques de filtrage logarithmique du système auditif humain en utilisant un banc de filtres logarithmiques de 200 Hz à 4 kHz. Tout d'abord, le signal de parole est segmenté en fenêtres de courte durée puis la fenêtre de Hamming est appliquée à chaque trame, aprés on applique pour chaque trame la transformée de Fourier rapide (FFT) afin de calculer le spectre de puissance ensuite on fait passer le signal obtenu par un banc de filtres à fréquence logarithmique. Le banc de filtres composé de plusieurs filtres passe-bande dont la sortie de chaque filtre donne le spectre $S(m)$ de la bande de fréquences m. Finalement, l'énergie est calculée à la sortie du banc de filtres comme suit :

$$LFPC(m) = \frac{10\,Log_{10}(S(m))}{N_m} \qquad \text{(V.5)}$$

Où N_m est le nombre de composantes spectrales dans le m^{ieme} banc de filtres.

La figure V.6 montre le schéma bloc des étapes de calcul des coefficients LFPCs.

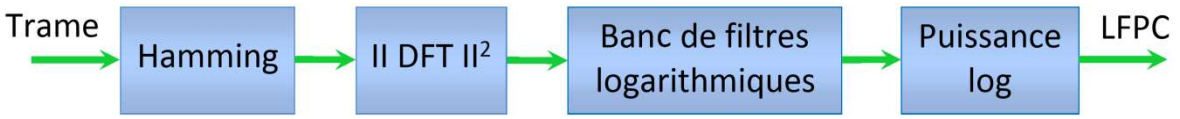


Figure V.6 : Schéma bloc de l'algorithme d'extraction des coefficients LFPCs.

V.3.1.3. Coefficients LPCCs (linear predictive cepstral coefficients)

Les coefficients cepstraux prédictifs linéaires (LPCC) sont extraits pour représenter l'enveloppe spectrale du signal de parole. Les coefficients sont obtenus en utilisant les prédictions faites à partir des p échantillons précédents du signal [210]. En raison de leur simplicité et du fait qu'ils donnent des informations spécifiques aux différentes classes liées aux caractéristiques du conduit vocal, ils sont largement utilisés en traitement de la parole.

La figure V.7 montre le diagramme des différentes étapes de calcul des coefficients LPCCs. Pour l'extraction des coefficients LPCCs, le signal de parole est segmenté en trames, ensuite la fenêtre de Hamming est appliquée à chaque trame puis une analyse d'auto-corrélation est effectuée pour chaque trame du signal fenêtré. L'analyse LPC est effectuée en utilisant l'algorithme de Durbin [211, 212] afin de convertir chaque séquence d'auto-corrélation en un ensemble de paramètres LPC (linear predicive coding). Finalement, on effectue la conversion des paramètres LPC en coefficients cepstraux en utilisant la relation suivante :

$$LPCC = LPC_i + \sum_{k=1}^{i-1}\left(\frac{k-i}{i}\right)LPCC_{i-k}LPC_k \tag{V.6}$$

Où $1 \leq i \leq p$, p étant l'ordre du prédicteur.

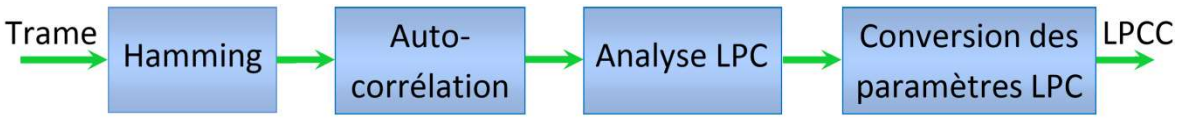


Figure V.7 : Schéma bloc de l'algorithme d'extraction des coefficients LPCCs.

V.3.2. Classification par modèles de Markov cachés

Un HMM permet de modéliser une unité acoustique ou une classe dans le cas de classification de la voix ou de la parole. Le HMM peut modéliser n'importe quelle unité acoustique (mot, phonème, etc.) selon l'application. Dans ce travail, un modèle HMM distinct a été utilisé pour chaque classe. On a opté pour un ensemble de 3 états pour chaque modèle HMM ce qui permet un compromis entre la précision et le temps de calcul.

Formellement, la modélisation HMM d'une unité acoustique peut être entièrement définie par l'ensemble des paramètres Λ, comme illustré dans la figure V.8 [213].

$\Lambda = (\pi, A, B)$ où π est la matrice de probabilité initiale, A est la matrice des transitions entre états et B est l'ensemble des probabilités d'émission des observations. Sachant que dans l'étude actuelle B est représenté par un ensemble de fonctions de densité de probabilité (PDF). Les PDFs sont la somme de gaussiennes connues sous le nom de modèles de mélange gaussiens (GMMs) définies comme [214] :

$$b_q(x_t) = \sum_{i=1}^{I} \omega_{q,i} \mathrm{N}(x_t, \mu_{q,i}, \Theta_{q,i}) \tag{V.7}$$

où x_t est la sortie au temps t, I est le nombre de composants du mélange, $\omega_{q,i}$ est le coefficient de mélange pour le $i^{ème}$ mélange de l'état $q^{ème}$ et N est une fonction de densité multivariée gaussienne :

$$\mathrm{N}(x_t, \mu_{q,i}, \Theta_{q,i}) = \frac{1}{\sqrt{(2\pi)^D \left|\Theta_{q,i}\right|}} \exp -\frac{1}{2}(x_t \mu_{q,i})^T \Theta_{q,i}^{-1}(x_t - \mu_{q,i}) \tag{V.8}$$

$\mu_{q,I}$: est le vecteur moyen du $i^{ème}$ mélange du $q^{ème}$ état.

$\Theta_{q,I}$: est la matrice de covariance pour le $i^{ème}$ mélange du $q^{ème}$ état.

D : est la dimension de l'espace vectoriel.

$|\Theta_{q,i}|$: est le déterminant de $\Theta_{q,i}$.

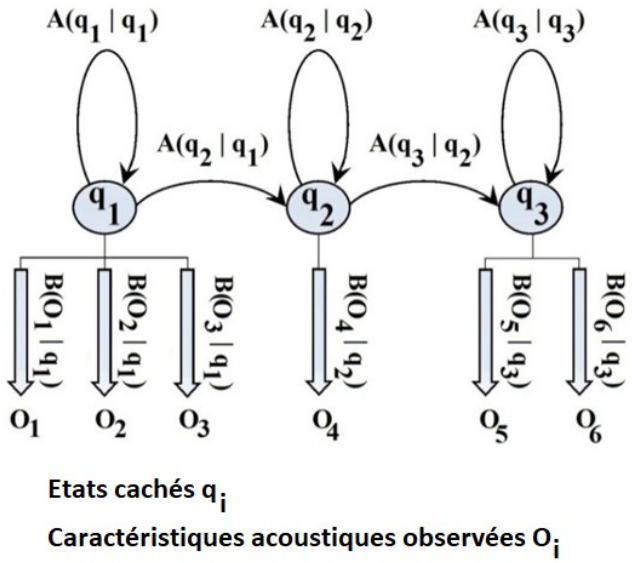


Figure V.8 : Modèle HMM à 3 états de gauche à droite, chaque état peut émettre une ou plusieurs observations O_i selon une probabilité gouvernée par B.

Pour la phase d'apprentissage du HMM, l'algorithme de Baum-Welch [213, 215] est utilisé. Le nombre d'itérations dans l'algorithme de Baum-Welch est de 5 avec 4 mélanges gaussiens (GMMs) pour chaque état. Ces valeurs ont été choisis afin d'atteindre une optimisation maximale en termes de précision et temps de calcul.

L'algorithme de Viterbi [186, 213, 216] est utilisé dans la phase test ou classification du HMM. Le but de cet algorithme est de trouver la séquence optimale d'états à partir d'un HMM et d'une séquence d'observations en suivant une perspective probabiliste. En d'autres termes, l'algorithme de Viterbi recherche la séquence d'états du HMM qui maximise la probabilité d'observations du HMM. Il crée un diagramme en treillis en plaçant les états $q(i)$ du HMM sur l'axe vertical et la séquence d'observations $O(t)$ sur l'axe horizontal, pour chaque nœud (i, t). L'algorithme de Viterbi calcule la probabilité $\delta_t(i)$, qui est la probabilité d'atteindre l'état i via une séquence optimale d'états, après avoir émis les t premières observations pour chaque nœud du diagramme à partir de la première colonne ($t = 1$).

V.3.3. Classification par séparateur à vaste marge

La plupart des problèmes de classification rencontrés en pratique sont des cas multi-classes. La décision n'est plus binaire et un seul hyperplan ne suffit pas. La méthode la plus simple et la plus ancienne utilisée dans la classification SVM multi-classes est la méthode un-contre-tous. Dans ce travail les deux noyaux linéaire et gaussien RBF (radial basis function) sont utilisés. Le noyau gaussien RBF peut être représenté mathématiquement comme :

$$K(x, y) = \exp\left(-\gamma \left|x - y\right|^2\right) \tag{V.9}$$

Où γ (largeur du noyau gaussien) est une constante à déterminer de façon empirique, elle contrôle la bande passante de la gaussienne. Plus γ est élevé, plus la gaussienne est étroite.

x et y sont des vecteurs dans l'espace d'entrée, c'est-à-dire des vecteurs de caractéristiques calculées à partir d'échantillons d'apprentissage et $|x\text{-}y|^2$ est la distance entre deux les deux points x et y.

Le tableau V.1 montre que les données des deux classes sont déséquilibrées, environ trois fois plus de stimuli sont dans la classe des voix perturbées que dans la classe modale. Le classificateur SVM permet de traiter ce problème en pondérant la constante de marge souple C pour les deux classes dans la fonction objective, qui devient en modifiant l'équation (III.47) :

$$\frac{1}{2}\|\mathrm{w}\|^2 + C_N \sum_{i=1}^{l_N} \xi_i + C_p \sum_{i=1}^{l_P} \xi_i \qquad \text{(V.10)}$$

Où w est le vecteur normal à l'hyperplan séparateur optimal (Chapitre III), l_N, l_P et C_N, C_P sont respectivement le nombre de stimuli et la constante de marge souple pour les classes modales et perturbées.

Comme il y a environ trois fois moins d'échantillons dans la classe modale, sa marge souple est pondérée par 3, soit $C_N = 3C$ et $C_P = C$, afin de ne pas favoriser les stimuli pathologiques.

Pour le cas multi-classes, la constante de marge souple C n'a pas été pondérée ($C_L = C_M = C_H = C$).

Le classificateur SVM, a été entrainé en prenant $\gamma = 101.5$ et $C = 1$. Les effets des paramètres γ et C sur la précision de la classification ont été étudiés dans la littérature [217, 218].

V.3.4. Classification hybride HMM-SVM

Le modèle hybride HMM-SVM est bien connu pour ses performances de généralisation et sa capacité à traiter des données volumineuses. L'idée clé du modèle hybride HMM-SVM est de convertir la sortie du SVM qui est une distance en une probabilité d'émission $b_i(V_k)$ d'état masqué puis de les injecter dans le modèle HMM [207, 170, 219].

La fonction décrivant le plan de séparation, mesure la distance de l'élément x_i à la marge. L'inclusion de l'élément x_i dans une classe dépend du signe ($f(x_i)$) et plus la distance à la marge est grande, plus la probabilité d'appartenir à la classe est grande [219].

La sortie du SVM, qui est une mesure de la distance entre les échantillons et l'hyperplan de classification, n'a aucune relation avec la probabilité *a posteriori* $p(y = +1 \mid x)$. Pour résoudre ce problème, ces deux variables sont combinées à l'aide de la fonction sigmoïde [207, 219, 220]:

$$\begin{aligned} p(y = +1 \mid x) &= g(f(x), A, B) \\ &\equiv \frac{1}{1 + \exp(A.f(x) + B)} \end{aligned} \qquad \text{(V.11)}$$

Le calcul des paramètres A et B consiste à construire l'ensemble d'apprentissage (p_i, t_i), dont $p_i = g(f(x), A, B)$, $t_i = (y_i + 1)/2$, on peut alors obtenir les valeurs de A et B via l'estimation

du maximum de vraisemblance, c'est-à-dire la fonction d'erreur d'entropie croisée minimisée :

$$-\sum t_i \log(p_i) + (1-t)\log(1-p_i) \tag{V.12}$$

En reliant les sorties de probabilité SVM aux états cachés correspondants du HMM, la probabilité d'émission $b_i(V_k)$ de l'état caché peut être représentée par la probabilité *a posteriori* $p(y = +1 \mid x)$. La progression de l'automate est représentée sur la figure V.9 [207, 213].

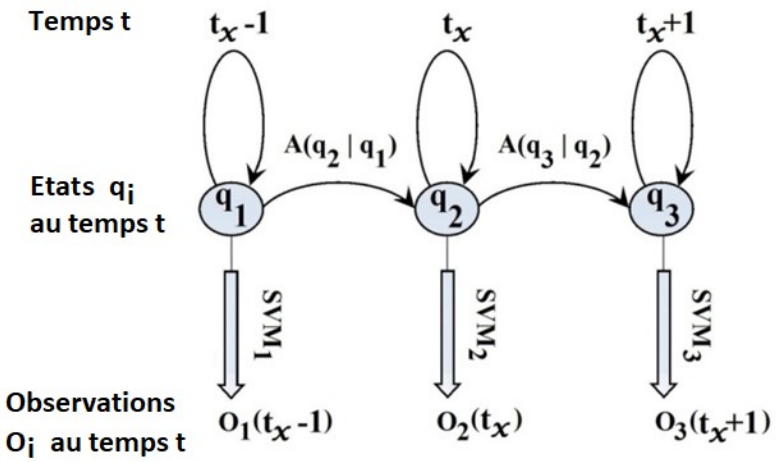


Figure V.9 : Schéma illustrant le remplacement des probabilités de sortie des états cachés q_i par les probabilités postérieures obtenues par le SVM.

En se basant sur les règles de Bayes [221], nous obtenons les définitions suivantes :

$P(class\ j \mid x_i) = \dfrac{P(x_i \mid class\ j)\,P(class\ j)}{P(x_i)}$: Probabilité *a posteriori*

P ($class\ j \mid x_i$) : Probabilité que le vecteur d'entrée x_i reçoit la classe du stimulus j

P ($class\ j$) : Probabilité précédente du stimulus j

P (x_i) : Probabilité *a priori* du vecteur acoustique x_i

V.3.4.1. Apprentissage

Le but de l'apprentissage du SVM est de connecter sa sortie aux états cachés du HMM en convertissant la sortie du SVM à l'aide de la fonction sigmoïde donnée par (V.11) en

probabilité d'émission des états cachés du HMM [219, 222], avec la formation d'un HMM pour chaque classe. Comme chaque état représente une partie temporelle du signal vocal, dans cette étude le nombre d'états cachés dans chaque HMM est fixé à 3. Nous considérons pour une classe que tous les états sont combinés sur le même HMM car le premier et le dernier état de chaque HMM n'a pas de transition vers un autre état. Par conséquent, 3 SVMs doivent être formés pour chaque classe, comme illustré à la Figure V.9 [222].
Construire un kit d'apprentissage est la première étape dans l'entrainement du SVM. Une fois que le vecteur de caractéristiques de chaque classe est obtenu en utilisant les coefficients MFCCs, LFPCs ou LPCCs, nous formons trois SVMs pour chaque classe de parole avec un SVM pour chaque état. Le SVM et le HMM sont entrainés séparément. Pour le HMM, la méthode classique d'apprentissage par l'algorithme de Baum-Welch est utilisée. La matrice d'état initial et la matrice de transition d'état sont initialisées de manière empirique. L'architecture générale du système hybride HMM-SVM est illustrée à la figure V.10 [207, 222].

V.3.4.2. Classification

Après avoir effectué l'apprentissage des HMM et SVM, nous ferons la classification d'un stimulus avec trois SVMs pour trouver les probabilités d'émission, chaque tranche temporelle concorde a un SVM_i différent (i=1, 2, 3). Sachant qu'à chaque état caché correspond un SVM_i, cette probabilité d'émission est injectée dans l'un des 3 états correspondants dans le HMM qui représente une classe [205, 216].
La probabilité d'émission de chaque état caché est remplacée par la sortie d'un SVM_i, comme le montre la figure V.9, où un SVM_i représente le SVM correspondant à l'état i du HMM.
On suit une approche probabiliste utilisant l'algorithme de Viterbi pour calculer la probabilité de la séquence d'observation $P(Q \mid \Lambda)$ pour chaque modèle HMM, en prenant en considération les paramètres MFCCs, LFPCs ou LPCCs (observations) de chaque stimulus et la probabilité *a posteriori* produite par le SVM comme résultat de classification.

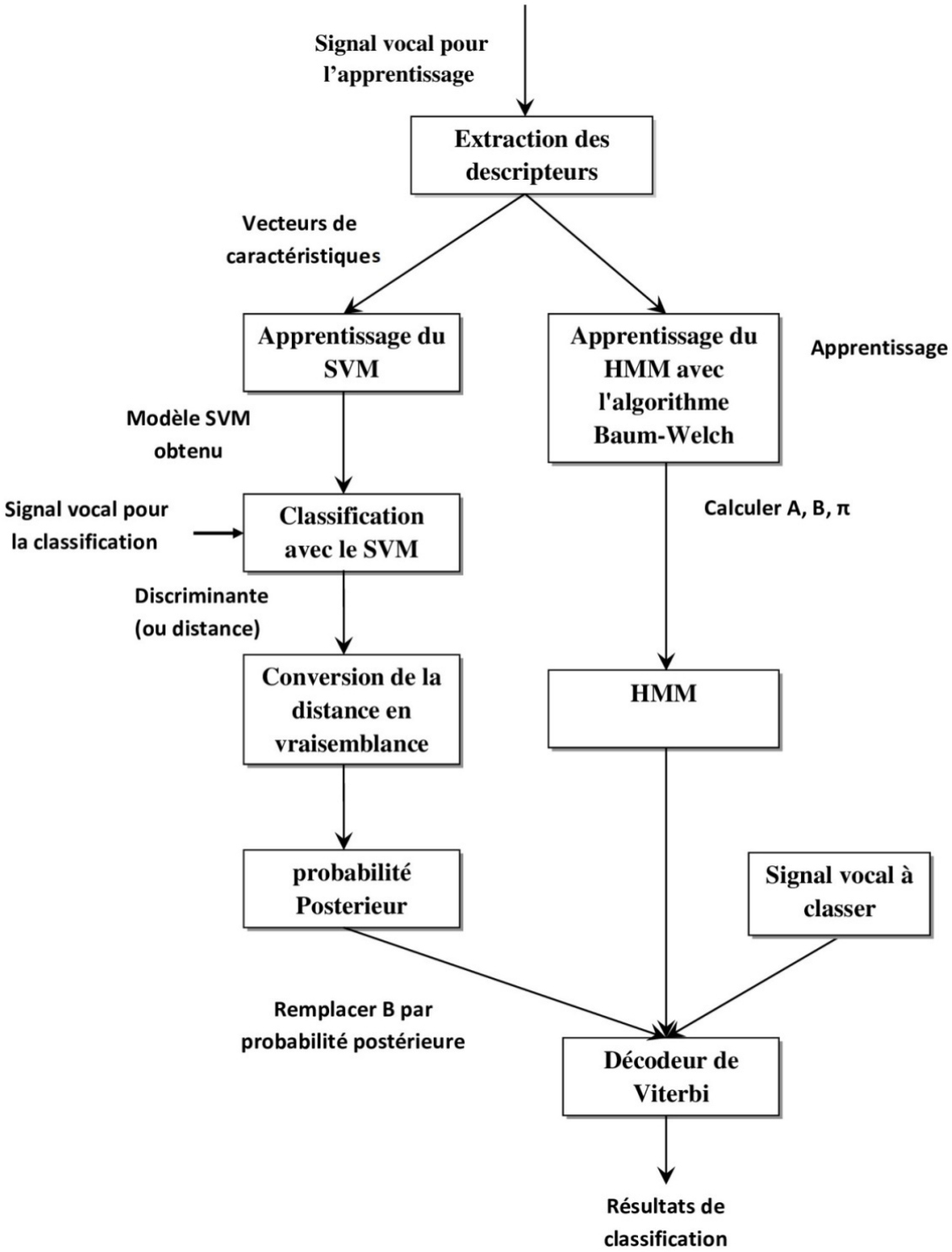


Figure V.10 : Architecture générale du système hybride HMM-SVM.

V.4. Critères d'évaluation des performances

Pour comparer les performances des différentes méthodes de classification, la précision, la sensibilité et la spécificité de la classification des voix modales et pathologique ont été utilisées comme critère de performance [223]. Ils sont calculés comme suit :

$$\text{Précision} = (TP + TN) / (TP + FP + TN + FN) \quad \textbf{(V.13)}$$

$$\text{Sensibilité} = TP / (TP + FN) \quad \textbf{(V.14)}$$

$$\text{Spécificité} = TN / (TN + FP) \quad \textbf{(V.15)}$$

Avec les notations suivantes :

TP : vrais positifs (true positive)

TN : vrais négatifs (true negative)

FP : faux positifs (false positive)

FN : faux négatifs (false negative)

V.5. Configuration expérimentale

La configuration informatique utilisée pour construire le classificateur proposé est un système basé sur un processeur Intel Core i5 3470 ($3^{ième}$ génération) avec une vitesse d'horloge de 3,2 GHz, 4 Go de mémoire DDR3. Le langage Matlab 2016 a été utilisé pour implémenter les modèles. Pour le corpus de voyelles [a], la meilleure performance en termes de temps de calcul (temps moyen) pour la classification des stimuli en utilisant les MFCCs comme descripteurs acoustiques est de 0.76 s pour la classification en deux classes et de 1.42 s pour la classification en trois classes comme indiqué dans le tableau V.3 [207]. Pour le corpus de concaténation des phrases suivies de la voyelle [a], la meilleure performance en termes de temps de calcul en utilisant les MFCCs comme descripteurs acoustiques est de 0.95 s pour la classification en deux classes et de 1.75 s pour la classification en trois classes comme indiqué dans le tableau V.4 [207]. La performance en termes de temps de calcul est influencée par le type de descripteurs et le nombre de classes.

Tableau V.3 : Temps de classification moyen des voyelles [a] en fonction du type de descripteurs et le nombre de classes.

Descripteurs acoustiques	Temps moyen de classification en deux classes (s)	Temps moyen de classification En trois classes (s)
MFCCs	0.76	1.42
LFPCs	1.32	1.59
LPCCs	1.49	1.64

Tableau V.4 : Temps de classification moyen des concaténations de phrases suivies de la voyelle [a] en fonction du type de descripteurs et le nombre de classes.

Descripteurs acoustiques	Temps moyen de classification en deux classes (s)	Temps moyen de classification En trois classes (s)
MFCCs	0.95	1.75
LFPCs	1.65	1.99
LPCCs	1.87	2.05

V.6. Résultats et discussion

- ***Effet de la longueur de trame***

L'effet de la longueur de trame sur les performances du classificateur hybride HMM-SVM en termes de précision a été analysé en variant la longueur de trame de 10 ms à 50 ms par pas de 10 ms. Les tableaux V.5 et V.6 montrent, respectivement, les précisions de classification HMM-SVM en deux et trois classes obtenues pour le corpus de parole continue pour différentes valeurs de la longueur de trame et différents descripteurs acoustiques [207, 213, 218]. Pour chaque type de descripteurs acoustiques, on utilise les 12 premiers coefficients comme entrée du classificateur. La précision la plus élevée est obtenue avec les coefficients MFCCs pour une longueur de trame de 30 ms. Les histogrammes des figures V.11 à V.14 montrent la performance de classification en deux et trois classes du classificateur hybride HMM-SVM pour une longueur de trame de 30 ms. Dans ce qui suit, la longueur de trame est fixée à sa valeur optimale de 30 ms.

Les performances du classificateur HMM-SVM en termes de précision, sensibilité et spécificité sont comparées à celles des classificateurs SVM et HMM utilisés séparément pour les corpus de voyelles [a] et de parole continue.

Tableau V.5 : Précision de classification en % du classificateur HMM-SVM en deux classes du corpus de la concaténation des phrases suivies de la voyelle [a] pour différentes valeurs de la longueur de trame et différents descripteurs acoustiques.

	Longueur de trame (ms)				
Descripteurs	10	20	30	40	50
MFCC	93.58	95.24	97.35	96.26	95.57
LFPC	91.75	93.33	94.95	92.78	90.21
LPCC	88.82	90.63	91.09	89.25	87.74

Tableau V.6 : Précision de classification en % du classificateur HMM-SVM en trois classes du corpus de parole continue pour différentes valeurs de la longueur de trame et différents descripteurs acoustiques.

	Longueur de trame (ms)				
Descripteurs	10	20	30	40	50
MFCC	90.68	91.14	92.01	91.35	90.46
LFPC	85.56	87.91	90.30	86.34	84.28
LPCC	83.67	84.49	86.94	85.35	82.29

- ***Classification en deux classes***

Les tableaux V.7 et V.8 montrent, respectivement, les résultats de classification en deux classes en utilisant les classificateurs SVM, HMM et HMM-SVM pour les corpus de voyelles [a] et de parole continue [207, 213, 218].

Tableau V.7 : Performances de classification en deux classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité en % pour le corpus de voyelles [a].

Méthode de classification		Descripteur	Précision	Sensibilité	Spécificité
SVM	Noyau linéaire	MFCC	**87.74**	**92.27**	**84.78**
		LFPC	85.52	89.96	82.44
		LPCC	84.96	88.08	82.12
	Noyau RBF	MFCC	**89.63**	**95.15**	**86.11**
		LFPC	87.74	92.27	84.78
		LPCC	86.52	91.46	84.44
HMM		MFCC	**94.26**	**95.23**	**88.22**
		LFPC	91.48	93.54	86.56
		LPCC	88.04	92.42	87.88
Hybride HMM-SVM		MFCC	**97.52**	**97.96**	**90.44**
		LFPC	95.96	95.58	89.12
		LPCC	92.08	93.34	88.76

Tableau V.8 : Performances de classification en deux classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité en % pour le corpus de parole continue.

Méthode de classification		Descripteur	Précision	Sensibilité	Spécificité
SVM	Noyau linéaire	MFCC	**75.22**	**78.76**	72.76
		LFPC	72.34	76.72	70.72
		LPCC	70.68	74.44	**75.44**
	Noyau RBF	MFCC	**77.36**	**79.94**	73.19
		LFPC	75.92	78.18	70.43
		LPCC	72.84	75.30	**75.86**
HMM		MFCC	**89.16**	**89.65**	81.27
		LFPC	81.52	86.05	78.19
		LPCC	79.04	85.71	**86.38**
Hybride HMM-SVM		MFCC	**97.35**	**98.73**	92.46
		LFPC	94.95	95.81	89.62
		LPCC	91.09	93.62	**95.24**

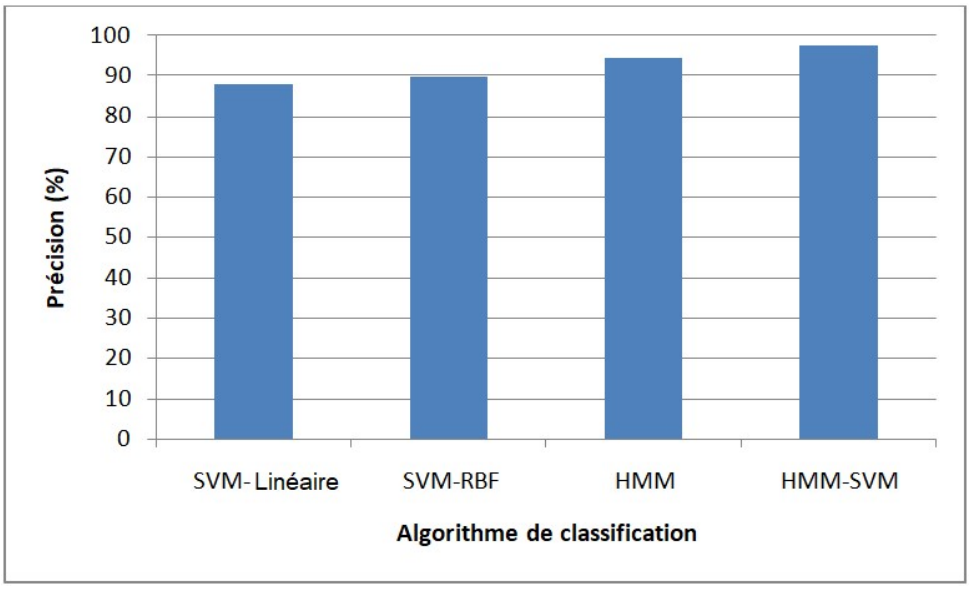


Figure V.11 : Comparaison des performances de classification des voyelles [a] en deux classes en termes de précision avec les MFCCs comme descripteurs acoustiques.

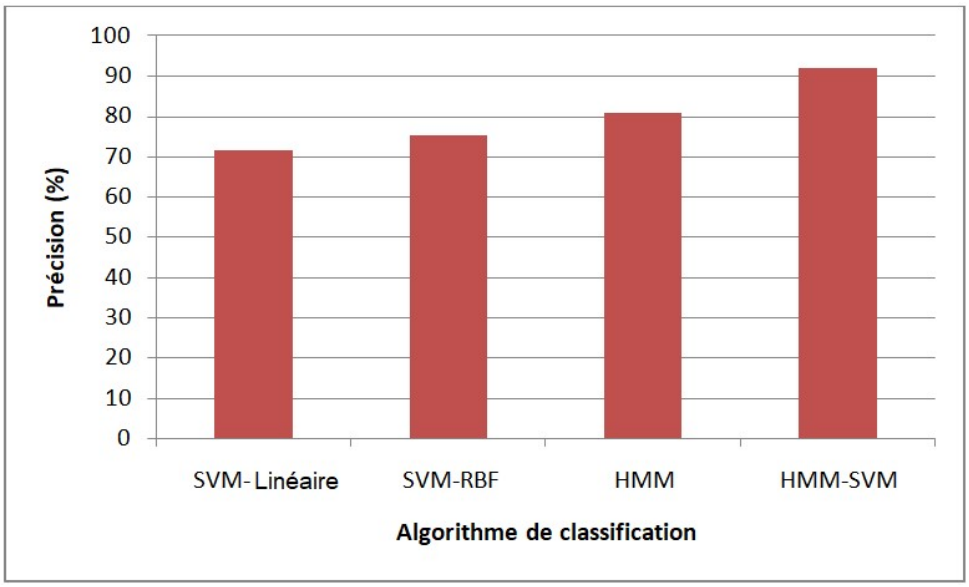


Figure V.12 : Comparaison des performances de classification de la parole continue en deux classes en termes de précision avec les MFCCs comme descripteurs acoustiques.

On observe dans les tableaux V.7 et V.8 que les meilleurs résultats sont obtenus en utilisant les coefficients MFCCs pour les trois classificateurs à l'exception de la spécificité pour le corpus de parole continue. Les meilleurs résultats en termes de précision de classification obtenus par l'algorithme hybride HMM-SVM sont respectivement de 97.52% et 97,35% pour les corpus de voyelles [a] et de parole continue. La précision du classificateur hybride HMM-SVM proposé dépasse jusqu'à 31 % la précision des

classificateurs HMM et SVM avec des noyaux linéaires et RBF. De même, les valeurs de la sensibilité et la spécificité du classificateur hybride HMM-SVM sont supérieures à celles des classificateurs SVM et HMM aussi bien pour le corpus de parole continue que pour le corpus des voyelles [a].

Le classificateur HMM-SVM présente la même performance en termes de précision pour les corpus de voyelles [a] et de parole continue mais il résulte en une meilleure performance en termes de sensibilité et spécificité pour le corpus de parole continue que pour le corpus des voyelles [a].

- ***Classification en trois classes***

Les tableaux V.9 et V.10 montrent, respectivement, les résultats de classification en trois classes pour les corpus de voyelles [a] et de parole continue.

Les résultats présentés dans les tableaux V.9 et V.10 indiquent que les coefficients MFCCs permettent d'atteindre les meilleures performances pour les trois classificateurs à l'exception de la spécificité pour le corpus de concaténation des phrases suivies de la voyelle [a] [207, 213, 218]. Les meilleurs résultats en termes de précision de classification obtenus par l'algorithme hybride HMM-SVM sont respectivement de 93.92% et 92.01% pour les corpus de voyelles [a] et de parole continue. La précision du classificateur hybride HMM-SVM dépasse la précision des classificateurs HMM et SVM avec des noyaux linéaires et RBF. De même, les valeurs de la sensibilité et la spécificité du classificateur hybride HMM-SVM sont supérieures à celles des classificateurs SVM et HMM aussi bien pour le corpus de parole continue que pour le corpus des voyelles [a].

Le classificateur HMM-SVM présente la même performance en termes de précision pour les corpus de voyelles [a] et de parole continue mais il résulte en une meilleure performance en termes de sensibilité et spécificité pour le corpus de parole continue que pour le corpus de voyelles [a].

Tableau V.9 : Performances de classification en trois classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité en % pour le corpus de voyelles [a].

Méthode de classification		Descripteur	Précision	Sensibilité	Spécificité
SVM	Noyau linéaire	MFCC	**71.04**	**74.92**	**68.88**
		LFPC	69.92	73.16	68.24
		LPCC	68.16	72.68	66.52
	Noyau RBF	MFCC	**72.48**	**76.04**	**70.56**
		LFPC	71.04	74.92	69.88
		LPCC	69.92	74.16	67.24
HMM		MFCC	**81.96**	**77.58**	**72.12**
		LFPC	80.08	76.84	69.76
		LPCC	78.84	75.32	69.48
Hybride HMM-SVM		MFCC	**93.92**	**79.16**	**73.24**
		LFPC	90.16	77.18	71.52
		LPCC	89.68	76.14	70.96

Tableau V.10 : Performances de classification en trois classes des classificateurs HMM, SVM et HMM-SVM en termes de précision, sensibilité et spécificité en % pour le corpus de parole continue.

Méthode de classification		Descripteur	Précision	Sensibilité	Spécificité
SVM	Noyau linéaire	MFCC	**71.33**	**75.41**	67.97
		LFPC	70.01	73.77	65.09
		LPCC	67.02	71.54	**72.18**
	Noyau RBF	MFCC	**75.15**	**78.67**	70.94
		LFPC	73.55	76.99	68.18
		LPCC	70.41	73.98	**74.30**
HMM		MFCC	**80.69**	**81.49**	78.61
		LFPC	80.20	80.53	76.17
		LPCC	76.86	78.06	**79.34**
Hybride HMM-SVM		MFCC	**92.01**	**95.87**	88.13
		LFPC	90.30	92.39	85.61
		LPCC	86.94	90.78	**91.22**

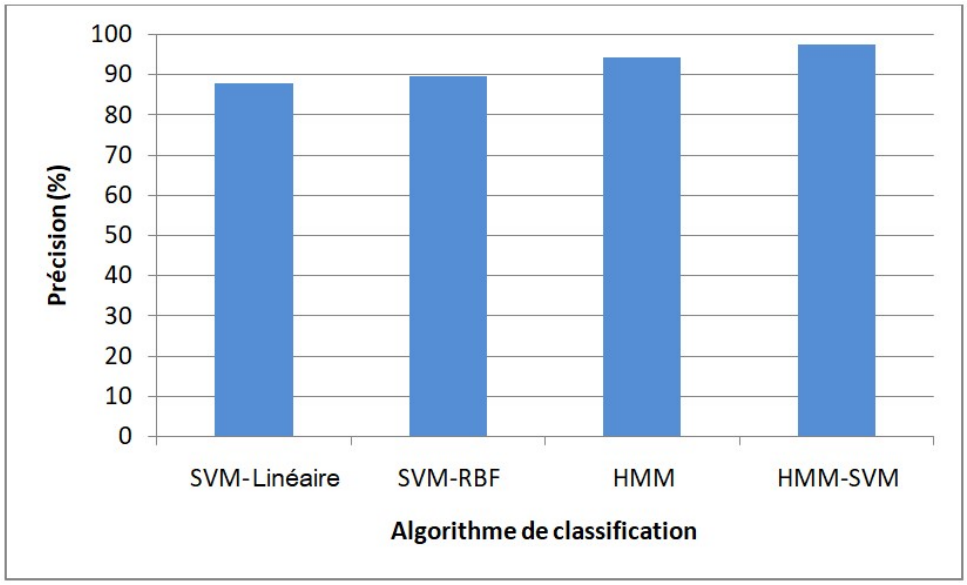


Figure V.13 : Comparaison des performances de classification des voyelles [a] en trois classes en termes de précision avec les MFCCs comme descripteurs acoustiques.

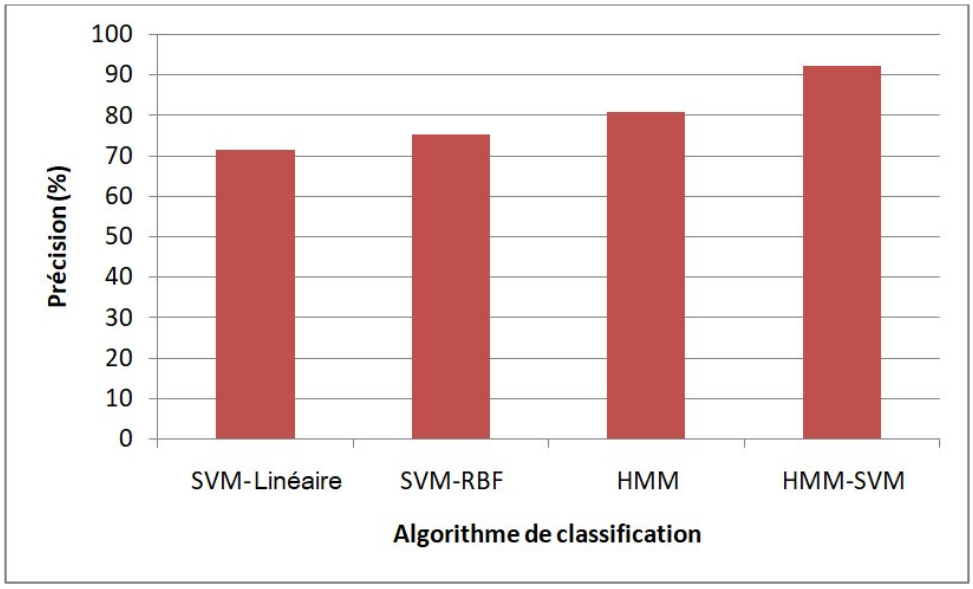


Figure V.14 : Comparaison des performances de classification de la parole continue en trois classes en termes de précision avec les MFCCs comme descripteurs acoustiques.

V.7. Conclusion

Les performances de la classification en deux classes et en trois classes des voix pathologiques basée sur les classificateurs SVM, HMM et HMM-SVM hybride ont été analysées sur deux corpus de voyelles [a] soutenues et de parole continue produites par des locuteurs normophoniques et dysphoniques. Les coefficients MFCCs, LFPCs et LPCC ont été utilisés comme entrées des classificateurs. Les meilleures performances ont été

obtenues avec les coefficients MFCCs pour les trois classificateurs. Les résultats expérimentaux montrent que le classificateur hybride HMM-SVM surpasse les classificateurs HMM et SVM avec des noyaux linéaires et RBF en termes de précision de classification, de sensibilité et de spécificité. La précision du classificateur hybride HMM-SVM proposé dépasse jusqu'à 32% la précision des classificateurs HMM et SVM avec des noyaux linéaires et RBF, ce qui est conforme avec les résultats rapportés dans la littérature dans le cadre de la reconnaissance automatique de la parole. Le classificateur hybride HMM-SVM présente des meilleures performances pour la classification de la parole continue que pour les voyelles [a] soutenues.

Conclusion générale

Conclusion générale

Dans cette thèse, on a proposé une méthode hybride combinant simultanément les HMMs et les SVMs pour évaluer et classer les voix pathologiques selon leurs degrés de perturbation en utilisant des descripteurs acoustiques obtenus à partir de la parole continue.
Les HMMs sont des automates de classification statistique utilisant les règles de Bayes pour calculer la probabilité d'appartenance aux différentes classes d'une voix cible. Ils sont très performants et particulièrement bien adapté pour la classification de la parole continue. Etant donné leur nature, les HMMs permettant de suivre les variations des observations en fonction du temps.
Les SVMs sont des méthodes de classification à apprentissage automatique supervisé, ils permettant la séparation des données de grande dimension par un hyperplan. Dans le cas linéaire, cet hyperplan est une droite, mais dans le cas non linéaire, la fonction noyau définit un espace hilbertien, dit auto-reproduisant et isométrique par la transformation non linéaire de l'espace initial et dans lequel la séparation devient linéaire. Ce noyau confère une grande flexibilité pour s'adapter à des situations de données non séparables.
L'utilisation des HMMs et SVMs est justifiée par leurs bonnes performances dans le domaine de la reconnaissance de la parole et de locuteur. Les coefficients MFCCs, LPCCs et LFPCs ont été utilisés comme descripteurs acoustiques.
Les performances de la classification en deux classes et en trois classes des voix pathologiques basée sur les classificateurs SVM, HMM et HMM-SVM hybride ont été analysées sur deux corpus comprenant des voyelles [a] soutenues et des concaténations de deux phrases suivies de la voyelle [a] produites par des locuteurs normophoniques et dysphoniques. Les meilleures performances ont été obtenues avec les coefficients MFCCs pour les trois classificateurs. Les résultats expérimentaux montrent que le classificateur hybride HMM-SVM surpasse les classificateurs HMM et SVM en termes de taux de précision de classification, de sensibilité et de spécificité aussi bien pour le corpus de concaténations des phrases suivies de la voyelle [a] que pour le corpus des voyelles [a].
Le classificateur hybride HMM-SVM présente des meilleures performances pour la classification de la parole continue que pour les voyelles soutenues. En effet, il est bien établi dans la littérature d'évaluation des voix pathologique que la parole continue peut être plus informative que les voyelles soutenues en termes de précision d'évaluation du degré d'enrouement si un indice acoustique adéquat est utilisé comme mesure objective. Les expériences réalisées ont montré qu'on obtient des meilleures performances en termes de

précision, sensibilité et spécificité pour le corpus de concaténations de deux phrases suivies de la voyelle [a] lorsque les coefficients MFCCs sont utilisés comme descripteurs acoustiques. En outre, le modèle hybride HMM-SVM profite des avantages des SVMs et des HMMs ce qui permet d'améliorer la précision de classification de façon significative, par rapport aux méthodes de classification utilisant les modèles SVMs et HMMs séparément.

Comme suite à ce travail, le modèle hybride HMM-SVM peut être utilisé pour la classification des voix pathologiques selon d'autres scores de l'évaluation perceptive tels que la raucité et le souffle. De même, le modèle hybride HMM-SVM peut être testé sur d'autres bases de données de la parole continue de plus grandes tailles. La méthode de classification à base du modèle hybride HMM-SVM peut être aussi proposée pour la classification de la parole dysarthrique pour l'évaluation des troubles de la parole.

Bibliographie

Bibliographie

[1] Y. Liu, M. Wu, “The Influence of Reverberation on the Perceptual Judgment of Cross-Lingual Speakers’ Timbre,” *IEEE/ACIS 19th International Conference on Computer and Information Science (ICIS)*, pp. 116-120, 2021.

[2] J. Hillenbrand, “Perception of aperiodicities in synthetically generated voices”, *J. Acoust. Soc. Amer.*, Vol. 83, N° 6, pp. 2361-2371, 1988.

[3] J. Hillenbrand, R. A. Clevand, R. L. Erickson, “Acoustic correlates of breathy vocal quality”, *J. Speech, Hear. Res.*, Vol. 30, pp. 448-461, 1987.

[4] V. Parsa, D. G. Jamieson, “Identification of pathological voices using glottal noise measures”, *J. Speech, Lang., Hear. Res.*, Vol. 43, pp. 469-485, 2000.

[5] L. Eskenazi, D. G. Childers, D. M. Hicks, “Acoustic correlates of vocal quality”, *J. Speech, Hear. Res.*, Vol. 33, pp. 298-306, 1990.

[6] A. Kacha, F. Grenez, and Schoentgen, J. “Estimation of dysperiodicities in disordered speech,” *Speech Communication*, Vol. 48, N° 10, pp. 1365-1378, 2006.

[7] A. Alpan, Y. Maryn, A. Kacha, F. Grenez and J. Schoentgen. “Multi-band dysperiodicity analyses of disordered connected speech,” *Speech Communication*, Vol. 53, N° 1, pp. 131-141, 2011.

[8] J. Schoentgen, “Spectral models of additive and modulation noise in speech and phonatory excitation signals,” *J. Acoust. Soc. Amer.*, Vol. 113, N° 1, pp. 553-562, 2003.

[9] F. Klingholz, “Acoustic recognition of voice disorders : A comparative study of running speech versus sustained vowels,” *J. Acoust. Soc. Amer.*, Vol. 87, N° 5, pp. 2218-2224, 1990.

[10] J. I. Godino-Llorente, P. Gómez-Vilda, and T. Lee, “Analysis and signal processing of oesophageal and pathological voices,” *EURASIP Journal on Advances in Signal Processing*, *Special Issue on Analysisand Signal Processing of Oesophageal and Pathological Voices*, 283504, 2009.

[11] L. Li, T. Watzel, L. Kürzinger and G. Rigoll, “Towards Constructing HMM Structure for Speech Recognition With Deep Neural Fenonic Baseform Growing,” *IEEE Access*, Vol. 9, pp. 39098-39110, 2021.

[12] L. Fan, “Audio Example Recognition and Retrieval Based on Geometric Incremental Learning Support Vector Machine System,” *IEEE Access*, Vol. 8, pp. 78630-78638, 2020.

[13] C. Peng, W. Chen, X. Zhu, B. Wan and D. Wei, “Pathological Voice Classification Based on a Single Vowel's Acoustic Features,” in *Proc. 7th IEEE International Conference on Computer and Information Technology (CIT 2007)*, pp. 1106-1110, 2007.

[14] A. Al-Nasheri et al., “Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions,” *IEEE Access*, Vol. 6, pp. 6961-6974, 2018.

[15] A. Esposito, N. Campbell, C. Vogel, A. Hussain and A.Nijholt, “Development of Multimodal Interfaces: Active Listening and Synchrony,” *Lecture Notes in Computer Science, Springer, Berlin, Heidelberg*.Vol 5967, 2010.

[16] V. Parsa, D. G. Jamieson, “Acoustic discrimination of pathological voice : Sustained vowels versus continuous speech,” *Journal of Speech, Language, and Hearing Research*, Vol. 44, pp. 327-339, 2001.

[17] H. T.Cordeiro, J. M. Fonseca and C. M. Ribeiro, “Reinke's Edema and Nodules Identification in Vowels Using Spectral Features and Pitch Jitter,” *Procedia Technology, Elsevier*, Vol. 17, pp. 202-208, 2014.

[18] G. Muhammad, M. Alsulaiman, A. Mahmood and Z. Ali, “Automatic voice disorder classification using vowel formants,” in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6, 2011.

[19] M. Zakariah, B. Reshma, Y. A. Alotaibi, Y. Guo, K. T. Trung and M. M. Elahi, “An Analytical Study of Speech Pathology Detection Based on MFCC and Deep Neural Networks,” *Hindawi*, 2022.

[20] K. Umapathy, S. Krishnan, V. Parsa and D. G. Jamieson,“Discrimination of pathological voices using a time-frequency approach,” *IEEE Transactions on Biomedical Engineering*, Vol. 52, N° 3, pp. 421-430, 2005.

[21] M. Markaki, Y. Stylianou,“Voice pathology detection and discrimination based on modulation spectral features,” IEEE *Transactions on Audio, Speech, and Language Processing*, Vol. 19, N° 7, pp. 1938-1948, 2011.

[22] E. Moore, M. A. Clements, J. W. Peifer and L. Weisser, “Critical analysis of the impact of glottal features in the classification of clinical depression in speech,” *IEEE Transactions on Biomedical Engineering*, Vol. 55, N° 1, pp. 96-107, 2008.

[23] P. Saidi, F. Almasganj, “Voice disorder signal classification usingm-band wavelets and support vector machine,” *Circuits Syst. Signal Process*., Vol. 34, pp. 2727-2738, 2015.

[24] V. Majidnezhad, “A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis,” *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 3, pp. 1-11, 2015.

[25] A. Al-nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki et al.,“Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions,” *IEEE Access*, Vol. 6, pp. 6961-6974, 2017.

[26] Tulics M G, Vicsi K. “The automatic assessment of the severity of dysphonia,” *International Journal of Speech Technology*,” Vol. 22, pp. 341-350, 2019.

[27] J. P. Haton, C. Cerisara, D. Fohr, Y. Laprie, K. Smaïli, Reconnaissance automatique de la parole du signal à son interprétation, DUNOD, 2006.

[28] S. Lehrer, Understanding Lung Sounds: Third Edition, W.B. Saunders Company, 2018.

[29] F. Le Huche, A. Allali, LA VOIX, Anatomie et physiologie des organes de la voix et de la parole , Tome 1, 4eme édition, Elsevier Masson SAS, 2010.

[30] M. Daumet, Elaboration de profils types en fonction de pathologies vocales à partir de critères d’analyse objectifs, par le logiciel vocalab, mémoire présenté pour l’obtention du certificat de capacité d’orthophoniste, université nice sophia antipolis, faculté de médecine, école d’orthophonie, 2015.

[31] J. C. Chen, Elements of human voice, World scientific publishing co, 2016.

[32] S. de Corbiere, E. Fresnel, C.Freche, LA VOIX: La corde vocale et sa pathologie, CCA, 2001.

[33] T. Dubuisson, Glottal Source Estimation and Automatic Detection of Dysphonic Speakers, PHD Thesis, University of Mons Doctoral School Musics Signal Processing, 2011.

[34] J.G. Švec, H.K. Schutte, C.J. Chen, I.R. Titze, “Integrative Insights into the Myoelastic-Aerodynamic Theory and Acoustics of Phonation. Scientific Tribute to Donald G. Miller,”*Journal of Voice*, 2021.

[35] M. Atkinson, S. McHanwell, Basic medical science for speech and language therapy students, Whurr Publishers, 2002.

[36] G. Fant, Acoustic theory of speech production, S-Gravenhage, 1960.

[37] T. B. Patel, H. A. Patil, “Significance of Source–Filter Interaction for Classification of Natural vs. Spoofed Speech,” *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, N° 4, pp. 644-659, 2017.

[38] J. Urbain, T. Dubuisson, S. Dupont, C. Frisson, R. Sebbe and N. D'Alessandro, "AudioCycle: a Similarity-Based Visualization of Musical Libraries," *in Proc. of ICME*, pp.1847-1848, 2009.

[39] G. Degottex, Glottal Source and Vocal-Tract Separation,PhD thesis, Université Pierre et Marie Curie, 2010.

[40] M. Boudjerda, Analyse du signal de parole pour l'évaluation automatique des voix pathologiques, Thèse de Doctorat en Sciences, Université Mohammed Seddik Benyahia, Jijel, 2018.

[41] G. Fant, J. Liljencrants and Q. Lin, "A four-parameter model of glottal flow," *STLQPSR*, Vol. 4, pp. 1-13, 1985.

[42] D. Klatt, L. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acous. Soc. Am*., Vol. 87, N° 2, pp. 820-857, 1990.

[43] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acous. Soc. Am.*, Vol. 49, N° 2, pp. 583-590, 1971.

[44] R. Veldhuis, "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation," *J. Acous. Soc. Am.*, Vol. 103, N° 1, pp. 566-571, 1990.

[45] P. Gómez-Vilda, R. Martínez, F. Díaz, C. Lázaro, A. Alvarez, V. Rodellar and V. Nieto, "Estimation of vocal cord biomechanical parameters by non linear inverse filtering of voice," *In: Proc*. of the 3rd Int. Conf. on Non-Linear Speech Processing (NOLISP'05), pp. 174–183, Barcelona, Spain, 2005.

[46] S. Fraj, Synthèse des voix pathologiques, Thèse de Doctorat, Université Libre de Bruxelles, 2010.

[47] B. Doval, C. D'Alessandro and N. Henrich, "The spectrum of glottal flow models," *Acta Acustica United With Acust.*, Vol. 92, N° 92, pp. 1026-1046, 2006.

[48] S. Sahoo, A. Routray, "A novel method of glottal inverse filtering," *IEEE/ACM Trans. Audio Speech Lang. Process*., Vol. 24, N° 7, pp. 1230-1241, 2016.

[49] J. B. Alonso, M. A. Ferrer, P. Henríquez, K. López-de-Ipina, J. Cabrera and C. M. Travieso, "A study of glottal excitation synthesizers for different voice qualities," Neurocomputing, Vol. 150, pp. 367-376, 2015.

[50] M. Atkinson, S. mchanwell, Basic Medical Science for Speech and Language Therapy Students, 2nd edition, Guildford, England, J & R Press, 2018.

[51] M. Remacle, Surgery of Larynx and Trachea, Springer, 2010.

[52] M. Anniko, M. Bernal-Sprekelsen, V. Bonkowsky, P. Bradley and S. Iurato, Otorhinolaryngology, Head and Neck Surgery. European Manual of Medicine, Springer, Berlin, Heidelberg, 2010.

[53] F. Le Huche, A. Attali. Pathologies vocales d'origine fonctionnelle, Tome 2, 3eme edition, Elsevier Masson SAS, 2010.

[54] W. Maetzler, R. Dodel, A. Jacobs, Neurogeriatrie. Springer, Berlin, 2019.

[55] 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.

[56] G. Pouchoulin, Approche Statistique pour l'Analyse Objective et la Caractérisation de la Voix Dysphonique, Thèse de doctorat, Académie d'aix-Marseille université d'Avignon et des pays de Vaucluse, 2008.

[57] J. S. Lee, E. Kim, M.W. Sung, et al. "A method for assessing the regional vibratory pattern of vocal folds by analysing the video recording of stroboscopy," *Med. Biol. Eng. Comput*. Vol. 39, pp. 273-278, 2001.

[58] P. Carding, M. Bos-Clark, S. Fu, P. Gillivan-Murphy, S.M. Jones and C. Walton, "Evaluating the efficacy of voice therapy for functional, organic and neurological voice disorders," *Clinical Otolaryngology*, *John Wiley & Sons Ltd*, 2016.

[59] M.G. Tulics, K. Vicsi, "The automatic assessment of the severity of dysphonia," *Int. J. Speech Technol.*, Vol. 22, pp. 341-350, 2019.

[60] M. Heikkinen, E. Penttilä, M. Qvarnström, et al. "Perceptual Assessment and Acoustic Voice Analysis as Screening Tests for Vocal Fold Paresis After Thyroid or Parathyroid Surgery." *World J. Surg.,Vol.* 45, pp. 765-773, 2021.

[61] J. McMurray, M. Hoffman, M. Braden, Multidisciplinary Management of Pediatric Voice and Swallowing Disorders, Springer, 2020.

[62] Clark A. Rosen, C. Blake Simpson, Operative Techniques in Laryngology, Springer, 2008.

[63] E. Shohamy, G. Iair, S. May, Language Testing and Assessment, Springer, 2017.

[64] C. McMullin, Transcription and Qualitative Methods: Implications for Third Sector Research. Voluntas, 2021.

[65] M. Hirano, Psycho-acoustic evaluation of voice : GRBAS scale for evaluating the hoarse voice. Clinical Examination of voice, Springer Verlag, 1981.

[66] 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," *Journal of Voice*, Vol. 1, pp 74-80. 1997.

[67] K. Johansson, S. Strömbergsson, C. Robieux, A. McAllister, “Perceptual Detection of Subtle Dysphonic Traits in Individuals with Cervical Spinal Cord Injury Using an Audience Response Systems Approach,” *Journal of Voice*, Vol. 31, N° 1, pp. 1-11, 2017.

[68] A. Zehnhoff-Dinnesen, B. Wiskirska-Woznica, K. Neumann, T. Nawka, Disorders of Language and Hearing Development, Springer Berlin Heidelberg, 2020.

[69] J. Kreiman, B. R. Gerratt, K. Precoda, and G. S. Berke, “Individual differences in voice quality perception,” *Journal of Speech and Hearing Research,* Vol. 35, pp 512-520, 1992.

[70] L. Anders, H. Hollien, P. Hurme, A. Sonninnen and J. Wendler, “Perceptual evaluation of hoarseness by several classes of listeners,” *Folia Phoniatrica et Logopaedica*, Vol. 40, pp. 91-100, 1988.

[71] J. Revis, L’analyse perceptive des dysphonies : approche phonétique de l’évaluation vocale. Thèse de Doctorat, Université de la Méditerranée, 2004.

[72] A. H. Poorjam, M. A. Little, J. R. Jensen and M. G. Christensen, “A ParametricApproach for Classification of Distortions in PathologicalVoices,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 286-290, 2018.

[73] M. G. Yildiz, S. Sagiroglu, N. Bilal, I. Kara, I. Orhan and A. Doganer, “Assessment of Subjective and Objective Voice Analysis According to Types of Sulcus Vocalis,” *Journal of Voice*, 2021.

[74] A. Darouie, M. Aghajanzadeh, P. Dabirmoghaddam, A. Salehi and M. Rahgozar, “The Design and Assessment of a Multiparametric Model for the Dysphonia Severity Index for Persian-speaking Populations,” *Journal of Voice*, Vol. 33, N° 2, pp. 226-231, 2019.

[75] J. Li, M. Hasegawa-Johnson, N. L. McElwain, “Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations,” *Speech Communication*, Vol. 133, pp. 41-61, 2021.

[76] A. Castellana, A. Carullo, S. Corbellini and A. Astolfi, “Discriminating Pathological Voice From Healthy Voice Using Cepstral Peak Prominence Smoothed Distribution in Sustained Vowel,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 67, N° 3, pp. 646-654, 2018.

[77] D. L. Poole, A.K., Mackworth, Artificial Intelligence : Foundations of Computational Agents, Cambridge University Press, 2nd édition, 2017.

[78] P. Kulkarni, P. Joshi, Artificial intelligence : building intelligent systems, PHI Learning, 2015.

[79] V. Chandra, A. Hareendran, Artificial intelligence and machine learning, PHI Learning, 2014.

[80] S. Zhongzhi, Advanced Artificial Intelligence, Series on intelligence science, Volume 1, World Scientific, 2011.

[81] W.S. McCulloch, W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133, 1943.

[82] A. Turing, "Computer machinery and intelligence," *Mind*, Vol 59, pp. 433-460. 1950.

[83] J. McCarthy, M. Minsky, N. Rochester and C. Shannon, Proposal for the Darmouth summer research project on Artificial Intelligence, Technical report, Darmouth College. 1955.

[84] R. S. Michalski, J. G. Carbonell, T. Mitchell and Y. Kodratoff, Apprentissage symbolique : une approche de l'intelligence artificielle, Tome 1 et 2, Editions Cépadues, Toulouse, 1993.

[85] S. R. Deb, Robotics Technology and Flexible Automation Paperback, MC Graw Hill India, 2nd edition, 2013.

[86] M. Kubat, An Introduction to Machine Learning, Springer, 2017.

[87] C. A. Azencott, Introduction au Machine Learning, DUNOD, 2018.

[88] A. El iraoui, Apprentissage Automatique : Apprentissage Automatique Broché, Éditions universitaires européennes, 2018.

[89] B. Boser, I. Guyon and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. of the fifth annual workshop on Computational learning theory COLT '92*, pp. 144-152, 1992.

[90] C. Cortes, V. Vapnik, "Support-vector networks," *Machine Learning*, Vol. 20, pp. 273-297, 1995.

[91] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., New York : Springer Press, 2000.

[92] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discoveryn*," Vol. 2, pp. 121-167, 1998.

[93] D. Pawar, S. Dhage, “Multiclass covert speech classification using extreme learning machine,” *Biomed. Eng. Lett.*, Vol. 10, pp. 217-226, 2020.

[94] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, “Efficient kNN Classification With Different Numbers of Nearest Neighbors,” *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 29, N° 5, pp. 1774-1785, 2018.

[95] W. Xing, Y. Bei, “Medical Health Big Data Classification Based on KNN Classification Algorithm,” *IEEE Access*, Vol. 8, pp. 28808-28819, 2020.

[96] D. Anggraeni, W. S. M. Sanjaya, M. Munawwaroh, M. Y. S. Nurasyidiek and I. P. Santika, “Control of robot arm based on speech recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and K-Nearest Neighbors (KNN) method,” in *Proc. International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation* (*ICAMIMIA*), 2017.

[97] A. R. Webb, K. D. Copsey, Statistical Pattern Recognition, Wiley, 2011.

[98] A. Sanchis, A. Juan and E. Vidal, “A Word-Based Naïve Bayes Classifier for Confidence Estimation in Speech Recognition,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, N° 2, pp. 565-574, 2012.

[99] S. Hamidi Ghalehjegh, R. C. Rose, “Linear Regression Based Acoustic Adaptation for the Subspace Gaussian Mixture Model,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, N° 9, pp. 1391-1402, 2014.

[100] Y. Hu, D. Wu and A. Nucci, “Fuzzy-Clustering-Based Decision Tree Approach for Large Population Speaker Identification,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, N° 4, pp. 762-774, 2013.

[101] V. M. Lelis, E. Guzmán and M. Belmonte, “Non-Invasive Meningitis Diagnosis Using Decision Trees,” *IEEE Access*, Vol. 8, pp. 18394-18407, 2020.

[102] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, “Classification and Regression Trees,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery , Vol. 1, N° 1, pp. 14 -23, 1984.

[103] B. Ban, J. Yang, P. Chen, J. Xiong and Q. Wang, “Ship Track Regression Based on Support Vector Machine,” *IEEE Access*, Vol. 5, pp. 18836-18846, 2017.

[104] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[105] A. J. Ochoa Reyes, A. O. Garcia and Y. L. Mue, “System for Processing and Analysis of Information Using Clustering Technique,” *IEEE Latin America Transactions*, Vol. 12, N° 2, pp. 364-371, 2014.

[106] K. P. Sinaga, M. Yang, “Unsupervised K-Means Clustering Algorithm,” *IEEE Access*, Vol. 8, pp. 80716-80727, 2020.

[107] V. P. Murugesan, P. Murugesan, “A new initialization and performance measure for the rough k-means clustering,” *Soft Comput.,* Vol. 24, pp. 11605-11619, 2020.

[108] E. Warns-Petit, E. Morignat, M. Artois, et al., “Unsupervised clustering of wildlife necropsy data for syndromic surveillance,” *BMC Vet. Res*., Vol. 6, N° 56, 2010.

[109] L. E. Baum, J.A. Egon, “An inequality with applications to statistical estimation for probabilistic functions of a markov process and to a model for ecology,” *Bull. Am. Meteorol*., Vol. 73, pp. 360-363, pp. 1967.

[110] M. Gogebakan, “A Novel Approach for Gaussian Mixture Model Clustering Based on Soft Computing Method,” *IEEE Access*, Vol. 9, pp. 159987-160003, 2021.

[111] Y. S. Popkov, Y.A.Dubnov, and A.Y. “Popkov, Entropy Dimension Reduction Method for Randomized Machine Learning Problems,” *Autom. Remote Control*, Vol. 79, pp. 2038-2051, 2018.

[112] G. R. Naik, Advances in Principal Component Analysis: Research and Development, Springer Singapore, 2018.

[113] H. Roopa and T. Asha, “A Linear Model Based on Principal Component Analysis for Disease Prediction,” *IEEE Access*, Vol. 7, pp. 105314-105318, 2019.

[114] U. Stańczyk, B. Zielosko, L. C. Jain, Advances in Feature Selection for Data and Pattern Recognition, Springer, 2017.

[115] N. Almusallam, Z. Tari, J. Chan, A. Fahad, A. Alabdulatif and M. Al-Naeem, “Towards an Unsupervised Feature Selection Method for Effective Dynamic Features,” *IEEE Access*, Vol. 9, pp. 77149-77163, 2021.

[116] Y. Wu, C. Wang, Y. Zhang, et al. “Unsupervised feature selection via joint local learning and group sparse regression,” *Frontiers Inf Technol Electronic Eng*., Vol. 20, pp. 538-553, 2019.

[117] A.V. Mokeev, V.V. Mokeev, “Pattern recognition by means of linear discriminant analysis and the principal components analysis,” *Pattern Recognit. Image Anal.*, Vol. 25, pp. 685-691, 2015.

[118] K. Zhu and T. Zhang, “Deep reinforcement learning based mobile robot navigation: A review,” *Tsinghua Science and Technology*, Vol. 26, N° 5, pp. 674-691, 2021.

[119] C. Huang et al*.,* “Deductive Reinforcement Learning for Visual Autonomous Urban Driving Navigation,” *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, N° 12, pp. 5379-5391, 2021.

[120] V. Francois-Lavet, et al., “An Introduction to Deep Reinforcement Learning,” *Foundations and Trends in Machine Learning*, Vol. 11, N° 3-4, pp. 219-354, 2018.

[121] Hugh Cartwright, Artificial Neural Networks, Springer US, 2020.

[122] B. E. Boser, V. N. Vapnik I. M. Guyon, A Training Algorithm for Optimal Margin Classifiers, University of California Berkeley, 1992.

[123] I. Sen, M. Saraclar and Y. P. Kahya, “A Comparison of SVM and GMM-Based Classifier Configurations for Diagnostic Classification of Pulmonary Sounds,” *IEEE Transactions on Biomedical Engineering*, Vol. 62, N° 7, 2015.

[124] I. Shahin, A. B. Nassif and S. Hamsa, “Emotion Recognition Using Hybrid Gaussian Mixture Model and Deep Neural Network,” *IEEE*, Vol. 7, 2019.

[125] M. Singla, K. K. Shukla, Robust statistics-based support vector machine and its variants : a survey, Neural Computing and Applications, Springer, 2019.

[126] V. N. Vapnik, A. Ya. Lerner, “Pattern recognition using generalized portraits,” *Automation and Remote Control*, Vol. 24, pp. 709-715, 1963.

[127] R. O. Duda, P. E. Hart, Pattern classification and scene analysis, New York : Wiley, 1973.

[128] V. Vapnik, A. Chervonenkis, “The uniform convergence of frequencies of the appearance of events to their probabilities,” *Teor. Veroyatnost. i Primenen*, Vol. 16, pp. 264-279, 1971.

[129] L. G. Valiant, “A Theory of the Learnable,” *Communications of the ACM*, Vol. 27, N° 11, pp. 1134-1142, 1984.

[130] J. Mercer, “Functions of positive and negative type and their connection with the theory of integral equations ,” *Philos. Trans. Roy. Soc. London, a*, Vol. 209, pp. 415-446, 1909.

[131] M. Aizerman, E. Braverman, and L. Rozonoer, “Theoretical foundations of the potential function method in pattern recognition learning,” *Automation and Remote Control*, Vol. 25, pp. 821-837, 1964.

[132] B. E. Boser, I.M. Guyon, and V.N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” in *Proc. of the 5th Annual Workshop on Computational Learning Theory (COLT’92),* Pittsburgh, pp. 144-152, 1992.

[133] C. Cortes, V. Vapnik, “Support-vector networks,” *Mach Learn*, Vol. 20, pp. 273-297, 1995.

[134] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer New York, NY, 1995.

[135] M. Abaszade, S. Effati, “Stochastic Support Vector Machine for Classifying and Regression of Random Variables,” *Neural Process. Lett.*, Vol. 48, pp. 1-29, 2018.

[136] V. K. Chauhan, K. Dahiya, and A. Sharma, “Problem formulations and solvers in linear SVM: a review,” *Artif. Intell. Rev*., Vol. 52, pp. 803-855, 2019.

[137] A. Tharwat, “Parameter investigation of support vector machine classifier with kernel functions,” *Knowl. Inf. Syst*., Vol. 61, pp. 1269-1302, 2019.

[138] A. Mir, J. A. Nasiri, “KNN-based least squares twin support vector machine for pattern classification,” *Appl Intell*, Vol. 48, pp. 4551-4564, 2018.

[139] M. Kafai, K. Eshghi, “CROification: Accurate Kernel Classification with the Efficiency of Sparse Linear SVM,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, N° 1, pp. 34-48, 2019.

[140] D. Zhang, Fundamentals of Image Data Mining : Analysis, Features, Classification and Retrieval, Springer, 2021.

[141] K. Tatsumi, T. Tanino, “Support vector machines maximizing geometric margins for multi-class classification,” *Top*, Vol. 22, pp. 815-840, 2014.

[142] L. Brand, L. Z. Baker, C. Ellefsen, J. Sargent and H. Wang, “A Linear Primal-Dual Multi-Instance SVM for Big Data Classifications,” in *Proc. IEEE International Conference on Data Mining (ICDM)*, pp. 21-30, 2021.

[143] I. Roman, R. Santana, and A. Mendiburu, et al. “In-depth analysis of SVM kernel learning and its components,” *Neural Comput & Applic*,” Vol. 33, pp. 6575-6594, 2021.

[144] A. Kacem, M. Daoudi, B. B. Amor, S. Berretti and J. C. Alvarez-Paiva, “A Novel Geometric Framework on Gram Matrix Trajectories for Human Behavior Understanding,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, N° 1, pp. 1-14, 2020.

[145] J. Mercer, “Functions of positive and negative type, and their connection with the theory of integral equations,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, Vol. 209, pp. 415-446, 1909.

[146] R. Herbrich, Learning Kernel Classifiers : Theory and Algorithms, MIT Press, 2002.

[147] G. Xu, B. G. Hu and J. C. Principe, “Robust C-Loss Kernel Classifiers,” *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 29, N° 3, pp. 510-522, 2018.

[148] B. B. Hazarika, D. Gupta, “Density-weighted support vector machines for binary class imbalance learning,” *Neural Comput & Applic*, Vol. 33, pp. 4243-4261, 2021.

[149] Y. C. Lee, J. S. Pang, and J. E. Mitchell, “Global resolution of the support vector machine regression parameters selection problem with LPCC,” *Euro J. Comput. Optim*., Vol. 3, pp. 197-261, 2015.

[150] A. Hekmatmanesh, H. Wu, F. Jamaloo, et al. “A combination of CSP-based method with soft margin SVM classifier and generalized RBF kernel for imagery-based brain computer interface applications,” *Multimed Tools Appl*., Vol. 79, pp. 17521-17549, 2020.

[151] W. Karush, Minima of functions of several variables with inequalities as side constraints. Master’s thesis, Dept. of Mathematics, Univ. of Chicago, 1939.

[152] H.W. Kuhn, A.W. Tucker, “Nonlinear programming,” *2nd Berkeley Symposium on Mathematical Statistics and Probabilistics*, University of California Press, pp. 481-492, 1951.

[153] L. Hamel, Knowledge discovery with support vector machines, Wiley Edition, 2009.

[154] A. Shigeo, Support Vector Machines for Pattern Classification, Springer-Verlag London Limited, 2005.

[155] V. N. Vapnik, Statistical Learning Theory, Edition Wiley, 1998.

[156] B. Scholkopf, A. J. Smola, Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2002.

[157] S. Knerr, L. Personnaz, J. Dreyfus, et al. “Single-layer ll learning revisited : A stepwise procedure for building and training a neural network,” *Optimization Methods and Software*, Vol. 1, pp. 23-34, 1990.

[158] F. de Morsier, D. Tuia, M. Borgeaud, V. Gass and J. Thiran, “Semi-Supervised Novelty Detection Using SVM Entire Solution Path,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 51, N° 4, pp. 1939-1950, 2013.

[159] D.M.J. Tax, R.P.W. Duin, “Data domain description by support vectors,” in *Proc. ESANN*, p. 251-256, 1999.

[160] A. Petropoulos, S. P. Chatzis, S. “Xanthopoulos, A hidden Markov model with dependence jumps for predictive modeling of multidimensional time-series,” *Information Sciences*, Vol. 412-413, pp. 50-66, 2017.

[161] Q. Dong and C. Pang, “Risk-Based Non-Myopic Sensor Scheduling in Target Threat Level Assessment,” *IEEE Access*, Vol. 9, pp. 76379-76394, 2021.

[162] B. Zaidi, S. Selouani, M. Boudraa and G. Hamdani, “Human/machine interface dialog integrating new information and communication technology for pathological voice,” in *Proc. Future Technologies Conference* (*FTC*), pp. 494-498, 2016.

[163] H. Wu *et* al*.,* “Simultaneous Extraction of Multi-Scale Structural Features and the Sequential Information With an End-To-End mCNN-HMM Combined Model for Fiber Distributed Acoustic Sensor,” *Journal of Lightwave Technology*, Vol. 39, N° 20, pp. 6606-6616, 2021.

[164] L.E. Baum, T. Petrie, “Statistical Inference for Probabilistic Functions of Finite State Markov Chains,” *The Annals of Mathematical Statistics*, Vol. 37, pp. 1554-1563, 1966.

[165] F. Jelinek, L. Bahl and R. Mercer, “Design of a linguistic statistical decoder for the recognition of continuous speech,” *IEEE Transactions on Information Theory*, Vol. 21, N° 3, pp. 250-256, 1975.

[166] J. K. Baker, “The DRAGON System - An Overview,” *IEEE Trans*, *Acoustics, Speech*, *and Signal Processing*, Vol. 23, N° 1, pp. 24-29, 1975.

[167] J.K. Baker, Stochastic Modeling as a Means of Automatic Speech Recognition, PHD Thesis, Computer Science Department, Carnegie Mellon University, 1975.

[168] H. Bourlard, C. J. Wellekens, “Links between Markov models and multilayer perceptrons,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, N° 12, pp. 1167-1178, 1990.

[169] L. R. Rabiner, S. E. Levinson and M. M. Sondhi, “On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition,” in *The Bell System Technical Journal*, Vol. 62, N° 4, pp. 1075-1105, 1983.

[170] L. R. Rabiner, B. H. Juang, “An Introduction to Hidden Markov Models,” *IEEE Acousfics, Speech & Signal Processing, Magazine*, Vol. 3, pp. 1-16, 1986.

[171] A. Poritz, A. Richter, “On hidden Markov models in isolated word recognition,” *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 705-708, 1986.

[172] L. R. Rabiner, B. Juang, S. E. Levinson and M. M. Sondhi, “Recognition of isolated digits using hidden Markov models with continuous mixture densities,” *AT&T Technical Journal*, Vol. 64, N° 6, pp. 1211-1234, 1985.

[173] L. R. Bahl, F. Jelinek and R. L. Mercer, “A Maximum Likelihood Approach to Continuous Speech Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, N° 2, pp. 179-190, 1983.

[174] A. Rosenberg and A. Colla, “A connected speech recognition system based on spotting diphone-like segments--Preliminary results,” *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987.

[175] C. Sènac, Contribution à l'étude du décodage acoustico-phonetique pour la reconnaissance automatique de la parole, PHD thesis, Universite Paul Sabatier, Toulouse. 1989.

[176] F. S. Samaria, A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in *Proc. of IEEE Workshop on Applications of Computer Vision*, pp. 138-142, 1994.

[177] A. Kundu, P. Bahl, “Recognition of handwritten script : a hidden Markov model based approach,” *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 928-931, 1988.

[178] C. Olivier, M. Avila, P. Courtellemont, T. Paquet and Y. Lecourtier, “Handwritten word recognition by image segmentation and hidden Markov models,” in *Proc. of IECON '93 - 19th Annual Conference of IEEE Industrial Electronics*, Vol. 3, pp. 2093-2097, 1993.

[179] L. Miclet , A. Cornuéjols, Apprentissage artificiel, concepts et algorithmes, Eyrolles, 2002.

[180] D. Jurafsky, J. H. Martin, Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, University of Colorado at Boulder, 2020.

[181] J.P. Haton, C. Cerisara, D. Fohr, Y. Laprie, K. Smaïli, Reconnaissance automatique de la parole Du signal à son interprétation, Dunod, 2006.

[182] A. Tavanaei, A.S. Maida, “Training a Hidden Markov Model with a Bayesian Spiking Neural Network,” *J. Sign. Process. Syst*., Vol. 90, pp. 211-220, 2018.

[183] J. Kim, S.Singh, E. D. Thiessen, et al. “A hidden Markov model for analyzing eye-tracking of moving objects,” *Behav. Res.*, Vol. 52, pp. 1225-1243, 2020.

[184] L. Baum, “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes,” in *Proc. Conference of the Third Symposium on Inequalities,* University of California, Los Angeles, Academic Press, pp. 1-8, 1972.

[185] F. Y. Xie, Y. M. Hu, B. Wu, et al. "A generalized hidden Markov model and its applications in recognition of cutting states," *Int. J. Precis. Eng. Manuf.*, Vol. 17, pp. 1471-1482, 2016.

[186] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, Vol. 13, N° 2, pp. 260-269, 1967.

[187] J. Lember, J. Sova, "Regenerativity of Viterbi Process for Pairwise Markov Models," *J. Theor. Probab.,* Vol. 34, pp. 1-33, 2021.

[188] D. Bepari, S. Koley and D. Mitra, "Empirical validation and performance of duty cycle–based DTMC model in channel estimation," *Ann. Telecommun.*, Vol. 75, pp. 229-240, 2020.

[189] H. Razik, Electrical Systems, From Diagnosis to Prognosis, John Wiley & Sons, 2020.

[190] Y.R. Yeminy, Y. Keller and S. Gannot "Single microphone speech separation by diffusion-based HMM estimation," *J. Audio Speech Music Proc.,* Vol. 16, 2016.

[191] M. Srinivasan, C. Shanmuganathan, S. M. K. Gupta, et al. "Multi-view representation based speech assisted system for people with neurological disorders," *J. Ambient Intell Human Comput*, 2021.

[192] A. Revathi, Y. Venkataramani, "Speaker independent continuous speech and isolated digit recognition using VQ and HMM," in *Proc. Conference on Communications and Signal Processing*, pp. 198-202, 2011.

[193] N. Bouguila, W. Fan, M. Amayri, Hidden Markov Models and Applications, Springer, 2022

[194] M. Dewar, C. Wiggins and F. Wood, "Inference in Hidden Markov Models with Explicit State Duration Distributions," *IEEE Signal Processing Letters*, Vol. 19, N° 4, pp. 235-238, 2012.

[195] M. T. Johnson, "Capacity and complexity of HMM duration modeling techniques," *IEEE Signal Processing Letters*, Vol. 12, N° 5, pp. 407-410, 2005.

[196] A. Coutrot, J. H. Hsiao and A. B. Chan, "Scanpath modeling and classification with hidden Markov models," *Behav Res.*, Vol. 50, pp. 362-379, 2018.

[197] S. Bhati, S. Nayak, and S. R. M. Kodukula, "Unsupervised Speech Signal-to-Symbol Transformation for Language Identification," *Circuits Syst. Signal Process.*, Vol. 39, pp. 5169-5197, 2020.

[198] X. Li and Z. Wang, “A HMM-based mandarin chinese singing voice synthesis system,” *IEEE/CAA Journal of Automatica Sinica*, Vol. 3, N° 2, pp. 192-202, 2016.

[199] L. Li, T. Watzel, L. Kürzinger and G. Rigoll, “Towards Constructing HMM Structure for Speech Recognition With Deep Neural Fenonic Baseform Growing,” *IEEE Access*, Vol. 9, pp. 39098-39110, 2021.

[200] A.Kacha, F.Grenez, and J. Schoentgen, “Estimation of dysperiodicities in disordered speech,” *Speech Comm*. Vol. 48, pp. 1365-1378, 2006.

[201] G. Muhammad, M. Melhem, “Pathological voice detection and binary classification using MPEG-7 audio features,” *Biomedical Sig. Proc. and Control*, Vol. 11, pp. 1-9, 2014.

[202] E.S. Fonseca, J.C. Pereira, “Normal versus pathological voice signals using wavelet analysis and least squares support-vector machines,” *IEEE Eng. in Medicine and Biology Magazine*, Vol. 09, pp. 0739-5175, 2009.

[203] J.D. Arias-Londono, J.I. Godino-Llorente, N. Saenz-Lechon, V.Osma-Ruiz, and G. Castellanos-Dominguez, “Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients,” *IEEE Transactions on Biomedical Eng*.,Vol. 58, N° 2, pp. 370-379, 2011.

[204] H. Frihia, H. Bahi, “HMM/SVM segmentation and labelling of Arabic speech for speech recognition applications,” *Int. J. of Speech Tech*.,Vol. 20 N° 3, pp. 563-573, 2017.

[205] E. Zarrouk, Y. B. Ayed, and F. Gargouri, “Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study,” *Int. J. of Speech Tech*., Vol. 17, N° 3, pp. 223-233, 2014.

[206] Y. Maryn, P. Corthals, P. V. Cauwenberge, N. Roy, M. De Bodt, “Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels,” *Journal of Voice*, Vol. 24, N° 5, 2010.

[207] R. Benhammoud, A. Kacha, "Automatic Classification of Disordered Voices Based on a Hybrid HMM-SVM Model," *J. Commun. Technol. Electron, Springer*, Vol. 66 (Suppl 2), pp. 139-148, 2021.

[208] S. 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, N° 4, pp. 357-366, 1980.

[209] T. L. Nwe, S. Foo, L. De Silva, “Classification of stress in speech using linear and nonlinear features,” in *Proc. Acoustics, Speech, and Signal Processing, Proceedings (ICASSP ’03), IEEE International Conference on Volume 2*, pp. 9-12, 2003.

[210] B. S. Atal, S. L. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the Speech Wave,” *Journal of the Acoustical Society of America*, Vol. 50, N° 2, pp. 637-655, 1971.

[211] F. Itakura, “Minimum prediction residual principle applied to speech recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 23, N° 1, pp. 67-72, 1975.

[212] S. Nagarajan, R. Sankar, “Efficient implementation of linear predictive coding algorithms,” in *Proc. IEEE Southeastcon '98 'Engineering for a New Era'*, pp. 69-72, 1998.

[213] R. Benhammoud, A. Kacha, "Automatic Classification of Disordered Voices with Hidden Markov Models," *in Proc. of IEEE, International Conference on Signal, Image, Vision and their Applications (SIVA)*, pp. 1-6, 2018.

[214] T. Virtanen, R. Singh, and B. Raj, Techniques for Noise Robustness in Automatic Speech Recognition, Wiley, United Kingdom, 2012.

[215] L. Baum, “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes,” in *Proc.of the Third Symposium on Inequalities, University of California, Los Angeles, Academic Press*, pp. 1-8, 1972.

[216] J. P. Haton, C. Cerisara, D. Fohr, Y. Laprie, and K. Smaïli, Reconnaissance automatique de la parole du signal à son interpretation, Dunod, Paris, 2006.

[217] A. Alpan, Objective assessment of disordered connected speech, PHD Thesis, polytechnic school of Brussels, 2012.

[218] R. Benhammoud, A. Kacha, "Classification automatique de la parole pour l’évaluation des troubles de la voix," CGE’10*, EMP, Algérie*, pp. 1-6, 17-18 Avril 2017.

[219] J. C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,” in *Proc. of Advances in Large-Margin Classifiers*, *MIT Press, Cambridge*, pp. 61-74, 1999.

[220] M. Gonen, A. G. Tanugur, and E. Alpaydin, “Multiclass Posterior Probability Support Vector Machines,” *IEEE Transactions on Neural Networks,* Vol. 19, N° 1, pp. 130-139, 2008.

[221] A. Cornuéjols, L. Miclet, Apprentissage artificiel concepts et algorithms, Eyrolles, Paris, 2003.

[222] J. Stadermann, G. Rigoll, “A Hybrid SVM/HMM Acoustic Modeling Approach to Automatic Speech Recognition,” in *Proc. of Interspeech-ICSLP of 8th Int. Conf. on Spoken Lang., Interspeech, ICSLP*, pp. 1-4, 2004.

[223] P. Shukla, A. Verma, Abhishek et al. “Interpreting SVM for medical images using Quadtree,” *Multimed Tools Appl*, Vol. 79, pp. 29353-29373, 2020.