



Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal et Communications



Thèse de Doctorat en Sciences Spécialité Electronique

Présentée par

Mr Khaled BAAZI

Magister en Electronique CRSTDLA - Alger

Thème

Classification Automatique de la Parole Pathologique par la Méthode des Arbres de Décisions chez des Locuteurs Présentant une Fente Palatine ou Labio-Palatine

Soutenue publiquement, le : 14/07/2022

Devant le jury composé de :

Présidente	Mme HAMAMI Latifa	Professeur	ENP - Alger
Rapporteur	Mme GUERTI Mhania	Professeur	ENP - Alger
Examineurs	Mme BENBLIDIA Nadjia	Professeur	USD. Blida
	Mme SAYOUD Siham	Professeur	USTHB
	Mr DJERADI Amar	Professeur	USTHB
	Mr LARBES Chérif	Professeur	ENP - Alger

ENP 2022



Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal et Communications



Thèse de Doctorat en Sciences Spécialité Electronique

Présentée par

Mr Khaled BAAZI

Magister en Electronique CRSTDLA - Alger

Thème

Classification Automatique de la Parole Pathologique par la Méthode des Arbres de Décisions chez des Locuteurs Présentant une Fente Palatine ou Labio-Palatine

Soutenue publiquement, le : 14/07/2022

Devant le jury composé de :

Présidente	Mme HAMAMI Latifa	Professeur	ENP - Alger
Rapporteur	Mme GUERTI Mhania	Professeur	ENP - Alger
Examineurs	Mme BENBLIDIA Nadjia	Professeur	USD. Blida
	Mme SAYOUD Siham	Professeur	USTHB
	Mr DJERADI Amar	Professeur	USTHB
	Mr LARBES Chérif	Professeur	ENP - Alger

ENP 2022

ملخص

يتعلق عملنا بالتصنيف التلقائي للكلام المضطرب (PP) من خلال طريقة أشجار القرار (AD) التي تقع ضمن الإطار العام للعمل الذي يهدف إلى تكييف أدوات معالجة الكلام التلقائية من أجل التقييم الموضوعي للكلام المضطرب. الهدف المطلوب هو التوصيف التلقائي للكلام المضطرب الذي ينتجه الأشخاص المصابون بالشق الحنكي ، بهدف استخدامه في تشخيص وتقييم وإعادة تأهيل هؤلاء المرضى في المستشفيات الجزائرية. لقد طورنا مدونة مؤلفة من كلمات تشتمل على الفونيم المفخم [t] مدمج في 1134 ملفاً صوتياً مع الحركة الطويلة [ā] وفي 1167 ملفاً صوتياً مع الحركة القصيرة [a]. تم تسجيل هذه المدونة من طرف ثمانية وعشرين (28) شخص مصاب بالشق الحنكي ± الشق الشفوي ومجموعة من مائة و ثلاثة (103) شخص سليم النطق. في البداية، أجرينا تحليلاً صوتياً لاستخراج الميزات الفيزيائية ذات الصلة F_0 و F_1 و F_2 و F_3 و E_0 و VOT و الممدد [CV] و [V] للحركة اللاحقة ، و Jitter درجة اضطراب التواتر (F_0) و درجة اضطراب الشدة Shimmer و HNR. ثم تم تطبيق نظام AD على أساس خوارزمية C4.5 لتمييز الكلام المضطرب (PP) من الكلام العادي (PN) ، النتائج التي تم الحصول عليها تظهر لنا أن أنظمة AD تتكيف بشكل جيد في التعرف وتصنيف الكلام المضطرب. سمحت لنا هذه الطريقة بالحصول على معدل تعرف ملموس TR % من PP مقارنة ب PN ، يقدر ب 83% في حالة الحركات الطويلة VL و 75% في حالة الحركات القصيرة VC. بإمكان عملنا المساهمة في إجراء التشخيصات التلقائية ، وإنشاء أنظمة خبيرة تؤدي إلى معدلات ملحوظة في تحديد الاضطرابات الصوتية والمساعدة في التدريب و في العلاج على النطق.

الكلمات المفتاحية: الشق الحنكي ، الكلام المضطرب ، التصنيف ، شجرة القرار AD ، الخوارزمية C4.5

Abstract

Our work concerns the automatic classification of Pathological Speech (PP) by the Decision Trees (AD) method which falls within the general framework of work aimed at adapting automatic speech processing tools for the objective evaluation of pathological speech. The targeted goal is the automatic characterization of pathological speech produced by people with facial clefts, with a view to its use in the diagnosis, evaluation and rehabilitation of these patients in an Algerian hospital environment. We have developed a corpus composed of words comprising the emphatic phoneme [t] combined in 1134 sound files with the Long Vowel (VL) [ā] and in 1167 sound files with the Short Vowel (VC) [a]. This corpus was recorded by twenty-eight (28) speakers operated on for cleft palate or lip-palate and a control group of one hundred and three (103) healthy speakers. Beforehand, we performed an acoustic analysis in order to extract the relevant features (F_0 , F_1 , F_2 and F_3 , E_0 , VOT, the durations [CV] and [V] of the subsequent vowel, the degree of disturbance of F_0 (Jitter), the degree of intensity disturbance (Shimmer) and the HNR (Harmonics to Noise Ratio). Then, an AD based on the C4.5 algorithm was applied in order to discriminate PP from Normal Speech (PN). The results obtained show us that the AD adapt well to the recognition and the classification of the PP. This method allowed us to have appreciable Recognition Rate TR% of the PP compared to the PN, for the case of VL 83% and 75% for VC. Our work can contribute to the conduct of automatic diagnoses, the establishment of expert systems leading to appreciable rates of identification of vocal anomalies and assistance in teaching in Speech therapy.

Keywords: Cleft palate, Pathological speech, Classification, Decision Tree, Algorithm C4.5.

Résumé

Notre travail concerne la classification automatique de la Parole Pathologique (PP) par la méthode des Arbres de Décisions (AD) qui s'inscrit dans le cadre général des travaux visant à adapter les outils de Traitement Automatique de la Parole pour l'évaluation objective de la PP. Le but ciblé est la caractérisation automatique de la PP produite par des patients porteurs de fentes faciales, en vue de son exploitation dans le diagnostic, l'évaluation et la réhabilitation de ces patients dans un milieu hospitalier algérien. Pour cela, nous avons élaboré un corpus composé de mots comprenant le phonème emphatique [t] combiné dans 1134 fichiers sonores avec la Voyelle Longue (VL) [ā] et dans 1167 fichiers sonores avec la Voyelle Courte (VC) [a]. Ce corpus a été enregistré par vingt-huit locuteurs opérés de fente palatine ou labio-palatine et un groupe témoin de cent trois (103) locuteurs sains. Au préalable, nous avons effectué une analyse acoustique afin d'extraire les traits pertinents (F_0 , F_1 , F_2 et F_3 , E_0 , VOT, les durées [CV] et [V] de la voyelle subséquente, le degré de perturbation de F_0 (Jitter), le degré de perturbation de l'intensité (Shimmer) et le HNR (Harmonics to Noise Ratio). Ensuite, un AD basé sur l'algorithme C4.5 a été appliqué afin de discriminer la PP par rapport à la Parole Normale (PN). Les résultats obtenus nous montrent que les AD s'adaptent bien à la reconnaissance et à la classification des PP. Cette méthode nous a permis d'avoir des Taux de Reconnaissance TR % appréciables des PP par rapport à la PN, pour le cas des [VL] 83 % et 75 % pour les [VC]. Notre travail pourra contribuer à la conduite de diagnostics automatiques, l'établissement de systèmes experts aboutissant à des taux appréciables d'identification des anomalies vocales et l'aide à la rééducation en Orthophonie.

Mots clefs : Fentes palatines, Parole pathologique, Classification, Arbre de Décision, Algorithme C4.5.

DEDICACES

" Louange à Dieu tout puissant, qui m'a permis de voir ce jour tant attendu."

*Si Dieu a mis le paradis sous les pieds des mères, ce n'est pas pour rien. Je dédie ce modeste travail à celle qui m'a donné la vie et s'est sacrifiée pour mon bonheur et ma réussite, à **Ma Mère HADJ HENNI Kheïra El Yakout**, le symbole de la bonté par excellence et l'exemple de dévouement. Aucune hommage ne saurait être assez éloquent pour exprimer ce qu'elle mérite pour tous les sacrifices qu'elle n'a cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Elle a fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Elle a tant souhaité que je parvienne à ce but. J'aurais bien aimé qu'elle soit parmi nous pour qu'ensemble nous partageons ce bonheur.*

***Maman**, ce travail est le résultat de ta prière et ta bénédiction ... Que Dieu, le miséricordieux, t'accorde son éternel paradis.*

*A **Mon Père DJILALI**, à qui je dois ce que je suis et ce que je serais, école de mon enfance et qui a été mon ombre durant toutes mes années d'études... Vous n'avez jamais cessé de déployer tous vos efforts afin de subvenir à mes besoins, m'encourager et m'aider à choisir le chemin de la réussite. Permettez-moi de vous exprimer mon grand amour, mon attachement et ma plus haute considération pour votre personne. Je suis très fière d'être votre fils et de pouvoir enfin réaliser, ce que vous avez tant espéré et attendu de moi... Ce modeste travail est le fruit de tous les sacrifices que vous avez consentis pour mon éducation et ma formation. Je vous aime papa et j'implore le Dieu, tout puissant, pour qu'il vous procure longévité et bonne santé.*

*A **Ma chère femme MEDINA**, qui a eu la grande patience de nous supporter, moi et mes humeurs, tout au long de mon travail de recherche, pour tout l'encouragement, le respect et l'amour que tu m'as offert, Je te dédis ce travail, qui n'aurait pas pu être achevé sans ton éternel soutien et optimisme. Tu es un modèle d'honnêteté et de loyauté. J'espère te combler et te rendre heureuse.*

*A **mon ami de toujours Lakhder** qui m'a toujours écouté et suivi avec l'attention dont j'avais besoin. Il a été là dans les moments de doute et a su me conseiller. Je lui témoigne toute ma gratitude.*

*A **Toute ma famille**, mes très chères sœurs *Sabah* et *Yasmina*, mes frères *Benyoucef*, *Hamza* et *Mahdi*, mes chers enfants *Malik*, *Melina*, *Amir (Mbappé)* et *Farouha*... qui m'ont aidé à garder le moral et l'énergie lorsqu'ils menaçaient de chuter. Aucun langage ne saurait exprimer mon respect et ma considération pour votre soutien et encouragements. Je vous dédie ce travail en reconnaissance de l'amour que vous m'offrez quotidiennement et votre bonté exceptionnelle. Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur.*

REMERCIEMENTS

A travers ces quelques lignes, je souhaite remercier toutes les personnes qui ont contribué à l'aboutissement de ce travail.

Tout d'abord, je tiens à remercier Mme GUERTI Mhania, ma directrice de thèse, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, sans qui cette présente thèse n'aurait pas vu le jour. Un grand merci pour vos nombreux conseils et encouragements prodigués tout au long de la réalisation de ce travail. Vos critiques étaient constructives.

Comme je tiens vivement à remercier Mme HAMAMI Latifa, Professeur au Département d'Electronique, Ecole Nationale Polytechnique d'Alger, pour avoir accepté de présider mon jury de thèse. Je suis très honoré de votre présence. Veuillez madame trouver, dans ce modeste travail l'expression de ma haute considération, de ma sincère reconnaissance et de mon profond respect.

J'adresse également mes remerciements aux membres du jury, Mme BENBLIDIA Nadjia, Professeur à l'université Saâd Dahleb - Blida, Mme SAYOUD Siham Professeur à l'USTHB, Mr DJERADI Amar Professeur à l'USTHB et Mr LARBES Chérif, Professeur à l'ENP, qui ont accepté de donner de leur temps pour lire et évaluer ce travail. Honorables Professeurs, je vous prie d'accepter l'expression de ma profonde gratitude et de ma sincère considération. Vous me faites un grand honneur en acceptant d'être membres de jury de ma thèse malgré vos multiples occupations.

Mes remerciements vont tout naturellement aux membres du service Chirurgie Infantile de l'Hôpital de Beni-Messous. Au Professeur Djaffar HANTALA, qui m'a autorisé d'accéder aux dossiers des patients, me permettant d'affiner mes analyses. Aux orthophonistes qui ont accepté ma présence lors de leurs consultations. Aussi, elles ont pris du temps pour m'expliquer certaines spécificités propres à leur travail, tels que les tests utilisés pour déceler certains troubles de la parole, et m'ont aidé à comprendre certaines techniques de rééducation.

Au Docteur Samir BENTEMRI, Maître Assistant en Maxillo-faciale à l'Hôpital de Mustapha Bacha, qui m'a accueilli dans son service lors des consultations de la face et m'a permis d'effectuer les enregistrements avec ses patients.

Je remercie de tout mon cœur tous les patients et les sujets témoins qui ont accepté d'être enregistrés.

Je remercie tout particulièrement Docteur Abd El Hak GASMI pour ses idées et propositions qui ont enrichi nos discussions et apporté un soutien considérable au présent travail.

Ce travail n'aurait pas abouti sans l'aide de toute personne ayant contribué de près ou de loin à sa réalisation, raison de notre gratitude. Enfin, à toute personne qui a manifesté un esprit de soutien à notre égard au cours de la réalisation de ce travail, nous disons merci.

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABREVIATIONS

Introduction générale.....	14
Chapitre 1 : Généralités Sur La Parole.....	18
1.1 Introduction.....	19
1.2 L'appareil phonatoire humain.....	19
1.2.1 La soufflerie.....	20
1.2.2 Le vibreur.....	20
1.2.3 Le pharynx.....	21
1.2.4 Le système articulateur.....	22
1.3 Caractéristiques articulatoires et acoustiques de la parole.....	23
1.3.1 La fréquence fondamentale.....	23
1.3.2 Les formants.....	25
1.3.3 La Durée.....	27
1.3.4 L'énergie.....	28
1.4 Evolution du système phonatoire en fonction de l'âge.....	29
1.4.1 Appareil respiratoire.....	29
1.4.2 Larynx.....	29
1.4.3 Cordes vocales.....	29
1.4.4 Conduit vocal.....	30
1.5 Production des sons du langage.....	31
1.5.1 Production des voyelles.....	32
1.5.2 Production des consonnes.....	34
1.5.2.1 Les Articulateurs.....	34
1.5.2.2 Les Lieux D'articulation.....	35
1.5.2.3 Le Mode D'articulation.....	35
1.5.2.3.1 Les occlusives.....	36
1.5.2.3.2 Les fricatives.....	37
1.5.2.3.3 Les consonnes sonantes.....	37
1.5.3 Semi-voyelles ou semi-consonnes.....	38
1.6 Généralités sur le système phonétique de l'Arabe Standard.....	39
1.6.1 Transcription Orthographique Phonétique (TOP).....	39
1.6.2 Les sons de l'Arabe Standard.....	41
1.6.2.1 Consonnes emphatiques.....	41
1.6.2.2 Gémination.....	42
1.6.2.3 Vocalisme de l'Arabe Standard.....	42
1.7 Conclusion.....	43
Chapitre 2 : Traitement Automatique De La Parole Pathologique.....	44
2.1 Introduction.....	45
2.2 Parole pathologique.....	45
2.3 La fente palatine.....	46
2.3.1 Définition de la fente palatine.....	48
2.3.2 Anatomie du palais normal.....	50
2.3.2.1 La voûte palatine.....	50
2.3.2.2 Le voile du palais.....	50

2.3.3	Principales parties du palais déformées par la fente palatine.....	51
2.3.3.1	Embryologie du Palais Primaire	52
2.3.3.2	Embryologie du Palais Secondaire	52
2.4	Les Différents Types De Fentes	53
2.4.1	Classification de V.VEAU	53
2.4.1.1	Division simple du voile.....	53
2.4.1.2	Division du voile et de la voûte sans bec-de-lièvre	53
2.4.1.3	Division du voile et de la voûte associée à une fente labio-alvéolaire unilatérale ..	53
2.4.1.4	Division du voile et de la voûte associée à une fente labio-alvéolaire bilatérale totale	54
2.4.2	Classification internationale des fentes.....	54
2.4.2.1	Fentes intéressant le palais primaire	54
2.4.2.2	Palais secondaire (voûte palatine osseuse et voile)	54
2.4.2.3	Association palais primaire et palais secondaire	55
2.5	Prise en charge des différentes fentes	56
2.5.1	Interventions chirurgicales.....	56
2.5.1.1	Traitement primaire	56
2.5.1.2	Traitement secondaire et séquelles	57
2.5.2	Prise en charge orthophonique	58
2.5.3	Prise en charge psychologique	58
2.5.4	Les conséquences les plus fréquentes des fentes sur l'articulation et la phonation.....	58
2.6	Paramètres acoustiques de la parole pathologique.....	60
2.6.1	Perturbation de F_0 ou Jitter.....	60
2.6.2	Perturbation de l'intensité ou Shimmer	61
2.6.3	Rapport Harmoniques/Bruit (HNR).....	62
2.6.4	Taux de Passage par Zéro (TPZ).....	63
2.6.5	Le Temps Maximum de Phonation (TMP)	64
2.7	Conclusion.....	64
Chapitre 3 : Analyse Acoustique de la Parole Pathologique.....		65
3.1	Introduction.....	66
3.2	Analyse acoustique de la parole pathologique	66
3.2.1	Corpus choisis	67
3.2.1.1	Description du phonème emphatique [t]	68
3.2.2	Conditions d'enregistrement	68
3.2.3	Matériel utilisé	69
3.2.4	Enregistrements.....	69
3.2.4.1	Période d'enfance	70
3.2.4.2	Période de la puberté	70
3.2.4.3	Age d'adulte	71
3.3	Mesure et analyse acoustique des enregistrements	71
3.3.1	Mesures et analyses des durées	72
3.3.2	Mesures et analyses formantiques.....	72
3.3.3	Extraction des paramètres acoustiques	73
3.3.3.1	Cas des Locuteurs Pathologiques Non Pris en Charge pour une rééducation orthophonique (LPNPC).....	73
3.3.3.1.1	Substitution du phonème [t] par [ʔ].....	75
3.3.3.1.2	Substitution du phonème [t] par [q]	76
3.3.3.1.3	Substitution du phonème [t] par [t]	77
3.3.3.2	Cas de Locuteurs Pathologiques Pris en Charge pour une rééducation orthophonique (LPPC).....	79
3.4	Etudes statistiques des résultats de l'analyse acoustique.....	80
3.4.1	Comparaison des paramètres acoustiques étudiés des locuteurs sains et pathologiques ..	80

3.4.1.1 Classe d'enfants.....	80
3.4.1.2 Classe d'adolescents.....	82
3.4.1.3 Classe d'adultes.....	84
3.4.2 Performances analytiques des différents paramètres étudiés.....	85
3.4.2.1 Présentation de la courbe roc.....	86
3.4.2.2 Résultats.....	89
3.5 Discussions générales.....	91
3.6 Conclusion.....	92

Chapitre 4 : Application Des Arbres De Décision à La Classification Des Paroles

Pathologiques.....	94
4.1 Introduction.....	95
4.2 Reconnaissance Automatique de la parole (RAP).....	95
4.2.1 K plus proches voisins.....	96
4.2.2 Technique des arbres de décision.....	96
4.2.3 Naïve ou décisions de Bayes.....	96
4.2.4 Hidden Markov Models (HMM).....	97
4.2.5 Réseaux de Neurones Artificiels (RNA).....	97
4.2.6 Programmation génétique.....	97
4.2.7 Machines à Vecteurs Support (ou SVM).....	98
4.3 Les arbres de décision (AD).....	98
4.3.1 Etat de l'art sur les AD.....	99
4.3.2 Définition d'un AD.....	99
4.3.2.1 Types d'arbres.....	101
4.3.2.2 Les arbres de décision dans l'apprentissage automatique et l'exploration des données.....	102
4.3.3 Principe de la construction d'un arbre de décision.....	102
4.3.3.1 Optimisation des nœuds.....	104
4.3.3.1.1 Entropie.....	105
4.3.3.1.2 Gain d'informations.....	105
4.3.3.1.3 Rapport de gain.....	105
4.3.3.1.4 Indice de Gini.....	105
4.3.3.2 Arrêt de l'arborescence.....	106
4.3.3.3 Elagage.....	106
4.3.4 Quelques algorithmes pour la construction d'un arbre de décision.....	107
4.3.4.1 Algorithme ID3.....	107
4.3.4.2 Algorithme C4.5.....	107
4.3.4.3 Algorithme CART.....	108
4.4 Conception et architecture du système de classification élaboré (SCAD).....	108
4.4.1 Corpus utilisé.....	109
4.4.2 Prétraitement du signal.....	110
4.4.2.1 Préaccentuation.....	110
4.4.2.2 Fenêtrage.....	110
4.4.2.3 Différents traitements effectués pour l'extraction des vecteurs acoustiques.....	111
4.4.2.3.1 Détection des voyelles.....	111
4.4.2.3.2 Suivi formantique par analyse prédictive.....	111
4.4.2.3.3 Détection du voisement par la fonction de distance (AMDF).....	114
4.4.3 Algorithme de base C4.5.....	115
4.4.3.1 Phase d'apprentissage.....	117
4.4.3.2 Phase de Classification.....	120
4.4.4 Résultats expérimentaux.....	121
4.5 D'autres cas de classifications.....	123
4.5.1 Résultats.....	124

4.6 Conclusion.....	125
Conclusions Générales et perspectives	126
Références bibliographiques	130

LISTE DES FIGURES

Figure 1-1 : Représentation schématique de l'appareil phonatoire humain [2]	19
Figure 1-2 : Principe de la génération d'un son par l'appareil phonatoire humain [3]	20
Figure 1-3 : Larynx [5].....	21
Figure 1-4 : Ouverture et fermeture des cordes vocales [6]	21
Figure 1-5 : Pharynx [9]	22
Figure 1-6 : Conduit vocal [5].....	22
Figure 1-7 : Schéma fonctionnel de l'appareil phonatoire humain [3]	23
Figure 1-8 : Spectrogramme et la courbe mélodique du mot [adab]	24
Figure 1-9 : Représentation spectrographique des trois voyelles brèves de l'Arabe Standard	26
Figure 1-10 : Durée du phonème [a]	27
Figure 1-11: Spectrogramme du mot : « adab ».	28
Figure 1-12 : Développement des cordes vocales [26]	30
Figure 1-13: Vue en coupe de l'appareil phonatoire lors de la production des voyelles brèves [29]	33
Figure 1-14: Triangle Vocalique du Français du point de vue acoustique et articulatoire [32]	34
Figure 1-15 : Articulateurs	35
Figure 1-16: Barre de voisement permettant de distinguer le [b] du [p]	36
Figure 1-17 : Spectrogramme de la consonne [m] représentant une opposition simple géminée	42
Figure 2-1 : Cas d'une fente palatine	48
Figure 2-2 : Différents types de Fente [47]	49
Figure 2-3: Différents types de Fente palatine [46]	49
Figure 2-4: Fente palatine simple [49]	50
Figure 2-5: Schéma des muscles du voile du palais [50]	51
Figure 2-6 : Les bourgeons faciaux et leur développement [51]	52
Figure 2-7 : Fente palatine simple [53]	53
Figure 2-8 : Division du voile et de la voûte palatine d'après Veau [53]	53
Figure 2-9 : Fente labio-alvéolo-palatine unilatérale totale [53]	54
Figure 2-10: Fente labio-alvéolo-palatine bilatérale totale [53]	54
Figure 2-11 : Fente bilatérale complète du palais primaire (3/3 alvéolaire) [53]	54
Figure 2-12 : Fentes concernant le palais secondaire [53]	55
Figure 2-13 : Association palais primaire et palais secondaire [53]	55
Figure 2-14 : Chirurgie d'une fente palatine [55]	56
Figure 2-15 : Chirurgie de fente labio-palatine bilatérale [46]	56
Figure 2-16 : Exemple d'apériodicité de la fréquence fondamentale	61
Figure 2-17 : Exemple d'instabilité de l'amplitude du signal	62

Figure 3-1: Illustration des mesures du VOT	72
Figure 3-2 : Comparaison entre l'articulation du phonème [t] à gauche et du phonème [ʔ] à droite	75
Figure 3-3 : Substitution du phonème [t] par [ʔ] dans [tā'ira] prononcé par des locutrices âgées de 19 ans	76
Figure 3-4 : Comparaison entre l'articulation du phonème [t] à gauche et du phonème [q] à droite	76
Figure 3-5 : Substitution du phonème [t] par [q] dans [tābība] prononcé par des locuteurs âgés de 29 ans	77
Figure 3-6: Comparaison de l'articulation du phonème [t] à gauche et du phonème [t] à droite	78
Figure 3-7 : Substitution du phonème [t] par [t] dans [tā'ira] prononcé par deux locuteurs âgés de 06 ans	78
Figure 3-8 : Exemple de courbe ROC	87
Figure 3-9 : Courbe ROC calculée pour la durée CV (voyelle courte) pour la classe Enfants-Masculins	88
Figure 4-1: Structure générale d'un Arbre de Décision [100]	101
Figure 4-2 : Arbre de classification	101
Figure 4-3 : Arbre de régression	102
Figure 4-4 : Organigramme de classification automatique de parole Normale / Pathologique	108
Figure 4-5 : Fenêtre de Hamming	111
Figure 4-6: Schéma de principe du Codage Prédicatif Linéair (LPC)	112
Figure 4-7 : F ₁ et F ₂ calculés par la méthode LPC	114
Figure 4-8 : Construction d'un arbre de décision par stratégie TDIDT	118
Figure 4-9 : Calcul des trois premiers Nœuds de l'AD	120

LISTE DES TABLEAUX

Tableau 1-1 : Corrélats Articulatoires, acoustiques et perceptifs [13]	38
Tableau 1-2 : Transcription Orthographique Phonétique des consonnes arabes et leurs lieux d'articulations [35]..	40
Tableau 3-1: Corpus utilisé	68
Tableau 3-2: Présentation des locuteurs pathologiques et sains ayant participé aux enregistrements	71
Tableau 3-3 : Analyse acoustique des enregistrements de la locutrice pathologique N°35	80
Tableau 3-4: Valeurs des paramètres acoustiques de la PP et PN de la classe (Enfant - M).....	81
Tableau 3-5 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Enfant - F).....	81
Tableau 3-6 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adolescent - M)	82
Tableau 3-7 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adolescent - F).....	83
Tableau 3-8 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adultes - M).....	84
Tableau 3-9 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adultes - F)	85
Tableau 3-10 : Interprétation des valeurs de l'AUC [83]	88
Tableau 3-11 : Valeurs AUC des différents paramètres étudiés pour les trois Classes - Cas des VC.....	89
Tableau 3-12 : Valeurs seuils AUC des différents paramètres étudiés selon l'âge et le sexe - Cas des VL	90
Tableau 3-13 : Valeurs seuils des différents paramètres acoustiques étudiés selon l'âge et le sexe (VL et VC)....	91
Tableau 3-14 : Evolution de la fréquence fondamentale F_0 et des formants F_1 , F_2 et F_3 , pour les deux sexes.....	91
Tableau 4-1: Fichiers sonores utilisés lors des deux phases (apprentissage et test)	109
Tableau 4-2 : fichiers sonores utilisés dans l'apprentissage pour les VL	117
Tableau 4-3: fichiers sonores utilisés dans l'apprentissage pour les VC	117
Tableau 4-4 : Calcul des GainRatio pour tous les indicateurs pour choisir l'attribut du nœud Zéro.....	118
Tableau 4-5 : Calcul des Gain Ratio pour tous les indicateurs pour choisir l'attribut du nœud suivant le nœud Zéro "Pitch" selon les valeurs Normaux.	119
Tableau 4-6 : Calcul des GainRatio de tous les indicateurs pour choisir l'attribut du nœud suivant le nœud Zéro "Pitch" selon ces valeurs Pathologiques.	119
Tableau 4-7 : Fichiers sonores utilisés dans le test pour les VL.....	121
Tableau 4-8 : Fichiers sonores utilisés dans le test pour les VC	121
Tableau 4-9 : Classification sans prise en charge des Seuils.....	121
Tableau 4-10 : Classification avec prise en charge des Seuils des classes d'âge et du sexe	122
Tableau 4-11 : Taux de Reconnaissance (TR) des PP sans prise en charge des Seuils.....	122
Tableau 4-12: Taux de Reconnaissance (TR) avec prise en charge des Seuils des classes d'âge et du sexe	122
Tableau 4-13 : Valeurs Seuils des paramètres acoustiques pour les différentes classes étudiées - Cas des VL.....	123
Tableau 4-14 : Classification pour les différentes classes étudiées pour les VC.....	124
Tableau 4-15 : Classification pour les différentes classes étudiées pour les VL.....	124
Tableau 4-16 : Taux de Reconnaissance pour les différentes classes étudiées pour les VC	124
Tableau 4-17 : Taux de Reconnaissance pour les différentes classes étudiées pour les VL	124

LISTE DES ABBREVIATIONS

AD	Arbres de Décision
AID	Automatic Interaction Detection
AMDF	Average Magnitude Difference Function
API	Alphabet Phonétique International
AS	Arabe Standard
AUC	Area Under Curve
CART	Classification and Regression Trees
CV	Consonne - Voyelle
FLP	Fente Labio-Palatine
FP	Fente Palatine
HNR	Harmonics to Noise Ratio
HMM	Hidden Markov Models
ID3	Iterative Dichotomiser
KNN	K- Nearest Neighbor
LP	Locuteur Pathologique
LPC	Linear Predictive Coding
LC	Locuteur de Contrôle
LPPC	Locuteur Pathologique Pris en Charge
LPNPC	Locuteur Pathologique Non Pris en Charge
PP	Parole Pathologique
PN	Parole Normale
RAP	Reconnaissance Automatique de la Parole
ROC	Receiver Operating Characteristic
Se_c	Sensibilité
Sp_c	Spécificité
SVM	Machines à Vecteurs Support
TAP	Traitement Automatique de la Parole
TDIDT	Top Down Induction of Decision Tree
TPZ	Taux de Passage par Zéro

TMP	Temps Maximum de Phonation
TOP	Transcription Orthographique Phonétique
V	Voyelle
VC	Voyelle Courte
VL	Voyelle Longue
VOT	Voice Onset Time
VVO	Vocalic Voiced Onset
VVT	Vocalic Voiced Termination

INTRODUCTION GENERALE

La production de la parole est un acte dont la complexité et la multitude des organes qui y prennent part sont souvent masquées par son caractère naturel et facile. En effet, la production de la parole n'est pas un acte involontaire, mais plutôt moteur planifié et particulièrement complexe qui nécessite une importante coordination ainsi qu'une succession de mouvements faisant intervenir aussi bien les systèmes anatomiques périphériques et neurologiques central.

Face à la complexité des mécanismes mis en œuvre dans la production de la parole, l'évaluation perceptive, quelle que soit sa forme, ne permet d'appréhender tous les dysfonctionnements et de comprendre leur répercussion sur la parole. Pourtant, actuellement elle reste l'évaluation la plus répandue en milieu clinique, malgré son caractère subjectif bien reconnue. Les avancées de ces dernières années en Traitement Automatique de la Parole (TAP) ont ouvert de nouvelles perspectives dans le développement de l'évaluation instrumentale pour la pratique clinique.

Le TAP comprend l'ensemble des techniques permettant de communiquer oralement avec une machine. Il présente un intérêt pratique indéniable. C'est un vaste domaine de recherche très complexe qui nécessite la collaboration des spécialistes dans de nombreux domaines : traitement du signal, théorie de l'information, linguistique, phonétique, acoustique, biologique, Informatique, etc [1].

La Reconnaissance Automatique de la Parole (RAP), un des domaines du TAP, a considérablement progressé ces dernières années. Elle permet à la machine de comprendre et de traiter des informations fournies oralement par un utilisateur humain. Elle consiste à employer des techniques d'appariement afin de comparer une onde sonore à un ensemble d'échantillons, composés généralement de mots mais aussi de phonèmes.

Cette thèse s'inscrit dans le cadre général des travaux qui visent à adapter les outils du TAP pour l'évaluation objective de la parole pathologique. Dans notre étude, nous nous intéressons à la classification automatique de la Parole Pathologique (PP) produite par des locuteurs porteurs de Fente Palatine (FP) ou Labio-Palatine (FLP) par rapport à la Parole Normale (PN).

Par Fente Palatine, nous désignons une pathologie de l'appareil phonatoire, qui concerne la fermeture incomplète du palais entraînant la présence d'une communication entre l'intérieur de la bouche (cavité buccale) et l'intérieur du nez (cavité nasale). C'est une malformation du bas du visage qui apparaît tôt durant le développement embryonnaire. Il y a différents types de fentes qui dépendent d'où se situe le manque de fusion.

Les travaux sur les FP ou FLP sont nombreux, mais ils manquent cruellement dans les études consacrées à la langue arabe. De nombreux chercheurs ont tenté de décrire les déficits

de communication, de locuteurs enfants et d'adultes, porteurs de FP ou FLP. Ainsi, ils relèvent des erreurs articulatoires, telles que celles dues à la sensation réduite liée à l'opération du palais, les insuffisances vélo-pharyngales et les anomalies subsistantes du palais dur.

L'objectif de ce travail est d'observer certains troubles articulatoires liés à cette malformation, d'examiner les réajustements effectués par les locuteurs porteurs de fente. Ainsi, à travers diverses analyses des productions de segments de la parole chez des locuteurs porteurs de fente, nous espérons pouvoir mettre en lumière certains résultats, révélateurs des stratégies employées par ces locuteurs pathologiques pour remédier aux perturbations. Les problèmes phonétiques rencontrés concernent plus particulièrement les consonnes occlusives. N'ayant pas de séparation totale et nette entre la cavité buccale et la cavité nasale, les locuteurs porteurs de fente sont dans l'impossibilité de produire une occlusive, puisque l'air doit être, pour de telles réalisations, retenu avant d'être relâché brusquement. Il nous a paru intéressant d'explorer, dans cette étude, l'impact de cette pathologie sur la production des occlusives en langue arabe, notamment le phonème emphatique [t].

Cette thèse est organisée en quatre chapitres :

- le premier décrit le processus de production de la parole et les différents composants qui y interviennent. Nous présentons l'appareil vocal humain, son fonctionnement et son évolution en fonction de l'âge. Des généralités sur les caractéristiques physico-acoustiques de la parole ont été également décrites. Ensuite, nous exposons brièvement le système phonologique qui caractérise l'Arabe Standard (AS) ;
- le deuxième est consacré à la description de diverses pathologies de la parole. Nous avons donné une attention particulière à la Parole Pathologique (PP) chez des sujets ayant un problème de fente faciale (palatine ou labio-palatine). Dans un premier temps, une présentation détaillée des fentes faciales a été faite ainsi que les différents types de classification proposés dans la littérature, puis nous faisons le point sur l'anatomie de certains articulateurs, recrutés en production de la parole et souvent touchés par les divisions palatines. Ensuite, nous exposons les paramètres acoustiques spécifiques à la caractérisation de la qualité de la PP, que nous avons exploités dans notre étude acoustique, tels que le degré de perturbation de la fréquence fondamentale (Jitter), le degré de perturbation de l'intensité (Shimmer), l'influence du Bruit sur les Harmoniques du signal (Harmonics to Noise Ratio HNR), le Taux de Passage par Zéro (TPZ) et l'énergie ;
- le troisième porte sur l'analyse acoustique de la parole pathologique chez des locuteurs opérés de fente palatine, et ce en fonction de l'âge, du sexe (Homme, Femme) et du

type de pathologie en vue de sa caractérisation. Ainsi, à travers une analyse acoustique des productions de segments de la parole chez ces locuteurs, nous essayerons de trouver les seuils de cette parole pathologique par rapport à la Parole Normale (PN) ;

- le dernier chapitre est consacré dans sa première partie aux généralités sur les systèmes de classification. Nous avons donné, en particulier, une importance aux Arbres de Décision (AD) que nous avons appliqués dans le cadre de notre travail ; la deuxième partie concerne la classification automatique de la PP par rapport à la PN par les AD. Nous avons expliqué d'une manière détaillée l'architecture du système de classification que nous avons élaboré. Les résultats obtenus ont été également exposés et commentés ;
- Nous finissons notre travail par des conclusions générales et perspectives et des références bibliographiques qui sont en rapport avec la thèse.

CHAPITRE 1 : GENERALITES SUR LA PAROLE

1.1 Introduction

Le processus de production de la parole est un mécanisme très complexe qui repose sur une interaction entre le système neurologique et physiologique. Il y a une grande quantité d'organes et de muscles qui entrent dans la production des sons des langues naturelles.

L'étude du mécanisme de la production de la parole humaine consiste donc à décrire, sur un plan physiologique, les mouvements des articulateurs vocaux qui produisent les sons et les mots, sur un plan acoustique, l'étude des sons des mots prononcés.

Chaque son d'une langue, et ce, quelle que soit la langue, met à contribution d'une façon qui lui est propre les mêmes organes phonatoires. Mais, les contraintes qui s'exercent sur le fonctionnement de ces organes sont universelles. C'est pourquoi la connaissance de ce mécanisme peut donc se révéler très utile dans l'apprentissage de la prononciation d'une langue.

Dans ce chapitre, nous décrivons brièvement la physiologie anatomique du mécanisme de production de la parole humaine. Ensuite, nous exposons brièvement le système phonologique qui caractérise l'Arabe Standard.

1.2 L'appareil phonatoire humain

On définit par appareil phonatoire l'ensemble des organes qui nous permettent d'émettre un son. Il peut être décomposé en quatre éléments qui collaborent étroitement pour produire la parole : la soufflerie, le vibreur, les résonateurs et le système articulateur (Figure 1-1). Tous ces éléments sont placés sous la dépendance du système nerveux central, qui en assure la bonne coordination.

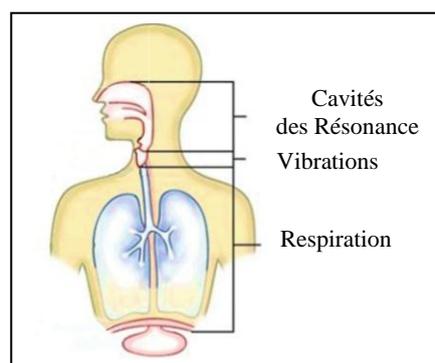


Figure 1-1 : Représentation schématique de l'appareil phonatoire humain [2]

Aucun de ces organes ne peut à lui seul produire un son, ils sont donc obligatoirement complémentaires, par conséquent si l'un d'entre eux dysfonctionne cela peut avoir des conséquences plus ou moins importantes dans l'émission du son.

La production d'un son par l'appareil phonatoire humain repose sur le principe suivant : la soufflerie (appareil respiratoire) crée une pression d'air engendrant au niveau d'un vibreur la création d'un son, qui est ensuite modulé par la partie résonateur du système (Figure 1-2).

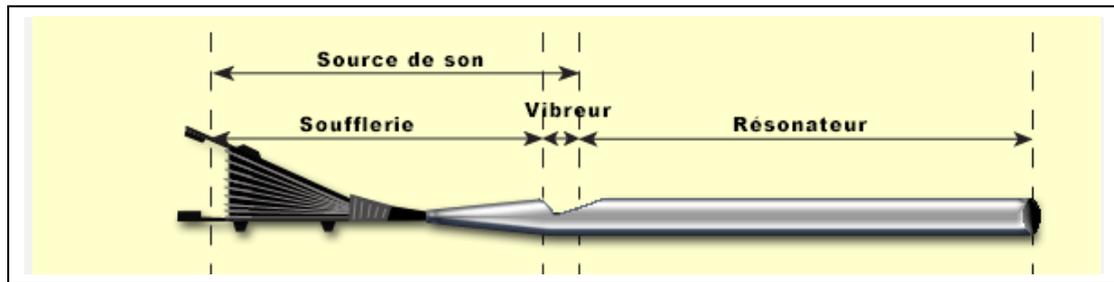


Figure 1-2 : Principe de la génération d'un son par l'appareil phonatoire humain [3]

1.2.1 La soufflerie

Elle consiste en un réservoir d'air stocké dans les poumons. Ces derniers actionnés par les muscles du thorax et de l'abdomen permettent en premier lieu d'alimenter les tissus du corps en oxygène par une alternance de cycles d'inspiration et d'expiration. Essentiellement, le temps expiratoire, alimente et contrôle l'émission sonore. Si la voix est le support acoustique de la parole, le souffle expiratoire est le support aérodynamique de la phonation. Le flux d'air expulsé des poumons se déplace jusqu'à arriver au niveau du larynx où plusieurs cas se présentent.

La difficulté d'évaluation de ce système est liée au fait qu'il n'existe pas une bonne et une mauvaise respiration, mais plutôt une respiration adaptée ou non [4].

1.2.2 Le vibreur

Le vibreur constitué du larynx, est un ensemble de cartilages réunis entre eux par des ligaments et des muscles, que logent les cordes vocales ; il couronne la trachée artère à la manière d'une cheminée. Les cordes vocales sont constituées de muscles et de ligaments recouverts d'une muqueuse. L'ensemble est attaché en avant au cartilage thyroïde et en arrière aux deux cartilages aryténoïdes (Figure 1-3).

L'espace compris entre les cordes vocales et la paroi postérieure du larynx est la glotte, elle joue le rôle de valve entre la trachée et le conduit vocal. Lorsqu'elle est largement ouverte, elle permet la respiration et aucun son n'est engendré ; lorsqu'elle n'est que faiblement ouverte, elle produit le chuchotement ; et quand elle est complètement fermée, il y a phonation.

Un petit opercule, l'épiglotte, vient fermer le larynx au moment de la déglutition des aliments. Il assure le rôle d'aiguillage entre les aliments ingérés (dirigés vers le tube digestif) et l'air inspiré (dirigé vers les poumons).

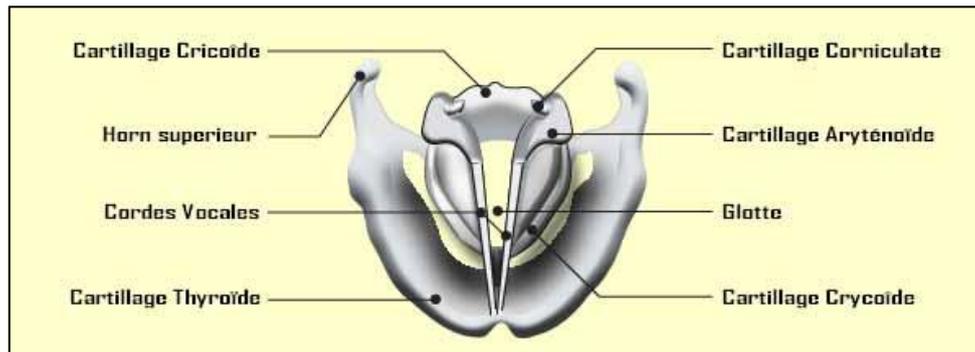


Figure 1-3 : Larynx [5]

Les mouvements des cartilages arythénoïdes permettent de modifier à la fois l'ouverture de la glotte et la longueur des cordes vocales en fonction de l'action désirée, ce qui détermine la hauteur du son fondamental (Figure 1-4).

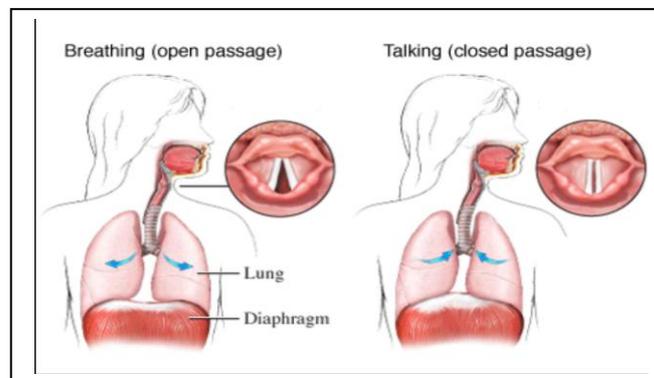


Figure 1-4 : Ouverture et fermeture des cordes vocales [6]

Les vibrations des cordes vocales donnent naissance à tous les sons voisés ou sonores. Pendant la respiration, les cordes vocales sont écartées et laissent libre passage de l'air. Pour que la phonation puisse s'effectuer, c'est-à-dire pour que les cordes vocales puissent vibrer, elles doivent s'accoler.

Le larynx n'est pas fixe dans le cou : il se déplace de haut en bas quand on parle. Il s'élève pour les sons aigus et s'abaisse pour les sons graves [7].

1.2.3 Le pharynx

Le pharynx est un conduit musculo-membraneux étendu de la base du crâne en haut à la sixième vertèbre cervicale en bas [8]. C'est un carrefour aéro-digestif situé entre la racine de la langue et les parois de l'œsophage, il relie les fosses nasales (la cavité nasale) au larynx (voie respiratoire) et la bouche (la cavité buccale) à l'œsophage (voie digestive) représentant ainsi une cavité de résonance (Figure 1-5).

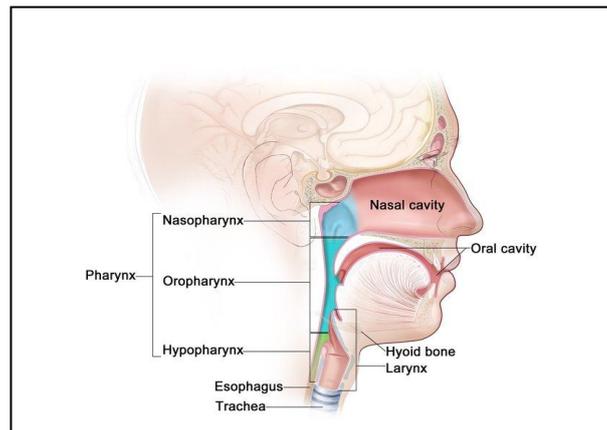


Figure 1-5 : Pharynx [9]

Il joue un rôle au cours de la respiration et de la phonation, et un rôle dans la déglutition. Par ailleurs, le pharynx intervient dans la physiologie de l'oreille par l'intermédiaire de la trompe d'Eustache [10], qui le met en communication avec l'oreille moyenne au niveau de la caisse du tympan.

1.2.4 Le système articulateur

Le système articulateur comprend (Figure 1-6) :

- La bouche dont les parties les plus importantes sont : le palais, le maxillaire inférieur, la langue, les dents et les lèvres;
- Les fausses nasales, pouvant être ouvertes ou fermées suivant la position qu'occupe le voile du palais, membrane qu'on aperçoit au fond de la bouche et dont l'extrémité inférieure est un petit appendice : la luette.

La position du voile du palais (appelé également Velum) détermine si l'air expiré durant la phonation s'échappe uniquement par la bouche (Velum relevé), par la bouche et le nez (vélum partiellement abaissé) ou uniquement par le nez (Velum totalement abaissé). Dans ce dernier cas, les sons produits sont nasalisés.

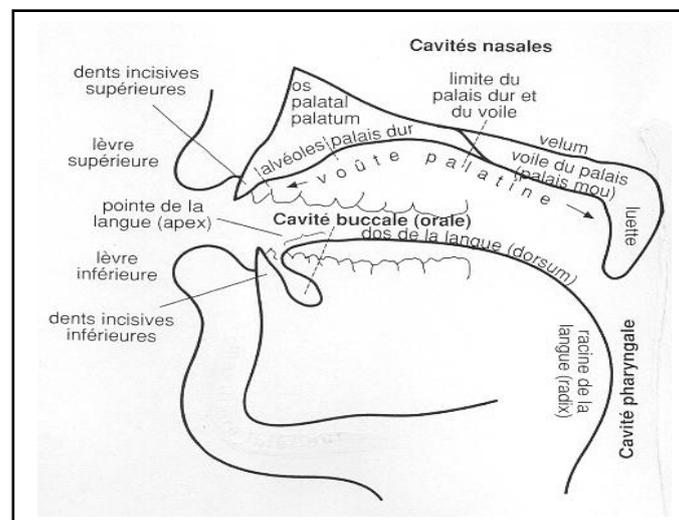


Figure 1-6 : Conduit vocal [5]

1.3 Caractéristiques articulatoires et acoustiques de la parole

La parole est l'une des formes les plus importantes de la communication humaine. Physiquement c'est un phénomène vibratoire résultant de deux composantes :

- *La source d'excitation* : elle débute, dans la phonation, par l'action des poumons. Ceux-ci libèrent, au rythme qui est sous le contrôle volontaire du locuteur, un souffle d'air, lequel passe par la trachée artère et traverse le larynx (Figure 1-7) ;
- *Un système résonnant*, appelé conduit vocal, s'étend de la glotte (larynx) aux lèvres. Il est composé de quatre cavités : pharyngale, buccale, labiale et nasale (Figure 1-6). Ces cavités ont des parois immobiles (parois nasales, palais osseux, mâchoire supérieure, dents) et des structures mobiles précédemment décrites (parois pharyngées, voile du palais, langue, mâchoire inférieure, lèvres). Toute cavité présente une fréquence caractéristique ou un son propre, qui peut être un son fondamental ou un de ses harmoniques, telle qu'elle résonne sous l'influence de ce son et le renforce. La fréquence caractéristique varie selon la forme et le volume des cavités, les plus grandes ayant en général un son propre plus grave que les plus petites.

Lorsqu'un locuteur génère un message oral, il procède à une série d'opérations mettant en jeu tous les niveaux linguistiques (lexical, syntaxique, sémantique, pragmatique, morphologique, phonologique et phonétique) : il y a donc, à partir du sens, production d'un ensemble de commandes qui, du système nerveux central jusqu'aux muscles, via le système nerveux périphérique, va permettre de piloter les évolutions du conduit vocal [11].

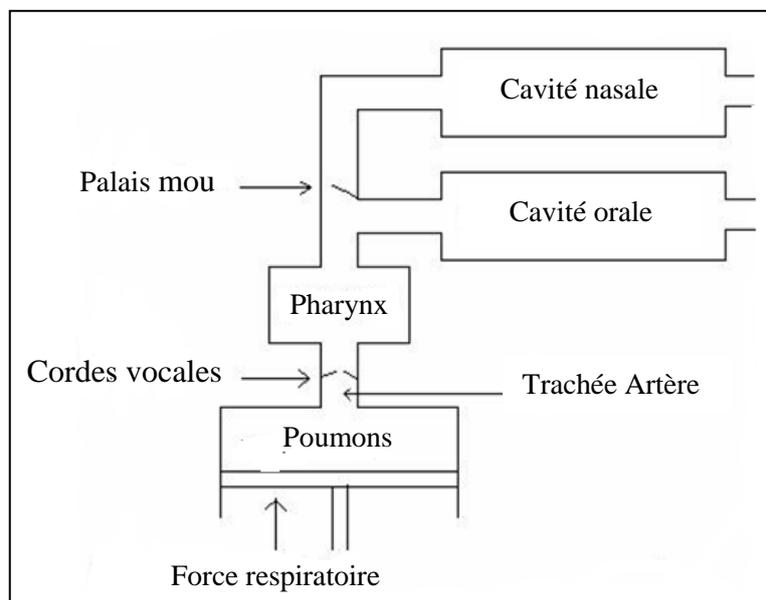


Figure 1-7 : Schéma fonctionnel de l'appareil phonatoire humain [3]

1.3.1 La fréquence fondamentale

L'accolement des cordes vocales constitue un obstacle à l'écoulement normal du souffle d'air venu des poumons, ce qui a comme conséquence de créer aussitôt une pression sous la glotte

(fermée). Lorsque cette pression, dite sous-glottique, atteint une force suffisamment grande, les cordes vocales cèdent et laissent passer le flux d'air. Après le passage de cet air, les cordes vocales reprennent rapidement leur position d'origine sous l'action de leur masse, de leur tension et de leur élasticité. Ce cycle de pression-relâchement recommence aussitôt que la pression sous-glottique s'est reconstituée. C'est en raison de l'extraordinaire rapidité des ouvertures et des fermetures qu'on dit que les cordes vocales vibrent. Ce qu'on appelle la fréquence d'un son, c'est le nombre de ces vibrations à la seconde, donc le nombre de mouvements d'ouverture-fermeture de la glotte. Si, par exemple, les cordes vocales s'ouvrent et se referment 220 fois par seconde, on produira alors un son fréquentiel de 220 Hertz (Hz).

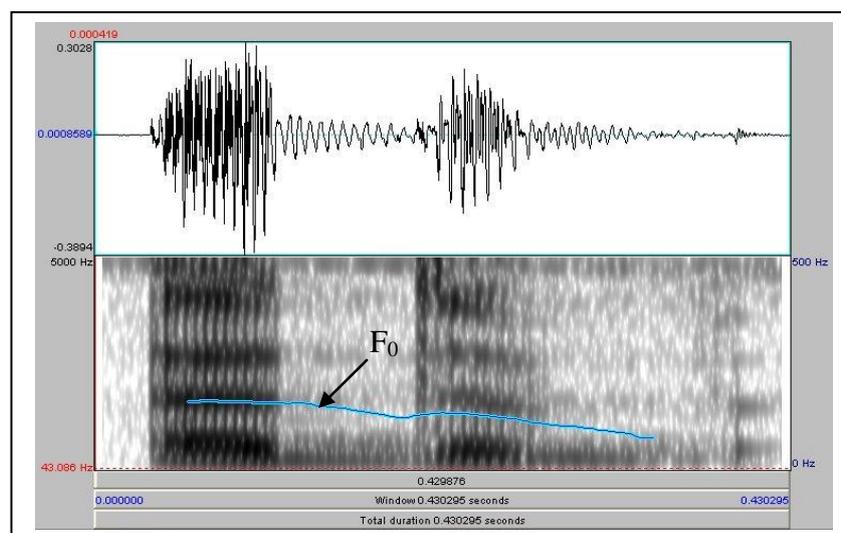


Figure 1-8 : Spectrogramme et la courbe mélodique du mot [adab]

La fréquence fondamentale ou F_0 est le corrélat acoustique de la fréquence laryngienne, due à la mise en vibrations des cordes vocales. Du point de vue perceptif, nous parlons de hauteur.

La fréquence fondamentale de la voix est propre à chaque individu. Elle est fonction de différents paramètres physiologiques tel que le volume et la masse de la glotte, la section de la trachée, sa longueur, etc.

Il a été démontré qu'une échelle logarithmique traduit mieux qu'une échelle linéaire la perception des hauteurs que nous pouvons avoir.

La période T_0 est comprise entre 2 et 17 ms, soit une fréquence F_0 située approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants [12].

La fréquence fondamentale n'est calculée que sur des parties voisées de la parole, c'est-à-dire principalement les voyelles, mais aussi quelques consonnes : ces localisations correspondent à des parties plus stables du signal, il devient possible d'en extraire des pseudo-

périodes. Il existe plusieurs algorithmes pour le calcul de F_0 . Ceux-ci ne donnent pas toujours des résultats identiques.

L'évolution de F_0 est essentiellement influencée par la nature du phonème, par sa position dans la syllabe et par son environnement phonétique immédiat (certains phonèmes, comme les occlusives voisées, contribuent à abaisser F_0 ; d'autres ont tendance à l'augmenter) (Figure 1-8).

Les algorithmes d'extraction du F_0 utilisent généralement une représentation temporelle ou spectrale du signal. Les méthodes temporelles utilisent la similarité du signal d'une période à l'autre pour identifier la période fondamentale T_0 . Il est parfois possible de la repérer manuellement directement sur le signal.

Dans le domaine fréquentiel, les algorithmes utilisent généralement les harmoniques de la fréquence fondamentale pour trouver le F_0 .

Le calcul de F_0 dépend du contexte sonore. Un milieu bruité détériore énormément les résultats des algorithmes de détection, il est donc important de soigner particulièrement les enregistrements destinés aux observations de ce paramètre.

1.3.2 Les formants

La production vocale est le résultat d'une coordination neuro-motrice des muscles de tous les organes impliqués dans la phonation depuis les muscles de la posture et de la respiration jusqu'aux muscles du larynx et de l'appareil articulatoire pharyngo-bucco-labial.

Avant d'être rayonné au niveau des lèvres, le signal acoustique se propage à travers le conduit vocal dans lequel il est filtré. En effet les cavités supraglottiques possèdent des fréquences de résonance qui renforcent certaines régions du spectre des sources excitatrices (Figure 1-6). Les maxima de la courbe de réponse en fréquence du conduit vocal sont appelés formants [13].

Le premier harmonique présentant la plus forte amplification (maxima d'énergie) sera assimilé au premier formant et noté F_1 , le second harmonique au deuxième formant et noté F_2 , etc. Le nombre des formants, selon les caractéristiques du résonateur (volume, forme et ouverture), est variable : d'un seul à une infinité (théoriquement). Néanmoins, du point de vue perceptif, seuls quelques-uns d'entre eux jouent un rôle central au niveau de la parole. Généralement, les formants F_1 , F_2 et F_3 sont bien marqués sur un spectrogramme ; les formants F_4 et F_5 sont plus difficilement visibles.

Les formants (F_1 , F_2 , F_3) permettent principalement de distinguer les voyelles entre elles (Figure 1-9). Ainsi, la forme, la dimension ainsi que la constitution physiologique des

résonateurs seront autant d'éléments qui détermineront l'ensemble des harmoniques mises en valeur.

Il est d'ailleurs important de noter que plus la taille du résonateur est grande, plus sa fréquence de résonance est basse, et par conséquent, plus la valeur du formant sera basse [14].

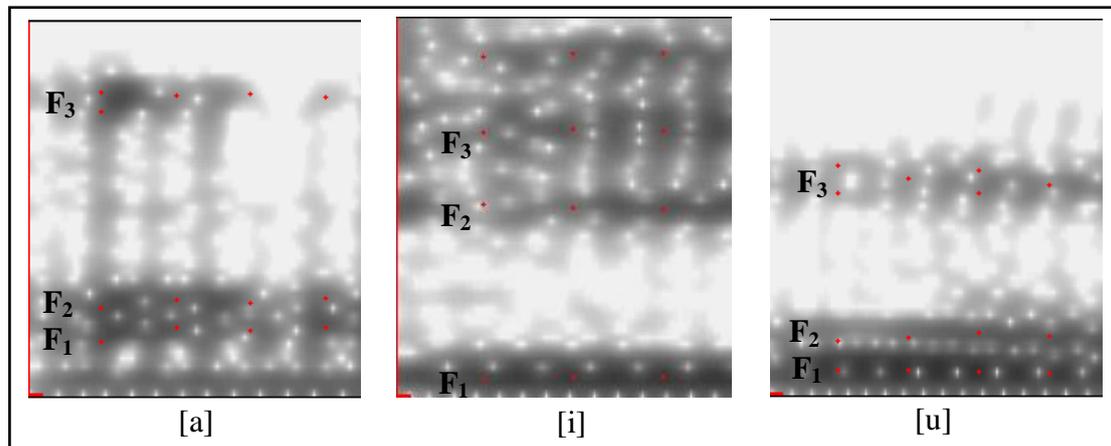


Figure 1-9 : Représentation spectrographique des trois voyelles brèves de l'Arabe Standard

Les formants peuvent être modifiés par quelques manœuvres articulatoires. La première règle est que tous les formants diminuent uniformément au fur et à mesure que la longueur du tube s'allonge.

Ces modifications de longueur peuvent être obtenues par un abaissement du larynx ainsi que par une projection ou au contraire une rétraction des lèvres. Ces modifications entraînent donc un glissement global des fréquences sans modification de l'intervalle entre les formants, donc sans modification de l'identification de la voyelle. La seconde façon de modifier les voyelles est d'arrondir les lèvres, ce qui correspond schématiquement à obturer légèrement la bouche comme font parfois certains instrumentistes à vent. Cette « couverture » aboutit à l'équivalent acoustique de l'augmentation de la longueur du tube et donc à un glissement des fréquences vers le grave. La combinaison entre la hauteur du larynx et la position et la forme des lèvres permet effectivement d'éclaircir ou d'assombrir le timbre de la voix. L'abaissement de la mâchoire peut également être utilisé. En particulier, F_1 peut être abaissé par un abaissement de la mâchoire. Par ailleurs, la contraction de la bouche abaisse F_1 et élève F_2 , ce qui donne une voyelle au spectre plus large avec un étalement de la puissance acoustique comme dans [i]. Inversement, la contraction du pharynx élève F_1 et diminue F_2 , ce qui rend la voyelle plus compacte comme [a].

1.3.3 La Durée

Parmi les paramètres classiques utilisés dans la recherche en prosodie, la durée est la plus difficile à préciser, car elle n'est pas directement associable à un corrélat biologique du système phonatoire [15].

Du point de vue de la production, la phonation est limitée par des contraintes physiologiques, notamment en ce qui concerne les consonnes. En effet, l'appareil vocal n'est pas capable de produire des plosives qui ont une durée importante. L'effort produit lors de l'articulation ne peut être infini, la séquence est toujours la même : effort et relâchement. Cet effort et ses variations sont véhiculés à travers le signal, ils permettent une segmentation temporelle de la parole en unités [16].

On associe à chaque segment de parole une durée. Avant de mesurer les durées, il faut cerner correctement les unités à mesurer. On distingue les durées des unités phonétiques, syllabes, phonèmes ou même distance entre voyelles et les durées des pauses (les temps de silence) (Figure 1-10).

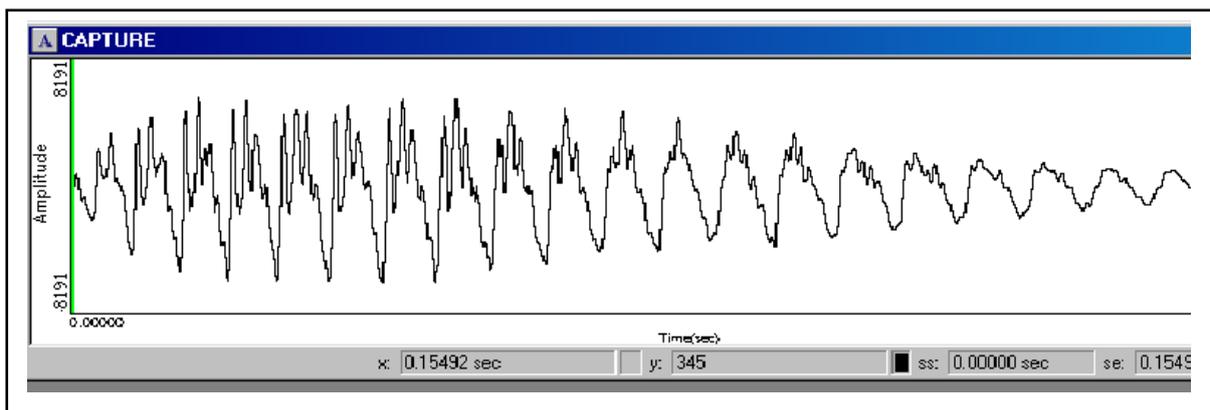


Figure 1-10 : Durée du phonème [a]

Comme les autres paramètres, les durées des unités choisies sont largement dépendantes du locuteur et du débit de parole. Toute mesure ne peut donner un modèle absolu pour une application brute des relevés.

La durée des différentes unités est le phénomène central pour la prosodie. En effet, chaque variation de fréquence fondamentale ou d'intensité s'établit sur un certain laps de temps, durée mesurable. Etudier l'organisation temporelle de la parole est incontournable. Etudier la durée c'est observer et modéliser les durées d'unités bien déterminées. Pour pouvoir comparer ou tirer des renseignements d'études passées, il est important de bien cerner à chaque fois l'unité choisie.

La durée est le paramètre le plus difficile à préciser car rien n'indique comment le système de contrôle, de production ou de perception de parole, mesure le temps. Les indices de durée

classiques supposent généralement la donnée d'une segmentation, des frontières des unités dont on désire mesurer la durée : par exemple d'une unité par le nombre de trames qui séparent ses frontières de début et de fin.

Il existe deux types de durées connues :

- Observées, se rapportant au mécanisme de production de la parole [17], elles correspondent à la mesure objective du temps de l'activation et de la coordination des organes de phonation [18] ;
- Perçues correspondent à la composante perceptive. Du point de vue de la production, la phonation est limitée par des contraintes physiologiques. Ce fait est observé notamment lors de la production des occlusives où la durée de réalisation ne peut être continue.

1.3.4 L'énergie

L'énergie de la parole est un paramètre couramment utilisé dans les systèmes de reconnaissance (Figure 1-11). C'est le paramètre le plus facile à calculer. L'énergie d'un signal échantillonné $(s_t)_{t=1,T}$ à support fini est définie par :

$$E = \sum_{t=1}^T s_t^2 \quad (1.1)$$

Etant donné sa dynamique et pour respecter l'échelle perceptive, elle est généralement exprimée en décibels :

$$E_{dB} = 10 \times \log_{10} \left(\sum_{t=1}^T s_t^2 \right) \quad (1.2)$$

Pour un signal échantillonné quelconque, on calcule l'énergie à court terme en prenant des portions de signal convoluées avec une fenêtre glissante (généralement assez étroite, de l'ordre de 5 à 10 ms).

Visuellement, l'énergie est corrélée à l'enveloppe temporelle du signal [19].

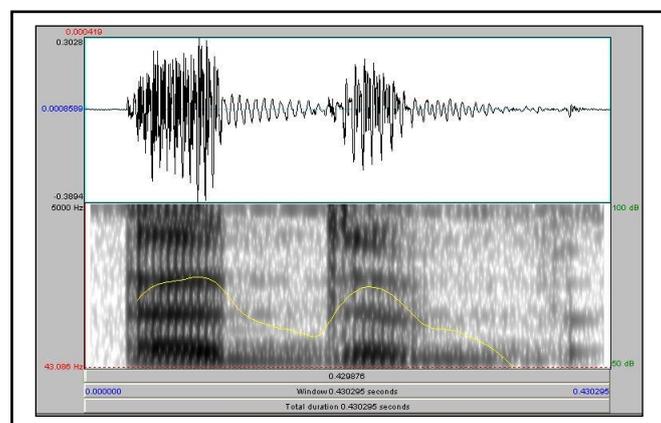


Figure 1-11: Spectrogramme du mot : « adab ».

1.4 Evolution du système phonatoire en fonction de l'âge

L'anatomie subit des évolutions, des adaptations et des ajustements précis, le système vocal et le système auditif évoluent pour travailler à l'unisson.

Les caractéristiques de la voix de l'enfant évoluent de la naissance à la puberté en raison des modifications affectant tous les éléments de l'appareil vocal.

La croissance du conduit vocal dépend essentiellement de celle du cerveau et du crâne, de la face et des vertèbres qui s'accompagne de la rotation basicranienne (du sphénoïde et du basi-occipital) [20].

Au cours de l'enfance le larynx et les cavités de résonance évoluent, en taille, bien sûr, mais aussi en position relative et en proportions [21].

Le développement de la voix de l'enfant entre 4 et 12 ans, nous conforte dans l'idée que filles et garçons suivent le même développement mais de façon décalée et non continue [22].

Dans notre étude, il nous a été primordial de connaître ces étapes du développement vocal pour en comprendre les limitations fonctionnelles.

1.4.1 Appareil respiratoire

Le souffle de l'enfant est conditionné par la forme du thorax. D'un point de vue anatomique, l'appareil respiratoire de l'enfant diffère de celui de l'adulte. De nombreuses modifications anatomiques vont avoir lieu tout au long de la croissance de l'enfant, jusqu'à donner à la cage thoracique sa forme adulte et aux poumons leur volume final [23].

Le développement de la cage thoracique et du maximum d'alvéoles pulmonaires atteint la maturation vers 18 - 20 ans [24].

1.4.2 Larynx

La position du larynx varie avec l'âge. À la naissance, le bord inférieur du cricoïde est au niveau des troisième et quatrième vertèbres cervicales, à 2 ans, il est au niveau de C5, à 5 ans, il est au niveau du milieu de C6, enfin à 15 ans il atteint sa place définitive, avec un bord inférieur du cricoïde au niveau du disque C6-C7 [21].

Le larynx descendant dans le cou le pharynx s'allonge et son volume augmente. De plus, la langue se postériorise et s'abaisse, ce qui libère la cavité buccale.

Le larynx augmente 60% pour les garçons et 34% pour les filles.[25]

L'abaissement du fondamental laryngé (son le plus grave) se fait très progressivement, parallèlement à la croissance du larynx et à sa descente dans le cou [23].

1.4.3 Cordes vocales

Les cordes vocales mesurent environ 5 à 8 mm à la naissance [21], avec une partie membraneuse à peu près aussi longue que la partie postérieure cartilagineuse, alors que chez

l'adulte la partie cartilagineuse ne représente qu'environ un tiers ou un quart de la longueur de la corde vocale (Figure 1-12).

Les cordes vocales vont s'accroître en largeur, en épaisseur et en longueur. À 6 ans, la longueur des cordes est de l'ordre de 8 mm aussi bien chez les filles que chez les garçons. À la puberté, sous l'effet des hormones sexuelles androgéniques, le larynx grandit très vite. Cette croissance est plus rapide et plus importante chez le garçon que chez la fille. À l'âge adulte, les cordes vocales mesurent environ 14 à 15 mm chez l'homme et 10 à 11,5 mm chez la femme [21].

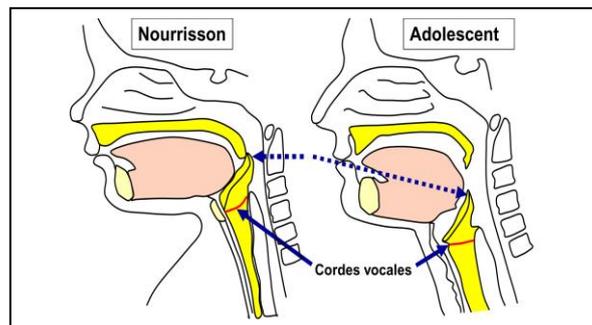


Figure 1-12 : Développement des cordes vocales [26]

Le champ de variation du fondamental laryngé croît de 5 demi-tons chez le nourrisson, à 14 -19 demi-tons chez les garçons de 12 ans et 16 - 22 demi-tons chez les filles de 12 ans.

Elle se modifie au fur et à mesure des modifications anatomiques du larynx. La descente et la croissance du larynx entraînent une diminution du fondamental laryngé moyen. Il est de 450 Hz à la naissance (pour les cris), de 286 Hz à 7 ans et 275 Hz à 8 ans [21].

À la puberté, les modifications de la voix sont plus nettes chez le garçon que chez la fille. La voix des filles baisse d'en moyenne 2,4 demi-tons (une tierce), avec un fondamental moyen de 210 Hz, alors que le fondamental laryngé du garçon va baisser d'en moyenne une octave en 3 à 6 mois, pour atteindre 120 à 130 Hz [21].

La mue s'accompagne souvent de changements brusques de hauteur avec de véritables changements de régime vibratoire au cours d'une même chaîne parlée.

La plupart des troubles vocaux s'accompagnent d'un abaissement du fondamental laryngé.

1.4.4 Conduit vocal

L'étude de l'évolution de la morphologie du conduit vocal est indispensable si l'on veut comprendre l'influence du remodelage du conduit vocal sur les productions acoustiques et les configurations articulatoires.

La croissance du conduit vocal en ontogenèse, loin d'être linéaire, est particulièrement importante durant les premières années de vie, période cruciale pour l'apprentissage de la production de la parole [27]. La longueur totale du conduit vocal double de la naissance à

l'âge adulte (80 mm en moyenne chez le nouveau né, 176 mm chez l'homme adulte et 154 mm chez la femme adulte) [26].

La taille et la conformation des cavités pharyngée, buccale et nasale, qui modifient le son produit par la vibration des cordes vocales, évoluent avec la croissance de la face [20]. la cavité pharyngale évolue de 20 - 30 mm à 1 mois, jusqu'à presque 80 mm en moyenne à l'âge adulte chez les hommes [26]. La cavité orale, quant à elle, évolue plus lentement et plus régulièrement que la cavité pharyngale, qui comprend deux pics de croissance marqués, un durant les premières années de vie et un second à l'adolescence [26].

Le velum a une taille importante à la naissance (environ 25 mm) et croît très peu au cours de l'ontogenèse (28 mm et 33 mm respectivement chez la femme et l'homme adulte). La mandibule, en revanche, a une croissance très régulière et atteint la taille adulte aux alentours de 17 ans [27].

L'évolution de la morphologie et de la taille de ces cavités s'arrête vers 20 ans et permet de fixer la majorité des caractéristiques de la voix qui aura évolué en parallèle [28].

1.5 Production des sons du langage

L'onde glottique subit des modifications majeures lors de son passage dans le conduit vocal dues aux phénomènes de résonance qui dépendent de la conformation des cavités supra-glottiques composant ce conduit.

Le conduit vocal est modulable au sens où sa configuration peut évoluer dans le temps. Ces modifications sont dues principalement à des mécanismes dits articulatoires qui font intervenir tout ou une partie d'organes du conduit vocal comme le pharynx, le voile du palais, la langue, la mâchoire inférieure et les lèvres. Ces derniers par rapprochement ou contact avec des parties fixes du conduit vont permettre :

- d'augmenter ou de réduire le volume d'une cavité ;
- d'augmenter ou de réduire le diamètre du conduit vocal. On parle alors de constriction du conduit et de son degré ;
- de fermer complètement le conduit vocal. On parle alors d'occlusion du conduit.

Ces multiples configurations du conduit vocal vont conduire à la réalisation des différentes productions sonores potentielles pour n'importe quelle langue. Les termes de mode et de lieu d'articulation sont fréquemment utilisés dans la littérature pour décrire et classer ces productions sonores. Ils permettent de caractériser la conformation du conduit vocal lors de la production d'un son ou d'une classe de sons. Le mode définissant le type d'obstruction du conduit phonatoire (c'est la façon avec laquelle l'obstruction est réalisée : totale avec les

occlusives, très resserrée avec les fricatives, etc.) et le lieu d'articulation qui représente le point de rapprochement ou de contact de l'articulateur avec les parties fixes du conduit vocal.

Les sons du langage peuvent être classés d'après les critères suivants (Tableau 1-1) [13] :

- l'opposition sonore-sourde selon que les cordes vocales vibrent ou non ;
- le mode d'articulation qui permet de montrer la façon dont l'air s'écoule à travers le canal respiratoire au moment de la phonation (Tableau 1-1) ;
- le lieu ou point d'articulation : endroit de la constriction maximale du conduit vocal ;
- l'opposition nasale-orale selon que le voile du palais est abaissé ou non.

En fonction des possibilités offertes par l'appareil vocal, chaque langue a adopté un ensemble particulier de sons distinctifs ou phonèmes (Les phonèmes sont les éléments sonores les plus brefs qui permettent de distinguer différents mots). On distingue deux classes liées au mode de production des sons, si le passage de l'air :

- se fait librement à partir de la glotte, on a affaire à une voyelle ;
- est obstrué, à partir de la glotte complètement ou partiellement, en un ou plusieurs endroits, on a affaire à une consonne.

1.5.1 Production des voyelles

Premier type de phonèmes, les voyelles sont caractérisées par un passage libre de l'air et les vibrations des cordes vocales.

Dans la prononciation d'une voyelle, on distingue quatre critères rentrant en jeu dans la configuration variable du conduit vocal :

- la langue se soulève toujours en un point créant ainsi deux résonateurs distincts séparés par un étranglement dont la situation par rapport à la séparation en palais dur et mou, permet de distinguer les voyelles antérieures et postérieures ou vélaires. Le changement de position ayant partiellement une incidence sur le deuxième formant F_2 , les voyelles antérieures seront caractérisées par un F_2 élevé (ex : voyelle [i]) contrairement aux voyelles postérieures qui présenteront un F_2 bas (ex : voyelle [u]) (Figure 1-12) ;
- L'ouverture plus ou moins grande du pavillon pharyngo-buccal (l'aperture de la mandibule) permet de distinguer les voyelles fermées ou ouvertes. Le premier formant F_1 étant déterminé par ce degré d'aperture et la hauteur de la langue, les voyelles ouvertes seront caractérisées par un F_1 élevé (ex : voyelle [a]) dû à une cavité pharyngale très petite suite à l'ouverture maximale de l'aperture et un abaissement de la langue contrairement aux voyelles fermées qui présenteront un F_1 bas (ex : voyelle [i]) dû à une cavité pharyngale très grande (Figure 1-13) ;
- la position des lèvres, soit neutre ou étirée, soit formant la cavité labiale par protrusion et arrondissement. Les voyelles seront qualifiées de labiales/arrondies ou de non labiales/non arrondies/étirées. La formation de la cavité labiale ayant pour conséquence un abaissement des valeurs de formants F_2 et F_3 , les valeurs de F_2 et F_3

des voyelles arrondies (ex : voyelle [y]) sont plus basses que les voyelles étirées (ex : [i]) ;

- Les mouvements du voile modifient les caractéristiques de la cavité buccale ainsi que du formant correspondant. Lorsque le voile s'abaisse légèrement, le résonateur rhino-sinusien entre en jeu et produit ainsi la nasalisation des voyelles. Les voyelles seront qualifiées d'orales (cavité buccale uniquement) ou de nasales (cavités buccale et nasale).

Du fait du couplage entre les deux cavités, l'analyse acoustique des voyelles nasales est largement plus complexe que celle des voyelles orales. Les observations rapportées dans la littérature font état de modifications spectrales des formants associés à la seule cavité orale interprétées par l'ajout d'anti-formants induits par la cavité nasale. Ces modifications peuvent se traduire par un affaiblissement de certains formants (notamment F_1) et un décalage des valeurs des premiers formants [13].

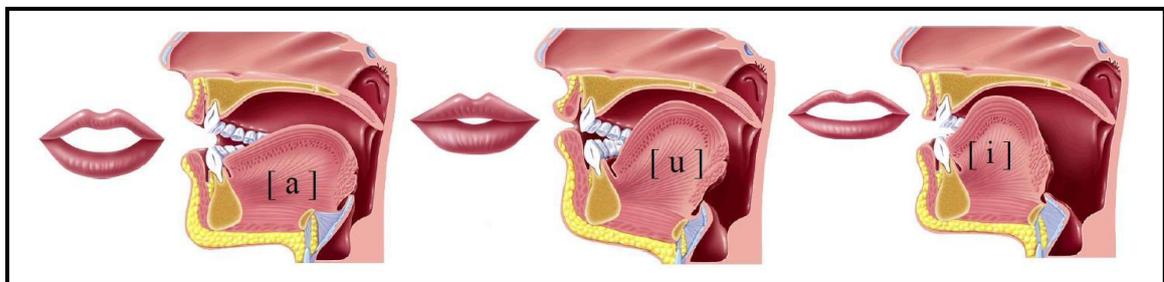


Figure 1-13: Vue en coupe de l'appareil phonatoire lors de la production des voyelles brèves [29]

Les voyelles se caractérisent principalement par la présence de zones de fréquences où les harmoniques sont particulièrement intenses (les formants). Les expériences ont montré que la position fréquentielle des trois premiers formants caractérise le timbre vocalique. L'explication qu'on donne aux trois premiers formants notés F_1 , F_2 et F_3 est la suivante [30,31]:

- F_1 naît dans la cavité comprise entre le larynx et le dos de la langue ;
- F_2 naît dans la cavité située entre le dos de la langue et les lèvres ;
- F_3 dépend de l'arrondissement des lèvres.

Dans la phonétique articulatoire classique, comme nous l'avons déjà signalé, les voyelles sont décrites en fonction de la position haute ou basse de la langue ainsi que de sa position reculée ou avancée : pour le [a] la langue est basse et reculée, pour le [i], elle est haute et avancée, pour le [u], la langue est haute et reculée. Il est commode de représenter une voyelle sur un plan F_1, F_2 . les voyelles [i], [a] et [u] sont disposées aux extrémités d'un triangle pointe en bas qui rappelle *le triangle articulatoire* de la phonétique classique (Figure 1-14). Ce triangle articulatoire représente très grossièrement la position moyenne de la langue dans la cavité buccale selon deux axes nommés « antérieur-postérieur » et « ouvert-fermé » [32].

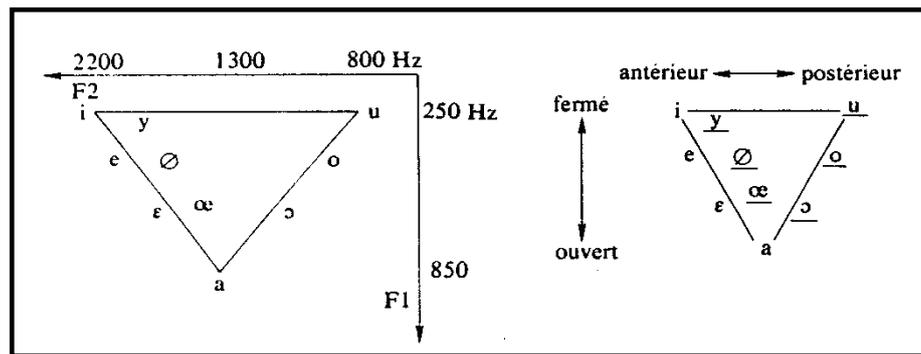


Figure 1-14: Triangle Vocalique du Français du point de vue acoustique et articuloire [32]

Le spectre vrai d'une voyelle comprend, en dehors de ses deux formants, plusieurs autres composants particuliers à un sujet, qui rendent sa voix reconnaissable parmi d'autres.

Nous pouvons donc interpréter l'augmentation de F_1 comme le résultat d'une ouverture articuloire et une augmentation de F_2 comme une antériorisation de l'articulation. L'arrondissement des lèvres se traduit par une baisse de F_3 [32].

1.5.2 Production des consonnes

Second type de phonèmes, les consonnes sont considérées comme une classe de phonèmes plus hétérogènes que les voyelles car elles présentent des distinctions articuloires plus larges. Le mode et le lieu d'articulation permettent de différencier chacune des consonnes du système consonantique de n'importe quelle langue. Le mode définissant le type d'obstruction du conduit phonatoire (c'est la façon avec laquelle l'obstruction est réalisée : totale avec les occlusives, très resserrée avec les fricatives, etc.) et le lieu d'articulation qui représente le point de rapprochement ou de contact de l'articulateur avec les parties fixes du conduit vocal.

Les consonnes sont, pour la plupart, constituées par des bruits engendrés par la traversée de l'air expiratoire à travers un rétrécissement marqué de la cavité buccale, en des endroits divers, selon la consonne articulée. Certaines consonnes y ajoutent une vibration concomitante des cordes vocales (sonores), d'autres non (sourdes).

1.5.2.1 Les Articulateurs

Ce sont toutes les parties mobiles du conduit vocal sur lesquelles on peut agir volontairement et qui sont fonctionnelles dans la production de sons de parole (Figure 1-15) :

- la langue : est sans aucun doute le plus important et le plus actif des articulateurs supra-glottiques. Elle œuvre pour modifier les formes du conduit vocal et ainsi les résonances caractéristiques de la cavité orale et du pharynx. Elle agit pour couper ou canaliser le courant d'air en contactant ou en avoisinant les dents, les alvéoles, le palais, le voile du palais ou le pharynx ;
- les lèvres : leur mouvement est conditionné par de grands muscles attachés beaucoup plus loin ainsi que par des muscles intrinsèques. La protusion (avancée des lèvres) et l'arrondi des lèvres se font toujours simultanément (s'il y a protusion, il y a arrondi) ;

- le velum (ou voile du palais ou palais mou ou luvette) : il entre en contact avec la racine de la langue et détermine deux configurations distinctes du flux d'air, passage par la bouche uniquement ou bien par la bouche et le nez ;
- le palais dur : en fonction des contacts qu'il aura avec la langue, il aura un rôle d'articulateur ;
- les dents : elles ne sont pas mobiles mais permettent le contact de la langue sur les alvéoles (derrière les dents, entre gencive et palais mou).

Palais dur et dents-alvéoles ne bougent pas : ils ne sont pas articulateurs, mais le contact avec la langue les rend articulateurs.

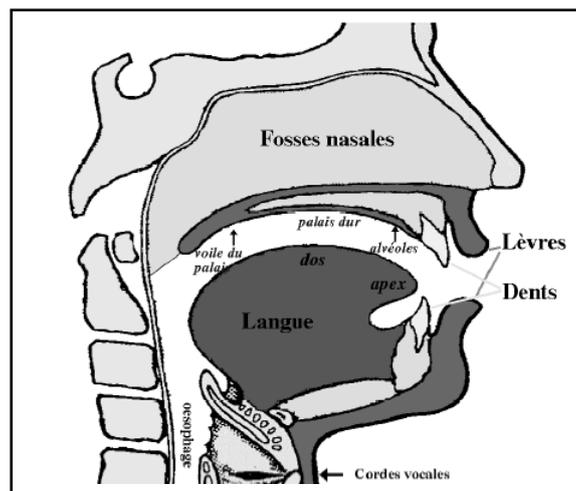


Figure 1-15 : Articulateurs

1.5.2.2 Les Lieux D'articulation

Endroit précis de la bouche où l'air rencontre un obstacle total ou partiel dans certaines productions de sons. C'est par rapport aux lieux d'articulation que sont nommées les différentes productions :

- bilabial : les deux lèvres sont en contact ;
- labio-dental : la lèvre inférieure s'appuie contre les incisives supérieures ;
- dental : pointe de la langue contre les incisives supérieures ou entre les incisives supérieures et inférieures ;
- alvéolaire : pointe de la langue contre les alvéoles ;
- post-alvéolaire : resserrement entre l'avant de la langue et les alvéoles ;
- palatal : la région coronaire se rapproche du palais ;
- vélaire : le dos de la langue s'appuie sur le voile du palais ;
- uvulaire : le dos de la langue se rapproche de la luvette ce qui peut provoquer une constriction ou une vibration de la luvette ;
- pharyngal : rapprochement de la racine de la langue de la paroi pharyngale ;
- glottal : Le son est formé au niveau de la glotte.

1.5.2.3 Le Mode D'articulation

Le mode d'articulation définit le type d'ouverture de la cavité buccale et permet de définir trois types de consonnes :

- occlusives, dans lesquelles l'obstruction du passage de l'air est presque totale ;
- fricatives, où l'obstruction n'est que partielle ;
- sonnantes pour les consonnes nasales, latérales, vibrantes.

1.5.2.3.1 Les occlusives

La réalisation des occlusives se fait sur deux temps : une tenue pendant laquelle le flux d'air est bloqué dans le conduit vocal (l'occlusion est maintenue) suivie d'une explosion et du relâchement de l'occlusion permettant la libération du flux d'air bloqué. Selon la présence ou non d'une périodicité liée aux mouvements des cordes vocales, nous pouvons distinguer deux types d'occlusives :

- les occlusives sourdes [p], [t] et [k] ;
- les occlusives sonores [b], [d] et [g].

Pour les occlusives sourdes, le conduit vocal étant totalement fermé, ces consonnes sont caractérisées par un silence durant toute la durée de l'occlusion. A la fin de l'occlusion, le contact de l'occlusion (lèvres, palais ou autre) est brutalement relâché donnant lieu à un bruit d'explosion dont la fréquence est caractéristique de la consonne. Il en va de même pour les occlusives sonores à la différence près que la vibration des cordes vocales donne lieu à une onde sonore (remplaçant le silence des occlusives apériodiques et visible sur un spectrogramme sous la forme d'une "barre de voisement") dont la fréquence est celle de la Fréquence Fondamentale (F_0) (Figure 1-16). Les caractéristiques de la barre de voisement ont une grande importance d'un point de vue perceptif. Le Voice Onset Time (V.O.T.) est le temps d'établissement du voisement, mesuré à partir de la barre d'explosion. En français, les occlusives sourdes ont un VOT positif (le voisement s'installe après l'explosion), alors que les sonores ont un VOT négatif (l'établissement du voisement précède l'explosion) [33].

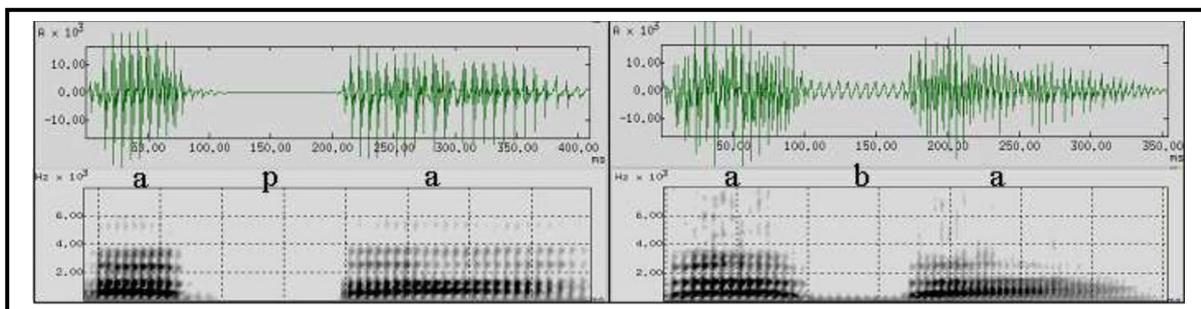


Figure 1-16: Barre de voisement permettant de distinguer le [b] du [p]

Les occlusives périodiques sont généralement plus courtes que les occlusives apériodiques. Leur bruit d'explosion est également de moindre intensité que celui de leur équivalente sourde.

Les occlusives peuvent aussi être classées selon leurs lieux d'articulation :

- occlusives bilabiales [p] et [b] ;
- occlusives alvéolaires [t] et [d] ;
- occlusives vélares [k] et [g].

1.5.2.3.2 Les fricatives

Les consonnes fricatives ou constrictives prennent naissance dans la production d'un bruit de friction qui résulte d'une turbulence aérodynamique en un ou plusieurs points du conduit vocal en raison de la présence d'un fort resserrement (ou constriction) dans le flot d'air expiratoire [32].

Tout comme les occlusives, on peut distinguer deux types de fricatives en fonction du mouvement ou non des cordes vocales :

- les fricatives sourdes [f], [s] et [S] ;
- les fricatives sonores [v], [z] et [Z].

Nous pouvons aussi différencier les fricatives en fonction de leurs lieux d'articulation:

- les fricatives labiodentales [f] et [v] ;
- les fricatives alvéolaires [s] et [z] ;
- les fricatives palatales [S] et [Z].

1.5.2.3.3 Les consonnes sonantes

Les consonnes sonantes possèdent comme les voyelles, une structure formantique, mais au contact des consonnes sourdes, les sonantes qui sont interséquentement sonores perdent leur voisement [32]. Cette classe regroupe deux sous-classes que sont les nasales et les liquides:

- les consonnes nasales qui sont souvent considérées comme des occlusives d'un point de vue articulaire étant donné l'obstruction complète du conduit vocal lors de leur réalisation. Le Français comprend deux consonnes nasales : une bilabiale [m] et une alvéolaire [n] ;
- la consonne approximante latérale [l] dont le lieu d'articulation est situé en avant du palais. Elle présente un rétrécissement du conduit vocal moins important que celui observé sur les fricatives ;
- la consonne [r] a un statut tout particulier dans le Français en raison des différentes formes (mode et lieu d'articulation) qu'elle peut prendre. Sa réalisation la plus répandue est uvulaire. Cependant, elle peut être vibrante, fricative ou se rapprochant d'une approximante selon le contexte et le phonème précédent.

Tableau 1-1 : Corrélats Articulateurs, acoustiques et perceptifs [13]

Trait	Articulation	Acoustique	Perception
Vocalisme	Vibration des cordes vocales	Source périodique	Timbre vocalique quasi - musical
Antériorité	Localisation de la langue	Spectre plus riche en hautes fréquences	Timbre plus clair
Postériorité	Localisation de la langue	Spectre plus riche en basses fréquences	Timbre plus sombre
Aperture	Ecartement des mâchoires	Rapprochement de F_1 et F_2	Son plus intense
Consonantique	Entrave au passage de l'air	Vibration aperiodique	Bruit
Occlusion	Fermeture plus tenue plus relâchement brusque	Vibration aperiodique impulsionnelle	Explosion
Constriction	Passage continu de l'air dans un rétrécissement du conduit vocal	Vibration aperiodique continue	Friction
Labiale	Utilisation des lèvres	Spectre favorable aux basses fréquences	Timbre plus sombre
Dentale	Localisation du mouvement aux dents	Spectre favorable aux hautes fréquences	Timbre plus clair
Palatale	Localisation du mouvement au palais	Spectre favorable aux moyennes fréquences	Timbre médian
Voisement	Vibrations des cordes vocales	Présence de flux laryngé	Assombrissement, sonorisation
Intensité	Accroissement du courant d'air et de la force articulatoire	Intensité plus franche	Force, intensité
Nasalisation	Abaissement du voile du palais	Modification de l'amplitude relative des formants et extra formants	Nasalité

1.5.3 Semi-voyelles ou semi-consonnes

Les semi-voyelles sont des phonèmes qui présentent une réalisation assez atypique et particulière entre les voyelles et les consonnes ce qui leur vaut les noms de " semi voyelles " ou de "semi-consonnes". Quand on les prononce, on entend le timbre d'une voyelle auquel s'ajoute le frottement d'une consonne spirante. Leur fréquence d'emploi est liée à la vitesse du débit de la parole, plus celui-ci est rapide, plus il y aura de semi-voyelles [34].

[j] [w] [y]

Exemple : [j] dans "miel" [mjɛl], [w] dans "roi" [Rwa], [y] dans "lui" [ly].

1.6 Généralités sur le système phonétique de l'Arabe Standard

Nous allons présenter une description générale du système phonétique de l'Arabe Standard (AS). La langue de référence est celle utilisée officiellement dans l'enseignement, la littérature, l'administration et les médias de l'ensemble du monde Arabe.

En effet, dans tout le monde arabe coexistent deux niveaux de langue [31] :

- le *dialectal*, qui diffère d'un pays à un autre et même d'une région à l'autre, s'emploie dans la conversation courante, reste cependant limité au niveau local ;
- le *standard* (appelé aussi classique, moderne ou littéral), est commun à tous les pays arabes, comprend l'Arabe ancien du Coran et la littérature classique. C'est la langue principale de l'écrit (journaux, correspondances) mais elle est aussi vivante à l'oral notamment dans les médias (radios, télévision) et dans les situations formelles (discours, conférences).

Selon Sibawayh , le célèbre Grammairien Arabe et l'auteur du premier livre de la tradition grammaticale arabe : Le *Kitāb*, l'alphabet arabe se compose de 35 [ḥarf] (comprenant à la fois les phonèmes caractérisant la langue [aʃl] et les variantes [far']). Cependant, les plus utilisés sont au nombre de 28, ils peuvent prendre des formes légèrement différentes selon qu'elles sont situées en position isolée, initiale, médiane ou finale dans le mot.

1.6.1 Transcription Orthographique Phonétique (TOP)

Les phonéticiens symbolisent les sons du langage au moyen de signes divers auxquels on attribue une valeur conventionnelle. L'Alphabet Phonétique International API, publié en 1888 et révisé en 2005, a pour objectif de transcrire les sons des langues du monde.

La Transcription Orthographique Phonétique (TOP), ou phonétisation, consiste à produire la prononciation correspondant au texte en entrée sous la forme d'une liste de phonèmes.

Un système de TOP est un automate paramétré appliquant un ensemble de règles de réécriture, qui permettent d'associer un phonème (ou un groupe de phonèmes) à un caractère (ou un groupe de caractères) orthographique(s) en prenant en compte le contexte gauche (caractères ou groupes de caractères précédant le segment à transcrire) et le contexte droit (caractère ou groupes de caractères suivant le segment à transcrire). Ces règles sont organisées de façon hiérarchique, des règles les plus particulières aux règles les plus générales.

Diverses transcriptions phonétiques sont utilisées selon les auteurs, celle que nous avons choisie est la plus fréquemment utilisée (Tableau 1-2).

Tableau 1-2 : Transcription Orthographique Phonétique des consonnes arabes et leurs lieux d'articulations [35]

Modes	Type de phonème	Phonèmes Arabes	Transcription Des Arabisants	Lieux d'articulation	
Occlusives	Voisées	ب	b	Bilabiale	
		د	d	Alvéodentale	
	Non Voisées	ق	q	Uvulaire	
		ت	t	Alvéodentale	
		ك	k	Postpalatale	
		ء	'	Glottale	
	Voisées Emphatiques	ض	d	Alvéolaire	
		ط	ʈ	Alvéodentale	
	Fricatives	Voisées	ز	Z	Sifflante dorsoalvéolaire
			ذ	ɖ	Interdentale
غ			g̃	Uvulaire	
ع			'	Pharyngale	
Non Voisées		س	s	Sifflante dentale	
		ث	ʈ	Interdentale	
		ف	f	Labiodentale	
		ش	ʃ	Chuichante palatale	
		خ	ħ	Vélaire	
		ه	h	Glottale	
		ح	ħ	Pharyngale	
		Voisées Emphatiques	ص	ʂ	Dorsoalvéolaire sifflante
ظ			ɖ	Interdentale	
Nasales		Voisées	م	m	Bilabiale
			ن	n	Alvéodentale
Liquide		Voisées	ل	l	Dentale
Affriquée	Voisées	ج	g̃	Alvéopalatale	
Vibrante	Voisées	ر	r	Apicvoalvéolaire	
Semi voyelles	Non Voisées	و	w	Bilabiale	
	Voisées	ي	y	Palatale	

La complexité de la TOP varie selon la langue traitée et selon l'application à laquelle elle est destinée. Par exemple, la langue française est difficile à transcrire en raison de sa forme orthographique qui est très différente de sa forme phonétique, contrairement à l'Arabe où la correspondance entre les graphèmes et les phonèmes est quasi-biunivoque [36].

1.6.2 Les sons de l'Arabe Standard

Le système consonantique joue un rôle primordial dans les langues sémitiques. L'Arabe Standard se compose de 28 consonnes, chacune d'elles correspondant à un phonème particulier.

- quinze de ces phonèmes ont une prononciation absolument identique à celle des phonèmes français ; ce sont : [b, t, ž, d, z, s, š, ġ, f, k, l, m, n, w, y] ;
- cinq autres ont une prononciation identique à celle des phonèmes existant dans des langues telles que l'Espagnole, l'Anglais et l'Allemand, il s'agit de : [t, d, h, ħ, r] ;
- 08 n'ont leur équivalent exact dans aucune autre langue européenne, ce sont : [ʔ, ħ, ʕ, q, s, ɗ, ɗ, t̤].

Chaque phonème se diffère de l'autre par le mode et le lieu d'articulation. Suivant [sibawayh], on distingue 16 lieux d'articulation [37]. Généralement, la prononciation des différents phonèmes arabes se révèle sur leurs points d'articulation, l'état des cordes vocales et l'état de la langue (Tableau 1-2). Selon ces critères, les consonnes peuvent être classifiées comme : labiales, dentales, palatales, vélares ou laryngales [38].

En outre, on peut les classifier selon leurs tenues comme occlusives ou spirantes (fricatives). Elles sont occlusives quand leur articulation est le résultat d'une détente et spirantes quand leur articulation peut être prolongée. De plus, si la prononciation d'une consonne s'accompagne d'une résonance du larynx, elle est considérée comme consonne sonore, ou sourde dans le cas contraire [39].

La particularité de la langue Arabe se trouve dans la présence des phonèmes arrières glottales [ʔ] et [ħ], pharyngales [ħ] et [ʕ], vélaire [q], de l'affriquée [ǧ] et du phonème emphatique [d̤].

1.6.2.1 Consonnes emphatiques

Le système phonétique de l'Arabe tire son originalité de la présence des consonnes emphatiques : la langue arabe est souvent appelée la langue du [d̤d̤ad]. Un phonème qui n'existe qu'en Arabe et qui est d'ailleurs difficile à prononcer [36]. Lors des réalisations des consonnes emphatiques, la langue ne couvre pas ou ne se plaque pas sur le palais : elle plie et s'incurve pour former un creux dans lequel la voix est pressée [11].

Les quatre phonèmes définis comme emphatiques sont [t̤], [d̤], [s̤] et [d̤] mais leur nombre varie d'un auteur à un autre. Ils sont caractérisés par une tonalité plus pleine et grave, car ils exigent la dépense d'un volume d'air important et une tension organique supérieure par rapport aux autres consonnes.

1.6.2.2 Gémination

La gémination en Arabe joue un rôle structural dans le développement morphologique du nom et du verbe. Elle correspond au phénomène de renforcement d'une articulation consonantique qui tend à prolonger la durée de la consonne tout en augmentant son intensité. L'absence d'une voyelle intermédiaire dans une séquence constituée de deux consonnes identiques permet l'insertion du phonème dans l'autre engendrant ainsi, le renforcement de l'articulation et le prolongement de la consonne. Toutes les consonnes arabes sont susceptibles d'être gémées sauf la hamza [ʔ].

Certaines expériences ont permis de constater que la consonne gémée ne présente pas deux mouvements articulatoires distincts, mais un mouvement unique, qui diffère de la consonne simple par sa stabilité articulatoire et par sa durée importante. Son spectre présente en général une énergie uniformément répartie et sans discontinuité.

Il existe deux sortes de gémination [40]:

- une gémation nécessaire qui a une fonction de différenciation, c'est à partir de cette gémation que dépend la signification même du mot. Elle peut donc à elle seule différencier les mots (Figure 1-17). Exemple [ḡamāl] : beauté et [ḡammāl] : chamelier ;
- une gémation euphonique est utilisée pour palier les groupes de sons qui paraissent durs à l'oreille.

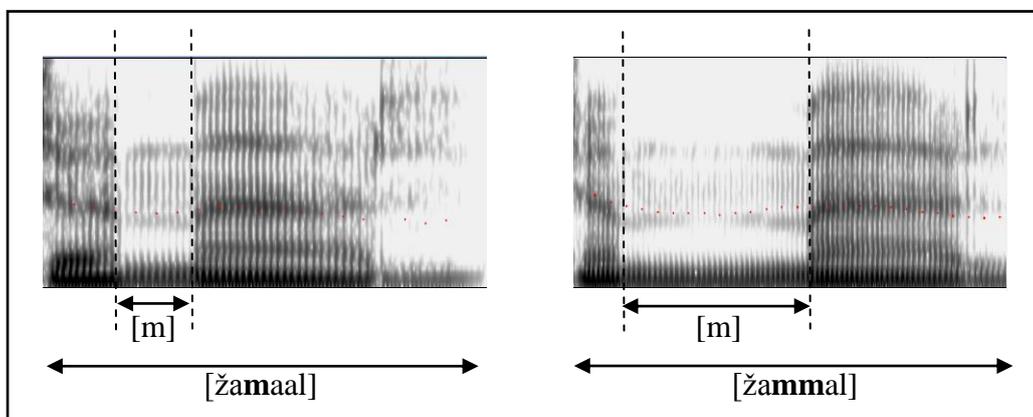


Figure 1-17 : Spectrogramme de la consonne [m] représentant une opposition simple gémée

1.6.2.3 Vocalisme de l'Arabe Standard

Le système vocalique de l'AS se compose de trois voyelles brèves : [a, u, i] s'opposant aux 3 voyelles longues : [ā, ū, ī]. La distinction entre courte et longue est semblable à la différence dans la longueur des notes musicales, qui sont jugées deux fois plus longues que les demi-notes.

Les Grammairiens Arabes nomment la voyelle brève [ḥaraka]. Pour les besoins de l'enseignement ou la nécessité de fixer la prononciation exacte d'un mot ou d'un texte on est

amené à l'exprimer par des signes diacritiques, écrits au dessus ou en dessous des consonnes. L'absence de voyelles génère une certaine ambiguïté à deux niveaux : le sens du mot et sa fonction dans la phrase. Les voyelles longues sont représentées par les lettres / *ā* / [ā], / *ū* / [ū] et / *ī* / [ī]. Elles sont écrites dans le mot en tant qu'éléments de l'épellation.

Nous avons pu constater, grâce à l'utilisation de l'articulographe AG100, qu'un signal très faible (énergie acoustique très basse ou presque sourd) requiert une articulation des organes. Ce qui nous amène à dire que le son vocalique ne produit pas le phonème mais c'est plutôt le mouvement des organes qui s'en charge. De ce fait, la voyelle ne fait qu'accompagner la consonne (subséquente ou concomitante) [41].

L'Arabe n'a que deux degrés d'aperture pertinents : [i, ī] et [u, ū] sont des voyelles fermées, tandis que [a, ā] sont des voyelles ouvertes. Elle possède également trois lieux d'articulation : [u, ū] sont postérieures (et arrondies) ; [i, ī] sont antérieures (et étirées) ; les voyelles [a, ā] sont médianes [13].

1.7 Conclusion

Les caractéristiques articulatoires et acoustiques du signal vocal ont été décrites dans ce chapitre. La dualité acoustico-articulatoire a été évoquée. Elle souligne la complexité du mécanisme de production vocale, car plusieurs paramètres peuvent agir simultanément dans ce processus.

La production vocale est le résultat d'une coordination neuromotrice des muscles de tous les organes impliqués dans la phonation depuis les muscles de la posture et de la respiration jusqu'aux muscles du larynx et de l'appareil articulatoire pharyngo-bucco-labial.

Nous avons défini les différents modes et lieux d'articulation des phonèmes de l'Arabe Standard, ainsi que les phénomènes particuliers comme l'emphase et la gémiation. Cette brève étude théorique nous a montré que la langue arabe présente des phonèmes spécifiques (vélares, pharyngales et glottales) et des phénomènes (emphase, gémiation, el-madd) que l'on ne retrouve pas dans certaines langues.

Il reste que l'objectif essentiel de la présentation de ce chapitre est que l'étude anatomique et physiologique des différents organes constitutifs de l'appareil phonatoire et l'étude acoustique de la production vocale sont un préalable indispensable à l'approche, la compréhension et la prise en charge des troubles de la voix et de la parole.

L'étude que nous présenterons dans le chapitre suivant montrera que toute variation articulatoire des signaux de la parole provoquera certainement des conséquences d'ordre acoustique et ainsi l'étude de l'aspect acoustique des sons s'avère utile.

**CHAPITRE 2 : TRAITEMENT AUTOMATIQUE DE LA
PAROLE PATHOLOGIQUE**

2.1 Introduction

L'expression « Parole Pathologique » est un raccourci pour désigner la parole produite par des locuteurs atteints de dysfonctionnements de la voix et/ou de la parole.

L'appareil vocal ou appareil de la phonation désigne l'ensemble des organes qui permettent à l'homme d'émettre des sons. Aucun de ces organes n'est exclusivement phonatoire, mais toute altération de leur intégrité entraînera une perturbation des sons émis. Donc la pathologie vocale n'est pas liée seulement à l'atteinte des cordes vocales, mais également au dysfonctionnement de la soufflerie et des cavités de résonance.

L'étude de la parole pathologique peut fournir des informations sur la production de la parole non pathologique, notamment au niveau de la variabilité, son étendue et ses limites de viabilité.

Dans cette étude nous nous intéressons à la compréhension de la production de la parole et de la parole pathologique chez des sujets ayant un problème de fente palatine.

Dans ce chapitre nous nous attacherons à la présentation des fentes palatines, puis nous proposerons de faire le point sur l'anatomie de certains articulateurs, recrutés en production de la parole et souvent touchés par les divisions palatines.

2.2 Parole pathologique

La voix, comme tout élément dépendant du corps humain, peut être atteinte par diverses pathologies. Celles-ci se divisent en deux catégories, les pathologies vocales d'origine fonctionnelle dues à une mauvaise utilisation du geste vocal et les pathologies vocales d'origine organique :

- Les pathologies d'origine fonctionnelle ou neurologique : pour ce genre de pathologies, l'appareil phonatoire est intact. Le trouble provient de déficiences fonctionnelles neurologiques telles que l'aphasie, les maladies de Parkinson et Alzheimer, etc. Ce groupe de pathologies est plus connu sous le nom de "dysarthries";
- Les pathologies d'origine organique ou physiologique : le trouble est causé par des déficiences au niveau de l'appareil phonatoire, telles qu'une ablation du larynx, une déformation de la langue, de la luvette, une anomalie congénitale des cordes vocales, etc. Le larynx et en particulier les cordes vocales sont à la base de la production de la voix, une atteinte de ces organes causera inévitablement un trouble de la voix.

Les troubles de la parole regroupent l'ensemble des déficiences affectant la production de la parole, dont voici quelques exemples :

- le bégaiement, défini par l'Organisation Mondiale de la Santé (OMS) comme un trouble du rythme de la parole dans lequel l'individu sait exactement ce qu'il veut dire,

mais est incapable de le dire en raison d'une répétition involontaire, d'une prolongation ou de l'arrêt d'un son ;

- la présence de malformations dans la cavité buccale, soit au niveau du palais (fente palatine), soit au niveau de la lèvre supérieure (fente labiale), soit au niveau des deux (fente labio-palatine), pouvant engendrer des difficultés d'ordre articulatoire et de nasalité et touchant les enfants soit, également, au niveau du frein de la langue, trop court, pour réaliser les mouvements attendus dans la réalisation de certains phonèmes, voire de la dentition ;
- l'apraxie de la parole, définie comme une altération d'origine neurologique affectant la programmation des mouvements propres à la parole ;
- la dysarthrie qui rassemble l'ensemble des troubles de la réalisation motrice de la parole dus à des lésions du système nerveux central ou périphérique ;
- la dysprosodie, définie comme l'atténuation ou la disparition complète des gestes prosodiques (rythme, intonation, timbre de la voix), se traduisant par une parole sans mélodie, monotone et un débit très lent ;
- la dyslalie infantile d'ordre fonctionnel définie comme un trouble de l'articulation, touchant les enfants lors de l'acquisition du langage. Elle se traduit par une mauvaise réalisation d'un ou plusieurs phonèmes (voire l'omission complète ou le remplacement du phonème par un autre plus facile à prononcer) due à des erreurs motrices systématiques et permanentes dans le placement des articulateurs en jeu. Ces troubles sont généralement considérés comme pathologiques chez l'enfant après l'âge de 4 ou 5 ans ;
- les troubles de l'articulation dus à une ablation partielle ou totale d'un ou plusieurs articulateurs, suite à une exérèse chirurgicale.

2.3 La fente palatine

Dès les premières semaines du développement du fœtus, les côtés gauche et droit de la lèvre supérieure et la voûte du palais grandissent normalement ensemble. Mais il arrive que ces deux parties ne se rencontrent pas, l'enfant naît avec une séparation dans la lèvre supérieure que l'on nomme fente labiale ou bec-de-lièvre. Lorsque cette séparation est localisée spécifiquement dans le conduit vocal, il s'agit d'une fente palatine.

La FP \pm FL est un phénomène embryonnaire résultant d'un échec de fusion au cours du développement maxillaire et palatin. Ces problèmes de fusion entraînent des déformations esthétiques et fonctionnelles, dont les troubles de la parole sont considérés comme les plus critiques [42].

Une fente palatine avec ou sans fente labiale (FP \pm FL) est la quatrième malformation congénitale la plus fréquente et la première anomalie craniofaciale la plus fréquente. Les taux d'incidence varient, selon la zone géographique, de 1 pour 650 naissances vivantes à 1 pour 750 naissances vivantes [43]. En Europe elle touche un enfant sur 650 [23]. Cette fréquence varie avec l'ethnie (3,6/1000 pour les indiens d'Amérique contre 0,3/1000 pour les noirs

américains), l'origine géographique, le niveau socio-économique des parents, le sexe (les fentes labiales affectent plus fréquemment les garçons, tandis que les fentes palatines concernent plus souvent les filles) [44].

Parmi les premiers travaux qui ont été menés dans l'étude de la production de la parole par des sujets porteurs de divisions palatines, nous citons :

Victor Veau (1931) était un grand chirurgien qui a été l'un des premiers à écrire beaucoup d'ouvrages sur les divisions palatines et les becs de lièvre [45].

Kieffer (1965) a fait une étude particulièrement intéressante, grâce à la méthode cinéradiographique, sur l'articulation des consonnes du français chez 2 sujets pathologiques (l'un présentait une fente palatine) comparée à celle de deux locuteurs sains d'où il a noté des changements articulatoires (type et lieu d'articulation), des durées d'arrêts non conformes à la norme, et des différences de mode (occlusions dans les constrictives) [44].

Borel-Maisonny (1969), qui a défini la terminologie des fentes, a présenté des films radiographiques sur lesquels les variations de position de la langue, à l'intérieur de la cavité buccale, extrêmement importantes est visibles [44].

Ensuite de nombreux auteurs ont tenté de décrire les déficits de communication de locuteurs porteurs de fente palatine. Ainsi, ils relèvent des erreurs articulatoires, telles que celles dues à la sensation réduite, liée à l'opération du palais, les insuffisances vélo-pharyngales, et les anomalies subsistantes du palais dur [44].

De nos jours, d'autres travaux ont analysé la voix et la parole des patients présentant une fente palatine à l'aide de nombreuses techniques dans le domaine du traitement automatique de la parole, nous citons :

Bocklet et al ont proposé deux systèmes automatiques différents qui modélisent l'espace articulatoire d'un locuteur en se basant sur l'analyse automatique des phonèmes des enfants atteints de FLP pour un diagnostic détaillé et un contrôle thérapeutique [46].

Rubin et al ont développé un système de reconnaissance simple capable de détecter les déficiences au niveau des phonèmes avec une grande précision dans le but d'améliorer la qualité de la thérapie en dehors du cabinet du médecin [47].

Vucovich et al ont développé un système d'apprentissage informatique pour évaluer la parole de locuteurs porteurs de fente. Ce système permet de faire la distinction entre la parole normale, le dysfonctionnement vélopharyngé et les erreurs de parole articulatoires en se basant sur l'interprétation des erreurs de résonance et des erreurs articulatoires dues à la fente [48].

Sudro et all ont utilisés les techniques d'augmentation de données pour simuler des données d'entraînement afin améliorer la reconnaissance de la parole pathologique et en particulier chez les enfants présentant une déficience articuloire due à une fente labiale et palatine (FLP). Cette étude révèle que les méthodes d'augmentation des données améliorent considérablement les performances de reconnaissance vocale de la parole de locuteurs porteurs de fente [49].

2.3.1 Définition de la fente palatine

Par Fente Palatine (FP), nous désignons une pathologie de l'appareil phonatoire, qui concerne la fermeture incomplète du palais (une absence de substance de la voûte buccale), ce qui entraîne la présence d'une communication entre l'intérieur de la bouche (cavité buccale) et l'intérieur du nez (cavité nasale) (Figure 2-1).



Figure 2-1 : Cas d'une fente palatine

Elle se produit au tout début de la grossesse, elle résulte d'une absence ou d'une insuffisance de fusion des différents bourgeons faciaux se déroulant de la 5^{ème} à la 7^{ème} semaine de vie intra-utérine [50]. En effet, lors de la formation du visage, de petites masses de tissus se dirigent vers la région du nez et de la bouche et se réunissent pour former la lèvre supérieure, le nez et le palais. Si ces tissus tardent à se rejoindre, ils risquent de ne pas se souder ensemble, ce qui crée la fente. A partir du moment où les lèvres et le palais se développent séparément, il est possible que l'enfant ait un bec-de-lièvre (fente labiale), un palais fendu (fente palatine), ou la combinaison des deux (fente labio-palatine) (Figure 2-2).

La fente labiale touche la lèvre, et parfois aussi la gencive et la narine. Elle peut être unilatérale (1 côté) ou bilatérale (2 côtés) (Figure 2-2).

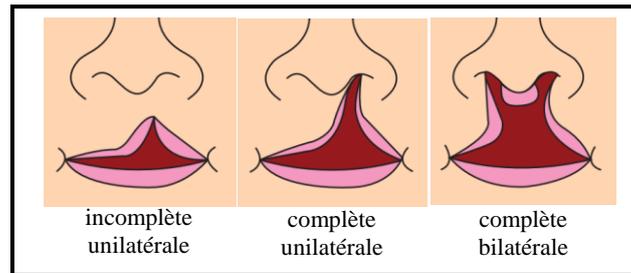


Figure 2-2 : Différents types de Fente [51]

La fente labio-palatine est la présence d'une fente labiale et d'une fente palatine (Figure 2-3). La fente labiale, associée ou non à une fente palatine est deux fois plus élevée chez le garçon, la division palatine est deux fois plus fréquente chez la fille. La prévalence est globalement deux fois plus fréquente chez les asiatiques et deux fois moins chez les africains que dans la population européenne [52].

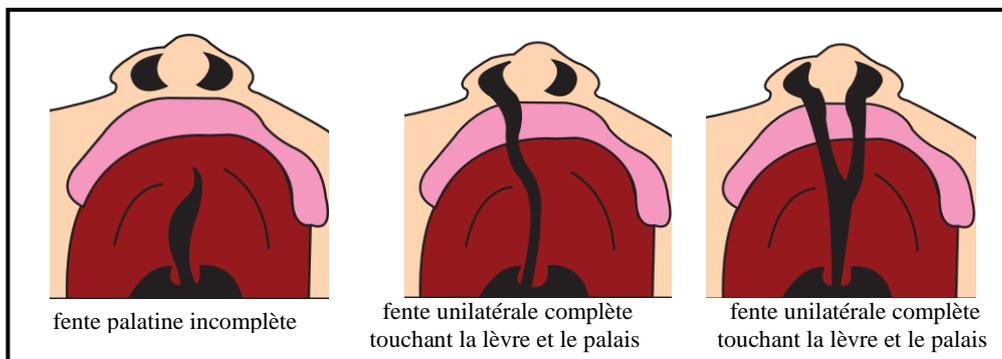


Figure 2-3 : Différents types de Fente palatine [46]

Une FP peut être plus ou moins étendue, et plus ou moins longue (Figure 2-3) et peut intéresser le seuil narinaire, la lèvre supérieure, la crête alvéolaire, la vôte palatine et le voile du palais. Tout se passe comme si le palais se fermait comme une fermeture éclair et qu'elle puisse se bloquer en un point quelconque entre la papille rétro-incisive qui marque le début du palais secondaire et l'extrémité de la lchette. Chez certains enfants, cette fente peut ne recouvrir qu'une portion du haut du palais, à l'arrière. Chez d'autres, cela peut être une séparation complète, allant de la partie antérieure à la partie postérieure de la cavité buccale touchant ainsi le palais au complet jusqu'à la lchette (Figure 2-4). Chez ces enfants, l'alimentation fait d'emblée problème : un obturateur spécialement conçu est appliqué dès la naissance sur le palais pour permettre l'alimentation du bébé. Cependant, comme l'enfant grandit rapidement, l'obturateur devra être remplacé plusieurs fois à quelques semaines d'intervalle.



Figure 2-4: Fente palatine simple [53]

Les causes des FP ne sont pas clairement établies. L'hérédité et la génétique pourraient toutefois jouer un rôle, puisque le risque est plus élevé chez les familles ayant des antécédents de FP. Certains autres facteurs comme le tabagisme, la consommation abusive d'alcool et la prise de certains médicaments en début de grossesse semblent aussi augmenter le risque de la FP [53].

2.3.2 Anatomie du palais normal

Il nous a paru important d'aborder la partie anatomique du palais dans la mesure où l'inhibition de certains muscles due à la présence de la fente et à l'intervention chirurgicale pourrait avoir des répercussions sur la production de la parole.

Le palais est le cloison qui sépare la cavité orale de la cavité nasale. Il est composé sur les 2/3 postérieurs du palais dur (également appelé voûte palatine ou palais osseux) et sur le 1/3 antérieur du palais mou ou voile du palais.

2.3.2.1 La voûte palatine

Le palais dur est une voûte osseuse qui présente une forme concave en bas, mesure environ 4 à 5 cm de large et 7 à 8 cm de long, chez l'adulte, avec une flèche de 1,5 cm. Sa forme ogivale lui donne sa solidité. Elle est formée par la réunion des lames horizontales des deux os maxillaires supérieurs d'une part, et la réunion des lames horizontales des deux os palatins d'autre part. Cette voûte est recouverte d'une muqueuse particulière comportant des glandes salivaires accessoires. C'est une fibro-muqueuse, un tissu inextensible très adhérent au maxillaire. Sa couche profonde est le périoste du maxillaire, un tissu capable de fabriquer de l'os.

2.3.2.2 Le voile du palais

Dans le prolongement postérieur du palais dur, est situé le palais mou, partie flexible du palais qui se termine au centre par la luette (ou uvule). Deux replis se détachent de cette partie : un en avant vers la langue, le repli ou l'arc palato-glosse, et un en arrière vers le pharynx, le repli ou l'arc palato-pharyngien. Entre les deux arcs existe une formation lymphoïde appelée la tonsille.

Le palais mou est constitué d'une charpente fibreuse sur laquelle s'insèrent les muscles pairs chargés de sa mobilité et d'une muqueuse recouvrant les deux faces du voile pourvue de glandes palatines et de récepteurs de goût [14]. Les muscles sont au nombre de 5 : le muscle palato-glosse, le muscle tenseur du voile, le muscle palato-pharyngien, le muscle élévateur du voile, le muscle uvulaire (azygos de la lnette) (Figure 2-5). Chacun ayant un rôle très important dans la production de la parole.

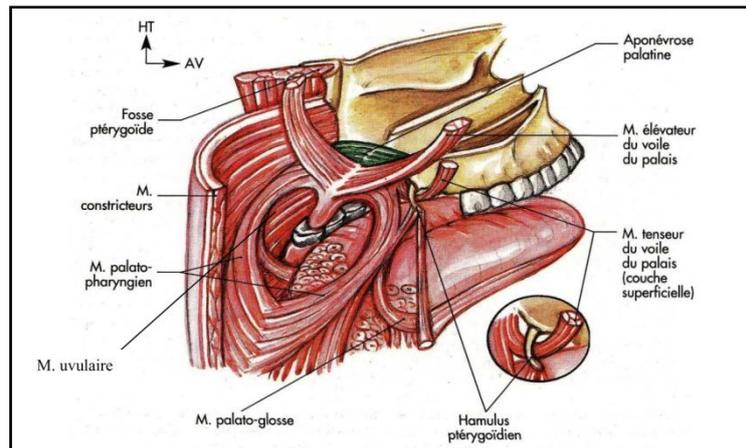


Figure 2-5: Schéma des muscles du voile du palais [54]

Le voile du palais participe à de multiples fonctions grâce à sa souplesse et à l'action des muscles qui le composent ; il joue un rôle dans la respiration, la succion, l'absorption des liquides, la déglutition et l'aération de l'oreille moyenne. Il joue un rôle important dans le mécanisme de la parole. Lorsque son fonctionnement est altéré, c'est l'ensemble du système qui est atteint, et donc la parole.

Le voile du palais représente un des éléments variables des cavités susglottiques, où se modifie le son laryngien afin d'assumer, dans la parole, des caractéristiques acoustiques déterminées que l'oreille reconnaît comme phonème. Son intégrité assure la bonne réalisation des phonèmes au niveau de l'articulation et de la résonance. En cas d'insuffisance vélaire, les défauts articulatoires et les projections d'air par le nez s'ajoutent à la variation de timbre.

2.3.3 Principales parties du palais déformées par la fente palatine

La face se construit en pièces détachées comme un puzzle à partir de bourgeons qui apparaissent et se développent autour de la bouche primitive (stomodeum). Ce développement suit un calendrier précis qui se déroule en deux étapes :

- le palais primaire qui correspond à la gencive se forme d'abord en même temps que la lèvre ;
- le palais secondaire qui sépare la bouche du nez en même temps que la cloison nasale sépare les deux fosses nasales.

La croissance vélaire est rapide entre 18 et 24 mois puis s'atténue jusqu'à 5 ans pour reprendre une croissance régulière jusqu'à 17 ans. Chez l'enfant porteur d'une fente palatine, le mécanisme fondamental de l'incompétence vélaire est surtout lié aux modifications anatomiques des muscles du voile dues à la fente.

2.3.3.1 Embryologie du Palais Primaire

Il correspond à la gencive, à la partie périphérique du palais celle qui est en avant de la papille incisive (le petit relief que l'on peut sentir avec la langue quelques millimètres derrière les incisives).

C'est le défaut de fusion du bourgeon maxillaire avec le bourgeon nasal interne qui sera responsable de la fente labio-narinaire et de celle du palais primaire. En effet, ces deux phénomènes embryonnaires étant consécutifs, ils ne sont que partiellement liés et il est possible d'avoir une fente labiale sans fente palatine ainsi qu'une fente palatine sans fente labiale.

2.3.3.2 Embryologie du Palais Secondaire

Le palais secondaire se développe après que les bourgeons constitutifs de la lèvre et du palais primaire n'aient fusionné. A ce stade, il existe une communication normale entre les cavités nasales et buccales qui ne sont pas encore cloisonnées.

Une cloison sagittale médiane va descendre du plafond des fosses nasales séparant le nez droit du nez gauche (en vert sur les photos ci-dessus) pendant que deux cloisons horizontales, les lames palatines (en rouge sur les photos ci-dessus), se développent à partir des régions latérales du palais primaire. Ces trois cloisons se réunissent sur la ligne médiane, d'abord en avant sur la papille rétro-incisive puis se soudent d'avant en arrière jusqu'à la luette, comme si l'on tirait une fermeture éclair (Figure 2-6). S'il n'y a pas eu du tout de soudure, la division palatine est totale (avec ou sans fente du palais primaire). Si la "fermeture éclair a coincé", la division palatine sera plus ou moins importante.

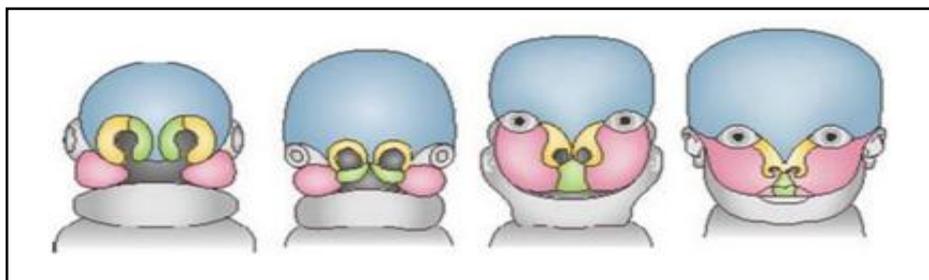


Figure 2-6 : Les bourgeons faciaux et leur développement.
De gauche à droite : 3 semaines, 7 semaines, 10 semaines, 3 mois [55]

La partie qui résulte de la fusion des deux lames palatines avec le septum correspond au palais osseux, en arrière du septum au palais mou ou voile du palais.

2.4 Les Différents Types De Fentes

Vu la multitude des formes des fentes faciales et la présence de plusieurs variétés anatomiques, classer les fentes fût nécessaire pour satisfaire les cliniciens, chirurgiens, et chercheurs d'une part, et de leur permettre d'autre part de communiquer entre eux en un langage universel.

2.4.1 Classification de V.VEAU

Le docteur V.VEAU (1931), un des premiers qui a écrit beaucoup d'ouvrages sur les divisions palatines, distingue quatre formes de divisions palatines congénitales [56].

2.4.1.1 Division simple du voile

La fente est localisée au palais mou (Figure 2-7:) ;

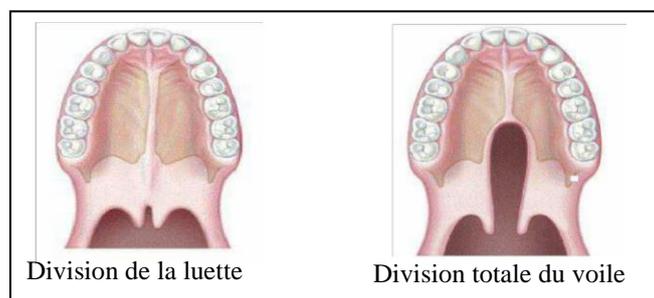


Figure 2-7 : Fente palatine simple [57]

2.4.1.2 Division du voile et de la voûte sans bec-de-lièvre

La solution de continuité est médiane, elle ne dépasse jamais le canal palatin antérieur ; l'arcade alvéolaire est intacte (Figure 2-8);

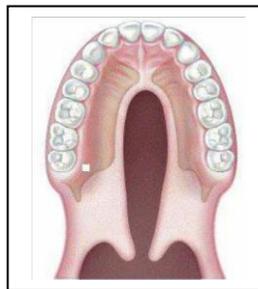


Figure 2-8 : Division du voile et de la voûte palatine d'après Veau [57]

2.4.1.3 Division du voile et de la voûte associée à une fente labio-alvéolaire unilatérale

L'arcade alvéolaire est intéressée, la langue passe entre le massif des incisives et la canine (Figure 2-9).

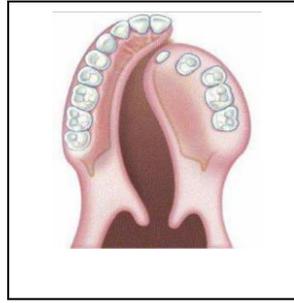


Figure 2-9 : Fente labio-alvéolo-palatine unilatérale totale d'après Veau [57]

2.4.1.4 Division du voile et de la voûte associée à une fente labio-alvéolaire bilatérale totale

Le tubercule médian est complètement isolé (figure 2-10).

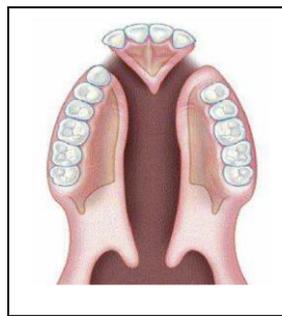


Figure 2-10: Fente labio-alvéolo-palatine bilatérale totale d'après Veau [57]

2.4.2 Classification internationale des fentes

La confédération internationale de chirurgie plastique a adopté depuis 1969 ; la classification ci-après [53, 57, 58]:

2.4.2.1 Fentes intéressant le palais primaire

On subdivise ces fentes en :

- Fente unilatérale incomplète du palais primaire (1/3,2/3,3/3 de la lèvre) ;
- Fente unilatérale complète du palais primaire (1/3, 2/3 alvéolaire) ;
- Fente bilatérale complète du palais primaire (3/3 alvéolaire) (Figure 2-11) ;
- Fente unilatérale incomplète (1/3, 2/3 ; 3/3 de la lèvre) Fente unilatérale complète (1/3, 2/3 alvéolaire).

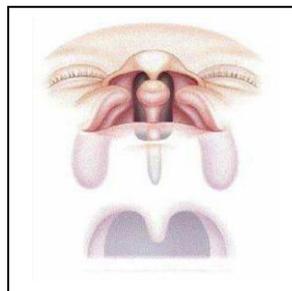


Figure 2-11 : Fente bilatérale complète du palais primaire (3/3 alvéolaire) [57]

2.4.2.2 Palais secondaire (voûte palatine osseuse et voile)

On subdivise ces fentes en deux types :

- Fente incomplète du palais secondaire (Figure 2-12 a) ;
- Fente complète du palais secondaire (Figure 2-12 b).

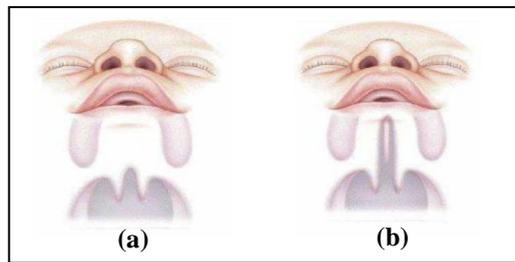


Figure 2-12 : Fentes concernant le palais secondaire [57]

2.4.2.3 Association palais primaire et palais secondaire

On distingue respectivement (Figure 2-13 a, b et c) :

- Fente unilatérale complète du palais primaire et secondaire (a) ;
- Fente bilatérale du palais primaire et secondaire (b) ;
- Fente unilatérale incomplète du palais primaire et secondaire (c).

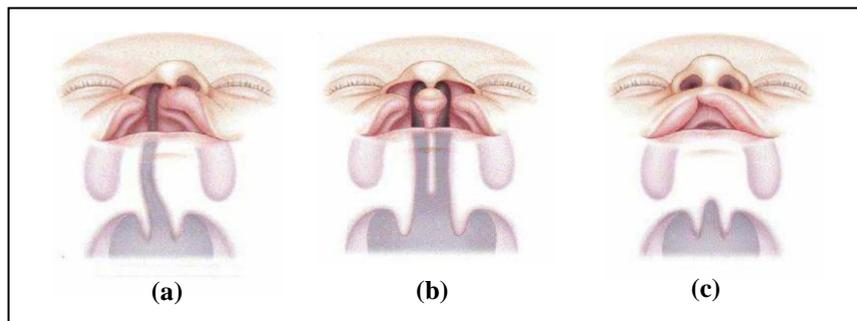


Figure 2-13 : Association palais primaire et palais secondaire [57]

Les fentes peuvent donc être plus ou moins étendues, et plus ou moins longues. Tout se passe comme si le palais se fermait comme une fermeture éclair et qu'elle puisse se bloquer en un point quelconque entre la papille rétro-incisive qui marque le début du palais secondaire et l'extrémité de la luette.

Du point de vue chirurgical, les fentes du voile larges se referment aussi facilement que les autres grâce à l'élasticité du voile.

Une forme très importante de division palatine est celle où il existe, une division qui s'arrête au milieu du voile, une division qui s'étend jusqu'à la voûte osseuse et la zone pellucide en avant de la fente.

Il n'existe pas de critère clinique permettant de différencier les divisions du voile seul et celles de la voûte et du palais mais un critère anatomique : la fente osseuse.

Les divisions totales s'étendent jusqu'au trou palatin antérieur. Il arrive que la fente soit totale, alors qu'en arrière de ce trou se trouve un pont de parties molles plus ou moins étendues.

Il se peut qu'il y ait une absence des lames palatines osseuses. Dans ce cas, la voûte n'est pas ogivale, mais est un plafond horizontal. Notons également la division palatine par enclavement de la langue entre les lames palatines ; la langue se trouve dans les fosses nasales.

2.5 Prise en charge des différentes fentes

La prise en charge des fentes faciales doit être pluridisciplinaire :

2.5.1 Interventions chirurgicales

Le traitement chirurgical des fentes faciales se veut avant tout fonctionnel. Le but est de normaliser la croissance de l'enfant, c'est pourquoi l'établissement d'un calendrier chirurgical est primordial. La chirurgie prend place dans une prise charge pluridisciplinaire et doit donc être prise en compte dans le projet thérapeutique global du patient [58].

La chirurgie de la division vélo-palatine va viser à rendre au palais son anatomie et sa fonction, il s'agit de fermer l'ouverture entre le palais et la cavité nasale. Elle consiste à faire une incision de part et d'autre de la division pour déplacer chaque bord vers le centre de la cavité buccale. Cette intervention a pour but de reconstruire le palais, joignant les muscles, et crée un palais de longueur suffisante pour que l'enfant puisse manger et parler de façon correcte (Figure 2-14). En l'absence d'intervention, une prothèse pourra aider au développement d'un langage intelligible.

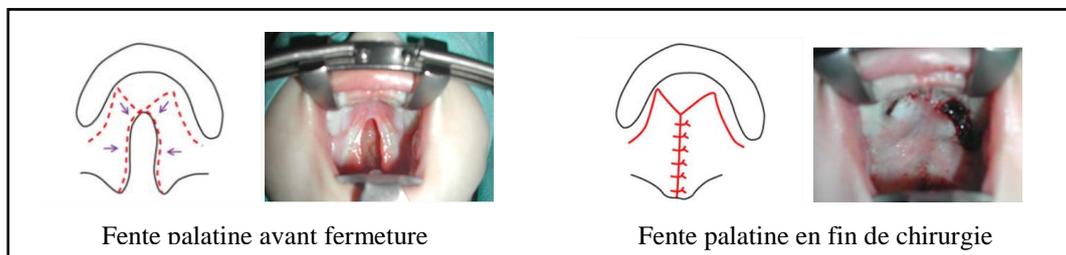


Figure 2-14 : Chirurgie d'une fente palatine [59]

Les différents traitements qui peuvent être effectués dans le domaine de la chirurgie pour les fentes sont [58, 60]:

2.5.1.1 Traitement primaire

En période néonatale, on réalise une rhinochéiloplastie : réparation complète de la lèvre et du seuil narinaire (Figure 2-15).

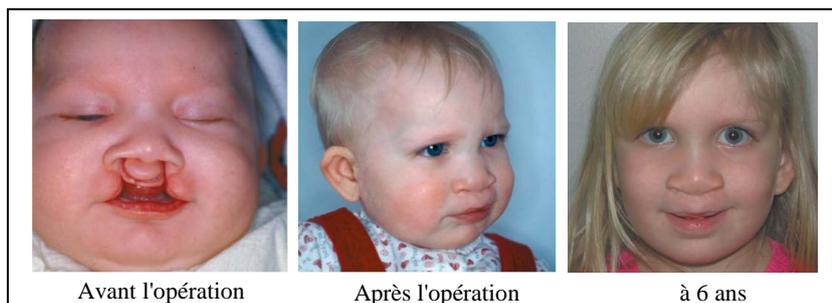


Figure 2-15 : Chirurgie de fente labio-palatine bilatérale [50]

A 10 mois, l'uranostaphylorrhaphie (chirurgie primaire du palais secondaire) est réalisée. Celle-ci regroupe deux opérations chirurgicales :

- la staphylorrhaphie : chirurgie du voile du palais qui permet d'assurer une fonction vélo-pharyngée efficace ;
- l'uranorrhaphie : chirurgie du palais osseux qui vise à assurer l'étanchéité entre les cavités buccale et nasale.

Il est essentiel de pratiquer cette opération avant que l'enfant n'atteigne sa première année puisque c'est à cet âge qu'il utilisera activement son voile du palais.

Entre 4 et 6 ans, on effectue une gingivopérioplastie (chirurgie du palais primaire). Celle-ci consiste en la fermeture de la communication bucco-nasale antérieure, elle est réalisée lorsque la denture lactéale est en place.

Il peut y avoir des voiles un peu court, des voiles qui ne sont pas assez musclés. Cela provient de l'anatomie du bébé. Dans certaines formes, ils ont des muscles très fins. Ils ont alors du mal à avoir une musculature efficace pour contenir l'air qui pourrait passer par le nez et donner une voix nasonnée. C'est le travail des orthophonistes. Parfois, une petite retouche chirurgicale est nécessaire lorsque l'enfant en a fini avec l'orthophonie.

2.5.1.2 Traitement secondaire et séquelles

- Une incompétence vélo-pharyngée persistante nécessitera une pharyngoplastie de type sphinctéroplastie ou un allongement vélaire par double plastie en Z de type Furlow.
- Une rhinoplastie (reconstruction de la forme du nez et de la cloison nasale) sera réalisée si le patient rencontre des problèmes fonctionnels ou pour des raisons esthétiques, à la demande de celui-ci.
- Une chirurgie orthognatique pourra être réalisée si la croissance des dents est défavorable. Cela permet d'obtenir un positionnement dentaire normal en intervenant sur les maxillaires. Cette chirurgie est réalisée dans moins de 15% des cas.
- Des retouches esthétiques au niveau de la lèvre peuvent également être réalisées.

Dans certains cas, même suite à des opérations chirurgicales réussies, il reste des trous résiduels, appelés « fistules ». Une fistule peut avoir un effet perturbateur sur le langage, dans la mesure où elle augmente le flux d'air nasal et fait baisser la pression d'air intra-orale, lors de la production de consonnes occlusives orales, provoquant une déformation de ces sons. Un changement dans le lieu d'articulation peut entraîner des erreurs de production des sons. En effet, le mouvement de la langue peut avoir pour but de bloquer la fistule afin d'avoir une pression normale et de maintenir les caractéristiques d'une consonne, consonne dont le lieu articulaire pourrait alors être dévié.

La chirurgie, contrairement à la fusion naturelle qui met en jeu certains processus biologiques embryonnaires, crée des cicatrices et la biologie de ces cicatrices peut interférer

avec les mécanismes de la croissance. C'est pourquoi aucun procédé n'est parfait, tous ont des avantages et des inconvénients.

2.5.2 Prise en charge orthophonique

L'intervention orthophonique dans le traitement des fentes labio-palatines débute dès l'âge d'un an, par la mise en place de guidances parentales. Elle se poursuit par des évaluations phonétiques régulières mais espacées jusqu'à l'adolescence. Les guidances parentales précoces permettent de stimuler et de favoriser la fonction vélaire et l'éclosion de la parole et du langage. Des conseils et des exercices adaptés à l'âge et aux possibilités des enfants sont donnés au cours de trois sessions (à 18, 24, 30 mois). Le suivi orthophonique objectif débute vers l'âge de trois ans par des bilans réguliers et des nasométries. La phonation est ainsi quantifiée et qualifiée.

2.5.3 Prise en charge psychologique

L'annonce de la découverte d'une fente représente toujours un moment de choc pour les parents. C'est un événement pénible, qui se surajoute à celui de la naissance et entraîne souvent des troubles psychiques chez les parents. Certains parents ressentent un sentiment de culpabilité, de perte de maîtrise et se sentent affectés au niveau de leurs représentations parentales, ce qui risque d'avoir une incidence sur le développement psychoaffectif de l'enfant. Un soutien aux parents peut alors s'avérer utile et est offert dès le diagnostic prénatal. L'enfant, ou l'adolescent, peut aussi être affecté au niveau de ses compétences sociales, avec pour résultat des manifestations d'anxiété, une mauvaise estime de soi ou des affects dépressifs. La consultation psychologique offerte par des spécialistes à chaque consultation pluridisciplinaire permet de repérer des situations à risque et d'organiser si nécessaire une prise en charge adaptée à chaque enfant.

2.5.4 Les conséquences les plus fréquentes des fentes sur l'articulation et la phonation

Les problèmes les plus critiques chez les patients porteurs de fente est un bien-être fonctionnel et socio-émotionnel insuffisant qui contribue à la réduction de la qualité de vie chez ces enfants [61].

Même après la réparation chirurgicale, les muscles du voile du palais reconstruits ne fonctionnent pas d'emblée au maximum de leur efficacité. « La déglutition, les cris entraînent bien ces muscles, mais pour la fonction tubaire et pour la délicate mise en place des points articulatoires, les progrès sont plus lents. » [55].

L'incompétence vélo-pharyngée, due à l'anomalie des muscles vélaire, a comme conséquences des altérations qui portent généralement sur le timbre de la voix et sur son intensité parfois [56]. Les troubles les plus fréquents sont :

- La rhinolalie ouverte est une modification de la voix due à une exagération de la perméabilité nasale. Les phonèmes occlusifs et les voyelles orales sont nasalisés ; les fricatives ne peuvent être correctement articulées ;
- Le nasonnement correspond à une modification du timbre vers les fréquences graves par adjonction de la cavité nasale à la cavité buccale ;
- La déperdition nasale. C'est un bruit de souffle qui s'entend sur les consonnes (plus audible sur les occlusives que sur les fricatives) ;
- La raucité peut exister en compensation d'une insuffisance vélo-pharyngée. Elle peut aller de l'impression de voix éraillée jusqu'à la désonorisation complète de tous les phonèmes ;
- Le manque d'intensité de la voix peut être une compensation ou une conséquence de l'incompétence vélo-pharyngée ;
- des syncinésies : ce sont des contractions visibles des muscles du visage et ce, surtout au niveau du nez et de la lèvre supérieure, qui témoignent de l'effort que fait le sujet pour essayer de pallier à la fuite d'air nasal.

Au niveau de l'articulation les troubles qui peuvent être entraînés par l'incompétence vélo-pharyngée sont :

- des phénomènes de compensations, qui sont des mécanismes qui consistent à tenter de compenser l'insuffisance vélaire. La fuite d'air nasale est massive et empêche toute possibilité de pression intra-buccale. Le sujet cherche alors des points d'occlusion et de constriction en amont du vélo-pharynx : on peut alors observer :
 - ✓ le coup de glotte : c'est une attaque dure qui peut remplacer les occlusives. Les cordes vocales s'accolent, l'air s'accumule dans la trachée et, à la séparation des cordes vocales, on entend une explosion glottale ;
 - ✓ le souffle rauques : les constrictives sont remplacées par une constriction produite dans la région sus-laryngée, l'air passant en sifflant entre les cordes vocales très rapprochées.
- remplacement de phonèmes par des phonèmes voisins par leur point d'articulation ou leur mode d'articulation (par exemple [t] devient [k], [z] est désonorisé...). La postériorisation des occlusives (souvent [k]-[g] pour [t]-[d]) se rencontre particulièrement chez les fentes bilatérales totales. Elle sera rééduquée avec attention ;
- articulation atypique des phonèmes (par exemple [t] interdental, [s] et [z] sont schlintés...)
- sigmatisme nasal : les constrictives sont nasalisées car la langue se recule et oblige l'air à passer par le nez.

En ce qui concerne la longueur des phrases, ces enfants ont tendance à utiliser des phrases plus courtes, peut-être à cause d'une moins bonne utilisation de l'air expiratoire.

2.6 Paramètres acoustiques de la parole pathologique

La Fente Labiale et Palatine (FLP) est l'anomalie crâniofaciale la plus courante. La FLP affecte la résonance, la voix et la parole. Outre les troubles de la résonance et de la parole les plus fréquemment rapportés, plusieurs rapports ont abordé des anomalies acoustiques dans la voix des patients atteints de FLP. Cependant, il n'y a que quelques rapports axés sur le traitement vocal dans cette population [62].

La qualité d'une voix est communément caractérisée sur le plan acoustique par trois paramètres qui sont la hauteur, l'intensité et le timbre. La hauteur de la voix est fixée par la fréquence fondamentale (F_0) de l'onde glottique. L'intensité de la voix traduit, quant à elle, l'amplitude des vibrations des cordes vocales et est principalement contrôlée par la variation de la pression sous-glottique. Plus la pression sous-glottique sera grande, plus l'intensité vocale le sera également, plus la voix sera considérée comme forte du point de vue perceptif. Finalement, le timbre de la voix est peut-être l'élément essentiel dans le principe d'individualité chez l'être humain, puisqu'il nous permet d'identifier une personne à l'écoute de sa voix. Il est directement lié au nombre et à l'intensité des harmoniques présents dans l'onde acoustique, ce qui en fait sa richesse.

Une Parole Normale (PN) est analysée essentiellement par l'observation des paramètres principaux :

- La fréquence fondamentale F_0 ;
- La durée des sons pour étudier le débit d'air et la fluidité de la parole ;
- L'intensité ou l'énergie qui permet de distinguer un son fort d'un son faible ;
- Les formants.

Dans le domaine de la Parole Pathologique, l'analyse acoustique fait appel à d'autres paramètres aussi importants tels que le degré de perturbation de F_0 ou Jitter et le degré de perturbation de l'intensité ou Shimmer, qui sont très exploités pour la caractérisation de la qualité de la PP [63, 64]. Ces deux paramètres sont habituellement mesurés sur les voyelles soutenues, et leurs valeurs au-dessus d'un certain seuil sont considérées comme étant liées à des PP [65].

2.6.1 Perturbation de F_0 ou Jitter

Le Jitter décrit un désordre vocal, causé selon les auteurs par un dysfonctionnement neurologique, aérodynamique ou biomécanique.

Le jitter est principalement affectée par le manque de contrôle des vibrations des cordes ; les voix des patients atteints de pathologies ont souvent un pourcentage plus élevé de jitter [66].

Le Jitter est une mesure des perturbations de la fréquence fondamentale du signal sonore. On se place ici au niveau de la période du son, donc du cycle vibratoire, et on observe les différences de durée entre une ou plusieurs périodes (Figure 2-16).

Dans Praat, il y a cinq mesures différentes du Jitter. Nous utiliserons dans notre étude le *J factor*. Il s'agit de la moyenne de toutes les différences, en valeur absolue, entre les durées de deux périodes consécutives du signal rapportée à la durée moyenne d'une période du signal, le résultat est donc un rapport exprimé ici en %.

Le Jitter *J* réfère à la variation cycle par cycle de F_0 dans une trame du signal :

$$J = \frac{1}{N-1} \sum_i |P(i) - P(i+1)| \quad (2.1)$$

Avec :

$$P_{moy} = \frac{1}{N} \sum_i P(i)$$

Alors :

$$J_f = 100 \times \frac{J}{\frac{1}{N} \sum_i P(i)} (\%) \quad (2.2)$$

P : période
N: Nombre de périodes.

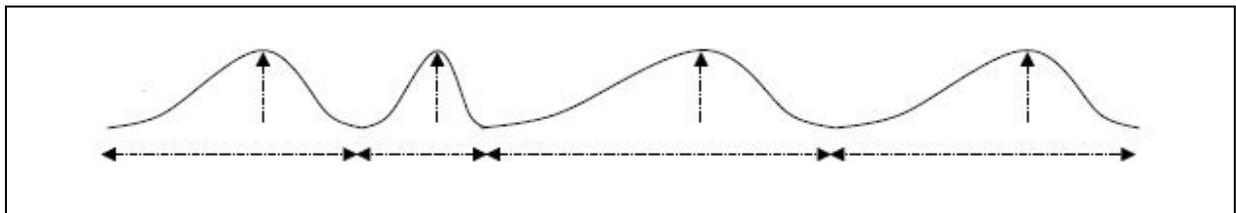


Figure 2-16 : Exemple d'apériodicité de la fréquence fondamentale

En cas de pathologie vocale, le Jitter augmente et une valeur supérieure à 1.04 % correspond à une parole pathologique [67].

Il est intéressant de noter que les perturbations de F_0 sont majorées à l'attaque et en final du son. Plusieurs auteurs préconisent, donc, d'analyser le Jitter à partir de la partie médiane de la voyelle.

2.6.2 Perturbation de l'intensité ou Shimmer

Le Shimmer est une mesure des perturbations à court terme de l'amplitude du signal sonore. Il permet la quantification de la perturbation de l'intensité.

Le shimmer change avec la réduction de la résistance glottale et des lésions de masse sur les cordes vocales et est corrélé à la présence d'émission de bruit et de respiration [68].

Le Shimmer *S* réfère à la variation cycle par cycle de l'intensité dans une trame du signal. Il se calcule par la moyenne de la différence d'amplitude entre deux cycles consécutifs de vibrations (Figure 2-17).

$$S = \frac{1}{N-1} \sum_i |A(i) - A(i+1)| \quad (2.3)$$

Comme pour le Jitter factor J_f , le Shimmer factor S_f est un bon indice pour explorer la stabilité de l'intensité. Pour le calculer, on divise la moyenne des différences, en valeur absolue, entre l'amplitude maximale de deux périodes successives, par la moyenne des amplitudes maximales de chaque période. Ainsi, il nous permet d'estimer en pourcentage l'irrégularité de l'intensité d'un ensemble de cycles vibratoires.

$$S_f = 100 \times \frac{S}{\frac{1}{N} \sum_i A(i)} (\%) \quad (2.4)$$

Avec : $A_{moy} = \frac{1}{N} \sum_i A(i)$

Alors : $S_f = 100 \times \frac{S}{A_{moy}} (\%) \quad (2.5)$

Le Shimmer absolu moyen, exprimé en dB, est la moyenne des rapports d'amplitudes entre deux cycles consécutifs de vibrations des cordes vocales. De même que pour le Jitter factor, le Shimmer factor relativise le Shimmer moyen et augmente en cas d'anomalie laryngée :

$$S_{dB} = \frac{1}{N-1} \sum_i |20 \log_{10}(A(i) - 20 \log_{10}(A(i+1)))| \quad (2.6)$$

$$S_{dB} = \frac{1}{N-1} \sum_i |20 \log_{10}(A(i)/A(i+1))|$$

$$S_{f_{dB}} = 100 \cdot \frac{S}{\frac{1}{N} \sum_i 20 \log_{10}(A(i))}$$

Avec : $A_{dB_{moy}} = \frac{1}{N} \sum_i 20 \log_{10}(A(i))$

Alors : $S_{f_{dB}} = 100 \cdot \frac{S_{dB}}{A_{dB_{moy}}} \quad (2.7)$

$A(i)$: Amplitude du pic P_i ;
 N : Nombre de pics glottiques.

En cas de pathologie vocale, le Shimmer augmente et une valeur supérieure à 3.81% correspond à une parole pathologique [67]. Ce seuil correspond à 0.35 dB. Ainsi, le seuil Normal/Pathologie correspond à 3.81 % et 0.35 dB pour le Shimmer.

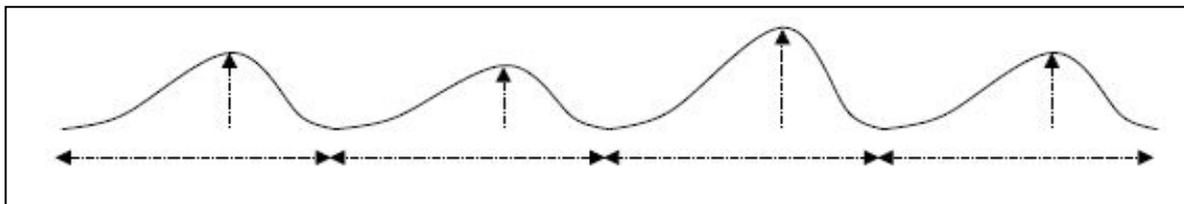


Figure 2-17 : Exemple d'instabilité de l'amplitude du signal

2.6.3 Rapport Harmoniques/Bruit (HNR)

Afin de pouvoir caractériser la qualité vocale, la connaissance énergétique entre harmoniques et bruit dans un signal de parole peut-être une information importante.

L'instabilité du signal glottique se manifeste comme un bruit qui lui est superposé. Elle est évaluée au moyen du rapport entre l'énergie harmonique dans le spectre du signal et l'énergie du bruit.

Pour cela, nous utilisons la méthode dite "HNR" (Harmonics to Noise Ratio) proposée par Yumoto et al., qui permet de calculer le rapport "énergie des harmoniques / énergie du bruit". Il a été démontré que de manière globale, 95% des voix saines vont avoir un HNR supérieur à 7.4 dB [67]. La mesure du HNR est considérée comme un indice de raucité, c'est-à-dire qu'elle renseigne sur le rapport bruit/harmoniques dans une voyelle. Traditionnellement, plus le ratio baisse, plus le signal est envahi par le bruit [68]. Globalement, le HNR est défini par :

$$HNR = \text{Energie}(F_0 + H_i) / \text{Energie}(\text{Bruit}) \quad (2.8)$$

$$HNR_{dB} = 10 \log_{10}(E_h/E_b) = 10 \log_{10}\left(\frac{\sum_{n=0}^{N-1} (H[n])^2}{\sum_{n=0}^{N-1} (B[n])^2}\right) \quad (2.9)$$

Avec :

- E_h : énergie des composantes harmoniques et périodiques du signal de parole ;
- E_b : énergie des composantes apériodiques ou bruit ;
- N : taille du signal.

Un HNR bas indique une voix asthénique et une dysphonie. Autrement dit, avec une valeur inférieure à 7 dB dans HNR est considéré comme pathologique [68].

Selon le logiciel Praat, nous pouvons considérer que nous sommes en présence d'une voix pathologique pour un rapport H/N très inférieur à 20 dB pour le phonème [a].

Le HNR diminue avec l'âge alors que l'instabilité vocale augmente [69].

2.6.4 Taux de Passage par Zéro (TPZ)

Le Taux de Passage par Zéro (TPZ) est défini par le nombre de changements de signe du signal vocal sur 20 ms (il a été prouvé, expérimentalement, que les sons voisés changent moins fréquemment de signe que les sons non voisés), dans le but de localiser les régions non voisés. Par conséquent, si $TPZ > 70$ on admet qu'il y a absence de voisement [70]. Cette caractéristique est fréquemment utilisée pour la classification de la parole Voisée/Non Voisée ou pour identifier les consonnes fricatives.

Le taux de passage par zéro court terme est estimé par la formule :

$$Z_n = \frac{1}{2} \sum_m |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (2.10)$$

Avec :
$$sgn[x(m)] = \begin{cases} 1, & \text{si } x(m) \geq 0 \\ -1, & \text{si } x(m) < 0 \end{cases} \quad (2.11)$$

et
$$w(n) = \begin{cases} \frac{1}{N}, & 0 \leq n \leq N \\ 0, & \text{ailleurs} \end{cases} \quad (2.12)$$

Du fait de sa nature aléatoire, le bruit possède un TPZ supérieur à celui de la parole voisée.

Le Taux de Passage par Zéro constitue un outil important pour la classification Voisée/Non Voisée et pour la détection du début et la fin de la parole dans un signal vocal.

2.6.5 Le Temps Maximum de Phonation (TMP)

Pour que les résultats de l'évaluation soient plus fiables, nous avons ajouté une mesure aérodynamique qui est le Temps Maximum de Phonation (TMP). La mesure la plus fréquemment utilisée est simple, non invasive et nécessite seulement un chronomètre.

Le TMP est la mesure du temps maximal d'émission vocale sur un [a] tenu, à une hauteur et une intensité confortables. La longueur du TMP dépend à la fois de la capacité pulmonaire et de la qualité d'accolement des cordes vocales. C'est un bon indicateur du rendement de la source vocale puisque plus la fuite glottique, est conséquente et plus le TMP est court.

Généralement exécutée sur un [a], la normale est comprise entre 15 et 25 secondes. En fait, la limite inférieure est la norme intéressante. Elle est un peu moins élevée chez la femme : 15 s chez l'homme, 10 s chez la femme [4].

2.7 Conclusion

La fente palatine peut avoir une infinité de formes anatomiques, leur combinaison augmente encore cette variabilité. Chaque patient est un cas particulier qui justifiera une prise en charge spécifique adaptée à sa forme particulière.

La cause exacte des fentes n'est pas claire mais nous savons que les facteurs génétiques et environnementaux peuvent jouer un certain rôle. Cependant, dans la plupart des cas, il n'y a pas de cause connue. C'est pourquoi la "faute" ne doit pas être incriminée, ni à l'un, ni à l'autre des parents.

Les cas légers peuvent consister en une encoche du palais mou, situé au fond de la bouche. Chez les cas modérés, la fente est un peu plus longue et s'étend du fond de la bouche à la jonction des palais mou et dur. Dans les cas graves, la fente couvre tout le plafond de la cavité buccale.

La chirurgie de la division vélo-palatine va viser à rendre au palais son anatomie et sa fonction.

L'évaluation perceptuelle est souvent utilisée et les mesures objectives HNR, jitter et Shimmer n'ont pas été largement explorées chez les enfants atteints de FLP. Par conséquent, l'évaluation de ces paramètres vocaux nous permet de mieux comprendre et diagnostiquer les difficultés d'élocution des enfants atteints de fente labiale et palatine [67].

**CHAPITRE 3 : ANALYSE ACOUSTIQUE DE LA PAROLE
PATHOLOGIQUE**

3.1 Introduction

Actuellement, l'évaluation perceptive (à l'oreille) reste le standard utilisé dans la pratique clinique pour le diagnostic et le suivi thérapeutique des patients. Cette approche est néanmoins reconnue comme étant subjective et coûteuse en temps et ses limites la rendent inadaptée pour le suivi de l'évolution des patients ayant une parole pathologique.

Face à ces limites, les professionnels expriment constamment leur besoin de méthodes objectives d'évaluation de la parole pathologique. L'analyse objective de la voix a été rendue possible grâce à l'avènement de l'outil informatique. Les outils de Traitement Automatique de la Parole (TAP) ont été rapidement considérés comme des solutions potentielles pour répondre à cette demande. Actuellement, divers logiciels permettent une analyse et une représentation de différents paramètres vocaux et peuvent être utilisés tant en clinique qu'en recherche. La mesure de ces paramètres vocaux permet de faire des comparaisons entre les Paroles Pathologiques (PP) et de suivre l'évolution chez des locuteurs pathologiques.

La présente étude tend à souligner d'éventuels traits caractéristiques de la production de la parole chez des locuteurs opérés de fente palatine, et ce en fonction de l'âge, du sexe (Homme, Femme) et du type de pathologie. Nous allons faire une étude acoustique de ce type de PP et essayer de trouver les seuils de cette PP par rapport à la Parole Normale (PN). Nous présentons la démarche suivie pour effectuer les analyses statistiques.

Rappelons que les enfants atteints par cette pathologie ont des difficultés à produire certains sons, et en particulier les occlusives. Du fait de leur fente palatine, ces locuteurs souffrent d'insuffisance vélaire. En conséquence, l'étude de la production de la parole dans les cas d'insuffisance vélaire est particulièrement intéressante. Notre étude étant axée directement sur ce phénomène, nous nous sommes intéressés à faire l'étude de la prononciation du phonème emphatique [t̪] chez des locuteurs porteurs de fente palatine avant et après la rééducation orthophonique. En effet, tel que nous le voyons dans la présente étude, la fente palatine empêche le locuteur d'atteindre une pression orale nécessaire à la production de cette consonne emphatique. Ces locuteurs auront, donc, recours à diverses articulations compensatoires et chacun d'eux aurait des stratégies articulatoires adaptées à son anatomie. Cela devrait nous permettre de justifier notre prédilection pour l'étude de l'occlusive emphatique [t̪] et, de manière préférentielle, son analyse en contexte des voyelles antérieures.

3.2 Analyse acoustique de la parole pathologique

L'analyse acoustique de la voix peut fournir des données instrumentales concernant les anomalies vocales. Ces résultats peuvent être utilisés pour surveiller l'évolution clinique en cas de troubles de la voix [71].

Les enfants atteints de Fente Labio-Palatine (FLP) présentent souvent diverses anomalies de la parole, dues à des anomalies structurelles. Il existe des études approfondies sur l'évaluation des aspects d'articulation et de résonance de la parole et très peu d'études se sont concentrées sur les paramètres de la voix chez les enfants atteints de FLP [62].

Notre étude vise à explorer les traits discriminatifs de la voix de locuteurs présentant une fente réparée. Une analyse acoustique a été réalisée à partir des enregistrements de paroles pathologiques de locuteurs porteurs de différents types de fente comparés à des enregistrements de paroles de locuteurs normaux. Plusieurs indices acoustiques sont relevés pour caractériser la parole produite par les locuteurs porteurs de fente. En plus des paramètres acoustiques classiques tels que la fréquence fondamentale, le suivi des trois premiers formants, la durée de prononciation et l'intensité, d'autres indices sont également pris en compte pour la caractérisation de la qualité de la parole pathologique, car ils permettent une meilleure appréciation du degré de perturbation de la voix pathologique. Parmi ces indices acoustiques, nous avons: le Jitter (mesure du degré de perturbations de la fréquence fondamentale), le Shimmer (mesure du degré de perturbations de l'amplitude) et le HNR (mesure du rapport harmoniques/bruit pour estimer le degré d'altération de la qualité de la voix).

Notre approche de l'analyse du signal acoustique se situe, toutefois, dans une démarche articulatoire-acoustique, à savoir extraire des informations articulatoires de l'analyse du signal acoustique qui nous permet de décrire et évaluer les dysfonctionnements dans ce type de parole pathologique.

3.2.1 Corpus choisis

En effet, les enfants avec fente palatine choisiraient leurs mots en fonction de leur répertoire phonétique. Ils produiraient plus de mots commençant par des nasales et des voyelles et moins de mots avec des occlusives orales. Cela limiterait donc leur vocabulaire en expression [66].

Du fait de leur fente palatine, ces locuteurs pathologiques souffrent d'insuffisance vélaire qui leurs cause des difficultés à produire certains sons, et en particulier les occlusives. En effet, la fente palatine empêche ces locuteurs pathologiques d'atteindre la pression orale nécessaire à la production de cette classe de son. Dans la présente étude, nous nous sommes intéressés à faire l'analyse acoustique et articulatoire du phonème emphatique [t] chez ces locuteurs avant et après la rééducation orthophonique. Rappelons que les sujets ayant une fente palatine auront recours à diverses articulations compensatoires et chacun d'eux aurait

des stratégies articulatoires adaptées à son anatomie. Certaines études ont démontré le lien entre la taille et l'emplacement de la fente et le lieu de l'articulation [45].

Notre corpus est constitué de six mots contenant la consonne emphatique [t̤] combinée dans trois mots avec la voyelle courte [a] et avec la voyelle longue [ā] dans les trois autres, en début du mot (Tableau 3-1).

Tableau 3-1: Corpus utilisé

Le phonème [t̤] suivi par [a]	Le phonème [t̤] suivi par [ā]
[t̤ab̤ba] (طبيبة)	[t̤āwila] (طاولة)
[t̤arīq] (طريق)	[t̤ā'ira] (طائرة)
[t̤amāṭim] (طماطم)	[t̤āwūs] (طاووس)

Nous pourrions par la suite dégager les séquences sonores contenant le phonème emphatique [t̤] combiné avec les deux voyelles [a] et [ā]. Nous essayons ainsi de les analyser et de les comparer avec les mêmes séquences prononcées par des locuteurs sains ayant les mêmes caractéristiques, à savoir le sexe, l'âge et la région (dans la mesure du possible). Les résultats obtenus seront exploités plus tard dans les prochains contrôles pour l'évaluation du patient.

3.2.1.1 Description du phonème emphatique [t̤]

Le phonème [t̤] est une plosive alvéodentale emphatique non aspirante et non voisée. Elle comporte une forte tension des différents organes articulatoires. Son articulation relève d'un phénomène buccal qui consiste en un report en arrière de la racine de la langue provoquant une constriction pharyngale et en un abaissement et un creusement du dos de la langue créant ainsi une augmentation du volume de la cavité buccale. Acoustiquement, Ce travail pharyngal, qui nécessite la mise en œuvre d'une certaine énergie se traduit par l'élévation de la transition de F_1 et la baisse de la transition de F_2 de la voyelle précédente et suivante.

Le phonème [t̤] apparaît sur le spectrogramme comme un *burst* de durée relative 20-30 ms, plus long avec les voyelles longues et qui est sous forme d'une barre verticale suivie d'un intervalle sans bruit remarquable, donc plus en durée. Avec [a] et [ā], la concentration est dans la région de F_1 - F_2 entre 1500-2400 Hz. Devant [t̤], le F_2 du [a] est vers 1150-1250 Hz et celui du [ā] est à 1050 -1150 Hz, ce qui différencie sensiblement le [a] du [ā]. Le [a] et [ā] ont un formant F_1 vers 650 Hz [30].

3.2.2 Conditions d'enregistrement

Afin de bien éclairer nos démarches expérimentales, rappelons en priorité les conditions d'enregistrement de nos Locuteurs Pathologiques (LP) et Locuteurs de Contrôle (LC).

Le présent travail a été réalisé en collaboration avec le Service de Chirurgie Infantile à l'hôpital de Beni-Messous à Alger, dans lequel se sont déroulés les enregistrements acoustiques des enfants ayant une fente palatine.

Dans notre étude, il n'a pas été possible, en raison des contraintes hospitalières, d'acquérir des données préopératoires pour les locuteurs porteurs de fente. Les enregistrements n'ont pu être réalisés qu'à partir d'un mois au minimum après l'intervention chirurgicale. En effet, ces enregistrements ont eu lieu dans la salle de consultation des orthophonistes, avec l'accord des parents et de l'enfant. Ces enfants arrivent très tôt le matin pour voir trois médecins différents (chirurgien, spécialistes en ORL et orthophoniste). Par conséquent, ce ne sont donc pas des conditions favorables pour les enregistrements, malgré ce contexte médical contraignant, les locuteurs se sont appliqués pour effectuer correctement les tâches requises. Ces conditions particulières expliquent que certains enfants n'avaient pas réalisé l'enregistrement du corpus complet comme nous l'aurions souhaité et nous n'avons pas voulu insister, afin que les conditions d'enregistrement de données ne soient pas pour eux trop déplaisantes.

Les enregistrements des locuteurs de contrôle ont été effectués sous deux conditions. Des élèves âgés de 5 à 11 ans, de l'école 1^{er} Novembre 1954 à Hammamet -Alger, ont été enregistrés dans une petite salle silencieuse, grâce à l'autorisation de l'administration de l'établissement et des parents des locuteurs. D'autres locuteurs de différents âges, les enfants accompagnés avec leurs parents, ont été enregistrés dans une salle silencieuse au sein du CRSTDLA - Université ALGER 2, tout en respectant les mêmes conditions d'enregistrements.

3.2.3 Matériel utilisé

L'analyse ne saurait se faire que sur un enregistrement de bonne qualité, c'est-à-dire fidèle et avec peu de bruit parasite. Il n'y a pas de consensus sur le positionnement du microphone ou de l'enregistreur par rapport à la bouche de l'enfant, ni de corpus standardisé pour l'évaluation des troubles de la voix ou de l'articulation.

Pour un enregistrement de bonne qualité, c'est-à-dire fidèle et avec peu de bruit parasite, nous avons utilisé le TASCAM DR-05, un enregistreur portable qui permet d'enregistrer des séquences audio en format (.wav) 16 bits avec une fréquence d'échantillonnage de 44100 KHz.

3.2.4 Enregistrements

Les participants aux enregistrements ont été divisés en un groupe clinique de Quarante six locuteurs porteurs de fente palatine et un groupe témoin de quatre vingt neuf Locuteurs de Contrôle.

Les locuteurs pathologiques sont divisés en deux groupes :

- Vingt huit Locuteurs Pris en Charge (LPC) qui ont bénéficié d'une rééducation orthophonique de différentes durées ;
- Dix huit Locuteurs Non Pris en Charge (LNPC) qui n'ont pas bénéficié de rééducation orthophonique.

Les locuteurs pathologiques enregistrés sont classés selon l'âge qu'ils avaient lors de l'enregistrement et selon le type de fente palatine.

Les Locuteurs de Contrôle sont qualifiés de locuteurs sains par opposition aux Locuteurs Pathologiques. Ces locuteurs ne présentent aucun antécédent connu au niveau de la production ou de la perception de la parole. Notons également que les enfants ont été sélectionnés en fonction de leur bonne élocution.

Afin de comparer les différents locuteurs pathologiques avec leurs correspondants sains de la façon la plus précise possible, nous les avons répartis selon le sexe et l'âge, en se basant sur les différentes étapes du développement de la voix et de l'appareil phonatoire chez l'être humain, à savoir durant la période :

- d'enfance ;
- de puberté (adolescence) ;
- de maturité (adulte).

3.2.4.1 Période d'enfance

À 6 ans, la voix du garçon et celle de la fille sont souvent indiscernables. En revanche, les garçons ont des voix plus fortes que celles des filles. La fille est en période de pré-mue vers 7 à 8 ans avec les mêmes changements que pour les garçons, mais un an avant [72].

L'enfant acquiert progressivement les différents phonèmes utiles dans sa langue maternelle. L'âge d'acquisition est très variable d'un enfant à l'autre, mais en moyenne les voyelles sont acquises à 3 ans, les consonnes occlusives à 4 ans, les liquides à 6 ans et les fricatives à 7 ans [21]. Quant à l'installation du rôle phonétique du voile du palais, la mise en place des points d'articulation est entre 2 ans à 6 ans [73]

3.2.4.2 Période de la puberté

La puberté correspond à l'entrée de l'enfant dans l'adolescence. Elle se signale par l'action d'hormones provoquant de nombreuses transformations telles que la maturation des organes génitaux, apparition de poils et, une des étapes les plus remarquées, le changement du timbre de la voix appelé " MUE " [25].

La MUE est dépendante du reste des transformations due à la puberté. Elle apparaît chez les garçons vers 12 - 14 ans , elle dure 6 mois à un an et apparaît chez les filles vers 9 à 11 ans et peut durer jusqu'à 3 ans [25].

Avant la puberté (12 à 18 ans pour les garçons, 9 à 16 ans pour les filles), les différences vocales entre filles et garçons sont discrètes, situées plutôt au niveau du timbre de la voix. A partir de la puberté, la voix affirme une différenciation sexuelle. Les modifications de la voix sont plus nettes chez le garçon que chez la fille au moment de la puberté [24].

3.2.4.3 Age d'adulte

L'évolution de la morphologie et de la taille de leurs cavités s'arrête vers 20 ans et permet de fixer la majorité des caractéristiques de la voix qui aura évolué en parallèle [74]. De l'adolescence au début de la vieillesse, la voix varie peu [25].

Ainsi, nous présentons dans le Tableau 3-2 les locuteurs pathologiques et les locuteurs sains qui ont participé aux enregistrements.

Tableau 3-2: Présentation des locuteurs pathologiques et sains ayant participé aux enregistrements

Groupe	Féminin			Groupe	Masculin		
	Loc Path		Loc de Contrôle		Loc Path		Loc de Contrôle
	PC	NPC			PC	NPC	
Enfants 06 à 08 ans	04 05 à 08 ans	06 2,5 à 06 ans	16 05 à 08 ans	Enfants 06 à 12 ans	05 05 - 10 ans	11 05 à 11 ans	17 05 à 11 ans
Adolescents 09 à 18 ans	07 09 à 13 ans	-	16 09 à 13 ans	Adolescents 13 à 18 ans	03 14 ans	-	12 13 à 17 ans
Adultes 19 ans et plus	05 19 à 37 ans	-	30 19 à 39 ans	Adultes 19 ans et plus	04 25 - 29 ans	-	12 19 à 31 ans

Pour les Locuteurs de Contrôle de la classe enfant, nous avons pu avoir l'accord de la direction de l'école 1^{er} Novembre 1954 à Hammamet -Alger, pour faire l'enregistrement avec les élèves. On nous a réservé une petite salle pour faire les enregistrements dans des conditions plus au moins favorables et semblables à celles de la salle où les enregistrements ont été réalisés à l'hôpital, avec les locuteurs pathologiques. Pour les adolescents et les adultes, les enregistrements ont été réalisés au sein du CRSTDLA dans une salle qui répond aux mêmes conditions d'enregistrements de la parole pathologique.

3.3 Mesure et analyse acoustique des enregistrements

Différents événements acoustico-articulatoires sont visibles dans le signal acoustique. A partir de ce dernier il est possible de déceler des événements phonétiques permettant de remonter aux configurations articulatoires [75].

Toutes les mesures ont été faites à l'aide du logiciel " Praat " [76].

3.3.1 Mesures et analyses des durées

Pour chaque séquence Consonne Voyelle [CV], nous avons mesuré les durées des :

- VOT (Voice Onset Time) ;
- Syllabes [CV] ;
- Voyelles [V].

Pour les occlusives, l'indice principal est le VOT (Voice Onset Time), qui est l'intervalle séparant le début de l'explosion du début des vibrations des cordes vocales. Selon D. Klatt, le VOT ou délai d'établissement du voisement est défini comme la durée allant du relâchement consonantique jusqu'à l'apparition de la structure formantique stable de la voyelle subséquente [77]. La mesure retenue va de l'explosion consonantique au début de la structure formantique stable de la voyelle suivante. Il est à noter que le VOT défini par Klatt est toujours positif (Figure 3-1).

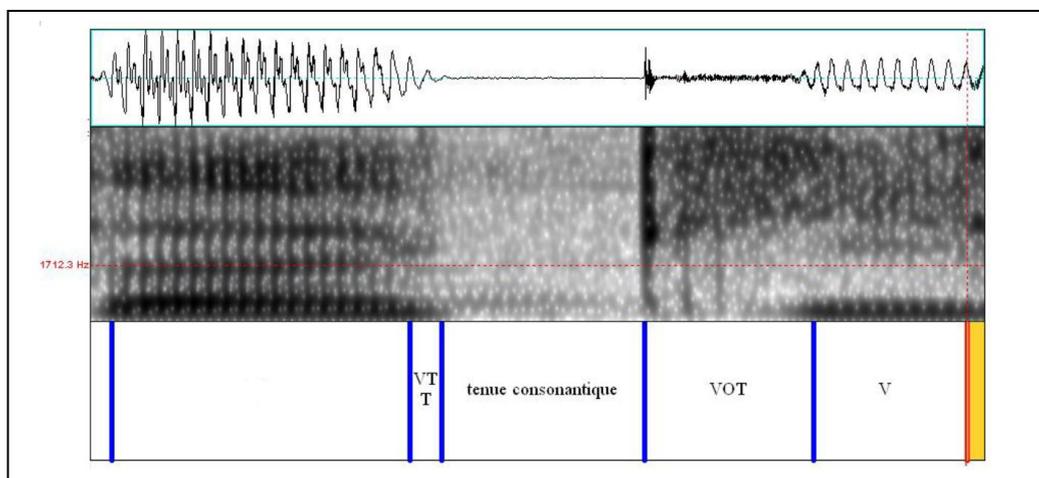


Figure 3-1: Illustration des mesures du VOT

Pour la durée absolue de la voyelle, elle a été quantifiée entre VVO et VVT, autrement dit entre respectivement le début et la fin d'une structure formantique vocalique clairement définie.

3.3.2 Mesures et analyses formantiques

Nous avons modifié les différents paramètres attribués par défaut à l'aide du logiciel Praat. Ces paramètres sont :

- le nombre de formants qui doit être supérieur à celui nécessaire à l'étude pour un meilleur suivi. Nous avons besoin de trois formants seulement, mais nous avons choisi de monter le nombre jusqu'à 5 ;
- la fréquence formantique maximale (Maximum du formant) qui correspond à l'étendue du spectre que l'on veut couvrir. Elle dépend du nombre de formants sélectionnés et du locuteur. Selon les recommandations de Sidney Wood, chez un locuteur masculin, un formant est approximativement détecté dans chaque bande de 1000 Hz environ. Par conséquent, si l'on veut observer 5 formants, il faut élever la fréquence maximale du spectre à 5000 Hz environ. Pour un sujet féminin, il faut augmenter de 10% cette

valeur et de cette manière à étendre la fréquence maximale spectrale à 5500 Hz. Compte tenu du nombre de formants analysés et des informations que l'on possède sur F_0 et les valeurs formantiques chez l'enfant nous avons choisi une valeur maximale de 8000 Hz (environ un formant dans chaque bande de 1600 Hz) ;

- La taille de la fenêtre (Window length) spécifie la fenêtre du signal utilisée pour calculer chaque point. En d'autres termes, elle pourrait correspondre à la résolution temporelle. La valeur proposée par défaut est 0.025 s et a rarement besoin d'être modifiée. Une taille de fenêtre trop petite peut donner des valeurs formantiques fausses, par précaution nous avons augmenté la taille de cette fenêtre à 0.04 s.

Pour chaque voyelle adjacente à la consonne étudiée, nous avons relevé à l'aide du logiciel Praat, au milieu de sa partie stable, les valeurs suivantes :

- la fréquence fondamentale F_0 ;
- l'énergie moyenne E_0 ;
- les trois premiers formants F_1 , F_2 et F_3 .

3.3.3 Extraction des paramètres acoustiques

Une analyse acoustique a été réalisée sur la parole de deux groupes de locuteurs porteurs de fente palatine : le premier contient des locuteurs en poste opératoire et n'ayant pas encore bénéficié de rééducation orthophonique et le deuxième contient des locuteurs porteurs de fente palatine mais qui ont été pris en charge dans un service orthophonique avec des durées différentes de rééducation. Cette analyse a pour objectif essentiel la caractérisation physico-acoustique de ce type de PP, permettant une meilleure connaissance des paramètres pertinents de discrimination de ce type de PP.

Le signal acoustique a été segmenté manuellement, en utilisant l'outil d'analyse Praat, suivant la forme du signal et du spectrogramme correspondant. Ces fichiers sont exploités pour l'extraction des paramètres acoustiques.

3.3.3.1 Cas des Locuteurs Pathologiques Non Pris en Charge pour une rééducation orthophonique (LPNPC)

Les locuteurs porteurs de fente palatine n'ayant pas bénéficié de rééducation orthophonique LPNPC, se trouvent face à l'impossibilité de produire certains sons. Pour y remédier, ils vont le plus souvent avoir recours à des productions compensatoires. Nous allons détailler celles-ci dans cette section.

Une articulation compensatoire est une production réalisée par le locuteur en compensation d'une consonne qu'il ne parvient pas à produire. Il tentera alors de prononcer le son qui sera le plus proche possible de la cible recherchée, sur le plan phonétique.

Pour les LPNPC, nous avons constaté que les problèmes résultants dans la production de la parole sont des complications d'ordre articulatoire et de nasalité. De plus, le velum, qui joue

un rôle essentiel dans l'orientation du flux d'air expiratoire lors de la phonation, ne fonctionne pas. Le dysfonctionnement et l'absence de fermeture efficace de l'espace vélo-pharyngé ont pour conséquence un passage de l'air expiratoire simultanément par la bouche et par le nez, qui apporte un degré plus ou moins important d'hypernasalité et un manque de pression intrabuccale dans les productions de la parole.

Après avoir présenté la démarche suivie pour faire les enregistrements avec nos locuteurs pathologiques, l'analyse acoustique des enregistrements des LPNPC nous a permis de constater que :

- les problèmes d'acquisition et de production des enfants porteurs de fente palatine s'expriment à tout âge de la croissance du locuteur, la diversité de ces problèmes s'étend aussi bien au niveau articulaire qu'aérodynamique ;
- la perturbation induite par un dysfonctionnement au niveau du palais a une incidence directe sur la mise en place des bases phonologiques ;
- l'agencement de l'occlusive [t] est différent à l'intérieur de la cavité buccale, entre les locuteurs sains et les locuteurs pathologiques. L'organisation de l'espace, liée à la position des articulateurs, diffère d'où ces locuteurs pathologiques produisent davantage d'erreurs articulaires dues au déplacement du lieu d'articulation de la consonne qui est difficile à articuler ;
- les locuteurs pathologiques usent de processus phonologiques allant vers la simplification. Ils modifient ainsi la structure des mots et des syllabes par suppression, ou ils substituent des classes de sons à d'autres et assimilent des sons à d'autres.
- les processus phonologiques, identifiés comme étant utilisés par les locuteurs pathologiques, incluent : la nasalité, les assimilations vélaires, les pertitions nasales et les insertions glottales. En fonction de la localisation de la fente dont ils ont été opérés, nos locuteurs pathologiques procèdent régulièrement à :
 - ✓ l'antériorisation des sons due à un contact lingual plus en avant dans la cavité buccale pour les locuteurs porteurs de fente palatine complète qui substituent le phonème [t] par [t] ;
 - ✓ la postériorisation des sons due à un contact lingual plus en arrière dans la cavité buccale pour les locuteurs porteurs de fente labio - palatine complète qui substituent le phonème [t] par [q] ;
 - ✓ le remplacement du son voulu par un autre, plus facile à prononcer en procédant à un changement du mode articulaire. En effet, l'occlusive [t] est remplacé par un son glottique [ʔ] ;
 - ✓ ou encore une pression inadéquate pour la production de certains sons.
- Certains sujets contractent leurs pharynx ou leurs larynx, ce qui se traduit par un bruit particulier : le souffle rauque.
- des phénomènes de compensation au niveau du visage sous forme de syncinésies ou mouvements de la musculature faciale.

- de façon générale, nous avons relevé une difficulté de prononciation des sons emphatiques de l'AS. Ces sons sont soit ignorés, soit prononcés amputés du trait de l'emphase.

3.3.3.1.1 Substitution du phonème [t̤] par [ʔ]

La glottale [ʔ] (hamza), appelée aussi coup de glotte, est décrite comme une plosive glottale non voisée. Pendant son articulation, il se produit une très importante constriction au niveau du laryngo-pharynx, entre la paroi pharyngale postérieure et la racine de la langue à la base de laquelle, dans un mouvement ascendant, suivi de l'accolement de l'épiglotte à la racine de la langue (cette dernière reste solidaire de la racine de la langue) (Figure 3-2).

La glottale [ʔ] présente un rétrécissement maximal au niveau du tube du larynx et une forte constriction au niveau racine- épiglotte. La configuration de la cavité pharyngale se présente sous forme de cloche [30] (Figure 3-2).

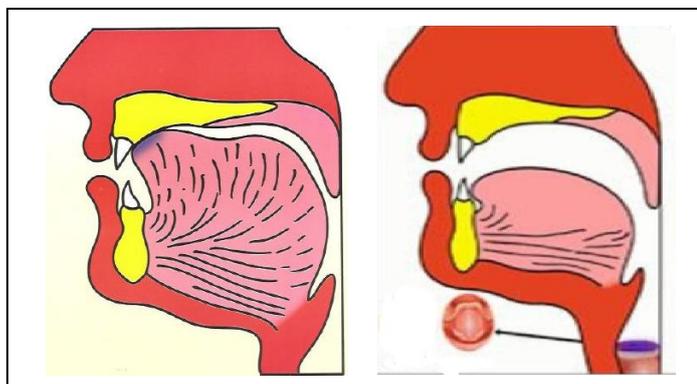


Figure 3-2 : Comparaison entre l'articulation du phonème [t̤] à gauche et du phonème [ʔ] à droite

La structure acoustique du phonème [ʔ] est très dépendante du contexte de production et de sa position à l'intérieur du mot. Initialement, la glottale [ʔ] apparaît sur le spectrogramme sous forme variée. Dans quelques cas, elle est sous forme d'un burst suivi d'un petit intervalle de silence de durée 15-20 ms ou bien suivi d'un faible bruit [78] (Figure 3-3).

En présence des voyelles [a] et [ā], le [ʔ] apparaît comme un burst de durée 20-30 ms dans la région de fréquences hautes 1500-1700 Hz. Les mesures des formants du [a] et [ā] sont respectivement : $575 < F_1 < 650$ Hz, $1180 < F_2 < 1300$ Hz et $2300 < F_3 < 2400$ Hz [78].

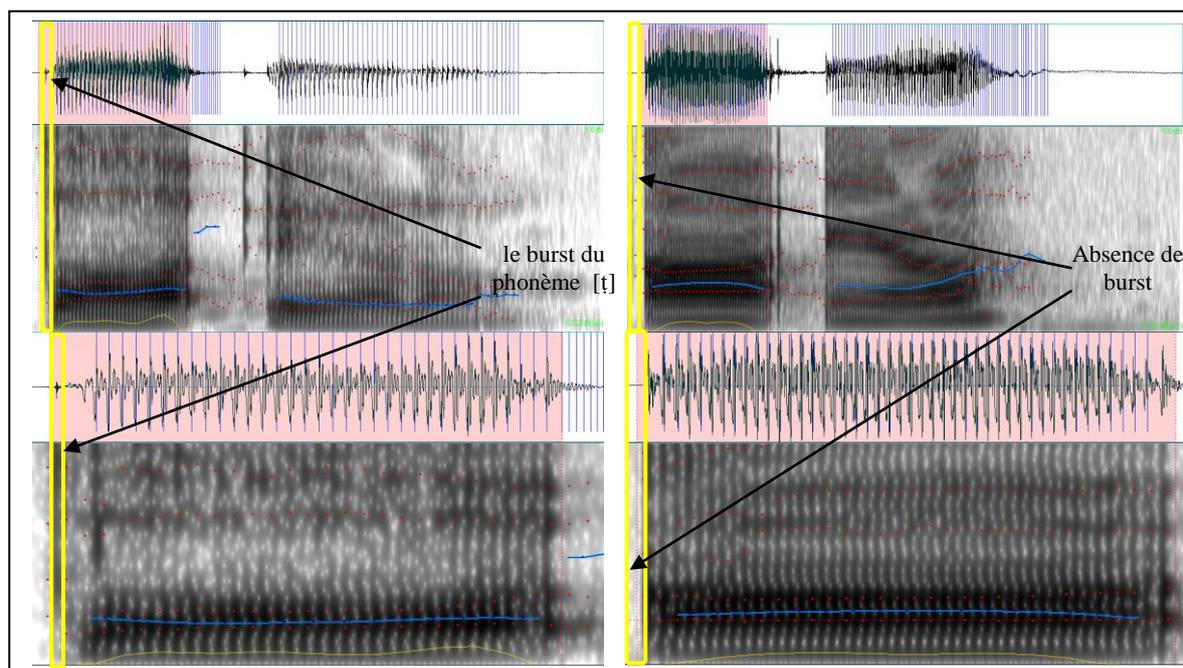


Figure 3-3 : Substitution du phonème [t] par [ʔ] dans [tā'ira] prononcé par des locutrices âgées de 19 ans

3.3.3.1.2 Substitution du phonème [t] par [q]

Les locuteurs ayant une fente labio-palatine, gênés par la fente qui se situe au niveau antérieur de la cavité buccale, vont utiliser davantage la masse postérieure de leurs langues, entraînant d'avantage de production des sons à l'arrière de la cavité buccale, évitant ainsi au maximum toute production antérieure.

Le phonème [q] [qaf] est une plosive uvulaire non voisée et non aspirante. Son articulation se fait avec le dos de la langue contre la luette. Il se produit une constriction pharyngale initiale, dans un mouvement circulaire de la langue qui aboutit à un contact uvulaire. La largeur du contact est par ailleurs nettement étendue. L'occlusion entraîne une accentuation de la constriction pharyngale. le travail du voile du palais et de l'uvule se montre compact pour la consonne [q] et l'accolement se fait avec une force très visible pour cette consonne (Figure 3-4).

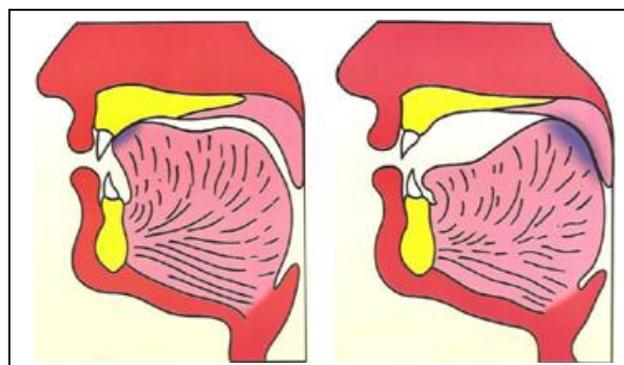


Figure 3-4 : Comparaison entre l'articulation du phonème [t] à gauche et du phonème [q] à droite

Sur le plan acoustique, la consonne [q] apparaît comme un burst important qui commence en basses fréquences et qui monte à 3000 Hz. Celui-ci est suivi d'un petit intervalle de silence de durée moyenne de 30-40 ms sans bruit, ce qui indique l'absence d'aspiration. La consonne [q] produit un début de F_2 du [a] très bas, autour de 1200 Hz, celui du [ā] est à environ 1150 Hz, par contre les débuts de F_1 sont légèrement relevés sous l'influence du [q] (figure 3-5). En moyenne, la durée relative du silence du [q] est d'environ 130 ms [78].

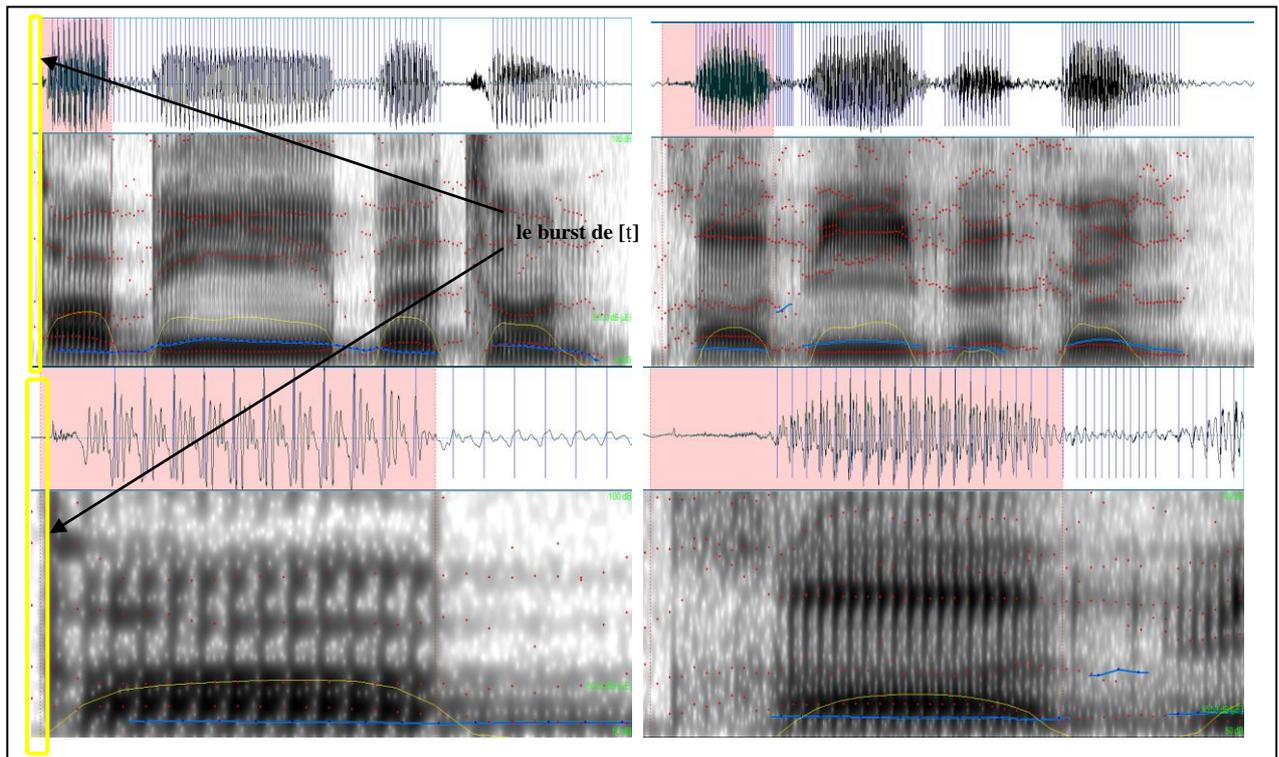


Figure 3-5 : Substitution du phonème [t] par [q] dans [tabība] prononcé par des locuteurs âgés de 29 ans (sain à gauche, pathologique à droite)

3.3.3.1.3 Substitution du phonème [t] par [q]

La consonne [t] est une plosive aspirante alvéodentale non voisée qui se réalise par un blocage de l'air avec la pointe de la langue derrière les incisives supérieures. Pendant l'articulation de la consonne [t], le niveau du voile contracté se situe plus haut et l'accolement rhino-pharyngal plus en arrière. Ce recul de la paroi se généralise sur toute sa longueur. A cela s'ajoute un mouvement considérable vers l'avant de la racine de la langue, sans pour autant influencer sa partie antérieure. De cet avancement de la racine et le recul de la paroi-pharyngale résulte une large cavité pharyngale (Figure 3-6).

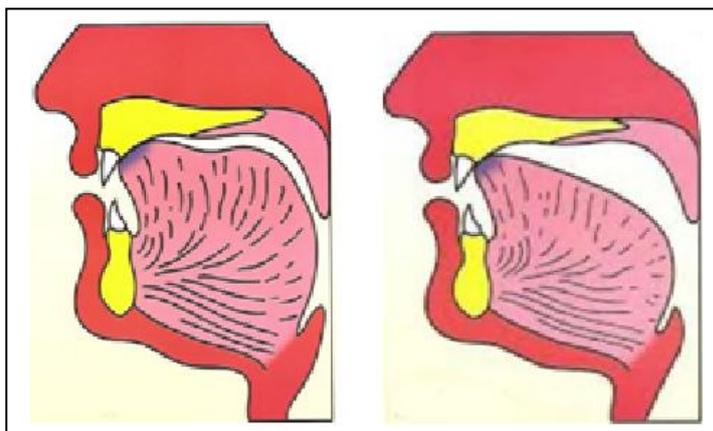


Figure 3-6: Comparaison de l'articulation du phonème [t̥] à gauche et du phonème [t] à droite

En comparant les deux profils de la phase centrale de la réalisation de chacune des deux consonnes [t̥] et [t] (Figure 3-7), nous observons une même cavité antérieure et un pharynx qui constitue le lieu de différenciation incontestable entre une emphatique et une non emphatique. La consonne [t̥] est réalisée avec un pharynx rétréci, [t] avec un pharynx large. La cavité pharyngale est réduite presque de moitié pour l'emphatique [t̥] à cause du recul de toute la racine et l'avancement de la paroi-pharyngale. L'accolement apical et le contour de la langue allant de l'apex à la partie dorsale, sont identiques.

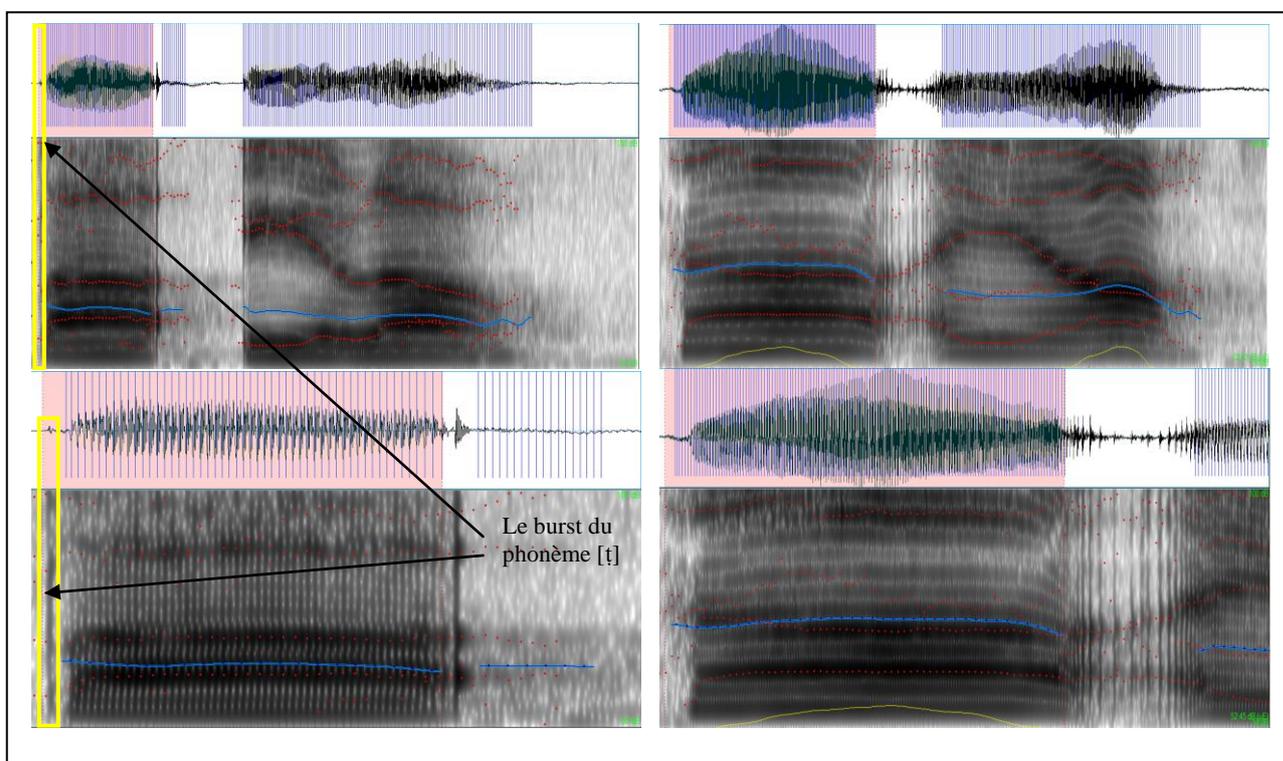


Figure 3-7 : Substitution du phonème [t̥] par [t] dans [t̥ā'ira] prononcé par deux locuteurs âgés de 06 ans (sain à gauche, pathologique à droite)

3.3.3.2 Cas de Locuteurs Pathologiques Pris en Charge pour une rééducation orthophonique (LPPC)

La rééducation orthophonique est dans tous les cas indispensable afin d'obtenir un fonctionnement correct du voile du palais qui permettrait de lutter contre la déperdition nasale et éviter le nasonnement, particulièrement audible dans l'émission des voyelles. Il s'agit d'exercices de souffle dirigé et d'exercices de mobilisation du voile. Tout le problème est d'arriver à faire percevoir au patient les mouvements vélares.

Les locuteurs porteurs de fente palatine doivent, en conséquence, avoir recours à certaines stratégies de réajustement, et usent le plus souvent de manœuvres compensatoires pour produire des sons qui, même s'ils sont relativement «faux» par rapport à une «cible» articulaire habituelle, se rapprochent au maximum des propriétés acoustiques des sons désirés et sont, en conséquence, perçus comme «justes» par l'auditeur.

En effet, chaque locuteur est contraint d'adapter ses gestes articulatoires en fonction de la configuration propre de sa cavité buccale. Rappelons que selon la forme du conduit vocal, un changement de la position de la langue entraînerait plus ou moins de variations acoustiques [79].

Dans cette partie, nous avons fait une analyse comparative entre une parole normale de référence de Locuteurs de Contrôle (LC) et une parole pathologique de Locuteurs Pathologiques (porteurs de fente palatine) Pris en Charge (LPPC) à travers des séances périodiques de rééducation orthophonique (Tableau 3-3).

En addition aux paramètres acoustiques ordinaires tels que la fréquence fondamentale, les formants et l'énergie (intensité), nous avons mesuré, pour chaque occurrence, la durée [V] de la voyelle [a] qui suit la consonne [t], en prenant pour repères l'apparition et la disparition d'une structure formantique clairement définie. Pour l'occlusive sourde [t], nous avons mesuré la durée [CV] et le VOT, celui de D. Klatt [80], qui va de l'explosion acoustique, due au relâchement, à l'apparition de la structure formantique claire de la voyelle subséquente. D'autres indices sont également mis en valeur, car ils permettent de discriminer la PP de la PN. Ces indices acoustiques sont le Jitter, le Shimmer et le HNR.

Tableau 3-3 : Analyse acoustique des enregistrements de la locutrice pathologique N°35

Type de Voyelle	Code Patient	VOT	CV	V	E ₀	F ₀	F ₁	F ₂	F ₃	Jitter	Shimmer	HNR
	4035F0041	0,0049	0,5286	0,4633	58,21	267,35	1115	1657	1974	0,299	1,649	22,434
	4035F0042	0,0055	0,4862	0,3743	59,399	265,7	1214	1804	2334	0,315	2,978	20,251
	4035F0043	0,0049	0,3708	0,2764	59,741	262,07	1174	1775	2510	0,242	2,481	22,463
	4035F0771	0,0082	0,53	0,4434	55,108	250,87	1070	1501	2504	0,374	2,185	20,783
[VL]	4035F0772	0,0077	0,4653	0,4269	54,453	245,35	1149	1483	2701	0,411	2,213	20,220
	4035F0773	0,0054	0,4858	0,3943	55,174	249,09	1182	1435	2480	0,327	2,013	22,188
	4035F1421	0,0114	0,5367	0,4785	59,526	268,05	1180	1651	2457	0,293	2,268	22,658
	4035F1422	0,0057	0,5386	0,5014	58,351	268,17	1175	1542	2461	0,267	1,766	23,696
	4035F1423	0,0057	0,4225	0,3756	58,354	278,54	1206	1671	2685	0,388	2,498	21,168
	4035F1971	0,0088	0,273	0,232	58,089	262,38	1210	1761	2544			
	4035F1972	0,0091	0,2343	0,2031	58,672	255,5	1197	1760	2521			
	4035F1973	0,0114	0,1882	0,1555	57,016	264,85	1188	1880	2724			
	4035F1711	0,0244	0,1411	0,0884	55,751	262,44	1140	1814	2654			
[VC]	4035F1712	0,0901	0,2027	0,0924	53,808	251,84	1099	1738	2481			
	4035F1713	0,0959	0,1983	0,0626	53,947	253,09	1115	1752	2472			
	4035F0221	0,0201	0,1513	0,0764	55,158	240,05	1272	1865	2540			
	4035F0222	0,0205	0,1298	0,0386	55,299	243,41	1249	1664	2533			
	4035F0223	0,0116	0,1132	0,0682	55,861	258,37	1169	1772	2654			

3.4 Etudes statistiques des résultats de l'analyse acoustique

L'analyse des résultats a été procédée de la manière suivante.

3.4.1 Comparaison des paramètres acoustiques étudiés des locuteurs sains et pathologiques

Les valeurs des paramètres étudiés obtenues de l'analyse acoustique sont présentées en moyenne \pm écart type. Nous comparons les moyennes des valeurs des paramètres acoustiques étudiés de la parole pathologique avec celles de la parole saine et ce pour les mêmes classes Masculin et Féminin.

Le test utilisé est celui de U Mann Whitney qui consiste à comparer deux moyennes de variables quantitatives à distribution non gaussienne. Dans le cadre de l'apprentissage supervisé, il convient surtout parce qu'il est en relation avec le critère AUC (Area Under Curve) associé à la courbe ROC [81].

La distribution normale a été vérifiée soit graphiquement soit par le test de Shapiro-Wilk.

3.4.1.1 Classe d'enfants

Les résultats obtenus de l'analyse acoustique des enregistrements réalisés par les locuteurs de la classe enfants Masculin et Féminin sont représentés dans les tableaux 3-4 et 3-5.

Tableau 3-4: Valeurs des paramètres acoustiques de la PP et PN de la classe (Enfant - M)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	Nbre de Locuteurs	N° de mots	Moy	Ecart-type	Nbre de Locuteurs	N° de mots	Moy	Ecart-type
[VC]	VOT	5	23	0000,015	0,006	20	184	0000,012	000,006
	CV	5	23	0000,168	0,056	20	184	0000,126	000,031
	V	5	23	0000,104	0,043	20	184	0000,082	000,027
	E ₀	5	23	0054,593	5,422	20	184	0054,096	005,954
	F ₀	5	23	0282,334	26,696	20	184	0284,403	053,449
	F ₁	5	23	0858,826	170,04	20	184	0906,967	127,651
	F ₂	5	23	1399,956	174,296	20	184	1459,701	202,794
	F ₃	5	23	2884,304	734,279	20	184	2869,858	723,802
[VL]	VOT	5	22	0000,023	000,025	20	162	0000,013	000,0052
	CV	5	22	0000,259	000,081	20	162	0000,194	000,0550
	V	5	22	0000,165	000,053	20	162	0000,129	000,0520
	E ₀	5	22	0058,014	004,506	20	162	0056,814	005,1680
	F ₀	5	22	0278,257	054,563	20	162	0289,774	048,2790
	F ₁	5	22	0964,227	121,594	20	162	0957,371	128,0990
	F ₂	5	22	1405,091	150,671	20	162	1502,241	177,9160
	F ₃	5	22	3098,091	553,728	20	162	3378,111	374,0970
	Jitter	5	22	0000,831	001,035	20	162	0000,530	000,3390
	Shimmer	5	22	0005,268	001,752	20	162	0004,171	002,2340
HNR	5	22	0018,371	003,224	20	162	0018,920	003,7160	

Tableau 3-5 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Enfant - F)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	Locuteurs	N ^{bre} de mots	Moy	Ecart-type	Locuteurs	N ^{bre} de mots	Moy	Ecart-type
[VC]	VOT	4	24	0000,026	000,024	16	112	0000,013	000,007
	CV	4	24	0000,178	000,038	16	112	0000,144	000,037
	V	4	24	0000,113	000,045	16	112	0000,101	000,041
	E ₀	4	24	0052,534	003,877	16	112	0053,158	008,213
	F ₀	4	24	0254,156	009,451	16	112	0287,881	059,734
	F ₁	4	24	1122,667	144,312	16	112	0957,893	213,019
	F ₂	4	24	1799,500	196,755	16	112	1532,813	285,338
	F ₃	4	24	3069,625	629,643	16	112	2763,786	838,875
[VL]	VOT	4	35	0000,021	000,0172	16	88	0000,015	000,008
	CV	4	35	0000,417	000,1070	16	88	0000,247	000,064
	V	4	35	0000,309	000,1180	16	88	0000,155	000,054
	E ₀	4	35	0058,397	005,3390	16	88	0055,907	007,467
	F ₀	4	35	0277,861	038,5370	16	88	0288,148	042,521
	F ₁	4	35	1141,772	087,9040	16	88	1003,136	153,192
	F ₂	4	35	1650,657	151,0800	16	88	1584,807	203,224
	F ₃	4	35	3291,028	916,7940	16	88	3445,012	652,184
	Jitter	4	35	0000,355	000,0960	16	88	0000,442	000,364
	Shimmer	4	35	0003,403	001,5260	16	88	0003,985	001,918
HNR	4	35	0019,894	002,8940	16	88	0019,717	002,555	

D'après les tableaux 3-4 et 3-5 on peut noter que :

- les durées de [CV] et [V] sont plus élevées pour les cas pathologiques ;
- le VOT est plus important pour les cas pathologiques masculins et féminins ;
- la fréquence fondamentale est presque égale pour les cas normaux des deux sexes. Elle est nettement inférieure pour les cas pathologiques féminins, pour les [VC] ;
- le premier formant est plus élevé chez les cas pathologiques féminins ;
- le deuxième et le troisième formant F_2 , F_3 sont inférieurs chez les cas pathologiques masculins pour les voyelles longues ;
- le deuxième formant F_2 est plus élevé pour les cas pathologiques féminins dans la réalisation des voyelles courtes ;
- la valeur de l'énergie est légèrement élevée pour les cas pathologiques féminins dans la réalisation des voyelles longues ;
- pour les paramètres pathologiques, nous avons relevé une valeur du Shimmer légèrement élevée pour les cas pathologiques masculins ;
- HNR est non significatif.

3.4.1.2 Classe d'adolescents

Les résultats obtenus de l'analyse acoustique des enregistrements réalisés par les locuteurs de la classe Adolescents Masculin et Féminin sont représentés dans les tableaux 3-6 et 3-7.

Tableau 3-6 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adolescent - M)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	N ^{bre} Locuteurs	N ^{bre} de mots	Moy	Ecart-type	N ^{bre} Locuteurs	N ^{bre} de mots	Moy	Ecart-type
VC	VOT	3	18	0000,011	000,003	12	90	0000,022	000,022
	CV	3	18	0000,133	000,055	12	90	0000,114	000,025
	V	3	18	0000,084	000,047	12	90	0000,073	000,022
	E_0	3	18	0061,936	008,164	12	90	0052,423	004,992
	F_0	3	18	0241,455	024,232	12	90	0164,886	045,973
	F_1	3	18	0938,888	157,405	12	90	0695,667	087,724
	F_2	3	18	1438,111	102,156	12	90	1261,256	086,453
	F_3	3	18	3003,667	445,011	12	90	2642,733	339,667
VL	VOT	3	18	0000,013	000,006	12	96	000,0169	000,006
	CV	3	18	0000,239	000,088	12	96	0000,183	000,047
	V	3	18	0000,146	000,068	12	96	0000,136	000,039
	E_0	3	18	0064,694	008,805	12	96	0055,487	005,265
	F_0	3	18	0256,616	043,486	12	96	0177,603	047,596
	F_1	3	18	0942,833	174,958	12	96	0712,823	084,303
	F_2	3	18	1294,222	136,196	12	96	1202,833	086,087
	F_3	3	18	3091,444	355,243	12	96	2697,032	351,304
	Jitter	3	18	0000,411	000,403	12	96	0000,783	000,583
	Shimmer	3	18	0003,860	001,908	12	96	0004,386	002,304
HNR	2	18	0022,313	003,016	12	96	0016,985	003,587	

D'après le Tableau 3-6 on peut noter que :

- le VOT est inférieur notamment pour les voyelles courtes ;
- les durées [CV] et [V] sont plus élevées pour les PP ;
- le pitch F_0 est supérieur pour les deux types de voyelles ;
- les valeurs de l'énergie est légèrement élevées pour les deux types de voyelles ;
- les valeurs des formants F_1 , F_2 et F_3 sont nettement supérieur dans la réalisation des deux types de voyelles ;
- Pour les paramètres pathologiques, nous avons constaté une valeur du shimmer inférieure et une valeur du HNR plus élevée que celle des cas sains.

Tableau 3-7 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adolescent - F)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	N ^{bre} de Locuteurs	N ^{bre} de mots	Moy	Ecart-type	N ^{bre} de Locuteurs	N ^{bre} de mots	Moy	Ecart-type
VL	VOT	7	54	0000,015	000,007	16	148	0000,011	000,004
	CV	7	54	0000,211	000,064	16	148	0000,214	000,064
	V	7	54	0000,106	000,049	16	148	0000,143	000,057
	E_0	7	54	0056,345	004,451	16	148	0059,388	008,139
	F_0	7	54	0278,573	028,678	16	148	0275,186	051,922
	F_1	7	54	0920,722	121,002	16	148	0966,108	116,928
	F_2	7	54	1479,518	146,901	16	148	1511,567	138,433
	F_3	7	54	3213,796	702,223	16	148	3379,845	382,523
	Jitter	7	54	0000,466	000,259	16	148	0000,524	000,579
	Shimmer	7	54	0004,012	001,864	16	148	0003,802	001,955
HNR	7	54	0018,638	003,263	16	148	0019,647	003,646	
VC	VOT	7	49	0000,015	000,006	16	140	0000,0112	0,006
	CV	7	49	0000,141	000,036	16	140	0000,1400	0,028
	V	7	49	0000,084	000,033	16	140	0000,0910	0,029
	E_0	7	49	0055,673	004,372	16	140	0057,4780	7,138
	F_0	7	49	0265,778	025,491	16	140	0258,3870	47,775
	F_1	7	49	0945,021	186,648	16	140	0950,1360	124,044
	F_2	7	49	1533,592	210,083	16	140	1575,9930	168,969
	F_3	7	49	2913,551	865,029	16	140	3296,5720	452,593

D'après le Tableau 3-7 on peut noter que :

- la valeur du VOT est supérieure pour les deux types de voyelles ;
- la durée [CV] est supérieure pour la PP ;
- la durée des voyelles longues est inférieure par rapport au cas sains ;
- le pitch F_0 est stable ;
- E_0 de la PP est légèrement inférieure ;
- le premier formant F_1 est inférieur pour les VL ;
- le deuxième formant F_2 est inférieur pour les VL et les VC
- HNR non significatif.

3.4.1.3 Classe d'adultes

Les résultats obtenus de l'analyse acoustique des enregistrements réalisés par les locuteurs de la classe Adultes Masculin et Féminin sont représentés dans les tableaux 3-8 et 3-9.

Tableau 3-8 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adultes - M)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	N ^{bre} de Locuteurs	N ^{bre} de mots	Moy	Ecart-type	N ^{bre} de Locuteurs	N ^{bre} de mots	Moy	Ecart-type
VC	VOT	4	41	0000,018	000,011	12	107	0000,017	000,007
	CV	4	41	0000,124	000,025	12	107	0000,128	000,034
	V	4	41	0000,068	000,026	12	107	0000,087	000,031
	E ₀	4	41	0056,883	004,876	12	107	0056,572	007,519
	F ₀	4	41	0132,175	023,598	12	107	0133,088	021,929
	F ₁	4	41	0682,488	206,085	12	107	0663,476	062,425
	F ₂	4	41	1357,732	157,012	12	107	1190,906	109,984
	F ₃	4	41	2703,464	443,917	12	107	2565,711	329,892
VL	VOT	4	41	0000,014	000,004	12	98	0000,016	000,005
	CV	4	41	0000,205	000,078	12	98	0000,206	000,067
	V	4	41	0000,143	000,073	12	98	0000,159	000,063
	E ₀	4	41	0058,171	002,613	12	98	0059,223	005,912
	F ₀	4	41	0138,306	017,166	12	98	0142,104	023,624
	F ₁	4	41	0650,028	057,404	12	98	0667,938	083,722
	F ₂	4	41	1236,200	092,617	12	98	1119,214	083,123
	F ₃	4	41	2547,457	324,777	12	98	2679,562	239,292
	Jitter	4	41	0000,754	000,415	12	98	0000,901	000,773
	Shimmer	4	41	0004,804	002,694	12	98	0004,236	001,957
HNR	4	41	0016,003	003,673	12	98	0016,723	004,828	

D'après le Tableau 3-8 on peut noter que :

- le pitch F₀ et l'intensité E₀ sont stables ;
- le VOT est inférieur pour les voyelles courtes ;
- la durée du [CV] est égale entre la PP et la PN ;
- la durée [V] de la PP est inférieure à celle de la PN ;
- le deuxième formant F₂ est plus élevé dans les deux cas VL et VC ;
- le troisième formant F₃ est inférieur dans le cas des VL ;
- HNR stable est non significatif.

Tableau 3-9 : Valeurs des paramètres acoustiques de la PP et PN de la classe (Adultes - F)

Type de Parole		PP				PN			
Type de voyelle	Paramètres Acoustiques	Locuteurs	N ^{bre} de mots	Moy	Ecart-type	Locuteurs	N ^{bre} de mots	Moy	Ecart-type
VC	VOT	5	38	0000,013	000,0060	30	291	0000,013	000,006
	CV	5	38	0000,124	000,0351	30	291	0000,117	000,027
	V	5	38	0000,081	000,0260	30	291	0000,075	000,039
	E ₀	5	38	0054,639	004,6390	30	291	0055,085	007,102
	F ₀	5	38	0223,063	011,8630	30	291	0219,975	035,175
	F ₁	5	38	0877,184	102,0620	30	291	0879,800	100,169
	F ₂	5	38	1379,526	120,9110	30	291	1470,845	120,724
	F ₃	5	38	2615,105	196,0870	30	291	2988,676	250,303
VL	VOT	5	27	0000,014	000,006	30	303	0000,013	000,008
	CV	5	27	0000,179	000,059	30	303	0000,208	000,061
	V	5	27	0000,128	000,052	30	303	0000,144	000,049
	E ₀	5	27	0057,518	004,209	30	303	0056,727	007,544
	F ₀	5	27	0242,872	030,343	30	303	0234,114	045,571
	F ₁	5	27	0864,667	081,607	30	303	0888,732	096,737
	F ₂	5	27	1381,296	125,313	30	303	1388,564	126,818
	F ₃	5	27	2542,592	246,637	30	303	3020,534	257,028
	Jitter	5	27	0000,841	000,758	30	303	0000,502	000,283
	Shimmer	5	27	0003,476	001,348	30	303	0003,416	001,813
	HNR	5	27	0018,625	003,059	30	303	0018,989	003,509

D'après le Tableau 3-9 on peut noter que :

- le pitch F_0 et l'intensité E_0 sont stables ;
- le VOT est égale à celui des cas sains pour les VC et légèrement supérieure pour les VL ;
- les durées des [CV] et [V] de la PP sont nettement inférieures à celle de la PN dans le cas des VL, contrairement au cas des VC où elles sont supérieures à celles de la PN ;
- le troisième formant F_3 est nettement inférieur dans le cas des VL ;
- la valeur du Jitter des cas pathologiques est supérieure par rapport à celle des cas sains
- HNR est non significatif.

3.4.2 Performances analytiques des différents paramètres étudiés

Le dépassement du seuil normal/ pathologique, pour un paramètre donné, est un moyen objectif qui permet de mettre en évidence les altérations vocales et de mesurer le degré de cette altération [82, 83].

Cette étape consiste en la détermination des performances des paramètres acoustiques comme test de diagnostic, qui passe par l'établissement de la courbe ROC qui est un outil graphique qui permet de représenter la capacité d'un test à discriminer entre la population des sujets pathologiques et des sujets sains.

3.4.2.1 Présentation de la courbe roc

Conçu en 1941, le calcul de l'aire sous la courbe ROC (**R**eceiver **O**perating **C**haracteristic) est une approche non paramétrique permettant de mesurer l'association entre une variable continue X et une variable binaire Y .

Soit $X|_{Y=g_1}$ et $X|_{Y=g_2}$ les sous-échantillons de X restreint aux deux groupes de Y , variable binaire. X peut être aussi bien une variable explicative ou le résultat d'une règle décisionnelle (souvent notée Z). On définit la spécificité comme le taux de classification associé aux observations du groupe g_1 ; et la sensibilité celui associé à celles du groupe g_2 (à noter, qu'en fonction des conventions, spécificité et sensibilité peuvent être inversées). La courbe ROC permet alors de présenter l'évolution de la spécificité et la sensibilité en fonction de l'évolution pas à pas du seuil (cutoff) appliqué sur X .

La courbe ROC permet de représenter l'évolution du taux de classification de Y à partir des valeurs de X . En faisant varier un cutoff $c \in [\min(X) ; \max(X)]$, on peut déterminer les deux indicateurs suivants : par l'appréciation de deux lois statistiques, qui sont la sensibilité et la spécificité :

- Sensibilité notée Se_c (aussi appelée Fraction de Vrais Positifs) : proportion d'individus positifs effectivement bien détectés par le test. Autrement dit, la sensibilité permet de mesurer à quel point le test est performant lorsqu'il est utilisé sur des individus positifs ;
- Spécificité notée Sp_c (aussi appelée Fraction de Vrais Négatifs) : proportion d'individus négatifs effectivement bien détectés par le test. Autrement dit, la spécificité permet de mesurer à quel point le test est performant lorsqu'il est utilisé sur des individus négatifs.

Le test est parfait pour les individus (positifs ou négatifs) lorsque la sensibilité vaut 1, équivalent à un tirage au hasard lorsque la sensibilité vaut 0.5. S'il est inférieur à 0.5, le test est contre-performant.

Les définitions mathématiques de la Sensibilité et de la Spécificité sont respectivement :

$$Se(i) = \frac{vp(i)}{vp(i)+fn(i)} \quad (3.1)$$

$$Sp(i) = \frac{vn(i)}{vn(i)+fp(i)} \quad (3.2)$$

- Vrais Positifs (VP) : Nombre d'individus déclarés positifs par le test et qui le sont effectivement.
- Faux Positifs (FP) : Nombre d'individus déclarés positifs par le test mais qui sont en réalité négatifs.
- Vrais Négatifs (VN) : Nombre d'individus déclarés négatifs par le test et qui le sont effectivement.
- Faux Négatifs (FN) : Nombre d'individus détectés négatifs par le test mais qui sont en réalité positifs.

La courbe ROC est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test pour toutes les valeurs seuils possibles. L'ordonnée représente la sensibilité et l'abscisse correspond à la quantité (1 - spécificité) [84] (Figure 3-8).

Ces taux de bonne classification sont basés sur la théorie de Bayes. On est donc dans un plan à deux dimensions, la diagonale passant par les points de coordonnées (0,0) et (1,1) est la droite de référence d'une classification aléatoire. Plus la courbe colle le côté haut-gauche du plan et plus la classification est de bonne qualité.

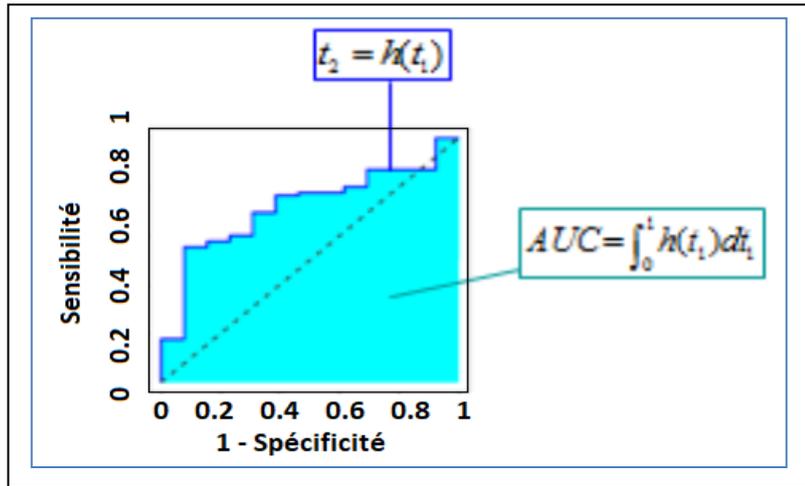


Figure 3-8 : Exemple de courbe ROC

L'aire sous la courbe ROC est alors l'intégrale :

$$AUC = \int_0^1 h(t_1) dt_1 \quad (3.3)$$

Le calcul de l'aire sous la courbe ROC dans le cas binaire, notée *AUC*, se base sur le nombre de paires concordantes/discordantes. Il s'agit d'une mesure empirique qui offre un très bon indicateur de la répartition de *X* par rapport aux deux groupes de la variable *Y*. En effet, dans le cadre d'une liaison entre *X* et *Y*, on s'attend à ce que les valeurs de *X* pour le groupe g_2 de *Y* soient généralement plus fortes que celles pour le groupe g_1 (ou l'inverse). L'*AUC* tient son nom d'indicateur empirique du fait qu'il ne travaille pas sur la forme continue des variables mais plutôt sous un aspect plus complexe de leur rang.

On rappelle, pour $X|_{Y=g_1}$ et $X|_{Y=g_2}$ distributions de *X* restreintes aux deux groupes de *Y*, de tailles respectives n_1, n_2 et i_1, i_2 observations des groupes g_1 et g_2 , une paire est dite :

- concordante, si $(X|_{Y=g_2})_{i_2} > (X|_{Y=g_1})_{i_1}$;
- discordante, si $(X|_{Y=g_2})_{i_2} < (X|_{Y=g_1})_{i_1}$;
- ex-aequo, si $(X|_{Y=g_1})_{i_1} = (X|_{Y=g_2})_{i_2}$.

Le calcul de l'*AUC* se fait alors par la formule :

$$AUC = \frac{1}{n_1 \cdot n_2} \left[\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \mathbf{1}_{(X|_{Y=g_1})_{i_1} > (X|_{Y=g_2})_{i_2}} + \frac{1}{2} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \mathbf{1}_{(X|_{Y=g_1})_{i_1} = (X|_{Y=g_2})_{i_2}} \right]$$

L'*AUC* varie dans $[\frac{1}{2} ; 1]$. Il se peut qu'elle soit inférieure à $\frac{1}{2}$, dans ce cas il faut rectifier sa valeur par $1 - AUC$. Cet effet s'explique par le fait que l'on a supposé que le groupe de référence admet en général des valeurs fortes par rapport à l'autre groupe alors qu'en réalité les données observées tendent vers le contraire (par exemple, on s'attendait à ce que X dans le groupe g_2 ait généralement des valeurs supérieures à celles du groupe g_1 alors que c'est l'inverse).

Lorsque l'*AUC* est autour de $\frac{1}{2}$ cela indique que les distributions de $X |_{Y=g_1}$ et $X |_{Y=g_2}$ sont confondues et que nous sommes dans un cas similaire à un tirage aléatoire de X par rapport à Y (la courbe ROC correspondante sera proche de la diagonale). Plus l'*AUC* s'approche de 1 et plus les deux distributions sont séparées (la courbe tend à ressembler à un carré de côtés de longueur 1). Le tableau 3-10 présente l'interprétation que l'on peut tirer en fonction de la valeur de l'*AUC* :

Tableau 3-10 : Interprétation des valeurs de l'*AUC* [85]

Valeurs AUC	Interprétation du Modèle
0.5	aléatoire
> 0.7	bon
0.87 - 0.9	bien discriminant
0.9 - 1	excellent

La courbe ROC (Receiver Operating Characteristic) représente en ordonnée la proportion de tests positifs parmi la population des sujets pathologiques (la sensibilité) en fonction de la proportion de tests positifs des sujets sains en abscisse (1- spécificité) pour les différentes valeurs seuil [86] (Figure 3-9).

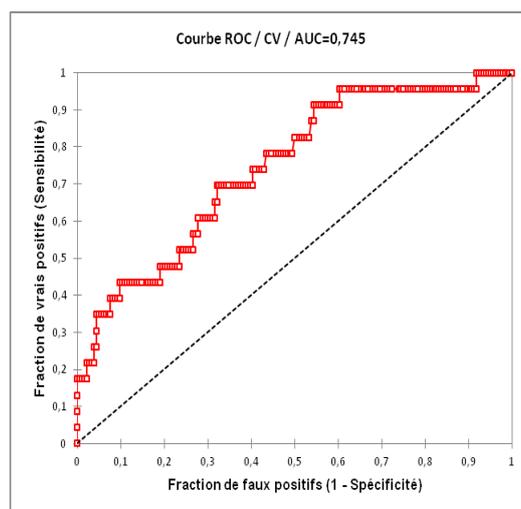


Figure 3-9 : Courbe ROC calculée pour la durée CV (voyelle courte) pour la classe Enfants-Masculins

AUC (Area Under Curve) est un indice calculé qui correspond à la probabilité pour qu'un événement soit classé positif par le test sur l'étendue des valeurs seuil possibles.

3.4.2.2 Résultats

Les résultats de l'application de la courbe ROC sur les différents paramètres acoustiques étudiés sont présentés dans le tableau 3-11.

Tableau 3-11 : Valeurs AUC des différents paramètres étudiés pour les trois Classes - Cas des VC

		1F	2F	3F	1M	2M	3M
VOT (ms)	<i>AUC</i>	0.674	0.714	0.537	0.663	0.779	0.516
	Seuil	0.019	0.012	<0.010	0.016	<0.012	0.017
	<i>Se_c</i>	0.455	0.707	0.368	0.464	0.571	0.537
	<i>Spe_c</i>	0.830	0.671	0.797	0.810	0.881	0.607
CV	<i>AUC</i>	0.650	0.564	0.534	0.672	0.649	0.528
	Seuil	0.164	0.158	0.181	0.138	0.136	<0.131
	<i>Se_c</i>	0.515	0.431	0.158	0.607	0.464	0.707
	<i>Spe_c</i>	0.768	0.786	0.983	0.679	0.826	0.439
V	<i>AUC</i>	0.510	0.501	0.604	0.614	0.608	0.683
	Seuil	<0.101	<0.051	0.075	0.089	0.1	<0.068
	<i>Se_c</i>	0.606	0.138	0.579	0.571	0.429	0.537
	<i>Spe_c</i>	0.5	0.957	0.615	0.701	0.881	0.766
E₀ (dB)	<i>AUC</i>	0.567	0.521	0.564	0.526	0.785	0.528
	Seuil	48.264	<61.586	<54.631	56.083	58.552	55.359
	<i>Se_c</i>	0.909	0.948	0.684	0.429	0.536	0.732
	<i>Spe_c</i>	0.304	0.257	0.584	0.712	0.927	0.439
F₀ (Hz)	<i>AUC</i>	0.688	0.564	0.550	0.535	0.921	0.581
	Seuil	<271.45	235.75	213.82	252.54	198.04	119.21
	<i>Se_c</i>	0.970	0.897	0.868	0.964	1	0.829
	<i>Spe_c</i>	0.545	0.257	0.426	0.272	0.817	0.383
F₁ (Hz)	<i>AUC</i>	0.758	0.604	0.501	0.560	0.852	0.484
	Seuil	1058	1055	867	<882	802	<595
	<i>Se_c</i>	0.788	0.362	0.658	0.464	0.821	0.366
	<i>Spe_c</i>	0.679	0.843	0.430	0.663	0.917	0.916
F₂ (Hz)	<i>AUC</i>	0.815	0.519	0.688	0.567	0.937	0.833
	Seuil	1715	<1406	<1477	<1513	1391	1206
	<i>Se_c</i>	0.818	0.293	0.868	0.786	0.786	0.976
	<i>Spe_c</i>	0.714	0.907	0.467	0.391	0.963	0.664
F₃ (Hz)	<i>AUC</i>	0.678	0.564	0.884	0.505	0.840	0.582
	Seuil	2323	<3023	<2891	2944	2891	3049
	<i>Se_c</i>	1	0.448	0.974	0.714	0.821	0.244
	<i>Spe_c</i>	0.384	0.807	0.687	0.429	0.826	0.991

Le signe (<) indique que les valeurs inférieures à cette valeur seuil sont pathologiques.

Tableau 3-12 : Valeurs seuils AUC des différents paramètres étudiés selon l'âge et le sexe - Cas des VL

		1F	2F	3F	1M	2M	3M
VOT (ms)	<i>AUC</i>	0.549	0.719	0.510	0.671	0.567	0.629
	Seuil	0.025	0.012	0.016	0.012	<0.012	<0.012
	<i>Se_c</i>	0.293	0.667	0.407	0.750	0.370	0.371
	<i>Spe_c</i>	0.875	0.689	0.746	0.593	0.813	0.888
CV	<i>AUC</i>	0.818	0.529	0.655	0.647	0.764	0.526
	Seuil	0.335	0.283	<0.155	0.214	0.197	<0.170
	<i>Se_c</i>	0.610	0.222	0.519	0.607	0.778	0.486
	<i>Spe_c</i>	0.886	0.878	0.792	0.660	0.732	0.704
V	<i>AUC</i>	0.820		0.614	0.639	0.599	0.592
	Seuil	0.170		<0.113	0.125	0.201	<0.100
	<i>Se_c</i>	0.756		0.593	0.786	0.407	0.457
	<i>Spe_c</i>	0.750		0.706	0.519	0.938	0.806
E₀ (dB)	<i>AUC</i>	0.683	0.568	0.504	0.504	0.729	0.567
	Seuil	58.013	<62.928	50.644	51.764	59.409	<62.765
	<i>Se_c</i>	0.683	0.952	1	0.929	0.593	1
	<i>Spe_c</i>	0.693	0.284	0.172	0.216	0.821	0.255
F₀ (Hz)	<i>AUC</i>	0.535	0.580	0.576	0.505	0.885	0.503
	Seuil	<273.82	262.77	236.57	<316.66	194.81	124.23
	<i>Se_c</i>	0.585	0.730	0.556	0.893	1	0.971
	<i>Spe_c</i>	0.636	0.547	0.614	0.278	0.732	0.255
F₁ (Hz)	<i>AUC</i>	0.798	0.556	0.604	0.573	0.825	0.583
	Seuil	1013	<915	<859	1006	819	<679
	<i>Se_c</i>	0.976	0.460	0.667	0.5	0.667	0.714
	<i>Spe_c</i>	0.511	0.736	0.617	0.704	0.938	0.469
F₂ (Hz)	<i>AUC</i>	0.562	0.541	0.534	0.638	0.736	0.832
	Seuil	1434	<1425	<1432	<1449	1297	1194
	<i>Se_c</i>	1	0.365	0.778	0.679	0.556	0.743
	<i>Spe_c</i>	0.250	0.770	0.366	0.593	0.875	0.755
F₃ (Hz)	<i>AUC</i>	0.627	0.527	0.910	0.687	0.859	0.641
	Seuil	3799	3735	<2700	<3378	2862	<2452
	<i>Se_c</i>	0.610	0.302	0.889	0.750	0.926	0.457
	<i>Spe_c</i>	0.750	0.838	0.921	0.605	0.804	0.878
Jitter	<i>AUC</i>	0.529	0.509	0.664	0.6	0.792	0.547
	Seuil	<0.654	0.269	0.606	0.568	<0.536	<0.599
	<i>Se_c</i>	1	0.810	0.481	0.571	0.852	0.543
	<i>Spe_c</i>	0.159	0.257	0.785	0.722	0.652	0.633
Shimmer	<i>AUC</i>	0.552	0.544	0.553	0.711	0.582	0.547
	Seuil	<5.018	4.865	2.453	4.292	<4.048	5.211
	<i>Se_c</i>	0.927	0.317	0.889	0.750	0.815	0.343
	<i>Spe_c</i>	0.239	0.784	0.317	0.654	0.438	0.786
HNR	<i>AUC</i>	0.509	0.563	0.528	0.583	0.804	0.529
	Seuil	21.706	<19.260	<21.449	<21.431	17.994	<19.121
	<i>Se_c</i>	0.293	0.587	0.852	0.893	0.926	0.857
	<i>Spe_c</i>	0.818	0.608	0.281	0.278	0.616	0.276

Le signe (<) indique que les valeurs inférieures à cette valeur seuil sont pathologiques.

Quand un test conduit à des résultats quantitatifs continus, il est nécessaire de définir un seuil (ou valeur seuil) permettant de classer le résultat en normal (PN dans notre cas) ou anormal (PP dans notre cas). Le choix de cette valeur seuil influencera la sensibilité et la spécificité du test et par conséquent ses valeurs prédictives [84].

Tableau 3-13 : Valeurs seuils des différents paramètres acoustiques étudiés selon l'âge et le sexe (VL et VC)

Voyelle Etudiée	Classe	VOT	CV	V	E ₀	F ₀	F ₁	F ₂	F ₃	Jitter	Shimmer
VL	1F	0,025	0,335	0,17	56,442	<245,05	1013	1434	3948	<0,241	<1,63
	2F	0,012	0,26	<0,181	<51,904	262,77	1017	<1836	3735	0,269	3,555
	3F	0,016	<0,106	<0,274	50,644	236,57	<741	<1292	<2089	0,487	2,453
	1M	0,013	0,214	0,136	53,179	<249,27	958	<1117	<3661	0,568	4,738
	2M	<0,026	0,197	0,201	59,409	227,14	821	1297	2862	<3,76	<1,828
	3M	0,038	<0,121	<0,199	<54,994	124,22	<651	1192	<3106	<0,389	4,995
VC	1M	0,008	0,117	0,089	59,501	<252,54	<942	<1069	<3133		
	2M	<0,005	0,182	0,1	59,075	198,04	858	1426	2891		
	3M	<0,017	0,084	<0,162	<55,318	119,2	<800	1203	3049		
	1F	0,02	0,164	0,069	<45,795	<237,08	1058	1548	2323		
	2F	0,013	0,158	0,106	56,044	235,75	1055	1584	<3703		
	3F	<0,018	0,181	0,075	<48,014	213,82	<978	<1819	<2184		

Le signe (<) indique que les valeurs inférieures à cette valeur seuil sont pathologiques.

3.5 Discussions générales

Les locuteurs adoptent donc diverses stratégies compensatoires, mais la capacité d'adaptation est propre à chacun, chaque sujet réagira ainsi de façon différente.

La déperdition nasale persiste et est importante, un souffle nasal devient audible et masque plus ou moins la parole.

Conformément à la littérature, concernant l'évolution du conduit vocal en général, les valeurs de la fréquence fondamentale F₀ et des trois formants F₁, F₂ et F₃ tendent à diminuer (Tableau 3-14).

 Tableau 3-14 : Evolution de la fréquence fondamentale F₀ et des formants F₁, F₂ et F₃, pour les deux sexes selon les trois tranches d'âge.

Formants (Hz)	Masculin			Féminin		
	Enfants	Adolescents	Adultes	Enfants	Adolescentes	Adultes
F ₀	289,774	177,603	142,104	288,148	275,186	234,115
F ₁	957,371	712,823	667,938	1003,136	966,108	888,733
F ₂	1502,241	1202,833	1119,214	1584,807	1511,567	1388,564
F ₃	3378,111	2697,032	2679,562	3445,012	3379,845	3020,534

D'après le tableau 3-14 on peut noter que : à l'âge de 6 ans, la fréquence fondamentale est presque identique pour les deux sexes **289,774 Hz** pour les garçons et **288,148 Hz** pour les filles. A la puberté, nous constatons une baisse plus importante chez les garçons **177,603 Hz** que chez les filles **275,186 Hz** et elle continue à baisser pour les adultes, de façon plus importante pour les hommes **142,104 Hz** que pour les femmes **234,115 Hz** ;

- les durées [CV] et [V] sont plus élevées pour les locuteurs pathologiques de la 1^{ère} tranche d'âge ; Les segments prolongés peuvent être dûs à une stratégie active pour

augmenter la pression de l'air oral et / ou améliorer la précision perceptuelle des segments de parole [87];

- les valeurs du VOT sont plus élevées chez les locutrices pathologiques à l'âge d'adolescence ;
- les locuteurs masculins pathologiques de la 2^{ème} Classe ont des valeurs élevées de E₀, F₀, F₁, F₂ ainsi que F₃ ;
- les locutrices pathologiques à l'âge d'adolescence ont des valeurs inférieures de F₁ et F₂ ;
- les locutrices pathologiques adultes ont des valeurs presque égales à celles des locutrices saines de cette tranche d'âge pour les différents paramètres étudiés ;
- à l'âge adulte, les valeurs du Jitter et du Shimmer tendant vers des valeurs de seuil normal. Ceci est probablement dû au fait que le patient est en mesure de mieux contrôler ses efforts par une meilleure connaissance des divers organes de la phonation de son nouveau mécanisme de production de la parole [88];
- relativement aux difficultés du contrôle de fermeture des occlusives, et au relâchement subséquent, le bruit de l'explosion du relâchement de l'occlusive pourrait affecter la voyelle suivante, retardant l'apparition d'une structure formantique claire, et augmentant de fait la durée du VOT [87];
- l'explosion de l'occlusive n'est pas nette, et la constrictive correspondant à l'occlusive voulue au niveau articulaire est souvent réalisée à la suite de cette occlusive, provoquant une durée plus importante du VOT ;
- la présentation détaillée des résultats, indique que ces résultats au niveau temporel sont à nuancer dans certains cas, étant donné la difficulté à systématiquement lier le type de pathologie à un comportement défini, et la grande variabilité observée surtout chez les plus jeunes sujets pathologiques. Cette variabilité tend à s'estomper avec l'âge, c'est-à-dire chez les sujets pathologiques adultes.

Nous avons constaté pour le cas des adultes que la parole pathologique devient plus fluide et assez intelligible après un an de rééducation. Une amélioration significative a été identifiée dans le placement de la langue, lors de la production du phonème [t], le système moteur de la parole a pu s'adapter à un changement dans la structure périphérique (transformation de la zone opérée) pour maintenir une « justesse » dans la production de la parole, ce qui pourrait s'expliquer par la compensation par l'organisme du manque de pression intra-orale normale et d'habitude d'articulation.

3.6 Conclusion

Une analyse acoustique a été réalisée sur des locuteurs porteurs de fente et des locuteurs sains. Les locuteurs pathologiques ont été divisés en deux groupes, ceux qui ont bénéficié d'une rééducation orthophonique et ceux qui n'ont pas encore bénéficié.

Les locuteurs ayant une fente ont recours à certains ajustements compensatoires qui leur sont nécessaires en raison de leur capacité réduite à produire des mouvements

vélopharyngaux correctement, le meilleur moyen pour ces locuteurs est de réorganiser leur système moteur en l'adaptant à leur propre structure, totalement unique. Ainsi, chaque locuteur aurait des stratégies articulatoires adaptées à son anatomie.

L'analyse acoustique a été réalisée à partir des enregistrements de paroles pathologiques de locuteurs porteurs de fente palatine comparés à des enregistrements de paroles de locuteurs normaux.

Plusieurs indices acoustiques sont relevés pour caractériser la parole produite par les locuteurs porteurs de fente. En plus des paramètres acoustiques classiques tels que la fréquence fondamentale, le suivi des trois premiers formants, l'intensité, les durées du VOT, [CV] et [V], d'autres indices sont également pris en compte pour la caractérisation de la qualité de la parole pathologique, car ils permettent une meilleure appréciation du degré de perturbation de la voix pathologique. Parmi ces indices acoustiques, nous avons : Le Jitter (mesure du degré de perturbation de la fréquence fondamentale), le Shimmer (mesure du degré de perturbation de l'amplitude) et le HNR (mesure du rapport harmoniques/bruit pour estimer le degré d'altération de la qualité de la voix). L'analyse de ces paramètres nous donnent des indications quant aux stratégies de réajustement adoptées par les locuteurs porteurs de fente au niveau articulatoire. Il s'agit donc d'informations précieuses pour notre système de reconnaissance automatique de ce type de parole pathologique. Elle permet notamment de représenter fidèlement le signal de parole par un choix adéquat de la dimension et de la nature des vecteurs d'entrée de notre système de reconnaissance.

La mesure des paramètres acoustiques choisis permet de faire des comparaisons entre les sujets et de suivre l'évolution de la production de la parole chez un patient porteur de fente. Cette analyse acoustique pourra appuyer la rééducation de ces patients et évaluer, de façon objective, l'évolution de leurs rééducation au cours du temps, en offrant des données concrètes à l'orthophoniste pour une meilleure prise en charge.

**CHAPITRE 4 : APPLICATION DES ARBRES DE DECISION
A LA CLASSIFICATION DES PAROLES PATHOLOGIQUES**

4.1 Introduction

La recherche sur les arbres de classification et de régression a connu une croissance rapide et les applications augmentent à un rythme encore plus rapide. L'interprétabilité des structures arborescentes est une raison importante de leur popularité parmi les praticiens, mais il en va de même pour une précision de prédiction raisonnablement bonne, une vitesse de calcul rapide et une large disponibilité des logiciels.

Ce chapitre s'articule essentiellement autour des principes de base des Arbres de Décision, avec une description détaillée de celui exploité dans le cadre de notre travail. Nous présentons les différentes techniques utilisées dans un système de Reconnaissance Automatique de la Parole (RAP). Nous exposons la conception et l'architecture de notre Système de Classification par la méthode de l'Arbre de Décision (SCAD) basé sur l'algorithme C4.5, pour reconnaître et classifier automatiquement la parole pathologique étudiée par rapport à la Parole Normale. Les résultats et leurs interprétations sont discutés à la fin du chapitre.

4.2 Reconnaissance Automatique de la parole (RAP)

L'utilisation de la parole comme mode de communication entre l'Homme et la Machine a été largement étudiée au cours des dernières décennies. Le Traitement Automatique de la Parole (TAP), c'est-à-dire à l'ensemble des techniques permettant de communiquer oralement avec une machine, présente un intérêt pratique indéniable. C'est un vaste domaine de recherche très complexe qui nécessite la collaboration des spécialistes dans de nombreux domaines : traitement du signal, théorie de l'information, linguistique, phonétique, acoustique, biologique, etc [1].

La Reconnaissance Automatique de la Parole (RAP), un des domaines du TAP, a considérablement progressé ces dernières années. Les études scientifiques sur la Reconnaissance Automatique de la Parole ont commencé au début des années 1950, mais ont donné leurs premiers résultats tangibles dans les années 70 avec le projet ARPA (Advanced Research Projects Agency) [89]. Leur principal but a été, et restera toujours, de concevoir un système robuste et efficace de reconnaissance de la parole, pour obtenir un outil s'intégrant naturellement dans la communication entre l'Homme et la Machine et facilitant son interaction.

La Reconnaissance Automatique de la Parole est l'un des axes les plus épineux dans le Traitement Automatique de la Parole (TAP). Elle consiste à faire interpréter à une machine un message prononcé par un locuteur humain en convertissant le signal acoustique de parole en un ensemble de mots dans un autre support, principalement graphique.

La RAP se scinde en deux domaines de recherche ou deux approches : la reconnaissance globale basée sur les méthodes utilisées en reconnaissance des formes et qui concerne surtout les mots isolés ou enchaînés, la reconnaissance analytique qui est basée sur la segmentation du signal en unités fondamentales dont les informations serviraient à déchiffrer n'importe quel segment de parole. Cette approche est surtout utilisée dans la parole continue.

Pour reconnaître et classifier automatiquement une parole, plusieurs méthodes sont utilisées.

4.2.1 K plus proches voisins

Plus connus en Anglais sous le nom K-Nearest Neighbor ou K-NN. Cette méthode diffère des méthodes traditionnelles d'apprentissage car aucun modèle n'est induit à partir des exemples. Les données restent telles quelles : elles sont simplement stockées en mémoire [90].

Pour prédire la classe d'un nouveau cas, l'algorithme cherche les K plus proches voisins de ce nouveau cas et prédit (s'il faut choisir) la réponse la plus fréquente de ces K plus proches voisins [91].

4.2.2 Technique des arbres de décision

Un arbre de décision est, comme son nom le suggère, un outil d'aide à la décision qui permet de répartir une population d'individus en groupes homogènes selon des attributs discriminants en fonction d'un objectif fixé et connu. Il permet d'émettre des prédictions à partir des données connues sur le problème par réduction, niveau par niveau, du domaine des solutions.

Chaque nœud interne d'un arbre de décision porte sur un attribut discriminant des éléments à classifier, ce qui permet de répartir ces éléments de façon homogène entre les différents fils de ce nœud. Les branches liant un nœud à ses fils représentent les valeurs discriminantes de l'attribut du nœud. En fin, les feuilles d'un arbre de décision sont les prédictions concernant les données à classifier.

C'est une méthode qui a l'avantage d'être lisible pour les analystes et qui permet de déterminer les couples <attribut, valeur> discriminants à partir d'un très grand nombre d'attributs et de valeurs.

Les arbres de décision ont été utilisés dans différentes tâches comme la classification d'images [92] et la détection humaine en temps réel [93].

4.2.3 Naïve ou décisions de Bayes

D'après le théorème de Bayes, ces méthodes sont qualifiées de "Naïve" ou "Simple" car elles supposent l'indépendance des variables. L'idée est d'utiliser des conditions de probabilité observées dans les données. On calcule la probabilité de chaque classe parmi les exemples.

Une variante des Naïve Bayes sont les réseaux Bayésiens : dans ce modèle, on ne suppose plus que les variables sont toutes indépendantes, et on autorise certaines à être liées. Cela alourdit considérablement les calculs et les résultats n'augmentent pas de façon significative.

4.2.4 Hidden Markov Models (HMM)

La plupart des applications en RAP utilisent la technologie basée sur les modèles de Markov cachés (HMM, pour Hidden Markov Model), qui sont capables de modéliser simultanément les caractéristiques temporelles et fréquentielles du signal vocal [94].

La théorie des HMM décrit comment passer d'état en état à l'aide de probabilités de transitions et comment chaque élément de la séquence peut être émis par un état du HMM à l'aide de probabilités d'observations par état. Tous les états ont des transitions unidirectionnelles vers tous les autres états, y compris vers eux-mêmes. Ils sont cachés et chacun émet des "observations" qui sont observables. On ne travaille donc pas sur la séquence d'états, mais sur la séquence d'observations générées par les états. Une description détaillée des principes de base et des différentes étapes d'application en RAP est assez disponible et peut être consultée sur internet et dans beaucoup d'ouvrages [95].

4.2.5 Réseaux de Neurones Artificiels (RNA)

Un Réseau de Neurones est un modèle de calcul dont le fonctionnement schématique est inspiré du fonctionnement des neurones biologiques. Chaque neurone fait une somme pondérée de ses entrées (ou synapses) et retourne une valeur en fonction de sa fonction d'activation. Cette valeur peut être utilisée soit comme une des entrées d'une nouvelle couche de neurones, soit comme un résultat qu'il appartient à l'utilisateur d'interpréter (classe, résultat d'un calcul, etc.) [90; 96].

La phase d'apprentissage d'un Réseau de Neurones permet de régler le poids associé à chaque synapse d'entrée (on parle également de coefficient synaptique). C'est un processus long qui doit être réitéré à chaque modification structurelle de la Base de Données traitée.

Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires, les neurones formels. Les principaux réseaux se distinguent par l'organisation du graphe (en couches, complets.. .), c'est-à-dire leur architecture, son niveau de complexité (le nombre de neurones) et par le type des neurones (leurs fonctions de transition) [97].

4.2.6 Programmation génétique

C'est une méthode générale qui peut être utilisée après n'importe quelle méthode précédente, par exemple avec les arbres de décisions. En entrée, un algorithme génétique reçoit une population de classificateurs non optimaux. Le but du programme génétique est de produire un classificateur plus optimal que chacun de ceux de la population d'origine.

D'une façon simple, cela consiste à extraire les meilleures parties de chaque classificateur d'origine et de les mettre ensemble pour produire un nouveau classificateur. Cela suppose de pouvoir comparer l'efficacité d'un classificateur.

Un résultat important de la méthode est qu'après chaque itération on obtient un meilleur classificateur que celui d'avant. On peut donc arrêter les itérations à tout moment, même si le résultat n'est pas l'optimum.

4.2.7 Machines à Vecteurs Support (ou SVM)

Cette technique a été initiée par V. Vapnik en 1995. Elle tente de séparer linéairement les exemples positifs des négatifs dans l'ensemble des exemples. Chaque exemple doit être représenté par un vecteur de dimension n .

Pour les SVM, leur principal objectif est de déterminer si un élément appartient à une classe ou pas, d'où leur principale exploitation dans le domaine de la classification. Nous disposons d'un ensemble de données et nous cherchons à séparer ces données en deux groupes. Le premier est l'ensemble de données appartenant à une classe, ces données sont étiquetées généralement par (+) et un autre ensemble qui contient les éléments qui n'appartiennent pas à la classe donc étiquetées (-). L'algorithme SVM permet de trouver un hyperplan séparateur entre ces deux groupes. Pour optimiser la séparation, cet algorithme cherche l'hyperplan pour lequel la distance entre la frontière des deux groupes et les points les plus proches est maximale. En d'autres termes, le principe des SVM est de séparer les exemples de deux classes avec cet hyperplan tout en gardant le maximum de marge entre les exemples et ce même hyperplan. C'est le principe de maximisation de la marge ou distance du point le plus proche à l'hyperplan. Pour cela, les SVM sont également appelés "Maximum Margin Classifier" [98].

Après avoir vu certains des modèles les plus pertinents de l'état de l'art pour la classification, il faut préciser qu'il n'existe pas une règle générale qui permet de savoir quel modèle utiliser. Ceci dépend du type de données disponibles ainsi que du domaine et des objectifs particuliers de l'application.

4.3 Les arbres de décision (AD)

Un Arbre de Décision (AD) est un schéma représentant les résultats possibles d'une série de choix interconnectés. Il commence généralement par un nœud d'où découlent plusieurs résultats possibles. Chacun de ces résultats mène à d'autres nœuds, d'où émanent d'autres possibilités. Le schéma ainsi obtenu rappelle la forme d'un arbre (Figure 4-1).

La popularité de la méthode des Arbres de Décision repose en grande partie sur sa simplicité, c'est une classe d'algorithmes qui fait partie de la famille des outils à valider par la

méthode d'apprentissage. Elle se base sur la représentation des choix sous la forme graphique d'un arbre avec les différentes décisions de classification placées logiquement dans les feuilles.

4.3.1 Etat de l'art sur les AD

La construction des arbres de décision à partir de données est une discipline déjà ancienne. Les statisticiens en attribuent la paternité à X. Morgan et Y. Sonquist (1963) qui, les premiers, ont utilisé les arbres de régression dans un processus de prédiction et d'explication (AID – Automatic Interaction Detection) [99].

A la fin des années 70 et début des années 80 deux groupes de chercheurs ont développé, de manière indépendante l'un de l'autre, la technique de classification d'arbres de décision. J.Ross Quinlan a développé un algorithme d'arbres de décision appelé ID3 (Iterative Dichotomiser -1979). Quinlan a été un acteur très actif dans la deuxième moitié des années 80 avec un grand nombre de publications où il propose un ensemble d'heuristiques pour améliorer le comportement de son système. Son approche a pris un tournant important dans les années 90 lorsqu'il présenta la méthode C4.5 qui est l'autre référence incontournable dès que l'on veut citer les arbres de décision (1993). Il existe bien une autre évolution de cet algorithme, C5.0, mais étant implémentée dans un logiciel commercial. De leur côté, en 1984, L.Breiman, J.Friedman, R.Olshen, et C.Stone, un groupe de statisticiens, ont développé un algorithme pour produire des arbres de décision binaires appelé CART (Classification and Regression Trees) [100]. Ces algorithmes ont été le réel début de la recherche et ils ont produit le développement de nouveaux et très intéressants algorithmes de classification basés sur des arbres de décision.

Dans cette partie, nous présentons les principes de construction des arbres de décision dans les problèmes de discrimination et classement. Nous commençons par un nombre de définitions pour fixer la terminologie. Ensuite nous présentons les principes de base de l'apprentissage statistique avec des arbres de décision (en classification et en régression).

4.3.2 Définition d'un AD

Les Arbres de Décision (AD) sont une catégorie d'arbres utilisée dans l'exploration de données et en informatique décisionnelle. Son fonctionnement repose sur des heuristiques construites selon des techniques d'apprentissage supervisées.

Les AD emploient une représentation hiérarchique de la structure des données et sont composés de nœuds et de feuilles reliés par des branches qui représentent nos principaux choix. Dans leur représentation graphique, la racine est placée en haut et les feuilles en bas.

L'accès à l'arbre se fait par la racine qui représente la question principale à laquelle on cherche à répondre. C'est de là que part toute notre réflexion.

Les principaux composants d'un modèle d'arbre de décision sont les nœuds et les branches et les étapes les plus importantes de la création d'un modèle sont le fractionnement, l'arrêt et l'élagage [101].

Un nœud dans l'arbre consiste à tester une condition particulière qui, en général, compare la valeur d'un attribut avec une constante, ou compare un ensemble de deux attributs, ou encore utilise des fonctions mathématiques d'un ou plusieurs attributs.

Il existe trois types de nœuds différents :

- Un nœud racine, également appelé nœud de décision, représente un choix qui entraînera la subdivision de tous les enregistrements en deux ou plusieurs sous-ensembles mutuellement exclusifs.
- Les nœuds internes, également appelés nœuds aléatoires, représentent l'un des choix possibles disponibles à ce point de l'arborescence; Chaque nœud interne porte sur un attribut discriminant des éléments à classifier qui permet de répartir ces éléments de façon homogène entre les différents fils de ce nœud ; le bord supérieur du nœud est connecté à son nœud parent et le bord inférieur est connecté à ses nœuds enfants ou nœuds feuilles.
- Les nœuds feuilles, également appelés nœuds d'extrémité, représentent le résultat final d'une combinaison de décisions ou d'événements, se sont les prédictions de l'AD concernant les données à classifier.

Un modèle d'arbre de décision est formé en utilisant une hiérarchie de branches. Les branches représentent des résultats aléatoires ou des occurrences qui émanent des nœuds racine et des nœuds internes. Chaque chemin depuis le nœud racine via les nœuds internes vers un nœud feuille représente une règle de décision de classification.

- une branche est orientée du "nœud parent" vers le " nœud enfant ". Elle correspond à une valeur pour la variable prédictive du nœud parent (ou un ensemble de valeurs) ;
- chaque nœud est connecté à un et un seul nœud parent, sauf le nœud racine qui n'a pas de parent ;
- chaque nœud peut être connecté à 0 ou n nœuds enfants ;
- un nœud qui n'a pas de nœuds enfants est appelé " nœud feuille" ou " feuille ", il correspond à la variable cible ;
- chaque nœud non-feuille correspond à une variable prédictive ;
- un nœud feuille est pur si les valeurs de la variable cible sont les mêmes pour tous les enregistrements de ce nœud.

Comme toutes les méthodes supervisées, un arbre de décision travaille sur une variable cible avec plusieurs variables prédictives. Le processus de décision est équivalent à une " descente " dans l'arbre, de la racine vers une des feuilles en vue de la prédiction d'un

résultat ou d'une classe. Les valeurs que peut prendre une variable dans un arbre de décision sont appelées instances ou attributs. A chaque étape un attribut est testé et un sous-arbre est choisi, le parcours s'arrête dans une feuille, une décision est prise (Figure 4-1). Les nœuds feuilles contiennent la classe appelée classe à prédire. L'arbre de décision construit des frontières continues par morceaux et qui se superposent les unes aux autres. Cette propriété fait de l'arbre de décision un outil qui s'adapte à toute forme de distribution de la variable réponse Y au détriment d'un manque de robustesse.

Après sa construction, un arbre de décision peut être traduit sous la forme d'un ensemble de règles de décision.

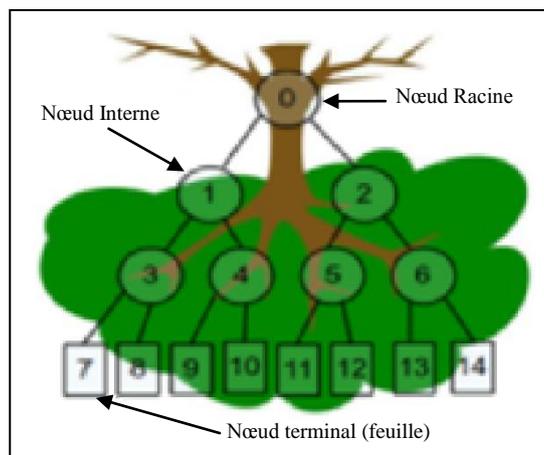


Figure 4-1: Structure générale d'un Arbre de Décision [102]

4.3.2.1 Types d'arbres

Il existe deux types d'arbre de décision, arbre de classification et arbre de régression. Lorsque la variable cible est une variable qualitative (peut être catégorielle), nous avons un arbre de classification, lorsqu'elle est quantitative (ou valeur réelle) nous avons un arbre de régression.

Les arbres de classification permettent de prédire à quelle classe la variable cible appartient, dans ce cas la prédiction est une étiquette de classe (Figure 4-2).

Dans les arbres de classification les groupes définis a priori par la variable catégorielle à expliquer, C'est-à-dire quelles variables classent le mieux les observations dans les groupes.

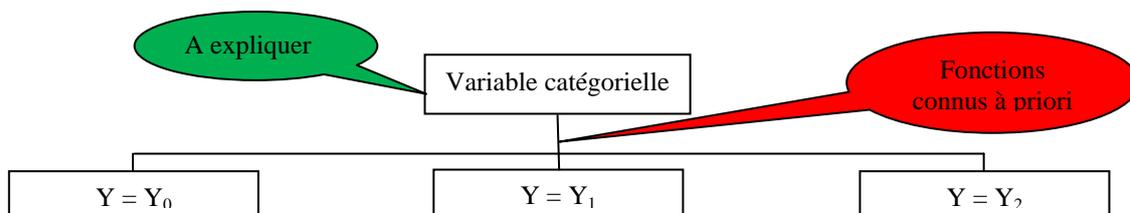


Figure 4-2 : Arbre de classification

Les arbres de régression permettent de prédire une quantité réelle (par exemple, la durée de séjour d'un patient dans un hôpital), dans ce cas la prédiction est une valeur numérique [103].

Dans les arbres de régression les groupes définis à posteriori pour la variable continue à expliquer c'est-à-dire quelles variables prédisent le mieux la variable à expliquer (Figure 4-3).

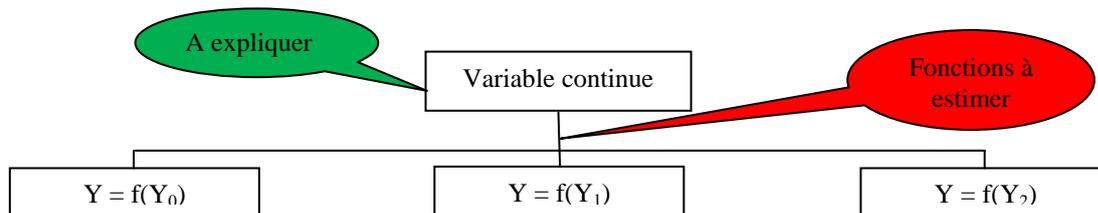


Figure 4-3 : Arbre de régression

4.3.2.2 Les arbres de décision dans l'apprentissage automatique et l'exploration des données

Un arbre de décision peut également servir à bâtir des modèles prédictifs automatisés, dont les applications peuvent concerner l'apprentissage automatique, l'exploration de données ou les statistiques. Cette méthode, appelée « apprentissage par arbre de décision », s'appuie sur les observations relatives à un élément pour prédire la valeur de cet élément.

Dans ces arbres de décision, les nœuds représentent les données plutôt que les décisions. Ce type d'arbre est aussi appelé arbre de classification. Chaque branche contient un ensemble d'attributs (règles de classification), associés à une étiquette de classe spécifique que l'on retrouve à l'extrémité de la branche.

Ces voies d'arbre de décision peuvent également être représentées sous forme de règles «si-alors» [104], où chaque décision ou valeur de donnée forme une clause, par exemple : « si les conditions 1, 2 et 3 sont remplies alors l'issue x sera le résultat avec une certitude de y.»

Chaque donnée supplémentaire aide le modèle à prédire avec davantage de précision à quel ensemble limité de valeurs le sujet en question appartient. Cette information peut alors être utilisée comme entrée dans un modèle de prise de décision plus vaste.

Parfois, la variable prédite est un chiffre réel, par exemple un prix. Les arbres de décision avec une infinité de résultats possibles sont appelés des arbres de régression.

4.3.3 Principe de la construction d'un arbre de décision

L'algorithme général de création d'un AD :

- Déterminer la meilleure caractéristique dans l'ensemble de données d'entraînement ;
- Diviser les données d'entraînement en sous-ensembles contenant les valeurs possibles de la meilleure caractéristique ;
- Générez de manière récursive de nouveaux arbres de décision en utilisant les sous-ensembles de données créés ;
- Lorsqu'on ne peut plus classifier les données, on s'arrête.

Les principaux buts de la construction des arbres de décision sont [102] :

- une meilleure généralisation des exemples de la base d'apprentissage ;
- une meilleure classification de nouveaux exemples ;
- une structure aussi simple que possible.

La première étape de l'arbre de décision est de définir notre problématique. Pour rappel c'est la racine de notre décision, c'est de là que tout part.

Au départ, les points de la base d'apprentissage sont tous placés dans le nœud racine. L'idée centrale est de diviser récursivement et le plus efficacement possible les exemples de l'ensemble d'apprentissage par des tests définis à l'aide des attributs jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant tous à une même classe.

Seules les variables d'entrée liées à la variable cible sont utilisées pour diviser les nœuds parents en nœuds enfants plus purs de la variable cible. Les variables d'entrée discrètes et les variables d'entrée continues (qui sont regroupées en deux ou plusieurs catégories) peuvent être utilisées.

L'arbre est construit par partition récursive de chaque nœud en fonction de la valeur de l'attribut testé à chaque itération (top-down induction). Le critère optimisé est l'homogénéité des descendants par rapport à la variable cible. La variable qui est testée dans un nœud sera celle qui maximise cette homogénéité.

Dans toutes les méthodes, on trouve les trois opérateurs suivants :

- Décider si un nœud est terminal, c'est-à-dire décider si un nœud doit être étiqueté comme une feuille ;
- Sélectionner un test à associer à un nœud ;
- Affecter une classe à une feuille.

Les méthodes vont différer par les choix effectués pour ces différents opérateurs, c'est-à-dire sur le choix d'un test (par exemple, utilisation du gain et de la fonction entropie) et le critère d'arrêt (quand arrêter la croissance de l'arbre, soit quand décider si un nœud est terminal) [105].

L'arbre le plus simple est celui qui permet de minimiser l'espérance du nombre de questions nécessaires à la classification d'un exemple d'apprentissage.

Le processus s'arrête quand les éléments d'un nœud ont la même valeur pour la variable cible (homogénéité).

L'algorithme générale d'un AD est donné comme suit :

```

Entrées : langage de description ; échantillon S
Début
    Initialiser à l'arbre vide ; // la racine est le nœud courant
    Répéter
    Décider si le nœud courant est terminal ;
        Si le nœud est terminal
            Alors affecter une classe ;
            Sinon sélectionner un test et créer le sous arbre ;
    Passer au nœud suivant non exploré s'il existe ;
Jusqu'à obtenir un arbre de décision ;
Fin

```

L'arbre de décision est une méthode classique d'apprentissage s'utilisant de manière similaire : le processus entier se compose de deux phases séparées, apprentissage et test. L'apprentissage consiste à construire un modèle représentant un ensemble des éléments, alors que le test utilise ce modèle pour évaluer un nouvel élément inconnu [102].

4.3.3.1 Optimisation des nœuds

Un arbre de décision parfait est un arbre de décision tel que tous les exemples de l'ensemble d'apprentissage soient correctement classifiés. L'objectif est de construire un arbre d'erreur de classification la plus petite possible. Mais malheureusement l'erreur apparente est une vision très optimiste de l'erreur réelle, c'est-à-dire qu'un arbre soit bien adapté à l'échantillon mais ait un pouvoir de prédiction faible. L'idéal serait de trouver un critère qui permette d'arrêter la croissance de l'arbre au bon moment.

Lors de la construction du modèle, il faut d'abord identifier les variables d'entrée les plus importantes, puis diviser les enregistrements au nœud racine et aux nœuds internes suivants en deux ou plusieurs catégories en fonction de l'état de ces variables. Les caractéristiques liées au degré de «pureté» des nœuds enfants résultants (c'est-à-dire la proportion avec la condition cible) sont utilisées pour choisir entre différentes variables d'entrée potentielles; ces caractéristiques comprennent l'entropie, l'indice de Gini, l'erreur de classification, le gain d'information, le rapport de gain et les critères de deux [102]. Ces mesures sont aussi appelées des mesures de pureté, parce qu'une petite valeur indique qu'une région, ou son nœud associé, contient principalement des observations d'une même classe.

Cette procédure de fractionnement se poursuit jusqu'à ce que l'homogénéité prédéterminée ou les critères d'arrêt soient satisfaits. Dans la plupart des cas, toutes les variables d'entrée potentielles ne seront pas utilisées pour construire le modèle d'arbre de décision et dans certains cas, une variable d'entrée spécifique peut être utilisée plusieurs fois à différents niveaux de l'arbre de décision.

4.3.3.1.1 Entropie

Soit une distribution de probabilité $P = (p_1, p_2, \dots, p_n)$. L'entropie de la distribution P est la quantité d'information qu'elle peut apporter. Elle est donnée par l'équation suivante :

$$E(P) = -\sum_{i=1}^n p_i \log p_i \quad (4.1)$$

log désigne la fonction logarithme en base 2.

Soit un ensemble de données T caractérisé par n classes (C_1, C_2, \dots, C_n) selon la variable cible. La quantité d'informations nécessaire pour identifier la classe d'un individu de l'ensemble T correspond à l'entropie $E(P)$, où P est la distribution de probabilité de la partition (C_1, C_2, \dots, C_n) :

$$P = \left(\frac{|C_1|}{|T|}\right), \left(\frac{|C_2|}{|T|}\right), \dots, \left(\frac{|C_n|}{|T|}\right) \quad (4.2)$$

$|C_i|$ désigne le cardinal de la classe i , c'est-à-dire le nombre d'éléments de la classe i .

L'entropie de T est déduite à partir de l'équation 4.1, qui donne :

$$E(T) = -\sum_{i=1}^n \frac{|C_i|}{|T|} \log \frac{|C_i|}{|T|} \quad (4.3)$$

Supposons que l'ensemble des données T est, en plus, partitionné ainsi (T_1, T_2, \dots, T_m) (m correspond au nombre de valeurs que peut prendre l'attribut X considéré). Dans ce cas, l'information nécessaire pour identifier la classe d'un individu de T_i devient :

$$E(X, T) = -\sum_{j=1}^m \frac{|T_j|}{|T|} E(T_j) \quad (4.4)$$

4.3.3.1.2 Gain d'informations

Soit un ensemble de données T , le gain d'informations de T par rapport à une partition T_j donnée est la variation d'entropie causée par la partition de T selon T_j .

$$\text{Gain}(X, T) = E(T) - E(X, T)$$

$$\text{Gain}(X, T) = E(T) - \sum_{j=1}^m \frac{|T_j|}{|T|} E(T_j) \quad (4.5)$$

4.3.3.1.3 Rapport de gain

Le rapport de gain est défini par :

$$\text{Rapport de Gain}(X, T) = \frac{\text{Gain}(X, T)}{\text{SplitInfo}(X, T)} \quad (4.6)$$

Avec : $\text{SplitInfo}(X, T) = E\left(\frac{|C_1|}{|T|}\right), \left(\frac{|C_2|}{|T|}\right), \dots, \left(\frac{|C_n|}{|T|}\right)$

4.3.3.1.4 Indice de Gini

L'indice de Gini a été introduit par X. Breiman en 2001 [106]. Cet indice mesure l'impureté, qui est un concept très utile dans la construction des arbres de décision : la qualité d'un nœud et son pouvoir discriminant peuvent être évalués par son impureté.

L'indice Gini est donné par la relation suivante :

$$Gini(T) = 1 - \sum_{j=1}^m \left(\frac{|T_j|}{|T|} \right)^2 \quad (4.7)$$

Une fois que l'ensemble des données T est caractérisé par une variable X partitionnée ainsi : (T_1, T_2, \dots, T_m)

$$Gain\ de\ Gini(X, T) = \sum_{j=1}^m \left(\frac{|T_j|}{|T|} \right) Gini(T_j) \quad (4.8)$$

Si l'ensemble de données est partitionné en deux partitions (T_1, T_2) , alors :

$$Gain\ de\ Gini(X, T) = \frac{|T_1|}{|T|} Gini(T_1) + \frac{|T_2|}{|T|} Gini(T_2) \quad (4.9)$$

4.3.3.2 Arrêt de l'arborescence

La complexité et la robustesse sont des caractéristiques concurrentes des modèles qui doivent être prises en compte simultanément lors de la construction d'un modèle statistique.

Plus un modèle est complexe, moins il sera fiable lorsqu'il sera utilisé pour prédire les enregistrements futurs. Une situation extrême consiste à construire un modèle d'arbre de décision très complexe qui s'étale suffisamment pour rendre les enregistrements de chaque nœud feuille 100% purs (c'est-à-dire que tous les enregistrements ont le résultat cible). Un tel arbre de décision serait trop ajusté aux observations existantes et aurait peu d'enregistrements dans chaque feuille, de sorte qu'il ne pourrait pas prédire de manière fiable les cas futurs et, par conséquent, aurait une faible généralisabilité (c.-à-d. Manque de robustesse). Pour éviter que cela ne se produise, des règles d'arrêt doivent être appliquées lors de la création d'un arbre de décision pour éviter que le modèle ne devienne trop complexe. Les paramètres courants utilisés dans les règles d'arrêt comprennent :

- le nombre minimum d'enregistrements dans une feuille ;
- le nombre minimum d'enregistrements dans un nœud avant le fractionnement ;
- la profondeur (c'est-à-dire le nombre d'étapes) de toute feuille à partir du nœud racine.

Les paramètres d'arrêt doivent être sélectionnés en fonction de l'objectif de l'analyse et des caractéristiques de l'ensemble de données utilisé. En règle générale, il est recommandé d'éviter le sur-ajustement et le sous-ajustement en définissant la proportion cible d'enregistrements dans un nœud feuille entre 0,25 et 1,00% de l'ensemble de données d'apprentissage complet [102].

4.3.3.3 Elagage

Dans certaines situations, les règles d'arrêt ne fonctionnent pas bien. Une autre façon de construire un modèle d'arbre de décision est de développer d'abord un grand arbre, puis de l'élaguer à la taille optimale en supprimant les nœuds qui fournissent moins d'informations supplémentaires [102].

Une méthode courante de sélection du meilleur sous-arbre possible parmi plusieurs candidats consiste à considérer la proportion d'enregistrements avec prédiction d'erreur (c'est-à-dire la proportion dans laquelle l'occurrence prévue de la cible est incorrecte). D'autres méthodes de sélection de la meilleure alternative consistent à utiliser un ensemble de données de validation (c.-à-d., Diviser l'échantillon en deux et tester le modèle développé sur l'ensemble de données d'apprentissage sur l'ensemble de données de validation), ou, pour les petits échantillons, la validation croisée (c.-à-d. Diviser l'échantillon en 10 groupes ou «plis», et test du modèle développé à partir de 9 plis sur le 10^{ème} pli, répété pour les dix combinaisons et calcul de la moyenne des taux ou des prédictions erronées).

Il existe deux types d'élagage, pré-élagage et post-élagage. Le pré-élagage utilise des méthodes d'ajustement à comparaisons multiples pour éviter la génération de branches non significatives. La post-élagage est utilisée après la génération d'un arbre de décision complet pour supprimer les branches d'une manière qui améliore la précision de la classification globale lorsqu'elle est appliquée à l'ensemble de données de validation.

4.3.4 Quelques algorithmes pour la construction d'un arbre de décision

Il existe plusieurs algorithmes pour construire un arbre de décision, dont nous pouvons citer les plus populaires.

4.3.4.1 Algorithme ID3

L'algorithme récursif ID3 considère en entrée un ensemble d'attributs " A ", un ensemble d'attributs cibles c (aussi appelés classes) et un ensemble d'échantillons " E ". Cet algorithme fonctionne exclusivement avec des attributs qualitatifs. A chaque étape de la récursion , il détermine l'attribut qui maximise le gain d'informations. Cet attribut est alors retenu pour la branche en cours puisqu'il permet de classer plus facilement l'ensemble des données à ce niveau de l'arbre [107].

ID3 a été conçu pour prendre en compte de nombreux attributs et nombreux exemples cherchant à construire des arbres relativement simple mais ne garantissant pas de produire le plus simple.

4.3.4.2 Algorithme C4.5

L'algorithme ID3 a été amélioré pour qu'il permette de traiter des attributs quantitatifs sous l'appellation C4.5. L'algorithme C4.5 utilise une mesure d'impureté de nœud basée sur l'entropie appelée rapport de gain " gain ratio", dont le but est de limiter la prolifération de l'arbre en pénalisant les variables qui ont beaucoup de modalités. L'algorithme C4.5 a été amélioré sous l'appellation C5.0, mais étant implémentée dans un logiciel commercial, il n'est pas possible d'en avoir le détail.

Des preuves empiriques montrent que C4.5 possède une excellente vitesse et une bonne précision de prédiction, mais ses arbres sont souvent beaucoup plus grands que ceux des autres méthodes [108].

C4.5 reste le meilleur algorithme de classification parmi plusieurs, il prend la première place dans le top 10 des algorithmes d'exploration de données identifiées par la Conférence internationale IEEE sur l'exploration de données (de ICDM) en Décembre 2006 [109].

4.3.4.3 Algorithme CART

L'algorithme récursif CART (Classification And Regression Trees) permet la construction d'un arbre de décision binaire puisque chaque nœud ne peut avoir que deux fils. Son principe s'appuie sur l'indice de Gini pour la segmentation de l'arbre qui vise à construire les feuilles. Il faut noter que plus l'indice de Gini est faible, plus le nœud est pur (tous les éléments du nœud appartiennent à la même classe).

Chaque méthode doit déterminer quelle est la meilleure façon de répartir les données à chaque niveau. Les méthodes courantes pour ce faire comprennent l'indice d'impureté de Gini, le gain d'information et la réduction de la variance.

4.4 Conception et architecture du système de classification élaboré (SCAD)

Pour construire notre système de classification, nous avons suivi les étapes suivantes: Prétraitement du signal, extraction des paramètres acoustiques, apprentissage puis discrimination automatique des PN et PP (Figure 4-4).

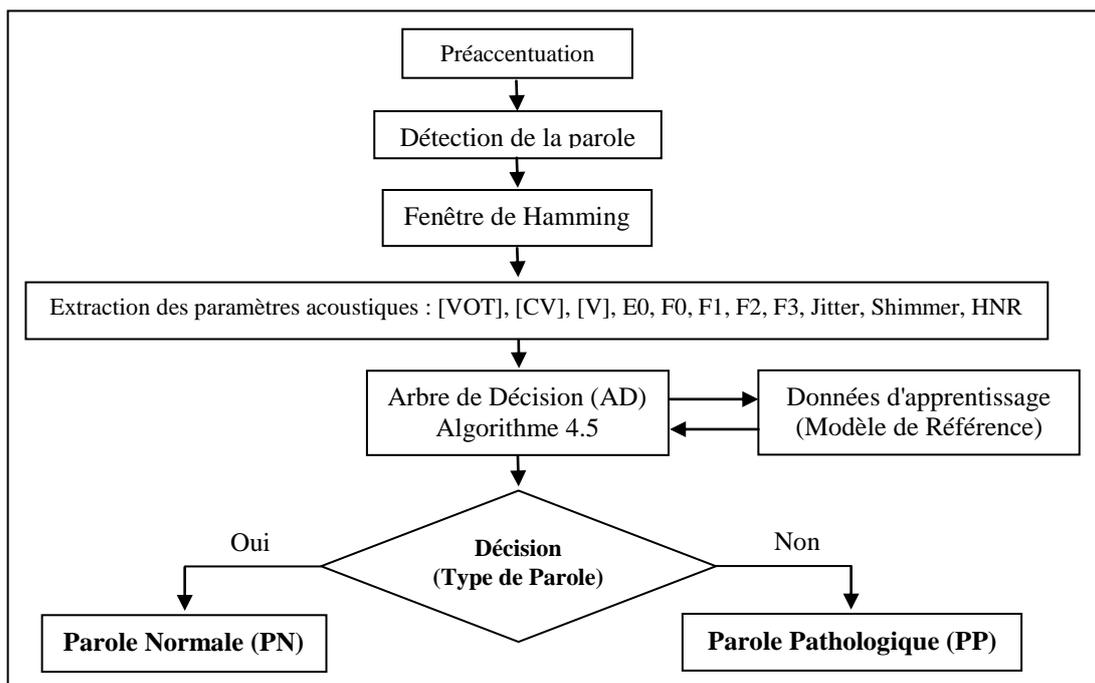


Figure 4-4 : Organigramme de classification automatique de parole Normale / Pathologique

Une des plus importantes phases de notre système est le choix adéquat des paramètres acoustiques à exploiter comme vecteurs d'entrée car il s'agit de la base du système de classification. En addition aux paramètres acoustiques ordinaires tels que la fréquence fondamentale, les formants et l'énergie (intensité), d'autres indices sont également mis en valeur, car ils permettent de discriminer la PP de la PN. Ces indices acoustiques sont le Jitter, le Shimmer et le HNR. Le dépassement du seuil normal/pathologique, pour un paramètre donné, est un moyen objectif qui permet de mettre en évidence les altérations vocales et de mesurer le degré de cette altération.

Pour mieux discriminer la PP étudiée par rapport à la PN, nous avons opté pour des paramètres assez représentatifs de la PP et dont nous avons montré l'importance dans la partie analyse acoustique. Pour la mise au point des outils du prétraitement des fichiers sons (préaccentuation, détection de parole utile et fenêtrage), ainsi que l'apprentissage et la classification automatique de la PP par rapport à la PN, nous avons utilisé le langage de programmation Matlab 2007. Il est aussi important à noter que les paramètres des vecteurs acoustiques F_0 , F_1 , F_2 , F_3 , E_0 , [VOT], [CV], [V], Jitter, Shimmer et HNR ont été conçus également à partir de ce langage.

4.4.1 Corpus utilisé

Pour extraire les fichiers sons, nous avons exploité le même corpus utilisé pour l'analyse acoustique. Ce corpus comprend de la PP et de la PN. Nous avons segmenté manuellement les mots enregistrés pour obtenir des fichiers de taille moyenne de 400 à 600 ms et en nombre assez suffisant. Nous avons divisé notre corpus des enregistrements sonores en deux ensembles : un ensemble de fichiers sonores (PP et PN) à exploiter lors de la phase d'apprentissage (Tableau 4-1) et un autre ensemble de fichiers sonores (PP et PN) différent du premier à exploiter lors de la phase de tests de classification (Tableau 4-1). En effet, le système entraîné, il est nécessaire de tester sa fiabilité sur une autre base de données différente de celle utilisée pour l'apprentissage. Ce test permet d'apprécier les performances du système.

Tableau 4-1: Fichiers sonores utilisés lors des deux phases (apprentissage et test)

Type de Parole	Voyelles Longues et Courtes					VL		VC			
	PP		PN		Total	PP		PN		Total	
	F	M	F	M		F	M	F	M		
Fichiers d'apprentissage	90	61	363	253	767	87	70	363	271	791	
Fichiers Test	41	30	177	119	367	42	27	178	129	376	

4.4.2 Prétraitement du signal

Les bons résultats qu'un classificateur automatique peut fournir reposent, en grande partie, sur la phase de prétraitement. Les données issues d'un mauvais prétraitement vont mettre en péril la qualité du classificateur. Cette phase consiste en une succession de traitements sur les données brutes afin d'extraire de l'information et de ne garder que celle qui est utile à la classification. ce qui nous permettra de réduire la taille de l'information qui va être présentée au classificateur en éliminant celle qui est non pertinente qui peut être une source de confusion pour le classificateur, ce qui se traduit par un gain en temps.

On procède tout d'abord à un filtrage numérique pour éliminer au maximum le bruit de fond. Ce dernier est surtout localisé en dessous de 100 Hz (et principalement à 50 Hz, fréquence à laquelle tournent les ventilateurs de l'ordinateur).

Dans cette étape, le signal analogique capté au moyen d'un microphone sera transformé en composantes numériques plus faciles à traiter. Dans le cadre de notre travail, le signal de parole est échantillonné à une fréquence de 44025 Hz avec une précision de 16 bits, afin de conserver l'information spectrale jusqu'à environ 5 kHz. Comme les Hautes Fréquences sont souvent atténuées, nous les renforçons artificiellement à l'aide d'un filtre numérique.

4.4.2.1 Préaccentuation

Le signal échantillonné est pré-accentué pour relever les Hautes Fréquences qui sont moins énergétiques que les Basses Fréquences. Le filtre de préaccentuation qui est souvent non récursif du premier ordre permet d'égaliser les aigus toujours plus faibles que les graves. Il est souvent sous la forme suivante :

$$H(z) = 1 - a(z)^{-1} \quad ; \quad a = 0.9 \quad (4.10)$$

4.4.2.2 Fenêtrage

La parole constitue un signal réel, continu, d'énergie finie et non stationnaire, c'est pourquoi la globalisation de son analyse à long terme est impossible. Sa multiplication par une fenêtre devient indispensable. Cette opération est réalisée sur une fenêtre glissante, chaque trame couvrant une durée de 20 ms sur laquelle le signal est supposé quasi-stationnaire. Le pas d'analyse entre deux trames successives est de 5 ms.

Le découpage du signal en trames produit des discontinuités aux frontières de trames, qui se manifestent par des lobes secondaires dans le spectre. Ces derniers sont candidats à une sommation ultérieure. Pour compenser ces effets de bord, nous multiplions en générale préalablement chaque trame par une fenêtre de pondération de type Hamming (Figure 4-5) dont l'expression mathématique est de la forme suivante :

$$H(k) = 0.54 - 0.46 \cos\left(\frac{2\pi}{N} k\right) \quad (4.11)$$

pour $k = 0, 1, 2, \dots, N-1$

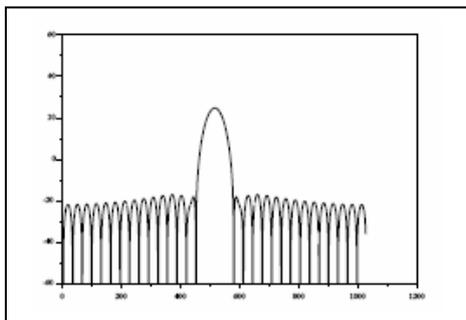


Figure 4-5 : Fenêtre de Hamming

4.4.2.3 Différents traitements effectués pour l'extraction des vecteurs acoustiques

Les méthodes que nous préconisons se basent sur une analyse temporelle du signal. Une observation minutieuse de la structure temporelle (forme d'onde), selon un certain nombre de critères, permet une segmentation fiable et précise.

Quatre algorithmes de segmentation sont présentés dans ce travail, la segmentation en :

- Voyelle / Consonne, identifie la présence d'une voyelle et calcule les trois premiers formants F_1 , F_2 et F_3 ;
- voisée / Non-voisée, permet de détecter la présence d'une excitation glottique.

4.4.2.3.1 Détection des voyelles

Les voyelles se distinguent des autres classes phonétiques par leur structure formantique ainsi qu'une énergie concentrée dans le bas du spectre. On ne tient pas compte des pics de très basse fréquence (formant glottal) vers 200 Hz et on se limite en même temps à visualiser et étudier uniquement les trois premiers formants. Dans la courbe de l'énergie, les pics dont les abscisses correspondent à des noyaux vocaliques doivent vérifier les conditions suivantes :

- leur hauteur doit être assez importante ;
- ils doivent dominer les zones de la courbe de part et d'autre ;
- ils doivent se présenter sur un intervalle de temps suffisant.

Nous proposons un algorithme de séparation consonnes voyelles basé sur la différence entre les niveaux d'énergie entre consonnes et voyelles, ainsi qu'un algorithme de reconnaissance par analyse des formants.

4.4.2.3.2 Suivi formantique par analyse prédictive

Les procédures de l'extraction des trois premiers formants et leur suivi sont accomplis grâce à un algorithme basé sur la LPC (Linear Predictive Coding). Il s'agit d'une technique de codage en vue de réduire la quantité d'information nécessaire pour produire la parole. Elle a été

utilisée avec performance dans plusieurs applications en relation avec des systèmes de Traitement Automatique de la Parole. Elle se fonde sur les connaissances de la production de la parole et suppose que le modèle de production est linéaire (Figure 4-6). Chaque échantillon de parole est décrit comme étant une somme pondérée linéairement des P échantillons qui le précèdent immédiatement et d'un signal d'erreur $e(n)$.

Le signal prédit est noté par :

$$\hat{S}_n = \sum_{k=1}^P a_k \cdot S(n-1) \quad (4.12)$$

\hat{S}_n : signal prédit ;
 a_k : coefficients de prédiction (coefficients du filtre AR) ;
 P: ordre de prédiction.

Un coefficient est calculé pour chaque échantillon afin de minimiser les erreurs entre le signal original et le signal prédit. Ces coefficients sont réactualisés régulièrement toutes les 5 à 20 ms.

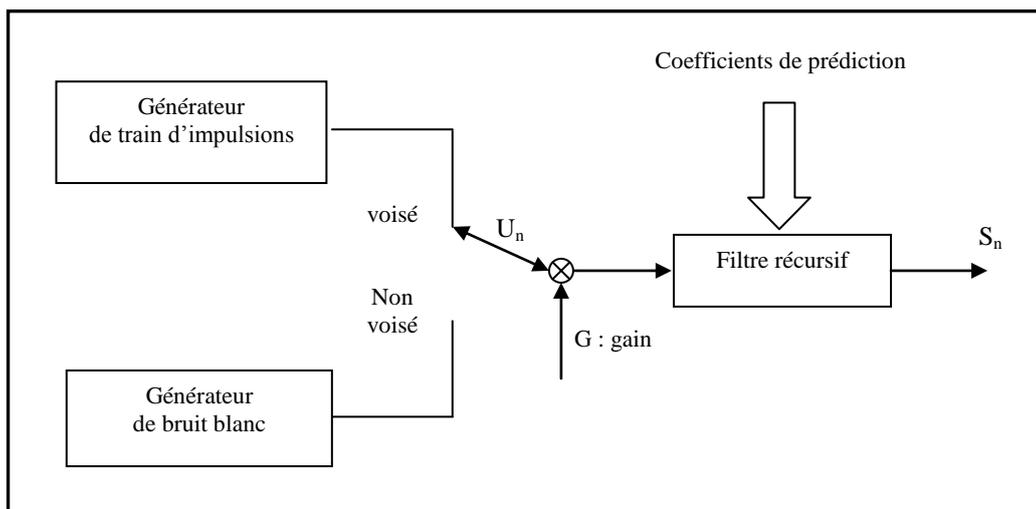


Figure 4-6 : Schéma de principe du Codage Prédicatif Linéair (LPC)

En considérant que le signal de la source est modélisé par un signal linéaire impulsionnel et que le filtre du conduit vocal est approché par un filtre tout pôle (filtre autorégressif), l'erreur de prédiction entre le signal réel $S(n)$ et le signal estimé \hat{S}_n sera donné par :

$$e(n) = S(n) - \hat{S}(n) = S(n) - \sum_{k=1}^P a_k \cdot S(n-1) \quad (4.13)$$

Ainsi, le calcul des coefficients a_k se fait par le biais d'une minimisation à court terme de $e(n)$. Pour cela, le critère retenu est celui d'une minimisation de l'erreur quadratique moyenne par la méthode des moindres carrés.

L'énergie résiduelle de prédiction est définie par la somme :

$$E_p = \sum_{k=1}^p e^2(n) = \sum_{n1}^{n2} (S(n) + \sum_{k=1}^p a_k \cdot S(n-1))^2 \quad (4.14)$$

Si les coefficients a_k sont choisis tels qu'ils minimisent l'énergie résiduelle de prédiction, il suffit pour les obtenir de poser :

$$\frac{\delta E_p}{\delta a_k} = 0 \quad i = 1, 2, \dots, p \quad (4.15)$$

Les étapes de calcul de l'équation (4.15) conduisent à la résolution des p équations de Yule -Walker suivantes :

$$\sum_{k=1}^p a_k \Phi_n(i, k) = -\Phi_n(i, 0) \quad i = 1, 2, \dots, p \quad (4.16)$$

$$\Phi(i, k) = \sum_{n=n_1}^{n_2} S(n-1) S(n-k) \quad (4.17)$$

Dans ce cas $\Phi(i, k)$ n'est autre que la fonction d'autocorrélation évaluée pour $(i-k)$ soit :

$$\Phi(i, k) = R(i-k) \quad (4.18)$$

Le système (4.18) s'écrit alors sous la forme suivante :

$$\sum_{i=1}^p a_k R(i-k) = -R(i) \quad i = 1, 2, \dots, p \quad (4.19)$$

Ce système est caractérisé par la matrice $R(i-k)$ dite matrice de Toeplitz, elle est symétrique et les éléments de la diagonale sont égaux.

Les coefficients $a(i)$ sont les coefficients du polynôme du filtre AR modélisant le conduit vocal. Nous pouvons ainsi obtenir la densité spectrale lissée du signal $S(n)$ (enveloppe spectrale due au conduit vocal). Elle est donnée par :

$$S(k) = \frac{1}{|H(k)|^2} \quad (4.20)$$

Avec
$$H(k) = \sum_{i=0}^{N-1} A(i) e^{-j\left(\frac{2\pi jk}{N}\right)} \quad (4.21)$$

Où $A = (1, a(1), a(2), \dots, a(p), 0, 0, \dots, 0)$

Le polynôme $A(z)$ possède des racines complexes conjuguées. A deux racines conjuguées z_0 et \bar{z}_0 correspond un formant dont la fréquence centrale F et la bande B à -3 dB sont données par :

$$F = \frac{1}{2\pi T} \arctg\left(\frac{\text{Re}(z_0)}{\text{Im}(z_0)}\right) \quad (4.22)$$

$$B = -\frac{1}{\pi T} \text{Log}|z_0| \quad (4.23)$$

Pour le calcul des contours correspondant aux trois premiers formants, une analyse à court terme est donc réalisée sur une fenêtre glissante de longueur $N=256$ échantillons ($\sim 20\text{ms}$) sur laquelle le signal est supposé quasi-stationnaire, pour un pas de recouvrement égal à $N/2$ (128 échantillons).

Chaque trame est représentée par 14 coefficients LPC dont l'estimation est faite par la fonction LPC intégrée à Matlab qui résout uniquement les équations de Yule-Walker en utilisant l'algorithme rapide Levison-Durbin (Figure 4.7).

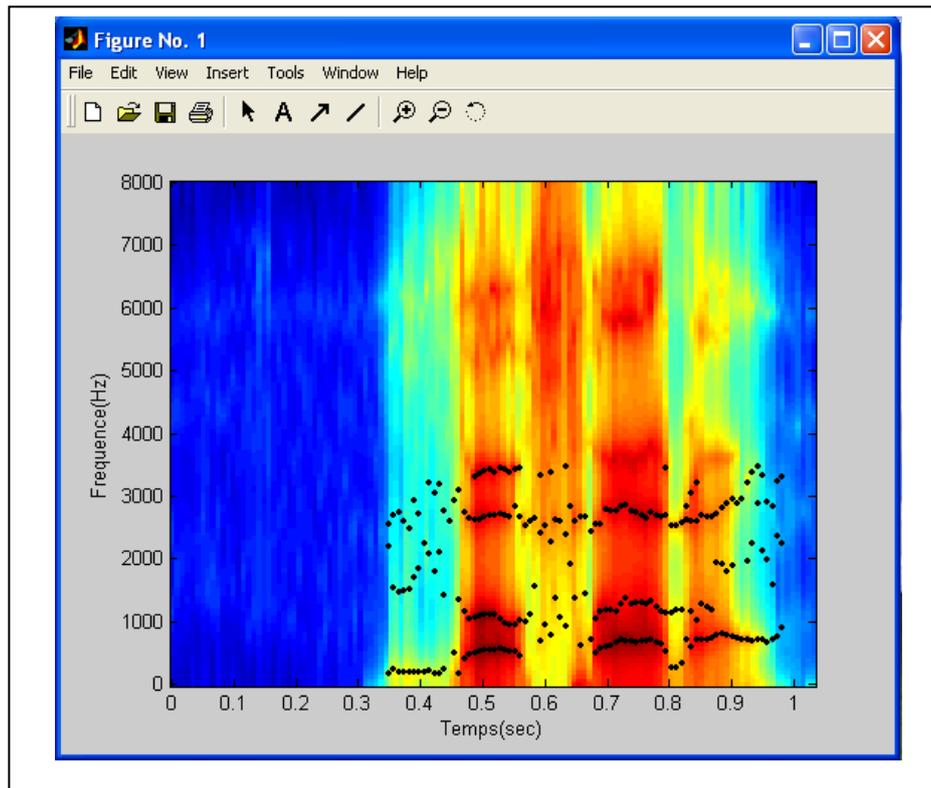


Figure 4-7 : F_1 et F_2 calculés par la méthode LPC

4.4.2.3.3 Détection du voisement par la fonction de distance (AMDF)

Le terme AMDF provient de la littérature anglo-saxonne « **A**verage **M**agnitude **D**ifference **F**unction » qui signifie fonction de la valeur moyenne de la valeur absolue de la différence. Dans cette perspective on admet que la différence d'une fonction avec elle-même décalée, caractérise le degré de similitude de celle-ci avec sa courbe décalée.

La fonction AMDF est basée sur le même principe que la fonction d'autocorrélation, elle ne mesure pas la ressemblance de deux portions de signal, mais leur différence. Là où la ressemblance est maximale, la différence est minimale et la fonction AMDF doit présenter des minima là où la fonction d'autocorrélation présente des maxima. Donc le critère de variation d'amplitude à court terme au lieu de calculer la corrélation entre deux signaux, utilise la valeur absolue des différences point par point.

La fonction de périodicité de l'AMDF est :

$$FP_{AMDF}(\tau) = \sum_{i=1}^n |s_i - s_{i+\tau}| \quad (4.24)$$

Il est clair que si l'onde était parfaitement périodique on devrait observer :

$$D(iP) = 0, \quad i = 0, 1, \dots$$

La pratique montre que l'estimation de P par la recherche d'un creux dans la suite $D(k)$ reste aisée malgré la non-stationnarité. Cette fonction est ensuite normalisée par n ou par $\sum_{i=1}^n |s_i|$ pour que la valeur de l'AMDF puisse être comparée à un seuil absolu dans le but de décider si le signal est périodique ou non.

La fonction de périodicité présente un minimum au niveau des multiples de la période. Cette méthode n'utilise pas l'hypothèse de stationnarité du signal. D'ailleurs, l'ambiguïté entre les pics $T_0, 2T_0, 3T_0, \dots$ est souvent atténuée par la non-stationnarité du signal analysé : plus le décalage est grand, plus le signal, de part sa non-stationnarité, intègre des différences par rapport à la trame de départ. Le signal présente alors plus de différences pour un décalage de $2T_0$ que pour un décalage de T_0 .

Une conséquence de la non utilisation de l'hypothèse de stationnarité est que le choix de la taille des fenêtres du signal et du signal décalé est libre : l'AMDF utilise deux fenêtres de taille fixe, mais il est possible de concevoir des algorithmes avec des tailles de fenêtres variables, par exemple égale au décalage testé.

La solution apportée au problème des erreurs grossières et la rapidité de calcul font de l'AMDF une méthode couramment employée.

4.4.3 Algorithme de base C4.5

Le rôle de notre système de classification est de classer des objets qui vont avoir plusieurs caractéristiques (attributs). En fonction de ces attributs que l'on va explorer et tester, on va finir par attribuer une classe à chaque objet. Chacun de ces objets est appelé exemple. On va donc manipuler des exemples composés d'attributs pour leur attribuer une classe. Il existe deux types d'exemples : les exemples étiquetés et les exemples non étiquetés. Les exemples étiquetés sont ceux qui sont pré-classés et qui donc vont être utilisés pour que le programme sache comment classer les exemples non étiquetés. La méthode qui va être utilisée pour déterminer comment étiqueter les exemples non étiquetés sur base des exemples étiquetés s'appelle l'algorithme d'apprentissage.

Pour construire notre AD, nous avons utilisé l'algorithme d'apprentissage C4.5 qui est une extension et amélioration de l'algorithme ID3.

Pour décider quelle est la meilleure caractéristique afin de diviser l'ensemble de données, C4.5 utilise une extension du gain d'information connue par rapport de gain. Lorsqu'on a une caractéristique ait un grand nombre de valeurs, elle sera favorisée par le gain d'information.

On dispose en entrée d'un ensemble d'apprentissage A. On utilise la fonction entropie définie précédemment. Soit p une position, on choisit alors le test qui maximise le gain en utilisant la fonction entropie pour mesurer le degré du mélange. La fonction Gain, ainsi définie, privilégie les attributs ayant un grand nombre de valeurs.

Elle est donc pondérée par une fonction qui pénalise les tests qui répartissent les éléments en un trop grand nombre de sous-classes. Cette mesure de la répartition est nommée SplitInfo et est définie par :

$$SplitInfo(p, test) = -\sum_{j=1}^n P' \left(\frac{j}{p} \right) \times \log \left(P' \left(\frac{j}{p} \right) \right) \quad (4.25)$$

Dans laquelle n est l'arité de test, et $P' \left(\frac{j}{p} \right)$ est la proportion des éléments présents à la position p prenant la j^{ème} valeur de test.

Contrairement à l'entropie, la définition précédente est indépendante de la répartition des exemples à l'intérieur des différentes classes.

La valeur de *SplitInfo* ne dépend que de la répartition entre les différentes valeurs possibles pour le test. Cette fonction a des valeurs grandes lorsque le test a un grand nombre de valeurs possibles avec peu d'éléments pour chacune des valeurs. En effet, considérons le cas extrême d'un attribut n-aire avec un exemple par classe, la fonction vaut alors log n. A l'inverse, considérons le cas d'un attribut binaire pour lequel les exemples sont répartis uniformément entre ces deux valeurs, la fonction vaut alors 1.

La nouvelle fonction de gain, appelée rapport de gain notée "GainRatio", est alors définie par :

$$GainRatio(p, T) = \frac{Gain(p, T)}{SplitInfo(p, T)} \quad (4.26)$$

En position p (non maximale), on choisit le test qui maximise le GainRatio (option par défaut de C4.5). On peut modifier les options pour utiliser le Gain : Affecter une classe à une feuille. On attribue la classe majoritaire. S'il n'y a aucun exemple on attribue la classe majoritaire du nœud père.

L'arbre de décision est une méthode classique d'apprentissage s'utilisant de manière similaire : le processus entier se compose de deux phases séparées, apprentissage et test. L'apprentissage consiste à construire un modèle représentant un ensemble des éléments, alors que le test utilise ce modèle pour évaluer un nouvel élément inconnu [110].

4.4.3.1 Phase d'apprentissage

L'apprentissage automatique est une tentative de comprendre et reproduire cette faculté d'apprentissage dans des systèmes artificiels. Il s'agit, très schématiquement, de concevoir des algorithmes capables, à partir d'un nombre important d'exemples, d'en assimiler la nature afin de pouvoir appliquer ce qu'ils ont ainsi appris aux cas futurs [111].

La phase d'apprentissage permet au système de lire les paramètres de référence, représentant les fichiers audio enregistrés utilisés dans l'analyse acoustique destinés à cette phase. Ces vecteurs de références sont obtenus à partir de modèles acoustiques qui permettent de caractériser les différents sons prononcés. Le principe est de fournir à notre classificateur une série d'exemples x et de résultats y , et ensuite trouver un modèle adapté au résultat escompté, pour avoir un bon taux de reconnaissance et surtout une bonne généralisation. Grâce aux exemples appris, le système est capable de traiter des exemples distincts, encore non rencontrés, mais similaires.

Pour notre classificateur, nous faisons un apprentissage en utilisant une matrice donnée de vecteurs acoustiques extraits à partir d'un ensemble de 767 fichiers acoustiques pour les VL (151 fichiers PP et 616 fichiers PN) (Tableau 4-2) et 791 fichiers acoustiques pour les VC (157 fichiers PP et 634 fichiers PN) (Tableau 4-3).

Tableau 4-2 : fichiers sonores utilisés dans l'apprentissage pour les VL

Type de Parole	PP		PN	
Genre	F	M	F	M
N ^{bre} de fichiers	90	61	363	253
Total des fichiers	767			

Tableau 4-3: fichiers sonores utilisés dans l'apprentissage pour les VC

Type de Parole	PP		PN	
Genre	F	M	F	M
N ^{bre} de fichiers	87	70	363	271
Total des fichiers	791			

Dans notre AD, les différents paramètres acoustiques étudiés, pour les deux types de voyelles VC et VL, représentent l'ensemble d'attributs (indicateurs) et chaque attribut aura deux valeurs possibles : Pathologique ou Normale. Nous avons introduit deux paramètres en plus qui sont l'âge et le sexe.

Au départ, on définit le premier attribut à utiliser pour notre AD (Nœud Zéro), nous calculons pour chaque indicateur : l'entropie, le Gain d'information, le SplitInfo et le Gain Ratio (Tableau 4-4).

Tableau 4-4 : Calcul des GainRatio pour tous les indicateurs pour choisir l'attributs du nœud Zéro

	Différents Indicateurs (Attributs)							
	VOT	CV	V	F ₀	E ₀	F ₁	F ₂	F ₃
Entropie	0,6768	0,6821	0,7082	0,6436	0,6809	0,6917	0,6713	0,6746
Gain d'Information	0,0421	0,0368	0,0107	0,0753	0,0380	0,0272	0,0476	0,0443
SplitInfo	0,873	0,861	0,914	0,943	1,000	0,950	0,991	0,958
Gain Ratio	0,048	0,043	0,012	0,080	0,038	0,029	0,048	0,046

Nous remarquerons que la valeur de **0,08** est la valeur maximale du Gain Ratio, impliquant que notre premier nœud démarrera par l'indicateur " **Pitch**" qui a en sortie deux branches, l'une pour les valeurs pathologiques du pitch et l'autre pour ces valeurs normaux.

L'algorithme C4.5 utilise l'approche Top- Down, la stratégie la plus utilisée sous le nom "Top Down Induction of Decision Tree (TDIDT)" (Figure 4-8), qui consiste à construire un arbre depuis sa racine vers ses feuilles en partitionnant successivement la base d'apprentissage par une technique de validation suivant une classe d'algorithmes Illustrée suivant la procédure suivante :

Début

Si tous les exemples de « X » sont de la même classe

Alors Création de la feuille portant le nom de cette classe

Sinon Choisir un attribut pour la création du nœud suivant le test associé pour définir le Xdroite du Xgauche

Construire Arbre (Xdroit)

Construire Arbre (Xgauche)

Fin

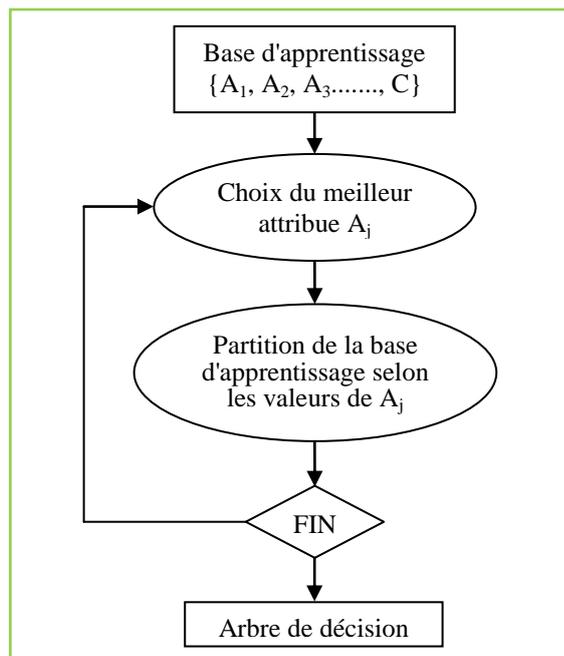


Figure 4-8 : Construction d'un arbre de décision par stratégie TDIDT

L'étape suivante dans la construction de notre AD est de choisir les attributs des deux nœuds qui relient les deux branches sortants du nœud zéro en calculant pour chacun des deux branches l'entropie, le Gain d'information, le SplitInfo et le GainRatio pour tout le reste des indicateurs (à l'exception du nœud Père dont l'indicateur est le Pitch dans ce cas) (Tableau 4-5 et 4-6).

Tableau 4-5 : Calcul des Gain Ratio pour tous les indicateurs pour choisir l'attribut du nœud suivant le nœud Zéro "Pitch" selon les valeurs Normaux.

	Différents Indicateurs (Attribus)							
	VOT	CV	V	F ₀	E ₀	F ₁	F ₂	F ₃
Entropie	0.2446	0.2476	0.2515	-	0.2395	0.2455	0.2214	0.2517
Gain d'Information	0.0073	0.0043	0.0003	-	0.0124	0.0063	0.0305	0.0001
SplitInfo	0.847	0.810	0.866	-	0.990	0.866	0.994	0.883
Gain Ratio	0.009	0.005	0	-	0.013	0.007	0.031	0

Nous remarquerons que la valeur de **0,031** est la valeur maximale du Gain Ratio, impliquant que le premier nœud fils du nœud zéro dont l'indicateur "**Pitch**" (Figure 4-9) aura comme indicateur "**F₂**" et qui aura en sortie lui aussi deux branches, l'une pour les valeurs pathologiques du deuxième formant F₂ et l'autre pour ces valeurs normales (Figure 4.9).

Tableau 4-6 : Calcul des GainRatio de tous les indicateurs pour choisir l'attribut du nœud suivant le nœud Zéro "Pitch" selon ces valeurs Pathologiques.

	Différents Indicateurs (Attribus)							
	VOT	CV	V	F ₀	E ₀	F ₁	F ₂	F ₃
Entropie	0.8009	0.8122	0.8509	-	0.8209	0.8374	0.8212	0.8040
Gain d'Information	0.0633	0.0520	0.0133	-	0.0434	0.0268	0.0430	0.0602
SplitInfo	0.887	0.887	0.936	-	0.995	0.979	0.964	0.983
Gain Ratio	0.071	0.059	0.014	-	0.044	0.027	0.045	0.061

Nous remarquerons que la valeur de **0,071** est la valeur maximale du Gain Ratio, impliquant que le deuxième nœud fils du nœud zéro dont l'indicateur "**Pitch**" aura comme indicateur "**VOT**" et qui aura en sortie lui aussi deux branches, l'une pour les valeurs pathologiques du VOT et l'autre pour ces valeurs normales (Figure 4-9).

Le processus de décision continue de cette manière dans notre AD (de la racine vers l'une des feuilles) : à chaque étape un attribut est testé et aura toujours deux branches avec deux valeurs distinctes, valeurs normales et valeurs pathologiques et un sous-arbre est choisi. Ce processus se répète d'une manière récursive et itérative jusqu'à ce qu'aucune séparation supplémentaire des données ne soit réalisable ou qu'une classification unique puisse être appliquée à chaque membre des sous-groupes dérivés c'est à dire jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant tous à une même classe. Le parcours s'arrête dans une feuille (une décision est prise).

Par conséquent, les nœuds terminaux à la fin des branches de l'arbre de décision représentent les différentes classes. Les attributs des feuilles de notre AD sont des valeurs qui correspondent à une classification dans l'ensemble { PP, PN }.

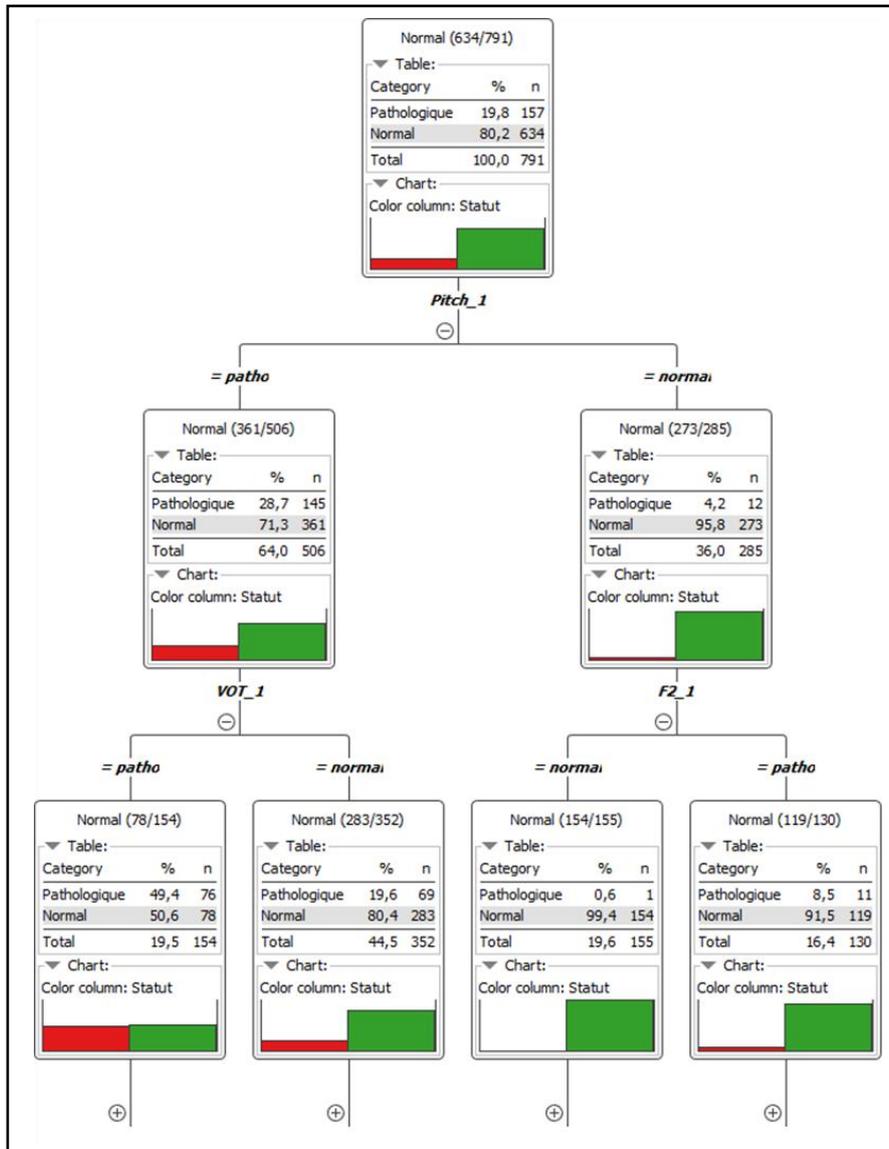


Figure 4-9 : Calcul des trois premiers Nœuds de l'AD

4.4.3.2 Phase de Classification

Cette phase permet de se rendre compte du pouvoir de généralisation du classificateur, c'est-à-dire sa capacité à obtenir de bons résultats sur n'importe quel ensemble de données provenant de la même distribution.

Une fois qu'on a fait grandir l'arbre, la phase de test consiste simplement à faire passer chacune des données de test par l'arbre de décision. A partir du nœud racine, chaque nœud appliquera une fonction de triage à l'exemple x qui sera envoyé, selon le résultat de cette fonction, vers un des nœuds suivants dans la hiérarchie. Ce processus est répété jusqu'à arriver à un nœud feuille qui assignera sa classe d'appartenance.

Pour valider nos tests de classification, nous avons utilisé un ensemble de fichiers sonores des locuteurs qui ont participé à nos enregistrements, contenant 367 fichiers acoustiques pour les VL (71 fichiers PP et 296 fichiers PN) (Tableau 4-7) et 376 fichiers acoustiques pour les VC (69 fichiers PP et 307 fichiers PN) (Tableau 4-8).

Tableau 4-7 : Fichiers sonores utilisés dans le test pour les VL

Type de Parole	PP		PN	
Genre	F	M	F	M
N ^{bre} de fichiers	41	30	177	119
Total des fichiers	367			

Tableau 4-8 : Fichiers sonores utilisés dans le test pour les VC

Type de Parole	PP		PN	
Genre	F	M	F	M
N ^{bre} de fichiers	42	27	178	129
Total des fichiers	376			

Pour les tests de notre classificateur, nous avons suivi les étapes allant de la lecture des sons acoustiques jusqu'aux tests de classification, en incluant les différentes étapes de la détection des frontières des mots, de la préaccentuation, et de l'extraction des caractéristiques. Les paramètres extraits lors de l'analyse acoustique seront introduits dans notre classificateur automatique qui permettra de déterminer la nature de la parole prononcée.

Le comportement de notre classificateur automatique est évalué en termes de pourcentage de classification correcte de l'ensemble des fichiers tests présentés à l'entrée. La méthode classique pour calculer le taux de reconnaissance (TR) des fichiers tests est donnée par l'équation classique :

$$TR (\%) = \left(\frac{\text{Cas correctes}}{\text{Total}} \right) \times 100 \quad (4.27)$$

4.4.4 Résultats expérimentaux

Nous avons réalisé une classification automatique de la PP de locuteurs porteurs de fente palatine par rapport à la PN. Dans une première étape, nous avons fait une classification quantitative des valeurs obtenues de l'analyse acoustique directement (Tableau 4-9).

Tableau 4-9 : Classification sans prise en charge des Seuils

Voyelles Courtes et Longues	VC		VL	
Statut de la parole	PP	PN	PP	PN
PP	46	23	48	22
PN	39	267	38	257

Dans une seconde étape, nous avons fait une classification avec association du sexe, de l'âge à la classification de ces types de fente étudiés, en introduisant les seuils des paramètres

acoustiques obtenus de l'étude statistique faite avec la courbe ROC sur l'analyse acoustique de la PP (Tableau 4-10).

Tableau 4-10 : Classification avec prise en charge des Seuils des classes d'âge et du sexe

Voyelles Courtes et Longues	VC		VL	
	PP	PN	PP	PN
Statut de la parole				
PP	52	17	58	12
PN	30	276	33	262

Pour le cas des VC, nous remarquons que sur 306 fichiers sonores correspondant à la PN, 267 ont été reconnus correctement alors que 39 fichiers ne sont pas reconnus et ont été considérés comme PP. De même, sur 69 fichiers de PP, 46 ont été reconnus et 23 reconnus comme PN. En introduisant les seuils des classes d'âge et du sexe à notre système de classification, nous remarquons que sur 306 fichiers sonores correspondant à PN, 276 ont été reconnus correctement alors que 30 fichiers ne sont pas reconnus et considérés comme PP. De même, sur 69 fichiers de PP, 52 ont été reconnus et 17 reconnus comme PN.

Pour le cas des VL, nous remarquons que sur 295 fichiers sonores correspondant à la PN, 257 ont été reconnus correctement alors que 38 fichiers ne sont pas reconnus et considérés comme PP. De même, sur 70 fichiers de PP, 48 ont été reconnus et 22 reconnus comme PN. En introduisant les seuils des classes d'âge et du sexe à notre système de classification, nous remarquons que sur 295 fichiers sonores correspondant à la PN, 262 ont été reconnus correctement alors que 33 fichiers ne sont pas reconnus et considérés comme PP. De même, sur 70 fichiers de PP, 58 ont été reconnus et 12 reconnus comme PN.

En récapitulant les résultats, nous obtenons les résultats suivants (Tableau 4-11 et 4-12), comme Taux de Reconnaissance TR (%) :

Tableau 4-11 : Taux de Reconnaissance (TR) des PP sans prise en charge des Seuils

Voyelles Courtes et Longues	VC		VL	
	PP	PN	PP	PN
Statut de la parole				
PP	0.66	0.33	0.68	0.32
PN	0.13	0.88	0.13	0.87

Tableau 4-12: Taux de Reconnaissance (TR) avec prise en charge des Seuils des classes d'âge et du sexe

Voyelles Courtes et Longues	VC		VL	
	PP	PN	PP	PN
Statut de la parole				
PP	0.75	0.25	0.83	0.17
PN	0.1	0.9	0.11	0.89

Selon les résultats obtenus, nous constatons que le système proposé, basé sur les Arbres de Décision (AD), donne un bon pourcentage de reconnaissance de la PP par rapport à la PN.

Notre système de classification répond positivement au bon choix des paramètres acoustiques étudiés composant les Vecteurs d'Entrée VE.

Nous avons obtenu de meilleurs résultats en introduisant les seuils des différents paramètres acoustiques obtenus par la courbe ROC en plus de ceux des différentes classes d'âge ainsi que celui du sexe VE : (S_{VOT} , S_V , S_{CV} , S_{E0} , S_{F0} , S_{F1} , S_{F2} , S_{F3} , $S_{\text{classe d'âge}}$, S_{sexe}) pour les VC (Tableau 4-12) et VE: (S_{VOT} , S_V , S_{CV} , S_{E0} , S_{F0} , S_{F1} , S_{F2} , S_{F3} , S_{Jitter} , S_{Shimmer} , S_{HNR} , $S_{\text{classe d'âge}}$, S_{sexe}) pour les VL (Tableau 4-12), avec un taux d'erreurs de 0.17% pour les VL et 0.25% pour les VC (Tableau 4-12).

4.5 D'autres cas de classifications

Pour évaluer les performances de notre système de discrimination entre la PP et la PN, nous avons essayé d'appliquer la même procédure de classification mais cette fois en changeant à chaque fois les paramètres acoustiques composant le vecteur d'entrée (VE) de notre système en fonction des deux critères séparés âge et sexe mais séparés. Les classes étudiées sont :

- l'ensemble total des locuteurs (M+F) tout âge confondu ;
- l'ensemble total des locuteurs (M+F) répartis selon les tranches d'âge : enfants, adolescents et adultes;
- l'ensemble total des locuteurs répartis selon le sexe : Locutrices et locuteurs de différents âge.

Nous avons calculé les seuils des paramètres acoustiques étudiés pour ces différentes classes de locuteurs (Tableau 4-13), en utilisant la courbe ROC. A chaque étape, nous faisons un apprentissage en utilisant une matrice donnée de vecteurs acoustiques extraits à partir d'un ensemble de fichiers apprentissage contenant des fichiers de PP et de PN, puis nous passons à la classification automatique, en utilisant un autre ensemble de fichiers de paroles différents de celui d'apprentissage, composé de PP et de PN. Ensuite, nous comparons les Taux d'Erreurs TR pour ces différentes variantes, ceci pour les VC et les VL.

Tableau 4-13 : Valeurs Seuils des paramètres acoustiques pour les différentes classes étudiées - Cas des VL

Classification des locuteurs	Paramètres acoustiques	VOT	CV	V	E ₀	F ₀	F ₁	F ₂	F ₃	Jitter	Shimmer	HNR
Selon l'âge	Classe Globale	0,015	0,303	0,168	53,734	243,35	1018	1196	2726	0,339	2,402	17,414
	Enfants (M+F)	0.012	0.314	0.166	58.013	316.56	1006	1425	2942	0.359	3.464	15.858
	Adolescents (M+F)	0.012	0.176	0.093	52.401	262.77	819	1302	3302	0.529	1.655	16.609
	Adultes (M+F)	0.007	0.151	0.101	54.994	149.57	859	1352	2715	0.459	2.4	19.121
Selon le sexe	Classe Globale F	0,016	0,33	0,17	54,19	248,71	1018	1473	3729	0,436	2,453	20,797
	Classe Globale M	0,015	0,227	0,194	53,819	124,22	971	1193	2416	0,3	5,087	17,414

4.5.1 Résultats

Les résultats trouvés sont illustrés dans les tableaux 4 -14 jusqu'à 4 -17 :

Tableau 4-14 : Classification pour les différentes classes étudiées pour les VC

Statut de la parole	Classe Globale avec prise en charge des Seuils		prise en charge des Seuils des classes d'âge		prise en charge des Seuils selon le sexe	
	PP	PN	PP	PN	PP	PN
PP	43	26	40	29	52	17
PN	18	288	25	281	28	278

Tableau 4-15 : Classification pour les différentes classes étudiées pour les VL

Statut de la parole	Classe Globale avec prise en charge des Seuils		prise en charge des Seuils des classes d'âge		prise en charge des Seuils selon le sexe	
	PP	PN	PP	PN	PP	PN
PP	56	14	54	16	50	20
PN	24	271	26	269	23	272

En récapitulant les résultats, nous obtenons les résultats suivants, comme Taux de Reconnaissance TR (%) :

Tableau 4-16 : Taux de Reconnaissance pour les différentes classes étudiées pour les VC

Type de parole	prise en charge des Seuils des classes d'âge et du sexe	Classe Globale avec prise en charge des Seuils	prise en charge des Seuils des classes d'âge	prise en charge des Seuils selon le sexe
PP	0.25	0.62	0.42	0.25
PN	0.1	0.06	0.08	0.09

Tableau 4-17 : Taux de Reconnaissance pour les différentes classes étudiées pour les VL

Type de parole	prise en charge des Seuils des classes d'âge et du sexe	Classe Globale avec prise en charge des Seuils	prise en charge des Seuils des classes d'âge	prise en charge des Seuils selon le sexe
PP	0.17	0.2	0.23	0.29
PN	0.11	0.08	0.09	0.08

En comparant les résultats obtenus de la classification de la PP par rapport à la PN pour les différentes classes, nous pouvons dire que les meilleurs résultats sont ceux obtenus en introduisant les seuils des différents paramètres acoustiques obtenus par la courbe ROC des classes d'âge et du sexe.

Nous pouvons ainsi déduire des résultats trouvés que les paramètres acoustiques utilisés dans les VE : (S_{VOT} , S_V , S_{CV} , S_{E0} , S_{F0} , S_{F1} , S_{F2} , S_{F3} , $S_{classe\ d'âge}$, S_{sexe}) pour les VC (Tableau 4-12) et VE: (S_{VOT} , S_V , S_{CV} , S_{E0} , S_{F0} , S_{F1} , S_{F2} , S_{F3} , S_{Jitter} , $S_{Shimmer}$, S_{HNR} , $S_{classe\ d'âge}$, S_{sexe}) pour les VL (Tableau 4-12), sont très pertinents dans la discrimination des PP [112].

En se basant sur ces résultats, nous pouvons dire que le système de reconnaissance proposé, à base d'arbres de décision, donne un bon pourcentage de reconnaissance de PP par rapport aux PN. De plus, en comparant les différents résultats obtenus pour les différentes

classes étudiées, nous pouvons dire que les meilleurs résultats sont ceux obtenus avec prise en charge des Seuils des différents paramètres acoustiques obtenus par la courbe ROC des classes d'âge et du sexe en tant que caractéristiques acoustiques. Cela montre que ces dernières caractéristiques sont pertinentes dans la discrimination de la parole pathologique par rapport à la parole normale.

4.6 Conclusion

Les résultats obtenus nous montrent que les arbres de décision s'adaptent bien à la reconnaissance et à la classification des voix pathologiques.

L'arbre de décision est une des méthodes de prise de décision les plus efficaces. Elle permet non seulement de présenter visuellement des informations mais aussi de les hiérarchiser. C'est un outil qui facilite grandement nos décisions et limite le sentiment de surcharge informationnelle. Cette méthode de décision est couramment utilisée dans le milieu informatique, médical et scientifique pour explorer et exploiter des données au quotidien.

Les arbres de décisions sont facilement interprétables, toutefois, la capacité de prévision qu'ils ont est presque toujours dépassée par les autres modèles de classification. Cette caractéristique a limité son utilisation jusqu'au début des années 2000, puis ils ont été repris comme élément de base d'une nouvelle technique, appelée la forêt de décision. Cette nouvelle technique utilise de manière combinée les arbres de décision et la théorie statistique pour réduire la variance du classeur en calculant la moyenne d'un ensemble d'arbres de décision en générant des classeurs avec une très bonne capacité de prévision.

CONCLUSIONS GENERALES ET PERSPECTIVES

Notre travail s'intéresse essentiellement à l'analyse acoustique et la classification automatique de la Parole Pathologique (PP) produite par des locuteurs porteurs de Fente Palatine (FP) ou Labio-Palatine (FLP). Il s'agit d'un trouble pathologique de l'appareil phonatoire, appelé fente palatine ou labio-palatine et qui touche un nombre important de personnes dans le monde.

La problématique des fentes palatines est très répandue. Certains troubles de la parole liés à ce type de pathologie sont persistants. Rappelons que les locuteurs porteurs d'une fente palatine sont le plus souvent sujets à des insuffisances vélares. Ces dernières empêchent les locuteurs de réunir la pression intra-orale nécessaire à la production de certains phonèmes notamment les occlusives. Il est possible de parler de ratage de cible articulaire chez ces locuteurs mais il faut surtout noter que leur cible n'est pas la même que celle des locuteurs sains. En effet, il semble que chaque locuteur établisse sa cible en fonction de la nature de la reconstruction du palais, ce qui correspondrait donc à une stratégie individuelle.

Cette étude est importante dans les milieux Hospitaliers Algériens, car elle nous permet de mettre à la disposition de l'orthophoniste un ensemble de données concrètes qui pourront l'aider à la prise en charge des patients, notamment les enfants qui trouvent de grandes difficultés scolaires et qui doivent avoir une prise en charge précoce pour anticiper sur d'éventuels échecs scolaires et éviter ainsi les conséquences graves voire dramatiques sur le développement psychique et également le devenir social et professionnel de ces enfants.

Pour réaliser notre objectif, nous avons exploité un corpus de 448 fichiers sonores enregistrés dans le Centre Hospitalo-Universitaire (CHU) de Beni Messous à Alger contenant les cas pathologiques en comparaison avec leurs homologues sains pris comme normes de contrôle. Nous avons préalablement segmenté les signaux de parole manuellement, en utilisant l'outil d'analyse Praat. Ce corpus est divisé en deux sous corpus l'un comportant 308 fichiers sonores pour la phase d'apprentissage du système et l'autre de 140 fichiers pour la phase du test.

Dans une première étape, nous avons réalisé une analyse acoustique sur ce corpus contenant de la Parole Pathologique étudiée. Pour cela, nous avons extrait les paramètres acoustiques tels que la fréquence fondamentale F_0 , les trois premiers formants F_1, F_2 et F_3 qui permettent de caractériser l'influence que subit le son lors de son passage à travers les cavités de l'appareil phonatoire, les durées du VOT, des séquences (C: Consonne et V: Voyelle) [CV] et [V] et l'énergie E_0 . Nous avons également fait appel à d'autres paramètres importants dans le domaine de la PP, à savoir les degrés de perturbation de F_0 ou Jitter et de l'intensité ou Shimmer ainsi que le HNR (Harmonics to Noise Ratio). Les résultats obtenus montrent une présence de nombreux troubles chez ces enfants porteurs de fentes dans la réalisation du

phonème [t] tels que les perturbations au niveau des formants et la prolongation des durées de sa réalisation par rapport à leurs similaires sains. La mesure de ces paramètres vocaux permet de faire des comparaisons entre les sujets et de suivre l'évolution de la production de la parole chez un patient porteur de fente.

Les résultats obtenus montrent une perturbation des formants F_1 , F_2 et F_3 et une stabilité des valeurs de F_0 . Un autre aspect important observé est l'augmentation de la valeur du VOT pour les trois types de fentes, due aux difficultés de contrôle de fermeture des occlusives et au relâchement subséquent. Ces difficultés de contrôle de fermeture se sont également reflétées dans les durées de [CV] et de la voyelle suivante [V] qui sont plus élevées.

Cette analyse acoustique nous a permis de vérifier la pertinence des mesures acoustiques pour évaluer les perturbations dans la voix des patients porteurs de fente. Elle permet ainsi d'améliorer la rééducation de ces patients et évaluer, de façon objective, l'évolution de leurs rééducation au cours du temps, en offrant des données concrètes à l'orthophoniste pour le guider dans la conception de programmes appropriés pour ces cas.

De plus, dans cette étude, la résonance nasale et l'émission nasale ont été évaluées perceptivement à l'aide d'une classification précédemment publiée qui n'a pas été entièrement validée quant à sa fiabilité. Il sera nécessaire de réaliser de futures études utilisant des évaluations instrumentales, y compris des mesures de nasométrie et d'aérodynamique pour évaluer la résonance et l'émission nasales.

Dans une dernière étape, nous avons réalisé une classification automatique de la PP par la méthode des AD qui est une des méthodes de prise de décision les plus efficaces, car elle permet non seulement de présenter visuellement des informations mais aussi de les hiérarchiser. Cette méthode de décision est couramment utilisée dans le milieu informatique, médicale et scientifique pour explorer et exploiter des données au quotidien. L'avantage principal des AD est qu'ils permettent de faire une représentation graphique du processus de décision facile à interpréter pour toute personne et beaucoup de personnes les considèrent comme un miroir du processus humain de prise de décision.

Dans notre AD, les différents paramètres acoustiques étudiés, pour les deux types de Voyelles Courtes (VC) et Longues (VL), représentent l'ensemble d'attributs (indicateurs) et chaque attribut aura deux valeurs possibles : Pathologique ou Normale. La raison pour laquelle, nous avons utilisé l'algorithme C4.5 qui est l'un des meilleurs algorithmes de classification [108].

Les résultats obtenus nous montrent que les AD s'adaptent bien à la reconnaissance et à la classification des voix pathologiques. Notre système de classification basée sur les AD,

utilisant ces caractéristiques acoustiques (VOT, [CV], [V], E₀, F₀, F₁, F₂, F₃, Jitter, Shimmer et HNR), nous a donné une performance appréciable avec une précision de reconnaissance de 74% des PP et 81% des PN (Paroles Normales).

En perspectives, il est évident que les résultats obtenus doivent être validés sur un nombre élevé de locuteurs et de Paroles Pathologiques pour aboutir à des résultats plus représentatifs et plus fiables du système de classification élaboré. D'autres paramètres acoustiques et articulatoires mériteraient d'être explorés afin de compléter nos résultats et d'apporter davantage de poids et de précisions aux interprétations proposées dans ce travail.

Le but ciblé par ce travail étant de proposer une approche de détection automatique des phonèmes anormaux dans la parole produite par un locuteur porteur de fente. Nous envisageons d'intégrer cette application dans un système d'aide au diagnostic médical.

Certaines études ont démontré le lien entre la dimension et l'emplacement de la fente et le lieu de l'articulation du son prononcé. Seule l'utilisation des données cinéradiographiques pourra apporter des renseignements précis sur la forme du conduit vocal, l'articulation des phonèmes ainsi que leurs caractéristiques.

Cette étude mériterait d'être poursuivie avec un nombre de locuteurs plus élevé en tenant compte du genre (Masculin, Féminin), de l'âge des locuteurs ainsi que du type de la pathologie afin d'élaborer une plus large Base de Données (BD) orientée vers la parole pathologique.

REFERENCES BIBLIOGRAPHIQUES

-
- [1] Y. A. Alotaibi, « Investigating spoken Arabic digits in speech recognition setting », *Information Sciences*, vol. 173, n° 1-3, p 115-139, 2005. DOI: 10.1016/j.ins.2004.07.008
- [2] Clemedicine, « Production de la Parole » [en ligne]. consulté le: 07/12/ 2021 . Disponible sur : <https://clemedicine.com/1-production-de-la-parole>
- [3] C. Fressard, F. Pellegrino, et G. Puech, « 3ème Partie - Chapitre 1 : L'appareil de production de la parole » [en ligne]. consulté le 18/11/2021. Disponible sur: <http://www.cyberphon.ddl.cnrs.fr/Partie3/P3C1.htm>
- [4] V. Woisard-Bassols, « Bilan clinique de la voix », *Encyclopedie Medico Chirurgicale* [en ligne]. *EMC - Oto-rhino-laryngologie*, vol.15, n° 3, p. 20-753, 2000
- [5] H. A. Guerrero López, « Caractérisation de la voix de l'enfant sourd appareillé et implanté cochléaire: approches acoustique et perceptuelle et proposition de modélisation ». Doctorat : Sciences du langage. Montpellier : Université Paul Valéry, 2010, p.359
- [6] T. Winslow, *Pharynx Anatomy* [en ligne]. National Cancer Institute. 2018. Consulté le: 23/11/ 2021. Disponible sur: <https://visualsonline.cancer.gov/details.cfm?imageid=9254>
- [7] C. Gabriel, « Chapitre 9: production de la parole et voix humaine corde vocale - Page d'accueil du site de Claude Gabriel » [en ligne]. Consulté le : 13/10/ 2021. Disponible sur : <https://www.yumpu.com/fr/document/read/16606261/corde-vocale-page-daccueil-du-site-de-claude-gabriel>
- [8] D. Dehesdin et O. Choussy, « Anatomie du pharynx », *EMC - Oto-rhino-laryngologie*, [20-491-A-10], 1998
- [9] VocalC, *SCHOOL VOCALIZE*. [En ligne], consulté le nov. 18, 2021. Disponible sur: <https://www.vocalize.fr/wp-content/uploads/2014/11/vocalchords.gif>
- [10] S. Périé, G. Monceaux, B. Angelard, et J. L. St Guily, « Pathologie neurologique du pharynx », *EM-Consulte*, 1998. consulté le : 13/10/2021. Disponible sur : <https://www.em-consulte.com/article/1295/pathologie-neurologique-du-pharynx>
- [11] A. Hadj Salah, « *Linguistique arabe et linguistique générale : essai de méthodologie et d'épistémologie du 'Ilm al-'arabiyya* », Doctorat : Linguistique, Université de Paris-Sorbonne, Paris - France, 1979
- [12] C. Richard, « *Etude de l'encodage des sons de parole par le tronc cérébral dans le bruit* », Doctorat : Sciences Cognitives Linguistique, Université Lumière Lyon 2 , 2010. Disponible sur : http://theses.univ-lyon2.fr/documents/lyon2/2010/richard_c#p=0&a=top
- [13] M. Guerti, K. Ferrat, « Synthèse De La Parole En Arabe Standard. Cas Des Phénomènes Spécifiques à La Langue », *Al-Lisaniyyat. Revue Algérienne de Linguistique*. Alger, Vol. 5, n° 1 , p. 216-237, 2007
- [14] C. Fredouille, « *Traitement Automatique et Troubles de la Voix et de la Parole: champs d'application, contraintes et limites* », Doctorat : Informatique, Université d'Avignon et du

Pays de Vaucluse, 2016. Disponible sur : <https://hal.archives-ouvertes.fr/tel-02102837/document>

[15] F. Emerard, « Les diphones et le traitement de la prosodie dans la synthèse de la parole », *Bulletin de l'Institut de Phonétique de Grenoble St-Martin-d'Hères*, Inist-CNRS. vol. 6, p. 103-147, 1977

[16] G. Vannier, « *Etude des contributions des structures textuelles et syntaxiques pour la prosodie: application à un système de synthèse vocale à partir du texte* », Doctorat : Sciences et techniques, Université de Caen, 1999. Disponible sur : <https://www.theses.fr/1999CAEN2032>

[17] P. A. Barbosa, « *Caractérisation et génération automatique de la structuration rythmique du français* », Doctorat : Signal, Image, Parole, Université de Grenoble INPG, 212p. 1994

[18] G. Droua-Hamdani, « *Prédiction de la durée segmentale des phonèmes de l'Arabe standard* », Mémoire de Magister : Electronique, CRSTDLA, Alger, Algérie, 2004

[19] F. Charpentier, « *Traitement de la parole par analyse-synthèse de Fourier: Application à la synthèse par diphones* », Doctorat : Sciences Appliquées, ENST, Paris. p. 159, 1988.

[20] L.-J. Boë *et al.*, « Nouvelles pistes pour revisiter la production de la parole et son développement: données, modèles, représentation (New tracks to revisit speech production and speech development: data, models and representation)[in French] », in *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 1: JEP*, 2012, p. 593-600. Disponible sur : <https://hal.archives-ouvertes.fr/hal-00724679/>

[21] M. François, « Troubles de la voix et de l'articulation chez l'enfant », *EMC-Oto-rhinolaryngologie*, vol. 1, n° 3, p. 175-186, 2004. DOI: 10.1016/j.emcorl.2004.03.004

[22] T. Hacki et S. Heitmüller, « Development of the child's voice: premutation, mutation », *International journal of pediatric otorhinolaryngology*, vol. 49, p. S141-S144, 1999. DOI: 10.1016/S0165-5876(99)00150-0

[23] J. Hohlfeld *et al.*, « Prise en charge des fentes labio-maxillo-palatines au CHUV. », *Revue médicale suisse*, n° 191, p. 402, 2009. Disponible sur : https://cleft-palate.com/wp-content/uploads/2017/09/RMS_idPAS_D_ISBN_pu2009-07s_sa04_art04.pdf

[24] D. G. Chevaillier, « Quelques spécificités de la voix de l'enfant » [En ligne], in *5ème Rencontre des Maîtrises et filières Voix*. juill. 2007, p. 43. Consulté le: 13/10/ 2021. Disponible sur: <https://artchoral.org/wp-content/uploads/2017/06/IFAC-2017-Voix-de-l-enfant-Dr-G%C3%A9rard-Chevaillier.pdf>

[25] I. ABRY-DURAND, « LA MUE, une transformation naturelle de la voix - ses mécanismes et ses conséquences », *CPEM Savoie*, 2012. Consulté le: 13/10/2021. Disponible sur: http://www.ac-grenoble.fr/savoie/pedagogie/docs_pedas/mue/index.php?num=995

[26] B. Dalens, « Particularités physiologiques de l'enfant » [En ligne]. Université Laval, Québec. Cours, 2009. Consulté le: oct. 18, 2021. Disponible sur : <https://docplayer.fr/30611338-Particularites-physiologiques-de-l-enfant-bernard-dalens-md-phd-chul-du-chuq.html>

[27] G. Barbier, L.-J. Boë, A. Vilain, et G. Captier, « Croissance du conduit vocal de la naissance à l'âge adulte: données morphologiques longitudinales et modélisation articulatoire », in *9ème Rencontres des Jeunes Chercheurs en Parole 2011 (RJCP 2011)*, 2011, p. 5-8. Disponible sur : <https://hal.archives-ouvertes.fr/hal-00682208>

[28] SICARD, Etienne et MENIN-SICARD, Anne. « *Temps phonatoire, rapport s/z et a/z : repères en lien avec la prise en charge orthophonique* » Doctorat. INSA. Toulouse. 2022. Disponible sur : <https://hal.archives-ouvertes.fr/hal-03662213/>

[29] C. Fressard, F. Pellegrino, et G. Puech, « Notes - 1.3.1 - L'appareil Phonatoire Humain » [en ligne], Le blog de outilsrecherche. consulté le 23/11/ 2021. Disponible sur : http://outilsrecherche.over-blog.com/pages/Notes_131_Lappareil_Phonatoire_Humain-3083095.html

[30] K. Ferrat, « *Extraction des paramètres acoustiques en vue de la synthèse par règles de l'arabe standard* », Magister : Electronique acoustique. Alger, Ecole Normale Supérieure, 2001, p.124

[31] M. Djoudi, « *Contribution à l'étude et à la reconnaissance automatique de la parole en Arabe standard* », Doctorat : Informatique. Université Henri Poincaré-Nancy1, Nancy, France, 1991

[32] C. CNET-ENST, La parole et son traitement automatique. Masson, Sept, 1989

[33] C. Meunier, Phonétique acoustique. Solal, 2007. Disponible sur : <https://hal.archives-ouvertes.fr/hal-00250272/>

[34] Université de Leon. Espagne. « articulation, phonétique FLE, phonétique française, cours de prononciation » [En ligne]. Consulté le 18/10/2020. Disponible sur: <http://flenet.unileon.es/phon/phoncours.html>

[35] K. Baazi, « *Identification des consonnes emphatiques de l'Arabe standard - Etude Acoustico-articulatoire* », Mémoire de magistère : Electronique acoustique, Ecole normale supérieure; Alger, 2006

[36] S. Baloul, P.-Y. Le Meur. «Présentation du système de synthèse de la parole à partir du texte arabe voyellé d'Elan Speech». in : *International Conference Arabic and Information Technology*, 2002, Alger, Algeria. consulté le: 05/04/2017. Disponible sur : <https://hal.archives-ouvertes.fr/hal-01434589/>

[37] K. Taleb Ibrahim, Principes dans la linguistique (مبادئ في اللسانيات). Algérie: Casbah Edition, 2006

[38] A. Zargua, Asrar al-huruf (أسرار الحروف). Syrie: Dar al-hasad, 1993

[39] N. Boukadida, « *Connaissances phonologiques et morphologiques dérivationnelles et apprentissage de la lecture en arabe (Etude longitudinale)* », Doctorat : Psychologie, Université Rennes 2; Université de Tunis, 2008, 277p. Disponible sur : <https://tel.archives-ouvertes.fr/tel-00300544/document>

-
- [40] A. Hadj-Salah, « La notion de syllabe et la théorie cinetico impulsienne des phonéticiens arabes », *Al-Lisaniyyat. Revue Algérienne de Linguistique*, Alger, vol. 1, n° 1, p. 63-83, 1971
- [41] G. Droua Hamdani et M. Abbas, « Haraka Et Sukun : études Cinétique Et Acoustique », *AL-Lisaniyyat*, vol. 18, n° 17, p. 71-82, 2012
- [42] W. M. Sweeney, S. T. Lanier, C. A. Purnell, et A. K. Gosain, « Genetics of cleft palate and velopharyngeal insufficiency », *Journal of pediatric genetics*, vol. 4, n° 01, p. 009-016, 2015. DOI: 10.1055/s-0035-1554978
- [43] YSUNZA, Pablo Antonio, REPETTO, Gabriela M., PAMPLONA, Maria Carmen, et al. « Current controversies in diagnosis and management of cleft palate and velopharyngeal insufficiency », *BioMed research international*, vol. 2015, Article ID 196240, 2015. DOI: 10.1155/2015/196240
- [44] ANDREU, Lalao. « *Elaboration d'un livret d'information sur les fentes maxillaires à l'attention des parents* », Doctorat, Université Toulouse III-Paul Sabatier, 2013
- [45] M. Bechet, « *Perturbation de la production des occlusives chez des locuteurs présentant une division palatine ou labio-palatine* », Doctorat : Sciences du Langage, Université de Strasbourg, France, 2011, p. 615
- [46] BOCKLET, Tobias, RIEDHAMMER, Korbinian, NÖTH, Elmar, et al. « Automatic intelligibility assessment of speakers after laryngeal cancer by means of acoustic modeling ». *Journal of Voice*, 2012, vol. 26, no 3, p. 390-397. DOI: 10.1016/j.jvoice.2011.04.010
- [47] RUBIN, Zachary, KURNIAWAN, Sri, et TOLLEFSON, Travis. « Results from using automatic speech recognition in cleft speech therapy with children ». In : *International Conference on Computers for Handicapped Persons. Springer*, Cham, 2014. p. 283-286. DOI: 10.1007/978-3-319-08599-9_43
- [48] VUCOVICH, Megan, HALLAC, Rami R., KANE, Alex A., et al. « Automated cleft speech evaluation using speech recognition ». *Journal of Cranio-Maxillofacial Surgery*, 2017, vol. 45, no 8, p. 1268-1271
- [49] SUDRO, Protima Nomo, DAS, Rohan Kumar, SINHA, Rohit, et al. « Significance of Data Augmentation for Improving Cleft Lip and Palate Speech Recognition ». In : *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE*, 2021. p. 484-490
- [50] M. Devisse, A. Hacquard, J. Lelièvre, C. Lansonneur, Y. Gauvin, et S. Boisramé-Gastrin. « Forme fruste de fente labio-palatine: présentation d'un cas clinique », *Médecine Buccale Chirurgie Buccale*, vol. 18, n° 4, p. 367-370.2012. DOI: 10.1051/mbcb/2009037
- [51] A. Clift, S. Dyke, et M. Valverde, vous et votre bébé Nourrissons présentant une fente labio-palatine Ressource destinée aux professionnels de la santé et aux familles. Canada: Eastern Health Lighthouse, 2009. Disponible sur : <https://docplayer.fr/14461548-Vous-et-votre-bebe-nourrissons-presentant-une-fente-labio-palatine-ressource-destinee-aux-professionnels-de-la-sante-et-aux-familles.html>

-
- [52] Société Française d'Orthopédie Dento-Faciale, « Les fentes labio-palatines » [en ligne]. Consulté le 18/10/2021. Disponible sur : http://www.sfodf.org/avada_portfolio/les-fentes-labio-palatines
- [53] C. Chevrier, « *Intéactions gene-environnement dans l'étiologie des fentes orales* », Doctorat : Epidémiologie, Université Paris XI, Faculté de Médecine Paris-Sud, 2005. Disponible sur : <https://tel.archives-ouvertes.fr/tel-00669882/document>
- [54] P. Bonfils et J.-M. Chevallier, Livre : Anatomie - Tome 3. ORL [en ligne]. Lavoisier Médecine Sciences. Paris. 2003. Consulté le: 18/10/2021. Disponible sur : https://www.lalibrairie.com/livres/anatomie-ork-vol-3_0-1230536_9782257204585.html
- [55] M. Ferguson, « Development of the Face and Palate », *The Cleft Palate-Craniofacial Journal*, Vol. 32, n°6, 1995.
DOI: 10.1597/1545-1569_1995_032_0522_dotfap_2.3.co_2
- [56] E. Noirrit-Esclassan, P. Pomar, R. Esclassan, B. Terrie, P. Galinier, et V. Woisard, « Plaques palatines chez le nourrisson porteur de fente labiomaxillaire », *EMC-Stomatologie*, 2005, vol. 1, n° 1, p. 60-79.
DOI: 10.1016/j.emcsto.2004.12.001
- [57] P. Pellerin et al., « Prise en charge des fentes labio-maxillo-palatines au sein du service de chirurgie plastique du centre hospitalier universitaire de Lille », *Annales de chirurgie plastique esthétique*, 2002, vol. 47, n° 2, p. 106-115.
DOI: 10.1016/S0294-1260(02)00094-8
- [58] Centre de Référence maladies rares des Fentes et malformations faciales (MAFACE). Protocole National de Diagnostic et de Soins (PNDS), « Fentes labiales et/ou palatines » [en ligne]. Consulté le: 18/10/2021. Disponible sur : https://www.has-sante.fr/upload/docs/application/pdf/2021-11/pnds_fentes_labiales_etou_palatines-texte-novembre_2021.pdf
- [59] Centre de Compétence des Fentes Labiopalatines - Médipôle Garonne Toulouse, « Le palais : Traitement des Fentes Palatines » [en ligne]. Consulté le: 10/10/2021. Disponible sur: <https://www.fente-palatine.com/le-palais/>
- [60] M. Zeraatkar, S. Ajami, N. Nadjmi, S. A. Faghihi, et A. Golkari, « A qualitative study of children's quality of life in the context of living with cleft lip and palate », *Pediatric health, medicine and therapeutics*, vol. 10, p. 13, 2019. DOI: 10.2147/PHMT.S173070
- [61] H. Baylon, M. Roger, M. Bigorre, et P. Montoya, « Spécificités de la rééducation orthophonique dans les cas de fente palatine », *Rééducation orthophonique*, vol. 41, n° 216, p. 111-119, 2003
- [62] M. Segura-Hernández *et al.*, « Acoustic analysis of voice in children with cleft lip and palate following vocal rehabilitation. Preliminary report », *International journal of pediatric otorhinolaryngology*, vol. 126, n° Novembre, 2019. DOI: 10.1016/j.ijporl.2019.109618
- [63] M. Brockmann, M. J. Drinnan, C. Storck, et P. N. Carding, « Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and

fundamental frequency effects in a typical clinical task », *Journal of voice*, vol. 25, n° 1, p. 44-53, 2011. DOI: 10.1016/j.jvoice.2009.07.002

[64] J. Kreiman et B. R. Gerratt, « Perception of aperiodicity in pathological voice », *The Journal of the Acoustical Society of America*, vol. 117, n° 4, p. 2201-2211, 2005.
DOI: 10.1121/1.1858351

[65] Elemetrics Kay, Multi-Dimensional Voice Program. Model. Kay Elemetrics Corporation, Lincoln Park NJ, 2008

[66] N. Attuluri, M. Pushpavathi, P. Pandey, et S. Mahapatra, « Voice Perturbations in Repaired Cleft Lip and Palate », *Global Journal of Otolaryngology*, vol. 8, n° 1, Art. n° 1, 2017.

[67] E. Yumoto, W. J. Gould, et T. Baer, « Harmonics-to-noise ratio as an index of the degree of hoarseness », *The journal of the Acoustical Society of America*, vol. 71, n° 6, p. 1544-1550, 1982. DOI: 10.1121/1.387808

[68] FAUTH, Camille, VAXELAIRE, Béatrice, RODIER, Jean François, VOLKMAR, Pierre-Philippe et SOCK, Rudolph. « Etude du Voice Onset Time (VOT) dans des séquences VCV produites par des patients francophones souffrants de paralysies récurrentielles ». In : *SHS Web of Conferences*. EDP Sciences. Vol. 8, pp. 1251-1265. 2014. ISBN 2261-2424.
DOI: 10.1051/shsconf/20140801172

[69] L. Crevier-Buchman, S. Brihaye-Arpin, A. Sauvignet, C. Tessier, M. Monfrais-Pfauwadel, et D. Brasnu, « Dysphonies non organiques (dysfonctionnelles) », *Encyclopédie Médico-Chirurgicale Oto-rhino-laryngologie, 20-752-A*, vol. 15, p. 1-12, 2006.

[70] H. Sayoud, « *Reconnaissance automatique du locuteur: approche connexionniste* » Doctorat : Electronique, USTHB, Alger, Algerie, 2003

[71] G. Li, Q. Hou, C. Zhang, Z. Jiang, et S. Gong, « Acoustic parameters for the evaluation of voice quality in patients with voice disorders », *Annals of Palliative Medicine*, vol. 10, n° 1, p. 130-136, janv. 2021. DOI: 10.21037/apm-20-2102

[72] G. Chevaillier, « La voix de l'enfant et ses troubles fonctionnels », *Enfances Psy*, n° 1, p. 15-28, 2013.

[73] M. Chouteau et C. PLANTARD, « *Impact d'une fente faciale sur le langage oral d'enfants de 3 ans à 5 ans 1/2 porteurs de fentes labio-palatines et palatines isolées* ». Mémoire Certificat de Capacité d'Orthophonie, Université Lille 2, 2013. Disponible sur : <https://pepite-depot.univ-lille2.fr/nuxeo/site/esupversions/faab485c-6e32-4d79-99ec-62e5efd7e4c8>

[74] G. Du Crest, « Etude de l'évolution, en fonction de l'âge, du temps maximum phonatoire et du rapport S/Z chez une population d'enfants âgés de 3 à 10 ans », Mémoire en vue de l'obtention du CCO, Université de Lille, 2014

[75] C. Abry *et al.*, « Propositions pour la segmentation et l'étiquetage des sons du français », *14e Journées d'études sur la parole*, , 1985. Paris. ENST p. 156-163

-
- [76] P. Boersma et D. Weenink, « Praat: Doing phonetics by computer (Version 5.3.82)[Computer software] », Amsterdam: Institute of Phonetic Sciences, 2012
- [77] G. Weismer, « Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production », *Journal of Phonetics*, vol 7, no 2, p. 197-204, 1979.
DOI: 10.1016/S0095-4470(19)31041-1
- [78] M. Djoudi, « Étude phonétique de l'Arabe standard », Rapport technique, CRIN - INRIA Lorraine, France, 1989, 22p.
- [79] J. Brunner, S. Fuchs, et P. Perrier, « The influence of the palate shape on articulatory token-to-token variability », *ZAS Papers in Linguistics*, vol. 42, p. 43-67, 2005.
DOI: 10.21248/zaspil.42.2005.273
- [80] D. H. Klatt, « Voice onset time, frication, and aspiration in word-initial consonant clusters », *Journal of speech and hearing research*, vol. 18, n° 4, Art. n° 4, 1975.
DOI: 10.1044/jshr.1804.686
- [81] M. Tenenhaus, « Statistique: méthodes pour décrire, expliquer et prévoir », vol. 680. Dunod Paris, France, 2007
- [82] K.FERRAT, «Analyse acoustique et évaluation de la rééducation de la maladie de Parkinson dans le milieu hospitalier algérien (دراسة صوتية فيزيائية لكلام الشخص المصاب بمرض (الباركينسون في الوسط الاستشفائي الجزائري)», *Revue Al-Lissaniyat*, CRSTDLA, Algérie, pp. 14-15, 2009.
- [83] K. Ferrat et M. Guerti, «A study of sounds produced by Algerian esophageal speakers », *African health sciences*, vol. 12, n° 4, Art. n° 4, 2012. DOI: 10.4314/ahs.v12i4.9
- [84] H. Delacour, A. Servonnet, A. Perrot, J. F. Vigezzi, et J. M. Ramirez, « La courbe ROC (receiver operating characteristic): principes et principales applications en biologie clinique », in *Annales de biologie clinique*, 2005, vol. 63, n° 2, p. 145-154.
- [85] « Courbes ROC », XLSTAT, Your data analysis solution [en ligne]. Consulté le 27/09/2019. Disponible sur: <https://www.xlstat.com/fr/solutions/fonctionnalites/courbes-roc>
- [86] T. Perneger et A. Perrier, « Analyse d'un test diagnostique: courbe ROC, ou « Receiver Operating Characteristic » », *Revue des maladies respiratoires*, vol. 21, n° 2, p. 398-401, 2004. DOI:10.1016/S0761-8425(04)71302-9
- [87] M. Eshghi, J. S. Preisser, M. Bijankhan, et D. J. Zajac, « Acoustic-temporal aspects of stop-plosives in the speech of Persian-speaking children with cleft lip and palate », *International journal of speech-language pathology*, vol. 19, n° 6, p. 578-586, 2017.
DOI: 10.1080/17549507.2016.1221458
- [88] S. Morén, P. Å. Lindestad, M. Holmström, et M. Mani, « Voice quality in adults treated for unilateral cleft lip and palate: long-term follow-up after one-or two-stage palate repair », *The cleft palate-craniofacial Journal*, vol. 55, n° 8, p. 1103-1114, 2018.
DOI: 10.1177/1055665618764521
- [89] K. H. Davis, R. Biddulph, et S. Balashek, « Automatic recognition of spoken digits », *The Journal of the Acoustical Society of America*, vol. 24, n° 6, p. 637-642, 1952.

- [90] G. Calas, « Etudes des principaux algorithmes de data mining », *Spécialisation Sciences Cognitives et Informatique Avancée*, France: [SCIA] EPITA, 2009.
- [91] D. Coomans et D. L. Massart, « Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules », *Analytica Chimica Acta*, vol. 136, p. 15-27, 1982.
DOI: 10.1016/S0003-2670(01)95359-0
- [92] A. Bosch, A. Zisserman, et X. Munoz, « Image classification using random forests and ferns », in *2007 IEEE 11th international conference on computer vision*, 2007, p. 1-8.
DOI: 10.1109/ICCV10464.2007
- [93] A. Criminisi, J. Shotton, et E. Konukoglu, « Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning », *Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114*, vol. 5, n° 6, p. 12, 2011.
DOI:10.1561/06000000035
- [94] L. Lazli et M. T. Laskri, « Nouvelle méthode de fusion de données pour l'apprentissage des systèmes hybrides MMC/RNA », *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, vol. 3, p. 125-170, 2005.
- [95] S. Young, « HMMs and related speech recognition technologies », in *Springer Handbook of Speech Processing*, Springer, 2008, p. 539-558.
DOI: 10.1007/978-3-540-49127-9_27
- [96] K. FERRAT, K. BAAZI et M. GUERTI, « Reconnaissance des sons spécifiques de l'Arabe Standard par Réseaux de Neurones », *Revue Maghrébine des Langues*, vol. 7, no 1, p. 125-142, 2011.
- [97] P. Besse, « Apprentissage statistique & data mining », Toulouse: INSA, 2009.
- [98] A. Ganapathiraju, J. E. Hamaker, et J. Picone, « Applications of support vector machines to speech recognition », *IEEE transactions on signal processing*, vol. 52, n° 8, p. 2348-2355, 2004. DOI: 10.1109/TSP.2004.831018
- [99] R. Rakotomalala, « Arbres de décision », *Revue Modulad*, vol. 33, p. 163-187, 2005.
- [100] HAN, Jiawei, KAMBER, Micheline, et PEI, Jian. « Data mining: concepts and techniques », Waltham, MA. Morgan Kaufman Publishers, vol. 10, 2012
- [101] Y.-Y. Song et L. U. Ying, « Decision tree methods: applications for classification and prediction », *Shanghai archives of psychiatry*, vol. 27, n° 2, Art. n° 2, p. 130-135, 2015.
Doi: 10.11919/j.issn.1002-0829.215044
- [102] J. A. Lopez Gonzalez, « Exploration des arbres de décision et des support vector machines en vue d'applications dans l'analyse de texte », Doctorat : Mathématiques et Informatique Appliquées, Université du Québec à Trois-Rivières, 2016, 147p. Disponible sur: <https://depot-e.uqtr.ca/id/eprint/7925/1/031383759.pdf>

-
- [103] T. H. Dang, « *Mesures de discrimination et leurs applications en apprentissage inductif* », Doctorat : Informatique, Université Pierre and Marie Curie-Paris VI, Paris, France, 2007, 189p. Disponible sur : <https://tel.archives-ouvertes.fr/tel-00184691/document>
- [104] M. Q. Vu, L. Besacier, E. Castelli, et N. Y. Pham, « Extraction automatique de Questions dans les corpus de réunions et de dialogues », in *MajecSTIC 2005: Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC*, 2005, Rennes. p. 393-397. Disponible sur : <https://hal.inria.fr/inria-00000739/document>
- [105] V. Guijarro, « Les Arbres de Décisions L'algorithme ID3 », Cours, université de Montpellier, 2006.
- [106] L. Breiman, « Random forests », *Machine learning*, vol. 45, n° 1, p. 5-32, 2001.
- [107] B. Devéze et M. Fouquin, « DATAMINING C4.5 - DBSCAN », PROMOTION, SCIA Ecole pour l'informatique et techniques avancées, 2005.
- [108] W.-Y. Loh, « Fifty years of classification and regression trees », *International Statistical Review*, vol. 82, n° 3, p. 329-348, 2014. Doi.org/10.1111/insr.12016
- [109] X. Wu *et al.*, « Top 10 algorithms in data mining » [en ligne], *Knowledge and information systems*, vol. 14, n° 1, p. 1-37, 2008. Doi :10.1007/s10115-007-0114-2
- [110] M. Q. Vu, E. Castelli, A. Boucher, et L. Besacier, « Classification de parole en Question et Non-Question par arbre de décision », *SFC 05, 12èmes Rencontres de la Société Francophone de Classification*, 2005.
- [111] P. Vincent, « *Modèles à noyaux a structure locale* » [en ligne]. Doctorat : Informatique. Université de Montréal, 2003. [consulté le : 15/10/2018]. Disponible sur : https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/14543/Vincent_Pascal_2003_these.pdf?sequence=1
- [112] K. Baazi et M. Guerti, « Automatic speech classification in school children with cleft (and lip) palate » [en ligne], *International Journal of Medical Engineering and Informatics*, 2021. Doi: 10.1504/IJMEI.2021.10041983. Disponible sur : <https://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijmei>