

UNIVERSITE SAAD DAHLEB DE BLIDA

Faculté des Sciences de l'Ingénieur

Département d'Electronique

MEMOIRE DE MAGISTER

Spécialité : Signaux & Systèmes

PATHOLOGIE DU LANGAGE ARABE

Etude du sigmatisme occlusif

Par

Mohamed Abdelkader BENCHERIF

Devant le jury composé de :

A. Guessoum	Professeur, U. de Blida	Président
H. Salhi	Maître de conférences, U. de Blida	Examineur
N. Khorissi	Chargé de cours, U. de Blida	Examineur
M. Guerti	Maître de conférences, E.N.P. Alger	Rapporteur
Z. Benselama	Chargé de cours, U. de Blida	Co
Rapporteur		

Blida, Septembre 2005

RESUME

L'étude de la pathologie du langage Arabe rentre dans un cadre pluridisciplinaire, en effet, elle concerne la détection des pathologies langagières concernant la vibration des cordes vocales, l'articulation synchrone des mandibules, la détection de la longueur de la langue, les défauts de langue, les prononciations non articulés, etc.

Les différents défauts de prononciation sont généralement corrigés à l'aide d'un orthophoniste qui utilise des méthodes très simplistes, parfois lentes et lassantes pour le patient.

Notre travail rentre dans un cadre d'entraînement à la bonne prononciation des personnes souffrant de défauts langagiers et d'une aide à la décision à l'orthophoniste, par des méthodes graphiques et sonores, afin de suivre l'évolution du patient présentant un sigmatisme et de finement détecter les phonèmes à corriger, le système d'aide peut être aussi installé au niveau du patient lui-même et lui permettra de se corriger au fur et à mesure de sa thérapie.

ملخص

دراسة الأمراض الكلامية أو عوائق نطق اللغة العربية بالطرق المتداولة حالياً في المستشفيات جد مرهقة للمريض و المعالج. في هذا الإطار حاولنا إدخال طريقة تتركز على استعمال الحاسوب كمساعد للنطق بإعطاء علامة تدل على تحسن نطق المريض عبر صور و أمثلة للنطق.

حاولنا في هذا البحث دراسة الإعاقة الخاصة بتبديل نطق السين بحرف الثاء، كأن ينطق أحدهم "ثخصية" بدل "شخصية".

الميزة الأساسية لهذا النظام أن المريض يستطيع التمرن لوحده بدون تدخل مباشر للمعالج.

ABSTRACT

The study of the Arabic language pathology combines different fields of study, It relates to the detection of linguistic pathologies concerning the vibration of the vocal cords, the synchronous articulation of the mandibles, the tongue's length detection, the defects of mispronunciation, etc.

This various speech defaults are generally corrected using a speech therapist or orthophonist who uses some simplistic methods, sometimes slow and burying for the patient.

This work is devoted to persons suffering from Arabic linguistic pronunciation defects, and is intended to be a decision aided tool, by giving a score of good pronunciation, in order to follow the evolution of the patient presenting a sigmatism, and finely to detect the zones to be corrected, the system can be also be used by the patient itself in an isolated therapy session.

The use of interactive images and the patient feedback pronunciation gives some remarkable good results, in order to let the patient progress in his therapy.

DEDICACES

Je dédie ce travail à mes parents qui m'ont tant donné sans se lasser.

Je le dédie à ma mère qui m'a tant encouragé dans mes études.

Je dédie ce travail à mon épouse qui m'a tant soutenu, ainsi qu'à mes filles Houda et Bouchra pour leurs patiences, pour les nuits blanches passées loin d'elles, pour les week-ends ratés, enfin pour tous les manquements que je dois acquitter.

Mohamed...

REMERCIEMENTS

"It does not take much strength to do things, but it requires great strength to decide on what to do"

Elbert Hubbard (1859-1915)

Je tiens, en premier lieu, à remercier Dieu pour sa miséricorde, sa bonté, son obligeance et la ferveur qu'il m'a donné afin de m'engager dans la voie de la recherche.

Je remercie ma Directrice de Thèse M.Guerti, M.C. à l'ENP, mon Codirecteur Z.Benselama C.C. à l'université de Blida, pour tous leurs conseils et directives tout au long de mon parcours de Magister.

Je remercie H.Bessalah, Directeur du CDTA, pour toute son aide et son soutien.

Je remercie A.Guessoum Professeur; H.Salhi M.C. ; N.Khorissi C.C. de l'Université de Blida pour avoir accepté d'examiner ce travail.

Je remercie le corps enseignant et dirigeant de l'Institut d'Electronique et de Mathématiques pour tous les efforts qu'ils font pour accomplir leur si noble devoir qu'est l'enseignement.

Je remercie Boukrittia Aicha pour toute son aide, ainsi que le Docteur Ghozali de l'hôpital Frantz Fanon, Service de Rééducation fonctionnelle pour toutes les informations utiles qu'elle m'a transmis.

Je remercie tous mes collègues du CDTA, pour leur soutien et leur générosité, particulièrement et spécifiquement les membres de l'équipe «Arithmétique, Algorithme Architecture Série et parallèles - A3SP», pour tous les conseils qu'ils m'ont donnés et surtout pour le cadre agréable de travail.

Je remercie tous ceux qui croient que le savoir est universel et qui nous donnent tant de plaisir sur Internet en nous guidant et en nous simplifiant les parcours de la recherche scientifique.

TABLE DES MATIERES

RÉSUMÉ	1
REMERCIEMENTS	2
TABLE DES MATIERES	3
LISTE DES ILLUSTRATION, GRAPHIQUE ET TABLEAUX	5
INTRODUCTION	6
1. PRODUCTION DE LA PAROLE	
1.1. Introduction	8
1.2. Système phonatoire	9
1.3. Définitions	15
1.4. Domaines d'études	15
1.5. L'Alphabet Phonétique International "API"	16
1.7. Alphabet Phonétique Arabe	19
1.8. Système vocalique Arabe	25
1.9. Structure de la langue arabe en succession phonétique	26
1.10. Etudes réalisées	27
1.11. Conclusion	28
2. PATHOLOGIES DU LANGAGE PARLE	
2.1. Introduction	29
2.2. Définitions	29
2.3. Cordes Vocales saines	30
2.4. Pathologie des cordes vocales	31
2.5. Pathologies des autres canaux vocaux	38
2.6. Classification des pathologies	39
2.7. Défauts de la voix détectés par l'oreille	40
2.8. Pathologie concernée par notre travail	43
2.9. La rééducation orthophonique	44
2.10. Conclusion	45

3. TECHNIQUES D'ANALYSE DU SIGNAL VOCAL	
3.1. Introduction	46
3.2. Historique	47
3.3. Problématique générale	49
3.4. Méthodologie de Travail	50
3.5. Filtrage de préaccentuation	51
3.6. Méthodes d'analyse en Reconnaissance Automatique de la Parole	52
3.7. Travaux similaires	54
3.8. Analyse et extraction des paramètres	55
3.9. Modélisation des données par mélange de Gaussiennes GMM	69
3.10. Apprentissage par l'Algorithme de Baum Welch «Variante de l'algorithme EM»	69
3.11. Conclusion	72
4. RECONNAISSANCE PAR LES HMM	
4.1. Introduction	73
4.2. Description du corpus	74
4.3. Présentation du cas d'étude	79
4.4. Comparaisons avec différents locuteurs	80
4.5. Spectrogrammes et déplacement formantique	82
4.6. Méthodologie de travail	85
4.7. Modélisation par les HMM	85
4.8. Courbes de convergence de l'algorithme EM	89
4.9. Modèles HMM du mot [ʃ a ʒ s i j a t u n]	90
4.10. Modélisation de la phrase SA1 de la base TIMIT	91
4.11. Degré de Vraisemblance des locuteurs par rapport au modèle HMM3	92
4.12. Segmentation et reconnaissance phonémique	94
4.13. Degré de vraisemblance phonémique	96
4.14. Conclusion	97
CONCLUSIONS ET PERSPECTIVES	98
APPENDICE	
A. Liste des symboles	100
REFERENCES	101

LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

Figure 1.1	Vue en Coupe des poumons	9
Figure 1.2.	Inspiration et expiration de l'air	9
Figure 1.3.	Vue de face de la trachée artère	10
Figure 1.4.	Vue de côté du larynx	12
Figure 1.5.	Vue de haut de la langue	13
Figure 1.6.	Coupe sagittale du dispositif phonatoire	14
Figure 1.7.	Lieux d'articulation des phonèmes	19
Figure 1.8.	Lieux d'articulation des phonèmes : [أ] , [و] , [ي]	23
Figure 1.9	Lieux d'articulation des phonèmes [ش] , [ك] , [ق]	23
Figure 1.10.	Lieux d'articulation des phonèmes [ل] , [ض] , [ج]	23
Figure 1.11	Lieux d'articulation des phonèmes [ت] , [ر] , [ن]	23
Figure 1.12.	Lieux d'articulation des phonèmes [ط] , [د] , [ر]	24
Figure 1.13.	Lieu d'articulation des phonèmes [ظ] , [ذ] , [ض]	24
Figure 1.14.	Lieux d'articulation des phonèmes [ز] , [ص] , [س]	24
Figure 1.15.	Position des lèvres lors de la prononciation des phonèmes [م] , [ي] , [و] , [ف]	24
Figure 1.16.	Triangle vocalique de la langue Arabe	25
Figure 2.1.	Cordes vocales saines avec différents degrés d'aperture	30
Figure 2.2.	Nodules sur les cordes vocales	31
Figure 2.3.	Paralysie unilatérale des cordes vocales	31
Figure 2.4.	Cordes vocales pathologiques présentant un arc à la fermeture	32
Figure 2.5.	Polype très distingué sur l'une des cordes vocales à gauche	33
Figure 2.6.	Cordes vocales gonflées à gauche	33
Figure 2.7.	Détail d'un kyste de différents patients	33
Figure 2.8.	Mesure d'un Kyste vocal après son extraction	33
Figure 2.9.	Granulomes de différents patients	34
Figure 2.10	Papillomes chez différents patients	34
Figure 2.11	Sièges possibles des cancers	3:
Figure 2.12	Vue en coupe du larynx	36
Figure 2.13.	Bec de lièvre "Avant intervention / Après intervention"	38

Figure 2.14.	Palais totalement absent	38
Figure 2.15.	Diagramme de classement des pathologies	39
Figure 3.1.	Segmentation temporelle et Recouvrement	51
Figure 3.2.	Schéma global du processus de reconnaissance utilisé en RAP	54
Figure 3.3.	Chaîne d'analyse du signal produisant les coefficients MFCC	56
Figure 3.4.	Filtre triangulaire passe bande en Mel fréquence $B(f)$	57
Figure 3.5.	Filtre triangulaire passe bande en fréquence f	57
Figure 3.6.	Chaînes de Markov comme Automate probabiliste	59
Figure 3.7.	Graphe d'états d'une chaîne de Markov	61
Figure 3.8.	Modèle de transitions à trois états Markoviens gauche-droite.	64
Figure 3.9.	Modèle de transitions entre états markoviens gauche droite et leur probabilité de transition et d'émission respectives.	64
Figure 4.1.	Comparatifs visuels des prononciations (F1/F2/F6)	74
Figure 4.2.	Comparatifs visuels des prononciations (M1/M2/F6)	76
Figure 4.3.	Spectrogrammes « locutrice F4 »	77
Figure 4.4.	Spectrogrammes « locutrice M4 »	78
Figure 4.5.	Spectrogrammes « locuteur F6 »	79
Figure 4.6.	Segmentation phonémique pour différents locuteurs en utilisant 12 MFCC.	79
Figure 4.7.	Segmentation phonémique pour différents locuteurs en utilisant 39 paramètres.	86
Figure 4.8.	Courbe de convergence de l'algorithme EM (a) Modèle HMM1 ; (b) Modèle HMM2 ;	89
Figure 4.9.	Courbe de convergence de l'algorithme EM- Modèle HMM3	90
Figure 4.10.	Degré de vraisemblance que HMM3 ait généré les observations acoustiques par locuteur	92
Figure 4.11.	Figure 4.11. Segmentation du mot pathologique [ʃ a χ s i j a t un]	93
Figure 4.12.	Prononciations de la locutrice F6P, avant et après les feedback visuels et auditifs	94

Tableau 1.1.	Alphabet International Phonétique	17
Tableau 1.2.	Alphabet International Phonétique pathologique	18
Tableau 1.3.	Correspondance entre organes – lieux d'articulation	20
Tableau 1.4.	Transcription phonétique des phonèmes de la langue Arabe ainsi que leurs lieux d'articulation.	21
Tableau 1.5.	Récapitulatif des phonèmes Arabes ainsi que leurs lieux d'articulation selon El Khalil	22
Tableau 1.6.	Structure phonétique minimale du langage Arabe	27
Tableau 2.1.	Phonétique du mot présentant la prononciation pathologique	43
Tableau 3.1.	Durées primaires d'analyse ainsi que la durée de chevauchement	51
Tableau 3.2.	Méthodes d'analyse	52
Tableau 3.3.	Méthodes de réduction des données	53
Tableau 3.4.	Méthodes de classification / calcul de vraisemblance	53
Tableau 3.5.	Calcul du nombre de vecteurs acoustiques	59
Tableau 3.6.	Choix des paramètres	68
Tableau 3.7.	Choix du nombre de gaussiennes	69
Tableau 4.1.	Différents locuteurs ayant enregistrés le corpus F : Féminin et M Masculin	74
Tableau 4.2.	Occurrences du [j] en position initiale des mots	76
Tableau 4.3.	Occurrences du [j] en position médiane des mots	77
Tableau 4.4.	Occurrences du [j] en position finale des mots	78
Tableau 4.5.	TOP des mots sélectionnés pour la modélisation par les HMM/GMM	79
Tableau 4.6.	Mots de tests sélectionnés à partir de la base de données.	79
Tableau 4.7.	Paramètres du modèle HMM1	86
Tableau 4.8.	Paramètres du modèle HMM2	86
Tableau 4.9.	Modèles HMM de la phrase SA1 par 80% du corpus (80 locuteurs)	86
Tableau 4.10.	Taux de reconnaissance sur la base TIMIT (20 / 80 locuteurs)	86
Tableau 4.11.	Paramètres du modèle HMM3	90
Tableau 4.12.	Matrice de transition du modèle HMM3	90
Tableau 4.13.	Modèles HMM de la phrase SA1 par 80% du corpus (80 locuteurs)	91
Tableau 4.14.	Taux de reconnaissance sur la base TIMIT (20 locuteurs)	91

Tableau 4.15.	Paramètres du modèle HMM du phonème [ʃ]	95
Tableau 4.16.	Matrice de transition du modèle HMM du phonème [ʃ]	95
Tableau 4.17.	Paramètres du modèle HMM des phonèmes [χ]	95
Tableau 4.18.	Matrice de transition du modèle HMM du phonème [χ]	95
Tableau 4.19.	Vraisemblance phonémique	96

INTRODUCTION

Tu peux manquer d'argent, on te prêtera,
Tu peux manquer de manger, on te rassasiera
Mais si tu manques de parole, ...

Dans la nature humaine, parler est un besoin, l'un des plus élémentaires peut-être. Mais le plus distinctif de par sa forme, son impact et surtout de par sa variabilité. Les différents peuples communiquent dans leurs langues maternelles, apprennent à leurs progénitures dès leurs naissances les mots, les intonations, les phrases, les manières de dire les choses, de décrire les objets, les faits ainsi que les circonstances. Ceci d'une part, pour uniformiser le parler de la 'tribu' et d'autre part pour développer leur présence et renforcer leur existence.

Les différentes langues se différencient, certes, de par leur vocabulaire, leur grammaire et leur prononciation. Mais l'unicité du canal vocal et des différents articulateurs vocaux, les aligne à pied égal lorsqu'il y'a défaillance ou pathologie du système phonatoire.

Lors du traitement des pathologies langagières, les médecins essaient en premier lieu de détecter la pathologie par des méthodes non invasives, afin de préserver au maximum les conditions de fonctionnement de l'organe défectueux, si toutefois, une intervention chirurgicale n'est pas nécessaire, et que le problème est d'ordre phonétique articulaire, l'orthophoniste prend le relais. Ce dernier, dans le cadre de la thérapie de prononciation par correction assistée, utilise ses sens, entre autres, son ouïe, sa vision ainsi que sa propre prononciation, procédant selon une méthodologie bien définie.

A toutes ces étapes, l'orthophoniste utilise son ouïe pour entendre et évaluer la bonne prononciation. Cependant, la qualité de l'évaluation dépend de l'ouïe de l'orthophoniste, cette évaluation se trouve amoindrie avec la défaillance de l'ouïe de l'orthophoniste. Dans le but d'aider ce dernier à décider de la prononciation ainsi que de l'évolution de la guérison du patient et rendre les méthodes 'manuelles' qu'il utilise plus

automatiques, nous allons, à travers ce travail contribuer à la réalisation d'un système d'aide à l'évaluation de la prononciation d'une des pathologies du langage Arabe appelé para sigmatisme, celle-ci concerne le remplacement du [ʃ] et [s] par [θ] et [t]. Cette pathologie langagière d'ordre articulo-phonatoire est généralement traitée en mode phonème isolé, le patient répète le phonème jusqu'à s'accoutumer avec les mécanismes d'articulation, la seconde étape concerne la prononciation du phonème, en question, précédé ou suivi par les voyelles, dans le but de dégager la prononciation de mots au complet.

En vue de simplifier l'utilisation de notre système d'aide, nous nous sommes basés sur un mode interactif, en présence ou en l'absence d'un orthophoniste, le patient enregistre la prononciation de mots bien ciblés, à travers un microphone, et observe la segmentation de ce qu'il a prononcé en unités phonémiques. Le système lui donne un score de vraisemblance lui indiquant le taux de réussite de la prononciation et le point de défaillance par rapport à un corpus référence. Le patient peut écouter différents locuteurs, et visualiser la prononciation par des images animées.

L'évaluation de la prononciation des phonèmes, ou mots mal prononcé, a été modélisée par des méthodes probabilistes à base de chaînes de Markov et de mélanges de gaussiennes.

Nous nous sommes inspirés, dans ce travail, des méthodes d'apprentissage de la langue Anglaise prononcée par des Taiwanais ou des Japonais non natifs de l'Anglais, ces études sont orientés vers la détection des phonèmes présentant une mal prononciation intrinsèque à ces personnes, nous avons essayé de développer le caractère de similitude remarqué entre apprendre une langue en étant non natif de celle-ci et un patient présentant un schlissement.

CHAPITRE 1

PRODUCTION DE LA PAROLE

1.1. Introduction

L'un des éléments clés des études concernant le Traitement Automatique de la Parole (T.A.P.) est de comprendre les mécanismes fondamentaux nécessaires à la production des sons ou de la parole, de distinguer les différents facteurs intervenants, de séparer les comportements entrelacés des acteurs principaux en allant des poumons vers les lèvres,...

L'étude de la production de la parole est un domaine pluridisciplinaire, elle fait intervenir diverses théories de l'aérodynamique, de la réfraction des sons, de l'humidification, de la résonance, de l'informatique.

- ◆ le comportement répétitif ou périodique des poumons dans leur infatigable mouvement de contraction et d'expansion;
- ◆ le comportement aérodynamique du conduit vocal lors du passage de l'air à travers ses différentes constriction comparé à un modèle multitube;
- ◆ le degré d'humidification des parois mandibulaires ainsi que de toutes les surfaces en contact de l'air;
- ◆ le système de synchronisation nerveux, qui contrôle les réflexes de déplacement de la mâchoire inférieure, des lèvres;
- ◆ le comportement mécanique lors de la synchronisation des mouvements afin d'éviter tout contact indésirable à la bonne production de la voix.

Nul aspect ne peut être isolé ou ignoré lors de la production vocale, si tel événement apparaît, une pathologie langagière se fait ressentir et tout le système tombe dans un déséquilibre.

1.2. Système phonatoire

Le système phonatoire humain est un ensemble d'organes vivants, transmettant l'air suivant une cadence bien rythmée, afin d'aboutir à un son audible d'un côté et assurer la continuité de la production de la phonation d'un autre côté.

1.2.1. Les poumons

Les poumons sont le principal acteur de la production vocale, ils fournissent l'énergie nécessaire qui est l'air, avec un débit bien déterminé. Les muscles responsables de la respiration travaillent inlassablement en arrière plan, afin d'autoriser l'entrée ou la sortie des séquences ou flux d'air, selon une temporisation parfaite en vue d'assurer la synchronisation globale avec les autres intervenants en aval, voir figure 1.1.

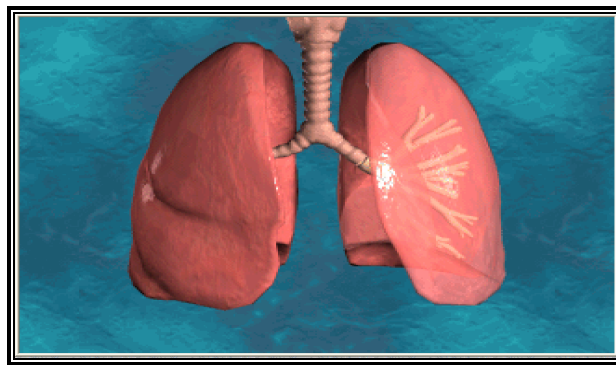


Figure 1.1. : Vue en coupe des poumons [2]

Le mouvement des poumons dans leur infatigable geste se résume en des équilibres perpétuels de pression entre l'intérieur du corps humain et l'environnement extérieur, les différentes phases sont illustrées dans les figures 1.2 (a) et (b).

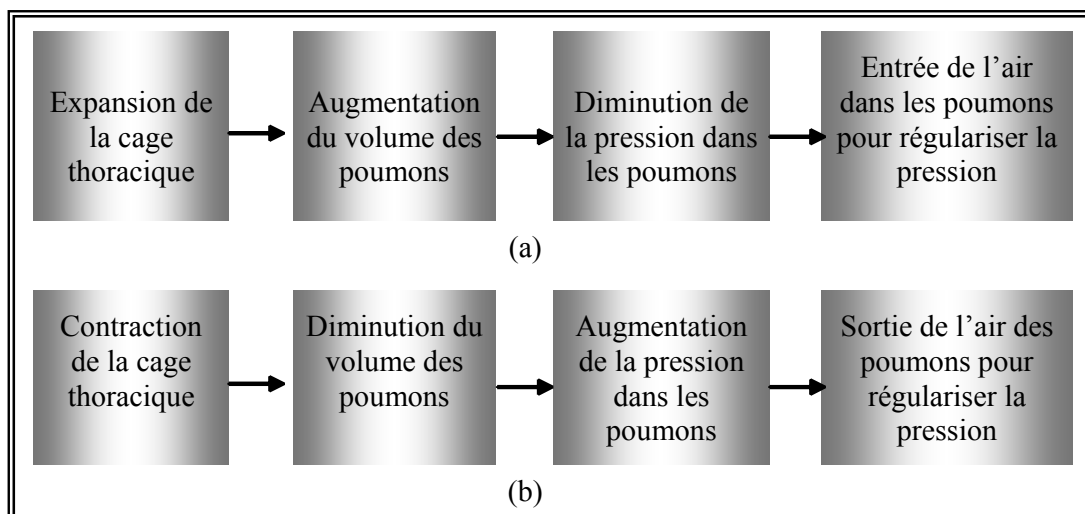


Figure 1.2 : (a) et (b) Inspiration et expiration de l'air [1]

Le réseau tubulaire très ramifié des deux poumons occupe la plus grande partie de la cage thoracique. Le poumon gauche qui se constitue du lobe supérieur et du lobe inférieur, est plus petit que le poumon droit, auquel s'ajoute encore un troisième lobe.

La surface des poumons ainsi que la paroi thoracique sont couvertes d'un pelage; les deux membranes constituent ensemble la plèvre. Elles sont superposées à plat et se confondent à la base des poumons. Entre les deux membranes, du liquide est stocké afin qu'elles puissent glisser l'une contre l'autre sans friction.

La surface intérieure des poumons est d'environ 70 m² au total - ce qui représente donc environ la grandeur d'un terrain de squash ! Chaque poumon contient environ 300 millions de vésicules pulmonaires qui se groupent autour des bronchioles. Elles sont alimentées par des capillaires et forment ensemble cette grande surface qui est nécessaire afin que les poumons puissent remplir leur fonction: de transporter à l'extérieur le gaz carbonique résultant lors de la combustion de substances nutritives. Un adulte absorbe normalement 400 à 500 ml d'air par inspiration. On respire à peu près 12 à 16 fois par minute. Lorsqu'on fait des efforts et quand on est physiquement actif, le volume est considérablement accru, ainsi que la fréquence respiratoire. C'est à travers la paroi de l'alvéole pulmonaire que diffusent l'oxygène et le gaz carbonique, selon des mécanismes simples d'équilibre chimique. Toutes les maladies affectant la paroi alvéolaire retentissent donc sur la qualité des échanges gazeux respiratoires. [3]

1.2.2. La trachée artère

La trachée-artère, longue d'environ douze centimètres, touche au larynx et se ramifie au niveau de la quatrième vertèbre dorsale vers les deux bronches principales.

Les tissus élastiques et musclés de l'artère sont soutenus par 16 à 20 boucles de cartilage en forme de fer à cheval et recouverts à l'intérieur par une muqueuse avec des cils vibratiles. Ils transportent les particules de poussière qui entrent par l'air inspiré en arrière dans le pharynx [3], comme illustré dans la figure 1.3.

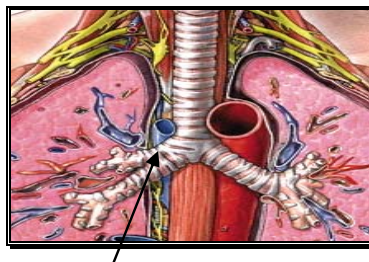


Figure 1.3. : Vue de face de la trachée artère [4]

1.2.3. Dispositif laryngé

1.2.3.1. Fonctions physiologiques

Le larynx assure trois fonctions dans la :

- ◆ respiration, puisqu'il fait partie intégrante des voies respiratoires ;
- ◆ déglutition, en fermant l'accès aux voies respiratoires sub glottiques.
- ◆ production de sons, à cause de son rôle phonatoire important, bien que non vital [5].

1.2.3.2. Anatomie du larynx

Le larynx se compose de deux parties essentielles, la charpente et la musculature laryngées, qui se présentent comme suit

La charpente laryngée est composée de trois cartilages :

- ◆ le cartilage cricoïde, qui a une forme de bague, et situé dans la partie inférieure du larynx, en contact avec la trachée artère;
- ◆ le cartilage thyroïde, qui forme la « pomme d'Adam » dans sa partie antérieure;
- ◆ le cartilage épiglottique, d'une position plus centrale et supérieure.

Cette charpente est maintenue à l'aide des membranes fibreuses crico-thyroïdienne, thyroïdienne et hyoépiglottique. L'attache supérieure se fait avec le pharynx via l'os hyoïde, le cartilage thyroïde et la membrane fibreuse thyroïdienne.

La musculature laryngée a pour but de mettre en mouvement le larynx ou de modifier son ouverture pour jouer sur la production des sons. Elle est composée, entre autres :

- ◆ du muscle crico-thyroïdien, qui, en se contractant, fait basculer le cartilage thyroïde vers l'avant, ayant pour effet de tendre les cordes vocales ;
- ◆ des muscles thyro-aryténoïdiens supérieur, interne et externe ;
- ◆ des muscles crico-aryténoïdien latéral et postérieurs ;
- ◆ du muscle inter aryténoïde (impair).

Les trois dernières catégories de muscles ont des points de fixation en commun : les cartilages aryténoïdes, situé derrière le cartilage thyroïde, et auxquels sont attachées les cordes vocales. Leurs mouvements de translation et de rotation permettent de moduler l'ouverture et l'accolement des ces cordes.

Le larynx est innervé par des branches du nerf crânien, le nerf vague. Le nerf laryngé supérieur est mixte, mais essentiellement sensitif, alors que le nerf laryngé inférieur est moteur. Tous les muscles du larynx sont innervés, à l'exception du muscle crico-thyroïdien, voir figure 1.4.

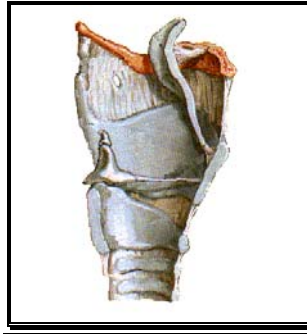


Figure 1.4 : Vue de profil du larynx

Les cartilages du larynx font des mouvements de rotation et de contraction pour permettre différentes figures géométriques des cordes vocales. Ces dernières s'ouvrent et se ferment pour permettre l'action de respirer et de modeler l'air sortant en son audible, en assurant une pression aérodynamique par des forces élastiques. Le mouvement continu ou périodique d'aperture et de fermeture assure la phonation et produit les sons voisés et non voisés. [6]

1.2.4 Articulations complexes

La vibration des cordes vocales ne suffit pas à produire un son intelligible, tout un système articuloire en aval, assure la propagation de l'air vibrant ou non vibrant, les systèmes essentiels, nous pouvons citer :

1.2.4.1. L'épiglotte

C'est une structure cartilagineuse reliée au larynx qui coulisse vers le haut quand les voies aériennes sont ouvertes, et aide à obstruer l'entrée de la trachée au moment de la déglutition. Elle descend légèrement vers le bas, afin d'entrer en contact avec le larynx qui s'élève, formant ainsi un verrou au-dessus du larynx. Il se peut que de temps à autre, lorsqu'on mange trop vite, des aliments liquides ou solides ingérés pénètrent dans le larynx avant que l'épiglotte n'ait pu se rabattre sur celui-ci. De tels cas peuvent s'avérer très dangereux du fait que les voies respiratoires peuvent se boucher et empêcher l'air de pénétrer dans les poumons. [5]

1.2.4.2. La luette

La luette ou uvule est une saillie allongée mobile qui termine le voile du palais et qui contribue, lorsqu'elle se détache de la paroi pharyngale, à permettre à l'air provenant des poumons et du larynx de se diriger non seulement vers la bouche, mais également vers les fosses nasales. Lorsque la luette s'appuie sur la paroi pharyngale, elle empêche l'air de pénétrer dans les fosses nasales et ne le laisse s'échapper que par la bouche (articulations orales).

1.2.4.3. La langue

La langue est une masse musculaire divisée en trois parties, voir figure 1.5.

- ◆ la pointe (apex) qui sert d'articulateur pour les articulations apicales, le dos pour les articulations pré médio ou post-dorsales, et la racine dans le cas des articulations radicales. Elle constitue l'articulateur principal des différents sons.
- ◆ La langue permet le blocage d'air venant des poumons pour produire les consonnes occlusives, le resserrement de la cavité buccale inhérent à la production des consonnes constrictives, lorsqu'elle demeure suffisamment éloignée de la voûte du palais, elle permet la réalisation des différentes voyelles. [7]

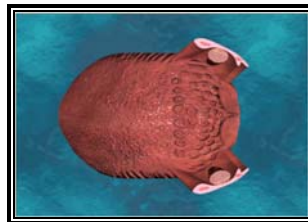


Figure 1.5 : Vue de haut de la langue [2]

1.2.4.4. Les lèvres

Ce sont les parties charnues qui bordent extérieurement la bouche. Elles s'amincissent pour se joindre aux commissures. La lèvre supérieure est limitée par le nez, alors que la lèvre inférieure est limitée par le sillon mentonnier. Lorsqu'elles sont projetées et arrondies, les lèvres forment une cavité qui sert de résonateur lors de la réalisation des voyelles arrondies et des consonnes labialisées. En revanche, lorsque les lèvres sont rétractées, les voyelles sont non arrondies et les consonnes non labialisées.

La lèvre supérieure peut également agir comme lieu d'articulation ([p, b, m]), alors que la lèvre inférieure peut agir comme articulateur ([p, f, v]). [7]

1.2.4.7. Les dents

Les dents bien que pas très coopératives à la phonation, leur absence rend le système phonatoire mécaniquement défectueux, en atrophiant les ouvertures des lèvres et la prononciation des labio-dentales...

Chacun des ces organes est à la base de la production d'un son élémentaire appelé phonème. Ce dernier est la contribution distribuée du système phonatoire. La participation de chaque intervenant dépend de la langue prononcée. Les nasalisations, les roulements des [r].... lorsqu'elles sont exagérées sont à la base des défauts langagiers et sont considérées comme pathologies nécessitant un traitement de réapprentissage de la prononciation.

La figure 1.6 illustre le dispositif phonatoire complet.



Figure 1.6. : Coupe sagittale du dispositif phonatoire [2]

Avant d'entamer les définitions de la phonétique articulatoire et de la phonologie, il y'a lieu de définir les éléments nécessaires à la compréhension du domaine d'étude de ces sciences.

1.3. Définitions

- ❖ **Phonème** : C'est un son idéal avec une articulation complète de l'appareil phonatoire, le langage Arabe comprend deux types de phonèmes, les voyelles et les consonnes.
- ❖ **Allophone** : Ce sont les différentes variantes d'un phonème comme /a/ dans marche et arrête.
- ❖ **Biphones, Triphones** : Séquence de deux ou trois phonèmes.

1.4. Domaines d'études

La **phonétique** étudie les sons du langage dans leur réalisation concrète appelé '**phonèmes**', indépendamment de leurs fonctions linguistiques. Il s'agit donc du 'son' en tant que 'son', sans se soucier du sens.

La **phonétique articulatoire** est une sous branche de la phonétique, elle étudie les mouvements des organes phonatoires lors de l'émission du message. C'est-à-dire comment un être humain fait pour produire tel ou tel phonème. Quels sont les organes qui sont utilisés, comment sont-ils disposés et comment parviennent-ils à articuler ce phonème?

La phonologie est la science qui étudie les phonèmes du point de vue de leur fonction dans le système de communication linguistique. La phonologie étudie les éléments phoniques qui distinguent, dans une même langue, deux éléments de sens différents. Elle se différencie donc de la phonétique qui étudie les éléments phoniques indépendamment de leurs fonctions dans la communication.

Les principes de base de la théorie de la phonologie actuelle sont l'héritage de L.F. Saussure pour qui un **phonème**, c'est-à-dire la plus petite unité, n'avait de valeur que par opposition aux autres phonèmes.

On distingue habituellement 2 grands domaines de la phonologie :

La **phonématique**, qui étudie les unités distinctives minimales ou les phonèmes, en nombre limité dans chaque langue, les traits distinctifs ou traits pertinents qui opposent entre eux les différents phonèmes d'une même langue.

La **prosodie**, qui étudie les traits suprasegmentaux, c'est-à-dire les éléments phoniques qui accompagnent la réalisation de deux ou plusieurs phonèmes et qui ont aussi une fonction distinctive: l'accent, le ton, l'intonation. [8]

1.5. L'Alphabet Phonétique International "API"

L'Alphabet Phonétique International est un alphabet universel utilisé pour la transcription phonétique des mots. Il permet d'indiquer la prononciation d'un mot, ce qui est utile lorsque celle-ci n'est pas évidente, ou lorsqu'il s'agit d'un mot étranger au lecteur. Cette transcription se note entre crochets droits.

Il a été initialement développé par des phonéticiens Britanniques et Français sous les auspices de l'Association Phonétique Internationale, fondée à Paris en 1886 par Paul Passy (cette association et son alphabet sont plus connus sous le sigle API). La plupart des lettres sont empruntées à l'alphabet latin ou en dérivent, certaines sont d'origine grecque, et quelques caractères sont sans rapport apparent avec les lettres ordinaires. Le principe général est d'employer un symbole unique pour chaque segment sonore de la parole, en évitant les combinaisons de lettres. Des signes diacritiques peuvent être combinés avec les symboles de l'API pour transcrire des valeurs phonétiques légèrement modifiées ou des articulations secondaires. Il existe également des symboles spéciaux pour noter des phénomènes suprasegmentaux, comme les tons mélodiques [5].

Révisé en 1990 et en 1993, voit tableau 1.1, l'API comprend 118 caractères principaux, 76 diacritiques et 23 marques de tons.

Tableau 1.1. : Alphabet International Phonétique [9]

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)											
CONSONANTS (PULMONIC)											
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ʀ					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)			SUPRASEGMENTALS		TONES & WORD ACCENTS	
Clicks	Voiced implosives	Ejectives	Primary stress	Secondary stress	LEVEL	CONTOUR
◌ ʘ Bilabial	ɓ Bilabial	ʼ as in:	ˈ founəˈtɪʃən	ˌ	˥ Extra high	˩ Rising
◌ ǀ Dental	ɗ Dental/alveolar	ɓ' Bilabial	ˈ eː	ˌ	˨ High	˨˩ Falling
◌ ǃ (Post)alveolar	ɟ Palatal	ɗ' Dental/alveolar	ˈ e˞	ˌ	˨ Mid	˩˨ High rising
◌ ǂ Palatoalveolar	ɠ Velar	ɟ' Velar	ˈ ɪ.ækt	ˌ	˨ Low	˩˨ Low rising
◌ ǁ Alveolar lateral	ʄ Uvular	ʂ' Alveolar fricative	ˈ	ˌ	˨ Extra low	˩˨ Rising-falling etc.
			◌ ˌ Minor (foot) group	◌ ˌ	˩ Downstep	˩˨ Global rise
			◌ ˌ Major (intonation) group	◌ ˌ	˩ Upstep	˩˨ Global fall
			◌ ˌ Linking (absence of a break)			

Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS		
ɱ Voiceless labial-velar fricative	ɕ ʑ Alveolo-palatal fricatives	
ɰ Voiced labial-velar approximant	ɹ Alveolar lateral flap	
ɰ Voiced labial-palatal approximant	ɥ Simultaneous ʃ and x	
ħ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.	
ʕ Voiced epiglottal fricative		
ʔ Epiglottal plosive		

DIACRITICS		
◌ ◌ Voiceless	◌ ◌ Breathy voiced	◌ ◌ Dental
◌ ◌ Voiced	◌ ◌ Creaky voiced	◌ ◌ Apical
◌ ◌ Aspirated	◌ ◌ Linguolabial	◌ ◌ Laminal
◌ ◌ More rounded	◌ ◌ Labialized	◌ ◌ Nasalized
◌ ◌ Less rounded	◌ ◌ Palatalized	◌ ◌ Nasal release
◌ ◌ Advanced	◌ ◌ Velarized	◌ ◌ Lateral release
◌ ◌ Retracted	◌ ◌ Pharyngealized	◌ ◌ No audible release
◌ ◌ Centralized	◌ ◌ Velarized or pharyngealized	
◌ ◌ Mid-centralized	◌ ◌ Raised	
◌ ◌ Syllabic	◌ ◌ Lowered	
◌ ◌ Non-syllabic	◌ ◌ Advanced Tongue Root	
◌ ◌ Rhoticity	◌ ◌ Retracted Tongue Root	

Dans le but d'inclure les pathologies langagières, l'API a intégré un alphabet dit "Disordered speech Alphabet" (Alphabet pathologique), voir tableau 1.2.

Tableau 1.2. : Alphabet phonétique pathologique

ExtIPA SYMBOLS FOR DISORDERED SPEECH (Revised to 1997)										
CONSONANTS (other than those on the IPA Chart)										
	bilabial	labiodental	dentolabial	labioalv.	linguolabial	interdental	bidental	alveolar	velar	velophar.
Plosive		<u>p</u> <u>b</u>	<u>p̥</u> <u>b̥</u>	<u>p̥</u> <u>b̥</u>	<u>t̥</u> <u>d̥</u>	<u>t̥</u> <u>d̥</u>				
Nasal			<u>m̃</u>	<u>m̃</u>	<u>ñ</u>	<u>ñ</u>				
Trill					<u>r̃</u>	<u>r̃</u>				
Fricative: central			<u>f̃</u> <u>ṽ</u>	<u>f̃</u> <u>ṽ</u>	<u>θ̃</u> <u>ð̃</u>	<u>θ̃</u> <u>ð̃</u>	<u>h̃</u> <u>ɦ̃</u>			<u>f̃</u> <u>ɰ̃</u>
Fricative: lateral+central								<u>ɬ</u> <u>ɮ</u>		
Fricative: nasal	<u>ɱ̃</u>							<u>ɳ̃</u>	<u>ɳ̃</u>	
Percussive	<u>w</u> <u>W</u>						<u>ɸ</u>			
Approximant: lateral					<u>l̥</u>	<u>l̥</u>				

DIACRITICS								
↔	labial spreading	<u>ɸ̃</u>	strong articulation	<u>f̥</u>	~	denasal	<u>ɱ̃</u>	
˘	dentolabial	<u>v̥</u>	weak articulation	<u>v̥</u>	̣	nasal escape	<u>ṿ</u>	
˙	interdental/bidental	<u>ɸ̣</u>	\	reiterated articulation	<u>p̣p̣</u>	̤	velopharyngeal friction	<u>ɰ̤</u>
=	alveolar	<u>ɬ̣</u>	♯	whistled articulation	<u>ɬ̣</u>	↓	ingressive airflow	<u>p↓</u>
˘	linguolabial	<u>ɸ̣</u>	→	sliding articulation	<u>θ̣</u>	↑	egressive airflow	<u>!↑</u>

CONNECTED SPEECH		VOICING	
(.)	short pause	˘	pre-voicing
(...)	medium pause	˘˘	post-voicing
(...)	long pause	(˘)	partial devoicing
<i>f</i>	loud speech [<i>{f loud f}</i>]	(˘)	initial partial devoicing
<i>ff</i>	louder speech [<i>{ff louder ff}</i>]	(˘)	final partial devoicing
<i>p</i>	quiet speech [<i>{p quieter p}</i>]	(˘)	partial voicing
<i>pp</i>	quieter speech [<i>{pp quieter pp}</i>]	(˘)	initial partial voicing
<i>allegro</i>	fast speech [<i>{allegro fast allegro}</i>]	(˘)	final partial voicing
<i>lento</i>	slow speech [<i>{lento slow lento}</i>]	=	unaspirated
<i>crescendo, rallentando</i> , etc. may also be used		h	pre-aspiration

OTHERS			
()	indeterminate sound	(())	extraneous noise ((2 sylls))
(̥), (̃)	indeterminate vowel, plosive, etc.	!	sublaminal lower alveolar percussive click
(̥), (̃)	indeterminate voiceless plosive, etc.	!!	alveolar & sublaminal click ('cluck-click')
()	silent articulation (f), (m)	*	sound with no available symbol

© 1997 ICPLA Reproduced by permission of the International Clinical Phonetics & Linguistics Association.

Parmi les conséquences les plus distinguées de la transcription phonétique est que tout phonème peut être prononcé par divers locuteurs mêmes non natifs de la langue transcrite.

1.6. Alphabet Phonétique de l'Arabe standard

La langue Arabe standard comprend 34 phonèmes, dont 3 voyelles courtes, 3 voyelles longues et 28 consonnes.

1.6.1 Correspondance Organes –lieux d'articulation

L'articulation complexe du système phonatoire donne lieu à une segmentation par région, voir figure 1.7. Cette dernière permet à un locuteur de distinguer l'emplacement exacte ou l'adresse du phonème, ce qui par la suite permet de :

- distinguer entre les phonèmes;
- distinguer entre les articulateurs ayant participé à l'articulation;

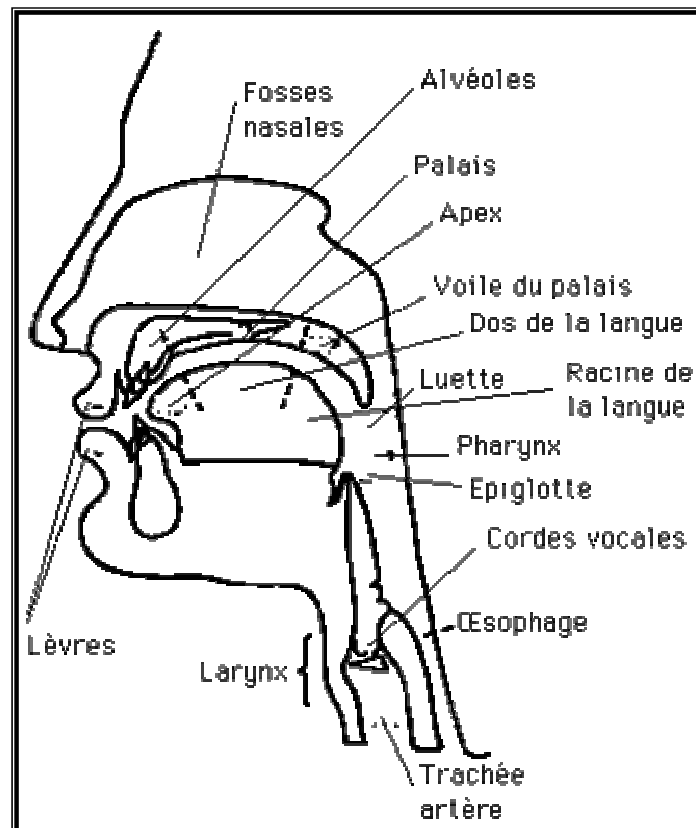


Figure 1.7. : Lieux d'articulation des phonèmes

Le tableau 1.3 montre les correspondances de la segmentation articulatoire de la figure 1.7

Tableau 1.3. : Correspondance organes - lieu d'articulation [10]

Organes	Manifestations selon le point d'articulation
Lèvres	Labiale
Dents	Dentale
Alvéoles des dents	Alvéolaire
Palais	Palatale
Voile du palais	Vélaire
Luette	Uvulaire
Pointe de la langue	Apicale
Dos de la langue	Dorsale
Pharynx	Pharyngale
Cordes vocales	<ul style="list-style-type: none"> ❖ sonore ou voisée (vibration des cordes vocales) ❖ sourdes ou non voisée (pas de vibration des cordes vocales)
Trachée artère	
Oesophage	

La langue Arabe standard contient des lettres différentes de celles du Français ou de l'Anglais, le tableau 1.4 illustre les différents phonèmes Arabes ainsi que leur lieux d'articulation et leur transcription phonétiques.

1.6.2 Segmentation d'El Khalil

El Khalil bien avant l'API, relate les lieux d'articulation des phonèmes de la langue Arabe, voir tableau 1.5, cette segmentation fut utilisée par les disciples pour bien prononcer le Coran.

Tableau 1.5. : Récapitulatif des phonèmes Arabes ainsi que leur lieux d'articulation selon El Khalil

[11]

5			17			
4			16	/ /		
			15			
3			14	/ /		
			13	/ /		
			12	/ /		
			11			
			10			
			9			
			8			
			7	/ /		
			6			
			5			
2			4	/		
			3	/		
			2	/		
1			1	-	-	-

Pour des raisons d'apprentissages, plusieurs phonéticiens et/ou orthophonistes utilisent les lieux d'articulation ou la manière d'articuler selon des schémas illustratifs comme mentionné dans les figures 1.8. à 1.14. [11].

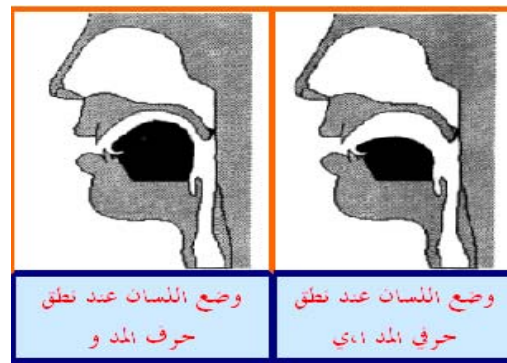


Figure 1.8. : Lieux d'articulation des phonèmes : [أ] , [و] , [ي]



Figure 1.9. : Lieux d'articulation des phonèmes [ق] , [ك] , [ش]



Figure 1.10. : Lieux d'articulation des phonèmes [ل] , [ض] , [ج]



Figure 1.11. : Lieux d'articulation des phonèmes [ن] , [ر] , [ت]



Figure 1.12. : Lieux d'articulation des phonèmes [ر] , [د] , [ط]



Figure 1.13. : Lieux d'articulation des phonème [ظ] , [ذ] , [ض]



Figure 1.14. : Lieux d'articulation des phonèmes [س] , [ص] , [ز]



Figure 1.15. : Position des lèvres lors de la prononciation des phonèmes [م] , [ي] , [و] , [ف]

1.8. Système vocalique Arabe

1.8.1. Réalisation des voyelles

C'est la classe de phonèmes issues de la vibration continue des cordes vocales, sans obstruction, nous distinguons 3 voyelles [a], [o], [i], appelées dans l'ordre :

الكسرة / الضمة / الفتحة

La réalisation phonétique des voyelles est très variable et dépend à la fois de :

- ❖ l'origine géographique des locuteurs ;
- ❖ l'environnement consonantique et de la place de la voyelle dans le mot ;
- ❖ la place de l'accent du mot (tendance à abréger les longues non accentuées chez beaucoup d'arabophones). [12]

1.8.2 Triangle vocalique

Afin de situer les voyelles en terme de degré d'aperture des cordes vocales et en terme de durée de voisement, celles-ci sont mentionnées dans un triangle vocalique contenant :

- deux plans représentant la durée de voisement, de la plus faible à la plus forte
- Le bas des triangles représente le degré d'aperture maximale, le haut des triangles représente le degré d'aperture minimal, comme illustré dans la figure 1.16.

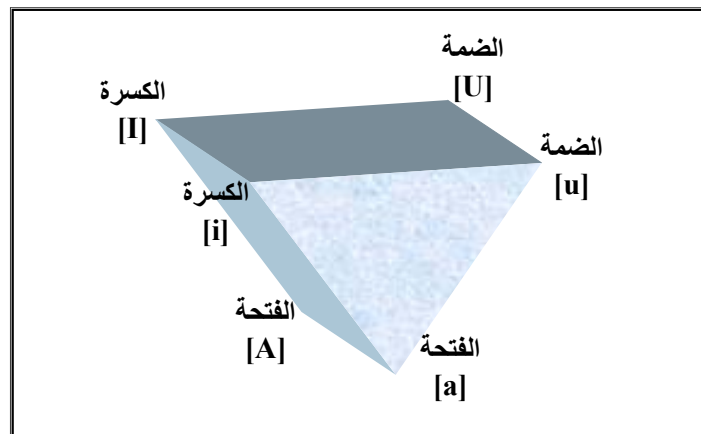


Figure 1.16. : Triangle vocalique de la langue Arabe

1.8.3. Les consonnes

Les consonnes sont tous les autres phonèmes ne nécessitant pas de vibration périodique des cordes vocales, voir tableau 1.3.

1.8.4 Redoublement des consonnes

Toutes les consonnes peuvent être redoublées. on dit aussi géminer une consonne, on prolonge et on renforce l'articulation de cette consonne. La gémination est indiquée par un signe graphique spécifique appelé "chadda" (ّ).

La gémination joue un rôle très important en morphologie. Il est donc essentiel de bien l'entendre et de bien la réaliser.

درس	: [darasa]	"il a étudié"
درّس	: [darrasa]	"il a enseigné"

1.8.5. Les emphatiques [ص], [ظ], [ط], [ض]

Il s'agit d'un trait articulatoire spécifique de la "pharyngalisation" : le son produit est plus grave que pour le son non emphatique correspondant. On l'obtient en modifiant la forme du résonateur buccal dans sa partie arrière par rétraction et exhaussement de la racine de la langue.

Il est très important de bien distinguer le son emphatique du non emphatique correspondant :

سيف	[sayf]	"épée"
صيف	[Sayf]	"été"

Souvent la présence d'une consonne emphatique dans un mot "contamine" l'environnement, consonantique et vocalique, et c'est toute la syllabe qui est emphatisée.

1.9. Structure de la langue Arabe en succession phonétique

Dans un besoin d'analyser toutes les occurrences d'un phonème et de cerner tous les cas possibles de composition phonétique, le langage Arabe obéit à une structure phonétique minimale, car les autres structures complexes ne sont que la concaténation de ces formes de base, comme défini dans le tableau 1.6.

Tableau 1.6 : Structure phonétique minimale du langage Arabe [13]

Structure syllabique	Correspondance
[CV]	Consonne – voyelle courte
[CVV]	Consonne – voyelle longue
[CVC]	Consonne – voyelle courte – Consonne
[CVVC]	Consonne – voyelle longue – Consonne
[CVCC]	Consonne – voyelle courte – Consonne – Consonne

1.10. Etudes réalisées

Le contexte se définit d'après le besoin, notre intérêt concerne les lieux d'articulation donc, il est judicieux de s'intéresser à la manière dont ces phonèmes sont réalisés pour comprendre leurs particularités et ainsi déterminer les caractéristiques communes et discriminantes, cette approche est fortement utilisée en segmentation phonémique [14], [15], [16], [17], [18].

Nous pouvons distinguer, d'après les propriétés articulatoires relatives aux différents lieux d'articulation les particularités acoustiques, énergétiques, spectrales, cesptrales, etc.... des phonèmes et ainsi décider des propriétés qui les discriminent, le résultat n'en sera que plus robuste.

La distinction entre phonèmes ou segmentation au niveau phonémique est une approche très répandue en Traitement de la Parole continue, car en reconnaissant les éléments de base on peut :

- ◆ distinguer entre les mots prononcés dans une même langue et reconstituer les phrases dans un contexte de parole continue, ou localiser les mots les plus significatifs dans un contexte de reconnaissance de mots isolés ou de nombres isolés comme pour le portable, la voiture, etc. [19][20],[21].
- ◆ distinguer entre différentes langues voire même entamer une reconnaissance directe de la langue pour engager une procédure de paiement par téléphone, par exemple, pour différents locuteurs non natifs de la langue du pays, ou pour des

renseignements en différentes langues avec un aiguillage intelligent, cas du projet Raphaël, ou les locuteurs des deux cotés du système ne parlent pas la même langue. [22], [23].

D'autres études comme celle portant sur la logopédie, introduit une notion phonémique très intéressante en terme de comparatif à des mots de référence selon le critère de Borel-Maisonny [24], [25].

Les applications n'en finissent pas, et l'amélioration des méthodes discriminatives en est l'un des piliers fondamentaux.

1.11 Conclusion

L'étude du Traitement Automatique de la Langue nécessite un pré requis dans différents domaines, à travers ce chapitre, nous avons essayé d'introduire quelques notions en anatomie articulaire par une étude simplifiée du système phonatoire et des différents articulateurs, ainsi que leurs rôles respectifs dans la production vocale, suivi de la structure de la langue Arabe et de sa représentation dans l'Alphabet Phonétique Internationale.

Dans le prochain chapitre, nous allons introduire les pathologies du système phonatoire, en montrant l'impact de chaque maladie sur la production vocale. Les conséquences de ces maladies, si celles-ci ne sont pas prises en compte à temps, cas du bégaiement, du chuintement, du schlintement ou des différents types de cancer.

CHAPITRE 2

PATHOLOGIES DU LANGAGE PARLE

2.1 Introduction

Lorsqu'on parle de pathologies, une référence intuitive nous fait penser à un médecin, c'est un réflexe très logique, toutefois le domaine pathologie fait intervenir différents scientifiques, par exemple l'utilisation de l'effet doppler en médecine, l'utilisation, en imagerie médicale, du scanner en 3D ainsi que des rayons X. Ces appareils sont le fruit de la technologie qui implique d'autres disciplines telle l'électronique, la mécanique, la physique, etc.

Le traitement des pathologies langagières se situe essentiellement à détecter la zone à traiter, mesurer l'ampleur de la maladie, ou pathologie, intervenir en post ou en pré chirurgie, corriger par un orthodontiste ou par un orthophoniste. Tous ces spécialistes, diffèrent par leur degré d'intervention relatif au stade d'évolution de la pathologie et de son emplacement.

Ce chapitre est une introduction aux pathologies relatives à la parole, ou en d'autres termes « au langage vocal » et aux pathologies subséquentes des différents « articulateurs » lors de la production de la voix et son altération au cours de son cheminement à travers le conduit buccal ou nasal.

2.2. Définitions

Les atteintes peuvent concerner les organes périphériques, atteintes qui gênent la production de la parole : le bec de lièvre, la division palatine, l'insuffisance vélaire, les malformations linguales, labiales ou laryngées.

Il s'agit d'anomalies consistant en des erreurs mécaniques et constantes dans l'exécution du mouvement propre à un phonème [26].

L'articulation est la capacité à articuler les sons de façon permanente et systématique, ce qui nécessite des mouvements précis de la mâchoire inférieure, de la langue, des lèvres, des joues, du voile du palais.

Le trouble d'articulation isolé est donc l'incapacité à prononcer ou à former un certain phonème correctement. C'est une erreur constante, systématique et mécanique pour un phonème donné. Cette erreur est plutôt de type praxique¹ [27].

La Production de la voix normale est basée la qualité de sa production, son intensité, son débit. Une voix pathologique présente une altération d'un ou de plusieurs de ces paramètres [28].

On peut d'ores et déjà classer les troubles du langage en régions pathologiques, c'est ce qui montre que la voix peut être altérée ou modifiée tout le long de sa production, voire disparaître, phénomène décrit par l'apparition d'une aphonie ou absence de voix complètement, surtout lors du cancers des cordes vocales.

2.3. Cordes Vocales saines

Avant de commencer l'étude des pathologies affectant la voix, il serait important de voir l'apparence de cordes vocales saines, car elles sont l'organe le plus important dans la production vocale.

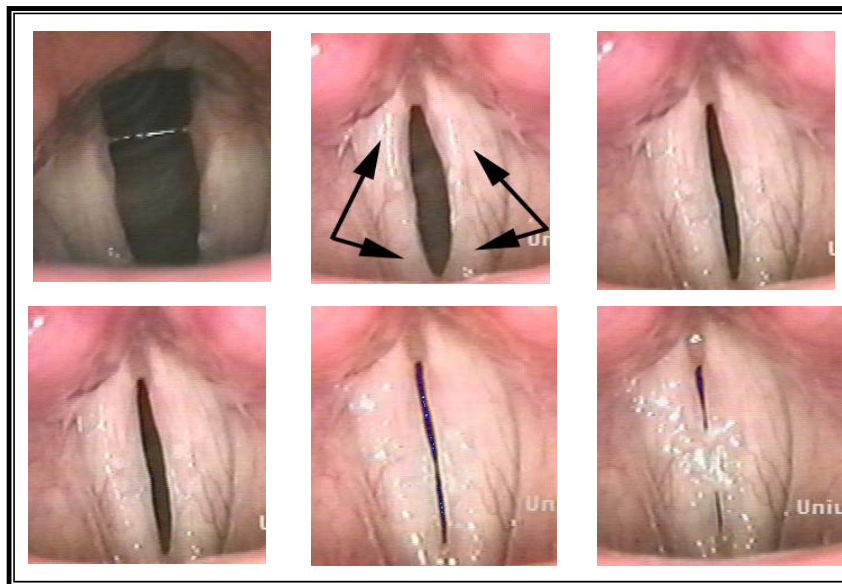


Figure 2.1 : Cordes vocales saines avec différents degrés d'aperture [29]

¹ La praxie se définit comme un mouvement coordonné normal vers un but suggéré.

L'atteinte des cordes vocales par n'importe quelle maladie ou par simple irritation, lorsqu'on crie très fort pendant un événement quelconque, agit essentiellement sur leur manière de vibrer, soit en atténuant le mode vibratoire par une paralysie ou en les rendant plus enroué ou plus âpre, comme illustré dans la figure 2.2.

Les différentes pathologies les plus importantes concernant les cordes vocales sont illustrées ci-après :

2.4. Pathologie des cordes vocales

Nous commençons notre étude sur les pathologies du langage par le lieu le plus important dans la phonation, les cordes vocales.

2.4.1 Nodules

Ce sont des nœuds durs en forme de pointe de flèche situés sur la partie vibratoire de contact des deux cordes vocales, voir figure 2.2.

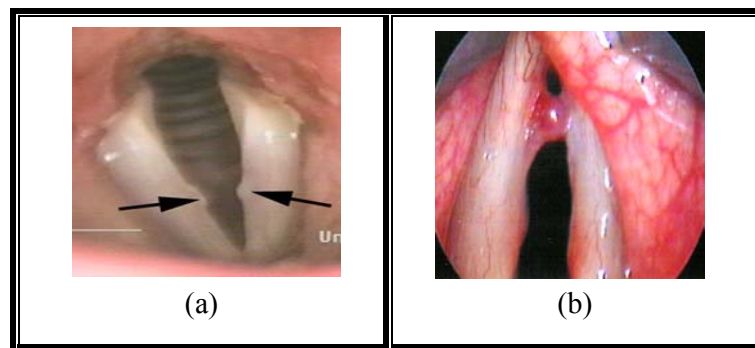


Figure 2.2 : (a, b) Nodules sur les cordes vocales

2.4.2 Paralysie des cordes vocales

Ce phénomène apparaît lorsque l'une des cordes, devient non élastique, comme illustré dans la figure 2.3, ou s'arrête presque de bouger; Cette situation est généralement traitée par intervention chirurgicale soit par une lipo-injection ou thyroplastie de la corde paralysée pour permettre une fermeture complète [29].

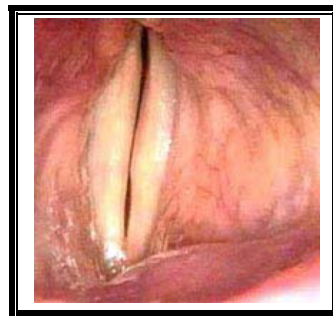


Figure 2.3. Paralysie unilatérale des cordes vocales

2.4.3. Cordes vocales arquées

Lorsque les deux cordes vocales ne se ferment pas complètement, La surface restante affecte la production phonatoire, ceci peut rendre la voix très fragile. [29], [30]

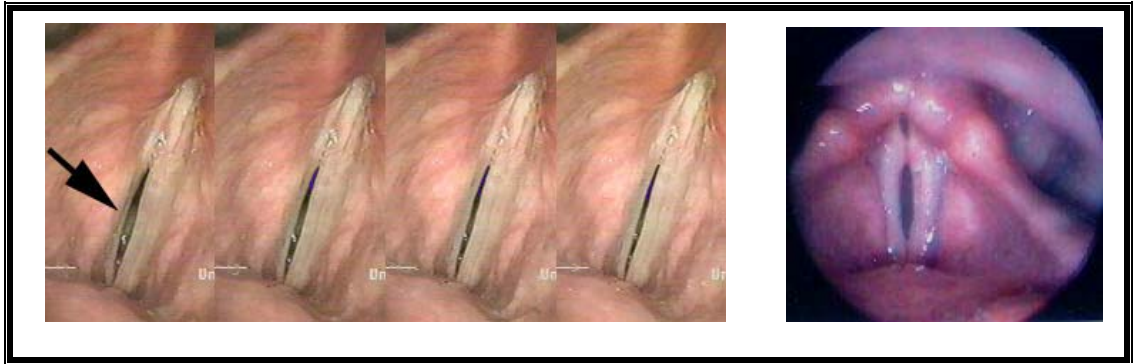


Figure 2.4 : Cordes vocales pathologiques présentant un arc à la fermeture

2.4.4. Polypes dans les cordes vocales

Les polypes, voir figure 2.5 dérangent la fermeture et les vibrations des cordes vocales, cette situation cause l'enrouement, la fatigue des cordes et une diminution du mode musicale.

2.4.5. Odème de Reinke

L'œdème de Reinke est marqué par l'accumulation d'œdème et de fibrose dans la totalité de la corde vocale, voir figures 2.5 et 2.6. L'étiologie est essentiellement le tabac associé ou non à l'alcool. Il s'agit d'une lésion bénigne mais qui peut être associée à un cancer développé ailleurs dans les Voies Aéro digestives Supérieures "V.A.D.S." dans 5 à 10 % des cas. L'accumulation de l'œdème peut conduire à un véritable ballonnement des cordes accompagné de dyspnée. La répartition est globalement identique suivant les deux sexes, mais la population féminine consulte plus facilement du fait de la répercussion de cet œdème chronique sur la voix. En effet le fait marquant de cette dysphonie est l'abaissement de la hauteur vocale. Ce trouble vocal est généralement bien accepté chez les hommes auxquels il donne un caractère "viril". A l'opposé chez la femme, la gêne est manifeste. La patiente est fréquemment appelée Monsieur au téléphone.

Le traitement est microchirurgical et consiste à inciser la corde vocale sur sa face supérieure et à aspirer la glue. L'abstention tabagique prévient la récurrence.

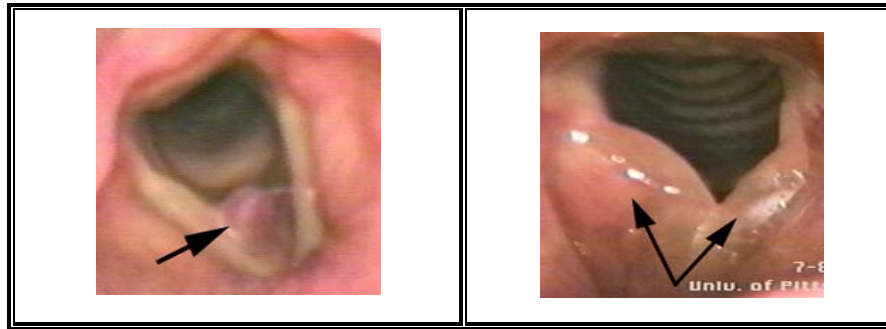


Figure 2.5. : Polype très distingué sur l'une des cordes vocales à gauche [29]

Figure 2.6. : Cordes vocales gonflées à gauche

2.4.6. Kyste localisé au niveau des cordes vocales

Le kyste muqueux, voir figures 2.7 et 2.8, est formé par l'obstruction de canal excréteur d'une glande muqueuse de la corde vocale. Macroscopiquement, on observe une voussure plus ou moins allongée. Plus le kyste est ancien et plus le liquide paraît épais. Le traitement est microchirurgical

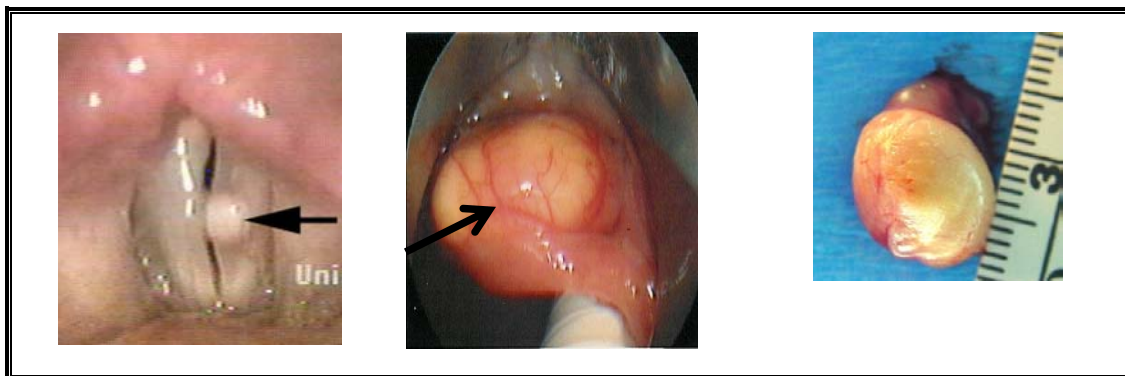


Figure 2.7 : Détail d'un kyste de différents patients

Figure 2.8 : Mesure d'un Kyste vocal après son extraction

2.4.7. Granulomes dans les cordes vocales

Petits amas granulomateux, inflammatoire « constitué de chair », c'est-à-dire de tissu conjonctif se développant sur la muqueuse du larynx et à ses dépens. Quelquefois ils sont observés au niveau de la trachée après une intubation du larynx et de la trachée pour une ventilation assistée, le plus souvent ayant eu lieu au cours d'une anesthésie générale ou un coma.

La flèche sur la figure 2.9. Indique un tissu épais et irrégulier sur les cordes vocales. [31], [32]

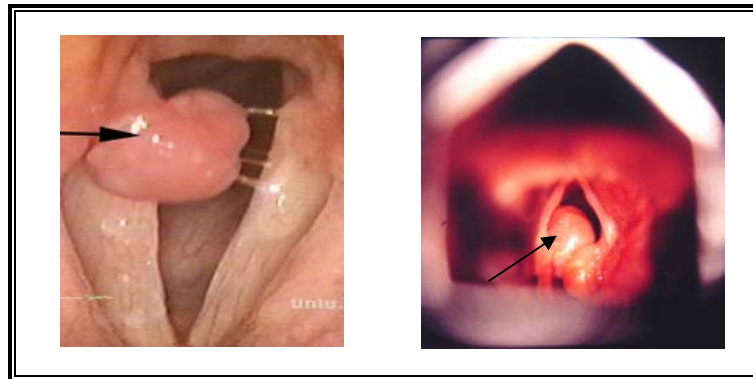


Figure 2.9 : Granulomes de différents patients

2.4.8. Papillomes laryngés

Les différentes flèches sur la figure 2.10 indiquent l'évolution des papillomes, dans le larynx, ceux-ci sont causés par une infection virale.

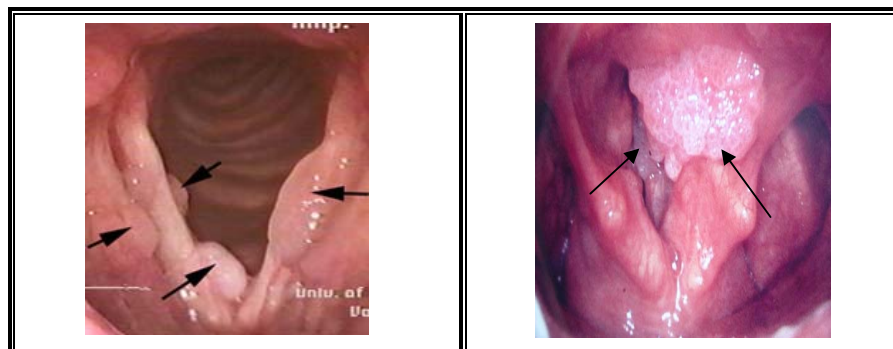


Figure 2.10 : Papillomes chez différents patients

Des études bien détaillées des pathologies indiquées sont décrites par « The National Center for Voice and Speech » et par la clinique Otolaryngology-Head & Neck Surgery [29] [32].

2.4.9. Cancers des Voies Aero-digestives Supérieures "V.A.D.S."

Le cancer, maladie parfois non bénigne, détectable après biopsie, se localise dans différents points du corps humain, entre autres dans le larynx et les différentes cavités vocales et nasales, appelées Voies Aérodigestives Supérieures, voir figure 2.11. [33]

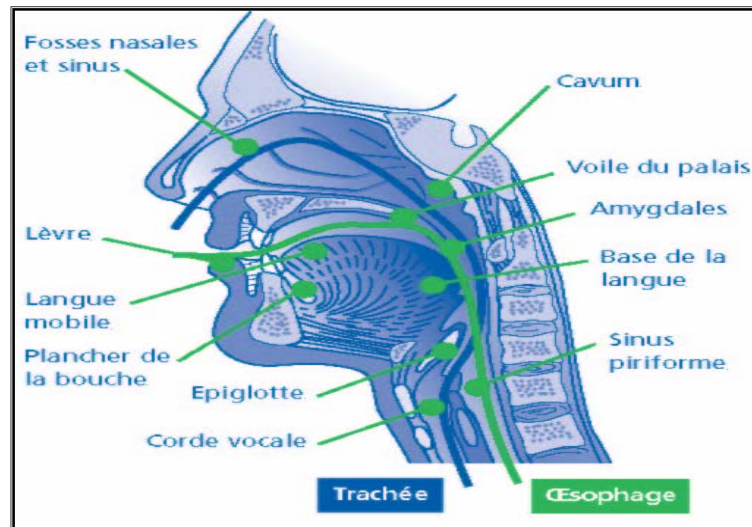


Figure 2.11. Sièges possibles des cancers

2.4.10 Cancer du larynx

Ce cancer, favorisé par l'alcool et le tabac, se voit surtout chez l'homme après 50 ans.

Symptômes

Le signe révélateur est une dysphonie progressive : le patient se plaint d'un enrouement. La dyspnée (gêne à la respiration), la dysphagie (difficulté pour avaler), sont beaucoup plus tardives.

Tout enrouement chronique nécessite un bon examen laryngologique direct.

L'ORL examine sous anesthésie locale le larynx grâce au miroir laryngé. Voir figure 2.14. La tumeur est ainsi observée. Le médecin apprécie ensuite la mobilité du larynx et recherche des ganglions palpables. L'examen endoscopique (laryngoscopie) recherche une localisation cancéreuse oesophagienne et permet la biopsie à la pince de la lésion.

Diagnostic différentiel

- ❖ tumeurs bénignes du larynx (polypes des cordes vocales, nodules vocaux...);

- ❖ tuberculose laryngée;
- ❖ laryngite chronique (qui peut dégénérer) ;
- ❖ atteinte neurologique des cordes vocales.

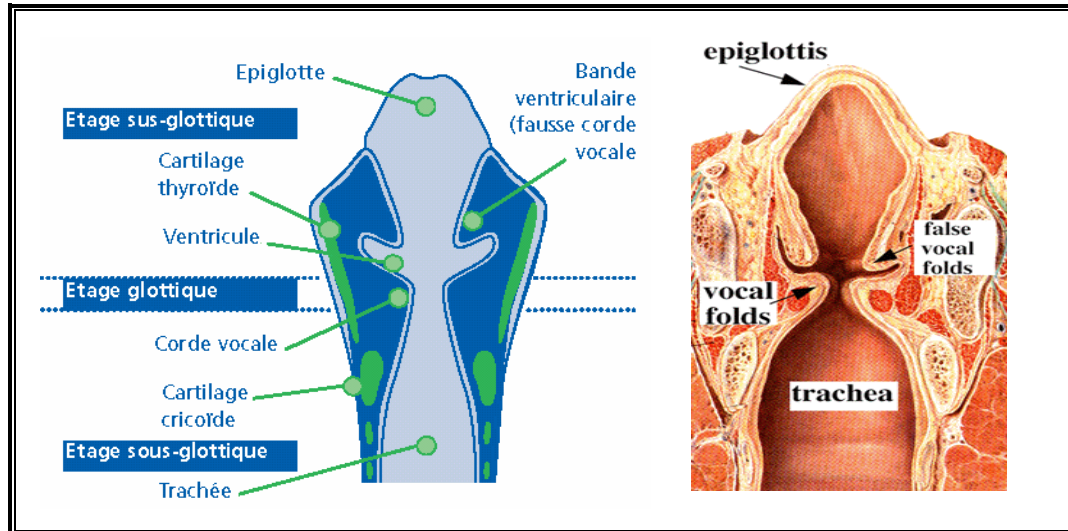


Figure 2.12. Vue en coupe du larynx [33]

2.4.11 Cancer Du Pharynx (Oropharynx et Hypopharynx)

Le cancer du pharynx apparaît en général après les symptômes suivants :

- ◆ une gêne ou une douleur d'un côté de la gorge;
- ◆ une sensation permanente d'un corps étranger ou d'angine traînante d'un seul côté ;
- ◆ une douleur à une oreille ;
- ◆ une difficulté à avaler, une gêne à la déglutition d'un côté, parfois sans douleur;
- ◆ une sensation de brûlure d'un côté de la gorge ;
- ◆ une modification progressive de la voix qui devient couverte, voilée ou rauque ;
- ◆ apparition d'une boule dans le cou qui correspond à un ganglion.

2.4.12. Cancer De La Bouche

- ◆ Le cancer de la bouche apparaît en général après les symptômes suivants :

- ◆ une gêne ou une douleur d'un côté de la bouche;
- ◆ une zone bourgeonnante ou creusée saignante, ne guérissant pas après un traitement d'une anomalie dentaire;
- ◆ un changement de la muqueuse persistant dans la bouche (tache rouge foncé ou blanche ressemblant à un aphte, mais à bords irréguliers);
- ◆ une gêne au port d'un dentier;
- ◆ une douleur à une oreille;
- ◆ une difficulté à avaler;
- ◆ une sensation de chaud, au froid, au vinaigre, au citron.

2.4.12 Cancer des Cordes Vocales (Glottiques)

Ces cancers se révèlent par une modification progressive de la voix qui devient couverte, voilée, rauque (dysphonie). Cette modification persiste et s'aggrave progressivement. Elle est parfois précédée d'épisodes transitoires de laryngite où complique une laryngite chronique ancienne, fréquente chez les fumeurs et/ou les personnes travaillant en atmosphère chaude et sèche, ou chargée de poussières.

2.4.13. Cancer Des Cordes Vocales (Sub - Glottiques)

Ils siègent au niveau de l'épiglotte, par :

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

2.5. Pathologies des autres canaux vocaux

2.5.1 Bec de lièvre

C'est une déformation prénatale; voir figure 2.14, s'attaquant à la lèvre supérieure, d'origine génitale, pouvant être corrigée par une intervention chirurgicale.[34]

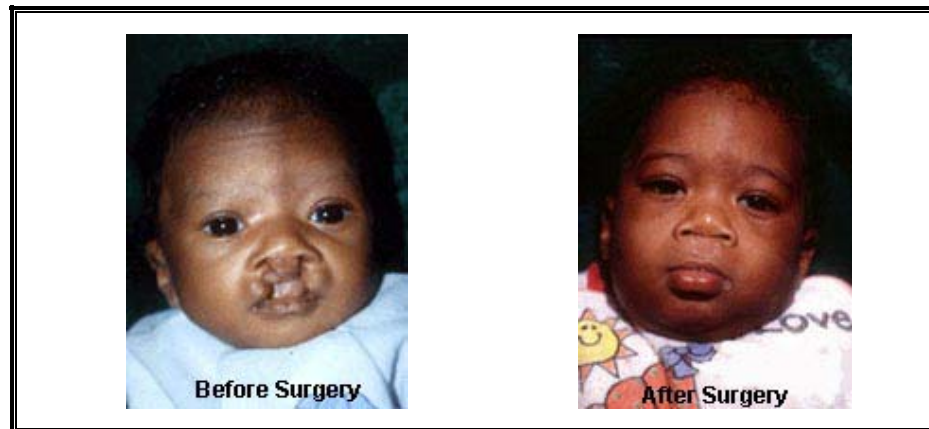


Figure 2.13. Avant intervention / Après intervention

2.4.2 Palais enclavé

L'une des conséquences directes de l'absence du palais, voir figure 2.15 est l'hypernasalité, c'est une pathologie de la résonance de la voix, causé par le dysfonctionnement du mécanisme vélopharyngéale, celle-ci provoque un :

- ❖ nasonnement ouvert ou hyperrhinophonie : le voile du palais ne ferme pas le passage de l'air à la cavité nasale dans le cas de division palatine ou d'opérations des végétations notamment ;
- ❖ nasonnement fermé ou hyporhinophonie : pas de nasalisation pour les consonnes et les voyelles nasales. [34], [45].



Figure 2.14. : Palais totalement absent

Une intervention chirurgicale est à la base de la correction de cette pathologie;

Une étude très intéressante, portant sur les remarques d'enfants parlant l'Arabe avec un palais enclavé, est développée dans [35].

D'autres définitions, causes et traitements sont bien traités dans le guide vocologique ou « Guide to Vocolgy » émis par le Centre National de la Voix et de la Parole, le N.C.V.S. [36].

2.6. Classification des pathologies

La première partie a concerné les pathologies des organes intervenant dans la production ou l'altération de la voix, celles-ci sont classées comme suit:

- ❖ **Disfonctionnement fonctionnelle** : L'organe existe mais, il y'a eu soit un mauvais apprentissage, soit une maladie en cours d'évolution ce qui présente un symptôme de pathologie de la parole, si la détection de la pathologie n'est pas effectuée à temps [37].
- ❖ **Disfonctionnement organique** : L'organe existe ou est absent, cas du palais enclavé ou laryngectomie, mais ne peut exécuter la tâche préconçue, soit par atrophie cas de la langue trop courte, soit par surdimensionnement cas du volume du palais démesurée,

Les différents défauts émanant de ces pathologies sont classés selon le diagramme de la figure 2.16.

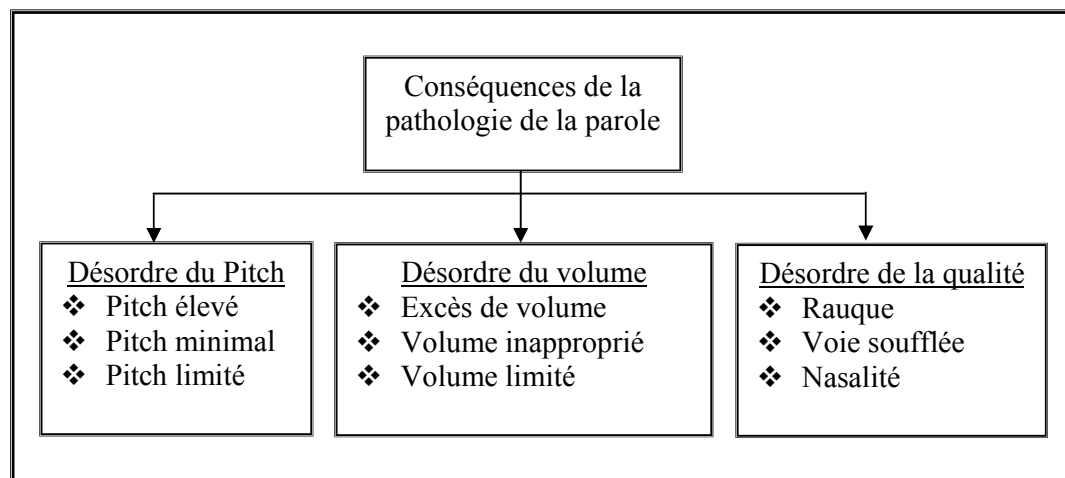


Figure 2.15. : Diagramme de classement des pathologies [38]

2.7. Défauts de la voix détectés par l'oreille

Les différents défauts de la voix détectés par l'oreille humaine sont :

2.7.1. Le blésemment ou zézaiement

Le blésemment ou zézaiement est un défaut de prononciation qui consiste en la substitution de [ʃ] (une consonne chuintante) par [s] (une sifflante) et de [g] ou [j] (Consonnes chuintantes) par [z] (sifflante) [39].

2.7.2. Chuintement

Le chuintement est la prononciation du [s] et du [z] à la manière du [ʃ] et du [j] Français [39].

Exemple :

- J'ai pris l'autobus jusqu'à la gare Saint-Lazare
- J'ai pris l'autobuch juchqu'à la gare Chaint-Lajare.

2.7.3. Le rhotacisme

Le rhotacisme (terme formé à partir du grec ρ, [r]) est une modification phonétique complexe, consistant en la transformation d'un phonème en [r]; Dans d'autres langues comme le Français c'est avec le [z] que le [r] [40].

Pour la langue arabe c'est la confusion entre le [r] et le [ʁ].

Donc, au lieu de prononcer 'ريح' [rihø].

La personne atteinte de rhotacisme prononce "ريح" [ʁihø],

2.7.4. Nasonnement

C'est l'altération du son de la voix; le nasonnement provient de la diminution de la résonance nasale par suite de l'obstruction du nez, de la présence de végétations adénoïdes, etc., et produit une déformation des syllabes nasales, [an], [on], [in], et des consonnes nasales, telles que [m], que l'on prononce [b] [41].

Exemple : En prononçant [pa] l'air ne doit pas passer par le nez.

Le nasonnement est l'inverse du rhume, ou l'air ne peut pas passer par le nez.

2.7.5. Bégaiement

C'est le trouble de la communication affectant le débit et le rythme de la parole se traduisant par:

- ◆ une forme clonique : répétition;
- ◆ une forme tonique : blocage;
- ◆ des troubles associés;

Si rien n'est entrepris, sur 4 enfants de 2 à 5 ans commençant à bégayer, 1 restera bègue à l'âge adulte. Il est nécessaire d'intervenir le plus tôt possible pour ne pas prendre le risque de la chronicisation [42], [43].

Les signes d'appel et manifestations du bégaiement se présentent comme suit :

- ◆ répétition de sons ou syllabes supérieures à 3 (ex: tou tou tou toupie) ;
- ◆ prolongation de sons ;
- ◆ blocage de syllabes ;
- ◆ répétitions de mots, de parties de phrases ;
- ◆ reprise d'énoncés ;
- ◆ hypertonie, blocages respiratoires lors de la prise de parole;
- ◆ comportement ou modification du comportement : colères, retrait, timidité, énurésie;
- ◆ Comportement verbal: refus ou repli ;
- ◆ antécédents de bégaiement dans la famille [27].

2.7.6. Le clicchement

Le clicchement est un défaut de prononciation se caractérisant par le fait d'ajouter le son [ll] (double L) mouillé, positionné après certaines consonnes.

Une consonne mouillée est articulée avec le son j. Par exemple l dans grisaille.

Un exemple de clicchement : prononcer chilluchoter au lieu de chuchoter [39].

2.7.7. Le gammacisme

Défaut de prononciation se caractérisant par la difficulté voire l'impossibilité de prononcer les consonnes gutturales, [k] à la place de [g].

2.7.8. Retard de parole

Le retard de la parole est l'altération de phonèmes ou de groupes de phonèmes, par leur mise en ordre séquentielle à l'intérieur d'un même mot, le stock phonétique étant acquis. C'est la forme du mot dans son ensemble qui ne peut être reproduite.

Un parler bébé qui perdure au-delà de 4 ans est caractérisé par :

- ◆ des omissions mots raccourcis ou élidés : fleur/feur, herbe/è;
- ◆ des inversions brouette / bourette;
- ◆ des assimilations : lavabo lalabo ou vavabo;
- ◆ des interversions : kiosque / kiokse;
- ◆ des simplifications : parapluie / papui ...;
- ◆ des substitutions : train / crain, fleur fieur;
- ◆ des élisions de syllabes finales : pelle pè, assiette assiè...

Et de façon plus globale :

- ◆ des problèmes de perception auditive;
- ◆ une mauvaise structuration de la perception du temps;
- ◆ une mauvaise structuration de la chronologie des sons;
- ◆ des difficultés motrices diverses;
- ◆ une attention auditive labile;
- ◆ une immaturité psychoaffective;
- ◆ un refus de grandir ...

Fréquemment, un retard de langage et/ou un trouble d'articulation peuvent être associés au retard de parole [27].

2.8. Pathologie concernée par notre travail

2.8.1. Sigmatisme

Terme issu de la lettre grecque sigma, c'est la difficulté que présentent certaines personnes à prononcer le phonème [s]. Cette affection ne doit pas être confondue avec le zézaïement qui est un défaut de prononciation d'une personne prononçant le son [s] comme étant [z], le son [ʒø] comme [s] ou le son [s] comme [sø]. On dit également zozoter.

Ce défaut est généralement relié à une déviation par la langue, dans le processus d'écoulement d'air.

Il y'a deux types de sigmatismes

- ◆ **Sigmatisme latéral ou schlintement** : L'air s'échappe sur le côté de la bouche;
- ◆ **Sigmatisme interdental ou zozotement** : la langue vient buter contre les incisives supérieures ou se place entre les dents et produit une interposition linguale lors de l'émission des phonèmes [s] et [z] [39].

2.8.1.1 Sigmatisme et constrictives

Le sigmatisme intervenant sur les consonnes constrictives peut avoir plusieurs appellations, suivant son origine :

- ◆ Le sigmatisme nasal est dû à un positionnement de la langue qui rend impossible le passage de l'air par la cavité buccale ;
- ◆ Le sigmatisme dorsal est également dû à un soulèvement de la langue excessif ;
- ◆ Le sigmatisme occlusif est le remplacement systématique de toute consonne constrictive par la consonne occlusive dont le point d'articulation est le plus proche [44].

Ce sigmatisme concerne le remplacement de [ʃ], [j], [s], [z] par [θ] et [d] ou par [f] et [v] [45].

L'exemple suivant, tableau 2.1. illustre l'un des mots du corpus que nous avons pu enregistrer, et la prononciation défectueuse qui est transcrite phonétiquement :

Tableau 2.1. : Transcription Phonétique du mot présentant la prononciation pathologique

	[ʃ a χ s i j a]	[θ a χ t i j a]	نخثية
		[θ a χ ʃ i j a]	نخشثية

2.9. La rééducation orthophonique

Les techniques de rééducation orthophonique utilisées sont multiples et sont en général adaptées au cas de chaque enfant : elles dépendent à la fois de la nature de son trouble, du caractère et de la problématique de l'enfant, de ses possibilités intellectuelles, d'où une prise en charge individuelle.

La rééducation peut être assez technique ou davantage orientée vers la P.R.L. (Pédagogie Relationnelle du Langage)

2.9.1. Au niveau articuloire

Chaque fois qu'un phonème n'est pas correctement émis, le rééducateur sert de modèle acoustique et articuloire à l'enfant. Il s'agit de créer de nouveaux automatismes audio kinesthésiques correspondant à une position organique correcte (**le travail devant la glace permettant à l'enfant de la visualiser et d'en prendre conscience**).

2.9.2. Bain de langage

L'orthophoniste essaie de mettre le sujet à l'aise en l'introduisant dans une sphère de langage au moyen de jeux : lotos, jeux de mémorisation, imagiers, lexicata.

2.9.1.3. Travail de perception

Cette phase concerne l'aide à la discrimination et à la mémorisation auditive, ainsi le sujet se rappelle de ce qu'il faut bien prononcer.

2.9.2 La Dimension Relationnelle

Les liens privilégiés qui se nouent entre l'enfant et le rééducateur, ne sont pas qu'un moyen d'appliquer la technique. Ils ont aussi pour but l'épanouissement de l'enfant. La relation duelle est gratifiante pour lui. Elle lui permet de se confier, de ressentir du plaisir dans un échange avec un adulte.

En lui montrant qu'il est capable de décider et de réaliser lui-même, en lui faisant constater ses propres progrès et en le valorisant, l'orthophoniste aide l'enfant à prendre confiance en lui, en ses capacités, et à trouver la volonté de surmonter ses échecs.

D'autre part, il est évident que la possibilité de s'exprimer oralement ne suffit pas, encore faut-il avoir envie de communiquer. Seule une situation duelle affectivement sécurisante et valorisante peut faire naître et grandir chez l'enfant ce besoin de la communication [46].

2.10. Conclusion

A travers ce chapitre, nous avons cité un éventail assez large de pathologies langagières, ayant trait aux variations phonétiques.

Le choix du sigmatisme occlusif, pathologie concernant les défauts de prononciation du [ʃ] et [s] prononcés comme [θ] et [d], porte essentiellement sur la disponibilité d'un corpus maladif ainsi qu'une tendance à mettre une méthodologie d'évaluation par un système d'aide, qui sera bénéfique à l'orthophoniste et au patient.

Dans le chapitre suivant, différentes techniques d'analyse du signal vocal sont présentées, en vue de définir un processus permettant :

- ❖ d'aider l'orthophoniste en lui donnant un score de bonne ou de mal prononciation du patient.
- ❖ de faire travailler le patient à la maison, en s'enregistrant, en se réécoutant et en observant les diverses articulations à réaliser afin de bien prononcer.

CHAPITRE 3

TECHNIQUES D'ANALYSE DU SIGNAL VOCAL

3.1. Introduction

Dans ce chapitre, nous allons présenter différentes techniques d'analyse du signal vocal, en mettant en évidence notre choix concernant les méthodes utilisées lors de la segmentation phonémique, à travers des comparatifs avec des travaux similaires.

Nous introduirons ensuite l'utilisation des chaînes de Markov et les mélanges de gaussiennes pour la modélisation des mots ou phonèmes prononcés. Ces techniques d'analyse serviront, dans un système interactif, à évaluer la mal prononciation d'un phonème ou d'un mot et à donner un score de vraisemblance avec un corpus préenregistré de différents locutrices et locuteurs saints.

Le système décide du pourcentage du défaut dans la prononciation et détermine le point de défaillance en remontrant les lieux d'articulation par un mode feedback visuel et auditif.

3.2. Historique

La recherche sur la parole a débuté depuis des siècles, les musulmans se sont intéressés à la prononciation, prenons à titre d'exemple cette décomposition phonémique de la langue arabe, il y'a de cela plusieurs siècles, comme l'a mentionné El Khalil [47].

مخارج الحروف عند الخليل سبعة عشر مخرجاً. وعند سيبويه وأصحابه ستة عشر، لإسقاطهم الجوفية. وعند الفراء وتابعيه أربعة عشر، لجعلهم مخرج الذلقية واحداً. ويحصر المخارج الحلق واللسان والشفتان، ويعمها الفم .
 فللحلق ثلاثة مخارج، لسبعة أحرف: فمن أقصاه الهمزة، والألف، لأن مبدأه من الحلق، ولم يذكر الخليل هذا الحرف هنا، والهاء .
 ومن وسطه العين والحاء المهملتان .
 ومن أدناه الغين والحاء .
 ولللسان عشرة مخارج لثمانية عشر حرفاً: فمن أقصاه مما يلي الحلق وما يحاذيه من الحنك الأعلى القاف. دونه قليلاً مثله الكاف .
 ومن وسطه الحنك الأعلى الجيم والشين والياء .
 ومن وسطه ووسط الحنك الأعلى الجيم والشين والياء .
 ومن إحدى حافتيه وما يحاذيها من الأضراس، من اليسرى .
 صعب ومن اليمنى أصعب، الضاد .
 ومن رأس حافته وطرفه ومحاذيها من الحنك الأعلى من اللثة اللام .
 ومن رأسه أيضاً ومحاذيه من اللثة النون .
 ومن ظهره ومحاذيه من اللثة الراء .
 هذا على مذهب سيبويه ، وعند الفراء وتابعيه مخرج اللثة واحد .
 ومن رأسه أيضاً وأصول الثنيتين العلين الطاء والتاء والذال .
 ومن رأسه أيضاً وبين أصول الثنيتين الطاء والذال والتاء .
 ومن طرفي الثنيتين وباطن الشفة السفلى الفاء .
 وللشفتين الباء والميم والواو .
 والغنة من الخيشوم من داخل الأنف، هذا السادس عشر .
 وأحرف المد من جو الفم وهو السابع عشر

- ◆ En 1791, W.Von Kempelen a construit une machine mécanique qui mimique la voix humaine;
- ◆ En 1939, l'exploit de H. Dudley avec le « Voder », basé sur des composants électriques faisait l'analyse et la synthèse de la voix humaine, les sons étaient analysés puis reproduits, il fallait parfois une semaine pour faire fonctionner le système avec ces petits instruments de musique spéciaux;
- ◆ En 1975 Baker et Jelinek en 1976, travaillant dans les laboratoires d'IBM, proposèrent qu'au lieu de stocker différentes occurrence du mot en mémoire, parfois des milliers de mots avec différentes prononciations dans différents contextes, le système utilise un modèle abstrait des unités à reconnaître pas nécessairement les mots, en intégrant les automates finis ou chaînes de Markov cachés "HMM, Hidden Markov Models", qui utilisent les méthodes d'entraînement proposés par Baum en 1975 ou de Viterbi en 1967, d'une façon similaire à la DTW;

Toutefois, ce qui est intéressant dans les chaînes de Markov cachés par rapport à la comparaison dynamique par déformation temporelle est la prise en compte des différentes occurrences d'un mot intégrant la variabilité intra locuteurs et inter locuteurs, ceci a permis d'aller jusqu'à l'apprentissage au niveau phonémique en intégrant les probabilités d'une séquence de mots dans un modèle linguistique, par son extension au bigrammes, trigrammes, etc..,

En termes de produits industriels, Le domaine du traitement de la parole a connu une montée fulgurante en termes de recherches et de produits industriels, et ceci essentiellement grâce à :

- ◆ En 1980 D. Klatt conçut l'un des meilleurs synthétiseurs de parole pour l'Anglais Américain à ce temps;
- ◆ Rabiner en 1989, qui a établi l'une des études les plus intéressantes à nos jours, concernant les HMM, ou il expose la méthodologie de travail d'apprentissage et de test des chaînes de Markov;
- ◆ En 1990 le système PSOLA a vu le jour au Centre National d'Etudes des Télécommunications - CNET par E. Moulines et F. Charpentier;
- ◆ De 1989 à 1998, le projet DARPA du département de défense des Etas-unis, concernait les technologies du langage humain a donné une lancée incroyable aux HMM qui par la suite sont devenus le standard des systèmes de reconnaissance.

Divers travaux concernant la Reconnaissance Automatique de la Parole, "RAP" ont été réalisés en utilisant les Réseaux de Neurones "RN", qui demandent comme les chaînes de Markov cachés des bases de données énormes pour l'apprentissage, ne gèrent malheureusement pas l'information temporelle comprise dans les phonèmes, allongement d'un mot ou d'un phonème serait reconnu comme différent pour les réseaux de neurones alors que pour les HMM, ce n'est qu'une transition vers le même état qui est réalisée. [48],[49].

Les travaux intégrant les RN et les HMM appelés approche hybride ont fait leur preuves dans différents travaux comme, [24], [50],[17],[51],[55], [56],[60].

3.3. Problématique générale

La reconnaissance de la parole invoque l'analyse, la synthèse, l'identification du langage, du locuteur, prenant en compte les différents aspects liés à l'environnement lors de la production de celle-ci, toutefois l'évolution des systèmes de RAP est « freinée » ou alourdie par les problèmes suivants :

- ✚ variabilité acoustique, due au fait que le même phonème prononcé dans différents contextes, en présence de phonèmes voisins, présente des réalisations acoustiques différentes, sans oublier la prosodie qui modifie le sens de la phrase, ainsi que les conditions environnantes qui faussent parfois les entrées au système;
- ✚ variabilité du locuteur, en effet, le locuteur change de voix d'une façon non ponctuelle lorsqu'il est malade, de mauvaise humeur où content, etc.
- ✚ variabilité linguistique, lorsque la même requête peut être prononcée de différentes manières, avec différents mots.
 - ◆ donnez moi de l'eau.
 - ◆ puis je avoir un peu d'eau?
 - ◆ je pourrais avoir un verre d'eau.
 - ◆ je puis avoir de l'eau (erreur de prononciation **peux** est remplacé par **puis**)
- ✚ variabilité phonétique, lorsque le même mot est prononcé de plusieurs manières différentes par différents locuteurs ayant divers accents sociolinguistiques.

Les problèmes additionnels, ou plutôt technologiques viennent des canaux de transmission, des nouveaux systèmes embarqués qui nécessitent des systèmes de compression, de l'effet Lombard [52] qui apparaît lorsque les personnes parlent avec une voix basse dans des milieux bruités et l'augment en conséquent.

- ✚ L'un des problèmes majeurs est l'absence de silence, contraire au texte qui est segmenté en mots avec des espaces, la parole connaît très peu ce phénomène, prenons par exemple l'Anglais des texans, qui peut être interprété comme un flux d'air incompréhensible, ceci est généralement dû au chevauchement des mots et non des phonèmes comme dans le premier point cité, qui peuvent induire le système de RAP soit à se planter ou à interpréter un mot à la place d'un autre.

3.4. Méthodologie de Travail

L'analyse des données issues du signal parole est très complexe, ceci est dû à la multitude ou la redondance de l'information vocale, c'est l'une des particularités de ce signal, en prenant en compte les informations citées auparavant, l'analyse sera orientée à des paramètres généraux discriminants, englobant différentes occurrences de l'information du point de vue temporel, spectral, spatial, perceptuel et/ou prosodique, ...

Dans notre travail, nous ne présentons pas toutes les bases mathématiques des méthodes utilisés, mais nous essaierons de justifier notre choix, en expliquant le trajet des données à travers les différents étages de traitement.

3.4.1 Acquisition des données

L'acquisition des données présente en elle-même une chaîne de traitement, ou l'erreur même infime se verra propagée sur les étages en amont, incombant ainsi des erreurs relativement considérables, celles-ci pourront fausser l'information et donner lieu à des estimations totalement erronées.

3.4.1.1 Enregistrement

La phase d'enregistrement concerne la conversion du signal continu de la parole en une suite ou succession d'impulsions normalisées, à une fréquence de 16 KHz, cette fréquence d'échantillonnage est suffisante pour prendre en compte la variabilité du signal vocal enregistré.

3.4.1.2 Segmentation et chevauchement

Après l'acquisition et le stockage du corpus d'étude, suivant un codage bien défini, chaque mot ou phrase du signal enregistré est segmenté en fenêtres de durée fixe, obéissant aux deux contraintes suivantes :

- ✚ stationnarité du signal parole (moyenne et variance constantes durant la trame ou la fenêtre temporelle d'analyse);
- ✚ durée supérieure à l'inverse de la fréquence fondamentale [53].

Le tableau 3.1. récapitule les différentes durées utilisées en littérature:

Tableau 3.1 : Durées primaires d'analyse ainsi que la durée de chevauchement

Méthodes	Durée	Chevauchement
Lawrence Rabiner en 1989 [49]	45 ms	30 ms
A.R El Obeid Ahmed et al en 1998 [54]	20 ms	5 ms
Yasuhi Tsubota et al en 2002 [57]	20 ms	10 ms
Alizera A. Dibazar et al en 2002 [58]	10 ms	2.5 ms
Georg Stemmer et al en 2001 [64]	20 ms	10 ms

La durée d'analyse sélectionnée dans notre travail est de 20 ms.

La conversion étant faite, il faudra éviter les effets de bord de la segmentation fenêtré, car les segments de 20ms sont contigus, donc lors du changement d'une fenêtre temporelle à une autre, l'apparition de transitions se fera remarquer, donc on doit recourir à un chevauchement ou à un recouvrement de 10 ms.

Le schéma suivant illustre les différentes zones :

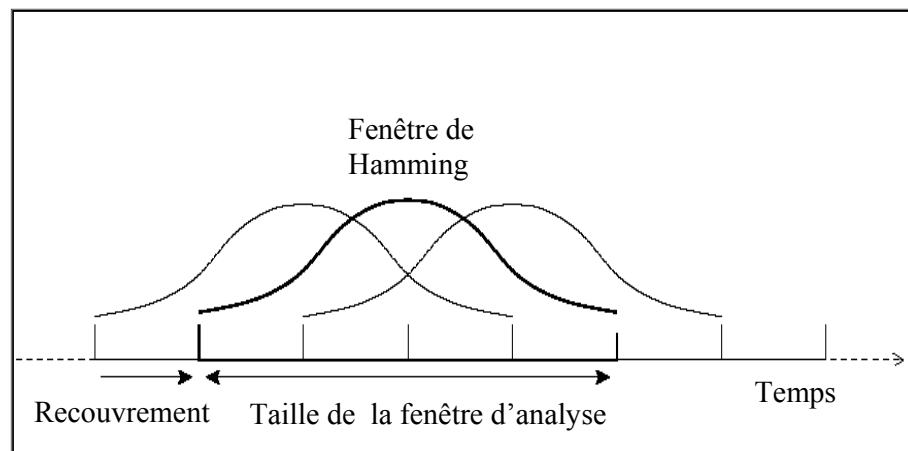


Figure 3.1 : Segmentation temporelle et Recouvrement

3.5. Filtrage de préaccentuation

Pourquoi filtrer un signal si on veut récupérer toutes les informations ?

Il existe deux explications pour l'utilisation du module de préaccentuation [65]. Pour la première, la partie voisée du signal de la parole présente une accentuation spectrale approximative de (-20) dB par décade. Le filtre de préaccentuation permet de compenser

cette accentuation avant d'analyser le spectre, ce qui améliore cette analyse. La deuxième considère que l'audition est plus sensible dans la région du spectre autour des 1 KHz. Le filtre de préaccentuation va donc amplifier cette région centrale du spectre.

En général le filtre de préaccentuation est de la forme :

$$y'_n = y_n - \alpha \cdot y_{n-1} \quad (1)$$

Où n représente le $n^{\text{ième}}$ échantillon calculé et $\alpha \in [0,9 \text{ et } 1]$.

Le découpage du signal en trames produit des discontinuités aux frontières des trames, qui se manifestent par des lobes secondaires dans le spectre; ces effets parasites sont réduits en appliquant aux échantillons de la trame une fenêtre de pondération comme par exemple la fenêtre de Hamming [61].

$$y''_n = y'_n * w_n \quad (2)$$

$$\text{Où } w_n = 0.54 - 0.46 \cdot \cos\left(2 \cdot \pi \cdot \frac{n}{N-1}\right) \quad \text{avec } 0 \leq n \leq N-1 \quad (3)$$

$$\text{Par la suite, on posera } x_n = y''_n \quad (4)$$

3.6. Méthodes d'analyse en Reconnaissance Automatique de la Parole

Cette partie est un récapitulatif des méthodes d'analyse, voir tableau 3.1, trouvées en littérature, ces différentes recherches se sont basées sur l'extraction des paramètres acoustiques, se différenciant par la signification physique, la complexité des calculs adjoints, le contexte d'étude, la portabilité du système, etc.

Tableau 3.2. : Techniques d'analyse

Domaine d'analyse	Techniques d'analyse	Abréviation
Temporel	Coefficients de codage prédictif linéaire	LPC
Cepstral	Coefficients Cepstraux Echelle Mel	MFCC
	Coefficients de Prédiction Linéaire Cepstrale	LPCC
Perceptif	Coefficients de Prédiction Perceptuelle	PLP
	Technique Spectrale Relative	RASTA-PLP
Spectral/Cepstral	Coefficients Cepstraux linéaires en fréquence	LFCC

Dans diverses études, le nombre des vecteurs acoustiques récupéré est excessivement immense [24], [54], il y'a lieu de réduire ce nombre en utilisant l'une des techniques suivantes, voir tableau 3.3.

Tableau 3.3. : Méthodes de réduction des données

Méthodes de réduction des données	Abréviation
Quantification Vectorielle	VQ
Analyse en Composantes Principales	PCA
Analyse Discriminante Linéaire	LDA

La dernière étape de traitement des vecteurs acoustiques est la classification ou le calcul de vraisemblance maximale, les méthodes les plus utilisées sont montrés dans le voir tableau 3.4.

Tableau 3.4. : Méthodes de classification / calcul de vraisemblance

Méthodes de Classification	Abréviation
Réseau de Neurones	ANN
Modèle de Markov caché	HMM
Systeme à base de règle (systemes experts)	
Comparaison par déformation temporelle	DTW
Transformée en ondelettes	DWT
Modèle à mélange de gaussiennes	GMM

Le schéma de la figure 3.2. illustre les différentes méthodologies d'analyse du signal vocal :

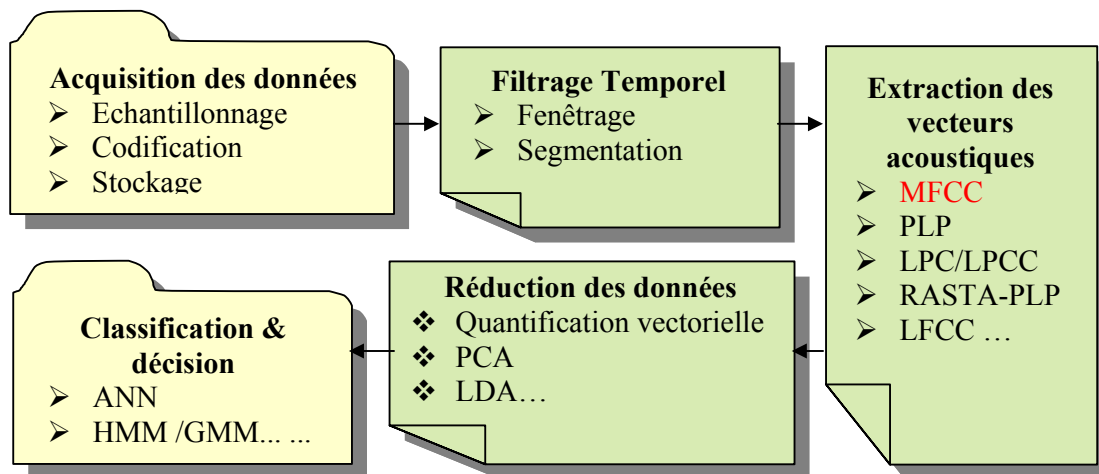


Figure 3.2 : Schéma global du processus de reconnaissance utilisé en RAP

3.7. Travaux similaires

L'étude du signal parole pathologique est un apport fondamental au développement de la recherche scientifique, les travaux dans ce sens sont orientés vers la détection des pathologies avant qu'elles ne nuisent à l'organisme. Ceci est fortement utilisé dans les hôpitaux à des fins préopératoires [62], [63], [35], ou à des fins d'apprentissage, contexte de notre travail [24], le travail de détection des erreurs de prononciation est aussi utilisé dans l'apprentissage des langues, exemple de l'Anglais pour les coréens ou les japonais. [51] [67].

3.8. Analyse et extraction des paramètres

Le corpus n'est pas utilisé directement, mais transformé en vecteurs acoustiques discriminants, stables au bruit, ayant une signification perceptuelle, dans ce sens la conversion vers un espace spectral est très intéressante.

3.8.1 Choix des Coefficients et corrélation

La seconde étape, décisive, est de faire sortir des coefficients représentant le signal, soit dans le domaine temporel, acoustique ou spectral. [17] a utilisé douze caractéristiques pour former les vecteurs acoustiques qui sont définis comme suit : l'énergie totale, le centre de gravité du spectre, la durée, les énergies de basses et hautes fréquences, les transitions formantiques, la détection de silence, la détection de voisement, variation de

l'énergie dans les sous bandes fréquentielles, variation du centre de gravité spectrale, fréquences maximales (peak), variation des fréquences maximales et enfin le taux de passage par zéro (TPZ), toutes ces informations forment les vecteurs acoustiques en entrée à un système à base de Réseau de Neurones ou à un système markovien de type HMM.

D'autres travaux ont utilisé les paramètres de prédiction linéaire LPC, qui ont été convertis en paramètres Cepstraux LPCC moins corrélés par l'utilisation des multi grammes et de la quantification vectorielle pour aboutir à un modèle de Markov caché [68].

Donc, le choix des paramètres dépend essentiellement du degré de reconnaissance que l'on veut appréhender telle que la reconnaissance des mots isolés, de la parole continue,...

Le signal, ainsi obtenu, de l'étape de stockage doit être valorisé par l'extraction de caractéristiques consiste à réduire l'information présente dans le signal de parole et à le transformer en une séquence de vecteurs acoustiques robustes aux variations acoustiques, tout en restant représentatif du contenu lexical, c'est-à-dire des phonèmes et des mots qui sont prononcés. [69]

Ce choix fait prévaloir la robustesse de l'analyse et facilite l'interprétation des résultats. Dans notre travail, nous utilisons les paramètres MFCC, paramètres Cepstraux connus pour leur robustesse au bruit et leur dé corrélation, ce choix sera expliqué dans la section suivante.

3.8.2. Coefficients Cepstraux Echelle Mel /MFCC

Les coefficients LPC ont un défaut assez remarquable c'est leur corrélation, donc pour différentes trames du signal, on obtient des coefficients très proches, ceci évidemment, n'aide pas le système de discrimination basée sur ces coefficients, alors il y'a eu recours à passer dans un domaine cepstral pour déterminer des coefficients assez discriminants et robustes au bruit engendré par le conduit vocal, ou tout autre source de bruit environnant.

Il y'a lieu de considérer le point suivant:

- ♦ la contribution à la perception des sons de la parole des hautes fréquences est plus faible que celle des basses fréquences, donc un changement d'échelle s'avère nécessaire.

3.8.3. Echelle Mel

A l'origine de l'échelle « Mel », nous pouvons citer le psychologue américain **Stanley Smith Stevens (1906-1973)**, fondateur du Laboratoire de Psychoacoustique de Harvard, ainsi que Volkman et Newman en 1937 [8].

L'échelle Mel est définie par :

$$B(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \quad (5)$$

Où f est la fréquence en Hz et $B(f)$ la fréquence suivant l'échelle de Mel.

Le calcul des coefficients cepstraux est effectué selon le schéma de la figure 3.3:

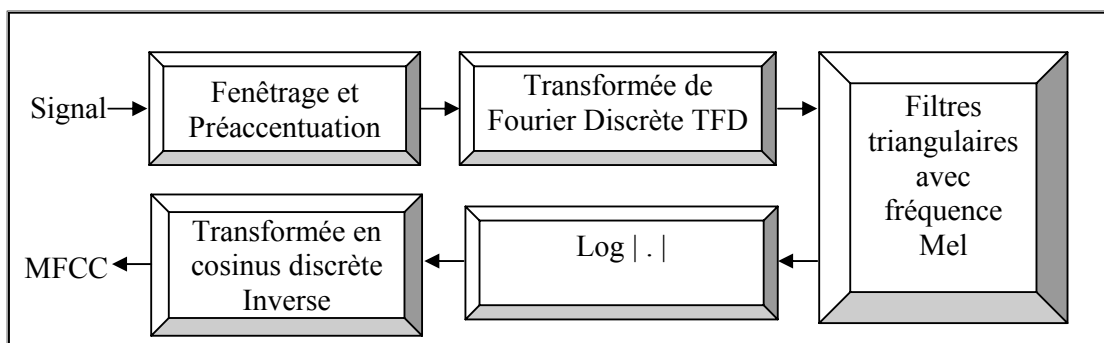


Figure 3.3 : Chaîne d'analyse du signal produisant les coefficients MFCC [61]

Soit un signal $x[n]$ discret avec $0 \leq n < N$, où N représente le nombre d'échantillons d'une fenêtre d'analyse, la transformée de Fourier est alors définie par :

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-\frac{j \cdot 2 \cdot \pi \cdot n \cdot k}{N}} \quad \text{avec } 0 \leq k < N \quad (6)$$

Le spectre du signal est filtré par des filtres triangulaires, figure 3.5, dont les bandes passantes sont équivalentes en domaine fréquence Mel. Les points de frontière $B(m)$ des filtres en échelle de fréquence sont calculés à partir de la formule :

$$B(m) = B(f_1) + m \cdot \frac{B(f_h) + B(f_b)}{M + 1} \quad \text{avec } 0 \leq m < M + 1 \quad (7)$$

- M désigne le nombre de filtres
- f_h désigne la fréquence la plus haute du signal
- f_b désigne la fréquence la plus basse du signal

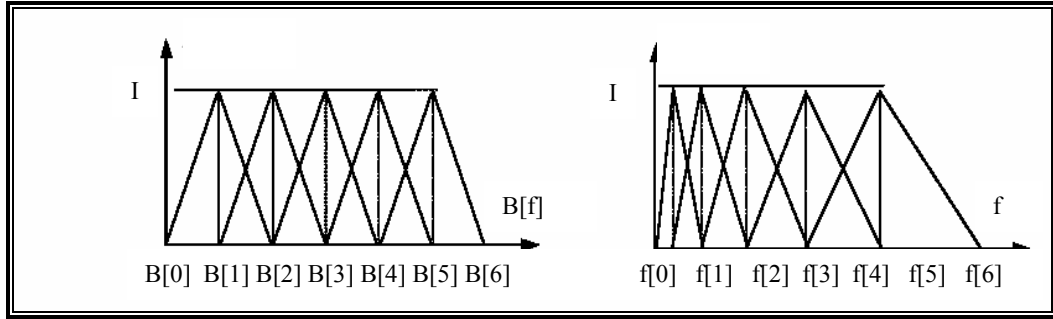


Figure 3.4 : Filtre triangulaire passe bande en Mel fréquence B(f)

Figure 3.5 : Filtre triangulaire passe bande en fréquence f

Dans le domaine fréquentiel, les points $f(m)$ discrets correspondants sont calculés d'après:

$$f(m) = \left(\frac{N}{F_s} \right) \cdot B^{-1} \left[B(f_b) + m \cdot \frac{B(f_h) + B(f_b)}{M+1} \right] \quad \text{avec } 0 \leq m < M+1 \quad (8)$$

$$\text{et } B^{-1}(i) = 700 \cdot \left(10^{\frac{i}{2595}} - 1 \right) \quad (9)$$

Les coefficients des filtres sont calculés par :

$$H_m[k] = \begin{cases} 0 & \text{si } k \leq f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & \text{si } f(m) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & \text{si } f(m) \leq k \leq f(m+1) \\ 0 & \text{si } k \geq f(m+1) \end{cases} \quad (10)$$

Ensuite on multiplie les énergies de $X(k)$ par les coefficients $H_m[k]$ et on calcule leur logarithme :

$$E[m] = \log \left[\sum_{k=0}^{N-1} |X(k)|^2 \cdot H_m[k] \right] \quad \text{avec } 0 \leq m < M \quad (11)$$

Les coefficients MFCC de fréquence en échelle MEL sont obtenus par la transformée inverse des coefficients en sortie des filtres.

Remarque : Le nombre de MFCC est moins grand que le nombre de filtres, donc une transformée en cosinus est plutôt utilisée.

$$c[n] = \sum_{m=0}^{M-1} E[m] \cdot \cos \left(\frac{\pi \cdot n \cdot (m + \frac{1}{2})}{M} \right) \quad \text{Avec } 0 \leq m < M \quad (12)$$

Ces coefficients ont été utilisés pour la reconnaissance automatique de digits en Anglais en conditions bruitées [20], pour la détection et reconnaissance des sons pour la surveillance médicale [70], pour la comparaison de paramètres de reconnaissance des phonèmes de la langue Arabe [54], pour la détection d'un signal de parole pathologique [58], pour la comparaison de plusieurs modélisations acoustiques pour des systèmes de reconnaissance embarqués [71], pour la reconnaissance/ vérification du locuteur [72], pour la conception d'un système hybride pour l'identification de traits phonétiques complexes de la langue Arabe [56], lors de la reconnaissance et la vérification de l'Anglais par des étudiants japonais dans un système d'apprentissage automatique. [57]

On ne peut citer toutes les études établies, utilisant les MFCC, toutefois le nombre de coefficients est en général pris égal à 13, puis parfois réduit à 12, en considérant deux points essentiels :

- ◆ le premier coefficient C_0 représente l'énergie de la trame et ne peut réellement contribuer à la segmentation ou la reconnaissance;
- ◆ les coefficients de 1 à 12 représentent l'enveloppe cepstrale plus ou moins lissée, les hautes variations fréquentielles étant éliminés.

A cette étape, le nombre de vecteurs acoustiques est énorme, alors nous modélisons leur distribution par un mélange de gaussiennes, qui ont démontré leurs preuves dans le domaine de la modélisation statistique d'un nombre très important de vecteurs observés.

En vue de maximiser l'information utile, la variation première et seconde des coefficients MFCC est fortement recommandée [54], Toutefois la problématique de dimensionnement apparaît; Pour cette raison, l'adjonction de la variation des coefficients Cepstraux Δ MFCC et de leur accélération $\Delta\Delta$ MFCC est un point à bien évaluer, d'un point de vue complexité de calcul et dimensionnement des matrices de co-variances.

A titre d'exemple, si chaque fenêtre temporelle d'une durée de 20 ms avec un recouvrement de 10 ms avec une fréquence d'échantillonnage de 16 KHz soit (16 échantillons par ms), le nombre de paramètres à traiter par trame est mentionné dans le tableau 3.4.

Tableau 3.5. : Calcul des nombre de vecteurs acoustiques

Coefficients	Nombre	Nombre de vecteurs par trame de 10 ms
MFCC	12	$12 \times 16 = 192$
MFCC+ Δ MFCC	24	$12 \times 24 = 288$
MFCC+ Δ MFCC + $\Delta\Delta$ MFCC	36	$12 \times 36 = 432$

On reporte que dans d'autres littératures, le nombre de coefficients est beaucoup plus important [70].

3.8.4. Chaînes de Markov Cachés HMM

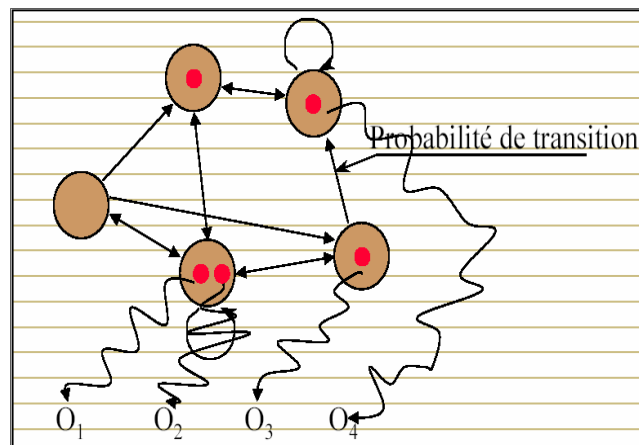


Figure 3.6. : Chaînes de Markov comme Automate probabiliste

Les chaînes de Markov trouvent leurs applications dans différents domaines de la physique, par exemple en expliquant les mouvements browniens, les radiations cosmiques, la radioactivité ainsi que la génétique, la parole, les fluctuations des stocks d'entreprise, les marches aléatoires, ...

Les chaînes de Markov² sont considérées comme des exemples de processus stochastiques, incluant les processus de Markov, de Poisson, etc. [74].

² Andreï Andreïevitch Markov (Андрей Андреевич Марков), mathématicien russe (1856-1922) né à Ryzan, en 1874 étudiant à l'université de St Petersburg sous la tutelle de Tchebychev, en 1886, il fût membre à l'académie des sciences de St Petersburg puis professeur à l'université, en retraite en 1905, mais continua à enseigner les probabilités jusqu'à sa mort. Markov était considéré comme le meilleur joueur d'échecs de St. Petersburg (Sheynin 1988, p. 341).

3.8.4.1. Définition 1

L'étude des processus stochastiques a commencé au début du XX^{ème} siècle grâce à un mathématicien Russe, Markov Andreï Andreïevitch. Son étude statistique du langage l'a conduit à formuler l'hypothèse Markovienne, qui peut se résumer ainsi:

« L'évolution future d'un système ne dépend que de son état présent ».

Autrement dit, cette hypothèse implique que l'état courant du système contient toute l'information apportée par le passé. C'est donc une hypothèse très forte, mais qui semble relativement logique. En pratique, on constate que de nombreux systèmes enfreignent cette condition. Cependant, en affinant le modèle, on peut souvent le rendre markovien.

3.8.4.2. Définition 2

Les modèles de Markov et, plus particulièrement, les modèles de Markov à états cachés, plus connus sous le nom de HMM, Hidden Markov Models, permettent de synthétiser la connaissance contenue dans un corpus par apprentissage. Cette connaissance sera synthétisée, dans les modèles de Markov, par représentation probabiliste au sein de plusieurs graphes, chaque graphe correspondant à une classe du corpus d'apprentissage qui, en reconnaissance automatique de la parole, peut correspondre à un phonème ou à un mot [49].

3.8.5. Comparaison de la DTW avec les HMM

D'un point de vue général, la différence de fonctionnement entre la méthode de l'alignement temporel et les modèles de Markov n'est pas fondamentale. Dans le cas des modèles de Markov, la forme à analyser est comparée à chacune des classes constituées en graphe. Dans le cas de la DTW, par contre, la forme à analyser est comparée à chacune des formes de référence dont le rattachement à une classe, et donc à une signification symbolique, ne sera utile qu'à un stade postérieur du traitement. Le modèle de Markov étant une représentation probabiliste des formes de référence, il ne s'agit plus ici de trouver un chemin de taille minimale mais de trouver une probabilité de cheminement dont la valeur est maximale [75].

3.8.6. Approche mathématique

Un modèle de Markov caché est défini par les données suivantes :

3.8.6.1. Graphe d'état

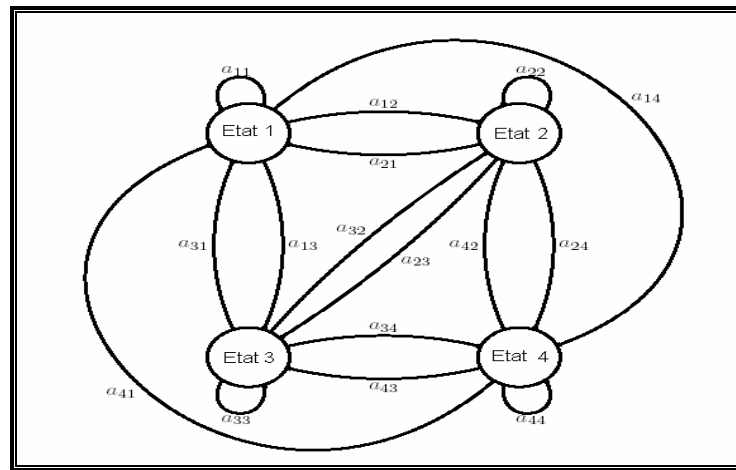


Figure 3.7. Graphe d'états d'une chaîne de Markov

3.8.6.2. Matrice de Transition

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}_{4 \times 4}$$

Les éléments sont définis tels que :

$$a_{ij} = p(q_{t+1} = j / q_t = i) \quad 1 \leq i, j < N \quad (13)$$

3.8.6.3. Matrice d'observation ou d'émission

Cette matrice est un peu spéciale, car sa dimension dépend des observations soient elles continues ou discrètes.

Ses éléments sont définis telle qu'à chaque état, une ou plusieurs observations sont associées à cet état, définis par :

$b(o_j / q_i)$: Représente la probabilité d'observer ou d'émettre l'observation j en étant à l'état q_i .

❖ **Cas discret** : Les matrices contiennent des vecteurs d'observations de dimension finie de $N \times M$.

- ❖ **Cas continu** : Les matrices contiennent les paramètres des probabilités de distribution des observations par état, celles-ci peuvent être des gaussiennes, des multi gaussiennes ou toute autre distribution propre au contexte des données.

$$B = \left\{ b_j(o_t) \right\}_{j=1}^N, \text{ tel que : } b_j(o_t) = p(o_t / q_t = j) \quad (14)$$

3.8.6.4. Vecteur d'initialisation

Ce vecteur donne au modèle les probabilités de transition initiale

$$\blacktriangleright \pi : \text{vecteur de dimension } N \times 1, \text{ tel que : } \pi_i = p(q_1 = i) \quad (15)$$

En résumé, le modèle des chaînes de Markov est noté comme suit :

$$\lambda = (A, B, \pi) \quad \text{Avec } A : N \times N; \quad B : N \times M \quad \text{et } \pi : N \times 1$$

3.8.7. Les trois problèmes fondamentaux des HMM

Différents formulations existent en littérature, mais nous avons choisi celle de P. Langlais [73].

3.8.7.1. Evaluation

Sachant ou ayant des vecteurs d'observation $O = \{O_1, O_2, \dots, O_T\}$ et $\lambda = \{A, B, \pi\}$ comment évaluer $P(O|\lambda)$?

Comment trouver le modèle qui a pu générer la séquence observée ?

3.8.7.2. Retirer le H de Hidden Markov Models

Ayant les vecteurs d'observation $O = \{O_1, O_2, \dots, O_T\}$ ainsi que les paramètres du $\lambda = \{A, B, \pi\}$ du modèle, comment trouver la séquence (cachée) optimale d'états qui explique au mieux ces observations ?

3.8.7.3. Apprentissage

Sachant un corpus d'entraînement O , comment ajuster les paramètres λ du modèle pour maximiser $p(O|\lambda)$?

3.8.7.4. Pourquoi problèmes ?

Prenons un exemple de m modèles, $(\lambda_i, \forall i \in [1, m])$ qui modélisent chacun une entité donnée (un mot ou un phonème, par exemple..), Soit O une observation dont on veut connaître l'identité $i = \arg\text{Max}_i(p(O/\lambda_i))$

Soit $Q=q_1, q_2, q_3 \dots q_T$ une séquence d'états du modèle de Markov caché pouvant « expliquer » O :

$$P(O/\lambda) = \sum_{\text{Tous les } Q} p(O, Q/\lambda) = \sum_{\text{Tous les } Q} p(O/Q, \lambda, \pi) \cdot P(Q/\lambda) \quad (16)$$

Avec :

$$p(O/Q, \lambda) = \prod_{t=1}^T p(o_t/q_t, \lambda) = b_{q_1}(o_1) \times b_{q_2}(o_2) \dots b_{q_T}(o_T) \quad (17)$$

$$\text{et } p(Q/\lambda) = \pi_{q_1} \cdot b_{q_1}(o_1) \times b_{q_2}(o_2) \dots b_{q_T}(o_T) \quad (18)$$

Le problème relié à la complexité de calcul qui est de l'ordre de $(2^T - 1) \cdot N^T$ multiplications et $(N^T - 1)$ additions,

Donc, si on prend $N=5$ états, $T=100$ observations, on a 1.5698×10^{72} Multiplications et 7.8886×10^{69} additions.

Pour complètement définir les HMM, il y'a lieu de résoudre les 3 problèmes cités :

Prob. 1 : Etant donné une séquence d'observations, par exemple, une séquence de vecteurs acoustiques, et un modèle défini par le triplet (A, B, π) : Quelle est la plus grande probabilité que le modèle donné génère la séquence observée. Ceci est résolu par l'algorithme Forward.

Prob. 2 : Etant donné une séquence d'observation et un modèle HMM, quelle est la séquence d'états qui a générée cette séquence observée. Ceci est résolu par l'algorithme de Viterbi.

Prob. 3 : Etant donné une séquence d'observations, quels sont les ajustements à faire sur les paramètres du modèle pour avoir la plus grande probabilité de générer cette séquence. C'est la phase d'apprentissage, ceci est résolu par l'algorithme backward-forward.

Il faut également définir une topologie qui va déterminer les transitions possibles entre les différents états. D'après les études déjà effectuées, un modèle de 5 états entièrement connecté est choisi pour la reconnaissance de mots, toutefois pour des phonèmes des modèles à 3 états seront suffisants, une topologie de type gauche droite, puisque que

chaque état représente un état au sein d'un phonème. Chaque mot impose ainsi une suite de phonèmes déterminée. [49]

La langue Arabe compte 34 phonèmes, chacun d'entre eux, peut avoir son propre modèle de Markov.

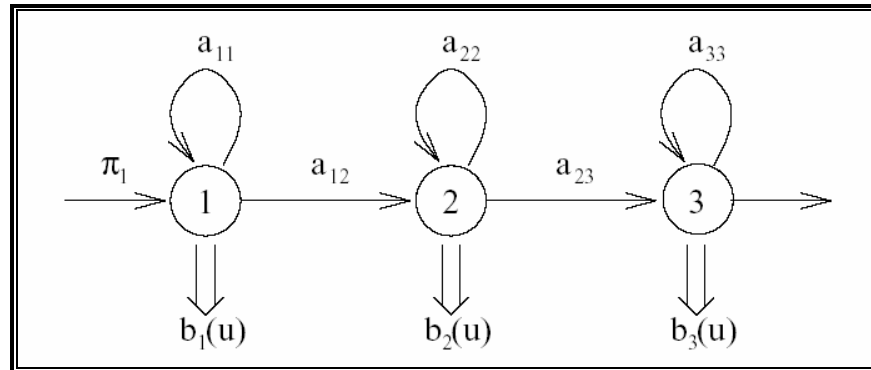


Figure 3.8 Modèle de transitions à trois états Markoviens gauche-droite.

Remarque : Les modèles de Markov permettent de suivre une évolution temporelle globale en même temps qu'ils fournissent une évaluation locale. La figure 3.8 illustre ce principe.

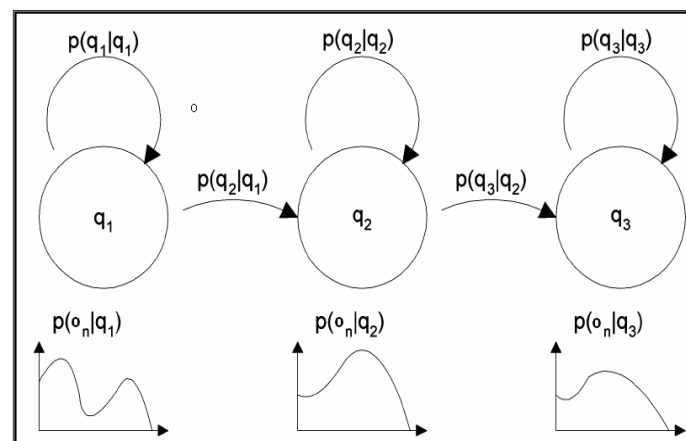


Figure 3.9. Modèle de transitions entre états markoviens gauche-droite et leur probabilité de transition et d'émission respectives.

Notre approche s'adapte bien aux modèles de Markov. Notamment, les modèles qui sont conçus pour reconnaître n'importe quelle phrase ou séquence de mots lors du traitement de la parole continue. Malheureusement, il n'est pas possible de construire un modèle pour chaque mot du dictionnaire, qui parfois atteignent des centaines de milliers, toutefois, dans notre contexte de pathologie langagière, nous avons ciblé un apprentissage

supervisé qui compte des mots sélectionnés définissant la mal prononciation, nous modélisons chaque mot par une suite d'états.

Remarque : Les probabilités des matrices de transition et d'observation sont indépendantes du temps.

3.8.8. Comment résoudre les trois problèmes ?

3.8.8.1 Solution au prob. 1 :

a. Evaluation par l'algorithme « Forward »

Hypothèse, nous disposons de k modèles de Markov $\{\lambda_1, \lambda_2 \dots \lambda_k\}$, représentant les mots ou les phonèmes à analyser ayant $\{q_1, q_2 \dots q_L\}$ états possibles, et d'un ensemble d'observations $O = \{o_1, o_2, \dots o_N\}$ [69].

Trouvons le modèle qui a généré cette séquence, tel que : $P(O/\lambda)$ est maximale.

« Une façon de calculer cette probabilité est d'énumérer tous les chemins du modèle :

$$P(O/\lambda) = \sum_{Q \in \text{Modèle}} p(Q, O/\lambda) \quad (19)$$

Où la somme porte sur tous les chemins Q de longueur L , dans le modèle, la complexité est toutefois considérable, alors une récurrence avant est utilisée de la manière suivante :

$$P(O/\lambda) = \sum_{l=1}^L p(q_l^n, O/\lambda), \forall n \in [1, N] \quad (20)$$

Chaque terme de cette somme exprime la probabilité que X soit émis par le modèle λ en passant par l'état q_l à l'instant n , qui peut se factoriser comme suit :

$$p(q_l^n, O/\lambda) = p(q_l^n, O_1^n/\lambda) \cdot p(O_{n+1}^N / q_l^n, O_1^n, \lambda) \quad (21)$$

Où O_1^n représente une séquence partielle de vecteurs d'observation $\{o_1, o_2, \dots o_n\}$

Définissons une nouvelle variable, représentant la probabilité que le modèle λ ait généré la séquence partielle O_1^n , en se trouvant dans l'état q_l à l'instant n .

$$\alpha_n(l/\lambda) = p(q_l^n, O_1^n/\lambda) \quad (22)$$

$$\text{et } \alpha_n(l/\lambda) = p(q_l^n, O_1^n/\lambda) = \sum_{k=1}^L p(q_l^{n-1}, q_k^n, O_1^{n-1}, o_n/\lambda) \quad (23)$$

$$\alpha_n(l/\lambda) = \sum_{k=1}^L p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) \cdot p(q_k^{n-1}, O_1^{n-1} / \lambda) \quad (24)$$

$$\alpha_n(l/\lambda) = \sum_{k=1}^L p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) \cdot \alpha_{n-1}(k/\lambda) \quad (25)$$

$$L' \text{initialisation de cette forme r\u00e9currente se fait par : } \alpha_n(l/\lambda) = \pi(l) \quad (26)$$

Toutefois le terme $p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda)$ n'est pas facilement calculable, il faudrait le simplifier :

$$p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) = p(q_l^n / q_k^{n-1}, O_1^{n-1}, \lambda) \cdot p(o_n / q_l^n, q_k^{n-1}, O_1^{n-1}, \lambda) \quad (27)$$

Les mod\u00e8les de Markov sont d'ordre 1, donc les \u00e9tats ne d\u00e9pendent que de l'\u00e9tat pr\u00e9c\u00e9dent la transition et sont conditionnellement ind\u00e9pendants du pass\u00e9, alors

$$p(q_l^n / q_k^{n-1}, O_1^{n-1}, \lambda) \rightarrow p(q_l^n / q_k^{n-1}, \lambda) \quad (28)$$

Les observations sont conditionnellement ind\u00e9pendantes du pass\u00e9, elles ne d\u00e9pendent ni des observations du pass\u00e9 ni des \u00e9tats HMM pr\u00e9c\u00e9dents, alors

$$p(o_n / q_l^n, q_k^{n-1}, O_1^{n-1}, \lambda) \rightarrow p(o_n / q_l^n, \lambda) \quad (29)$$

$$\text{Alors } p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) = p(q_l^n / q_k^{n-1}, \lambda) \cdot p(o_n / q_l^n, \lambda) \quad (30)$$

Cette \u00e9quation d\u00e9finit, les probabilit\u00e9s d'\u00e9mission $p(o_n / q_l^n, \lambda)$, celles-ci seront consid\u00e9r\u00e9es dans notre cas multi gaussienne.

Et $p(q_l^n / q_k^{n-1}, \lambda)$ les probabilit\u00e9s de transition \u00e0 l'int\u00e9rieur du mod\u00e8le.

L'\u00e9quation $\alpha_n(l/\lambda) = \sum_{k=1}^L p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) \cdot \alpha_{n-1}(k/\lambda)$ est r\u00e9crite de la mani\u00e8re suivante :

$$\alpha_n(l/\lambda) = p(o_n, q_l) \sum_{k=1}^L \alpha_{n-1}(k/\lambda) \cdot p(q_l^n / q_k^{n-1}, \lambda) \quad (31)$$

En r\u00e9sum\u00e9 cette r\u00e9currence nous permet de calculer la probabilit\u00e9 de toute une s\u00e9quence \u00e0 partir des probabilit\u00e9s d'\u00e9mission et de transition locales.

b. Evaluation par l'algorithme « Backward »

Une nouvelle variable est d\u00e9finie telle que, c'est la probabilit\u00e9 que le mod\u00e8le λ g\u00e9n\u00e8re le restant de la s\u00e9quence $O = \{o_{n+1}, o_2, \dots, o_N\}$ au d\u00e9part de l'\u00e9tat q_l .

Par la m\u00eame m\u00e9thode de calcul de l'algorithme forward,

$$\beta_n(l/\lambda) = p(O_{n+1}^N / q_l^n, O_1^n, \lambda) = \sum_k \beta_{n+1}(q_l^{n+1} / q_k^n, \lambda) \cdot p(o_{n+1} / q_k) \quad (32)$$

Ou la somme sur k porte sur les successeurs possibles de q_l , l'initialisation de cette nouvelle r\u00e9currence est donn\u00e9e par :

$$\beta_n(l/\lambda) = \pi_{l_F}(\lambda) \quad (33)$$

Qui représente la probabilité de rejoindre l'état final de λ à partir de q_l

Etant donné (19) et la définition de α , nous pouvons écrire :

$$P(O/\lambda) = \sum_{l=1}^T p(q_l^N, O_1^N / \lambda) = \sum_{l=1}^L \alpha_N(l/\lambda) \quad (34)$$

La somme se limite aux états finaux possibles dans le modèle λ , et pouvant correspondre à la fin du mot ou du phonème, pour l'estimation de $P(O/\lambda)$, et donc aussi de la reconnaissance, nous avons donc :

$$P(O/\lambda) = \sum_{l=1}^L \sum_{n=1}^T \alpha_n(l/\lambda) \beta_n(l/\lambda) = \sum_{\{Finaux\}} \alpha_N(I/\lambda) = \sum_{\{Initiaux\}} \beta_N(F/\lambda) \quad (35)$$

3.8.8.2 Solution au probl. 2 : Meilleur chemin par l'algorithme de Viterbi

Le modèle étant connu, quelle est le meilleur chemin, qui a donné la meilleure vraisemblance, donc quels sont les états qui ont réellement participé au calcul de la probabilité $P(O/\lambda)$?

Nous disposons du modèle défini par $\lambda = \{A, B, \pi\}$ et des observations $O = \{o_1, o_2, \dots, o_N\}$, trouvons le meilleur chemin maximisant $P(O/\lambda)$.

D'après Bellman, « Toute politique optimale est issue de sous politiques optimales »

L'approche Viterbi est plus simple à calculer, car au lieu de faire la somme sur tous les chemins, le chemin ayant une probabilité maximale est pris, donc l'équation $p(q_l^n, O/\lambda) = p(q_l^n, O_1^n / \lambda) \cdot p(O_{n+1}^N / q_l^n, O_1^n, \lambda)$, peut être réécrite comme suit :

$$\bar{p}(q_l^n, O_1^n / \lambda) = \max_k \bar{p}(q_k^{n-1}, O_1^{n-1} / \lambda) \cdot p(q_l^n, o_n / q_k^{n-1}, O_1^{n-1}, \lambda) \quad (36)$$

Où : $\bar{p}(q_k^{n-1}, O_1^{n-1})$ représente la probabilité du meilleur chemin possible allant de l'état initial $q_{Initial}$ à l'état q_l , en ayant émis les n premiers vecteurs O_1^n de la séquence O .

Utilisant l'indépendance et la supposition de Markov d'ordre 1, (36) devient

$$\bar{p}(q_l^n, O_1^n / \lambda) = \max_k \bar{p}(q_k^{n-1}, O_1^{n-1} / \lambda) \cdot p(q_l^n / q_k^{n-1}, \lambda) \cdot p(o_n / q_l) \quad (37)$$

En passant au logarithme, afin d'éviter l'underflow,

$$-\log[\bar{p}(q_l^n, O_1^n / \lambda)] = \min_k \{-\log[\bar{p}(q_k^{n-1}, O_1^{n-1} / \lambda)] - \log[p(q_l^n / q_k^{n-1}, \lambda)] - \log p(o_n / q_l)\}$$

$$\text{La probabilité } \bar{p}(O/\lambda) = \bar{p}(q_F^N, O_1^N / \lambda) \quad (38)$$

Où q_F^N est l'état final du modèle λ . [69]

C'est ce point qui fait que les HMM ressemblent à la DTW, il y'a dualité entre les logarithmes et les distances euclidiennes.

3.8.8.3 Solution au prob. 3 : Ajuster le modèle

Ayant un corpus d'entraînement O , comment ajuster les paramètres λ du modèle pour maximiser $P(O/\lambda)$? [73].

Ce stade, en général est le premier à réaliser car nous ne disposons que des vecteurs d'observation, et nous voudrions chercher un modèle qui pourrait « s'ajuster » à ces données, pour les modéliser avec la meilleure vraisemblance pour une future reconnaissance.

Le choix du modèle HMM, du nombre de gaussiennes est déterminant, voir tableau 3.5., il dépend de plusieurs facteurs, entre autres du :

- modèle du HMM : droite - gauche pour la parole [49].
- nombre de gaussiennes modélisant les observations par état;

Tableau 3.6. : Choix des paramètres des HMM/GMM

Etudes réalisées	Nombre d'états par HMM	Nombre de Gaussiennes
Lawrence Rabiner, 1989 [49]	Pour chaque mot : Mono locuteur : 5 à 8 états Multi locuteurs 8 à 10	3 à 5 par état 9 par état
Projet Raphaël, 1998 [22] Traduction multi langue	3 états par phonème	16 gaussiennes par état
M. A. Mokhtar et al, 1996 [77]	Language arabe complet : 41 états	
A. R. Elobeid Ahmed et al, 1998 [54]	3 états par Phonème	Pas de Gaussienne mais un Dictionnaire de 128 valeurs
A.A. Dibazar et al, 2002 [58]	3 états pour le phonème [a]	3 gaussiennes par état

3.9. Modélisation des données par mélange de Gaussiennes GMM

Avant d'aborder l'apprentissage, quelques définitions s'avèrent utiles, dans le cas de notre étude, l'utilisation d'un nombre très important de données nous incite à modéliser

leurs distributions, l'une des techniques de modélisation est l'approche multi gaussiennes, ou l'ensemble des données est écrit en fonction de diverses gaussiennes de la manière suivante :

$$p(O / \Theta) = \sum_{i=1}^T c_i \cdot p_i(O / \theta_i) \quad (39)$$

$$\text{Où } \Theta = \{c_1, c_2, \dots, c_T, \theta_1, \theta_2, \dots, \theta_T\} \quad (40)$$

$$\text{Avec } \theta_i = \{\mu_i, \sigma_i\} \quad (41)$$

Représentant chaque gaussienne

On ne peut citer tous les travaux relatifs au choix des paramètres, dans le tableau 3.7., toutefois, nous présentons une comparaison entre différentes études concernant le nombre d'états par modèle et le nombre de gaussienne par état.

Tableau 3.7 : Choix du nombre de gaussiennes

Nombre d'états par mot	Nombre de gaussiennes par état
A chaque phonème du mot un état	4, 8, 16
A chaque phonème du mot 3 états	4, 8, 16
Prendre le mot le plus long comme référence	4, 8, 16

3.10. Apprentissage par l'Algorithme de Baum Welch «Variante de l'algorithme EM »

L'idée générale derrière cet algorithme est d'estimer les paramètres du modèle HMM, ainsi que les paramètres des mélanges de gaussiennes, ayant en main deux informations essentielles:

- ◆ les vecteurs d'observation en nombre suffisant ;
- ◆ le nombre d'états gouvernant les transitions à trouver;
- ◆ la topologie des transitions à utiliser. (gauche - droite; Ergodique)

L'algorithme EM (Expectation – Maximisation) Maximiser l'espérance,

- ❖ calcule l'espérance d'une variable aléatoire manquante par rapport à une variable aléatoire présente;
- ❖ maximise l'espérance trouvée en fonction des variables présentes, par une méthode récursive, jusqu'à ajuster les paramètres des chaînes de Markov notamment les probabilités de transition ainsi que les paramètres des mélanges de gaussiennes.

L'algorithme EM n'est pas démontré mathématiquement dans notre étude, toutefois pour tout détail complémentaire, se référer à [66].

Soit $\{O\}$ l'ensemble des variables aléatoires connues et $\{Y\}$ les variables aléatoires inconnues, nous supposons qu'il existe une densité de probabilité jointe $z = (O, Y)$ telle que :

$$p(z / \lambda) = p(Y / O, \theta).p(X / \lambda) \quad (42)$$

Définissons une nouvelle quantité Q , représentant l'espérance jointe z ,

a. Etape E de l'algorithme

$$Q(\lambda, \lambda^{t-1}) = E[\log p(O, Y / \lambda) / Y, \lambda^{t-1}] \quad (43)$$

Où λ^{t-1} représente le modèle utilisé à l'itération $t-1$ pour calculer λ à l'itération t .

b. Etape M de l'algorithme

Cette valeur est alors maximisée selon λ ,

$$\text{Donc l'algorithme calcule le modèle } \lambda^t = \arg \max_{\lambda} Q(\lambda, \lambda^{t-1}) \quad (44)$$

Donc à chaque itération, on cherchera si le nouveau modèle apporte une amélioration à l'ajustement des données, c'est-à-dire est ce que le modèle représente les données à l'étape t mieux qu'à l'étape $t-1$.

Dans notre cas, nous allons définir deux nouvelles valeurs qui serviront lors de l'ajustement du modèle, telle que :

$$\gamma_i(n) = p(q_i^n / O, \lambda) \quad (45)$$

Qui représente la probabilité d'être à l'état q_i à l'instant n , générant la séquence O .

$$\gamma_i(n) = p(q_i^n / O, \lambda) = \frac{p(q_i^n, O / \lambda)}{p(O / \lambda)} = \frac{p(q_i^n, O / \lambda)}{\sum_{j=1}^L p(O, q_j^n / \lambda)} \quad (46)$$

Remarquons que :

$$\alpha_i(n)\beta_i(n) = p(o_1, o_2, \dots, o_t, q_i^n / \lambda).p(o_{t+1}, o_{t+2}, o \dots, o_T / q_i^n, \lambda) = p(O, q_i^n / \lambda) \quad (47)$$

Alors (46) devient :

$$\gamma_i(n) = p(q_i^n / O, \lambda) = \frac{\alpha_i(n).\beta_i(n)}{\sum_{j=1}^L \alpha_j(n).\beta_j(n)} \quad (48)$$

On définit une seconde valeur telle que :

$$\xi_{i_j}(n) = p(q_i^n, q_j^{n+1} / O, \lambda) \quad (49)$$

Qui représente la probabilité d'être à l'état i à l'instant n et de passer à l'état j à l'instant $n+1$, ceci peut être reformulée comme suit :

$$\xi_{i_j}(n) = p(q_i^n, q_j^{n+1} / O, \lambda) \quad (50)$$

$$\xi_{i,j}(n) = p(q_i^n, q_j^{n+1} / O, \lambda) = \frac{p(q_i^n, q_j^{n+1}, O / \lambda)}{p(O / \lambda)} = \frac{\alpha_i(n) \cdot a_{ij} \cdot b_j(o_{n+1}) \cdot \beta_i(n)}{\sum_{j=1}^L p(O, q_j^n / \lambda)} \quad (51)$$

L'on peut remarquer que le terme $\sum_{n=1}^T \gamma_i(n)$ représente la valeur espérée d'être à l'état q_i pendant tout les instants n pour toutes les observations O donnant ainsi le nombre de transitions partant de l'état q_i .

Et que le terme $\sum_{n=1}^{T-1} \xi_{i,j}(n)$ représente le nombre de transitions de l'état q_i à l'état q_j pour toutes les observations O .

L'utilisation de l'algorithme EM pour estimer les nouveaux paramètres à chaque itération nécessite de mettre à jour les valeurs manquantes itérativement de la manière suivante :

La quantité

$$\tilde{\pi}_i = \gamma_i(1) \quad (52)$$

Qui est la fréquence relative de passage à l'état q_i à l'instant 1.

Ainsi que :

$$\tilde{a}_{i,j} = \frac{\sum_{n=1}^{N-1} \xi_{i,j}(n)}{\sum_{n=1}^N \gamma_i(n)} \quad (53)$$

Qui représente le nombre de transitions de l'état q_i à l'état q_j relatif au nombre de transitions sortant de l'état q_i .

Pour le mélange de gaussiennes, les paramètres à estimer sont les moyennes et variances des gaussiennes ainsi que le taux de participation de la gaussienne à l'état q_i noté :

$$\tilde{c}_{i,l} = \frac{\sum_{n=1}^N \gamma_{i,l}(n)}{\sum_{n=1}^N \gamma_i(n)} \quad (54)$$

Où l représente la $l^{\text{ième}}$ gaussienne modélisant les vecteurs d'observation à un état q_i .

$$\mu_{i,l} = \frac{\sum_{n=1}^N \gamma_{i,l}(n) \cdot o_t}{\sum_{n=1}^N \gamma_{i,l}(n)} \quad (55)$$

L'équation (55) représente la moyenne de chaque gaussienne à l'état q_i

$$\text{Et le terme } \sum_{il} = \frac{\sum_{n=1}^N \gamma_{i_l}(n) \cdot (o_t - \mu_{il})(o_t - \mu_{il})^T}{\sum_{n=1}^N \gamma_{i_l}(n)} \quad (56)$$

Représente la variance de chaque gaussienne à l'état q_i [66].

3.11 Conclusion

Dans ce chapitre, nous avons utilisé une approche pour pouvoir segmenter les mots de notre corpus, afin de faire des comparaisons à un niveau phonémique et donner un score de vraisemblance nécessaire à l'évaluation de la prononciation d'un patient souffrant de pathologie langagière, cas des parasigmatismes.

Le prochain chapitre concerne la description du corpus saint ou de référence et le choix des mots, où le parasigmatisme est très distinctif. La modélisation par les chaînes de Markov et les mélanges de gaussiennes sera appliquée à différents locuteurs composant notre base de données, différents comparatifs seront visualisés et des scores de vraisemblances montreront la fiabilité de la segmentation automatique des phonèmes, dans un cadre d'évaluation de la prononciation et de détection du point de défaillance langagière.

CHAPITRE 4

RECONNAISSANCE PAR LES HMM

4.1. Introduction

Le traitement du signal vocal est un processus très complexe en terme d'extraction de l'information utile, car il concerne l'une des problématiques les plus influentes dans cette phase. Le corpus sur lequel s'effectuent les tests de performance est fondamental étant donné que l'extraction des paramètres acoustiques est sujette aux différentes influences qui sont dues à la négligence lors de l'enregistrement, du réglage du niveau sonore du microphone, le bruit environnant, les défauts de prononciation non pathologiques, la prononciation incorrecte d'un mot avec ou sans gémation et lors de la segmentation manuelle, etc.

Tous ces défauts influent directement, d'une façon décisive, sur la phase de reconnaissance. L'absence d'une grande base de données sonores Arabes nous a conduit à faire les tests de reconnaissance sur la base TIMIT, afin de valider nos résultats en vue d'une segmentation correcte des phonèmes.

4.2 Description du corpus

L'un des éléments clés de toute recherche est sa base de données, dans cet ordre d'idée, nous avons essayé de développer un corpus constitué de 126 mots en Arabe, destiné essentiellement à mettre en évidence l'occurrence du phonème [ʃ].

4.2.1 Bruit environnant

Les enregistrements ont été faits dans une salle non acoustique, ceci afin d'inclure le bruit environnant, car nous préconisons d'utiliser notre système d'aide dans un environnement en conditions normales de bruit (dans un cabinet de médecin, à la maison, dans une salle de classe, etc.).

4.2.2. Locuteurs

Les 126 mots du corpus ont été enregistrés par une vingtaine de locuteurs (hommes et femmes), dont les caractéristiques sont les suivantes (Table 1).

Tableau 1 : Différents locuteurs ayant enregistré le corpus
F : Féminin et M : Masculin

Code	Sexe	Age (ans)	Observations
F1	F	40	
F2	F	27	
F3	F	8	
F4	F	31	
F5	F	32	
F6	F	28	
F6P	F	21	voix Pathologique
F7	F	27	
F8	F	35	
M1	M	22	Accent régional
M2	M	27	Accent régional
M3	M	43	
M4	M	23	Nasalisation
M5	M	30	Nasalisation
M6	M	16	
M7	M	45	
M8	M	35	
M9	M	35	
M10	M	25	
M11	M	36	
M12	M	43	
M13	M	45	
M14	M	30	
M15	M	37	

En vue de faire des essais sur un corpus présentant le sigmatisme occlusif (cf. 2.8.) , nous avons pu enregistrer un cas de sexe féminin âgée de 19 ans, « code F6P », en parole continue et en mots isolés afin de bien distinguer l'impact du sigmatisme sur les mots en débit lent, normal et rapide.

La pathologie a été confirmée par le Dr. Ghozali du service de rééducation fonctionnelle du CHU Frantz Fanon de Blida.

4.2.3. Constitution du corpus

Les différents mots sélectionnés ont été tirés du dictionnaire "Al Mawrid El Wasit" du Dr. Rouhi Balabaki et al [78]. Ces mots sont divisés en 3 classes d'occurrence :

Classe 1 : Début de mot (tableau 4.2)

- ◆ 22 occurrences de [ʃ] suivi de [a]
- ◆ 17 occurrences de [ʃ] suivi de [o]
- ◆ 17 occurrences de [ʃ] suivi de [i]

Classe 2 : Milieu de mot (tableau 4.3)

- 53 occurrences du [ʃ] avec les combinaisons suivantes :
 - ◆ [aʃa],[aʃi],[aʃo] ;
 - ◆ [iʃa],[oʃi],[oʃo].

Classe 3 : Fin de mot (tableau 4.4.)

- ◆ 12 occurrences du [v_iʃ] avec [v_i]≡ [a],[o],[i].

4.3 Présentation du cas d'étude

En vue de pouvoir réaliser la première étape de segmentation automatique des phonèmes, nous avons sélectionné des mots présentant le sigmatisme occlusif d'une façon très distinguée lors de la prononciation des phonèmes [ʃ] et [s]. Ces derniers présentent une grande ambiguïté de prononciation du fait du rapprochement de leur lieu d'articulation.

Tableau 4.5. : TOP des mots sélectionnés pour la modélisation par les HMM/GMM

Code	Mots en Arabe	TOP	Prononciations pathologiques	Transcriptions en Arabe standard
C1	شخصية	[ʃ a χ s i j a]	[θ a χ t i j a] [θ a χ ʃ i j a]	ثخنية ثخشية
C2	شمس	[ʃ a m s ø]	[θ a m ʃ ø] [θ a m θ ø]	ثمش ثمث

Nous avons segmenté manuellement en mots et en phonèmes les segments dont nous avons besoin. Chaque mot ou phonème a été étiqueté avec un code spécifique, comme illustré dans le tableau 4.6.

Tableau 4.6. : Mots de tests sélectionnés à partir de la base de données.

TOP	Locuteurs	Codification
[ʃ a χ s i j a]	F	F_Chaksiya1
	M	M_Chaksiya2
[ʃ a m s ø]	F	F_Chamsse5
	M	M_Chamsse3

Chaque mot ou phonème des mots cibles est ensuite modélisé par les HMM avec les GMM correspondantes.

Avant d'entamer le traitement des données par une modélisation markovienne, nous allons voir quelques aspects visuels des signaux sonores et faire des comparaisons, dans le but de comprendre le fait de prononcer le [θ] au lieu du [ʃ] et son impact sur les différents spectrogrammes (Déplacement formantique).

4.4. Comparaisons avec différents locuteurs

A titre de comparaison visuelle, nous avons choisi quelques locuteurs (masculin et féminin) lors de la prononciation du mot C1.

Nous remarquons le défaut du [ʃ] transformé en [θ] dans la figure 4.1. patiente F6P.

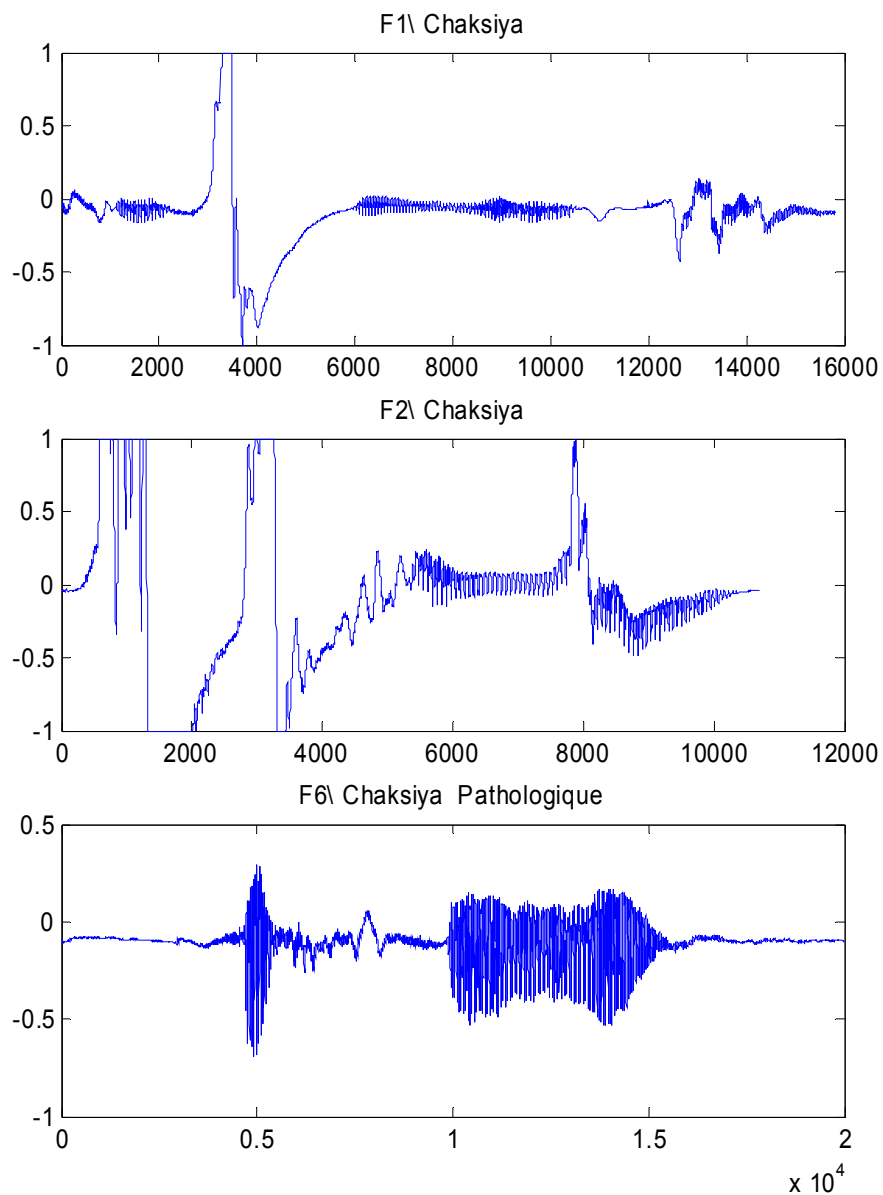


Figure 4.1 : Comparaisons visuelles des prononciations
(F1 / F2 / F6P)

La seconde comparaison est effectuée avec des locuteurs de référence, voir figure 4.2.

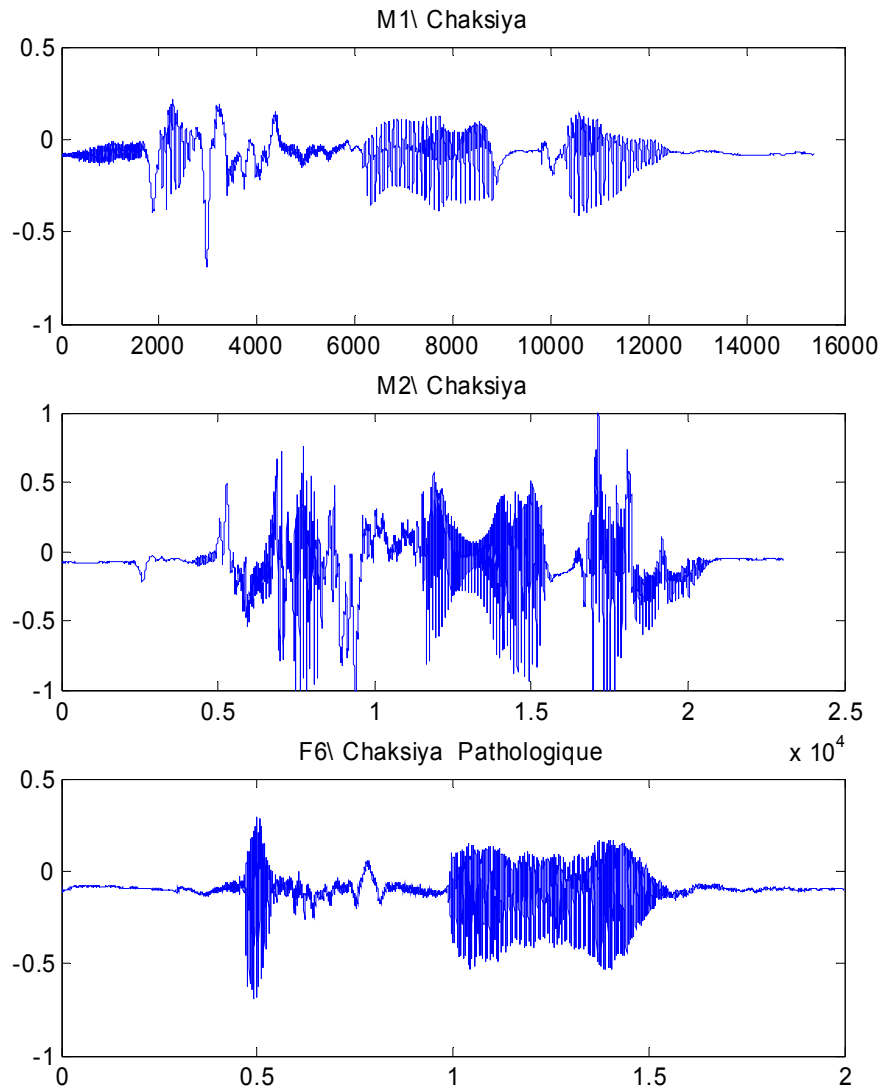


Figure 4.2. : Comparaisons visuelles des prononciations $\times 10^4$
(M1 / M2 / F6P)

Il est difficile de décider sur la pathologie visuellement, pour cela l'utilisation de méthodes utilisant des techniques d'analyse plus avancées est impérative.

4.5. Spectrogrammes et déplacement formantique

Le spectrogramme est un moyen de voir le déplacement formantique et dire si le bon phonème a été prononcé ou non, ceci se remarque par une analyse visuelle comparative des fréquences propres à chaque phonème, comme illustré dans les figures 4.3, 4.4 et 4.5.

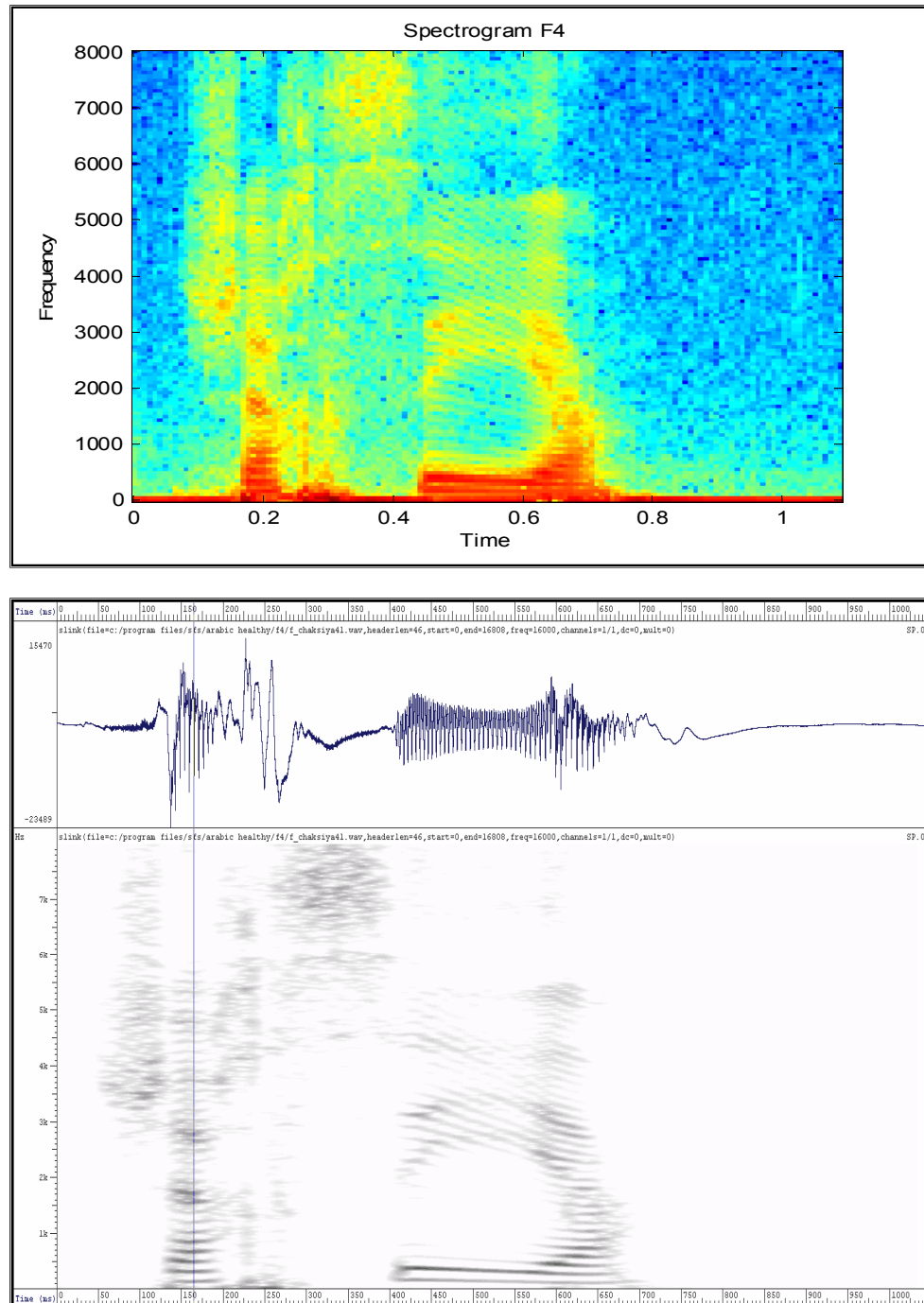


Figure 4.3. Spectrogrammes « locutrice F4 »

Locuteur M4 de référence :

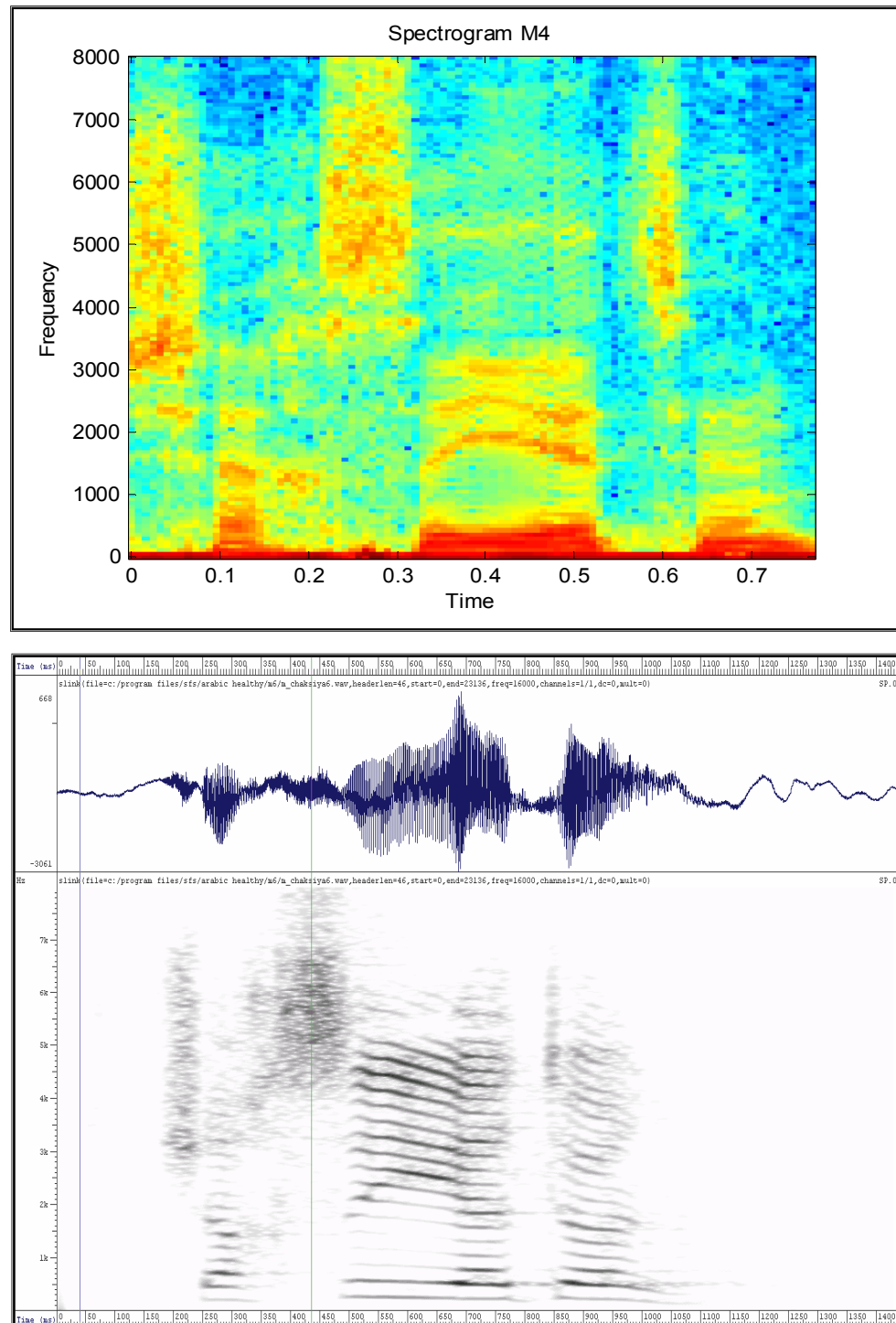


Figure 4.4. Spectrogrammes « locuteur M4 »

Patiente F6 présentant la pathologie

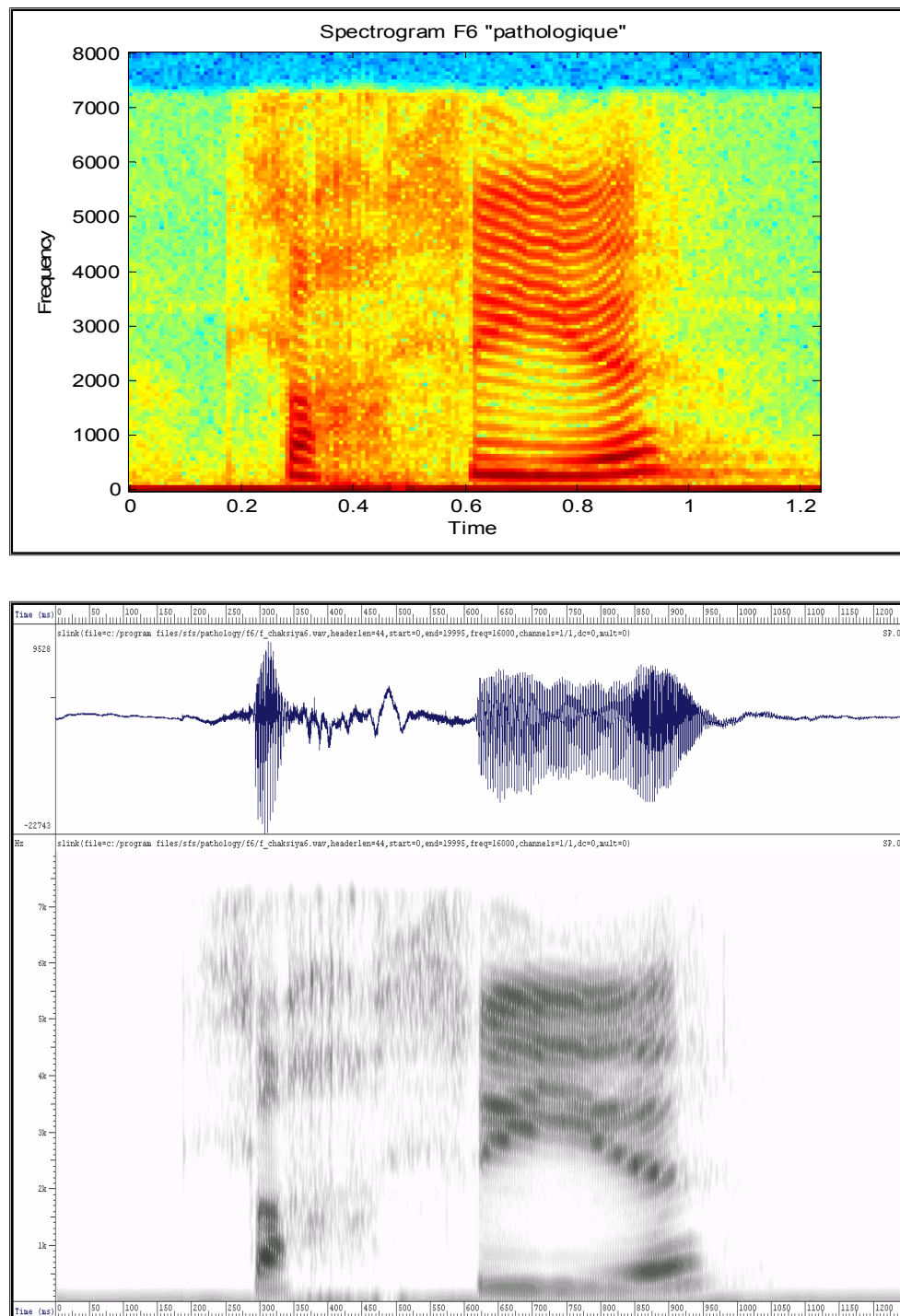


Figure 4.5. Spectrogrammes « locutrice F6 »

4.6. Méthodologie de travail

Après ces quelques comparaisons visuelles, nous avons opté pour la méthodologie suivante:

- ◆ modélisation du/des mots sélectionnés par les HMM avec 80% du corpus pour l'entraînement et 20% pour le test afin de préparer la segmentation phonémique pour la comparer auditivement avec la segmentation manuelle;
- ◆ validation de nos travaux avec une partie de la base TIMIT en terme de reconnaissance par mot (ensemble de phonèmes) ;
- ◆ modélisation du/des mots sélectionnés avec 100% du corpus pour l'apprentissage et attribuer les segments aux classes, selon le chemin de Viterbi pour aboutir à un ensemble de phonèmes segmentés automatiquement ;
- ◆ réapprentissage de nouveaux modèles HMM phonémiques avec 80% du corpus des phonèmes segmentés automatiquement et test avec les 20% restants ;
- ◆ segmentation du mot du corpus pathologique ;
- ◆ évaluation du mot et calcul de la vraisemblance.

4.7. Modélisation par les HMM

4.7.1. Modélisation par 80% du corpus total et test sur les 20% restant

En premier lieu, nous avons effectués une modélisation sur 80% du corpus, les 20% restant ont servi aux tests. Le nombre d'états des modèles de Markov est spécifique à longueur phonémique du mot, vu que les comparaisons se font au sein d'une même classe et la détection du phonème mal prononcé est sujette au mot pris comme exemple de test.

Chaque HMM/GMM est défini par :

- ◆ le modèle du HMM, dans notre cas d'étude tous les modèles sont de type gauche-droite
- ◆ le nombre d'états.
- ◆ l'état initial le plus probable
- ◆ le nombre de gaussiennes modélisant les données par état.
- ◆ le nombre de paramètres représentant les données.
- ◆ le nombre d'itérations de l'algorithme EM.

4.7.1.1. Modèles HMM du mot [ʃ a χ s i j a]

En premier lieu, nous allons modéliser le mot [ʃ a χ s i j a] par deux modèles à 6 états, avec un nombre de coefficients MFCC différents voir tableau 4.7. et 4.9, les matrices de transition obtenues sont mentionnées respectivement dans les tableaux 4.8. et 4.10.

Tableau 4.7. : Paramètres du modèle HMM1

Nombre d'états du HMM	Mélange de gaussiennes par état	Nombre d'itérations de l'algorithme EM	Coefficients MFCC/trame
6	8	15	12

Tableau 4.8. : Matrice de transition du modèle HMM1 :

	Etat 1	Etat 2	Etat 3	Etat 4	Etat 5	Etat 6
Etat 1	0.9683	0.0317	0.0000	0	0	0
Etat 2	0	0.9659	0.0319	0.0023	0	0
Etat 3	0	0	0.9690	0.0288	0.0022	0
Etat 4	0	0	0	0.9725	0.0275	0.0000
Etat 5	0	0	0	0	0.9738	0.0262
Etat 6	0	0	0	0	1.0000	0

Tableau 4.9 : Paramètres du modèle HMM2

Nombre d'états du HMM	Mélange de gaussiennes par état	Nombre d'itérations de l'algorithme EM	coefficients MFCC/trame
6	8	20	39

Tableau 4.10. : Matrice de transition du modèle HMM2

	Etat 1	Etat 2	Etat 3	Etat 4	Etat 5	Etat 6
Etat 1	0.9711	0.0289	0	0	0	0
Etat 2	0	0.9721	0.0279	0	0	0
Etat 3	0	0	0.9548	0.0452	0	0
Etat 4	0	0	0	0.9733	0.0267	0
Etat 5	0	0	0	0	0.9699	0.0301
Etat 6	0	0	0	0	0	1.0000

Remarquons que le modèle HMM2 s'affine de plus en plus vers un modèle de type gauche droite, par l'annulation des probabilités de transition état 2 vers état 4 ou état 5 du modèle HMM1.

Les figures 4.6. (a,b,c,d) montrent la segmentation phonémique selon 12 coefficients MFCC seulement, remarquons que les observations associées à l'état 6 ne sont pas distribuées proportionnellement sur tous les locuteurs.

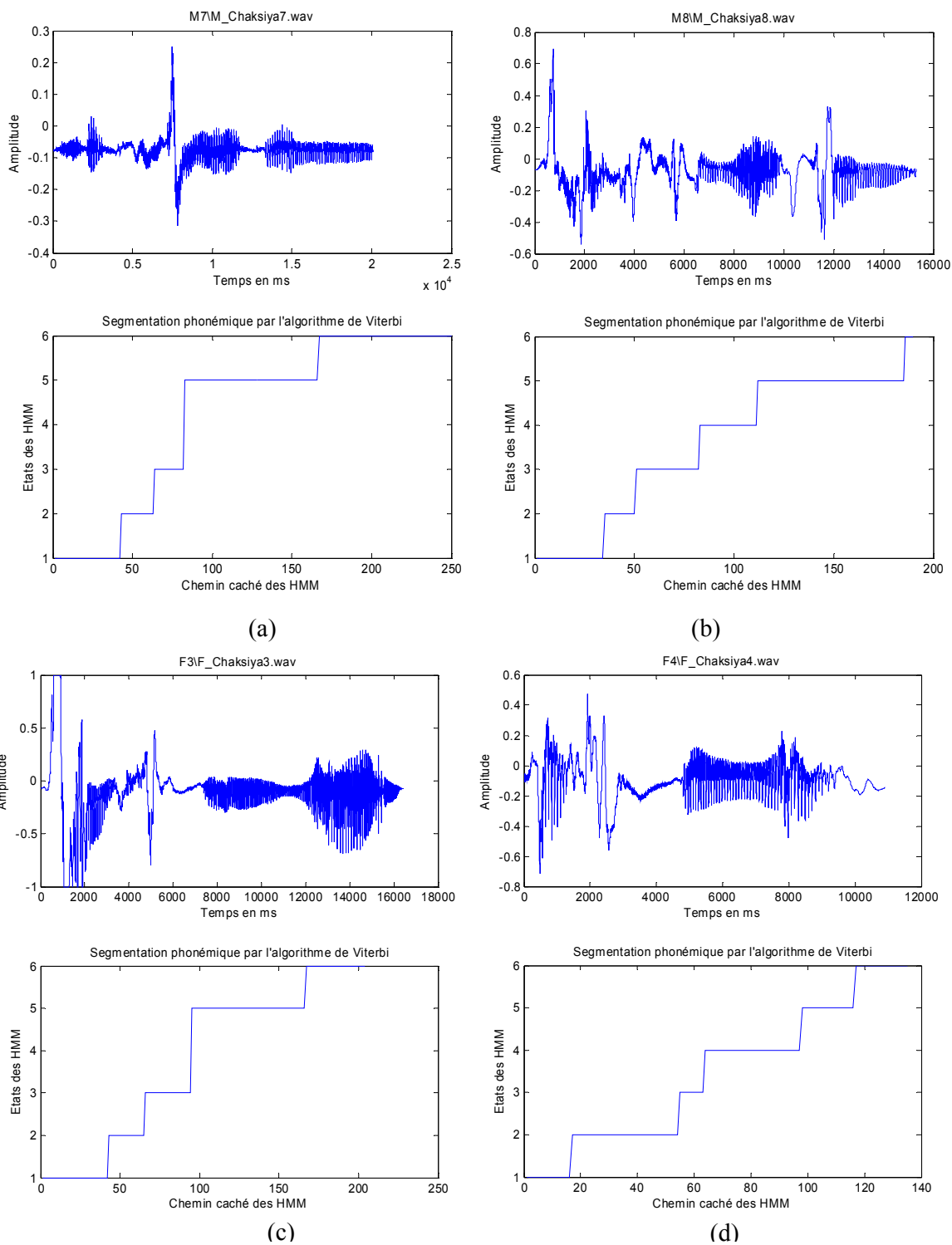


Figure 4.6. (a,b,c,d) : Segmentation phonémique pour différents locuteurs en utilisant 12 coefficients MFCC

Les 12 coefficients MFCC sont adjoints avec leurs dérivées premières et seconde ainsi que le log énergie de la trame complète.

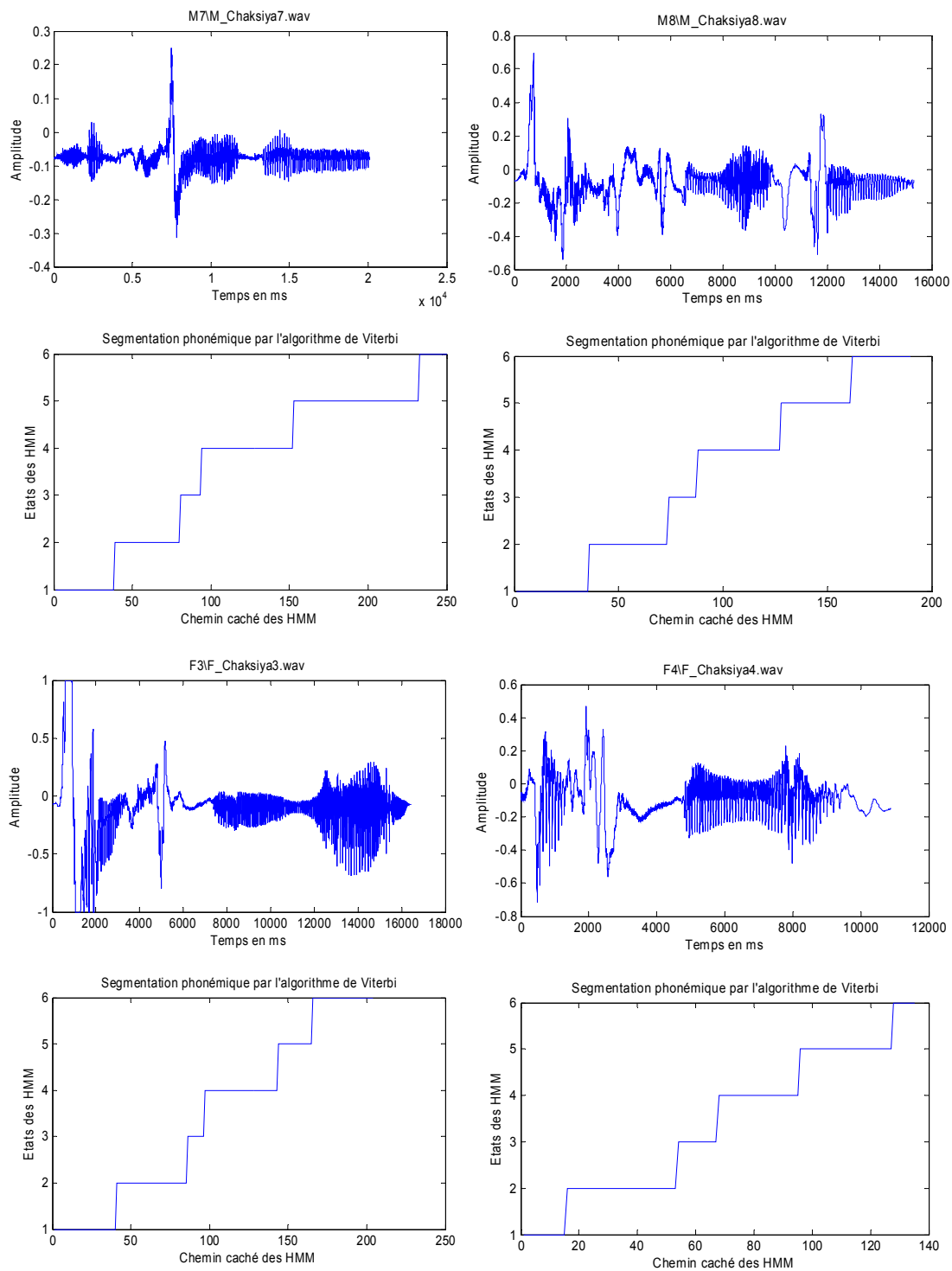


Figure 4.7. Segmentation phonémique pour différents locuteurs en utilisant 39 paramètres.

4.8. Courbes de convergence de l'algorithme EM

La figure 4.8. illustre la différence de convergence de l'algorithme EM après l'adjonction des dérivées première et seconde des MFCC, ainsi que de la log énergie des trames [79].

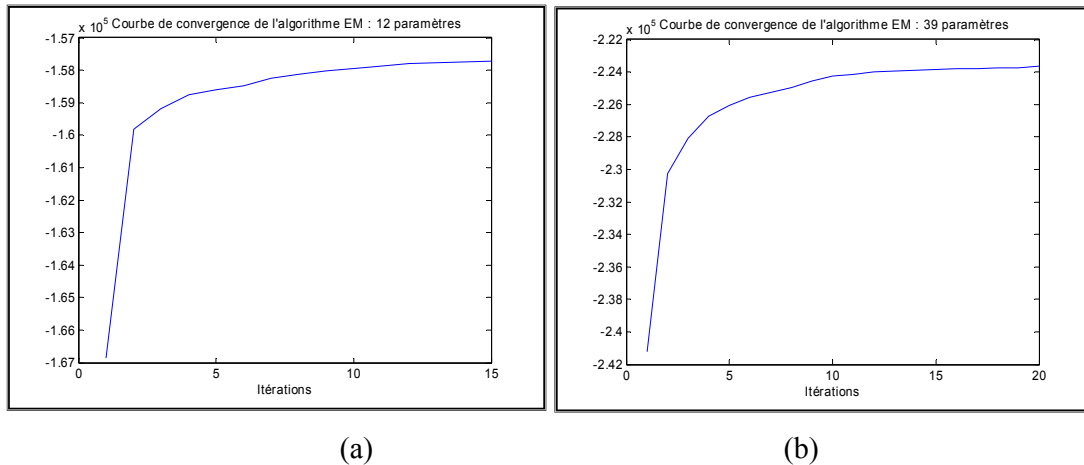


Figure 4.8. Courbe de convergence de l'algorithme EM

Modèle HMM1 (a) : 12 paramètres ; Modèle HMM2 (b) : 39 paramètres

Discussions

Nous avons remarqué que l'augmentation du nombre de coefficients agit directement sur la matrice de transition et fait tendre notre HMM à un modèle de type gauche droite, pour une meilleure segmentation phonémique. Toutefois, la segmentation n'est pas réellement bien réalisée, ceci est essentiellement dû au fait que le mot [ʃaxsijatun] ou tout autre mot Arabe est composé de phonèmes co-articulés, en plus de la terminaison [tun] qui est parfois ajoutée, et qui n'a pas été enregistrée par tous les locuteurs.

4.9. Modèles HMM du mot [ʃ a ʒ s i j a t u n]

Tableau 4.11. : Paramètres du modèle HMM3

Nombre d'états du HMM	Mixture de gaussiennes par état	Nombre d'itérations de l'algorithme EM	Coefficients MFCC/trame
8	8	20	39

Tableau 4.12: Matrice de transition du modèle HMM3

	Etat 1	Etat 2	Etat 3	Etat 4	Etat 5	Etat 6	Etat 7	Etat 8
Etat 1	0.96164	0.038358	0	0	0	0	0	0
Etat 2	0	0.96192	0.038077	0	0	0	0	0
Etat 3	0	0	0.96346	0.036541	3.8717e-187	0	0	0
Etat 4	0	0	0	0.94777	0.05223	5.9338e-228	0	0
Etat 5	0	0	0	0	0.96959	0.028894	0.0015207	0
Etat 6	0	0	0	0	0	0.95251	0.044989	0.0024992
Etat 7	0	0	0	0	0	0	0.96725	0.032747
Etat 8	0	0	0	0	0	0	0	1

La courbe de convergence de l'algorithme EM, figure 4.9 illustre la rapidité de convergence.

Notons toutefois, que le modèle contient assez d'informations (100% du corpus) pour être de type gauche droite, mais la segmentation phonémique n'est pas toujours bien réalisée, ceci est essentiellement dû à la segmentation manuelle, durant laquelle les silences sont parfois collés aux phonèmes.

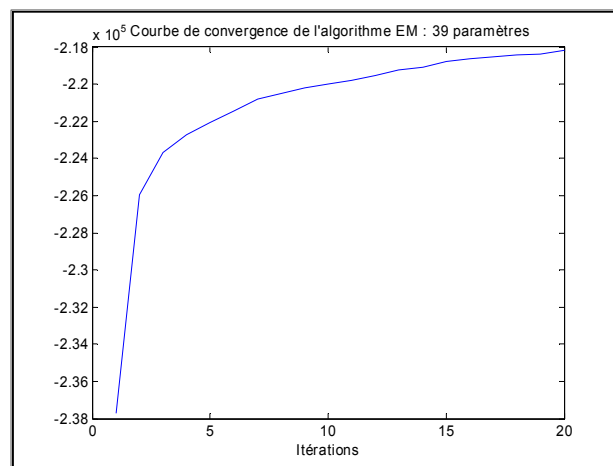


Figure 4.9. Courbe de convergence de l'algorithme EM- Modèle HMM3

4.10. Modélisation de la phrase SA1 de la base TIMIT

Cette étape est une validation de notre modélisation HMM, en effet, notre choix se porte sur la phrase SA1 : «She had your dark suit in greasy wash water all year», nous allons essayer de reconnaître les 11 mots de cette phrase par notre approche, voir tableau 4.13.

Tableau 4.13. : Modèles HMM de la phrase SA1 par 80% du corpus (80 locuteurs)

Mots	Modèle HMM	Matrice de Transition	Moyenne par état	Variance par état	Mélange de gaussiennes
She	Mot1	5*5	39*5*5	39*39*5*5	5 * 5
had	Mot2	4*4	39*4*5	39*39*4*5	4 * 5
your	Mot3	2*2	39*2*5	39*39*2*5	2 * 5
dark	Mot4	4*4	39*4*5	39*39*4*5	4 * 5
suit	Mot5	3*3	39*3*5	39*39*3*5	3 * 5
in	Mot6	1*1	39*1*5	39*39*1*5	1 * 5
greasy	Mot7	7*7	39*7*5	39*39*7*5	7 * 5
wash	Mot8	4*4	39*4*5	39*39*4*5	4 * 5
water	Mot9	5*5	39*5*5	39*39*5*5	5 * 5
all	Mot10	2*2	39*2*5	39*39*2*5	2 * 5
year	Mot11	3*3	39*3*5	39*39*3*5	3 * 5

Reconnaissance des mots de la phrase SA1 par les 20% du corpus restant (20 locuteurs) :

Tableau 4.14 : Taux de reconnaissance sur la base TIMIT (20 locuteurs)

Mots	Modèle HMM	Taux de reconnaissance
She	Mot1	95%
had	Mot2	100%
your	Mot3	100%
dark	Mot4	100%
suit	Mot5	100%
in	Mot6	95%
greasy	Mot7	100%
wash	Mot8	100%
water	Mot9	100%
all	Mot10	100%
year	Mot11	100%
Taux de reconnaissance Phrase complète		99%

La phrase SA1 est reconnue à 99% des cas, ceci montre qu'en terme de reconnaissance, l'approche HMM est adéquate à notre problème, toutefois la base TIMIT est un « clean speech » enregistrement sans bruit, donc l'approche est « stable » si la base contient le moins de bruit possible.

Cette condition ne peut se réaliser qu'avec des conditions d'enregistrement très rigoureuses. Cas très difficile à mettre en oeuvre, surtout lorsque le patient ou l'orthophoniste travaillent que dans le cabinet du médecin ou à la maison.

Nous allons, maintenant, extraire les phonèmes à modéliser de la structure HMM3, en considérant l'une des premières raisons possibles de la chute de reconnaissance, les moyens d'enregistrement non professionnels qui n'étaient pas disponibles au moment de réalisation de ce travail.

4.11. Degré de Vraisemblance des locuteurs par rapport au modèle HMM3

En vue de comparer le degré de vraisemblance de notre corpus par rapport au modèle HMM3, tout les locuteurs ont été utilisés dans la phase de test.

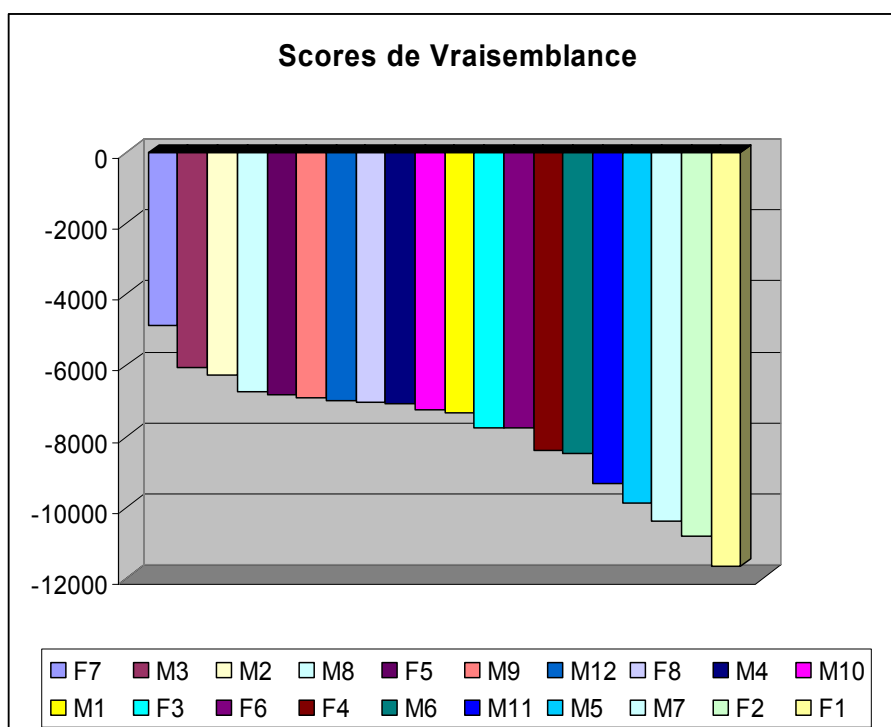


Figure 4.10. : Degré de vraisemblance que HMM3 ait généré les observations acoustiques par locuteur

Nous remarquons, dans la figure 4.10. qu'il y'a déjà dissemblance intra classe dans le modèle HMM3, nous pouvons dire que ce modèle est plus adapté à générer les observations du locuteur F7 et moins adapté à générer les valeurs de l'observation de F1.

4.11.1. Degré de vraisemblance du cas pathologique F6P

Avant de modéliser les phonèmes constituant $[\text{a} \chi \text{s i j a t u n}]$, calculons la probabilité que le HMM3 ait généré les observations de la locutrice F6 (pathologique).

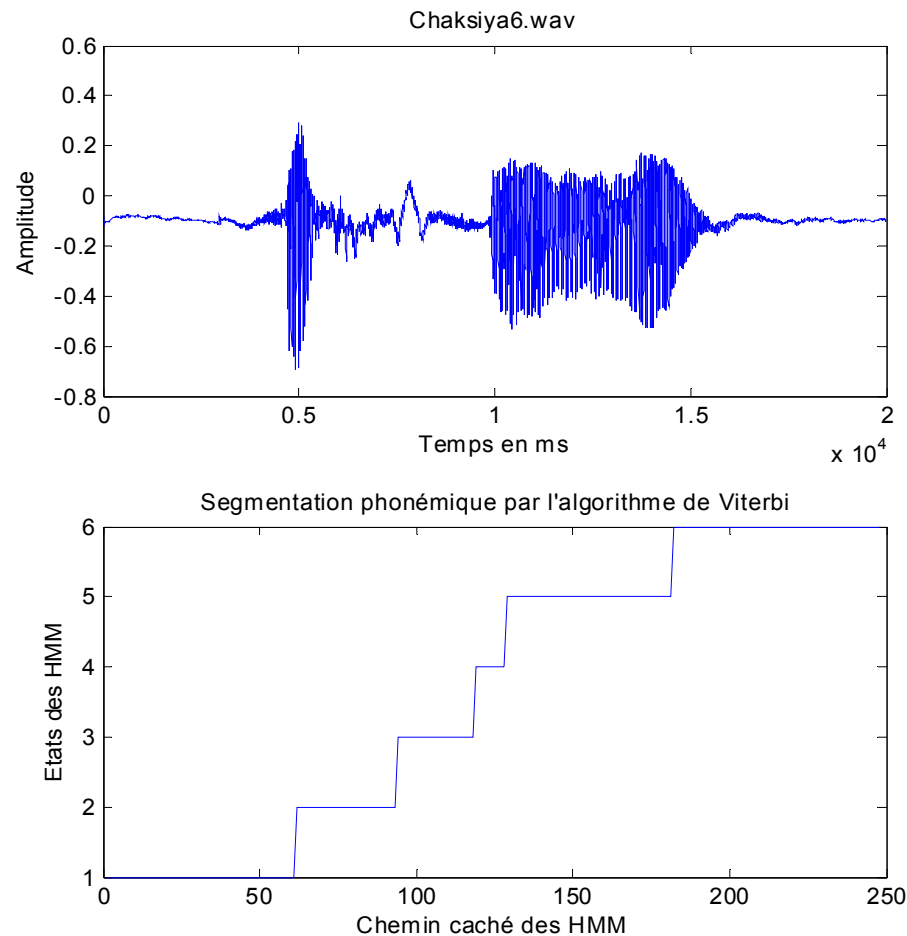


Figure 4.11. Segmentation du mot pathologique $[\text{a} \chi \text{s i j a t u n}]$

Remarque: La locutrice a prononcée $[\theta \text{ a} \chi \text{s i j a}]$ au lieu de $[\text{a} \chi \text{s i j a t u n}]$

Lorsque nous essayons de segmenter manuellement le mot prononcé, le phonème $[\theta]$ correspond **auditivement** au segment automatiquement segmenté par l'algorithme de Viterbi, c'est à dire qu'il y'a eu distinction phonémique ce qui tend à diminuer la probabilité d'entendre le $[\text{j}]$.

Le degré de vraisemblance ou loglikelihood calculée est de : -15361, ceci montre que la prononciation pathologique est nettement inférieure à toutes les prononciations de référence.

Remarquons que l'on segmente [Ja] au lieu de [j] et [tu] au lieu de [t], etc..., à cause de la co-articulation qui tend à minimiser l'effet de la segmentation phonémique automatique par rapport à la segmentation manuelle qui est basée sur du visuel et de l'auditif.

4.12. Résultat de l'apprentissage visuel et auditif

Après quelques heures de feedback visuels et auditifs, la patiente F6P a appris à prononcer le mot C1 « **correctement** », la figure 4.12 illustre le changement dans la prononciation.

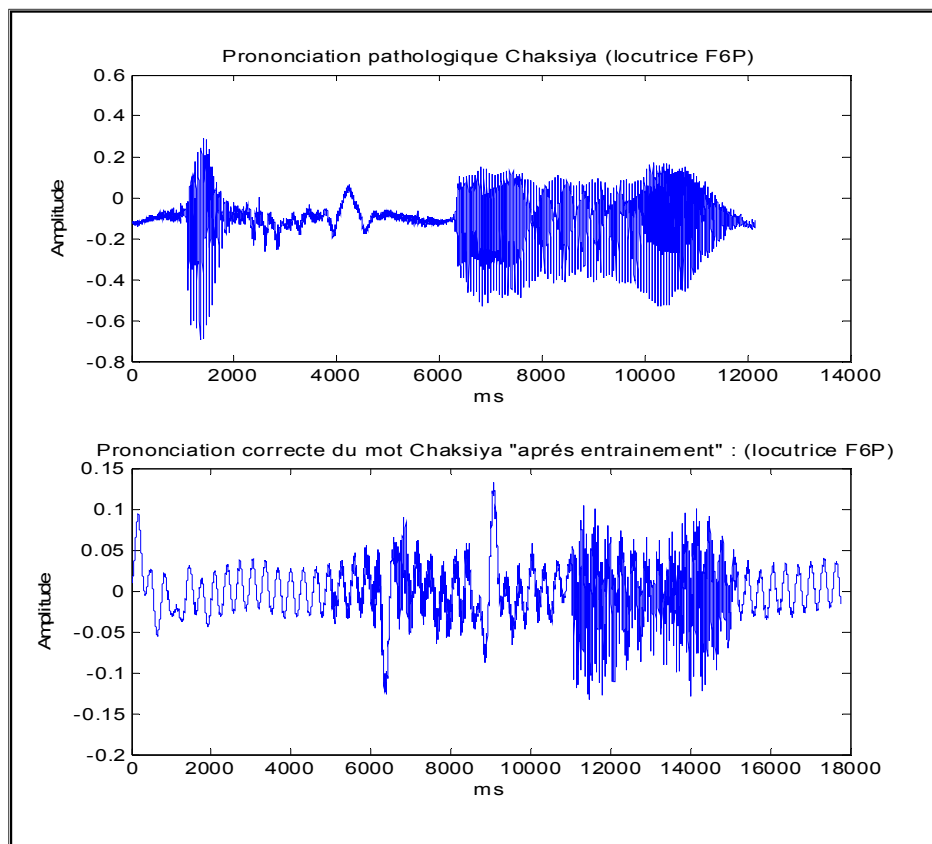


Figure 4.12 : Prononciations de la locutrice F6P, avant et après
les feedback visuels et auditifs

4.12. Segmentation et reconnaissance phonémique

A cette étape, les données issues du modèle HMM3 vont être utilisées, pour modéliser les HMM élémentaires des phonèmes : [ʃa],[χ],[si],[ja],[tu],[n].

La segmentation phonémique du mot C1, nous a permis de définir un nouveau modèle HMM propre au phonème [ʃa], voir tableaux 4.15. et 4.16., ainsi qu'un modèle HMM du phonème [χ], voir tableaux 4.17 et 4.18.

4.12.1. Modèle HMM du phonème [ʃa] de type gauche droite

Tableau 4.15. : Paramètres du modèle HMM du phonème [ʃ]

Nombre d'états du HMM	Mixture de gaussiennes par état	Nombre d'itérations de l'algorithme EM	Coefficients MFCC/trame
3	6	10	39

Tableau 4.16. : Matrice de transition du modèle du phonème [ʃ]

	Etat 1	Etat 2	Etat 3
Etat 1	0.8573	0.1427	5.528e-116
Etat 2	0	0.8632	0.1368
Etat 3	0	0	1

4.12.2. Modèle HMM du phonème [χ] de type gauche droite

Tableau 4.17. : Paramètres du modèle HMM du phonème [χ]

Nombre d'états du HMM	Mixture de gaussiennes par état	Nombre d'itérations de l'algorithme EM	Coefficients MFCC/trame
3	6	10	39

Tableau 4.18. : Matrice de transition du modèle HMM du phonème [χ]

	Etat 1	Etat 2	Etat 3
Etat 1	0.86994	0.13006	3.9418e-253
Etat 2	0	0.87782	0.12218
Etat 3	0	0	1

Les tests inter classes nous ont donnés 100% de reconnaissance, ceci s'explique par le fait que la modélisation est doublement stochastique, modélisation HMM des mots de référence donnant lieu à une premier niveau de segmentation phonémique puis re

modélisation des données par des HMM donnait lieu à des modèles phonémiques minimaux de HMM.

4.13 Degré de vraisemblance phonémique

La dernière étape de reconnaissance phonémique est illustrée par les degrés de vraisemblances entre la prononciation correcte et la prononciation corrigée de la locutrice F6P.

Nous avons segmenté manuellement le mot [ʃaχsija] présentant un sigmatisme occlusif au niveau du [ʃ] et la version « correcte » prononcée par la locutrice F6P lors des séances d'enregistrement, afin de calculer le degré de vraisemblance entre le modèle HMM doublement stochastique et les prononciations réelles.

Le comparatif a été fait avec les deux premiers phonèmes du mot, car nous considérons l'occlusion initiale pour cet exemple, voir tableau 4.19.

Tableau 4.19. : Vraisemblance phonémique

N°	Phonème segmenté manuellement	Prononciation avec sigmatisme	Prononciation correcte	Phonème [χ]	Phonème [ʃ]
1	[θ]	-1002.6	-917		HMM 2
2	[ʃ]	-3073.4	-2732.1		HMM 2
3	[a]	-204.3149	-273.1638	HMM 1	
4	[a]	-395.9116	- 623.6084	HMM 1	
5	[χ]	-1.0824	-1.4440	HMM 1	
6	[χ]	-1352.0	-1830.3	HMM 1	
7	[s]	-910.3196	-907.6913		HMM 2
8	[s]	-0.9458	-1.0357	HMM 1	
9	[i]	-1.6636	-1.3418		HMM 2
10	[i]	-1.2882	-1.0783		HMM 2
11	[j]	-2.4896	-2.1206		HMM 2
12	[j]	-1.3417	-1.1025		HMM 2

D'après ces données, nous pouvons confirmer que le fait que le patient se corrige lors des sessions d'apprentissage de la prononciation tend à augmenter le degré de vraisemblance (valeurs moins négatives) entre phonèmes prononcés, ceci aidera à améliorer le seuil d'écoute et de prononciation du patient lors des sessions d'orthophonie.

Discussions

Nous pouvons dire qu'à ce stade de segmentation et de reconnaissance, les phonèmes même proches en lieu d'articulation ont été classés selon le rapprochement du lieu d'articulation, voir tableau 1.4.

4.14 Conclusion

Dans ce chapitre, nous avons adoptée une méthodologie de détection du sigmatisme occlusif par l'utilisation des coefficients MFCC connus pour leur robustesse au bruit et une modélisation stochastique basée sur les HMM à base de GMM.

Différents tests et comparatifs ont été réalisés sur un corpus enregistré localement ainsi que des tests sur la phrase SA1 de la base TIMIT, ceci nous a permis de déduire des règles de travail relatives aux paramètres de l'enregistrement et au choix du corpus.

La segmentation phonémique automatique obtenue nous a permis de comprendre les difficultés de la co-articulation qui tend à minimiser la séparation des phonèmes, toutefois l'utilisation croisée de la segmentation manuelle en plus de la modélisation Markovienne nous a permis de résoudre une grande partie du problème posé.

CONCLUSION

La Reconnaissance Automatique de la Parole a connu un essor incommensurable par l'introduction de la modélisation stochastique, certainement, depuis les études établies par J. Fergusson et L. Rabiner.

La modélisation du signal vocal est régie par le choix des vecteurs acoustiques dont les performances doivent être des plus optimales, en terme de sensibilité au bruit, de complexité de calcul, de degré de perception, etc.

Dans ce travail, nous avons développé une application complémentaire au traitement de la parole qui concerne la segmentation phonémique orientée vers l'aide à la correction de la prononciation des sigmatismes occlusifs sur micro-ordinateur, cette approche est basée sur des "feedback" visuels et auditifs qui aident l'orthophoniste ou le patient à réentendre la prononciation incorrecte présentant la pertinence de la pathologie en lui indiquant les zones d'erreurs possibles.

La détection de la défaillance de prononciation du phonème [j] prononcé [θ], a été traitée ainsi que différents comparatifs ont été réalisés avec des corpus saints et pathologiques enregistrés au sein du laboratoire de recherche LATSI, ainsi qu'avec la base TIMIT en vue de valider les travaux pour une extension à d'autres maladies.

Nous avons pu segmenter automatiquement le phonème pathologique [θ] et extrait son emplacement dans le mot pathologique cible, les différents degrés de vraisemblance obtenus par rapport au modèle de référence montrent que la fiabilité des résultats est régie par les conditions d'enregistrement, ainsi que le corpus de référence qui modélise le mieux les différents modèles Markoviens des phonèmes à traiter.

A titre comparatif, la modélisation de la phrase SA1 de la base TIMIT nous a donné un taux de reconnaissance de 99% pour la reconnaissance de toute la phrase, ceci montre la fiabilité de l'utilisation des HMM/GMM dans cet axe de recherche.

Notre approche est basée sur une aide à l'apprentissage à l'aide d'images et de sons suivi de la génération d'un score de vraisemblance, ou « note d'appréciation » qui aide le patient à corriger sa prononciation au fur et à mesure.

Les résultats des expériences réalisées nous ont permis de tirer des règles relatives au choix du corpus de travail, à l'impact de la segmentation manuelle et les erreurs de reconnaissance issues des erreurs infimes d'approximation ainsi que le nombre et le type de vecteurs acoustiques à utiliser.

En termes de perspectives à ce travail, nous préconisons de modéliser les autres sigmatismes, les schlintements ainsi que les autres défauts pathologiques détectables par l'ouïe.

APPENDICE A

LISTE DES SYMBOLES ET DES ABREVIATIONS

API / IPA	:	Alphabet Phonétique International/International Phonetic Alphabet
CMC / HMM	:	Chaînes de Markov Caché / Hidden Markov Models
DARPA	:	Defense Advanced Research Projects Agency
DTW	:	Comparaison par déformation temporelle / Dynamic Time Warping
DWT	:	Transformée en ondelette / Discrete Wavelet Transform
E.M.	:	Algorithme Maximisation de l'Espérance
LFCC	:	Coefficients Cepstraux linéaires en fréquence
LDA	:	Analyse discriminante linéaire / Linear Discriminant Analysis
LPC	:	Codage linéaire prédictif / Linear predictive coefficients
LPCC	:	Coefficients de Prédiction Linéaire Cepstrale
LPCC	:	Coefficients de Prédiction Linéaire Cepstrale
MFCC	:	Coefficients Cepstraux d'échelle MEL en fréquence
MG / GMM	:	Mélange de Gaussiennes/ Gaussian Mixture Model
PCA	:	Analyse en Composantes Principales / Principal Component Analysis
PLP	:	Coefficients de Prédiction Perceptuelle
PSOLA	:	Pitch Synchronous OverLap and Add
RASTA	:	Technique Spectrale Relative
RNA/ ANN	:	Réseaux de neurones / Artificial Neural Networks
TAP	:	Traitement Automatique de la Parole
TPO	:	Transcription Phonétique Orthographique
TPZ	:	Taux de Passage par Zéro
VADS	:	Voies Aérodigestives Supérieures
VQ	:	Quantification Vectorielle
Δ MFCC	:	Variation première des MFCC
$\Delta\Delta$ MFCC	:	Variation seconde des MFCC

REFERENCES

1. http://www.sfu.ca/~saunders/l33098/L3/Respiration_02.html
2. Encyclopédie du corps humain, CD-ROM, TLC-Educsoft.
3. <http://www.kehlkopfoperiert.ch/F/f-lunge.htm>
4. http://cystic-fibrosis-symptom.com/lungs_trachea.htm
5. <http://fr.wikipedia.org/wiki>
6. http://sprojects.mmi.mcgill.ca/larynx/notes/n_frames.htm
7. <http://www.lli.ulaval.ca/labo2256/lexique>
8. <http://www.geocities.com/phlplacephntics/elements.htm>
9. <http://www2.arts.gla.ac.uk/IPA/fullchart.html>
10. <http://www.linguistes.com/phonetique/phon.html>
11. http://al-zahra.net/qurannet/old_qurannet/main/quran/lesson6/tajweed-MKH.htm
12. J. CANTINEAU « Cours de phonétique arabe », Klincksieck (1960)
13. T.Saidane, M. Zrigui et M.Benahmed, «La transcription orthographique-phonétique de la langue arabe», Société Tunisienne d'Electricité et du Gaz, Centre de production de Sousse; Laboratoire RIADI, Unité Monastir, Faculté des sciences de Monastir; Laboratoire RIADI, Ecole Nationale des Sciences de l'informatique, Tunis, Tunisie, (2004).
14. T. Koizumi, M. Mori, S. Taniguchi, and M. Maruya, «Recurrent neural networks for phoneme recognition » ; Dept. of Information Science, Fukui University, Japan, (1996).
15. O.Essa, «Using supra-segmentals in training H Markov Models for Arabic», Computer Science Dept, University of South California, (1998).
16. A.M.A.Ali, J.V.Der-spiegel, P. Mueller, G.Haentjens and J.Berman, «An acoustic-phonetic feature-bases system for automatic phoneme recognition in continuous speech », University of Pennsylvania, (1999).

17. A.Juneija and C.Epsy-Wilson, « Segmentation of continuous speech using acoustic-phonetic parameters and statistical learning », ECE Dept., University of Maryland, College Park, USA, (2003).
18. M. Chetouani, B.Gas and J.L. Zarader, « Coopération entre codeurs-Neuro prédictifs pour l'extraction de caractéristiques en reconnaissance de phonèmes», Laboratoire des Instruments et systèmes d'Ile de France, Université Paris IV, France, (2004).
19. Y. Lee, K.Papineni, S.Roukos, O.Emam, H.Hasny, "Language model based Arabic word segmentation", Proceedings of the 41st Annual meeting of the Association for Computational Linguistics, (July 2003), 399-406 p.
20. B. LE Viet, «Reconnaissance automatique de digits en anglais en conditions bruitées », Thèse de DEA d'informatique, systèmes et communications, INP de Grenoble, France, (2002).
21. C. Lévy, G. Linarés, P. Nocera et J.F Bonastre, «Reconnaissance de chiffres isolés embarquée dans un téléphone portable», Laboratoire Informatique d'Avignon, France, (2004).
22. M. Akbar et J.Caelen, « Parole et traduction automatique : le module de reconnaissance RAPHAEL», Université Joseph Fourier, Grenoble, France, (1998).
23. J. Cernocky, G. Baudoin, et G. Chollet, «ALISP : Quelques outils pour une analyse acoustico - phonétique de la parole indépendante de la langue», Revue Parole, Déc. (2000).
24. S. Poitoux, « Etude des mesures de confiance dans le traitement de la parole avec application en logopédie», Faculté polytechnique de Lausanne, Suisse (2002).
25. Z. Zemirli et N. Vigouroux, «Vers une modélisation de la durée des sons pour la génération automatique du rythme dans la synthèse de la langue Arabe», XXIII^{èmes} journées d'études sur la parole, Aussois, Institut de Recherche en Informatique de Toulouse, France, (2000).
26. M. Moscato et J. Wittwer, « La psychologie du langage », Collection 'Que sais je ?', Presses Universitaires de France, (1978), 75 p.

27. Fédération nationale des orthophonistes, « Semaine nationale de prévention des troubles du langage 27 au 31 Mai 2002 », Fédération Nationale des Orthophonistes, SPODS, France.
28. <http://www.geneva-link.ch/ceppim/final/ORL/Laphonation.htm>, Université de Genève.
29. <http://www.upmc.edu>, Université de Pittsburgh, Voice Center.
30. J. Koufman, «Bowling of the Vocal Cords», the visible voice, Vol. 3 N°.2, (April, 1994).
31. T. Murry, C.A. Rosen, «Vocal fold granuloma», University of Pittsburgh Voice Center, Department of Otolaryngology, Pittsburgh, Pennsylvania, (23 May 2001).
32. www.ghorayeb.com, Otolaryngology-Head & Neck Surgery.
33. www.ligue-cancer.net, «La ligue contre le Cancer: Information et prévention, les cancers des voies aéro-digestives», Ligue Nationale contre le cancer, Paris.
34. <http://hsc.virginia.edu>, Health System, University of Virginia.
35. K. Shahin, «Remarks on the speech of Arabic-Speaking Children with cleft palate, The University of British Columbia, 2002.
36. K. Verdolini, K. DeVore, S. McCoy et J. Ostrem, «Vocology Guide», National Centre for Voice and speech, (2002).
37. <http://www.aquacorp.be/logopedie/index.php>.
38. H. S. Venkatagiri, « Voice is one aspect of speech production », Department of Psychology Iowa State University (2003).
39. www.vulgaris-medical.com.
40. <http://www.infovoyager.com/francais/wikipedia/r/rh/rhotacisme.html>.
41. <http://dictionnaire.metronimo.com/term/53aa5da457a7acaaa2,xhtml>.
42. <http://www.begaiement.org/faqapb.htm>.
43. http://www.geneva-link.ch/ceppim/final/ORL/La_phonation.htm#trouble.
44. <http://encyclopaedic.net/franc/dy/dyslalie.html>.
45. <http://tecfa.unige.ch/tecfa/teaching/UVLibre/tp-iish/ex-9798/logopedie/trouble1.htm>.
46. <http://imerosendael.free.fr/orthopho.htm>.

47. <http://www.tadjweed.com/Makharijul7uroof.htm>.
48. www-2.cs.cmu.edu/~ref/mlim/chapter5.html.
49. L.R. Rabiner, Fellow IEEE, « A tutorial on Hidden Markov Models and selected Applications in Speech recognition », Proceedings of the IEEE, Vol. 77 N°. 2, (February 1989).
50. <http://jedlik.phy.bme.hu/~gerjanos/HMM/node2.html>.
51. O. Deroo, C. Ris, S. Gielen and J. Vanparys, «Automatic detection of mispronounced phonemes for language learning tools», Faculté Polytechnique de Mons, Faculté Universitaires Notre-dame de la Paix, Belgique, (2000).
52. <http://www.voiceacademy.org:8080/vaweb/glossary.html> (Effet lombard).
53. Y.Laprie, «Analyse spectrale de la parole», Cours, CRIN, Nancy, France, (2002).
54. A. R.Elobeid Ahmed, «Performance Tests on Several Parametric representations for an Arabic phoneme recognition system using HMM's», Transactions on Information and communications Technologies, Vol. 20, (1998).
55. D. J. Kershaw, «Phonetic Context-Dependency In a Hybrid ANN/HMM Speech Recognition System », Thèse de Doctorat, St. John's College, University of Cambridge, (January 28, 1997).
56. S. A. Selouani et J. Caelan, «Un système hybride pour l'identification de traits phonétiques complexes de la langue Arabe », (2004).
57. Y. Tsubota, T. Kawahara and M. Dantsuji, «Recognition and verification of English by Japanese students for computer-assisted language learning », School of informatics, Center for information and multimedia studies, Kyoto University, (2002).
58. A.A. Dibazar, S. Narayanan and T.W. Berger, «Feature analysis for automatic detection of pathological speech », Biomedical Eng. Dept, Elect. Eng. Dept., University of South California, (2002).
59. Joe Tebelskis, «Speech Recognition using Neural Networks», Thèse de Doctorat, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890, (1995).

60. C. Barras, «Reconnaissance de la parole continue: Adaptation au locuteur et contrôle temporel dans les modèles de Markov cachés », Thèse de Doctorat de l'université de Paris VI, (1996).
61. Q.C. Nguyen, «Reconnaissance de la parole en langue Vietnamiennne », Thèse de Doctorat, Institut National Polytechnique de Grenoble, (2002).
62. Jeffery A. Jones, «Learning to produce speech with an altered vocal tract », Acoustical Society of America, (2003).
63. M.O. Rosa, J.C. Pereira and M. Grellet, «Adaptative estimation of residue signal for voice pathology diagnosis », IEEE, (2000).
64. G. Stemmer, C. Hacker, E. Noth and H. Niemann, «Multiple time resolutions for derivatives of Mel-Cepstral Coefficients», Université Erlangen, Nuremberg, (2001).
65. W.J.Picone., «Signal Modeling Techniques in Speech Recognition», Proceedings of IEEE, vol. 9, (1993),1215-1247 p.
66. J.A. Bilmes, «A Gentle Tutorial of the EM Algorithm and its applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models», International Computer Institute, Berkeley, (April 1998).
67. H.Franco, L. Neuymer, M. ramos and H. Bratt, «Automatic detection of phone-level mispronunciation for language learning », SRI International, Speech technology and research Laboratory, USA, (1999).
68. J.Cernocky, G. Baudoin, D. Petrovaska-Delacretaz et G. Chollet, «Vers une analyse acoustico-phonétique de la parole indépendante de la langue, basée sur ALISP », FEI UVT, Bruno, Tchécoslovaquie ; ESIEE , France ; CNRS-LTICI-ENST , France, (22 Janvier 2002).
69. R. Boite, H. Boulard, T. Dutoit, J. Hancq et H. Leich, « Traitement de la parole » Presses polytechniques et universitaires romandes, (2000).
70. D.M.Istrate, «Détection et reconnaissance des sons pour la surveillance médicale », Thèse de Doctorat, pp. 95-129, CLIPS- IMAG, France, (2003).
71. C. Levy, G. Linares, P. Nocera, «Comparaison of several acoustic modelling techniques and decoding algorithms for embedded speech recognition systems », LIA/CERI, STEPMIND, France, (2003).

72. L. Lefort, T. Merlin, J. Bonastre et P. Nocera, «Le projet MTM–Reconnaissance de la parole et du locuteur sur une plateforme embarquée», Laboratoire Informatique d'Avignon, (2002).
73. P. Langlais, "Introduction VMM vs HMM", Université de Montréal, fiches de cours, (Janvier 2002).
74. <http://zapffe.mat-stat.uit.no/bi/hmm>
75. L. Buniét, «Traitement automatique de la parole en milieu bruite : étude de modèles connexionnistes statiques et dynamiques», Thèse de Doctorat de l'Université Henri Poincaré - Nancy 1, spécialité informatique, (1997).
76. C. Savariaux, P. Perrier, J. Lebeau, G. Magana et C.D. Pattoret « Production de parole après traitement de cancers de la cavité endobuccale», Institut de la communication parlée - INPG & Université Stendhal, (2000).
77. M. A. Mokhtar and A. Z. El-Abdin, «A model for the acoustic Phonetic structure of Arabic Language using a single Ergodic Hidden Markov Model», Electrical Engineering Faculty of Engineering, Alexandria University, Alexandria, Egypt, (1996).
78. M. Baalbaki and R. Baalabaki « Al mawrid el wasit , English-Arabic ; Arabic-English », Dar el ilm lilmlayin, 1996.
79. Barbara Resch, «Hidden Markov Models, A Tutorial for the Course Computational Intelligence », Signal Processing and Speech Communication Laboratory, USA, (2005).