République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



*Université de Jijel*

*Faculté des Sciences et de la Technologie*

*Département d'Electronique*

*Projet de fin d'Etudes pour l'obtention du Diplôme de*

*Master en Electronique*

*Option : Electronique et Systèmes de Communication*

*Thème :*

# Analyse et Evaluation du Signal de la Parole des Voix Pathologiques

***Réalisé par :***

***Mr. SADOU Zineddine***

***Mr. CHALOUCHE Housseyn***

***Proposé par :***

***Mr. BOUBAKIR Chabane***

***Promotion : Juin 2015.***

# Remerciements

***La Louange est à Allâh,*** *le Seigneur des mondes. Et que la prière et le salut soient sur celui qu'Allâh a envoyé en miséricorde pour l'univers, ainsi que sur sa famille, ses compagnons et ses frères jusqu'au Jour de la Rétribution.*

*Nous voudrions remercier tout d'abord Allâh, le tout puissant qui nous a donné la force, la volonté et le courage pour accomplir ce modeste travail.*

*Nous tenons à formuler notre gratitude et notre profonde reconnaissance à l'égard de notre promoteur Mr :* ***Chaabane Boubakir*** *qui a supervisé ce travail de recherche. Son soutien, sa disponibilité, sa patience, sa compréhension, ainsi que ses conseils judicieux tant lors de nos recherches que lors de l'écriture de ce mémoire. Ses connaissances et ses jugements nous ont permis d'acquérir des compétences essentielles en recherche.*

*Nous adressons également nos remerciements, à tous* ***nos enseignants****, qui nous ont donné les bases de la science, sans oublier d'exprimer nos remerciements au* ***Chef de Département*** *d'Électronique.*

*Nos remerciements aux* ***membres du jury*** *qui nous ont fait l'honneur d'accepter de lire et de juger ce mémoire.*

*Nous remercions l'ensemble des* ***collègues et amis*** *qui nous ont aidés et supporté durant ces dernières années, sans oublier toute personne ayant participée de près ou de loin à l'élaboration de ce modeste travail.*

# Dédicaces

*Je dédie ce travail à mes chers parents pour leurs encouragements et leur soutien moral et matériel durant toutes mes années d'études, que DIEU le tout puissant me les gardes.*

***Ma chère mère et Mon cher père***

*A mes chers frères,*

*A toute ma famille,*

*A tous mes collègues,*

*A mes chèrs amis,*

*Housseyn*

# Dédicaces

*Je dédie ce travail à mes chers parents pour leurs encouragements et leur soutien moral et matériel durant toutes mes années d'études, que DIEU le tout puissant me les gardes.*

***Ma chère mère et Mon cher père***

*A mes chers frères,*

*A toute ma famille,*

*A tous mes collègues,*

*A mes chèrs amis,*

*Zinedine*

Table des matières

# *Liste des figures*

# Liste des abréviations

| | |
|---|---|
| HNR | Harmonic to Noise Ratio (Rapport harmonique sur bruit) |
| GRBASI | Grade, Roughness, Breathiness, Asthenia, Strain, Instability |
| F0 | Fréquence fondamentale |
| GFA | Glottal Frequency Analyser |
| LPC | Linear predictive coding |
| CAPE-V | The Consensus Auditory-Perceptual Evaluation of Voice |
| TMP | Temps maximum de phonation |
| PFR | Phonatory Frequency Range |
| MAJ | Mean Absolute Jitter |
| RAP | Relative Average Perturbation |
| PPQ | Pitch Perturbation Quotient |
| ORL | Otorhinolaryngologie |
| MAS | Mean Absolute Shimmer |
| APQ | Amplitude Perturbation Quotient |
| DFT | Discrete Fourier Transform |

# Introduction générale :

La parole est le mode de communication privilégié entre les humains. C'est la faculté d'exprimer et de communiquer la pensée par l'intermédiaire d'un système de sons articulés.

Pour des raisons diverses, la mécanique vocale peut être sujette à un certain nombre de dysfonctionnements et de pathologies qui présentent une altération d'une ou de plusieurs paramètres acoustiques de la voix. Les conséquences de ces pathologies peuvent aller d'une simple voix enrouée à l'absence complète de la voix (aphonie).

Nous nous intéresserons beaucoup plus aux pathologies du larynx, à la parole générée par des locuteurs dysphoniques dans ce cas ou la parole synthétique qui modélise les troubles vocaux des pathologies du larynx.

Généralement, on distingue deux grandes classes des dysphonies soit d'origines morphologiques dues aux changements morphologiques de l'anatomie du larynx, essentiellement au niveau de la glotte, soit des dysphonies d'origines neurologiques, qui peuvent être dues à un mauvais contrôle de la respiration, d'une atteinte neurologique ou une difficulté psychologique.

La qualité vocale caractérise la performance d'un appareil phonatoire aussi bien des voix normales que des voix dysphoniques. Pour cela, plusieurs mesures, objectives et subjectives, peuvent être établies pour l'évaluation des dysphonies et des troubles de la voix.

L'objectif de notre travail est l'étude et l'évaluation du signal de la parole des voix pathologiques, l'implémentation des mesures objectives et la vérification de l'efficacité de ces méthodes sur des voyelles synthétiques et de la parole continue.

Ce travail est décomposé en trois chapitres et organisé de la façon suivante :

Au cours du premier chapitre, nous présenterons des généralités sur la parole humaine, nous allons parler également en détail sur les caractéristiques de la voix, sa production et ses catégories. Dans le même chapitre nous allons citer les différentes pathologies du larynx qui sont responsables de la plupart des dysphonies de la voix.

Le deuxième chapitre sera consacré d'une part, à l'étude théorique des différents indices dite 'acoustiques' caractérisant la parole et ses formules de calcul, d'autre part nous allons décrire les méthodes utilisées pour la caractérisation des troubles de la voix, soit subjectives comme l'analyse perceptive basée sur l'échelle du GRBASI, soit objectives comme : le Jitter, le Shimmer et le rapport harmoniques sur bruit.

Dans le troisième chapitre, nous présenterons une étude détaillée via la simulation de différents indices acoustiques avec discussion des résultats obtenus et ces interprétations. Les simulations seront faites avec des logiciels spécialisés (Matlab et PRAAT), sur des bases de données des voyelles et des phrases phonétiquement équilibrées synthétiques ou normales.

Enfin, une conclusion générale résumera les résultats et les interprétations de notre travail, suivie par les références bibliographiques utilisées.

# Chapitre I

# Généralités sur la parole pathologique

## I.1 Introduction :

L'homme écrit depuis cinq milles ans mais parle de puis long temps, c'est le seul moyen le plus simple et le plus efficace des modèles de communications, la parole n'est autre qu'un ensemble de voix articulées qui expriment ses pensées, ses désirs et même ses sentiments.

La recherche en traitement de la parole a débuté par un traitement proprement dit du signal vocal qui se base sur la phonétique et la linguistique. Les aspects de la phonétique sont décrits par deux sciences fondamentales à savoir, la phonétique acoustique et la phonétique physiologique qui s'occupe du rôle de nos organes phonatoires dans l'émission des sons; par contre, la première analyse les sons comme étant un ensemble des traits (fréquence, durée, intensité…) perçus par l'oreille humaine.

L'information d'un message parlé réside dans les fluctuations de la pression de l'air engendrées puis émises par l'appareil respiratoire et phonatoire.

Ce chapitre sera consacré aux notions générales sur le signal parole : production, Caractéristiques et pathologies.

## I.2 Production de la parole :

### I.2.1 Appareil phonatoire :

L'appareil phonatoire est l'ensemble des organes qui permettent de produire les sons constituant la voix. Il est essentiellement constitué de trois parties (Figure I.1) :

- Le niveau sous-glottique qui est constitué des poumons et de la trachée artère est souvent appelé « appareil respiratoire ». Il permet de réguler le débit et la pression d'air en entrée du système.
- Le niveau glottique (larynx avec les cordes vocales) qui intervient dans la production de sons voisés, où il joue le rôle d'un excitateur acoustique. Le débit d'air qui traverse la glotte est modulé par la vibration des cordes vocales, ce qui génère une onde acoustique qui se propage dans le conduit vocal et qui est rayonnée par les lèvres.
- Le niveau supra-glottique (pharynx et cavités buccale et nasale) : il joue le rôle d'un articulateur et permet la production des consonnes et des voyelles.

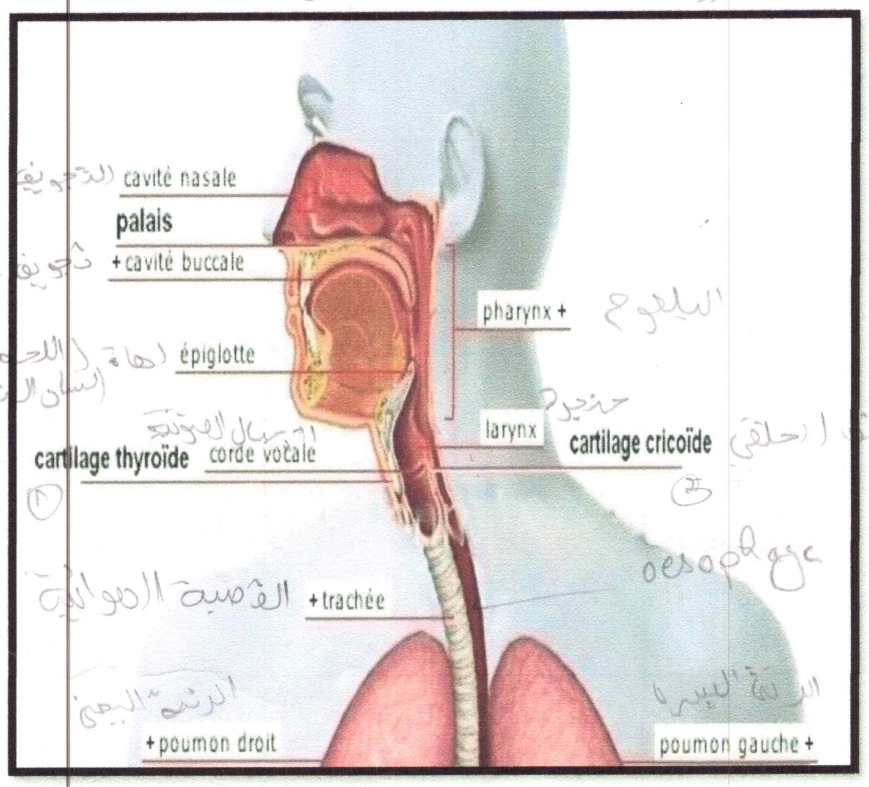


**Figure I.1 :** Anatomie de l'appareil phonatoire [1].

La synthèse vocale humaine se déroule suivant trois étapes successives :

**Inspiration et expiration :** L'inspiration consiste à stocker l'air dans les poumons qui jouent le rôle de réservoir. Durant la phonation, l'air suit le chemin inverse et arrive au niveau du larynx pour assurer la vibration des cordes vocales, c'est l'expiration.

**Vibration des cordes vocales :** Les états de l'appareil phonatoire déterminent les natures des sons produits. Lorsque les cordes vocales sont tendues, le flux d'air les fait vibrer, c'est ce qu'on appelle la phonation. Le flux d'air est découpé en un train d'impulsions quasi périodique qui résonne dans les différentes cavités.
Physiquement, le train d'impulsion quasi périodique subit une modulation en fréquence en passant par les différentes cavités, on obtient donc un son voisé. Lorsque les cordes vocales sont relâchées, l'air passe librement au niveau du larynx sans les faire vibrer. On obtient alors un son non voisé.

**Génération et amplification du son des cordes vocales :** Le conduit vocal permet la production des voyelles et des consonnes par l'intermédiaire d'un articulateur (pharynx, cavités buccale et nasale). Il fait office de résonateur, et permet de sélectionner les bandes de fréquences à renforcer par ajustement des fréquences et largeurs de bande des résonnances acoustiques. Dans le cas des consonnes, des sources aéro-acoustique de bruit sont générées selon la position des constrictions.

## I.2.2 Catégories de sons de la parole :

Selon l'état des cordes vocales, on distingue deux catégories :

**a) Sons voisés :**

Pendant certains sons, la glotte s'ouvre brusquement libérant ainsi la pression accumulée en amont sous forme d'impulsions quasi-périodiques, ces impulsions mettent les cordes vocales en vibration quasi-périodique.

Le spectre d'un son voisé présente des raies correspondantes à l'harmonique du fondamental (structure de pitch) c'est le cas des voyelles, l'enveloppe de ces raies présente des maximums appelés les formants, les trois ou quatre premiers formants sont essentiels pour caractériser le spectre vocal (figure I.2).

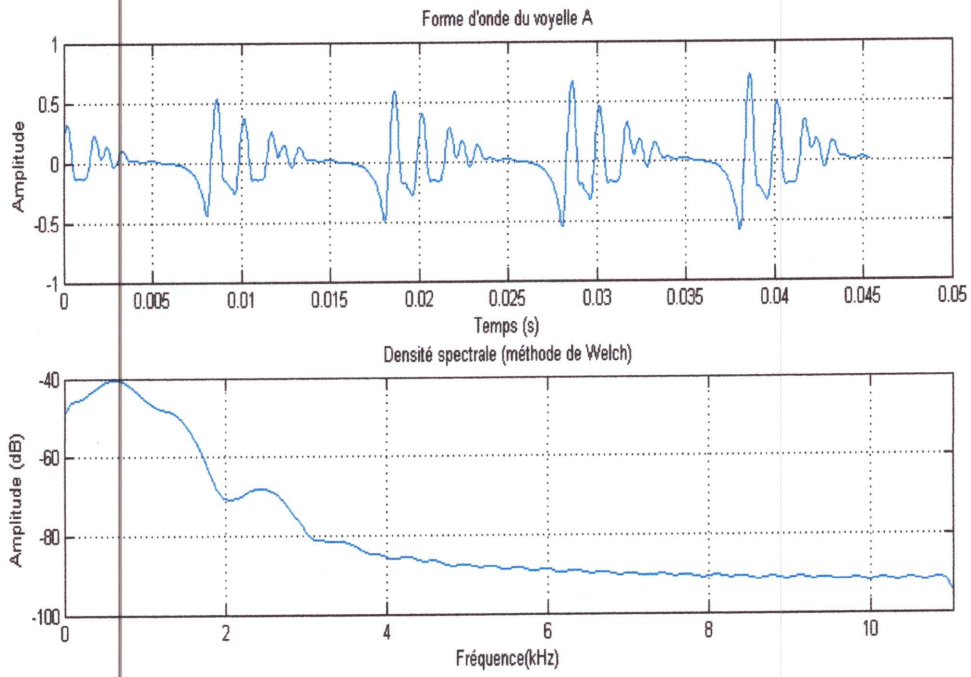


**Figure I.2 :** Son voisé et sa densité spectrale (voyelle « a »).

**b) Sons non voisés :**

Si les cordes vocales sont relâchées; soit une turbulence quasi aléatoire d'air est produite dans le conduit vocal par restriction de sa section, soit le conduit est momentanément fermé complètement pour augmenter la pression et rouvert instantanément produisant une transitoire décroissante. Les sons ainsi produits sont appelés sons non voisés. Le son non voisé ne présente pas une structure périodique, il peut être considéré comme un bruit blanc, ainsi son spectre ne présente pas une structure de pitch (figure I.3).

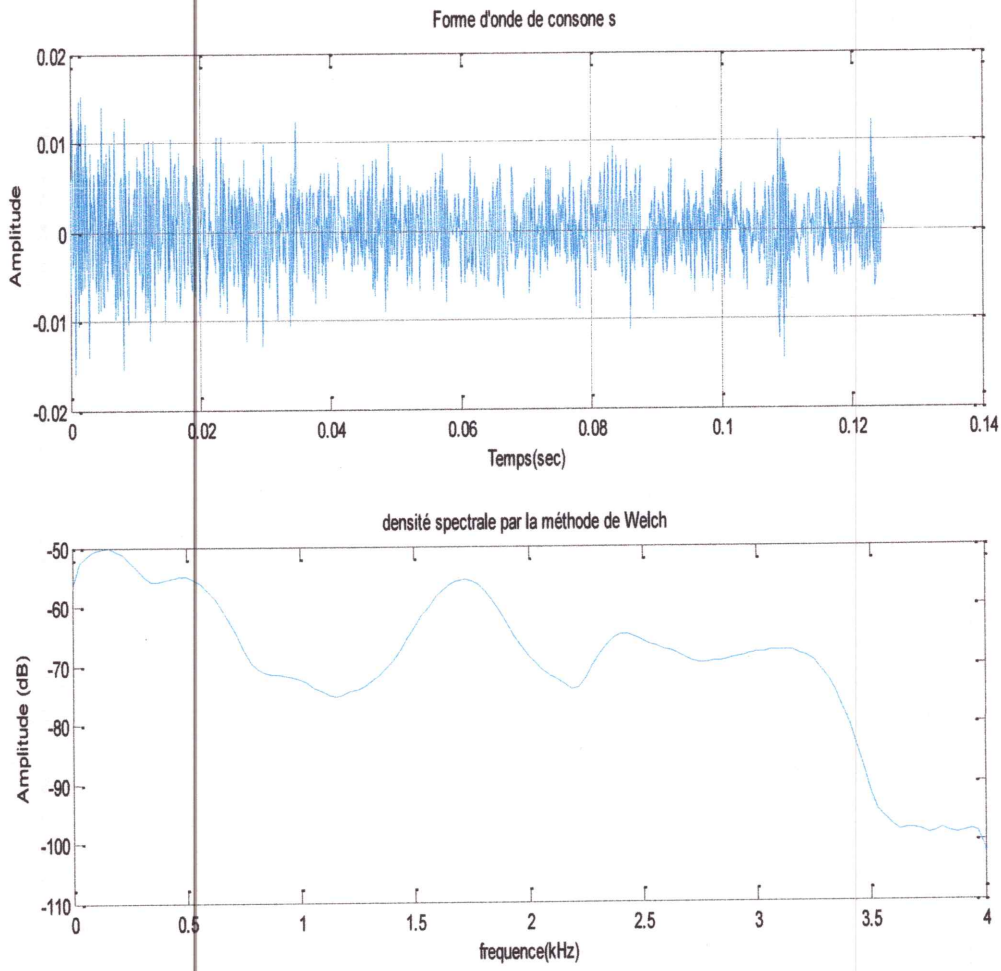


**Figure I.3 :** Son non voisé et son spectre (consonne « s »).

## I.2.3 Modélisation de production de la parole :

### a. Modèle source-filtre :

Dans la plupart du temps, quel que soit le traitement que l'on désire réaliser sur le signal de parole, on est amené à effectuer comme premier traitement une modélisation.

Pour cela, différents modèles ont été proposés en vue d'une description quantitative des facteurs associés au processus de production de la parole. Jusqu'à maintenant aucun des modèles ne peut produire toutes les caractéristiques observées de la parole et en particulier lors d'une évolution rapide du conduit vocal. L'un de ces modèles est le modèle source-filtre qui est représenté par la figure (I.4).

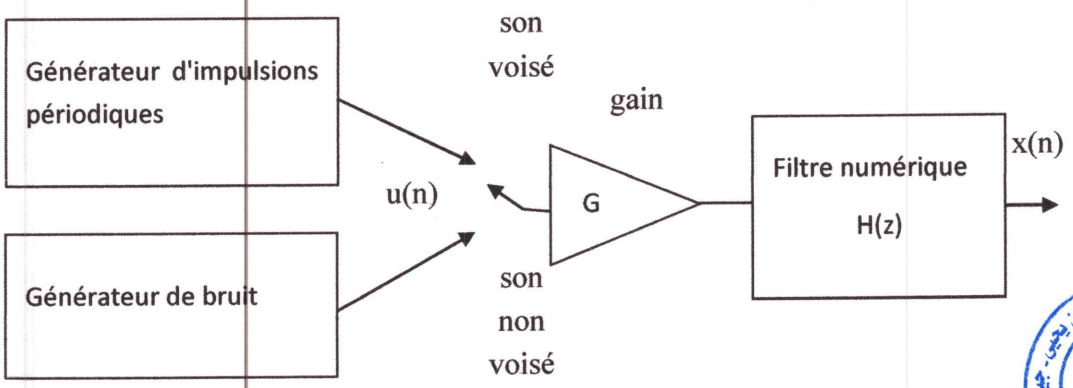




**Figure I.4 :** Modèle source-filtre du système de production de la parole.

**b. Modèle acoustique :**

L'onde acoustique, causée par l'écoulement pulsé de l'air à travers la glotte, obéit aux lois générales de la propagation dans un guide d'onde. La célérité de propagation ne dépend que des propriétés physiques du milieu de propagation telles que l'humidité, la chaleur, …etc.

Dans le cas des sons voisés, l'onde acoustique est formée de plusieurs dizaines d'harmoniques dont les fréquences sont situées aux multiples entiers de la fréquence de vibration des cordes vocales.

Les valeurs des fréquences des harmoniques d'un signal glottique ne changent pas au cours de sa propagation dans le conduit vocal. Tout mauvais fonctionnement de la vibration des cordes vocales, quelque soit son origine, implique une perturbation dans l'onde acoustique qui continue à se propager dans le conduit vocal. Au niveau des lèvres, une partie de ces ondes est rayonnée vers l'extérieur. Ce signal constitue le son de parole.

## I.3 Caractéristiques du signal de parole :

La voix représente le support acoustique de la parole, c'est un ensemble de sons produits par le larynx, lorsque l'air expiré fait vibrer les cordes vocales. Tous les sons simples peuvent être décrits de manière exhaustive par les caractéristiques suivantes :

### I.3.1 Le phonème :

Un phonème est la plus petite unité présente dans la parole et susceptible par sa présence de changer la signification d'un mot, il peut être une voyelle, consonne ou semi-consonne (semi-voyelle) comme il est montré dans la figure (I.5) :

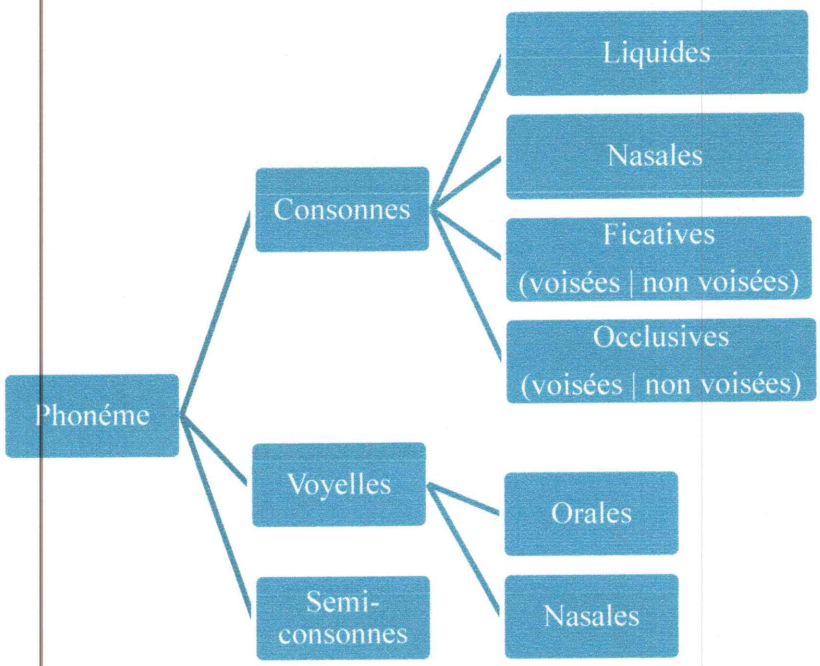


**Figure I.5 :** Classes des phonèmes.

**Les voyelles :** Elles sont des sons de parole produits lorsque le conduit vocal ne présente en aucun endroit un obstacle au libre passage de l'air. L'écoulement de l'air est latéral. La production des voyelles nécessite une source sonore voisée.

La forme du conduit vocal et donc la position de tous les articulateurs déterminent les caractéristiques des voyelles produites. On peut distinguer les voyelles orales, qui se prononcent avec le voile du palais relevé, ce qui ferme le passage nasal, et Les voyelles nasales qui se prononcent avec le voile du palais abaissé, ce qui laisse passer de l'air par la bouche et par le nez.

**Les consonnes :** Les consonnes sont des bruits, qui évoquent des explosions ou des frottements, produits par le souffle heurtant divers organes dans le conduit vocal ou la bouche. Elles ne peuvent pas constituer des syllabes à elles seules et elles sont de différentes classes :

- Fricatives (ou constrictives, spirantes) : un rétrécissement des parois produit un frottement, mais l'air passe, et ces consonnes peuvent durer : [f – v – s – z], raison pour laquelle on les appelle aussi continues.
- Nasales : Pour les consonnes nasales, même si la bouche est fermée, le souffle s'échappe par le nez, et les fosses nasales résonnent (m, n).
- Liquides : c'est une consonne fricative, ce terme est utilisé en fonction de l'impression produite. Dans les consonnes fricatives l'air s'échappe sur les côtés de la langue.

- Occlusives : la fermeture complète et l'ouverture brutale produisent un son de type explosif. On appelle aussi ces consonnes explosives, ou momentanées (pas de durée).

**Les semi-voyelles (semi-consonnes) :** Ce sont des phonèmes intermédiaires entre les voyelles et les 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 (par exemple : j, w....).

| MODE D'ARTICULATION | | | | LIEU D'ARTICULATION | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type de consonne selon le mouvement | Passage de l'air | | Vibration des cordes vocales | Bi-labiale | labio-dentale | Apico-dentale | Apico-alvéolaire | Pré-dorso-alvéolaire | Pré-dorso-pré-palatale | médio-palatale | Dorso-palatale ou vélaire | Post-dorso-uvulaire |
| OCCLUSIVE | ORAL | | NON-VOISEE | p | | t | | | | | k | |
| | | | VOISEE | b | | d | | | | | g | |
| | NASAL | | VOISEE | m | | n | | | | ɲ | (ŋ) | |
| CONSTRICTIVE | ORAL | TYPE DE CONSTRICTIVE | | | | | | | | | | |
| | | FRICATIVE | NON-VOISEE | | f | | | s | ʃ | | | |
| | | | VOISEE | | v | | | z | ʒ | | | |
| | | LATERALE | VOISEE | | | | l | | | | | |
| | | VIBRANTE | VOISEE | | | | | | | | | R |

**Tableau I.1 :** Classification des consonnes de la langue française [2].

### I.3.2 Prosodie :

C'est un terme qui désigne collectivement les trois composantes; pitch, durée et intensité.

#### a) Fréquence fondamentale (Pitch) :

La hauteur de la voix, au cours d'une conversation varie selon les personnes, elle dépend essentiellement de la dimension et de la tension des cordes vocales, ainsi que des dimensions des résonateurs. Elle peut être volontairement modifiée dans certaines limites, par l'intermédiaire des muscles respiratoires, en faisant varier la pression de l'air. L'association de ces éléments détermine la fréquence de vibration des cordes vocales, appelée « fréquence fondamentale » ou

« pitch », elle est variée selon l'âge et le sexe. Alors que la fréquence fondamentale de la voix parlée est :

- De 60 à 250 Hz pour une voix masculine.
- De 150 à 500 Hz pour une voix féminine.
- De 200 à 600 Hz pour une voix d'enfant.

La mesure de la fréquence fondamentale s'effectuant soit à partir d'un signal microphonique (Glottal Fréquency Analyser ou GFA) soit à partir d'un signal electrolaryngographique ou electroglottographique.

**b) Intensité :**

C'est l'amplitude du signal acoustique de la parole exprimée en dB. Généralement sa valeur moyenne est de 60 dB, et elle se divise en 3 cas:

- Chuchotement : les cordes vocales sont ouvertes et laissent passer l'air. La source sonore est une turbulence qui produit un son proche d'un bruit blanc.
- Voisement : les cordes vocales sont proches et vibrent.
- Murmure : les cordes vocales vibrent accolées.

**c) Le rythme :**

Souvent appelé « la durée », il est déterminé à partir de la durée, ou l'intervalle du temps, des silences et des phones d'une phrase.

### I.3.3 Quasi-périodicité :

Le signal de la parole est, par définition, considéré comme un signal non stationnaire et aléatoire, parce que les paramètres et la forme du conduit vocale et de la source changent au cours de la phonation, cependant pour un petit intervalle du temps, de 5ms à 30ms, il présente une structure quasi-périodique, ce qui permet d'appliquer la transformée de Fourier et le fenêtrage pour le traitement du signal de la parole.

### I.3.4 L'énergie :

L'énergie correspond au carré de l'amplitude du signal de pression. Elle est partie intégrante de la parole. On parle ici de l'énergie temporelle, par contre l'énergie spectrale

représente le timbre. Evidemment l'énergie des phonèmes est très différente suivant leur nature, une voyelle ouverte rayonnera plus de puissance qu'une consonne nasale.

$$E = \sum_{n=0}^{N-1} |s^2(n)| \quad \text{(I.1)}$$

Tel que: E est l'énergie de la trame de taille N et s(n) l'amplitude du signal.

### I.3.5 Les formants :

Les formants sont les maximums de l'enveloppe des raies, correspondants aux harmoniques du fondamental. Ils correspondent aux fréquences propres du conduit vocal.

### I.3.6 Le timbre :

C'est le taux de différents harmoniques, dépend de la qualité du son glottique, donc de l'état des cordes vocales, et de la liberté des caisses de résonance.

### I.3.7 La modulation :

C'est la variation progressive d'un ou plusieurs des facteurs précédents.

### I.3.8 L'articulation :

Se caractérise par l'adjonction à la voix modulée de variations brusques transitoires de ces mêmes grandeurs propres à l'homme, à l'encontre des facteurs précédents qui appartiennent à tous les animaux possédant un larynx.

## I.4 Perception de la parole :

Le processus de compréhension par l'auditeur d'un message oral peut être décomposé en deux niveaux; un niveau de transformation de l'information contenue dans le signal acoustique par l'oreille, qui transmet le résultat au cerveau (système auditif périphérique), et le niveau de la reconstitution dans le cerveau du message linguistique (système auditif central). Pour le système auditif périphérique, il est constitué de trois parties (figure I.6) :

**Oreille externe :**

Elle a pour rôle d'une antenne dans les systèmes de communications, elle est constituée de :

- Le pavillon auriculaire : il sert à capter et à concentrer les ondes sonores et d'amplifier les fréquences qui seront également amplifiées par le conduit auditif externe.
- Le conduit auditif externe : c'est juste un canal (tube) qui conduit à l'oreille moyenne.

**Oreille moyenne :**

Le rôle de l'oreille moyenne est double : elle doit à la fois protéger l'oreille interne et transformer les vibrations aériennes arrivant de l'oreille externe en vibrations mécaniques analysables par l'oreille interne.

**Oreille interne :**

Appelée aussi « labyrinthe », elle est composée de plusieurs vestibule, canaux semi-circulaires, cochlée, seule cette dernière joue un rôle dans l'audition, les autres contenant les organes de l'équilibration. Sous l'effet du son, la cochlée (qui contient un liquide) transforme les vibrations en flux électrique (via les cellules ciliées), puis transmettre ce dernier au cerveau (via le nerf auditif).

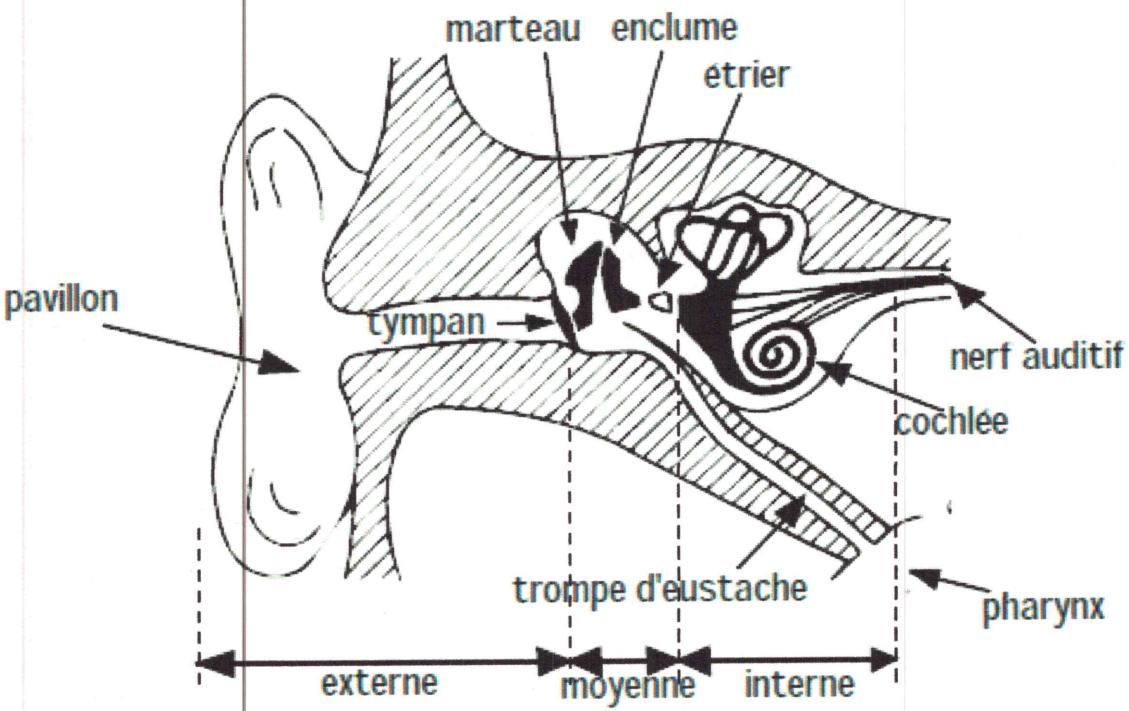


**Figure I.6 :** Système auditif périphérique [3].

## I.5 Synthèse de la parole :

Il est possible en effet de produire automatiquement la parole, soit en concaténant tout simplement des mots ou des parties de phrases préalablement enregistrées, soit par une modélisation physique du système phonatoire (système d'articulations), ou soit en vocalisant un message via la connaissance de l'évolution de quelques paramètres des formants, cette opération de produire la parole artificielle s'appelle : synthèse de la parole.

Il existe plusieurs types de synthétiseurs; synthétiseur a formants [4] [5], synthétiseur LPC et synthétiseur par modèle articulatoire. Les applications des systèmes synthétiseurs sont très nombreuses et couvrirent tous les domaines, on cite parmi eux:

- ✓ Télécommunications et systèmes informatique intelligents par exemple l'application SIRI de IPHONE.
- ✓ Education et l'apprentissage des langues et même les logiciels de traduction vocale.
- ✓ Audio-book, les jouets parlants et les robots de loisir.
- ✓ Etablissement d'une moyenne de communication avec les handicapés, et le meilleur exemple est le cas du célèbre "Stephen Hawking", le savant et génie britannique qui parle seulement à l'aide d'un synthétiseur électronique accroché à sa gorge.
- ✓ Les logiciels informatiques de l'écriture et de commande vocale tel que le logiciel « dragon naturally speaking », qui permet la saisie vocale des textes a plusieurs langues.
- ✓ Le contrôle vocal.

## I.6 Pathologies du larynx :

Il existe plusieurs catégories des pathologies du larynx (figure I .7) pouvant être à l'origine de plusieurs types de troubles vocaux.

Les conséquences de ces pathologies peuvent aller d'une simple voix enrouée à son absence complète (aphonie). Considérée comme un élément banal sans conséquences, la dysphonie est souvent négligée par les patients.

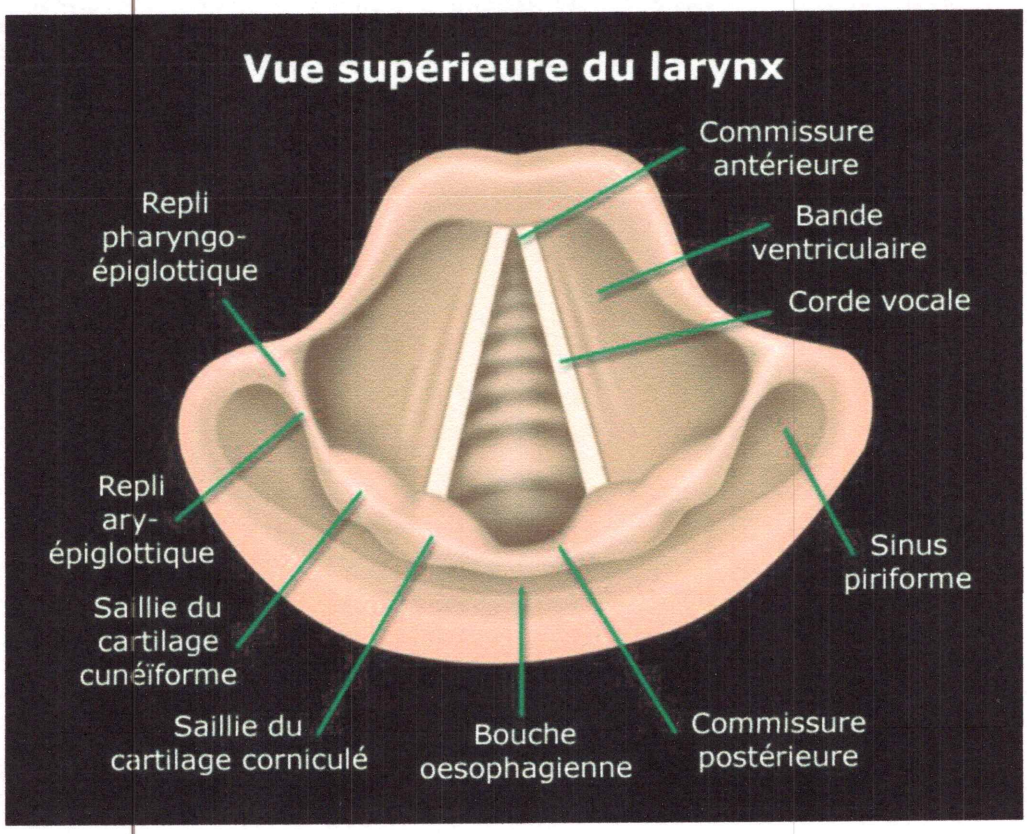


**Figure 1.7 :** Vue laryngoscopique du larynx [6].

Généralement, on distingue deux grandes classes des dysphonies :

### I.6.1 Dysphonies d'origines morphologiques :

Les changements morphologiques de l'anatomie du larynx, essentiellement au niveau de la glotte, sont la cause principale de ces dysfonctionnements vocaux. La modification de la structure glottique peut se traduire par la génération d'un excès de tissus biologiques ou des manques anatomiques provoqués par des gestes chirurgicaux.

En plus, les dysphonies d'origine organique peuvent être la conséquence d'une utilisation inadéquate ou faire suite à un déficit organique congénital. Elles sont causées par des changements anatomiques de la glotte provoqués par l'apparition de nodules, polypes et kystes qui sont des lésions bénignes des cordes vocales dus généralement à un forçage vocal permanent ou brutal.

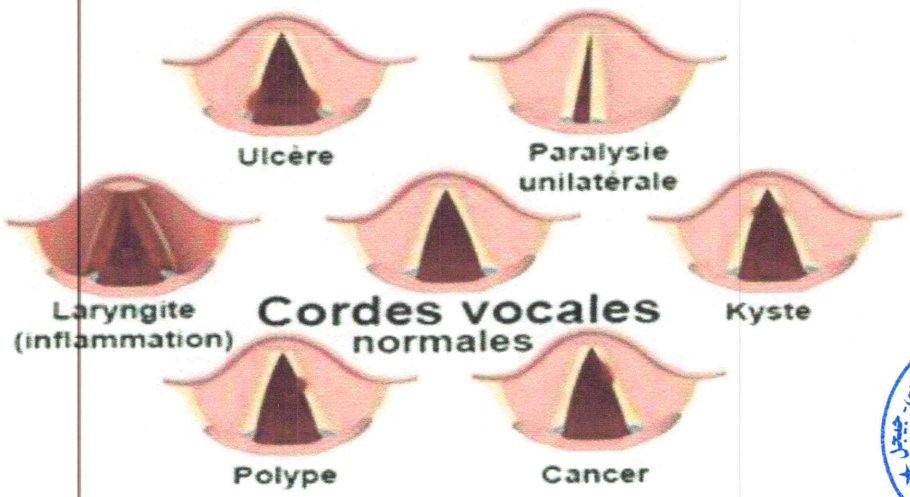




**Figure I.8 :** Différents types de pathologie [7].

Les nodules se présentent comme des petites masses blanches sur le front des cordes vocales au niveau de la jonction du tiers antérieur et du tiers moyen des celles-ci (Figure I.9). Les nodules peuvent être dus à une activité vocale excessive et à une phonation hyperfonctionnelle.

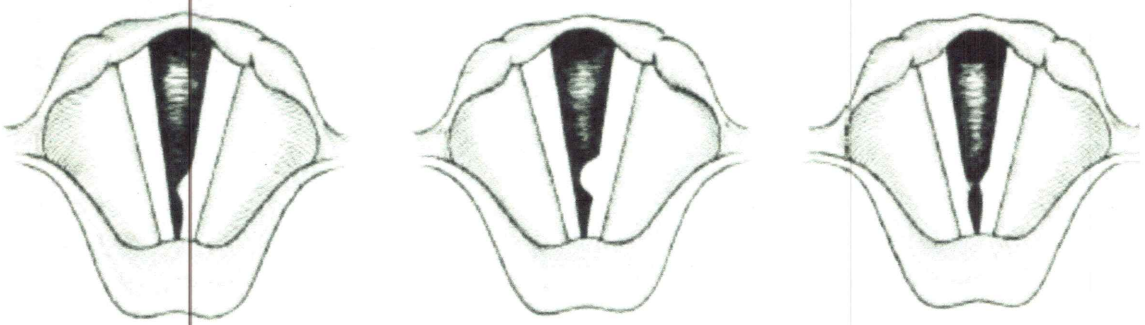

**Figure I.9 :** Divers types de nodules [8].

Le polype est une excroissance arrondie qui se forme sur la muqueuse qui tapisse les cordes vocales (Figure I.10). Il est généralement favorisé par une intoxication alcoolo-tabagique ou une exposition aux poussières. Les signes et symptômes du polype des cordes vocales comprennent une perte de la voix, une voix enrouée, un ton de voix grave ou une voix voilée.

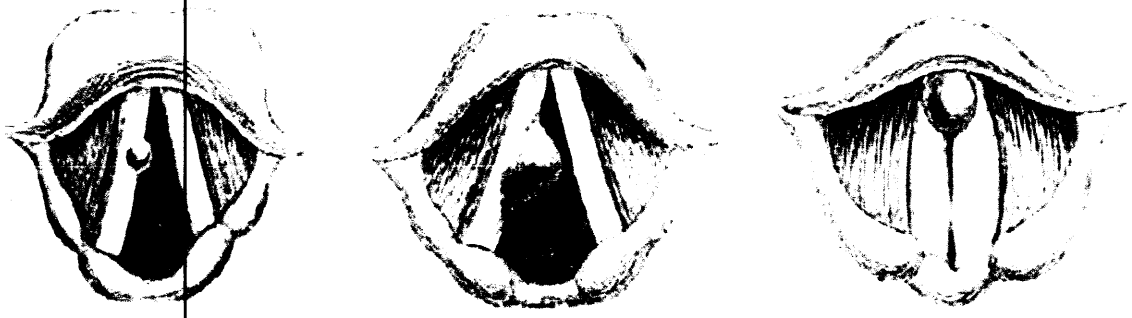

**Figure I.10 :** Différents types de polypes [9].

Les kystes apparaissent dans la couche superficielle à n'importe quelle partie des cordes vocales (Figure I.11). Lorsque le kyste croît, il exerce une pression sur les ligaments des cordes vocales sans affecter les couches adjacentes.

La présence de kystes se manifeste par une augmentation de la masse et un durcissement de la couverture. Comme conséquence, la glotte ne se ferme pas complètement durant la phonation.

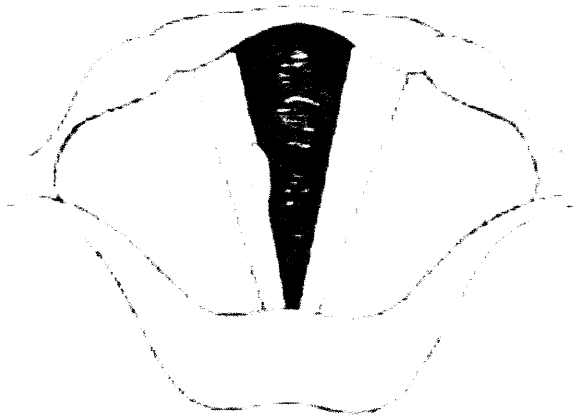

**Figure I.11 :** Kyste épidermique.

Les laryngites (Figure I.12) sont des inflammations aigues ou chroniques des cordes vocales, sont causées par des infections virales et/ou bactériennes. Ces laryngites augmentent la masse et la raideur des cordes vocales ce qui aboutit à la diminution de l'amplitude des vibrations de celles-ci. La voix est plus grave, rauque avec un timbre voilé, sourd et éraillé. Elle peut même disparaître totalement.

**Figure I.12 :** Laryngite herpétique.

Les traumatismes chirurgicaux, suite à l'ablation d'un cancer des cordes vocales, sont aussi parmi les causes du changement de l'anatomie du larynx.

### I.6.2 Dysphonies d'origines neurologiques :

Ces dysphonies peuvent être dues à un mauvais contrôle de la respiration ou encore d'une atteinte neurologique voire une difficulté psychologique. Dans ce cas, aucune lésion anatomique des cordes vocales n'est observée. Dans la référence [8], les auteurs définissent la pathologie vocale d'origine neurologique comme un trouble momentané ou durable de la fonction vocale ressenti comme tel par le sujet lui-même ou son entourage.

Les causes de ces dysphonies sont l'hypotonie et l'hypertonie de la musculature laryngée ou respiratoire. Ceci se traduit par une altération d'un ou plusieurs indices acoustiques de la voix tels que la fréquence fondamentale, le timbre et l'intensité. En présence de ces dysarthries ou dysphonies, une fatigue vocale et un malmenage ou surmenage vocal sont souvent observés.

L'hypotonie provoque un abaissement de l'intensité de la voix et de la fréquence fondamentale.

L'hypertonie, qui se manifeste par la difficulté à initialiser un acte volontaire du larynx, se traduit par des hésitations au démarrage du voisement, des émissions vocales discontinues et une augmentation de la fréquence fondamentale, un timbre sourd et voilé à cause du mauvais accolement des cordes vocales.

Les tremblements qui peuvent être de fréquence variable en fonction de leur cause, rendent la voix chevrotante [9].

Les dysphonies spasmodiques provoquent des changements brutaux de la hauteur de la voix qui peut s'interrompre, repartir, glisser et chevroter. Le timbre est désagréable.

Dans les paralysies laryngées, les cordes vocales demeurent dans une position plus ou moins ouverte. La voix est soufflée et rauque avec une importante fuite d'air, entrainant un essoufflement en fin de phrase.

## I. 7 Conclusion :

La majorité des notions et des caractéristiques du signal de la parole d'un sujet normal ont été présentées dans ce chapitre. Pour un sujet pathologique, plusieurs troubles vocaux sont souvent dus aux pathologies du larynx. Les changements morphologiques de l'anatomie du larynx engendrent des troubles vocaux ou dysphonies d'origines morphologiques, par contre un mauvais contrôle de la respiration, une atteinte neurologique ou une difficulté psychologique engendrent des dysphonies d'origines neurologiques.

# Chapitre II

# Méthodes d'évaluation des pathologies

## II.1 Introduction :

La qualité vocale caractérise la performance d'un appareil phonatoire aussi bien des voix normales que des voix dysphoniques. pour la caractérisation des troubles de la voix il existe plusieurs méthodes de diagnostic et de l'établissement d'un bilan vocal de la phonation du personne visé (patient), ces méthodes sont classées en deux types; soit objective appelée aussi instrumentale et qui se base sur les mesures des indices acoustiques, ou soit subjective appelée aussi « bilan fonctionnel », elle comporte trois parties essentiels: interrogatoire, analyse perceptive et l'autoévaluation [10]. Ce chapitre exposera ces différentes méthodes d'évaluation.

## II.2 Evaluation subjective (Bilan fonctionnel):

L'évaluation subjective consiste à des mesures ou des tests faits par le patient lui-même ou par un expert (phoniatres, médecin...). Elle est généralement sous forme de questionnaires et elle comprendra les deux types suivants:

### II.2.1 Interrogatoire (Auto-évaluation) :

L'interrogatoire du patient est sans doute la partie la plus longue car il est important de comprendre quelles sont ses plaintes et comment la voix est utilisée, l'interrogatoire permet dans le même temps de faire l'analyse perceptive de la voix.

Ce long entretien permet d'évaluer les plaintes du patient, d'apprécier ses contraintes professionnelles, d'évaluer son état psychologique et le retentissement éventuel des troubles vocaux, d'écouter la voix et d'en faire une analyse perceptive tout en appréciant le «geste vocal» (technique respiratoire, posture, détente notamment du cou et des mâchoires) et de demander au patient d'auto évaluer la qualité de sa voix.

### II.2.2 Evaluation perceptive :

C'est l'analyse la plus utilisée dans le monde, elle a pour but de faire juger la qualité vocale du patient par des phoniatres lesquels effectuent une description analytique des caractéristiques vocales sur la base d'échelles standardisées telles que : le CAPE-V ou, [11] le plus souvent, l'échelle GRBAS [12]. La GRBAS est une échelle perceptive basée sur l'évaluation de cinq paramètres acoustiques (Hirano, de 1981), par la suite complétée par « Djenkere » en 2001 pour devenir GRBASI [13], tel que:

**G** (*Grade*) : le grade général de pathologie, il est numéroté sur une échelle de 0 à 3 : le grade "0" représente la voix normale, "1" une dysphonie légère, "2" pathologie moyenne et "3" la dysphonie la plus grave.

**R** (Roughness) : le degré de raucité de la voix.

**B** (Breathiness) : le souffle.

**A** (Asthenia) : l'asthénie vocale ou faiblesse vocale.

**S** (Strain) : le forçage vocal.

**I** (Instability) : caractère instable.

Cette échelle peut être appliquée lors de la production d'une voyelle tenue, d'une phrase ou d'un texte généralement lus.

## II.3 Evaluation objective :

Souvent appelée évaluation ou analyse instrumentale, elle est conçue pour qualifier et surtout quantifier les dysphonies à partir des mesures acoustiques et/ou aérodynamiques [14]. Ces mesures sont le plus fréquemment réalisées sur une voyelle tenue, en général le /a/, à l'aide de différents capteurs conçus pour enregistrer et étudier de multiples paramètres de la production de parole. Il est toutefois souvent nécessaire de combiner différentes mesures complémentaires afin de tenir compte de l'aspect multidimensionnel de la production vocale [15].

Les mesures acoustiques (i.e. fréquence et amplitude, jitter et shimmer, analyse spectrale) révèlent les caractéristiques audibles de la dysphonie. Il s'agit principalement des mesures de la fréquence fondamentale et de l'intensité, de leur stabilité, ainsi que de l'analyse du spectre du son émis. Les mesures aérodynamiques, sans être à proprement parler des mesures de la voix, permettent d'évaluer les caractéristiques biomécaniques du système pneumo-phonatoire. Il s'agit principalement de mesures de débit, de pression et d'efficacité glottique.

Il existe aussi d'autres types d'évaluation objective, on site:

### II.3.1 Phonétogramme :

Le phonétogramme est l'un des mesures aérodynamiques, il sert à étudier la dynamique vocale qui représente de façon quantitative et qualitative le « champ de liberté de la voix ». C'est une évaluation globale de deux des paramètres de la voix : fréquence et intensité, ces deux derniers sont représentés dans un plan cartésien tel que les fréquences en abscisse et les intensités (entre 40 dB et 120 dB) en ordonnées (figure II.1).

En reliant les points mesurés, on obtient une sorte de forme géométrique ellipsoïdal aux extrémités rétrécis, dont la face inférieure représente les valeurs d'intensités les plus faibles et la face supérieure les valeurs les plus élevées. La distance entre les deux extrémités représente la dynamique tonale (en Hz) et l'épaisseur donne la dynamique de l'intensité en dB.

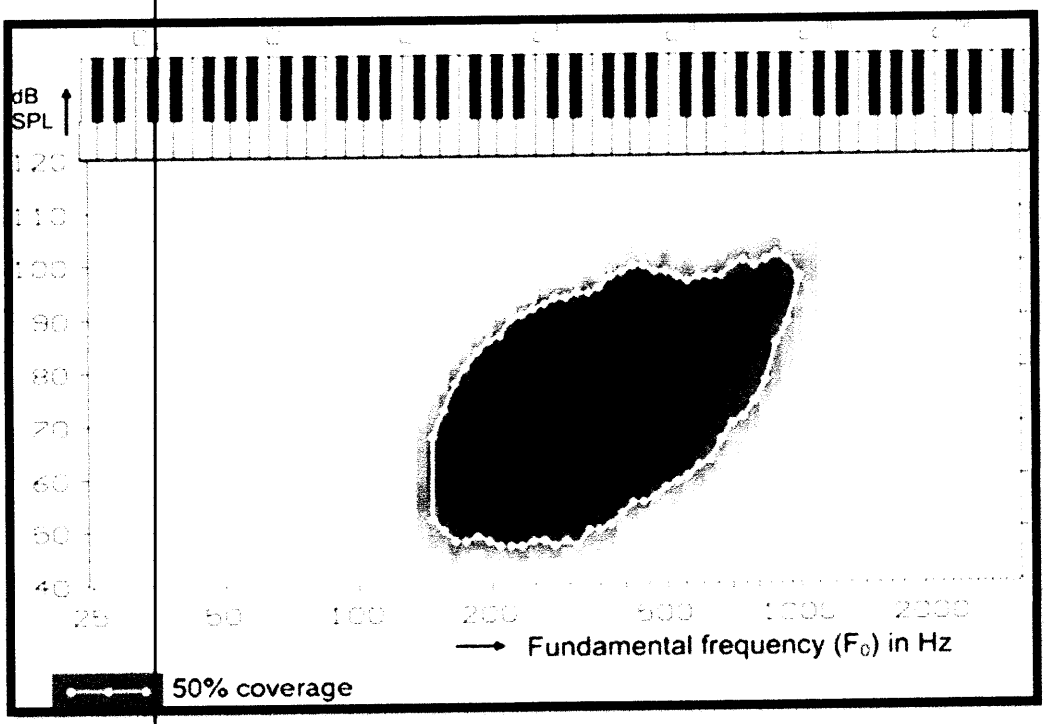


**Figure II.1 :** Phonétogramme d'un locuteur normal [16].

### II.3.2 Temps maximal de phonation :

Le temps maximum de phonation (TMP) consiste à mesurer la durée de la tenue d'une voyelle à une intensité et une fréquence données (valeur usuelle en général). On considère qu'il est significatif du rendement de la source vocale. Il est diminué en cas de fatigue vocale. Le test S/Z mesure la durée de l'émission des deux consonnes, l'une voisée ou sonore, l'autre non voisée ou sourde, et en fait le rapport.

Le TMP est plus grand chez les hommes que chez les femmes (à cause de leur grande capacité vitale), il est compris entre 25 et 35 secondes chez les hommes et de 15 à 25 secondes chez les femmes.

## II.4 Indices acoustiques pour la caractérisation des troubles de la voix :

Tous les paramètres (ou indices) acoustiques de la voix peuvent être altérés : la hauteur (ou fréquence ou tonalité), l'intensité, le timbre, mais aussi le débit ou l'articulation. C'est pour ça qu'on les utilise dans les mesures acoustiques non invasives (innocuité totale) pour l'évaluation de la parole. Elles donnent des résultats quantitatifs objectifs d'un échantillon de voix et sont maintenant informatisées et, donc, permettent d'identifier les dysphonies de la parole, ces indices acoustiques sont très nombreux et très difficile à citer. Comme la dysphonie

concerne essentiellement la source vocale, dans notre travail nous allons nous concentrer sur les paramètres directement liés à ce vibrateur (jitter, shimmer, HNR ....).

### II.4.1 Jitter (Vocal Jitter) :

Souvent appelé "la gigue vocale", il désigne des petites perturbations rapides des durées des cycles glottiques. C'est-à-dire les changements du pitch (ou de fréquence fondamentale) d'un cycle glottique à l'autre. Les origines exactes sont inconnues. Mais, on site généralement des causes neurologiques, l'écoulement turbulent de l'air à travers la glotte, la répartition inégale de mucus sur les plis vocaux, etc. La présence des pathologies amplifiée le jitter [17]. Il existe plusieurs représentations du jitter (tableau II.1) comme :

- La valeur moyenne de la fréquence fondamentale (Mean pitch).
- Valeur maximale (Maximum pitch detected) et minimale (Minimum pitch detected) du pitch, ainsi que l'écart-type (Standard Deviation of pitch contour).
- La gamme de fréquence phonatoire (Phonatory Frequency Range : PFR).
- Jitter absolu moyen (Mean Absolute Jitter : MAJ) : est la moyenne de la différence de fréquences entre deux cycles vibratoires du larynx consécutifs.
- Facteur de jitter (Jitter Factor) : est le jitter moyen rapporté à la fréquence fondamentale moyenne du signal.
- Rapport jitter (Jitter Ratio): parfois appelé jitter local (JITT) est la moyenne de toutes les différences, en valeur absolue, entre les durées de deux périodes consécutives du signal. On divise cette moyenne par la durée moyenne d'une période du signal, le résultat est donc un rapport, exprimé ici en %. Selon le manuel de Praat, le seuil normal/pathologique de ce critère est fixé à 1,04% [18].
- Perturbation moyenne relative (Relative Average Perturbation, RAP): comme le Jitter ratio, le RAP mesure les perturbations à court terme de la fréquence fondamentale. Ici, on compare la durée de chaque période Ti non pas à celle de la période suivante (Ti+1), mais à la moyenne de 3 périodes successives Ti-1, Ti et Ti+1, ce qui a théoriquement pour effet d'atténuer les variations volontaires de la fréquence vocale. Dans les faits, le Jitter ratio et le RAP sont des mesures à peu près équivalentes. Le RAP est également exprimé en %. Le seuil normal/pathologique est fixé à 0,68% dans Praat.
- Quotient de perturbation de pitch (Pitch Perturbation Quotient, PPQ5): est le quotient de perturbation à cinq points de périodes. Calculé comme la différence moyenne, absolue entre

une période et la moyen des autres quarte périodes voisines, tous divise par la période moyenne. De plus, le PPQ55 utilise 55 périodes de pitch.

| Description | Méthode de calcul |
|---|---|
| Pitch moyen | $F0_{av} = \frac{1}{n}\sum_{i=1}^{n} F_i$ (2.1) |
| Maximum et minimum de Pitch détectée | $F0_{hi} = max(F0)$ (2.2)<br>$F0_{lo} = min(F0)$ (2.3) |
| Ecart-type du contour de Pitch | $F0_{sd} = \frac{1}{n-1}\sum_{i=1}^{n}(Fi - \bar{F})^2$ (2.4) |
| Gamme de la fréquence phonatoire | $PFR = \frac{\log(\frac{F0_{hi}}{F0_{lo}})}{log2} \times 12$ (2.5) |
| Jitter absolu moyen | $MAJ = \frac{1}{n-1}\sum_{i=n-1}^{i}\lvert F_{i+1} - F_i\rvert$ (2.6) |
| Jitter (%) | $JITT = \frac{MAJ}{F0_{av}}$ (2.7) |
| Perturbation relative moyenne | $RAP = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left\lvert\frac{F_{i+1}+F_i+F_{i-1}}{3} - F_i\right\rvert}{F0_{av}} \times 100$ (2.8) |
| Quotient de perturbation de Pitch lissé sur 5 périodes de pitch | $PPQ_5 = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\left\lvert\frac{\sum_{k=i-2}^{i+2} F(k)}{5} - F_i\right\rvert}{F0_{av}} \times 100$ (2.9) |
| Quotient de perturbation de Pitch lissé sur 55 périodes de pitch | $PPQ_{55} = \frac{\frac{1}{n-54}\sum_{i=28}^{n-27}\left\lvert\frac{\sum_{k=i-27}^{i+27} F(k)}{55} - F_i\right\rvert}{F0_{av}} \times 100$ (2.10) |

**Tableau II.1:** Paramètres de perturbation du pitch [19].

La valeur du Jitter à une relation directe avec le degré de pathologie, c'est-à-dire plus la valeur de Jitter augmente, plus la pathologie est grave.

### II.4.2 Shimmer :

Le Shimmer est un coefficient, généralement exprimé en dB et caractérise les variations d'amplitudes entre deux cycles du larynx consécutifs. Comme le Jitter, le Shimmer est en proportionnalité directe avec le grade de pathologie et possède plusieurs types (tableau II.2), le plus utilisé est le Shimmer moyen (dB) donné par :

$$Shimmer\ (dB) = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log(\frac{A_i}{A_{i+1}})\right| \qquad (2.11)$$

Tel que : A est l'amplitude du signal et N le nombre d'échantillons. En plus des mesures suivantes :

- Amplitude moyenne (Mean Amplitude : $Amp_{av}$). Amplitude maximale et minimale détectée (Maximum and Minimum Amplitude detected). Ecart-type (Standard Deviation of Amplitude Contour).

- Shimmer moyen absolu (Mean Absolute Shimmer : MAS) et Shimmer (dB ou %).

- Perturbation relative moyenne d'amplitude (Amplitude Relative Average Perturbation : ARP).

- Shimmer (local shimmer $SHIM$) : On mesure ici les perturbations à court terme de l'amplitude du signal sonore. Pour ce faire, 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. Le seuil normal/pathologique est fixé à 3,81 % dans Praat.

- Quotient de perturbation d'amplitude (Amplitude Perturbation Quotient, APQ5): est une mesure des perturbations à court terme de l'amplitude du signal. Sur le même principe que le RAP et le ARP, il s'agit d'atténuer les effets des modulations volontaires d'intensité en comparant l'amplitude maximale de chaque période Ti à l'amplitude moyenne des pics des périodes Ti-2 à Ti+2, soit 5 périodes.

- Quotient de perturbation d'amplitude lissé sur 55 périodes de pitch (Amplitude Perturbation Quotient, APQ55).

| Description | Méthode de calcul |
|---|---|
| Amplitude moyenne | $Amp_{av} = \frac{1}{n}\sum_{i=1}^{n} A_i$ (2.) (2.12) |
| Maximum et minimum de l'amplitude détectée | $A_{hi} = max(A_i)$ (2.13)<br>$A_{lo} = min(A_i)$ (2.14) |
| Ecart-type du contour de l'amplitude | $A_{sd} = \frac{1}{n-1}\sum_{i=1}^{n}(A_i - \bar{A})^2$ (2.15) |
| Shimmer absolu moyen | $MAS = \frac{1}{n-1}\sum_{i=n-1}^{i}\lvert A_{i+1} - A_i\rvert$ (2.16) |
| Shimmer (dB) | $SHIM = \frac{1}{n-1}\sum_{i=1}^{n-1} 20 \times log\left(\frac{A_i}{A_{i+1}}\right)$ (2.17) |
| Perturbation relative moyenne d'amplitude | $ARP = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left\lvert\frac{A_{i+1}+A_i+A_{i-1}}{3} - A_i\right\rvert}{Amp_{av}} \times 100$ (2.18) |
| Quotient de perturbation d'amplitude lissé sur 5 périodes de pitch | $APQ_5 = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\left\lvert\frac{\sum_{k=i-2}^{i+2} A(k)}{5} - A_i\right\rvert}{Amp_{av}} \times 100$ (2.19) |
| Quotient de perturbation d'amplitude lissé sur 55 périodes de pitch | $APQ_{55} = \frac{\frac{1}{n-54}\sum_{i=28}^{n-27}\left\lvert\frac{\sum_{k=i-27}^{i+27} A(k)}{55} - A_i\right\rvert}{Amp_{av}} \times 100$ (2.20) |

**Tableau II.2 :** Paramètres de perturbation d'amplitude [19].

### II.4.3 Rapport harmoniques sur bruit HNR :

Le rapport harmoniques sur bruit ou (HNR : Harmonique to Noise Ratio), est la mesure la plus classique proposé par Yumoto en 1982 [20], considéré comme le rapport d'énergie des harmoniques par rapport à l'énergie du bruit et exprimé en dB. Le HNR diminue avec l'âge alors que l'instabilité vocale augmente.

L'équation du HNR dans le domaine fréquentiel est obtenue après la décomposition du signal en deux composantes ; une composante période (harmoniques H) et une composante apériodique (bruit N) :

$$HNR = 10\,log_{10}\left(\frac{\sum_{k=0}^{M-1} H^2(k)}{\sum_{k=0}^{M-1} N^2(k)}\right) \tag{2.21}$$

Il existe plusieurs méthodes pour calculer le HNR dans le domaine temporel, dans le domaine fréquentiel et dans le domaine cepstral. Par la suite nous allons présenter l'algorithme de De Krom [21], qui est le plus utilisé, le plus cité dans la littérature et qui se base sur l'analyse cepstrale.

## II.5 Analyse Cepstrale :

### II.5.1 Transformation homomorphique:

Le filtrage (ou transformation homomorphique) $\hat{A}[n] = D(A[n])$, est une transformation non linéaire généralement appliquée dans le traitement d'images et de parole pour convertir le signal obtenu par la convolution de deux signaux, vers un signal composé par la somme de deux signaux, comme il est montré par l'équation suivante [22] :

$$x[n] = u[n] * h[n] \rightarrow \hat{x}[n] = \hat{u}[n] + \hat{h}[n] \tag{2.22}$$

Dans le traitement de la parole, le filtrage homomorphique est appliqué pour séparer l'excitation glottique et la réponse du filtre de conduit vocal (figure II.2).

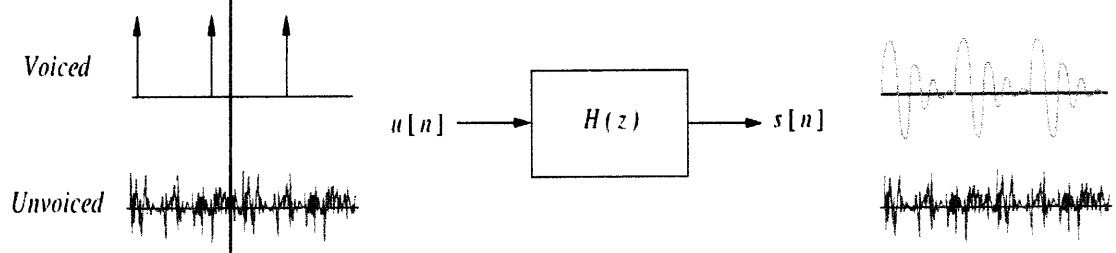


**Figure II.2:** Signal d'excitation, la réponse du filtre et le signal de parole.

### II.5.2 Le cepstre:

L'opération « cepstre » est une méthode numérique basée sur la transformation homomorphique permettant d'étudier séparément la source et le conduit vocal. Le cepstre est basé sur une connaissance du modèle de production de la parole (figure II.3), une modélisation du signal de parole consiste à définir ce signal comme le résultat de la convolution de la fonction de transfert du conduit vocal (filtre) par un signal d'excitation (source). Le but du cepstre est de séparer ces deux contributions (source et filtre) par application de la déconvolution.

Le cepstre réel a une amplitude réelle, non négative. Il n'utilise que l'amplitude du spectre de ce signal, il perd donc la partie de l'information contenue dans la phase et l'on ne peut donc pas reconstruire parfaitement le signal de départ à partir de ce cepstre. Sa relation est donnée par l'équation:

$$c_x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{jw})|\, e^{jwn} dw \qquad (2.23)$$

Pour le cepstre complexe :

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} [\log|X(e^{jw})| + j \arg(X(e^{jw}))]\, e^{jwn} dw \qquad (2.24)$$

Il faut noter que arg($Xe^{jw}$) représente la phase et qu'il est appelé complexe car il utilise le logarithme appliqué à la valeur complexe de la transformée de Fourier. Le cepstre complexe contient donc à la fois l'information d'amplitude et de phase du spectre du signal, ce qui va permettre notamment de reconstruire le signal de départ.

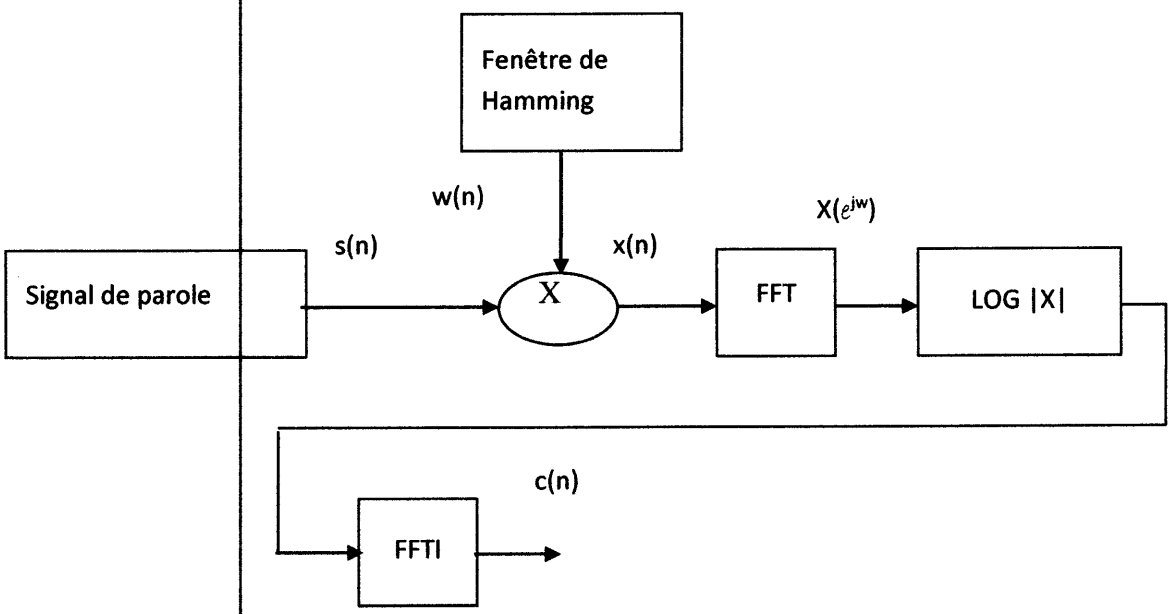


**Figure II.3 :** Schéma bloc général de détermination du cepstre [23].

**Propriétés du cepstre :**

- ✓ Il a une durée infinie, même si x[n] a une durée finie.
- ✓ Il est réel si x[n] est réel, autrement dit, les pôles et les zéros sont des paires conjugués complexes.
- ✓ Vocabulaire [24] :

| | | |
|---|---|---|
| ▪ Spectre | → | Cepstre |
| ▪ Fréquence | → | Quéfrence |
| ▪ Filtrage | → | Liftrage |
| ▪ Harmonique | → | Rahmonique |
| ▪ Période | → | Répiode |
| ▪ Phase | → | Saphe |
| ▪ Amplitude | → | Gamnitude |



## II.6 Détermination du pitch basée sur l'analyse Cepstrale :

Dans la plupart des algorithmes d'extraction du pitch, trois phases essentielles durant le traitement s'impliquent : le prétraitement, le traitement et le post-traitement (Figure II.4).

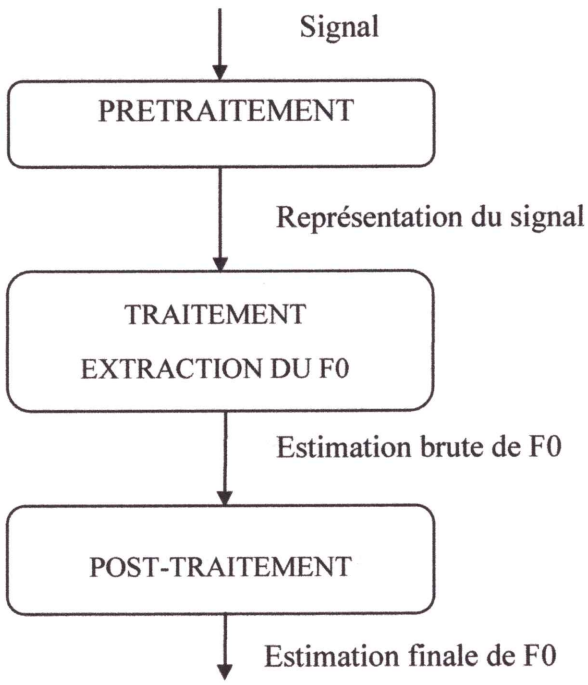


**Figure II.4 :** Schéma global d'un algorithme de détermination du pitch [25].

**a. Phase de prétraitement**

Est réservée à la préparation du signal de la parole. Elle consiste à choisir la durée des trames d'analyse et du recouvrement, afin de moins compromettre la condition de stationnarité exigée par les algorithmes de traitement et l'effet de bord lié aux fenêtres de pondération appliquées.

La durée de la trame est généralement choisie entre 20 et 50 ms avec un recouvrement de 30 à 75%, pour assurer la présence d'au moins une période du fondamental.

**b. Phase de traitement**

Est réservée à l'extraction de la fréquence fondamentale (F0) en utilisant la méthode cepstrale.

Le principe de la procédure de calcul de pitch fondé sur le cepstre est plutôt simple. On recherche dans le cepstre un pic dans la région autour de la période du pitch (P).

Si le pic est supérieur à un seuil fixé (P0), le segment de parole en entrée est probablement voisé, et la position autour du pic est la zone dans laquelle on peut estimer le pitch.

Si le pic n'est pas supérieur au seuil, il est alors probable que le segment de parole en entrée est non voisé. La figure (II.5) montre le schéma bloc d'estimation de pitch par la méthode cepstrale :

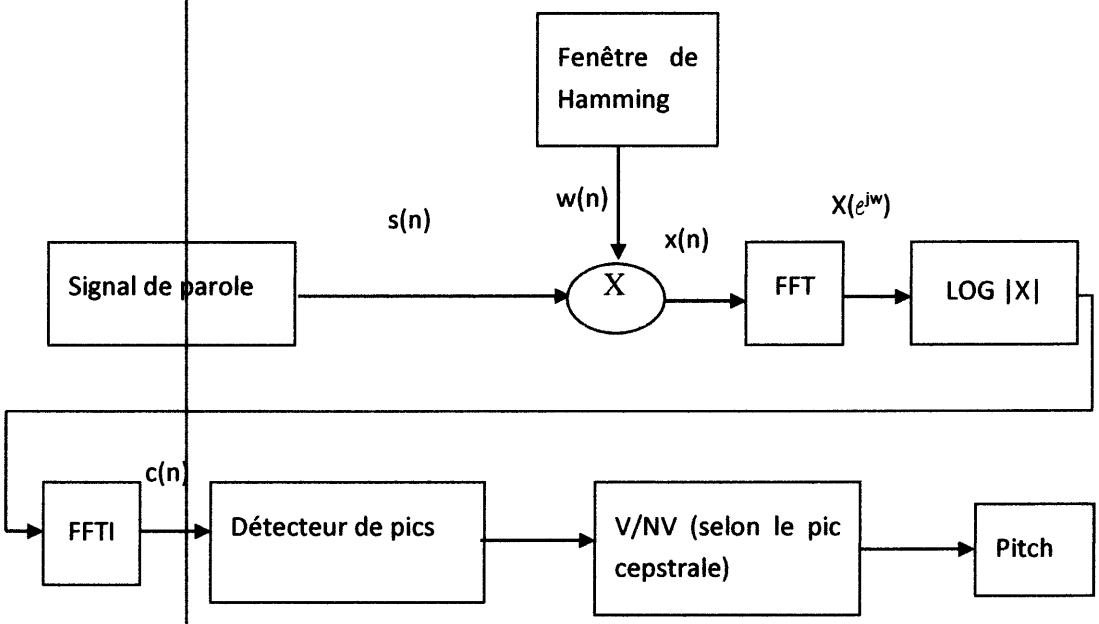


**Figure II.5** : Schéma bloc d'estimation de pitch par la méthode cepstrale [26].

La figure (II.6) présente le cepstre réel d'une trame de la voyelle tenue synthétique « a » :

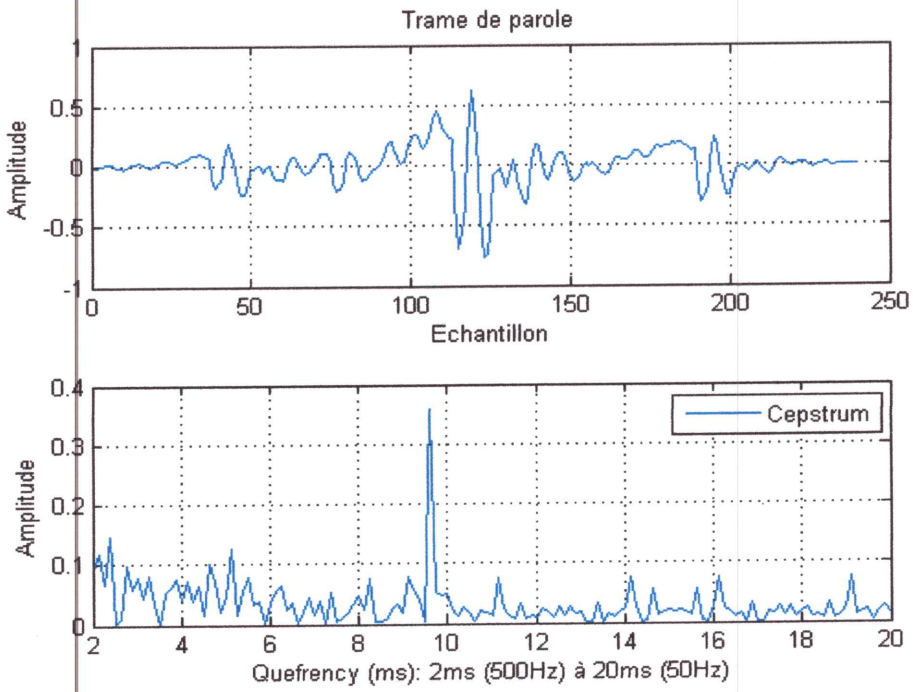


**Figure II.6** : Trame de 20ms du signal de parole et son cepstre (voyelle 'a').

La présentation du contour du pitch, sa valeur moyenne estimée par le cepstre et la forme d'onde du signal de parole complet sont illustrés sur le schéma de la figure (II.7) :

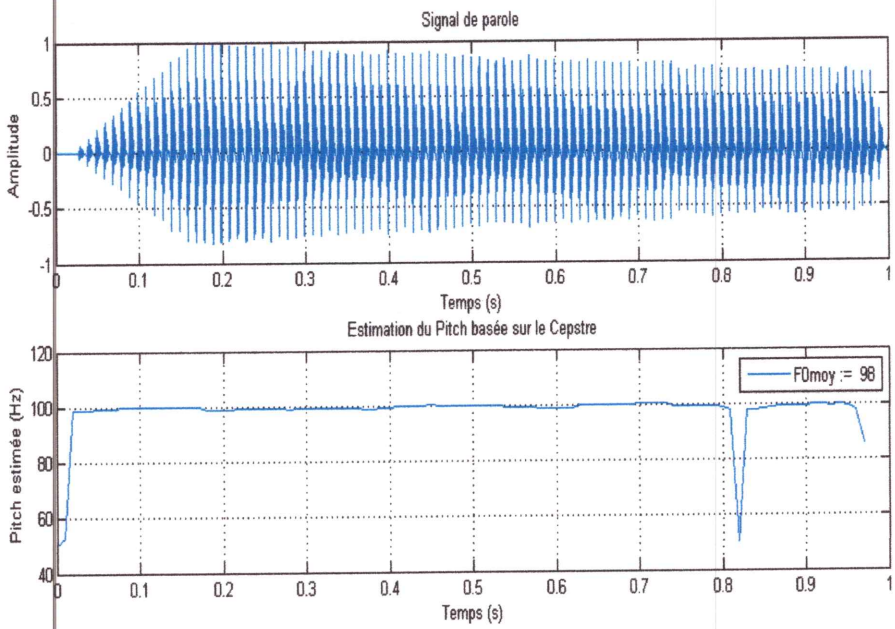


**Figure II.7** : Forme d'onde et contour du pitch estimé par la méthode cepstrale (voyelle 'a').

**c. Phase de post-traitement**

A pour but de diminuer les erreurs d'estimation. Qui sont de plusieurs types :

- les erreurs de voisement : lorsqu'une valeur de F0 a été trouvée sur une zone non- voisée, ou lorsqu'aucune n'a été trouvée sur une zone voisée.
- les erreurs grossières ("gross-errors") : la fréquence fondamentale correspond à une harmonique ou une sous-harmonique. Ce type d'erreur peut facilement être corrigé en tenant compte du voisinage ou en effectuant un lissage.
- Les erreurs fines : la valeur trouvée est située à plus ou moins 10% de la valeur réelle. Les techniques de post-traitement les plus courantes sont : le filtre médian, le lissage linéaire et la technique de la programmation dynamique.

## II.7 Technique basée sur le cepstre pour la détermination du HNR :

Dans [21], De Krom a proposé une technique de calcul du HNR dans les signaux de la parole basée sur l'analyse cepstrale. Cette méthode utilise une séparation entre l'énergie des harmoniques et l'énergie du bruit en utilisant un filtre en peigne (comb-liftering) dans le domaine cepstral.

### II.7.1 Formulation de base :

Généralement, un signal de parole est segmenté en plusieurs trames successives y(t), qui est le résultat de la convolution du signal d'excitation x(t) et la réponse impulsionnelle h(t) du conduit vocal :

$$y(t) = x(t) * h(t) \tag{2.25}$$

On calcule maintenant le spectre d'amplitude à court terme par une transformation de Fourier de la trame y (t). Afin d'éviter toute distorsion du spectre, nous multiplions y(t) par une fenêtre de (Hanning) w(t) et nous obtenons alors une trame de signal lissé $y_w(t)$ :

$$y_w(t) = [x(t) * h(t)] \times w(t) \tag{2.26}$$

Comme w (t) varie lentement par rapport à h(t), on peut récrire (2.26) comme suit :

$$y_w(t) \approx x_w(t) * h(t) \tag{2.27}$$

Où : $x_w(t) = x(t).w(t)$

L'application de la transformée de Fourier sue la trame fenêtrée du signal : $y_w(t)$, donne les coefficients réels et imaginaires qui seront utilisés pour calculer le spectre d'amplitude $Y_w(f)$ :

$$DFT[y_w(t)] \rightarrow Re[Y(f)] + I[Y(f)]$$

$$|Y_w(f)| = \sqrt{(Re[Y(f)]^2 + Im[Y(f)]^2)} \quad (2.28)$$

Le spectre d'amplitude $|Y_w(f)|$ peut être considéré comme le produit de $X_w(f)$, le spectre du signal fenêtré de la source d'excitation, et H(f) la réponse en fréquence du conduit vocal :

$$|Y_w(f)| = X_w(f).H(f) \quad (2.29)$$

En prenant le logarithme du spectre d'amplitude, la relation multiplicative entre le spectre de la source et la réponse en fréquence est changée en une relation d'addition :

$$\log|Y_w(f)| = \log(X_w(f).H(f))$$

$$\log|Y_w(f)| = \log X_w(f) + \log H(f) \quad (2.30)$$

Par la suite, le spectre original non filtré $\log|Y_w(f)|$ sera appelé O(f). Ainsi, le logarithme du spectre d'amplitude O(f) peut être considéré comme la somme de deux spectres : $\log X_w(f)$, le logarithme du spectre du signal pondéré de la source, et log H(f), la réponse fréquentielle du conduit vocal. De plus, le spectre de la source lui-même est constitué de deux parties; une série d'harmoniques régulièrement espacées, avec des amplitudes décroissantes avec la fréquence, et une partie de bruit distribué aléatoirement.

En regardant l'équation (2.30) précédente, le log du spectre est considéré comme la somme de deux signaux : l'enveloppe spectrale et le spectre de la source. L'enveloppe est la composante qui varie lentement, tandis que la série des harmoniques dans le spectre de la source, constitue la composante périodique, par contre le bruit dans le spectre de la source constitue une composante irrégulière qui varie rapidement. En prenant la transformée de Fourrier inverse du "log de spectre", on obtient le cepstre réel $C(\tau)$ :

$$IDFT[O(f)] \rightarrow Re[C(t)] + Im[C(t)] \quad (2.31)$$

Comme les coefficients imaginaires sont nuls, nous pouvons prendre les coefficients réels pour le cepstre réel :

$$C(\tau) = Re[C(\tau)] \quad (2.32)$$

$\tau$ : quéfrence en ms.

L'enveloppe spectrale, étant une partie qui varie lentement, contribue le plus aux basses quéfrences du cepstre. En revanche, la série périodique d'harmoniques, donnera lieu à quelques pics équidistants dans le cepstre, généralement appelées rahmoniques. Enfin, le bruit contribuera à diverses parties du cepstre. La structure fine du bruit contribuera principalement aux quéfrences les plus élevées, alors que son niveau global et sa forme brute contribueront à la composante DC du Cepstre (la première valeur du cepstre) et les composantes faibles des quéfrences. Cependant, comme l'allure du spectre de bruit n'est pas périodique et ne varie pas lentement, les harmoniques, l'enveloppe spectrale et l'énergie du bruit contribueront à des quéfrences tout à fait disjoints.

Les propriétés des quéfrences du cepstre permettent diverses opérations de filtrage dans le domaine cepstrale, appelées généralement liftrage "liftering". Dans notre cas, nous voulons faire une séparation entre l'énergie des harmoniques et l'énergie du bruit dans le spectre. A cette fin, il peut être utile de penser que le cepstre est la somme de deux composantes séparées, l'une contient seulement tous les pics des harmoniques (rahmoniques), tandis que l'autre se compose de tous les échantillons non-harmoniques du cepstre. La dernière composante peut être appelée "rahmonic comb-liftered part" : la partie du cepstre filtrée par un filtre en peigne. Si nous appliquons maintenant la transformée de Fourier à la partie comb-liftered du cepstre, nous obtenons un spectre qui a conservé son enveloppe spectrale et son énergie de bruit, mais qui est dépourvue d'harmoniques. Nous appellerons cela le spectre de bruit N(f), qui est en fait constitué à la fois des composantes du bruit eux-mêmes et de l'enveloppe spectrale. Etant donné que l'enveloppe spectrale contribue de la même façon dans le niveau du spectre original et dans celui du spectre de bruit obtenu, on peut définir le HNR pour une bande de fréquences β comme étant la différence de niveau en dB entre le spectre original et le spectre de bruit :

$$HNR_{\beta}(dB) = level\ O_{\beta}(f) - level\ N_{\beta}(f) \qquad (2.33)$$

Tel que : $o < \beta \leq f_{Nyquist}$

Le niveau du spectre pour une bande de fréquences β est défini comme le logarithme en base 10 de la somme des valeurs du spectre de puissance dans cette bande de fréquence.

### II.7.2 Etapes de détermination du spectre de bruit :

Pour la détermination du spectre de bruit N (f), on doit passer par plusieurs étapes. Tout d'abord, le filtre en peigne du cepstre (cepstrum comb-lifter) doit être défini. Une routine simple

de détection des pics (peak-picking) peut être utilisée pour localiser le premier pic proéminent des rahmoniques à τ ms, correspondant à la période du fondamentale. Les autres pics des rahmoniques seront recherchés aux intervalles multiples de τ ms sur toute la plage des quéfrences.

Une fois les pics des rahmoniques sont localisés, la bande passante du filtre en peigne est déterminée pour chaque pic de rahmonique en se basant sur le changement de signe dans la dérivée première de la fonction du cepstre à gauche et à droite de ce pic. Ce changement de signe se produit habituellement à, ou très près d'un échantillon du cepstre d'amplitude nulle. Pour les rahmoniques d'ordre élevé, la bande passante est généralement constituée de quelques échantillons. Les amplitudes de tous les échantillons du cepstre à l'intérieur des dents du filtre en peigne seront remisent à zéro, ce qui donne un cepstre de rahmonique comb-liftered $C_{cl}(\tau)$ :

$$C_{cl}(\tau) = 0, \qquad \tau \in [n.T - 0.5.CW, n.T + 0.5.CW]$$

Où CW est la largeur de peigne (CombWidth),

$$C_{cl}(\tau) = C(\tau), \tag{2.34}$$

Pour toutes les autres valeurs de τ où ($n \geq 1$) et T est la quéfrence en ms du premier pic.

Le cepstre filtré par le filtre en peigne (comb-liftered cepstrum : $C_{cl}(\tau)$) est transformé dans le domaine spectrale par la transformée de Fourier, ce qui donne une première approximation du spectre de bruit $N_{ap}(f)$ :

$$N_{ap}(f) = DFT[C_{cl}(\tau)] \tag{2.35}$$

Notant que l'information de l'enveloppe spectrale a été retenue dans cette approximation du spectre de bruit $N_{ap}(f)$, car le liftrage affecte seulement les rahmoniques.

### II.7.3 Procédure de correction de la ligne zéro (baseline correction procedure) :

Compte tenu de notre définition du bruit spectral, qui est une énergie spectrale indépendante de l'énergie harmonique ou de celle de l'effet de l'enveloppe spectrale, une soustraction du spectre de bruit estimé N (f) à partir du spectre original non filtré, doit produire un spectre de la source purement harmonique : Ha (f) avec comme niveau de référence la ligne zéro dB :

$$Ha(f) = O(f) - N(f) \tag{2.36}$$

Cependant, le spectre de la source harmonique obtenu après la soustraction du spectre de bruit estimé à partir du spectre original, ne reste pas au-dessus de la ligne zéro dB, mais il est plus ou moins symétrique autour d'elle. Nous appellerons ce spectre par : "le spectre d'harmonique approximé" $Ha_{ap}(f)$ et il est donné par :

$$Ha_{ap}(f) = O(f) - N_{ap}(f) \tag{2.37}$$

Le spectre de la source harmonique doit être centré par rapport à la ligne zéro. Notant que le niveau sous-estimer du spectre d'harmonique approximé (par rapport à la ligne zéro) par définition, correspond au niveau surestimer du spectre de bruit. Alors, si nous déterminons l'éloignement du spectre harmonique de la source par rapport à la ligne zéro dB, nous connaîtrons aussi l'éloignement du spectre du bruit estimé par rapport au spectre du bruit réel. La procédure de correction suit les étapes suivantes :

Pour chaque harmonique dans le spectre estimé de la source d'harmoniques $Ha_{ap}(f)$, la valeur absolue du minimum entre deux pics d'harmoniques successives est calculée, ce qui donne une série de valeurs d'éloignement par rapport à la ligne de base Bd (f) :

$$Bd(f) = \left|min\left[Ha_{ap}(f)\right]\right|$$

Où $$(n-1).F_0 < f \leq n.F_0 \; et \; n \geq 1 \tag{2.38}$$

Les valeurs de Bd(f) seront soustraites du spectre de bruit approximé $N_{ap}(f)$, ce qui donne le spectre de bruit estimé final N (f) :

$$N(f) = N_{ap}(f) - Bd(f) \tag{2.39}$$

Enfin, le HNR pour chaque bande de fréquence β sera calculé par la différence de niveau entre le spectre original $O(f)$ et le spectre de bruit $N(f)$ dans cette bande de fréquences, comme montré à l'équation (2.39).

La figure (II.8) résume toutes les étapes à suivre pour le calcul du HNR :

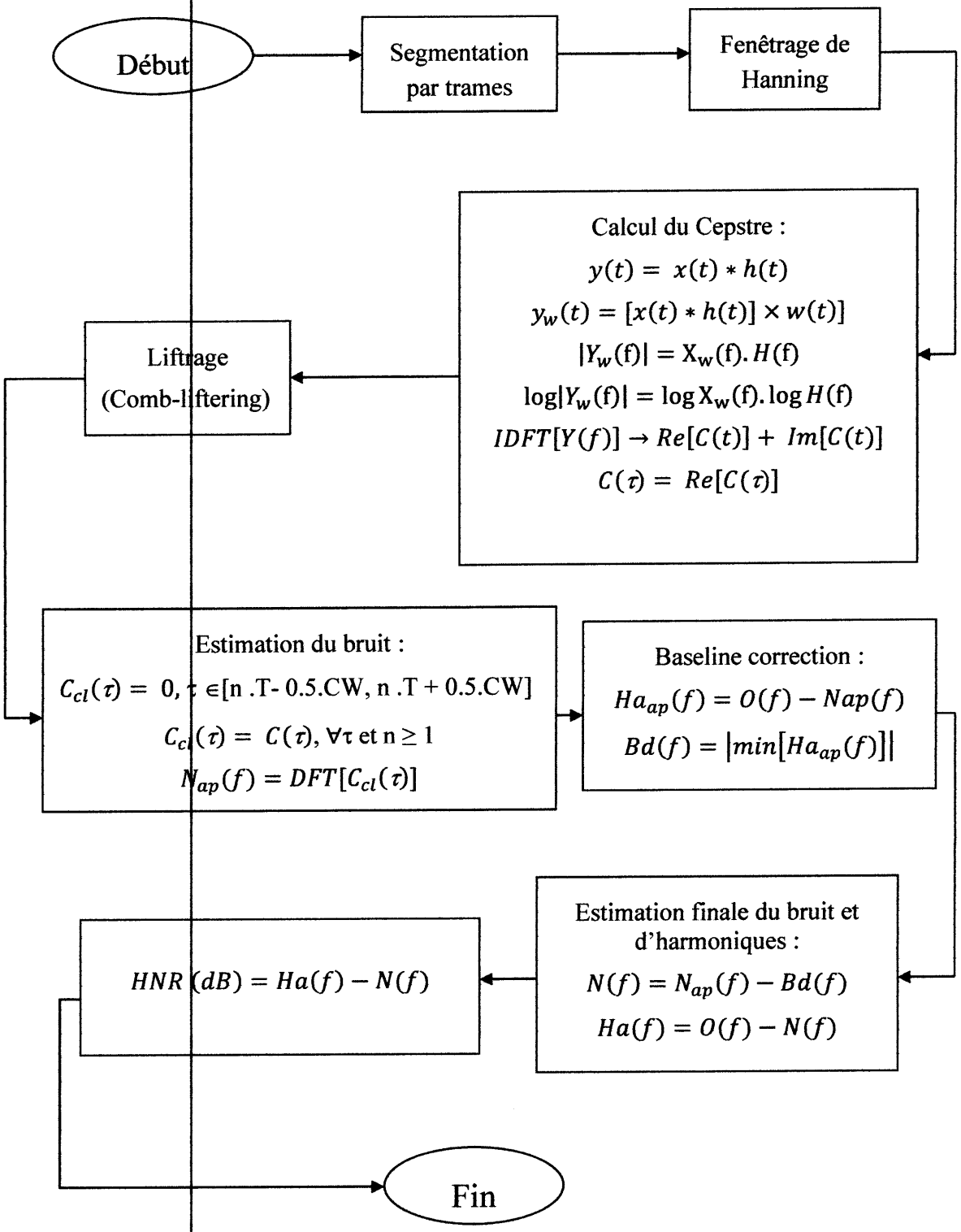


**Figure II.8 :** Organigramme des différentes étapes pour le calcul du HNR.

## II.8 Conclusion :

L'évaluation des pathologies de la voix peut être basée sur des méthodes subjectives et objectives. L'échelle GRB est le plus utilisé dans les tests subjectifs et le plus rencontré dans les bases de données des voix pathologiques, en particulier le grade (G). A noter qu'on dispose des bases de données avec un grade G de chaque stimulus par juge ou un grade moyen des juges.

Les variations de la fréquence fondamentale cycle-à-cycle, les variations de l'amplitude cycle-à-cycle et le rapport harmoniques sur bruit forment l'ensemble des indices acoustiques de l'évaluation objective. Ses définitions et ses relations ont été présentées dans ce chapitre et seront simulées et appliquées sur des fichiers et des bases de données des voix normales et pathologiques dans le chapitre suivant.

# Chapitre III

# Simulation des Indices Acoustiques et Evaluation des Pathologies

## III.1 Introduction :

Les pathologies vocales sont relativement communes affectant cinq pour cent de la population et se retrouvent à des degrés de progression et de gravité variable [27]. Le développement des méthodes non invasives pour le diagnostic de la pathologie de la voix ont été motivés par la nécessité de procéder à une analyse à la fois objective et efficace de la fonction vocale du patient. À l'heure actuelle un certain nombre d'outils de diagnostic sont disponibles pour les ORL (specialists de l'oto-rhino-laryngologie : étude de l'oreille, du nez et du larynx) et des orthophonistes y compris vidéostroboscopie et videokymography. Les méthodes de détection de pathologies vocales basées sur l'analyse acoustique ont été rapportées avec une précision de plus de 90% [21], le développement de ces systèmes de détection automatique a mis l'accent sur les algorithmes d'extraction de caractéristiques (pitch...) et les mesures de perturbation tels que la gigue vocale (jitter), le shimmer et le HNR. Ces indices permettent de fournir une base de discrimination pathologique /normal.

Au cours de ce chapitre, nous présenterons les corpus utilisés dans notre simulation, les expériences de l'évaluation subjective (perceptive) et objective (HNR, Jitter et Shimmer), la corrélation et les interprétations de ces résultats.

## III.2 Description des corpus et dispositifs utilisés :

Afin d'appliquer et de tester l'efficacité des algorithmes des indices acoustiques d'évaluation objective de la parole pathologique, on doit disposer de plusieurs fichiers de parole normale et pathologique issues des bases de données synthétiques ou naturelles. Ces bases doivent contenir des sons de plusieurs locuteurs masculins et féminins, plusieurs valeurs du pitch et différents niveaux de bruit, de jitter et de shimmer.

On dispose de trois bases de données appelées (Corpus A, B et C), constituées de plusieurs fichiers sons avec leurs évaluations subjectives par des juges.

### III.2.1 Corpus de simulation :

- **Corpus A :**

Le corpus A est composé de plusieurs fichiers de la voyelle soutenue synthétique /a/. Ils ont été générés par un synthétiseur articulatoire qui utilise des modèles de l'aire glottique et du flux d'air à travers la glotte [28]. Le corpus comprend 48 stimuli de voyelles synthétiques et composé de trois valeurs différentes du pitch (100 Hz, 120 Hz et 140 Hz).

Pour chaque valeur du pitch, on trouve quatre valeurs différentes du jitter et quatre niveaux différents du bruit. La durée de chaque voyelle est d'une seconde et la fréquence d'échantillonnage est de 22050 Hz avec un codage de 16 bits.

Huit thérapeutes et un phoniatre ont évalué perceptivement le corpus A selon le grade (G), la raucité (R) et le souffle (B) en utilisant les quatre degrés par échelle (0 : son normal, 1 : pathologie légère, 2 : pathologie moyenne, 3 : pathologie sévère). Seul le grade (G) sera utilisé dans le cadre de ce projet car il fournit une mesure globale de la qualité de la voix. Généralement, on donne le score moyen du grade de l'ensemble des neuf juges.

Le tableau suivant montre que les juges sont corrélés entre eux dans leurs évaluations perceptives du grade des voix synthétiques du corpus A.

| | J1 | J2 | J3 | J4 | J5 | J6 | J7 | J8 | J9 |
|---|---|---|---|---|---|---|---|---|---|
| J1 | | | | | | | | | |
| J2 | 0.6282 | | | | | | | | |
| J3 | 0.7491 | 0.6979 | | | | | | | |
| J4 | 0.6283 | 0.7511 | 0.7469 | | | | | | |
| J5 | 0.49 | 0.4511 | 0 .6064 | 0.5166 | | | | | |
| J6 | 0.7177 | 0.5678 | 0.7473 | 0.5937 | 0.4565 | | | | |
| J7 | 0.6526 | 0.8072 | 0.7523 | 0.8044 | 0.6095 | 0.5955 | | | |
| J8 | 0.6617 | 0.7256 | 0.7555 | 0.7273 | 0.7142 | 0.6074 | 0.7808 | | |
| J9 | 0.7708 | 0.7325 | 0.8101 | 0.7608 | 0.4908 | 0.7085 | 0.7593 | 0.7922 | |

**Tableau III.1 :** Valeurs de corrélations inter-juges du grade (G).

- **Corpus B et C :**

Les deux corpus B et C sont constitués de la parole naturelle. Les 251 stimuli de chaque corpus sont produits par 28 locuteurs normophoniques et 223 locuteurs dysphoniques dont 146 masculins et 77 féminins, âgés de 8 ans à 85 ans avec différents degrés de dysphonie. Les sujets dysphoniques qui sont des patients de l'hôpital Saint-Jean (Sint-Janshospitaal) de Bruges en Belgique, présentent des pathologies d'origine morphologique et des pathologies d'origine neurologiques.

Le corpus B est une base de données de la voyelle naturelle /a/ où chaque fichier est de durée 3 seconds, échantillonnés à 44100 Hz et codé sur 16 bits.

La concaténation de la voyelle [a] avec deux phrases en Néerlandais (« Papa en Marloes staan op het station. Ze wachten op de trein »), prononcées par les mêmes locuteurs que ceux du corpus B génère le corpus C.

Cinq juges ont participé à l'évaluation perceptive des stimuli. Les cinq scores des grades par stimuli ont été moyennés. On dispose donc, de la valeur moyenne du grade (G) de chaque fichier des deux corpus B et C.

### III.2.2 Dispositifs utilisés :

La simulation des programmes et les tests des différents indices acoustiques mentionnés précédemment sont effectués sur un PC Intel Core™ 2 Duo, E7500 @ 2.93 GHz, 4Go da RAM.

Les logiciels de simulation sont : MATLAB R2014a, version 8.3.0.532 sous Windows8 64bit et le logiciel PRAAT qui est un exemple de logiciels qui sert à l'extraction d'indices acoustiques sur des signaux vocaux. Il permet de réaliser de nombreuses mesures acoustiques comme les perturbations de la fréquence et de l'amplitude les plus courantes en pathologie vocale. Ainsi, le logiciel PRAAT est utilisé dans la suite de ce travail pour la comparaison [29].

## III.3 Simulations et résultats :

### III.3.1 Représentation temporelle et fréquentielle :

Au début, on a utilisé deux fichiers du corpus A avec une fréquence fondamentale de 100Hz. Le premier a été assigné un score moyen du grade le plus faible par les juges (petite variation du jitter, sans bruit) et le deuxième un score moyen du grade le plus élevé (grande variation du jitter, niveau élevé du bruit).

La figure suivante illustre les formes d'ondes et les spectrogrammes des deux signaux utilisés. On observe l'influence du bruit sur toutes les fréquences du spectre par rapport au signal normal et celle du jitter sur la périodicité de la forme d'onde.

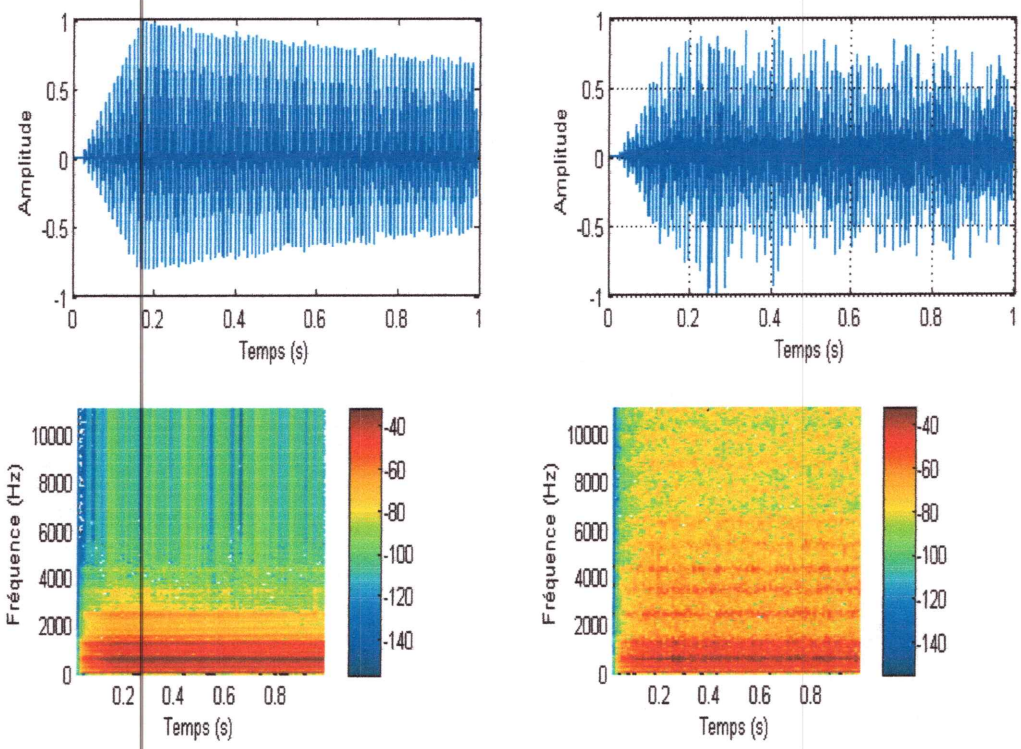


**Figure III.1** : Forme d'ondes et spectrogrammes d'une voyelle [a] synthétique normale (à gauche) et dysphonique (à droite).

### III.3.2 Résultats de simulation des indices acoustiques sur des fichiers :

Au début, nous avons élaboré un programme Matlab qui calcule les différentes valeurs du pitch dans chaque segment de 10 ms pour chaque fichier de la parole en cours de traitement. La méthode cepstrale de détermination du pitch a été utilisée où nous avons fixé la fréquence maximale du pitch à 500 Hz (2ms) et la fréquence minimale du pitch à 50 Hz (20ms). Ce programme fait appelle à deux fonctions. La première est réservée au calcul des différentes valeurs de la fréquence fondamentale et les différentes représentations du jitter comme :

- La valeur moyenne de la fréquence fondamentale.
- La valeur maximale, minimale et l'écart-type de la fréquence fondamentale.
- La gamme de fréquence phonatoire (PFR).
- Le jitter absolu moyen (MAJ).
- Le rapport du jitter (Jitter Ratio).
- Perturbation moyenne relative (RAP).
- Quotient de perturbation de pitch (PPQ).

L'autre fonction est consacrée au calcul des différentes mesures du shimmer comme :

- L'amplitude moyenne.
- L'amplitude maximale et minimale détectée.
- L'écart-type du contour de l'amplitude.
- Le shimmer absolu moyen.
- Le shimmer (%).
- Le shimmer (dB).
- Perturbation relative moyenne d'amplitude (ARP).
- Quotient de perturbation de l'amplitude (APQ).

Ensuite, nous avons implémenté sous Matlab l'algorithme de De Krom [21] du calcul du HNR présenté dans l'organigramme de la figure (II.8).

Nous rappelons que les seuils normal/pathologique selon le logiciel PRAAT sont donnés par le tableau suivant :

| Jiiter | | | |
|---|---|---|---|
| Jitter local | MAJ | RAP | PPQ |
| 1.04 % | 83.200 us | 0.68 % | 0.84 % |
| Shimmer | | | |
| Shimmer local | Shimmer (dB) | APQ | |
| 3.81% | 0.35 dB | 3.070 % | |

**Tableau III.2 :** Seuil normal/pathologique selon le logiciel PRAAT.

L'application des différents programmes Matlab des fonctions citées sur les deux fichiers sélectionnés avant donne les résultats présentés dans le tableau suivant, où nous avons présentés aussi les résultats obtenus avec le logiciel PRAAT à titre de comparaison.

Nous remarquons une très grande ressemblance entre les différents indices obtenus par les deux logiciels (tableau III.3). Pour le premier fichier de grade moyen G = 0.4444 (son normal), on remarque que toutes les valeurs obtenues des différentes représentations du jitter et du

shimmer sont en dessous du seuil de pathologie. Par contre, pour le deuxième fichier de grade moyen G = 2.6667 (son pathologique), les valeurs obtenues sont au-dessus du seuil de pathologie. Par exemple, le jitter = 3.37 % > 1.04 %, le shimmer = 19.37 % > 3.81 % et le shimmer (dB) = 1.71 dB > 0.35 dB,...

On remarque aussi que pour un son normal, la valeur du HNR est élevée (21.18 dB, pour le fichier 1), correspondant à un grade moyen faible d'une voyelle synthétique normale. Mais, pour un son pathologique, la valeur du HNR décroit avec l'augmentation du grade moyen (HNR = 3.24dB et G = 2.6667, pour le fichier 2), correspondant à une voyelle synthétique dysphonique.

| Fichiers | | Jitter | | | | Shimmer | | | | HNR (dB) | Grade moyen |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Jitter (%) | MAJ (s) | RAP (%) | PPQ 5 (%) | SHIM (%) | SHIM (dB) | APQ3 (%) | APQ5 (%) | | |
| Fichier 1 | Praat | 0.43 | 42.93E-6 | 0.26 | 0.27 | 2.94 | 0.33 | 1.17 | 1.25 | 21.18 | 0.4444 |
| | Matlab | 0.5 | 50.55E-6 | 0.29 | 0.35 | 1.72 | 0.23 | 1.12 | 1.9 | 38.73 | |
| Fichier 2 | Praat | **3.37** | **336.66E-6** | **2.01** | **2.19** | **19.37** | **1.71** | **11.79** | **12.48** | **3.24** | 2.6667 |
| | Matlab | **7.33** | **725.15E-6** | **4.81** | **6.02** | **3.83** | **0.43** | 2.53 | **4.42** | 12.20 | |

**Tableau III.3 :** Valeurs des indices acoustiques de deux fichiers de la voyelle /a/ synthétique.

On remarque aussi que malgré les petites variations entre les différentes valeurs des indices acoustiques du jitter et du shimmer obtenues avec nos simulations sous Matlab et celles du PRAAT, à cause de la différence des algorithmes de détermination du pitch pour le jitter et le choix de la valeur maximale de l'amplitude du signal utilisée pour le calcul du shimmer soit

directement de la forme d'onde ou à partir des amplitudes aux instants d'ouverture de la glotte. Les valeurs des indices du fichier (1) sont toutes en dessous des seuils de pathologie de chaque indice, confirmant que le fichier (1) est celui d'une voyelle normale. Par contre, celles du fichier (2) sont toutes au-dessus des seuils de pathologie, car le fichier (2) est celui d'une voyelle pathologique. Alors, la discrimination normale/pathologique est vérifiée dans les deux cas.

Une autre remarque importante, la valeur de l'HNR du premier fichier est considérable et plus proche au niveau de la norme (environ 40dB pour un jeun homme en bonne santé), par contre, celle du deuxième fichier est plus petite, soit de 12.20 dB sous Matlab ou 3.24 dB sous PRAAT, dans les deux cas elle reste loin (au-dessous) au niveau d'énergie de la parole normale.

### III.3.3 Résultats de simulation des indices acoustiques sur des bases de données :

Nous avons utilisé le corpus A décrit au paragraphe III.2.1. Nous rappelons que ce corpus est composé de 16 voyelles synthétiques formées par la combinaison de quatre valeurs de l'amplitude *b* du jitter et de quatre valeurs de l'amplitude *n* du bruit additif. Trois fréquences différentes sont utilisées (100Hz, 120Hz et 140Hz), ce qui donne les 48 fichiers du corpus A.

Les valeurs mesurées du jitter local (%), du HNR (dB) et les scores moyens des grades (Gmoy) de tous les juges sont présentés dans les tableaux III.4, III.5 et III.6.

| b | j | 0.05 | 0.15 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| 0 | Jitter local (%) | 0.429 | 1.082 | 2.164 | 2.855 |
| | HNR (dB) | 21.185 | 15.384 | 8.228 | 6.452 |
| | Gmoy | 0.44 | 0.778 | 2.11 | 2.22 |
| 0.1 | Jitter local (%) | 0.396 | 1.078 | 1.974 | 2.829 |
| | HNR (dB) | 21.655 | 14.933 | 9.327 | 6.692 |
| | Gmoy | 0.89 | 0.89 | 2.00 | 2.11 |

| | | | | | |
|---|---|---|---|---|---|
| 0.3 | Jitter local (%) | 0.481 | 1.223 | 2.275 | 2.991 |
| | HNR (dB) | 15.222 | 11.449 | 6.989 | 5.144 |
| | Gmoy | 1.89 | 1.89 | 2.33 | 2.33 |
| 0.6 | Jitter local (%) | 0.845 | 1.230 | 2.674 | 3.370 |
| | HNR (dB) | 8.011 | 7.194 | 3.966 | 3.238 |
| | Gmoy | 2.67 | 2.56 | 2.56 | 2.67 |

**Tableau III.4 :** Mesures par logiciel PRAAT du jitter local et HNR (Corpus A, F=100Hz).

| b | j | 0.05 | 0.15 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| 0 | Jitter local (%) | 0.294 | 0.931 | 1.644 | 2.754 |
| | HNR (dB) | 23.440 | 13.834 | 8.194 | 4.149 |
| | Gmoy | 0 | 0.33 | 1.22 | 1.78 |
| 0.1 | Jitter local (%) | 0.322 | 0.802 | 1.891 | 2.073 |
| | HNR (dB) | 21.006 | 14.367 | 7.228 | 5.000 |
| | Gmoy | 0.56 | 0.56 | 1.56 | 2.33 |
| 0.3 | Jitter local (%) | 0.343 | 0.888 | 1.916 | 2.092 |
| | HNR (dB) | 15.783 | 12.024 | 6.994 | 5.189 |
| | Gmoy | 1.33 | 1.67 | 2.00 | 2.22 |
| 0.6 | Jitter local (%) | 0.510 | 1.166 | 2.164 | 1.974 |
| | HNR (dB) | 10.466 | 8.504 | 4.745 | 4.703 |
| | Gmoy | 2.22 | 2.44 | 2.67 | 2.44 |

**Tableau III.5 :** Mesures par logiciel PRAAT de jitter local et HNR (F = 120 Hz).

| b | j | 0.05 | 0.15 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| 0 | Jitter local (%) | 0.325 | 0.938 | 1.904 | 2.594 |
| | HNR (dB) | 23.042 | 14.211 | 7.549 | 5.687 |
| | Gmoy | 0 | 0.44 | 1.44 | 1.78 |
| 0.1 | Jitter local (%) | 0.335 | 1.032 | 2.187 | 2.258 |
| | HNR (dB) | 21.563 | 13.655 | 6.876 | 6.666 |
| | Gmoy | 0.67 | 0.89 | 1.22 | 1.67 |
| 0.3 | Jitter local (%) | 0.407 | 1.006 | 1.905 | 2.590 |
| | HNR (dB) | 16.025 | 12.623 | 7.839 | 4.838 |
| | Gmoy | 1.44 | 1.22 | 1.56 | 2.33 |
| 0.6 | Jitter local (%) | 0.617 | 1.143 | 1.791 | 2.320 |
| | HNR (dB) | 10.980 | 9.317 | 6.518 | 5.138 |
| | Gmoy | 2.22 | 2.11 | 2.00 | 2. 22 |

**Tableau III.6 :** Mesures par logiciel PRAAT du jitter local et HNR (F = 140 Hz).

Pour un niveau de bruit nul, nous observons que plus le paramètre *b* est important, plus la raucité perçue par les juges est importante (Gmoy augmente) et la valeur du HNR diminue pour les trois fréquences. Plus le niveau de bruit (b) est important, la valeur du jitter local augmente d'avantage et celle du HNR diminue encore plus. Ces tableaux montrent, en outre, une correspondance entre les valeurs mesurées de tous les indices acoustiques et les valeurs du Gmoy.

Nous observons aussi, que lorsque le bruit additif est élevé (b=0.6), les valeurs du grade moyen des juges sont presque identiques quel que soit la valeur du coefficient j, pour les trois tableaux. Les voyelles sont noyées dans le bruit.

Pour des quantités du bruit additif plus faibles, les variations des valeurs du jitter et du HNR sont remarquables.

### III.3.4 Corrélation du HNR avec les scores des juges des corpus :

Comme dernier essai, nous avons calculé le degré de corrélation entre l'indice objectif (le HNR) et les mesures subjectives données par le grade moyen des juges des trois corpus utilisés (A, B et C). Le tableau suivant donne les valeurs du coefficient de corrélation de Pearson des trois corpus.

| Corpus | Corpus A | Corpus B | Corpus C |
|---|---|---|---|
| Coefficient de Pearson | -0.8576 | -0.6613 | -0.6939 |

**Tableau III. 7 :** Coefficients de corrélation entre le Grade et le HNR des trois corpus.

Nous observons que la corrélation est acceptable pour les trois corpus et que celle des voyelles synthétiques (Corpus A) est nettement plus grande que celles des corpus de la parole naturelle (corpus B et C).

## III.4 Conclusion :

Dans ce chapitre, la simulation montre que La parole normale naturelle ou synthétique, d'une part, présente des valeurs des indices acoustiques, modérées et au-dessous du seuil de la pathologie, et affirme aussi que le niveau du HNR est considérable et au-dessus du seuil. De l'autre part, elle expose que la parole pathologique (ou bruitée) a des valeurs du jitter et shimmer élevées et supérieures au seuil pathologique, et que HNR aussi reste très petit, ce qui rendre la parole moins compréhensible.

Les valeurs du jitter et du shimmer augmentent avec le dégrée de la pathologie (du grade, Gmoy), par contre le HNR diminue avec le niveau de coefficient du bruit ou du jitter, ainsi qu'il diminue avec le grade.

Les mesures objectives (jitter, shimmer et HNR) présentent une bonne corrélation avec l'évaluation subjective (Grade), chaque une complémenter l'autre.

# Conclusion générale :

Dans ce travail, nous avons présenté une analyse et une évaluation des dysphonies des voix pathologiques synthétiques ou naturelles, avec des mesures subjectives et objectives sur des fichiers de parole et des bases de données.

Nous avons remarqué que pour un sujet pathologique, plusieurs troubles vocaux sont souvent dus aux pathologies du larynx. Les changements morphologiques de l'anatomie du larynx engendrent des dysphonies d'origines morphologiques, par contre un mauvais contrôle de la respiration, une atteinte neurologique ou une difficulté psychologique engendrent des dysphonies d'origines neurologiques.

Les variations de la fréquence fondamentale cycle-à-cycle, les variations de l'amplitude cycle-à-cycle et le rapport harmoniques sur bruit forment l'ensemble des indices acoustiques de l'évaluation objective.

Afin d'évaluer et d'analyser ces pathologies, pour l'évaluation subjective nous avons utilisé le grade G de l'échelle GRBASI, et pour l'évaluation objective les mesures du Jitter, Shimmer et HNR sont utilisées.

Les différentes mesures subjectives et objectives assurent la discrimination normal / pathologique des signaux utilisés lors des essais sur des fichiers seuls ou sur des bases de données. De plus, nous avons remarqué que les valeurs du jitter et du shimmer augmentent avec le dégrée de la pathologie et en corrélation avec le grade (Gmoy), par contre le HNR diminue avec les variations du niveau de bruit ou les fluctuations cycle-à-cycle. Enfin, une bonne corrélation est obtenue entre le HNR et le grade pour les trois corpus de parole utilisés.

# *Bibliographie et Webographie*


[1] http://tpe-son-jvc.e-monsite.com/pages/emission-du-son/i-b-role-des-poumons.html

[2] http://www.linguistes.com/phonetique/phon.html.

[3] R. Boite, H. Bourlard, T. Dutoit, J. Hancq et H. Leich, *Traitement de la parole*, Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse, 488p, 1999.

[4] O. Calliope, *La parole et son traitement automatique*, Collection Masson, CNET-ENST, Paris, France, pp. 410-414, 1989.

[5] D. H. KLATT, "Software for cascade parallel formant synthesizer," JASA, pp. 971-995, 1980.

[6] H. Gray, *Anatomy of the human body*, Warren H. Lewis, Bartleby, New York, May 2000.

[7] http://www.docteurclic.com/dictionnaire-medical/laryngite-de-l-adulte.aspx.

[8] http://fr.slideshare.net/DrHSamir/polypes-des-cordes-vocales.

[9] Charles FRECHE et al., "La voix : La corde vocale et sa pathologie," Catherine Journiac, Collège International de Médecine et Chirurgie de l'hôpital américain de Paris, France, pp. 34, 2001.

[10] Inserm, *La voix : Ses troubles chez les enseignants : Expertise collective*, Institut national de la santé et de la recherche médicale, Paris, France, pp. 181-186, 2006.

[11] C. Busseuil et C. Chauvy, "Esthétique des voix dysphoniques : une approche perceptuelle," Doctorat en médecine : Orthophonie, Université de Montpellier 1, Montpellier, France, 2011.

[12] M. Hirano, "Head and neck surgery," in Objective Evaluation of the Human Voice: Clinical Aspects, Kurume University School of Medicine, Kurume, Japan, pp. 89-144, 1989.

[13] I. V. Bele, "Reliability in perceptual analysis of voice quality," Journal of Voice, vol. 19, n°4, pp. 555-573, 2004.

[14] R. J. Baken, R. F. Orlikoff, *Clinical measurement of speech and voice*, Singular Publishing Group, San Diego, USA, 2000.

[15] A. Ghio et al., "Approches complémentaires pour l'évaluation des dysphonies: bilan méthodologique et perspectives", Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), France, vol. 26, pp. 33-74, 2007.

[16] www. cairn.info

[17] J. Schoentgen, "Stochastic models of jitter," J. Acoust, Soc, Am, pp. 1631-1650, 2001.

[18] www.praat.org.

[19] Roslyn J. Morgan, Richard B.Reilly, "Telephony-Based voice pathology assessment using automated speech analysis," IEEE Transaction on biomedical engineering, vol. 53, n°3, pp. 471-472, 2006.

[20] Yomoto, E. Baer T. et Gauld W. J., "Harmonic-to-noise-ratio as an index of the degree of hoarseness", journal of the acoustical society of America, vol. 71, pp. 1544-1550, 1982.

[21] G. de Krom, "A cepstrum based technique for determining an harmonics-to-noise ratio in speech signals," J. Speech Hear. Res. Vol. 36 n° 2, pp. 254-266, 1993.

[22] http://web.mit.edu/.

[23] Rashmi Makhijani et al., "Speech enhancement using pitch detection approach for noisy environment," International journal of engineering science and technology (IJEST), Vol. 3 N°.2, pp. 1766, 2011.

[24] B.P. Bogert et al. "The quefrency analysis of Time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and shape cracking," Time series analysis, M. Rosenblatt, Wiley, NY, USA, pp. 209-243, 1963.

[25] Brioua Fathi, "Estimation du pitch d'un signal de parole," Projet de fin d'Etudes pour l'obtention du Diplôme de Master II en Optoélectronique, Université de Mohammed Sadik Ben Yahia, Jijel, Algérie, pp. 24-25, 2011.

[26] David Gerhard, "Pitch extraction and fundamental frequency: History and current techniques," Technical Report TR-CS, pp. 12, 2003, University of Regina, Saskatchewan, CANADA,

[27] W. Becker, H. H. Naumann and C. R. Faltz, *Ear nose and throat diseases*, New York, USA, 1994.

[28] Laghmizi Sabrina, "Analyse de signal de la parole pour l'évaluation automatique des voix pathologiques, " Mémoire de Master en physique, Université de Jijel, Algérie, 2013.

[29] http://www.fon.hum.uva.nl/praat.