

République Algérienne Démocratique et populaire
Ministère de l'Enseignement supérieur et de la recherche Scientifique



Université Mohammed Seddik BENYAHIA – Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique

MEMOIRE DE FIN D'ETUDES

Présenté en vue de l'obtention du diplôme
De **MASTER** en
Systèmes des télécommunications
Thème :

Analyse multivariée de la parole pour l'évaluation objective de
la qualité vocale

Encadré par :

 Pr. Abdellah KACHA

Réalisé par :

 M^{elle} Amel FENIZA
 M^{elle} Hassiba LAISSANI

Promotion : Juillet 2019



REMERCIEMENTS

Louange à Allah, de nous avoir donné la santé, la volonté et la patience pour réaliser et terminer ce travail.

Nous remercions nos très chers Parents qui nous ont apporté la force et l'aide tout au long de notre parcours universitaire en nous fournissant nos besoins afin d'étudier dans les meilleures conditions.

Tout notre respect, gratitude et reconnaissance à notre encadreur **Pr. Abdellah KACHA** qui nous a fourni le sujet de ce mémoire et nous a guidés de ses précieux conseils et suggestions, et la confiance qu'il nous a témoignés tout au long de ce travail.

Nous tenons à remercier aussi les membres de jury pour l'intérêt qu'ils ont porté à notre projet en acceptant d'examiner notre travail.

Nous adressons aussi nos remerciements à **Mr. Toufik BENKEDIDAH** chef de département d'électronique et à tous les enseignants de la filière électronique.

Enfin, nous adressons nos sincères sentiments de gratitude à toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Merci



DEDICACE

Nous dédions ce mémoire à

Nos chers parents

Nos frères et sœurs

Nos familles et nos amis

Avec amour **Amel** et **Hassiba**



TABLE DES MATIERES

Remerciements.....	i
Dédicace.....	ii
Table des matières	iii
Liste des abréviations	vii
Liste des figures	viii
Liste des tableaux.....	x
Liste des algorithmes.....	xi
Introduction Générale.....	01

Chapitre I : Production de la parole

I.1. Introduction.....	4
I.2. Production de la parole.....	4
I.2.1 Anatomie de l'appareil de production de la parole.....	4
I.2.1.1 l'Appareil phonatoire.....	4
I.2.1.2 Le système articulatoire.....	7
I.2.2. Composantes physiologiques du système de production de la parole.....	8
I.3. Catégories des sons de la parole.....	9
I.3.1. Sons voisés.....	10
I.3.2. Sons non voisés.....	10
I.4. Modèle source-filtre de production de la parole.....	11
I.5. Propriétés et caractéristiques du signal de la parole.....	13
I.5.1. Propriétés du signal de la parole.....	13
I.5.2. Caractéristiques du signal de la parole.....	13

I.5.2.1. Formes d'onde de la parole.....	13
I.5.2.2. Les phonèmes.....	14
I.5.2.3. La fréquence fondamentale.....	14
I.5.2.4. L'énergie.....	15
I.5.2.5. Le spectre de fréquence.....	15
I.5.2.6. Les formants.....	15
I.6. Troubles de la voix.....	16
I.6.1. Troubles fonctionnelles.....	16
I.6.2. Troubles organiques.....	18
I.7. Evaluation des troubles de la voix	18
I.7.1. Evaluation subjective.....	19
I.7.2. Evaluation objective.....	20
I.7.2.1. Evaluation des troubles de la voix via le signal de la parole.....	20
I.7.2.2. Evaluation basée sur l'estimation de la source glottique.....	20
I.8. Conclusion	21

Chapitre II : Décomposition modale empirique

II.1. Introduction.....	22
II.2. Décomposition modale empirique.....	22
II.2.1. Principe de la décomposition modale empirique.....	23
II.2.2. Processus de tamisage : principe algorithmique.....	24
II.2.3. Critère d'arrêt du processus de tamisage.....	28
II.3. Propriétés fondamentales de la décomposition modale empirique.....	29
II.3.1. Non-linéarité.....	29
II.3.2. Localité.....	29
II.3.3. Aspect multi-résolution.....	29
II.3.4. Auto-adaptativité.....	29
II.3.5. Quasi-orthogonalité.....	30
II.4. Conclusion.....	30

Chapitre III : Estimation des dyspériodicités vocales

III.1. Introduction.....	31
III.2. Méthodes d'estimation des dyspériodicités vocales.....	31
III.2.1. Analyse prédictive bidirectionnelle à long terme.....	31
III.2.1.1. Méthode d'analyse.....	31
III.2.1.2. Indice acoustique	32
III.2.1.3. Propriétés de l'erreur de prédiction à long terme comme une mesure de dyspériodicités.....	33
III.2.2. Analyse basée sur le variogramme généralisé.....	33
III.2.3. Analyse par décomposition modale empirique du spectre d'amplitude.....	36
III.2.3.1. Séparation des composantes du signal de la parole.....	36
III.2.3.2. Indice acoustique.....	38
III.3. Conclusion.....	39

Chapitre IV : Analyse multivariée par décomposition modale empirique

IV.1 Introduction.....	40
IV.2 Base de données.....	40
IV.3 Méthode d'analyse.....	40
IV.4 Exemple d'illustration.....	42
IV.5 Résultats et discussions	46
IV.6 Conclusion	51
Conclusion Générale	52
Références Bibliographiques.....	53



LISTE DES ABRÉVIATIONS

DFT	Discrete Fourier Transform
GRBAS	Segmental Excitation to Dysperiodicity Ratio
TAP	Traitement Automatique de la Parole
EMD	Empirical Mode Decomposition
AM-FM	Amplitude Modulation-Frequency Modulation
IMFs	Intrinsic Mode Functions
RSD	Rapport Signal à Dyspériodicité
RHB	Rapport Harmonique à Bruit

LISTE DES FIGURES

Figure I.1- Anatomie de l'appareil phonatoire.....	5
Figure I.2- Anatomie du larynx.....	5
Figure I.3- Anatomie de l'appareil articulatoire.....	7
Figure I.4- Les trois principales composantes du système de production de la parole.....	8
Figure I.5- Schéma récapitulatif du fonctionnement de l'appareil phonatoire.....	9
Figure I.6- Exemples de son voisé et son non voisé.....	10
Figure I.7- Modèle général de production de la parole.....	11
Figure I.8- Modèle source-filtre de production de la parole.....	12
Figure I.9- Les différentes classes des phonèmes.....	14
Figure I.10- Divers types de nodules.....	16
Figure I.11- Différentes types de polypes.....	17
Figure I.12- Laryngite herpétique.....	17
Figure I.13- Kyste épidermique.....	18
.	
Figure II.1- Décomposition du signal en une composante lente et une autre rapide.....	23
Figure II.2- Organigramme de l'algorithme de l'EMD.....	27
Figure IV.1- Signal de parole correspondant à une voyelle [a] produite par un locuteur normophonique.....	47
Figure IV.2- Trame courante de 5 ms extraite d'une voyelle [a] produite par un locuteur dysphonique qui a été assigné un score moyen de 1, la trame décalée, leurs spectres d'amplitudes respectifs ainsi que les quatre premières IMFS correspondantes.....	48
Figure IV.3- Boîte à moustaches des rapports signal à dyspériodicité locaux des quatre premières IMF pour un locuteur normophonique et un locuteur dysphonique.....	52

Figure IV.4- *Rapports signal à dyspériodicité en fonction es scores moyens pour les quatre premières IMF. La longueur de la trame est fixée à 7.5 ms.....54*

Figure IV.5- *Scores prédits en utilisant les rapports signal à dyspériodicité des quatre premières IMFs en fonction des scores moyens de l'évaluation auditive pour une longueur de trame de 7.5 ms.....56*



LISTE DES TABLEAUX

Tableau I.1- <i>Définition et fonction des cartilages du larynx.....</i>	6
Tableau III.1- <i>Catégorisation des modes oscillants en fonction de leurs fréquences moyennes.....</i>	38
Tableau IV.1- <i>Valeurs du coefficient de corrélation entre le rapport signal à dyspériodicité et les scores moyens de l'évaluation auditive des quatre premières IMFs pour différentes valeurs de la longueur de la trame d'analyse.....</i>	48
Tableau IV.2- <i>Valeurs du coefficient de corrélation entre les rapports signal à dyspériodicité obtenus avec le variogramme et les scores moyens de l'évaluation auditive pour différentes valeurs de la longueur de la trame d'analyse.....</i>	48



LISTE DES ALGORITHMES

Algorithme II.1- <i>Les différentes étapes de l'algorithme EMD.....</i>	25
Algorithme III.1- <i>Estimation des dyspériodicités vocales par le variogramme généralisé...35</i>	

Introduction générale

Introduction générale

La voix est l'ensemble des sons produits sous l'influence de l'air des poumons qui passe dans le larynx et fait vibrer les cordes vocales. La parole est la modification de ces sons en fonction de la forme du conduit vocale. Le système de production de la voix contient trois parties : les poumons ont le rôle de générer le flux d'air, le larynx transforme ce flux d'air en énergie acoustiques et les cavités supra-laryngées pour amplifier l'énergie acoustique.

La modification anormale de la voix correspond aux troubles de la voix qui sont des conséquences des pathologies organiques ou fonctionnelles des cordes vocales. Les maladies fonctionnelles comprennent tous les troubles qui peuvent aboutir à un changement du système vocal suite à une mauvaise utilisation de la voix. Ces maladies peuvent être traitées par une thérapie vocale. Une pathologie organique est causée par une maladie structurelle (anatomique ou physiologique), soit une maladie du larynx lui-même, soit par une maladie systémique distante qui modifie la structure ou la fonction du larynx. Les maladies laryngales peuvent être traitées via une thérapie de la voix ou par une intervention chirurgicale. Pour cette raison, une méthode robuste qui permet le diagnostic et l'évaluation du degré de sévérité de la pathologie est d'une grande importance pour le suivi de l'évolution de l'état du patient.

L'évaluation clinique de la voix est régulièrement basée sur la perception de la parole par l'auditeur. Par exemple, les cliniciens évaluent sur une échelle allant de 0 à 3 le degré d'anomalie globale perçue appelée grade pour suivre l'état de la voix des patients. Cette méthode d'évaluation est subjective sans le sens que le résultat d'évaluation dépend de l'auditeur. Un inconvénient majeur de l'évaluation subjective est la variabilité intra- et inter-juges [1, 2]. Les expériences ont montré que pour obtenir des évaluations reproductibles, les auditeurs (juges) doivent avoir une expérience substantielle dans l'évaluation du timbre de la voix.

Contrairement aux mesures subjectives, les mesures objectives sont obtenues à partir de l'analyse acoustique de la parole. Les mesures objectives sont d'une grande importance pour

l'évaluation clinique de la voix parce que l'analyse est non invasive et fournit un indice de perturbation de la voix qui permet aux cliniciens de suivre l'évolution des patients et documenter quantitativement le degré d'enrouement perçu. En dépit du nombre d'indices acoustiques proposés dans la littérature pour caractériser la parole des locuteurs dysphoniques, trouver des indices fiables et précis reste un problème posé.

Bien qu'il existe diverses conditions médicales qui peuvent affecter la voix, la plupart des troubles de la voix proviennent du système vocal et entraînent une augmentation de la dyspériodicité des sons voisés de la parole. Les dyspériodicités peuvent être causées par le bruit additif dû à la turbulence et par le bruit de modulation dû aux perturbations externes du signal d'excitation glottique ainsi que les dyspériodicités dues à la dynamique irrégulière des cordes vocales. Comme conséquence de ces dyspériodicités, l'énergie de la structure harmonique du spectre diminue en faveur de la structure non harmonique. Plusieurs indices acoustiques utilisés pour évaluer la fonction vocale reflètent la déviation de la forme d'onde de la parole voisée par rapport à la périodicité parfaite. Parmi les méthodes proposées récemment, la prédiction linéaire à long terme (LTP) [3] et le variogramme généralisé [4] se sont révélés efficaces pour estimer les dyspériodicités vocales à partir de la parole. La décomposition modale empirique (EMD : empirical mode decomposition) [5] s'est récemment révélée utile pour l'estimation des dyspériodicités vocales [6].

L'objectif de la présente étude est de proposer une méthode d'analyse multivariée des dyspériodicités vocales qui utilise le principe du variogramme généralisé appliqué aux fonctions de modes intrinsèques du spectre du signal de la parole obtenus par décomposition en modes empiriques. La méthode d'analyse multivariée est testée sur un corpus comprenant des voyelles [a] produites par des locuteurs normophoniques et dysphoniques présentant différents degrés de dysphonie.

Le mémoire est organisé comme suite :

Le premier chapitre est consacré à la description du système de production de la parole ainsi qu'à la présentation de l'anatomie et la physiologie de ce système. Le chapitre termine par la présentation de quelques troubles de la voix et leur évaluation.

Le deuxième chapitre est consacré à la présentation de méthode de décomposition modale empirique qui utilise un processus appelé tamisage (sifting) permettant de décomposer temporellement des signaux non stationnaires et issus de systèmes non linéaires en la somme de signaux mono-composantes.

Le troisième chapitre présente les méthodes d'estimation des dyspériodicités vocales à partir du signal de la parole et les indices acoustiques utilisés pour quantifier le degré d'enrouement de la voix.

Le quatrième chapitre se focalise sur la présentation de la méthode d'analyse multivariée basée sur la décomposition modale empirique du spectre de la parole ainsi que sur les résultats expérimentaux. La méthode d'analyse multivariée est testée sur une base de données comprenant des voyelles [a] produites par des locuteurs normophoniques et dysphoniques avec différents degrés de dysphonie et comparée à une méthode de référence.

Le mémoire se termine par une conclusion générale qui résume les résultats obtenus et les suggestions pour la suite du travail.

Chapitre

I

Production de la parole

I.1 Introduction

La parole est un signal complexe intégrant des informations acoustiques à différentes échelles temporelles [7]. Elle est produite par une série de mouvements des appareils respiratoires et articulatoires. Le processus de production de la parole peut être assimilé à un système dans lequel une ou plusieurs sources excitent un ensemble de cavités. La source est soit une onde de débit quasi-périodique générée au niveau des cordes vocales par vibrations quasi-périodique de ces dernières soit un bruit de friction ou d'explosion généré suite à un fort rétrécissement au niveau du conduit vocal ou par un brusque relâchement d'une occlusion du conduit vocal [8].

Ce chapitre est une description du système de production de la parole. On commence par une présentation de l'anatomie et la physiologie de ce système. Ensuite, on présente les différents types de sons, le modèle inspiré du mécanisme de production de la parole, puis on donne les caractéristiques du signal de la parole. Enfin, nous terminons par la présentation de quelques troubles de la voix et les méthodes d'évaluation de ces troubles.

I.2 Production de la parole

I. 2.1 Anatomie de l'appareil de production de la parole

La production de la parole fait intervenir, outre les composants du système nerveux central et périphérique, divers organes tels que le larynx, le pharynx, les poumons, la langue, les lèvres, les cavités buccale et nasale, etc. Cette section est une description des différents organes intervenant dans ce processus [9].

Le système de production de la parole peut être divisé en deux grands sous-systèmes qui sont l'appareil phonatoire et le système articulatoire.

I.2.1.1 Appareil phonatoire

L'appareil phonatoire (Figure I.1) est constitué d'un ensemble d'organes qui permettent de produire un son de base qu'on appelle la voix.

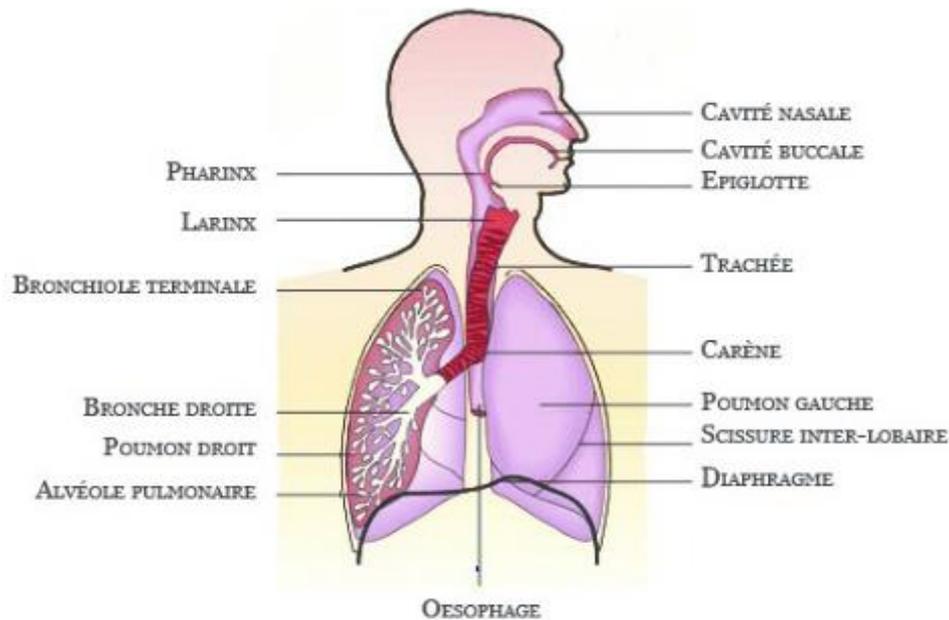


Figure I.1- Anatomie de l'appareil phonatoire.

L'appareil phonatoire comprend :

- ✓ Les poumons, qui fournissent le flux d'air nécessaire à la production des sons de la parole. L'air expirée des poumons monte vers la bouche par un passage commun qui s'appelle la trachée, dite aussi trachée artère [10].
- ✓ Le larynx, est une structure tubulaire située à la partie médiane et antérieure du cou, en arrière et sous la base de langue, surplombant la trachée, circonscrite en arrière et latéralement par l'hypo-pharynx et constituée de l'assemblage de cinq pièces cartilagineuses comme montré par la figure I.2. La définition et la fonction des cartilages dont données dans le tableau I.1

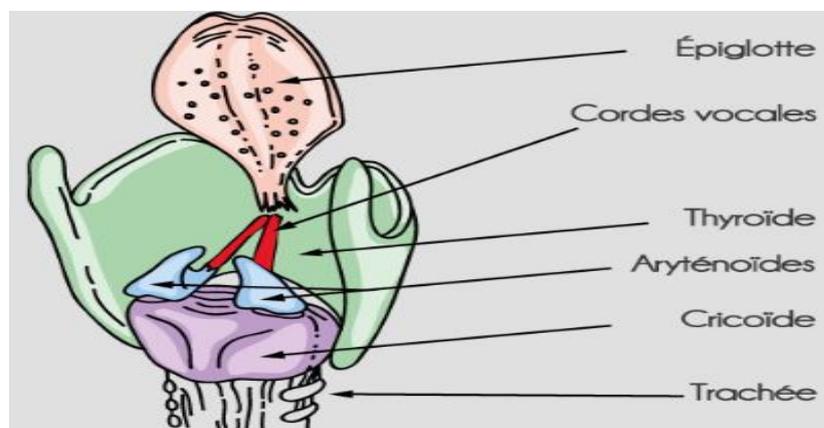


Figure I.2- Anatomie du larynx.

Tableau I.1- Définition et fonction des cartilages du larynx [11].

Cartilages	Définition	Fonction
Cartilage cricoïde	Réalisant un anneau complet situé juste au-dessus du premier anneau trachéal	Assurent la rigidité du larynx qui peut ainsi demeurer « ouvert » pour la fonction respiration
Cartilage thyroïde	Formé de deux lames réunies en avant par un angle dièdre, ouvert vers l'arrière, supporté par le segment antérieur du cartilage cricoïde	
Cartilage épiglottique	S'intégrant au-dessus et dans l'angle antérieur du cartilage thyroïde	Ferment le larynx lors des efforts de déglutition, de phonation et musculaire (Effort à glotte fermée tels que le soulèvement d'une charge
Cartilages aryénoïdes	Deux cartilages pairs (droit et gauche), mobiles et symétriques, placés au-dessus du segment postérieur du cartilage cricoïde fermant l'ouverture postérieure du cartilage thyroïde	

- ✓ Les cordes vocales sont des plis souples au niveau du larynx qui peuvent prendre différents états. Elles peuvent être ouvertes, fermées ou rapprochées. L'espace entre les cordes vocales (qui forme une sorte de triangle) permettant le passage de l'air provenant des poumons est appelé la glotte [9].
- ✓ L'épiglotte est un cartilage situé juste au-dessus de la glotte permettant de couvrir entièrement le passage d'air pulmonaire pour empêcher la nourriture de tomber dans la trachée et les poumons.

L'appareil phonatoire peut avoir deux fonctions différentes : la respiration et la phonation. Lorsque les cordes vocales sont écartées, l'air expiré passe librement à travers la glotte, on parle alors de la respiration. En revanche, si l'on rapproche les cordes vocales, l'expiration de l'air pulmonaire met en mouvement les cordes vocales (phonation) ce qui produit ainsi le son de base appelé aussi voix ou vocalisation.

I.2.1.2 Le système articulatoire

Le système articulatoire permet de modifier le son de base et de produire ainsi une gamme de sons différents (Figure I.3).

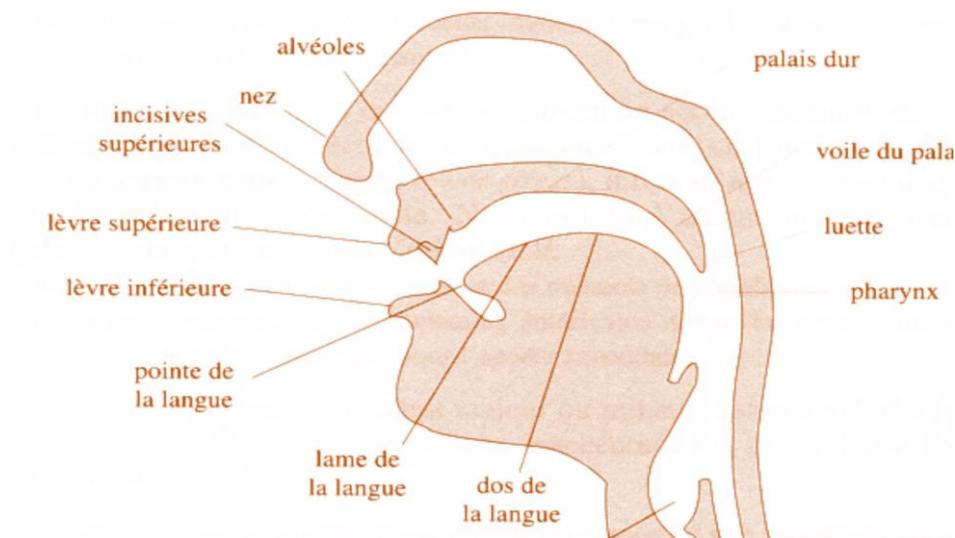


Figure I.3- Anatomie de l'appareil articulatoire.

Ce système est plus complexe et comprend :

- ✓ La mâchoire inférieure, relativement mobile, qui peut prendre différents degrés d'ouverture.
- ✓ Les lèvres, très mobiles peuvent prendre des configurations diverses.
- ✓ La langue, également très mobile, que l'on divise en deux parties principales : l'apex ou la pointe de la langue, et le dorsum, c'est-à-dire le dos ou la partie centrale de la langue.
- ✓ Les dents, qui fournissent des lieux de contact pour la langue et pour les lèvres.
- ✓ Les alvéoles, sorte de renflement dur juste derrière les dents supérieures, fournissent également un lieu de contact pour la langue.

- ✓ Le palais dur, également lieu de contact pour la langue. Le palais dur se divise en deux parties : la partie courbée juste derrière les alvéoles ou partie pré-palatale et la partie dure mais relativement plate ou palatale.
- ✓ Le palais mou, que l'on appelle également le voile du palais, ou velum. Le voile du palais est à la fois un lieu de contact et un organe relativement mobile qui est soit relevé, soit abaissé.
- ✓ L'uvule, que l'on appelle également la luette, est un prolongement fait de la partie élastique du voile du palais [10].

I.2.2 Composantes physiologiques du système de production de la parole

Il est pratique et approprié de diviser le système de production de la parole en trois composantes. Le larynx est une structure complexe qui est utilisé comme point de référence pour cette décomposition. Ces trois principales composantes physiologiques du système de production de la parole sont généralement décrites en référence à la glotte (Figure I.4).

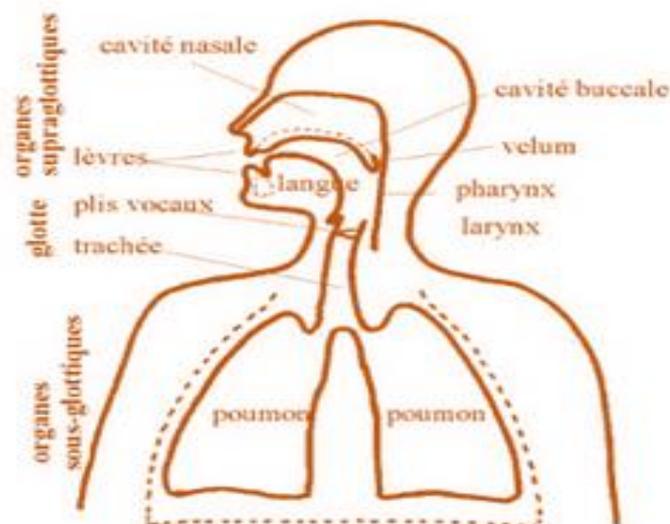


Figure I.4- Les trois principales composantes du système de production de la parole.

Sous le larynx, nous avons les composants sous-glottiques qui sont constitués des poumons et de la musculature respiratoire associée. La trachée ou la trachée-artère descend du larynx et se bifurque dans des voies respiratoires qui aboutissent chacune dans un poumon. Les composants sous-glottiques génèrent le flux d'air qui alimente le système de production de la parole.

Le rôle principal du larynx dans la production de la parole est de convertir un flux d'air relativement constant sortant des poumons en une série de bouffées d'air "quasi-périodiques". Le larynx le fait en fermant et en ouvrant rapidement les voies respiratoires en rapprochant ou en écartant les cordes vocales. Lorsque les cordes vocales se déplacent vers l'intérieur et l'extérieur, la glotte se ferme ou s'ouvre. Une bouffée d'air se produit chaque fois que le larynx s'ouvre brièvement (Figure I.5). Ces bouffées d'air contiennent une énergie acoustique aux fréquences audibles. Le larynx est donc une valve pouvant s'ouvrir et se fermer rapidement, générant une source d'énergie acoustique.

Les voies respiratoires supra-laryngées agissent comme un filtre acoustique variable qui agit sur la source d'énergie acoustique. Il s'agit d'un filtre acoustique variable, car le locuteur modifie la forme de son appareil vocal lorsqu'il parle [12]. Le conduit vocal supra-laryngé joue le rôle d'une caisse de résonance qui agit sur les fréquences émises et amplifie l'intensité sonore. La gamme des fréquences renforcées par le résonateur correspond à un formant. Ces formants caractérisent le timbre de la voix humaine [13].

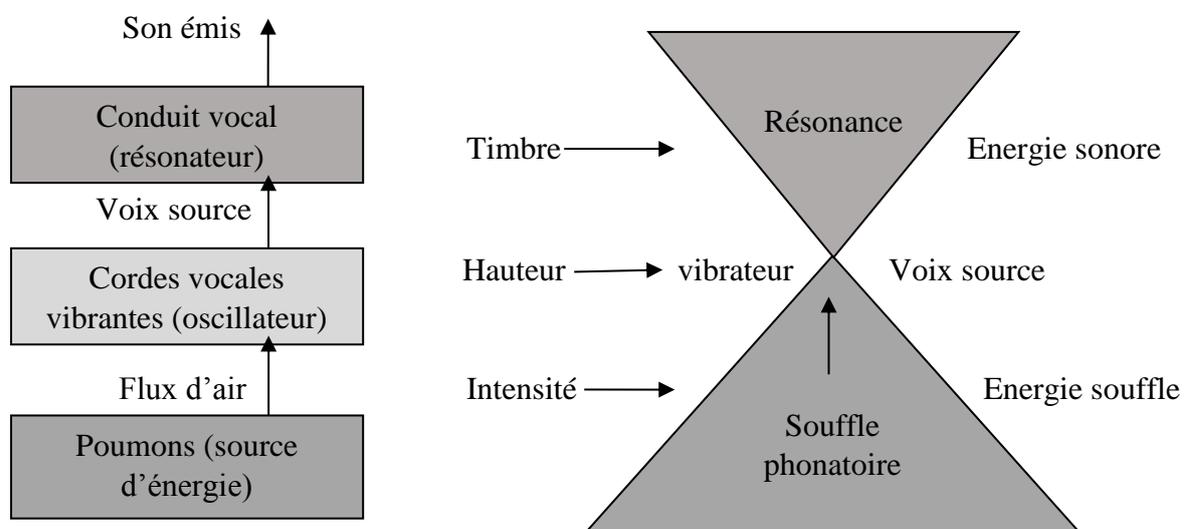


Figure I.5- Schéma récapitulatif du fonctionnement de l'appareil phonatoire.

I.3 Catégories des sons de la parole

Les sons de la parole sont divisés selon l'état des cordes vocales (vibration ou non) en deux catégories : les sons voisés et les sons non voisés.

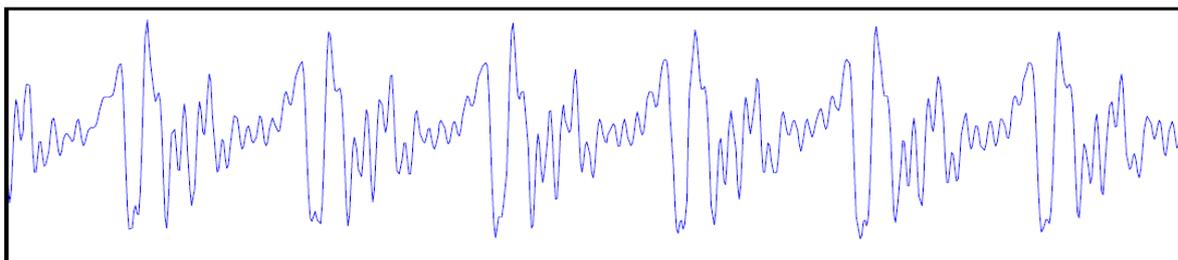
I.3.1 Sons voisés

Les sons voisés sont produits en forçant de l'air à travers la glotte avec la vibration des cordes vocales. Il en résulte des bouffées d'air quasi-périodiques qui excitent le conduit vocal. Les voyelles (/ a /, / i /, / u /, etc.) sont des exemples de sons voisés [14]. Ils ont un spectre de raies harmoniques d'une fondamentale de 100 Hz à 200 Hz pour les hommes et de 200 Hz à 400 Hz pour les femmes. Leur enveloppe spectrale et particulièrement ses maxima (appelés formants) sont spécifiques du son émis [15].

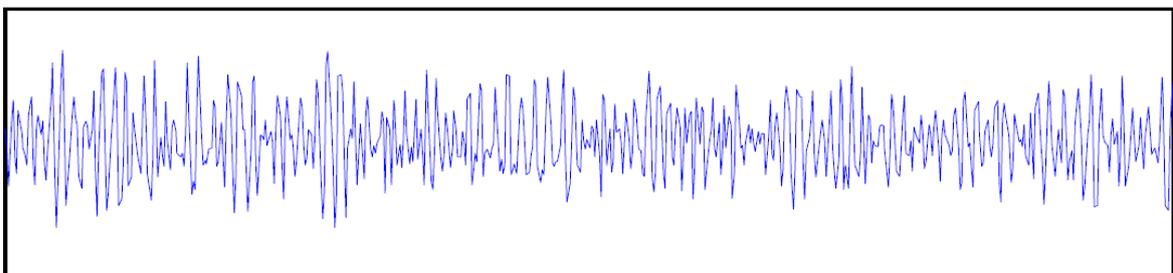
I.3.2 Sons non voisés

Lors de la production des sons non voisés, les cordes vocales ne vibrent pas mais elles peuvent être presque fermées provoquant ainsi des frottements audibles comme pour le son /h/. L'énergie acoustique est produite par la turbulence au niveau d'un ou de plusieurs passages d'air étroits dans la bouche (pointe de la langue contre ou entre les dents, langue contre le palais, etc.). Cette énergie turbulente a un spectre "lisse" comme celui du bruit, sans structure linéaire [16]. Ils sont caractérisés par un spectre continu, stationnaire ou transitoire, mais non uniforme [15].

La figure I.6 montre un exemple d'un son voisé et un son non voisé caractérisé par deux spectres différents.



(a) Son voisé



(b) Son non voisé

Figure I.6- Exemples de son voisé et son non voisé.

I.4 Modèle source-filtre de production de la parole

Le processus de production de la parole peut être représenté par le modèle source-filtre. Ce modèle se compose de trois parties : la source, le filtre et le rayonnement des lèvres comme illustré par la figure I.7. Le modèle source-filtre de production de la parole est représenté sur la figure I.8.

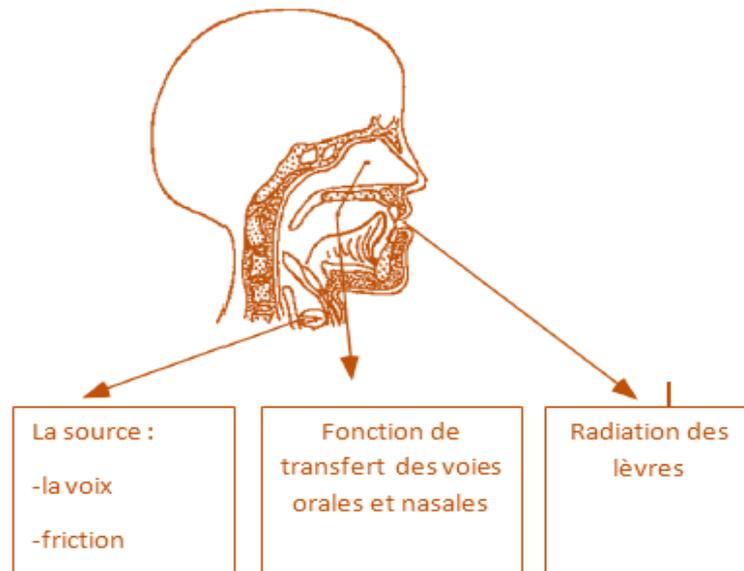


Figure I.7-Modèle général de production de la parole [17].

Pour les voyelles, la source est modélisée par un train d'onde périodique dont la forme, caractérisée par $G(z)$, simule la forme du signal glottique et dont la période est la durée d'un cycle glottique T_0 .

Pour les sons voisés, le signal d'excitation est :

$$U(n) = \sum_k \delta(n - kT_0) \quad (\text{I.1})$$

où T_0 est la durée du cycle vocale.

Pour les sons non-voisés, le signal d'excitation $U(n)$ est un bruit blanc de moyenne nulle et de variance unité.

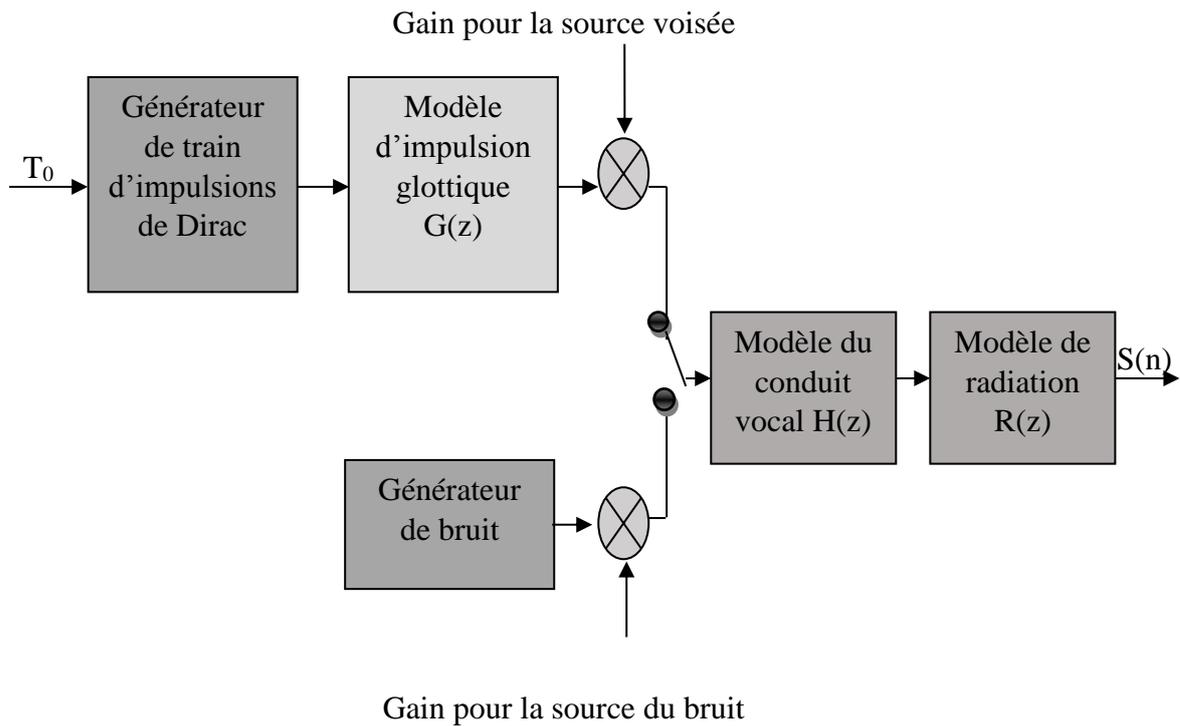


Figure I.8-Modèle source-filtre de production de la parole.

Le conduit vocal est généralement modélisé par un filtre tout pôle. On peut en effet montrer qu'un filtre tout pôle modélise la fonction de transfert d'un système de tubes acoustiques concaténés excité par une onde plane, ce qui est une approximation raisonnable du conduit vocal pour les sons non nasaux.

La fonction du transfert du conduit vocal a la forme :

$$H(z) = \frac{H_0}{1 + \sum_{k=1}^N b_k z^{-k}} = \frac{H_0}{\prod_{k=1}^p (1 - p_k z^{-1})} \quad (\text{I.2})$$

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

Dans le cas de la parole voisée, la source périodique (source glottique) est représentée par un modèle qui dépend de l'application, à condition que ce modèle soit tout pôle.

$$g(n) = (\alpha^n - \beta^n)u(n), \quad \beta < \alpha < 1, \quad \alpha \approx 1 \quad (\text{I.3})$$

où $u(n)$ représente la séquence échelon unité.

Le rayonnement aux lèvres est modélisé par un filtre proche d'un filtre dérivateur :

$$R(z) = 1 - z_0 z^{-1}, z_0 \approx 1, z_0 < 1 \quad (\text{I.4})$$

La valeur de z_0 est choisie légèrement inférieure à 1 afin de garantir la stabilité du filtre inverse $R^{-1}(z)$, qui est rencontré dans certaines applications.

En résumé, la transmittance globale entre le train d'impulsions de la figure I.8 et le signal émis est [18] :

$T(z) = G(z) H(z) R(z)$	pour les sons voisés
$T(z) = H(z) R(z)$	pour les sons non voisés

I.5 Propriétés et caractéristiques du signal de la parole

I.5.1 Propriétés du signal de la parole

Le signal de la parole a les propriétés inhérentes suivantes :

- ✓ On peut remarquer que le signal de la parole est une séquence de sons changeant continuellement.
- ✓ Il dépend fortement des sons produits.
- ✓ Il dépend fortement du contexte dans lequel les sons sont produits.

I.5.2 Caractéristiques du signal de la parole

Le signal de parole présente les caractéristiques suivantes :

I.5.2.1 Formes d'onde de la parole

Les signaux de parole dans le domaine temporel sont appelés formes d'onde de la parole. Les signaux acoustiques ou les sons sont émis sous forme de variations de pression à partir des lèvres tout en articulant des informations linguistiquement significatives.

L'amplitude de la forme d'onde de la parole varie avec le temps de manière complexe, y compris les variations du niveau global ou de l'intensité du son. La fonction de densité de probabilité des amplitudes de forme d'onde, sur une échelle de temps, peut être mesurée en tant que son en décibels [19].

I.5.2.2 Les phonèmes

Les sons qui permettent de distinguer les mots les uns des autres s'appellent phonèmes. Le phonème est la plus petite unité de son capable de produire un changement de sens par simple commutation ou échange de deux unités de même classe en un même point de la chaîne parlée tel que par exemple substituer [b] à [l] à [t] pour obtenir bouche, louche, touche [20]. Le phonème est ainsi caractérisé par quelques attributs articulatoires : ouvert-fermé, oral-nasal, etc., qui décrivent le type d'articulation (fricative, f, s / plosive, b, p / nasale, n, m) et le lieu d'articulation (labial, dental, alvéolaire, palatal, glottal) comme montré par la figure I.9 [21].

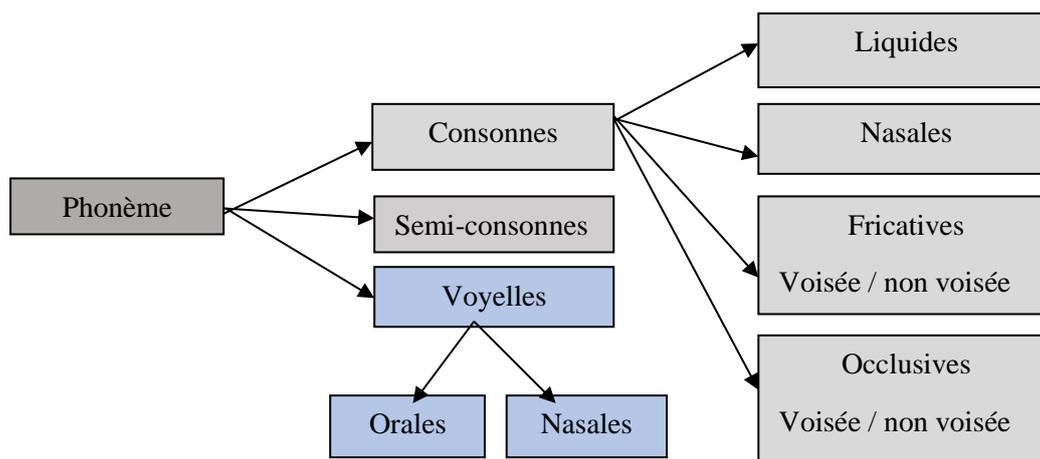


Figure I.9- Les différentes classes des phonèmes.

I.5.2.3 La fréquence fondamentale (Pitch)

La fréquence fondamentale est le nombre de cycles (ouverture-fermeture) produits en une seconde par les cordes vocales pendant la phonation. La fréquence fondamentale (notée F_0) est la fréquence de base (la plus basse) d'un son complexe. En moyenne, la fréquence fondamentale de la voix varie selon l'âge et le sexe [22] :

- ✓ Un homme adulte est de 100 à 120 cycles par seconde ou Hz.
- ✓ Une femme est de 200 à 220 hertz.

- ✓ Une voix aigüe de femme ou d'enfant peut avoir une fréquence fondamentale de 300 à 350Hz.
- ✓ Une voix d'homme très grave peut avoir une fréquence fondamentale de 60 Hz.

I.5.2.4 L'énergie

L'énergie est un paramètre essentiel due à la force du flux pulmonaire mais peut être modifiée par l'ouverture des lèvres créant la résonance. Elle modélise l'intensité avec laquelle le locuteur s'adresse à l'auditeur. L'énergie d'un signal sur une fenêtre de longueur N est définie par :

$$E = \sum_{n=0}^{N-1} |s^2(n)| \quad (I.5)$$

où $s(n)$ est le signal de la parole et N est la longueur de la trame [23].

I.5.2.5 Le spectre de fréquence

Le contenu spectral du signal de parole change avec le temps. Si la transformée de Fourier discrète (DFT) de la forme d'onde de la parole est prise sur un intervalle de temps, elle peut estimer la plage de fréquences globale qui couvre la puissance principale du signal de parole. On constate que la puissance de la parole se concentre principalement aux basses fréquences lorsque le spectre de fréquences de la parole est mesuré dans un environnement non bruité [19].

I.5.2.6 Les formants

Les formants représentent les maxima spectraux, ces points de résonances du conduit vocal, amenant à l'émergence de certaines fréquences, correspondent au plan acoustique à l'individualisation de zones formantiques ou formants. Trois formants principaux, dits vocaliques car variant en fonction des données linguistiques, sont clairement identifiés. Il s'agit du grave vers l'aigu [11] :

- ✓ Du premier formant(F1) produit par le pharynx.
- ✓ Du second formant (F2) émis par la cavité buccale.
- ✓ Du troisième formant (F3) attribué à la cavité labiodentale.

I.6 Troubles de la voix

Les troubles de la voix peuvent généralement être regroupés en deux catégories :

- ✓ Les troubles fonctionnels : ceux résultant d'une mauvaise utilisation du mécanisme vocal (par exemple, des nodules vocaux, polypes, kystes, laryngites).
- ✓ Les troubles organiques : elles sont causées par une maladie structurelle (anatomique) ou physiologique, soit une maladie du larynx lui-même, soit par une maladie systémique distante qui modifie la structure ou la fonction du larynx (par exemple, une paralysie du pli vocal) [24].

I.6.1 Troubles fonctionnelles

I.6.1.1 Les nodules

Sont les lésions les plus souvent rencontrées : ce sont des épaissements de la muqueuse du bord libre des plis vocaux. Ils sont bénins. Ils sont souvent bilatéraux, situés l'un en face de l'autre au niveau du premier tiers antérieur des plis vocaux. Ils peuvent être fibreux ou œdémateux. Le traitement est généralement rééducatif et on ne recourt à la chirurgie que si la rééducation échoue. La figure I.11 représente les différents types des nodules [25].

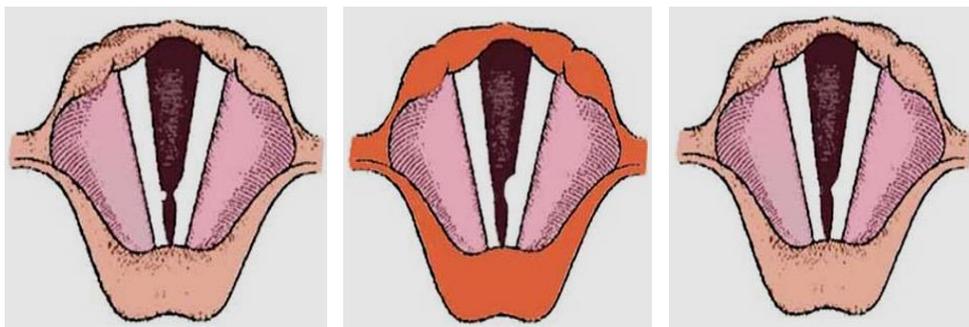


Figure I.10-*Divers types de nodules.*

I.6.1.2 Les polypes

Les polypes vocaux sont des lésions bénignes des plis vocaux qui ont tendance à être unilatérales et se produisent généralement au même endroit que les nodules vocaux, c'est-à-dire la jonction du tiers antérieur et du tiers moyen des plis vocaux (Figure I.11). De tels polypes

sont souvent précipités par un seul épisode vocal hyperfonctionnel (abus ou mauvaise utilisation), un peu comme ce qui cause la laryngite aiguë. Si la croissance est importante, elle peut couvrir la moitié à la moitié du pli vocal. Le polype, comme un nodule vocal, interfère avec les vibrations du pli vocal, provoquant ainsi un souffle et une dureté et une qualité de voix très enrouée.



Figure I.11-*Différentes types de polypes.*

I.6.1.3 Les laryngites

La laryngite est un terme général qui désigne un trouble de la voix aigu ou chronique pouvant avoir diverses étiologies, telles que des abus vocaux ou des infections bactériennes ou virales du larynx (Figure I.12). Le terme laryngite est utilisé pour décrire une inflammation de la muqueuse vocale qui provoque une dysphonie légère à sévère avec une hauteur tonale réduite et des pauses de phonation intermittentes (la voix s'interrompt de manière aléatoire). La laryngite est l'un des problèmes de larynx hyperfonctionnels les plus courants qui entraîne un trouble de la voix. La qualité vocale généralement entendue dans la laryngite est un enrrouement, c'est-à-dire une combinaison de dureté et de respiration.

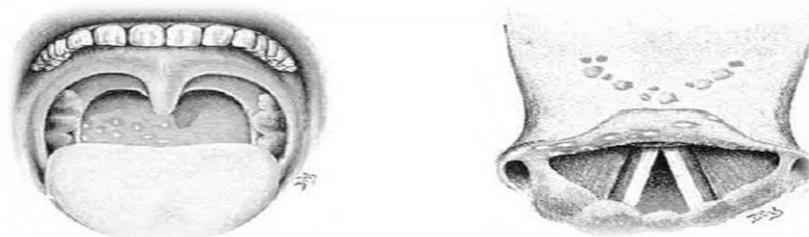


Figure I.12-*Laryngite herpétique.*

I.6.1.4 Les kystes

Les kystes des cordes vocales sont des masses bénignes constituées de structures ressemblant à des poches, que l'on trouve généralement en position unilatérale au milieu du

bord médial des cordes vocales comme montré par la figure I.13. Les symptômes comprennent l'enrouement et une diminution de la hauteur du son [26].

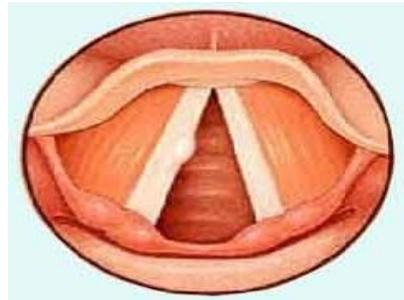


Figure I.13- *Kyste épidermique.*

I.6.2 Troubles organiques

La paralysie d'une corde vocale se manifeste par l'immobilité ou la diminution de mobilité d'une des deux cordes vocales en laryngoscopie indirecte. Elle est causée par la paralysie des muscles intrinsèques du larynx. Il s'agit d'un symptôme d'un trouble sous-jacent et non d'une maladie en soi. Les muscles intrinsèques de la corde vocale sont fournis par le nerf vague. Le terme vague signifie « vagabond », terme approprié pour décrire ce nerf en raison de sa longue trajectoire anatomique.

Les paralysies vocales unilatérales se produisent en raison d'un dysfonctionnement du larynx récurrent ou du nerf vague. Une paralysie unilatérale du larynx récurrent ou du nerf vague provoque une voix essoufflée. Le souffle de la voix est provoqué par une faille glottique qui permet à l'air de s'échapper lorsque le patient tente de parler. La production vocale normale dépend de la bonne fermeture de la glotte résultant de l'adduction bilatérale des cordes vocales. Cette adduction des cordes vocales, combinée à une pression atmosphérique sous-glottique, fait vibrer les cordes vocales, provoquant une phonation [27].

I.7 Evaluation des troubles de la voix

Les 3 paramètres acoustiques de la voix sont :

- ✓ La hauteur tonale ou fréquence fondamentale : exprimée en Hertz, représentée par le nombre de cycle vibratoire (ouverture - fermeture) par seconde.
- ✓ L'intensité : exprimée en décibels, correspond à l'amplitude des vibrations sonores.

- ✓ Le timbre : déterminer par l'intensité des harmoniques qui sont des multiples entiers de la fréquence fondamentale.

Le signal acoustique de la voix est très instable, et la voix varie en permanence dans son intensité et dans sa hauteur (intonation). Les 3 paramètres acoustiques peuvent être étudiés séparément, indépendamment les uns et des autres ou mesurés en même temps. Cela dépend des moyens d'analyse dont on dispose.

L'objectif de l'analyse du signal acoustique est d'extraire des indices pertinents permettant de déterminer les caractéristiques de la voix afin de renseigner sur l'état du larynx (cordes vocales) du locuteur.

L'évaluation des troubles de la voix se classe en deux approches :

- ✓ Evaluation subjective : basé sur une évaluation perceptive à l'oreille.
- ✓ Evaluation objective : permettre de porter un regard objectif sur le trouble et la qualité vocale grâce à des mesures acoustiques et aérodynamiques.

I.7.1 Evaluation subjective

Elle ne requiert qu'une bonne oreille, bien entraînée, celle de l'examineur. C'est une évaluation psycho-acoustique. L'écoute de la voix doit se faire pendant la voix conversationnelle, puis en demandant au patient de parler fort, un enregistrement est réalisé en même temps. Ceci donne une bonne idée de la hauteur de la voix, de l'intensité, du timbre, de l'intonation, des capacités à s'adapter à forte intensité et dans le chant.

On peut aussi utiliser des échelles d'évaluation pour quantifier de façon un peu plus précise la qualité de ce que l'on entend. Un des protocoles est l'échelle GRBAS. Chaque item est coté de 0 à 3 : 0 (normal), 1 (un peu altéré), 2 (modérément altéré), 3 (très altéré).

- ✓ **G** ou grade : appréciation globale de la qualité de la voix.
- ✓ **R** ou roughness or harshness : impression audible d'irrégularités des cycles vibratoires, de fluctuations anormales de la fréquence fondamentale.
- ✓ **B** ou breathiness : impression audible de fuite d'air en phonation.
- ✓ **A** ou asthenicity : fatigue vocale ou voix hypotonique
- ✓ **S** ou strain : voix forcée ou voix hypertonique [28].

I.7.2 Evaluation objective

L'évaluation objective utilise les techniques de traitement automatique de la parole (TAP). L'objectif est d'extraire des indices acoustiques pertinents permettant de déterminer les caractéristiques de la voix pour renseigner sur l'état du larynx du locuteur. Les cliniciens préfèrent des indices acoustiques qui sont corrélés avec les caractéristiques perceptuelles. Cependant, il serait vain de rechercher un accord acoustique-perceptuel lorsque l'évaluation perceptive n'est pas suffisamment fiable.

Les études de caractérisation des troubles de la voix sont consacrées au développement de méthodes d'analyse dédiées à l'estimation des dyspériodicités vocales dans le signal de parole.

Il existe plusieurs méthodes d'évaluation objective de la voix pathologique. Les méthodes d'analyse présentées dans cette section sont classées en deux classes selon qu'elles utilisent le signal de parole ou une estimation de la source glottique pour calculer l'indice acoustique.

I.7.2.1 Evaluation des troubles de la voix via le signal de parole

Les méthodes d'estimation des dyspériodicités vocales par utilisation de signal de la parole peuvent être classées en méthodes temporelles et méthodes spectrales. Plusieurs indices acoustiques utilisés pour quantifier le degré de perturbation de la parole produite par les locuteurs dysphoniques reflètent la déviation du signal de parole visée par rapport à la périodicité parfaite. Les causes de ces dyspériodicités sont différentes : vibrations non modales des cordes vocales, bruit de modulation comprenant les variations cycle à cycle de la durée de cycle (jitter) et de l'amplitude (shimmer) dues aux perturbations externes et bruit additif causé par une turbulence excessive.

I.7.2.2 Evaluation basée sur l'estimation de la source glottique

L'analyse du signal de la source glottique a été utilisée dans plusieurs applications de traitement automatique de signal de parole, et plus particulièrement pour la détection des troubles de la voix. En effet, les pathologies de la parole sont souvent associées à un dysfonctionnement des cordes vocales, par exemple, un nodule ou un polype. En plus, L'utilisation de la source glottique peut également être intéressante dans d'autres applications telles que la création d'un signal d'excitation artificiel basé sur la compréhension de la

production de signal de la source glottique pour les patients ayant subi une laryngectomie totale, et la réparation de la voix dysphonique où la parole alaryngée est re-synthétisée en utilisant un signal glottique synthétique [29].

I.8 Conclusion

Dans ce chapitre, nous avons étudié le système de production de la parole, nous avons décrit les organes responsables de la production sonore, et nous avons vu que la source du son se situe au niveau du larynx, qui joue un triple rôle dans ce système. Mais celui qui nous intéresse concerne son rôle lors de la respiration est la qualité des sons produits par la vibration des cordes vocales.

Le premier propos de notre mémoire, est d'analyser le signal de la parole pour renseigner sur l'état du larynx du locuteur. Les chapitres suivants présentent les méthodes d'analyse pour l'évaluation objective de la qualité vocale.

Chapitre

II

Décomposition modale empirique

II.1 Introduction

Le développement de méthodes pour le traitement des signaux non-linéaires et/ou non-stationnaires a fait l'objet de nombreux travaux de recherches, notamment au cours des vingt dernières années. De tels signaux ne peuvent pas être analysés de façon optimale par les méthodes usuelles de traitement du signal qui ont été développées pour des signaux linéaires [30]. Récemment, Huang et al ont introduit une méthode d'analyse locale auto-adaptative basée sur la décomposition en sous bandes appelée décomposition modale empirique (EMD : empirical mode decomposition) [31].

Dans ce chapitre, on présente la procédure de décomposition modale empirique qui utilise un processus appelé tamisage (sifting) permettant de décomposer temporellement des signaux multi-composantes en somme de signaux mono-composante. Les différentes étapes du processus de tamisage seront détaillées tout en donnant un exemple d'illustration de la décomposition. Enfin, on donne certaines propriétés fondamentales de la décomposition modale empirique.

II.2 Décomposition modale empirique

La Décomposition en modes empiriques est une méthode pilotée par les données permettant de décomposer les signaux complexes (non stationnaires et issus de systèmes non linéaires) en somme de signaux mono-composantes modulées en amplitude et en fréquence (AM-FM) appelées fonctions de modes intrinsèques (IMFs : intrinsic mode functions) [32]. Idéalement, ces IMFs sont des fonctions oscillantes, de moyenne nulle, et de plus, celle-ci reflètent les fréquences présentes localement dans le signal, des plus hautes fréquences aux plus basses fréquences. Le résidu est un terme de basse fréquence qui donne la tendance globale du signal [33]. L'extraction de ces composantes oscillantes est non-linéaire mais leur recombinaison est linéaire.

D'autre part, la décomposition en modes empiriques constituant la première étape de la transformée de Hilbert-Huang est une procédure sans base théorique. Elle est définie par un processus itératif dit de tamisage [30].

L'avantage principal de l'EMD est qu'elle ne fait pas intervenir des fonctions de base externes, les fonctions de base sont obtenues à partir du signal lui-même.

Par conséquent, l'analyse est adaptative contrairement aux méthodes traditionnelles où les fonctions de base sont prédéfinies [31].

II.2.1 Principe de la décomposition modale empirique

Dans son principe, l'EMD consiste à considérer localement un signal $s(t)$ comme la superposition d'une oscillation rapide et d'une tendance de fréquence plus basse.

Tout signal $s(t)$ peut être considéré comme la superposition d'une composante lente $r(t)$ (basse fréquence) appelée approximation et une composante rapide $d(t)$ (haute fréquence) appelée détail (Figure II.1) [34]. Ces composantes sont des IMFs (Fonctions Modales Intrinsèques) interprétées comme étant des ondes non stationnaires.

$$s(t) = d(t) + r(t) \quad (\text{II.1})$$

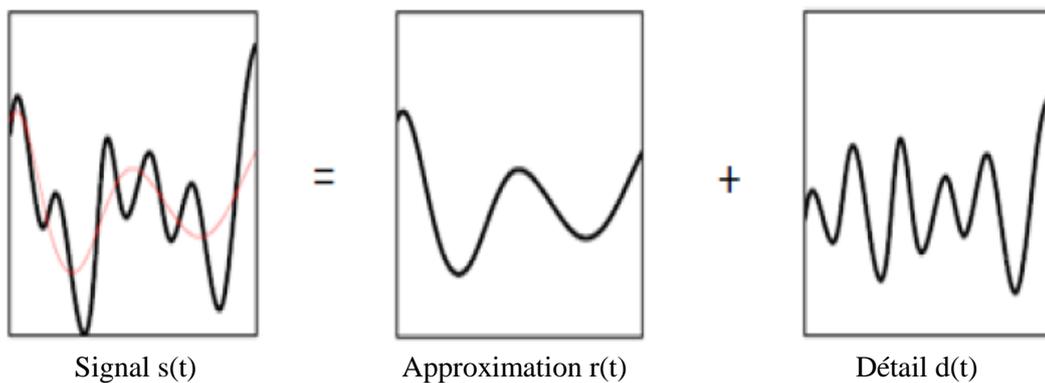


Figure II.1- Décomposition du signal en une composante lente et une autre rapide [34].

Pour ce faire, l'algorithme classique procède au calcul de l'enveloppe moyenne du signal à partir de l'interpolation des extrema locaux. Cette enveloppe est soustraite du signal puis la même procédure est appliquée à cette différence jusqu'à ce que l'enveloppe obtenue soit nulle ou quasi nulle.

Le résultat obtenu à l'issue de cette procédure donne la première IMF. Les modes suivants d'ordre supérieur sont calculés par le même processus dit de tamisage (sifting process en anglais) sur le résidu (différence du signal courant et de l'IMF).

L'algorithme de la décomposition peut s'arrêter à un ordre défini, ou lorsqu'il n'existe plus d'oscillations à extraire dans le résidu. Ainsi, pour tout signal $s(t)$, l'EMD donne la décomposition suivante :

$$s(t) = \sum_{j=1}^k \text{imf}_j(t) + r_k(t) \quad (\text{II. 2})$$

où imf_j est la $j^{\text{ième}}$ IMF du signal, r_k est le résidu final (un polynôme de faible degré), et k le nombre des IMFs ou de modes [32].

Comme k est un nombre entier fini, si on additionne toutes les IMF_j ainsi que le résidu $r(t)$, on reconstruit linéairement le signal original sans perte ou déformation de l'information initiale. Bien que l'EMD soit une approche non linéaire, la reconstruction du signal initial, à partir de ses modes, est linéaire [35].

II.2.2 Processus de tamisage : principe algorithmique

Le processus de tamisage (sifting process en anglais) correspond à l'extraction d'une IMF à partir d'un signal donné. La fonction de mode intrinsèque, ou IMF, ainsi obtenue peut être vue comme les détails du signal, de variation rapide. Pour cela, on extrait de façon itérative ces détails, jusqu'à conserver une IMF [36].

Dans ce qui suit nous détaillons les étapes de l'algorithme de la décomposition modale empirique, tout en montrant son aspect récursif. Etant donné un signal $s(t)$, sa décomposition par l'EMD est réalisée selon les étapes suivantes [37] :

Algorithme II.1-Les différentes étapes de l'algorithme EMD.

Étape 1) fixer $\varepsilon, j \leftarrow 1$ (j^{e} IMF)

Étape 2) $r_{j-1}(t) \leftarrow s(t)$ (résidu)

Étape 3) extraire la j^{e} IMF :

- a) $h_{j,i-1}(t) \leftarrow r_{j,i}(t), i \leftarrow 1$ (i , itération de la boucle de tamisage)
- b) Extraire les maxima et minima locaux de $h_{j,i-1}(t)$
- c) Calculer les enveloppes supérieure et inférieure $U_{j,i-1}(t)$ et $L_{j,i-1}(t)$ par interpolation des maxima et minima locaux de $h_{j,i-1}(t)$ respectivement
- d) Calculer l'enveloppe moyenne : $u_{j,i-1}(t) \leftarrow (U_{j,i-1}(t) + L_{j,i-1}(t)) / 2$
- e) Mettre à jour : $h_{j,i}(t) \leftarrow h_{j,i-1}(t) - u_{j,i-1}(t), i \leftarrow i+1$
- f) Calculer le critère d'arrêt :

$$SD(i) = \sum_{t=0}^T \frac{|h_{j,i-1}(t) - h_{j,i}(t)|^2}{(h_{j,i-1}(t))^2}$$

où T représente le nombre d'échantillons du signal.

- g) Répéter l'étape (b)-(f) tant que $SD(i) < \varepsilon$ et alors mettre :
IMF $_j(t) \leftarrow h_{j,i}(t)$

Étape 4) Mettre à jour le résidu : $r_j(t) \leftarrow r_{j-1}(t) - \text{IMF}_j(t)$

Étape 5) Répéter l'étape (3) avec $j \leftarrow j+1$ jusqu'à ce que le nombre de point d'extrema dans $r_j(t)$ soit inférieur à 2.

De manière plus littérale, les différentes étapes précédentes de l'algorithme de la décomposition modale empirique sont expliquées comme suite (Figure II.2) :

- ✓ Trouver les extrema locaux (maximum et minimum) du signal (étape : 3b).
- ✓ Estimer les enveloppes supérieures et inférieures par interpolation respective des maxima et minima locaux (étape : 3c) [38]. Le calcul de ces enveloppes se fait via une interpolation, dont la méthode la mieux adaptée à l'EMD est l'interpolation par les splines cubiques [39].

$$U_{j,i-1} = \left\{ h_{j,i-1}(t) : \frac{dh_{j,i-1}(t)}{dt} = 0, \quad \frac{d^2h_{j,i-1}(t)}{dt^2} < 0 \right\} \quad (\text{II.3})$$

$$L_{j,i-1} = \left\{ h_{j,i-1}(t) : \frac{dh_{j,i-1}(t)}{dt} = 0, \quad \frac{d^2h_{j,i-1}(t)}{dt^2} > 0 \right\}$$

- ✓ Estimer l'enveloppe moyenne locale à partir des enveloppes supérieur et inférieur (étape : 3d)
- ✓ Soustraire l'enveloppe moyenne du signal d'entrée (étape : 3e). Cela correspond alors à la première itération du tamisage. Le tamisage doit être réitéré jusqu'à ce que $h_{j,i}(t)$ soit une vrai IMF [38], c'est-à-dire respectant deux conditions :
 - Dans l'ensemble du jeu de données, le nombre d'extrema et le nombre de passages par zéro doivent être égaux ou tout au plus leur différence égale à 1.
 - En tout point, la valeur moyenne de l'enveloppe définie par les maximas locaux et de l'enveloppe définie par les minima locaux est égale à zéro [40].

Si $h_{j,i}(t)$ satisfait aux conditions d'une IMF, et le critère d'arrêt assure que ce signal obtenu vérifie bien les propriétés d'une IMF, tout en limitant le nombre d'itération. Alors on peut écrire $\text{IMF}_j(t) \leftarrow h_{j,i}(t)$.

- ✓ Vérifier si le résidu présente un nombre suffisant d'extrema (supérieur à 2) et réitérer sur le signal résultant le processus d'extraction de L'IMF (étape : 3) ; sinon, le résidu est considéré comme étant le résidu final $r(t)$ de l'équation (II.2). Idéalement, le processus d'extraction des IMFs est terminé lorsque le résidu ne contient plus d'extrema [38].

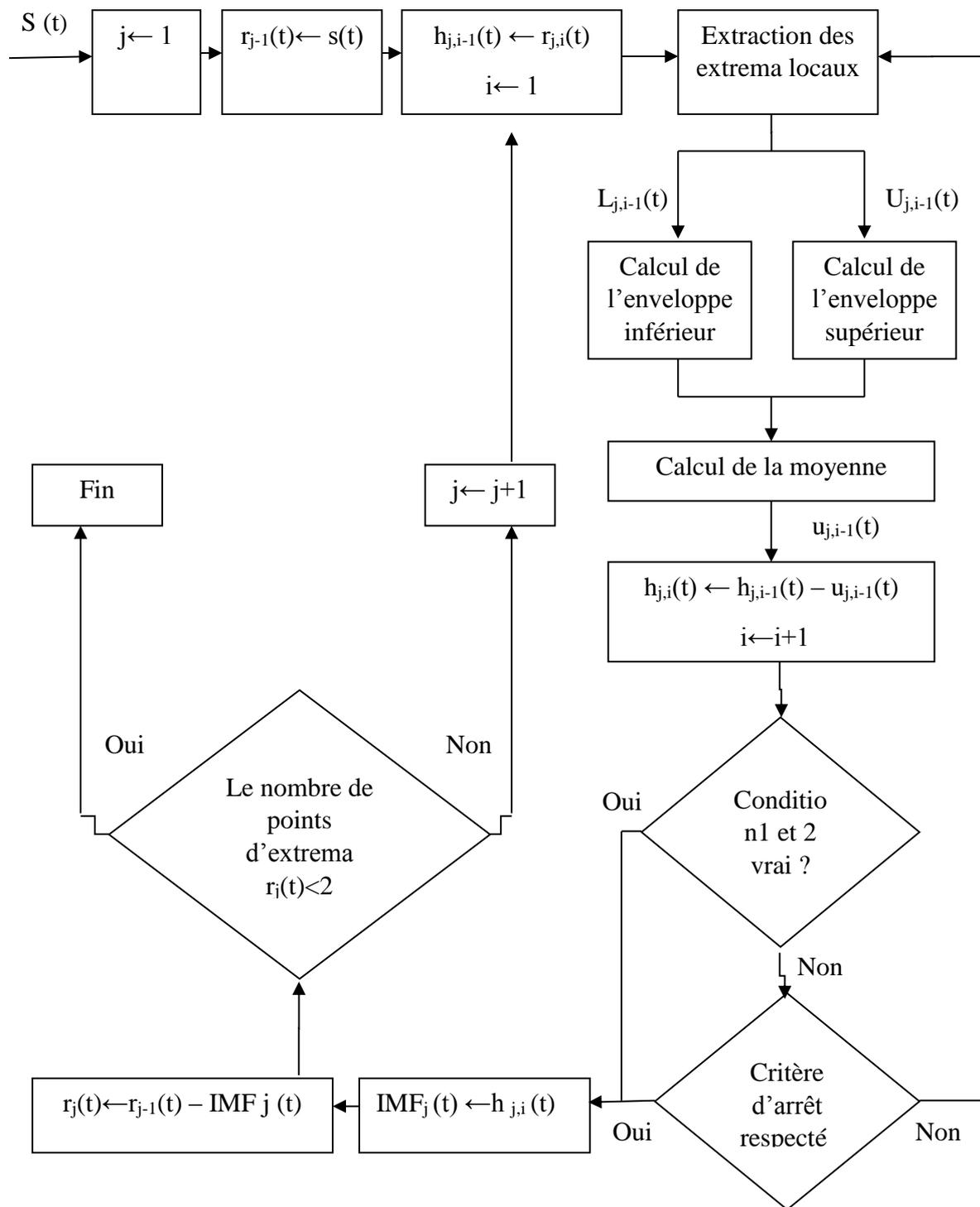


Figure II.2-Organigramme de l'algorithme de l'EMD [39].

Nous pouvons noter que la méthode présente deux boucles :

La boucle principale, indicée sur j . Elle s'interrompt lorsqu'il n'est plus possible d'extraire du résidu courant une autre IMF ($r(t)$ présentant moins de deux extrema). Ce qui permet de déterminer l'échelle de l'EMD (niveau de profondeur de la décomposition k).

La deuxième boucle, indicée sur i , est plus fondamentale. En effet, l'application stricte du concept (boucle sur j) conduirait à des fonctions $IMF_j(t)$ qui ne seraient pas nécessairement des IMFs : notamment les deux enveloppes L et U ne serait pas nécessairement symétrique par rapport à zéro. D'où la nécessité d'avoir une seconde boucle (indicée sur i) qui correspond au processus de tamisage [41].

II.2.3 Critère d'arrêt du processus de tamisage

Comme il existe deux boucles dans l'algorithme, il faut s'assurer que les deux doivent s'arrêter. La boucle principale indicée par j s'arrête lorsqu'il n'est plus possible de décomposer le résidu courant c.-à-d. que $r_j(t)$ possède moins de deux extrema. La boucle indicée par i est liée à un critère d'arrêt qu'il convient de définir de manière précise.

Ce critère d'arrêt noté $SD(i)$ est employé par Huang et al, reposant sur la déviation standard et défini par :

$$SD(i) = \sum_{t=0}^T \frac{|h_{j,i-1}(t) - h_{j,i}(t)|^2}{(h_{j,i-1}(t))^2} \quad (\text{II. 4})$$

Le test d'arrêt est valide lorsque la différence entre deux tamisages consécutifs est inférieure à un seuil prédéfini ε . Typiquement, la valeur ε permettant de stopper le tamisage est comprise entre 0.2 et 0.3. Cette valeur réalise un certain compromis. En effet, si ε est trop grand, l'EMD ne permet pas de séparer les différents modes présents dans le signal. Cependant si ε est trop petit, l'EMD risque d'aboutir à des composantes dont l'amplitude est quasiment constante et modulée par une seule fréquence (sur-décomposition du signal) [42].

II.3 Propriétés fondamentales de la décomposition modale empirique

II.3.1 Non-linéarité

L'EMD est une transformation non-linéaire, dans le sens où la décomposition de la somme de deux signaux diffère de la somme des décompositions de chaque signal. La cause principale de ce comportement est la dépendance de l'EMD des extrema locaux, du critère d'arrêt et de l'interpolation qui peuvent créer des non-linéarités. Cette propriété permet de mieux analyser des données issues de systèmes eux-mêmes non-linéaires [43].

II.3.2 Localité

Dans le cadre de l'EMD, l'échelle locale est définie par les extrema. Pour s'en convaincre, il suffit de rappeler que la moyenne locale est définie à partir des enveloppes du signal qui interpolent les maxima et les minima. La notion d'enveloppe étant intrinsèquement locale, on en déduit que la valeur, par exemple de l'enveloppe supérieure, en un point donné dépend essentiellement des deux maxima qui entourent ce point et dans une moindre mesure des autres maxima plus éloignés. Par conséquent, on peut considérer que les enveloppes, et donc la moyenne locale, sont définies localement à l'échelle correspondant à l'espacement entre les maxima/minima.

II.3.3 Aspect multi-résolution

L'EMD réalise une décomposition multi-échelles, ou multi-résolution, dans la mesure où elle explore successivement les échelles du signal de la plus fine, représentée par la première IMF, à la plus grossière, représentée par la dernière IMF ou le résidu. Par rapport à d'autres méthodes d'analyse multi-résolution, telle que la transformée en ondelettes, l'EMD présente un certain nombre de particularités. A l'instar de la transformée en ondelettes discrète, elle propose une décomposition en échelles discrètes dans la mesure où la décomposition est constituée d'un nombre fini de composantes.

II.3.4 Auto-adaptativité

L'EMD fournit une décomposition auto-adaptative contrairement à la transformée en ondelettes pour laquelle on est obligé de choisir une ondelette mère adéquate selon la nature ou la classe du signal à décomposer. L'avantage de l'EMD réside dans le fait que la décomposition ne dépend que des points caractéristiques qui sont des données intrinsèques au signal à décomposer en l'occurrence les maxima, minima ou points d'inflexion [44].

II.3.5 Quasi-orthogonalité

Les IMFs sont extraites au fur et à mesure de la décomposition. Etant déduites algorithmiquement, on ne peut les considérer comme des bases Hilbertiennes. Cependant, de nombreuses études ont montré le caractère quasi-orthogonal de la décomposition EMD. Intuitivement, cela vient du fait que deux ondes oscillant à des fréquences distinctes sont globalement orthogonales. Cependant, plusieurs phénomènes peuvent nuire à l'orthogonalité de la décomposition : les non-stationnarités d'une part, mais aussi le fait que la moyenne locale nulle n'entraîne pas l'annulation de l'intégrale entre deux extrema successifs [45].

II.4 Conclusion

Dans ce chapitre, nous avons étudié la technique de la décomposition modale empirique, qui est une approche de type multi-résolutions introduite pour le traitement des signaux non linéaires et/ou non stationnaires. L'EMD permet de décomposer un signal oscillant en des contributions de base appelées fonctions de modes intrinsèques (IMFs) qui sont des composantes modulées en amplitude et en fréquence (AM-FM) extraites successivement à partir des enveloppes inférieures et supérieures du signal. L'extraction de ces fonctions se fait par un processus itératif appelé « tamisage » définie par un algorithme spécifique.

Chapitre

III

Estimation des dyspériodicités vocales

III.1 Introduction

Le larynx est un organe essentiel dans la production de la voix. Dans beaucoup de situations, la présence de pathologies laryngales organiques ou fonctionnelles entraîne une altération du signal acoustique. Il existe une relation étroite entre la qualité de la voix perçue et les paramètres du signal acoustique qui reflètent la périodicité locale de ce dernier dans le cas des sons voisés. L'analyse du signal de la parole offre un moyen privilégié pour l'évaluation clinique de la voix en vue d'une documentation quantitative des pathologies du larynx. Une approche permettant de mettre en évidence la relation entre le degré de perturbation de la voix et le signal acoustique est d'utiliser une analyse de corrélation entre un indice acoustique obtenu à partir du signal de la parole et les scores de l'évaluation perceptive. L'indice acoustique quantifie numériquement les dyspériodicités vocales qui sont étroitement liées à la qualité de la voix perçue.

Les méthodes d'estimation des dyspériodicités vocales à partir du signal de la parole peuvent être divisées en deux classes : méthodes temporelles et méthodes fréquentielles. Parmi les méthodes temporelles récemment proposées dans la littérature, l'analyse prédictive bidirectionnelle à long terme et le variogramme généralisé seront présentés dans ce chapitre. Dans la classe des méthodes fréquentielles d'estimation des dyspériodicités vocales, on présente la décomposition en modes empiriques du spectre du signal de la parole. Un indice acoustique est utilisé pour quantifier le degré d'enrouement de la voix par une mesure directe des Dyspériodicités vocales.

III.2 Méthodes d'estimation des dyspériodicités vocales

III.2.1 Analyse prédictive bidirectionnelle à long terme

III.2.1.1 Méthode d'analyse

Dans le modèle de prédiction linéaire à long terme, la trame d'analyse courante de longueur N est prédite au moyen d'une trame décalée de même longueur [4]. Le prédicteur à long terme à un coefficient $\hat{x}(n)$ est exprimé comme suit :

$$\hat{x}(n) = ax(n - T), \quad n = 0, 1, \dots, N - 1 \quad (\text{III. 1})$$

L'indice n positionne les échantillons de la parole dans la trame d'analyse. Lorsque le délai T est positif, la trame d'analyse est positionnée à gauche de la trame courante, sinon elle est positionnée à droite.

Le délai de prédiction T est déterminé pour chaque terme d'analyse soit recherchant le délai qui maximise l'inter-corrélation entre la trame courante et la trame décalée soit au moyen d'une recherche exhaustive dans l'intervalle des valeurs entières $I = [-T_{\max}, -T_{\min}] \cup [T_{\min}, T_{\max}]$, où les bornes T_{\min} et T_{\max} sont les longueurs acceptables minimales et maximales, pour le cycle glottique, exprimées en nombre d'échantillons. Elles sont fixées à 2.5ms et 20ms, respectivement ($50\text{Hz} \leq F_0 \leq 400\text{ Hz}$). La valeur optimale du délai de prédiction est celle qui minimise l'énergie de l'erreur de prédiction.

Pour chaque trame d'analyse, le coefficient du prédicteur à long terme est estimé en minimisant l'énergie de l'erreur de prédiction E .

$$E = \sum_{n=0}^{N-1} (x(n) - ax(n-T))^2 \quad (\text{III. 2})$$

Ce qui donne :

$$a = \frac{\sum_{n=0}^{N-1} x(n)x(n-T)}{\sum_{n=0}^{N-1} x^2(n-T)} \quad (\text{III. 3})$$

La valeur optimale du délai de prédiction est celle qui minimise l'énergie de l'erreur de prédiction :

$$T_{opt} = \underset{T \in I}{\operatorname{argmin}} \left\{ \sum_{n=0}^{N-1} \left[x(n) - x(n-T) \frac{\sum_{n=0}^{N-1} x(n)x(n-T)}{\sum_{n=0}^{N-1} x^2(n-T)} \right]^2 \right\} \quad (\text{III. 4})$$

Le délai de prédiction optimal T_{opt} n'est pas contraint d'être égal à la longueur du cycle glottique. L'expression (III.5) est définie indépendamment du fait que la trame d'analyse soit voisée ou non voisée, régulier ou irrégulier. La longueur de la trame d'analyse ainsi que la longueur du décalage sont fixées à 2,5ms. La valeur instantanée de l'erreur de prédiction est calculée par :

$$e(n) = x(n) - ax(n - T_{opt}), \quad 0 \leq n \leq N - 1 \quad (\text{III. 5})$$

où le délai T_{opt} peut être positif ou négatif.

III.2.1.2 Indice acoustique

L'indice acoustique utilisé pour quantifier les dyspériodicités vocales dans le signal de parole est le rapport signal à dyspériodicité (RSD). Le RSD, exprimé en dB, est défini comme étant le logarithme du rapport entre l'énergie du signal et l'énergie de la dyspériodicité :

$$RSD = 10 \log \left[\frac{\sum_{n=0}^{L-1} x^2(n)}{\sum_{n=0}^{L-1} e^2(n)} \right] \quad (\text{III. 6})$$

où L est la longueur d'analyse totale.

III.2.1.3 Propriétés de l'erreur de prédiction à long terme comme une mesure de dyspériodicités

La minimisation de l'énergie de l'erreur de prédiction peut donner lieu à un coefficient de prédiction négatif incompatible avec la définition de la périodicité. En effet, l'erreur de prédiction (III.2) est égale à zéro chaque fois que les expressions (III.7) s'appliquent :

$$E = 0 \quad \begin{cases} x(n) = +gx(n-T) \\ x(n) = -gx(n-T) \end{cases} \quad (\text{III.7})$$

Le paramètre g étant strictement positif et constant pour une trame d'analyse donnée. Les solutions (III.7) montrent que, formellement, l'erreur de prédiction à long terme peut ne pas être une mesure de dyspériodicités. Ceci est dû à l'existence de la solution parasite,

$$x(n) = -gx(n-T).$$

Une autre ambiguïté de l'analyse prédictive à long terme est due à l'existence d'une intercorrélacion dans le calcul du coefficient de prédiction. Comme conséquence, l'énergie de l'erreur de prédiction est une mesure des dyspériodicités vocales pour un bruit vocal de faible niveau. Lorsque le signal est très bruité, le coefficient de prédiction (III.4) tend vers 0 et l'énergie de l'erreur de prédiction s'exprime comme :

$$E = \sum_{n=0}^{N-1} x^2(n) \quad (\text{III. 8})$$

Par conséquent, à des niveaux de bruit élevés, l'erreur de prédiction est un mélange de l'énergie du signal et de l'énergie des dyspériodicités. Ces difficultés peuvent être évitées grâce au variogramme généralisé

III.2.2 Analyse basée sur le variogramme généralisé

Pour un signal $x(n)$ périodique de période T_0 , on peut écrire [4] :

$$x(n) = x(n - kT_0), \quad k = \dots, -2, -1, 0, 1, 2, \dots \quad (\text{III. 9})$$

Une mesure de la déviation par rapport à la dyspériodicité, calculée sur un intervalle de longueur N , fournit une indication sur le degré d'irrégularité du signal. Pour les signaux stationnaires, l'énergie de la dyspériodicité peut être estimée par l'expression suivante :

$$\hat{\gamma} = \min_T \left\{ \sum_{n=0}^{N-1} [x(n) - x(n-T)]^2 \right\} \quad (\text{III. 10})$$

avec $-T_{max} \leq T \leq -T_{min}$ et $T_{min} \leq T \leq T_{max}$

L'expression entre les accolades est connue sous le nom de variogramme du signal $x(n)$.

Formellement, l'expression entre crochet dans (III.10) est équivalente à la différence entre la trame courante et la trame d'analyse décalée de longueur N . L'indice temporel n positionne les échantillons du signal de parole à l'intérieur de la trame d'analyse.

Le signal de la parole est localement stationnaire. L'amplitude du signal évolue d'une trame à la suivante à cause des attaques et des déclins, de l'intensité des segments et de l'accentuation. En introduisant un gain α_T pour tenir compte de ces lentes variations de l'amplitude du signal, la définition (III.9) devient :

$$x(n) = \alpha_{T_0} x(n - kT_0), \quad 0 \leq n \leq N - 1, \quad k = \dots, -2, -1, 0, 1, 2, \dots \quad (\text{III. 11})$$

En conséquence, le variogramme empirique généralisé peut être écrit comme suit :

$$\hat{\gamma} = \min_T \left\{ \sum_{n=0}^{N-1} [x(n) - \alpha_T x(n-T)]^2 \right\} \quad (\text{III. 12})$$

avec $-T_{max} \leq T \leq -T_{min}$ et $T_{min} \leq T \leq T_{max}$

Le gain α_T est contraint d'être positif. Il est défini de manière à garantir des énergies identiques dans la fenêtre d'analyse courante et la fenêtre décalée.

$$\alpha_T = \sqrt{\frac{E_0}{E_T}}, \quad E_0 = \sum_{n=0}^{N-1} x^2(n), \quad E_T = \sum_{n=0}^{N-1} x^2(n-T) \quad (\text{III. 13})$$

Les paramètres T_{min} , T_{max} , la longueur et le décalage de la trame d'analyse sont définis comme précédemment.

L'effet du bruit de quantification peut être réduit en sur-échantillonnant le signal. La valeur instantanée de la dyspériodicité est estimée comme suit :

$$e(n) = x(n) - \alpha_{T_{opt}} x(n - T_{opt}), \quad 0 \leq n \leq N - 1 \quad (\text{III. 14})$$

où T_{opt} désigne le délai qui minimise le variogramme généralisé pour la position courante de la trame. Rappelons que le délai T_{opt} peut prendre des valeurs positives ou négatives.

L'énergie de la dyspériodicité par trame est donnée par :

$$E = \sum_{n=0}^{N-1} [x(n) - \alpha_{T_{opt}} x(n - T_{opt})]^2 \quad (\text{III. 15})$$

Notons que pour le bruit blanc stationnaire, le gain (III.13) $\rightarrow 1$. Donc, l'énergie (III.15) est une mesure de la différence inter-frames indépendamment du niveau de bruit. Sachant que les trames courante et décalée ne sont pas corrélées dans le cas d'un bruit blanc, l'énergie estimée (III.15) se réduit à :

$$E \cong 2 \sum_{n=0}^{N-1} x^2(n) \quad (\text{III. 16})$$

L'énergie de la dyspériodicité estimée au moyen du variogramme généralisé peut donc être supérieure à l'énergie du signal.

L'estimation des dyspériodicités en utilisant le variogramme généralisé se résume comme suit :

Algorithme III.1-Estimation des dyspériodicités vocales par le variogramme généralisé.

Pour chaque trame d'analyse $k = 0, 1, 2, \dots, K - 1$

- ✚ Calculer le gain $\alpha_T, -T_{max} \leq T \leq -T_{min}, T_{min} \leq T \leq T_{max}$
- ✚ Calculer le délai optimal $T_{opt} = \arg \min_{T \in I} \gamma_T$
- ✚ Calculer la dyspériodicité associée à la trame $k, e(n) = x(n) - \alpha_{T_{opt}} x(n - T_{opt}), n = kN, \dots, KN+N-1$.

Le rapport signal à dyspériodicité exprimé par (III.7) est utilisé pour quantifier numériquement les dyspériodicités dans l'intervalle d'analyse totale.

III.2.3 Analyse par décomposition modale empirique du spectre d'amplitude

La décomposition en modes empiriques est appliquée au logarithme du spectre d'amplitude du signal de parole pour le décomposer en ses composantes harmonique, enveloppe et bruit [6].

III.2.3.1 Séparation des composantes du signal de la parole

Une trame de parole voisée $x(t)$ peut être modélisée comme la convolution d'une source périodique $e(t)$ avec la réponse impulsionnelle du conduit vocal $v(t)$

$$x(t) = e(t) * v(t) \quad (\text{III. 17})$$

où $*$ désigne le produit de convolution.

En multipliant la trame du signal $x(t)$ par une fenêtre de pondération et en prenant l'amplitude de sa transformée de Fourier, on trouve :

$$|X_w(f)| = |E_w(f) \times V(f)| \quad (\text{III. 18})$$

où $X_w(f)$ et $E_w(f)$ sont, respectivement, les spectres des trames du signal de parole et du signal d'excitation fenêtrées et $V(f)$ est la réponse fréquentielle du conduit vocal.

En prenant le logarithme de (III.18), l'opération de multiplication devient une addition :

$$\log|X_w(f)| = \log|E_w(f)| + \log|V(f)| \quad (\text{III. 19})$$

L'équation (III.19) montre que le logarithme du spectre d'amplitude du signal de parole est la somme de deux composantes spectrales : $\log|E_w(f)|$, le logarithme du spectre d'amplitude du signal d'excitation fenêtré et $\log|V(f)|$, l'enveloppe spectrale dû à la caractéristique de filtrage du conduit vocal. En raison de la présence d'un bruit d'aspiration à la glotte, les équations (III.17) et III.18) deviennent :

$$x(t) = (e(t) + n(t)) * v(t) \quad (\text{III. 20})$$

$$|X_w(f)| = |(E_w(f) + N_w(f))V(f)| \quad (\text{III.21})$$

où $n(t)$ et $N_w(f)$ désignent respectivement le bruit d'aspiration et le spectre d'amplitude du bruit d'aspiration pondéré.

En prenant le logarithme des deux membres de (III.21), il vient :

$$\log|X_w(f)| = \log|E_w(f) + N_w(f)| + \log|V(f)| \quad (\text{III. 22})$$

Le spectre du bruit est large bande, cependant, si le signal bruyant présente une structure harmonique, le bruit domine entre les emplacements harmoniques et le spectre de l'excitation de la source glottique est concentré aux emplacements harmoniques. En conséquence, le spectre d'excitation lui-même peut être considéré comme composé de deux parties : la première partie consiste en une série d'harmoniques régulièrement espacées ayant une amplitude décroissante avec la fréquence et la seconde partie est un bruit réparti de manière irrégulière.

Le logarithme du spectre d'amplitude peut être considéré comme composé d'un contour variant lentement en fréquence, noté $V_{dB}(f)$, en raison de la contribution du conduit vocal, une série d'harmoniques caractérisée par une structure périodique, notée $H_{dB}(f)$, et une partie irrégulière et variant rapidement, notée $N_{dB}(f)$, en raison du bruit à la glotte.

L'algorithme EMD fournit un outil permettant de séparer les trois composantes du logarithme du spectre d'amplitude. En effet, l'algorithme EMD joue le rôle d'un banc de filtres, de sorte que la décomposition du logarithme du spectre d'amplitude via l'algorithme EMD aboutit à plusieurs modes oscillants (IMFs) pouvant être regroupés en trois classes, chaque classe est affectée à une partie du logarithme du spectre d'amplitude comme le montre le tableau (III.1).

Tableau III.1- Catégorisation des modes oscillants en fonction de leurs fréquences moyennes

f_j (la fréquence moyenne de la composante IMF $_j$)	Appartenance des IMF $_j$
$f_j < S_1$	Partie Enveloppe
$S_1 < f_j < S_2$	Partie harmonique
$f_j > S_2$	Partie bruit

où $s_{j,j} = 1, 2$ sont des seuils qui dépendent de la fréquence fondamentale moyenne F_0 du signal de la parole et elles sont fixées de façon empirique.

Chaque partie du logarithme du spectre d'amplitude du signal de parole est estimée en sommant les IMFs appartenant à la classe correspondante. Il a été montré que les seuils optimaux sont $S_1 = 0.3/F_0$ et $S_2 = 4/F_0$.

III.2.3.2 Indice acoustique

L'indice acoustique utilisé pour quantifier les dyspériodicités vocales est le rapport harmonique à bruit (RHB). L'intervalle d'analyse totale est divisé en K trames temporelles et le rapport harmonique à bruit est calculé comme étant la moyenne des rapports harmonique à bruit RHB_i ($i=1, 2, \dots, K$) des K trames.

$$RHB = \frac{1}{K} \sum_{i=1}^K RHB_i \quad (\text{III. 23})$$

$$\text{où} \quad RHB_i = 10 \log \left[\frac{\sum_{k=0}^{M-1} H_i^2(k)}{\sum_{k=0}^{M-1} N_i^2(k)} \right], \quad i = 1, \dots, K \quad (\text{III. 24})$$

avec H_i et N_i indiquant respectivement le spectre d'amplitude de la composante harmonique et le spectre d'amplitude de la composante de bruit pour la trame i et M est le nombre de points de fréquence.

III.3 Conclusion

Dans ce chapitre, nous avons décrit deux méthodes temporelles d'estimation des dyspériodicités vocales via le signal de la parole. Les deux méthodes sont l'analyse prédictive bidirectionnelle à long terme et le variogramme généralisé. De même, nous avons présenté une méthode fréquentielle basée sur la décomposition en modes empiriques du spectre d'amplitude du signal de la parole. Pour la quantification numérique des dyspériodicités vocales, les indices acoustiques appelés rapport signal à dyspériodicité et rapport harmonique à bruit sont utilisés.

Chapitre

IV

Analyse multivariée par décomposition
modale empirique

IV.1 Introduction

Contrairement à une analyse univariée qui utilise une seule variable (indice acoustique) pour l'analyse des dyspériodicités vocales, une analyse multivariée combine plusieurs variables (indices acoustiques). Dans ce chapitre, on se propose d'effectuer une analyse multivariée des dyspériodicités vocales en utilisant la décomposition en modes empiriques du spectre d'amplitude du signal de la parole. La méthode d'analyse multivariée est appliquée à un corpus de voyelles produites par des locuteurs normophoniques et dysphoniques et sa performance est comparée à celle de la méthode d'analyse basée sur le variogramme généralisé.

IV.2 Base de données

La base de données utilisé dans notre étude incluant des voyelles [a] produites par 28 locuteurs normophoniques âgés de 19 à 40 ans et 223 locuteurs dysphoniques âgés de 8 à 85 ans avec différents degrés de dysphonie [46]. Tous les stimuli sont échantillonnés à 44100 Hz.

Cinq juges ont évalué auditivement le corpus des données. Les cinq juges sont des thérapeutes professionnels de la voix avec au moins cinq ans d'expérience dans l'évaluation clinique de la qualité de la voix. Chaque juge a évalué, de 0 à 3, l'élément "grade" de l'échelle GRABS. Le "Grade" représente le degré d'enrouement ou anomalie globale de la voix [47]. Chaque stimulus a été assigné comme score global la moyenne des cinq scores. L'enregistrement et l'évaluation auditive ont été effectués à l'hôpital général de Sint-Jan, à Bruges, en Belgique.

IV.3 Méthode d'analyse

Pour un signal périodique de période T_0 , on peut écrire :

$$x(n) = x(n - T_0) \quad (\text{IV.1})$$

En passant au domaine spectral, il vient

$$X(f) = X_{T_0}(f) \quad (\text{IV.2})$$

où $X(f)$ et $X_{T_0}(f)$ désignent, respectivement, les spectres d'amplitude de la trame courante et de trame décalée de T_0 .

En appliquant l'algorithme de décomposition en modes empiriques aux spectres d'amplitude de la trame courante et de la trame décalée, on peut écrire :

$$IMF_i(f) = IMF_{i,T_0}(f) \quad i = 1, 2, \dots \quad (\text{IV.3})$$

où $IMF_i(f)$ et $IMF_{i,T_0}(f)$ désignent, respectivement, les fonctions de modes intrinsèques des spectres d'amplitude de la trame courante et de trame décalée de T_0 .

Les sons de la parole voisée produits par des locuteurs normophoniques peuvent être considérés comme pseudo-périodiques. De manière similaire à la méthode d'analyse basée sur le variogramme généralisé, on introduit pour chaque paire IMF_i – IMF_{i,T_0} un gain a_{i,T_0} pour tenir compte de l'évolution lente de l'amplitude du signal de la parole à cause des attaques et des déclins. L'équation (IV.3) devient :

$$IMF_i(f) = a_{i,T_0} IMF_{i,T_0}(f) \quad i = 1, 2, \dots \quad (IV.4)$$

Le gain a_{i,T_0} est donné par-:

$$a_{i,T_0} = \sqrt{\frac{E_i}{E_{i,T_0}}} \quad i = 1, 2, \dots \quad (IV.5)$$

où E_i et E_{i,T_0} sont, respectivement, les énergies des IMFs des spectres d'amplitude des trames d'analyse courante et décalée,

$$E_i = \sum_f IMF_i^2(f) \quad (IV.6.a)$$

$$E_{i,T_0} = \sum_f IMF_{i,T_0}^2(f) \quad (IV.6.b)$$

La dyspériodicité correspondante à chaque paire d' IMF_i est donnée alors dans le domaine fréquentiel par :

$$e_i(f) = IMF_i(f) - a_{i,T_0} IMF_{i,T_0}(f), \quad i=1, 2, \dots \quad (IV.7)$$

Pour chaque paire d' IMF_i , on définit un rapport signal à dyspériodicité exprimé dans le domaine fréquentiel par

$$RSD_i = 10 \log_{10} \left[\frac{\sum_f IMF_i^2(f)}{\sum_f e_i^2(f)} \right], \quad i=1, 2, \dots \quad (IV.8)$$

L'intervalle d'analyse totale est divisé en K trames temporelles et le rapport signal à dyspériodicité associé à chaque paire d'IMF est calculé comme étant la moyenne des rapports signal à dyspériodicité locaux des K trames.

Les rapports signal à dyspériodicité définis entre les M premières paires d'IMFs sont utilisés comme variables prédicteur des scores de l'enrouement perçu :

$$\text{enrouement} = b + \sum_{i=1}^M a_i RSD_i \quad (IV.9)$$

Les paramètres a_i et b sont les coefficients de régression obtenus par minimisation de l'erreur quadratique moyenne dans la prédiction de l'enrouement perçu comme combinaison linéaire des rapports signal à dyspériodicité entre les M premières paires d'IMFs.

Les expériences réalisées ont montré que les quatre premières IMFs contiennent l'essentiel de l'information sur le bruit vocal. Par conséquent, le nombre M de variables prédictes est fixé à 4.

IV.4 Exemple d'illustration

Dans cet exemple, la méthode d'analyse multivariée basée sur la décomposition en modes empiriques est appliquée à la voyelle [a] représentée sur la figure IV.1 produite par un locuteur dysphonique qui a été assigné un score moyen de 1 et de fréquence fondamentale $F_0=246$ Hz. Le signal est échantillonné à 44100 Hz.

L'analyse est effectuée trame par trame sans chevauchement entre les trames et l'algorithme de décomposition en modes empiriques est appliqué aux spectres d'amplitude de la trame courante et de la trame décalée d'une période fondamentale T_0 .

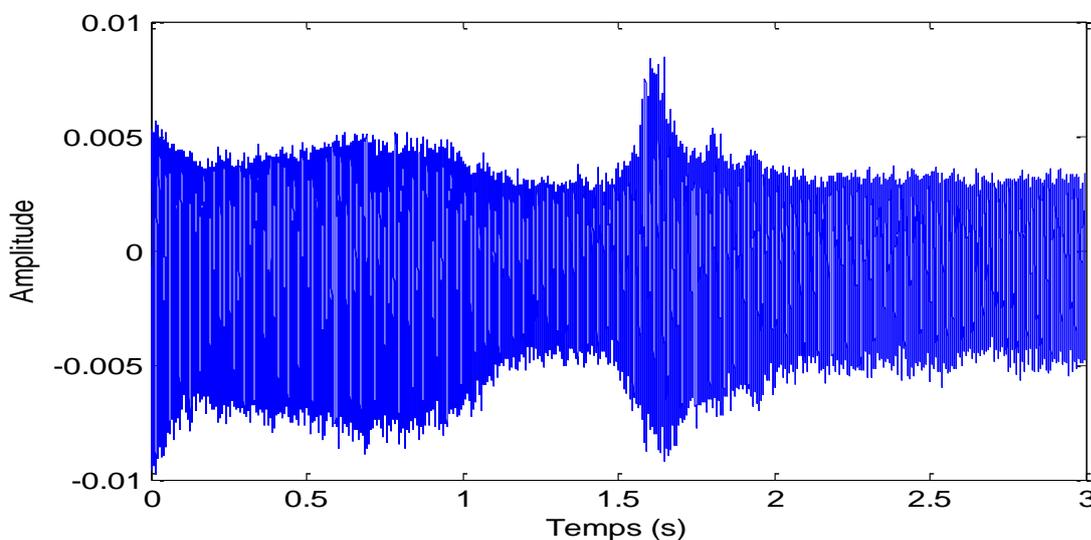
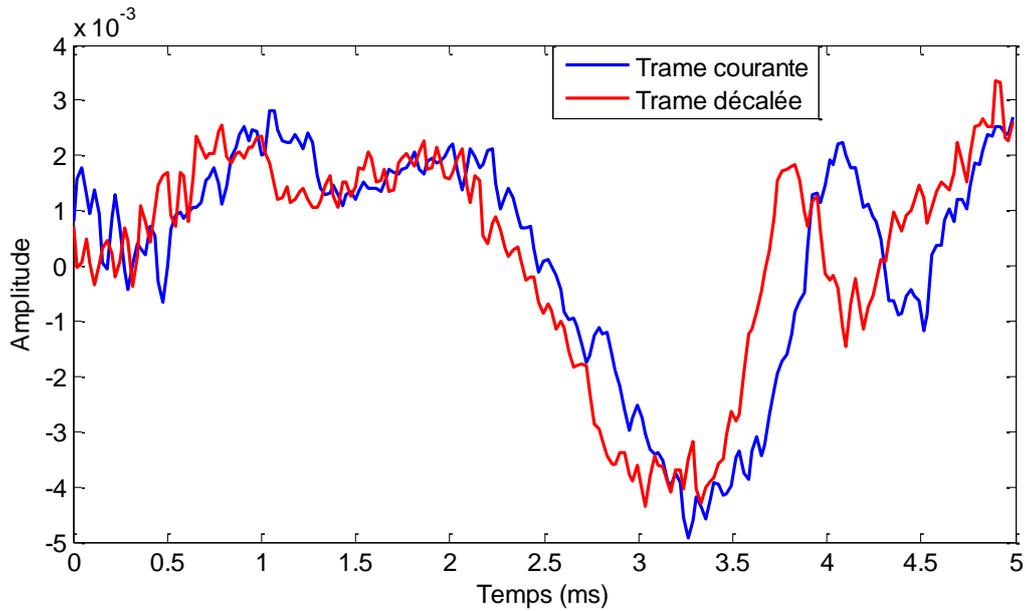
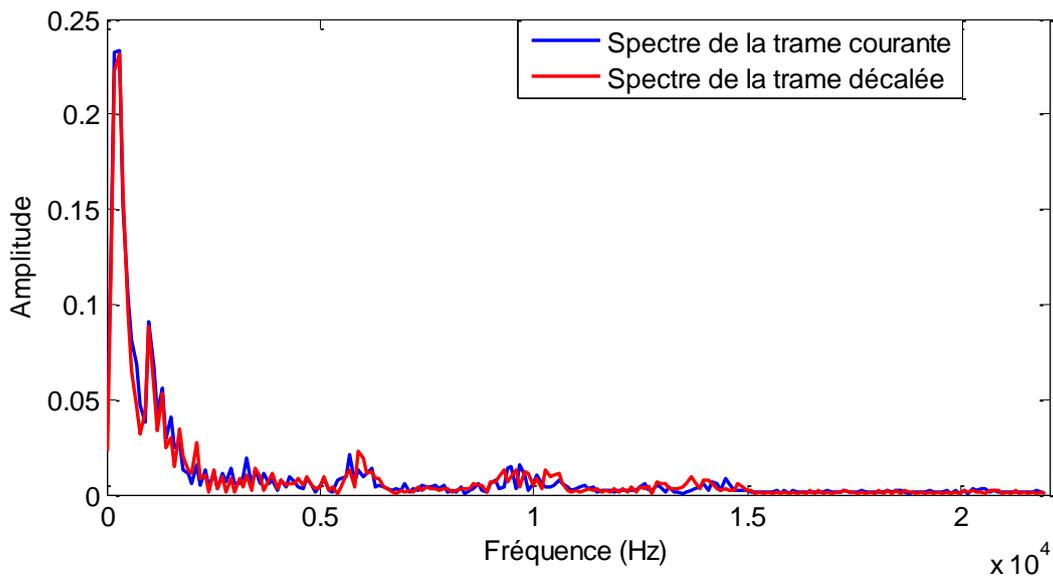


Figure IV.1- Signal de parole correspondant à une voyelle [a] produite par un locuteur dysphonique.

La figure IV.2 représente une trame courante de 5 ms extraite de la voyelle [a] produite par le locuteur dysphonique, la trame décalée, leurs spectres d'amplitudes respectifs ainsi que les quatre premières IMFs correspondantes. Les rapports signal à dyspériodicité correspondants aux quatre premières IMFs pour ce signal sont 17.1 dB, 10.4 dB, 8.8 dB et 14.8 dB.



Trame courante de 5 ms et trame décalée



Spectres d'amplitude des trames courante et décalée

Figure IV.2- Trame courante de 5 ms extraite d'une voyelle [a] produite par un locuteur dysphonique qui a été assigné un score moyen de 1, la trame décalée, leurs spectres d'amplitudes respectifs ainsi que les quatre premières IMFS correspondantes.

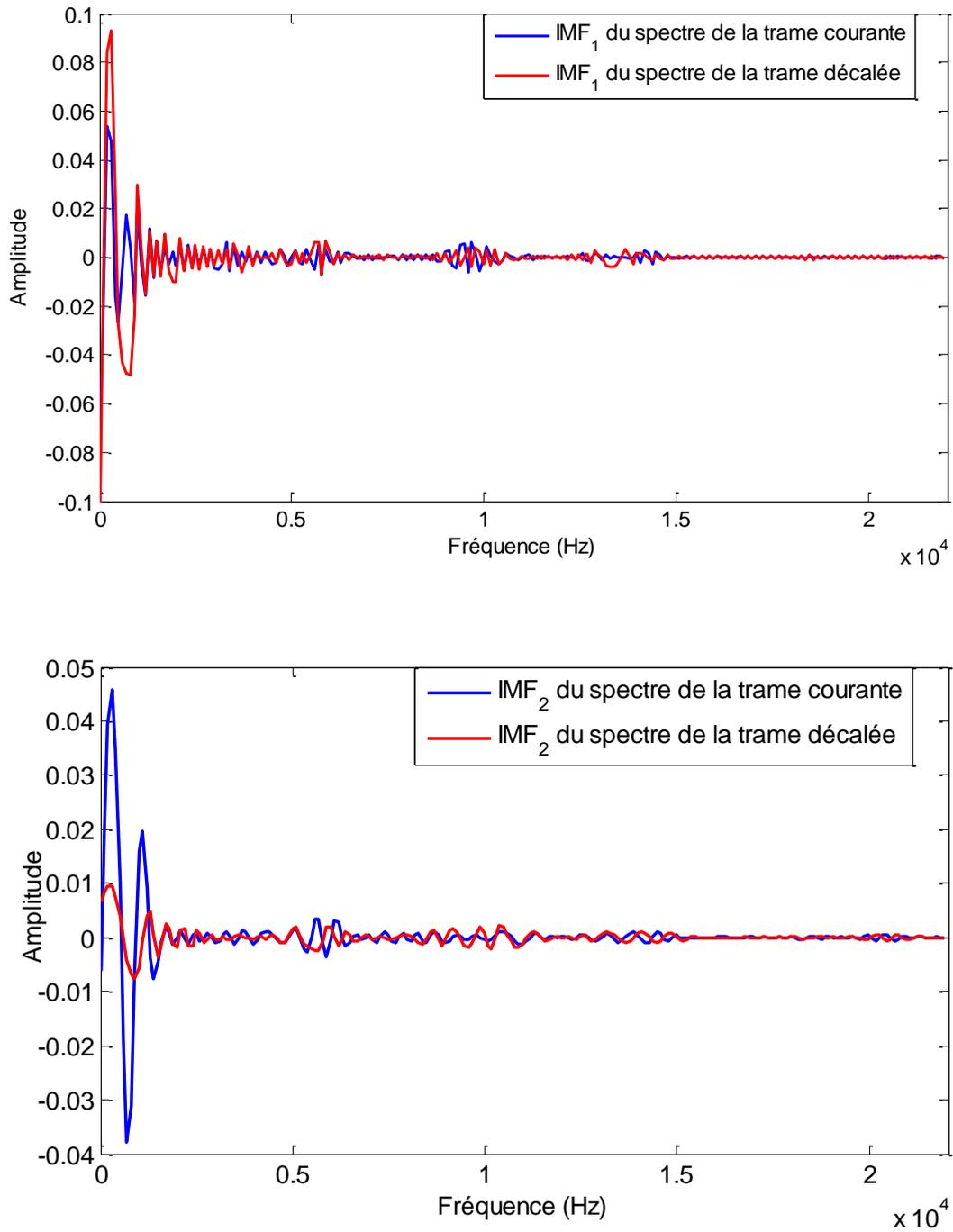
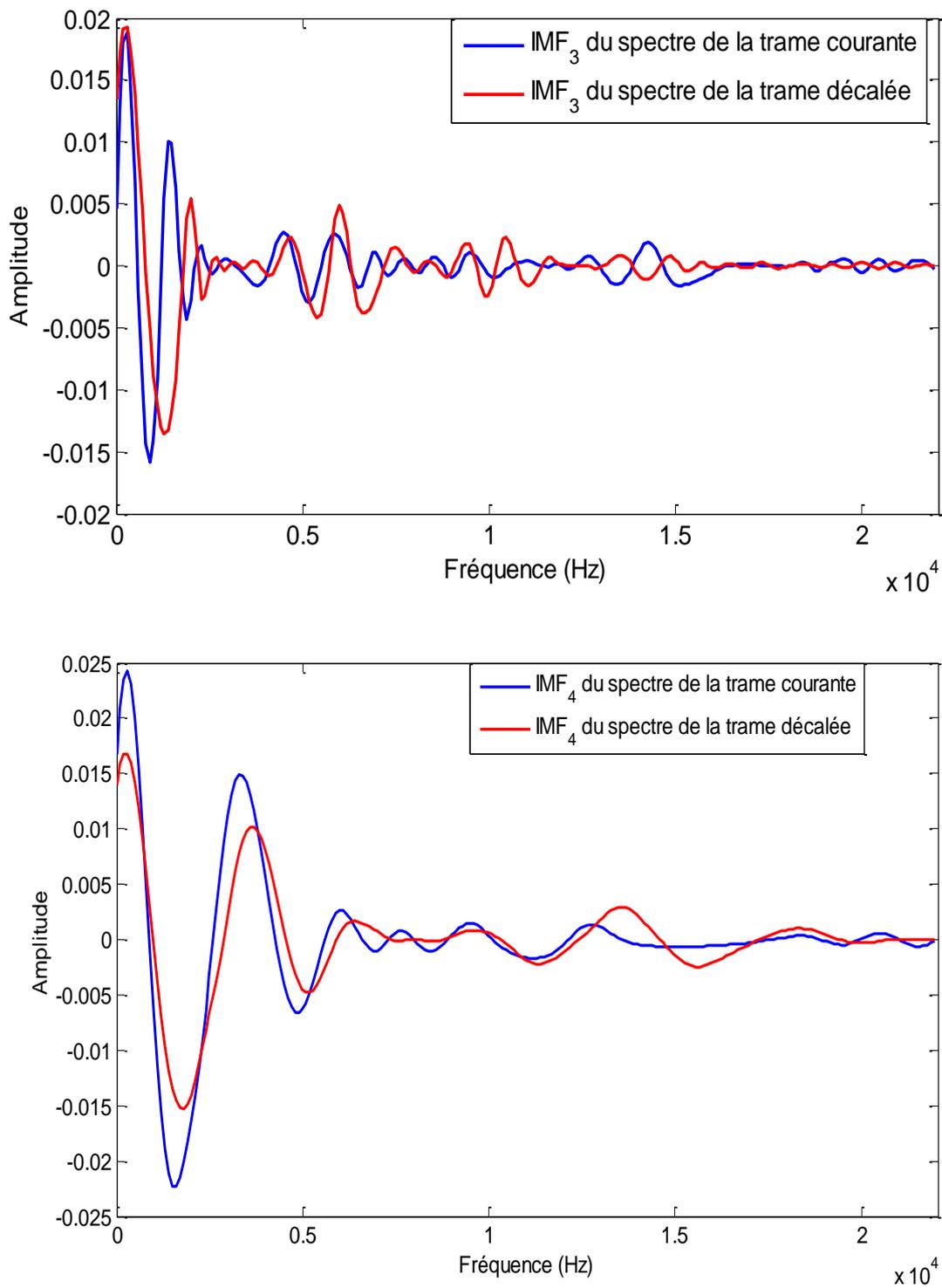


Figure IV.2 (suite)



Quatre premières IMFs des spectres des trames courante et décalée

Figure IV.2 (suite)

IV.5 Résultats et discussions

Le corpus de voyelles [a] produites par des locuteurs normophoniques et dysphoniques a été utilisé pour investiguer la performance de la méthode d'analyse multivariée des dyspériocités vocales et la comparer à celle du variogramme généralisé. L'analyse est effectuée trame par trame sans chevauchement entre les trames successives et l'algorithme de décomposition modale empirique est appliqué aux spectres d'amplitude des trames. Dans le calcul de l'indice acoustique, la bande de fréquence a été limitée à 1000 Hz. La fréquence fondamentale de chaque stimulus a été estimée en utilisant le logiciel Praat [48] et le délai de la trame décalée a été fixé à la période fondamentale. Lors du calcul du spectre de la trame, le nombre de points fréquentiels a été pris égal à deux fois la longueur de la trame. Pour investiguer l'effet de la longueur de la trame, l'analyse a été effectuée en utilisant des trames de 5 ms, 7.5 ms et 10 ms.

La figure IV.3 montre les quartiles des rapports signal à dyspériodicité locaux des quatre premières IMF pour un locuteur normophonique et un locuteur dysphonique qui a été assigné un score moyen de 2.8. On observe que la médiane des rapports signal à dyspériodicité locaux du locuteur normophonique est plus grande que celle du locuteur dysphonique. La différence entre les rapports signal à dyspériodicité locaux est plus importante pour les deux premières IMFs.

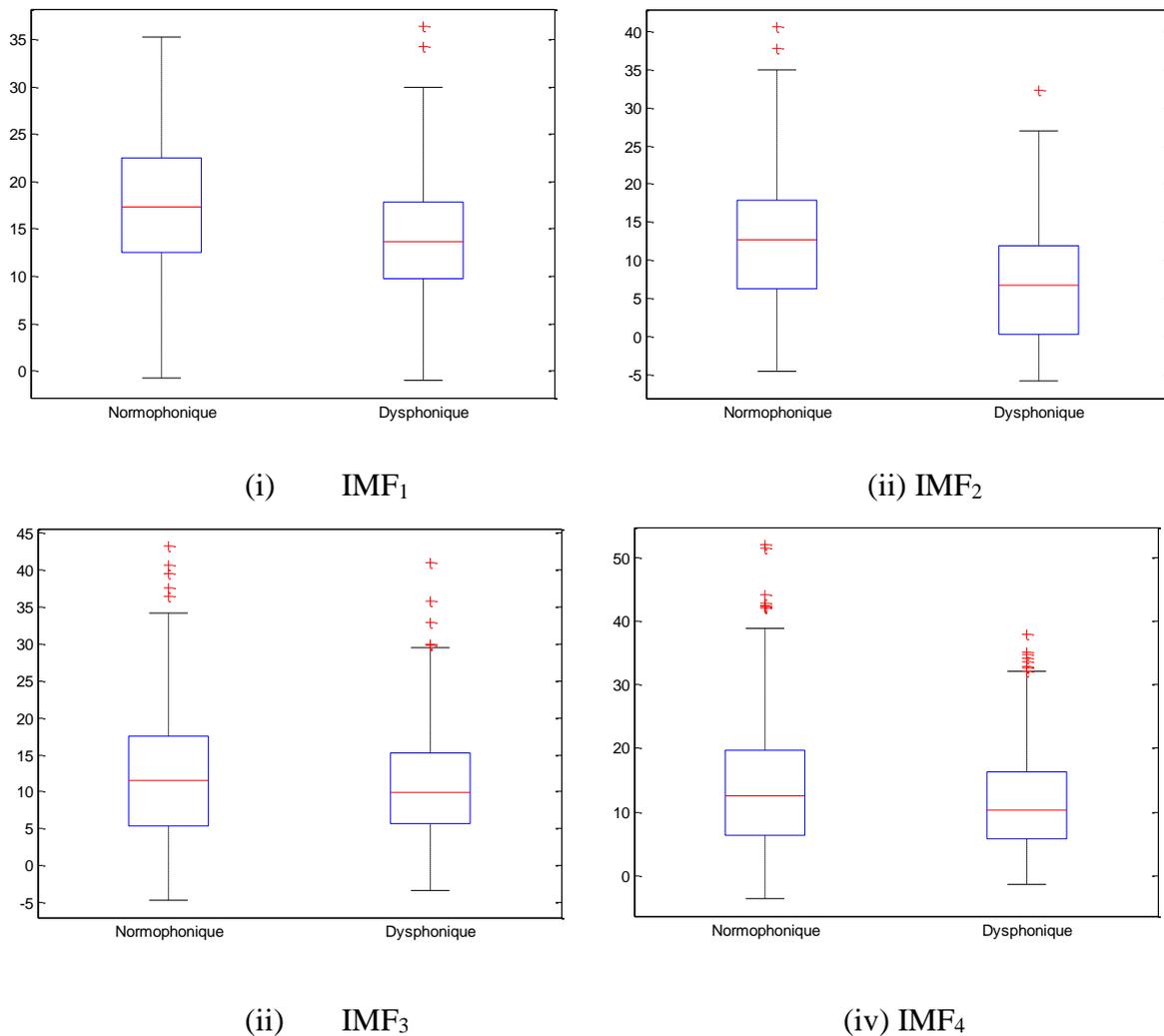


Figure IV.3- Boîte à moustaches des rapports signal à dyspériodicité locaux des quatre premières IMFs pour un locuteur normophonique et eu locuteur dysphonique.

Le tableau IV.1 montre les valeurs du coefficient de corrélation entre le rapport signal à dyspériodicité et les scores moyens de l'évaluation auditive des quatre premières IMFs pour différentes valeurs de la longueur de la trame d'analyse. La dernière colonne du tableau reporte le coefficient de corrélation multiple. La plus forte corrélation multiple est obtenue pour une trame d'analyse de longueur 7.5 ms. Le tableau IV.2 montre les corrélations entre les scores moyens de l'évaluation auditive et les rapports signal à dyspériodicité obtenus avec le variogramme généralisé. On remarque que l'analyse multivariée basée sur la décomposition en

modes empirique présente une meilleure performance en termes de corrélation puisque la plus forte corrélation obtenue avec le variogramme généralisé est de -0.67.

Tableau IV.1- Valeurs du coefficient de corrélation entre le rapport signal à dyspériodicité et les scores moyens de l'évaluation auditive des quatre premières IMFs pour différentes valeurs de la longueur de la trame d'analyse.

Longueur de la trame	Coefficient de corrélation				
	R _{IMF1}	R _{IMF2}	R _{IMF3}	R _{IMF4}	R _{multiple}
5 ms	-0.66	-0.69	-0.496	-0.48	0.71
7.5 ms	-0.67	-0.66	-0.66	-0.46	0.72
10 ms	-0.67	-0.64	-0.62	-0.45	0.71

Tableau IV.2- Valeurs du coefficient de corrélation entre les rapports signal à dyspériodicité obtenus avec le variogramme et les scores moyens de l'évaluation auditive pour différentes valeurs de la longueur de la trame d'analyse.

	Longueur de la trame		
	5 ms	7.5 ms	10 ms
Coefficient de corrélation	-0.66	-0.67	-0.67

La figure IV.4 montre les rapports signal à dyspériodicité en fonction des scores moyens pour les quatre premières IMF lorsque la longueur de la trame est fixée à 7.5 ms. On observe visuellement que les deux premières IMFs sont caractérisées par une plus forte tendance de décroissance linéaire des rapports signal à dyspériodicité lorsque les scores moyens de l'évaluation auditive augmentent.

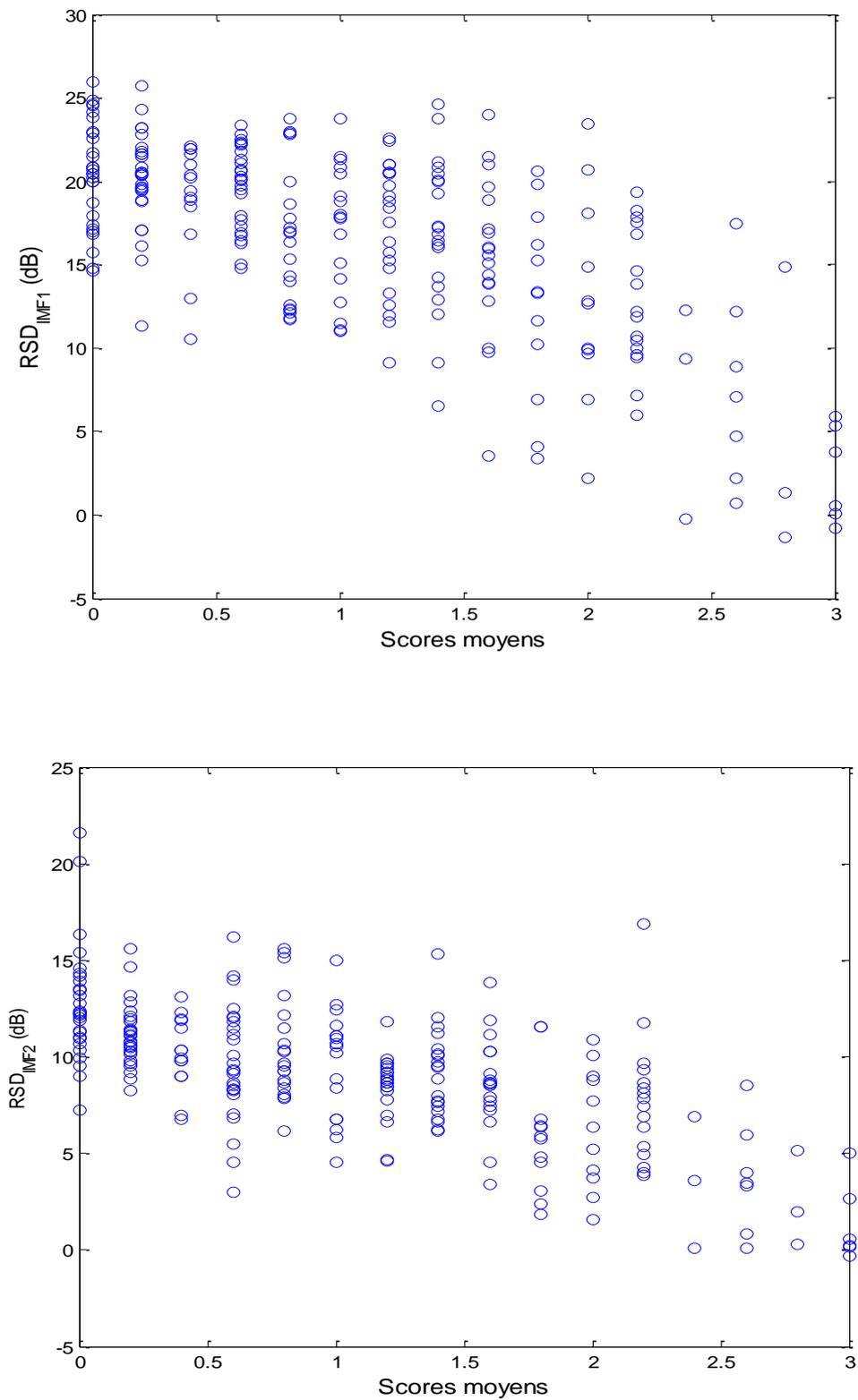


Figure IV.4- Rapports signal à dyspériodicité en fonction des scores moyens pour les quatre premières IMF. La longueur de la trame est fixée à 7.5 ms.

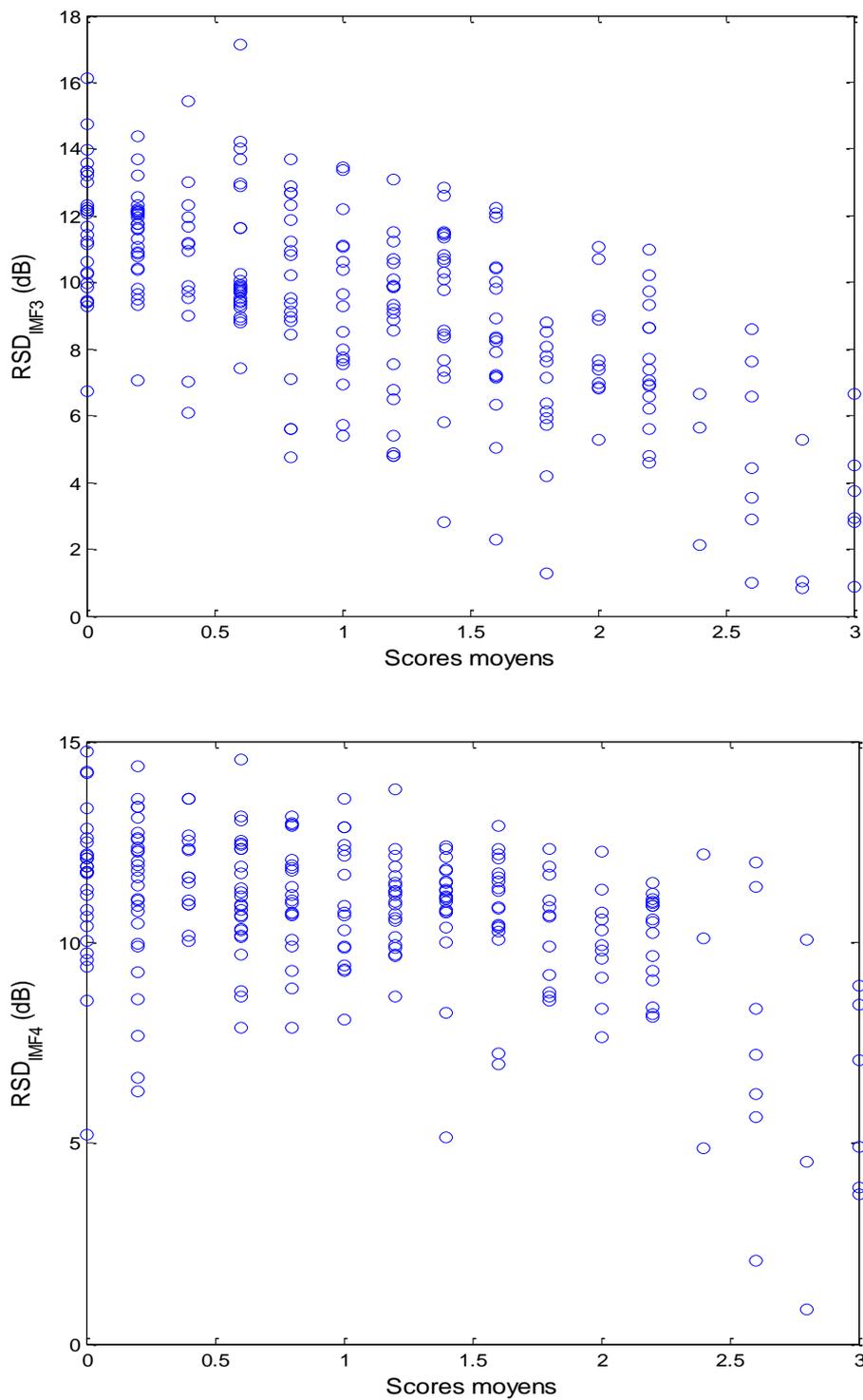


Figure IV.4 (suite)

La figure IV.5 montre les scores prédits en utilisant les rapports signal à dyspériodicité des quatre premières IMFs en fonction des scores moyens de l'évaluation auditive pour une

longueur de trame de 7.5 ms. On observe que les scores prédits tendent à augmenter lorsque les scores moyens de l'évaluation auditive augmentent et cette forte corrélation est traduite par la valeur $R_{\text{multiple}}=0.72$.

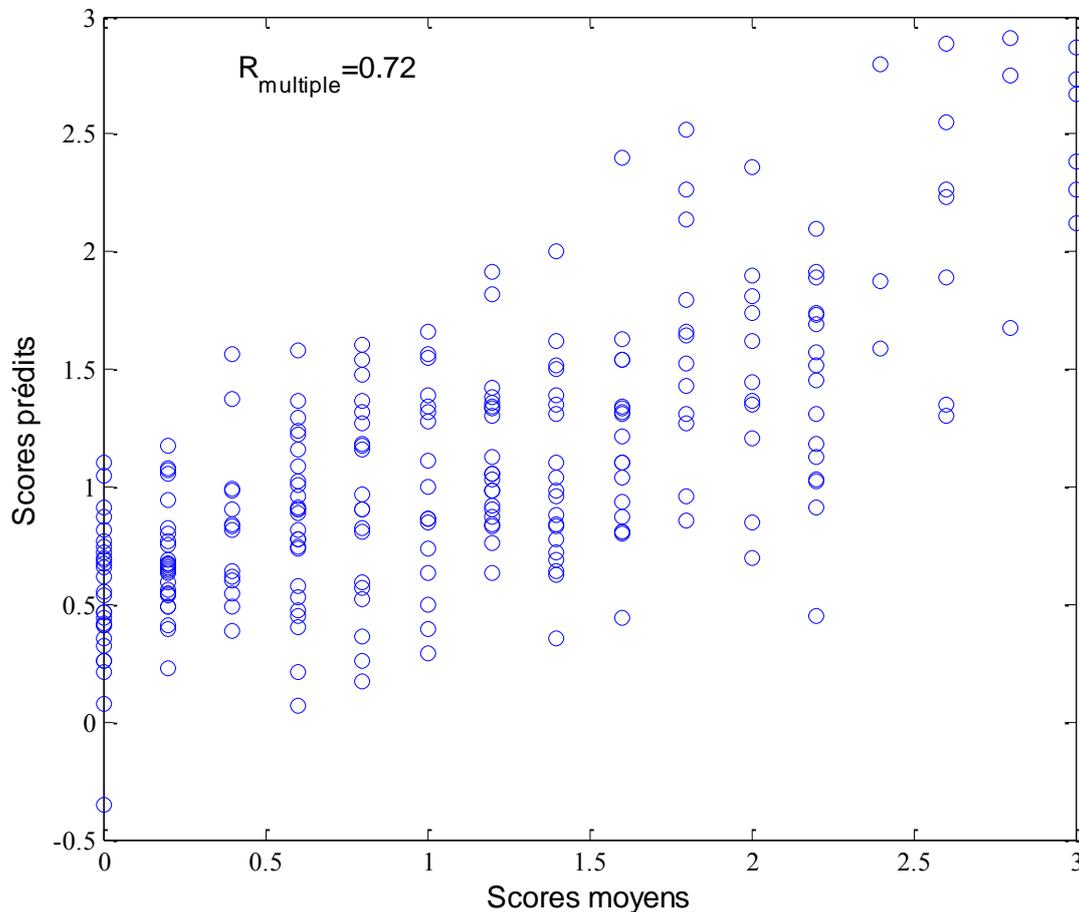


Figure IV.5- Scores prédits en utilisant les rapports signal à dyspériodicité des quatre premières IMFs en fonction des scores moyens de l'évaluation auditive pour une longueur de trame de 7.5 ms.

IV.6 Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de l'analyse multivariée des dyspériodicités vocales dans les voyelles [a] produites par des locuteurs normophoniques et dysphoniques en vue de l'évaluation objective de la qualité de la voix. L'analyse multivariée basée sur la décomposition modale empirique dans le domaine spectral présente une performance meilleure que celle de la méthode de référence basée sur le variogramme généralisé en termes de corrélation de l'indice acoustique avec les scores de l'évaluation auditive.

C

onclusion générale

Conclusion générale

Le travail réalisé s'inscrit dans le cadre d'une analyse multivariée de la parole pathologique pour l'évaluation objective de la qualité vocale. L'objectif de l'analyse multivariée est de combiner plusieurs indices acoustiques pour l'estimation des dyspériodicités vocales en utilisant la décomposition en modes empiriques du spectre d'amplitude du signal de la parole. L'analyse est effectuée trame par trame et l'algorithme de décomposition modale empirique est appliqué au spectre d'amplitude de chaque trame afin de le décomposer en fonctions de modes intrinsèques (IMFs). Un rapport signal à dyspériodicité vocale est ensuite calculé pour chaque IMF.

La méthode proposée a été appliquée à un corpus de voyelles [a] produites par 223 locuteurs normophoniques et 28 dysphoniques. Une analyse de régression linéaire multiple combinant les rapports signal à dyspériodicité des différentes IMFs a été utilisée pour prédire les scores de l'évaluation auditive et ses performances en termes de corrélation ont été comparées à celle de la méthode de référence basée sur le variogramme généralisé. Les résultats expérimentaux ont montré que l'analyse multivariée basée sur la décomposition modale empirique présente une meilleure performance comparée à la méthode de référence basée sur le variogramme généralisé.

Comme perspectives, les points suivants peuvent être suggérés :

- ✚ Application de la méthode d'analyse multivariée basée sur la décomposition modale empirique à la parole continue.
- ✚ Utilisation d'une méthode basée sur la maximisation de la fonction d'intercorrélation entre l'IMF du spectre de la trame courante et celle du spectre de la trame décalée pour la recherche du délai optimal comme alternative à l'estimation de la fréquence fondamentale.

Références bibliographiques



RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] J. Kreiman, B. R. Gerrat, “Validity of rating scale measure of voice quality”, *J. Acoust. Soc. Am.*, Vol. 104, pp. 1598-1608, 1998.
- [2] J. Kreiman, D. Vanlancker-Sidtis, B. R. Gerrat, “Defining and measuring voice quality”, *Workshop on voice quality*, pp. 115-120, Geneva, Switzerland, Aug. 2003.
- [3] F. Bettens, F. Grenez, J. Schoentgen, “Estimation of vocal dysperiodicities in connected speech by means of distant-sample bi-directional linear predictive analysis”, *J. Acoust. Soc. Am.*, Vol. 117, pp. 328-337, 2005.
- [4] A. Kacha, F. Grenez, J. Schoentgen, “Estimation of dysperiodicities in disordered speech”, *Speech Commun.* Vol. 48, pp. 1365-1378, 2006.
- [5] N.E. Huang, et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis”, *Proc. R. Soc. London Ser. A* 454 pp. 903–995, 1998.
- [6] A. Kacha, F. Grenez, J. Schoentgen, “Multiband vocal dysperiodicities analysis using empirical mode decomposition in the log-spectral domain”, *Biomedical signal processing and control*, Vol. 17, pp. 11-20, 2015.
- [7] S. Pinto and M. Sato, *Traité de neurolinguistique: Du cerveau au langage*, De Boeck Supérieur, Belgique, 2016.
- [8] G. Richard, *Traitement de la parole, Notes de cours*, Télécom Paris, 2010.
- [9] I. Laaridh, *Evaluation de la parole dysarthrique : Apport du traitement automatique de la parole face à l’expertise humaine*, Thèse de doctorat, Université d’Avignon et des pays de Vaucluse.
- [10] P. Léon, and P. Bhatt, *Structure du français moderne : Introduction à l’analyse linguistique*, 4^e édition, Canadian Scholars, Canada, 2017.

- [11] X. Renard and L. Collet, Précis d'audioprothèse : Production, phonétique acoustique et perception de la parole, Elsevier Masson, Paris, 2011.
- [12] P. Lieberman and S. E. Blumstein, Speech Physiology, Speech Perception, and Acoustic Phonetics, Cambridge University Press, Great Britain, 1988.
- [13] S. Tazarella and A. Mamecier, Perception et communication chez les animaux, De Boeck Supérieur, Belgique, 2005.
- [14] A.R. Jayan, Speech and Audio signal Processing, PHI Learning Private Limited, Delhi, 2016.
- [15] P.G Fontollet, système de télécommunications, PPUR presses polytechniques, Lausanne, vol.XVIII, 1996.
- [16] M.R Schroeder, H. Quast and H.w. Strube, computer Speech: Recognition, Compression, Synthesis, 2^e edition, Springer Science and Business Media, Germany, 2004.
- [17] M. Flohberger, Source filter-model. – Acoustic tube models.- Linear prediction.- Formant Synthesizer, Technische Universität Graz, 2003.
- [18] L. Cnockaert, Analyse du tremblement vocal et application à des locuteurs parkinsoniens, Thèse de doctorat, Université Libre de Bruxelles, 2007.
- [19] R. Thanki, K. Borisagar and S. Borra, Advance Compression and Technique for Speech Signals, Springer, India, 2017.
- [20] J.R Lapaire and W. Rotgé, Séminaire pratique de linguistique anglaise, Presses Univ. du Mirail, Toulouse, 1993.
- [21] L. Favier, Visualisation des caractéristiques de la parole dans le cadre de l'étude des interactions vocales avec un assistant virtuel, Rapport d stage 2A, Institut de recherche en Informatique de Toulouse, 2018.
- [22] M. Fortunata, Notions élémentaires sur l'expression orale du français, L'Harmattan, Paris, 2018.
- [23] J.L Roulin, Psychologie cognitive, 2^e édition, Bréal, Rome, 2016.
- [24] A.E. Aronson, Clinical Voice Disorders : An Interdisciplinary Approach, 3^e édition, Thieme, New York, 1990.
- [25] <https://www.institut-numerique.org>

- [26] P.T. Fogle, Essentials of communications Sciences and disorders, 2^e edition, Jones and Bartlett Learning, United States of America, 2017.
- [27] T. Balasubramanian, vocal cord paralysis current management trends, Geetha R, 2010.
- [28] S. de Corbiere, E. Fresnel, C. Freche, ‘‘ la voix : La corde vocale et sa pathologie’’, Collège international de médecine et chirurgie de l’American hospital of paris, France, 2001.
- [29] M. Boudjerda, Analyse du signal de la parole pour l’évaluation automatique des voix pathologiques, Thèse de doctorat, Département d’Electronique, Université de Jijel, 2018.
- [30] N. Di Palma, A. Batailly and M. Legrand, Méthodes de traitement du signal par décomposition en modes empiriques, [Rapport de recherche] Université McGill, 2018.
- [31] J.C CEXUS, A.O. BOUDRAA and S. BENRAMDANE, ‘‘Débruitage des signaux par approche EMD: Multi-EMD_{SG}’’, Institut de Recherche de l’École Navale, IRENav (EA 3634) École Navale, Lanvéoc-Poulmic, France.
- [32] O. Niang, M. Ould Guerra, A. Thioune, E. Deléchelle, and J. Lemoine et al, ‘‘A propos de l’orthogonalité dans la décomposition modale empirique’’, XXIIIème colloque GRETSI, Bordeaux (France), 2011.
- [33] A. Raad, S. Kass and J. Antoni, ‘‘ Etude de la décomposition en mode empirique combinée dans le cas des signaux de vibrations de roulements et d’engrenages dans diverses conditions’’, CFA 2016/ VISHON, Le Mans, 2016.
- [34] O. Beya, B. Jalil, E. Fauvet, O. Laligant, and L. Lewyanvoon, ‘‘Application de la décomposition modale empirique à l’étude de signaux cardiaques’’, Colloque nationale de la recherche en IUT CNRIUT’09, Lille (France), 2009.
- [35] O. BEYA, Analyse et reconnaissance de signaux vibratoires : Contribution au traitement et à l’analyse des signaux cardiaques pour la télémédecine, thèse de doctorat, Université de Bourgogne, 2014.
- [36] M. Kerebel, L. Pellissier and D. Stan, L’empirical mode décomposition, algorithmes et utilisations, mémoire de fin d’étude, Université de Cachan, 2011.
- [37] F. Mhamdi, M. Jaïdane, ‘‘Décomposition modale empirique et densité de probabilité des IMFs’’, Unité Signaux et Systèmes, Ecole Nationale d’Ingénieurs de Tunis, 2018.

- [38] S. Haddad, M. Grimes, T. Benkdidah, M. Mekideche, A. Bouhadjera, “La décomposition modale empirique pour la détection d’échos ultrasonores multiples”, NDT-Lab, Université de Jijel, Algérie.
- [39] A. Girard, “Réduction de bruit de signaux de parole mono-capteur basée sur la modélisation par EMD”, Mémoire de maîtrise, Département de génie électrique et de génie informatique, Université de Sherbrooke, 2010.
- [40] N.E Huang, Z. Shen, S.R. Long, M.C. Wu and H.H Shih et al, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”, The Royal Society vol. 454 (3), pp. 903-995, 1998.
- [41] M. Kemiha, Etude de l’algorithme de décomposition modale empirique (EMD) : application au débruitage et à la caractérisation des signaux de parole, Mémoire de magister, Département d’Automatique, Université de Jijel, 2011.
- [42] K. Khaldi, Traitement et analyse des signaux sonores par transformée de Huang (EMD), Thèse de doctorat, Département du signal et communications, Université européenne de Bretagne, 2012.
- [43] T. Oberlin, Analyse de signaux multicomposantes, Thèse de doctorat, Université de Grenoble, 2013.
- [44] A. Thioune, Décomposition modale empirique et décomposition spectrale intrinsèque : Application en traitement du signal et de l’image, Thèse de doctorat, Université de Paris Est Val-de-Marne Créteil, 2015.
- [45] T. Oberlin, Approches mathématiques pour la décomposition modale empirique, Rapport de stage, Université Joseph Fourier, Grenoble.
- [46] Y. Maryn, P. Corthals, P. V. Cauwenberge, N. Roy and M. D. Bodt, “Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels,” J. Voice, vol. 24, pp. 540–555, 2010.
- [47] M. Hirano, “The laryngeal examination,” in Speech Evaluation in Medicine, J. Darby, Ed. New York: Grune and Stratton, 1981.
- [48] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer, 2007, <http://www.praat.org>.