

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L 'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE



N° d'ordre :.....

Série :.....

Université Mohammed Seddik

BENYAHIA - Jijel

Faculté des Sciences et de la Technologie

Département : Electronique

Thèse Présentée

Pour l'obtention du Diplôme de Doctorat en Sciences

Spécialité : Electronique

Par :

Mounir BOUDJERDA

Thème :

Analyse du signal de parole pour l'évaluation
automatique des voix pathologiques

Soutenue le : **18 / 01 / 2018** , Devant le Jury composé de :

Mr. Ammar SOUKKOU	MCA	U. M.S.B.Jijel	Président
Mr. Abdellah KACHA	Professeur	U. M.S.B.Jijel	Rapporteur
Mr. Moussa BENYOUCEF	Professeur	U. M.B.Batna 2	Examineur
Mr. Rédha BENZID	Professeur	U. M.B.Batna 2	Examineur
Mr. Abdelhamid DAAMOUCHE	MCA	U. M.B. Boumerdes	Examineur

Remerciements

Ce travail de recherche s'est déroulé au sein du laboratoire de Physique de Rayonnement et Applications à la Faculté des Sciences et de la Technologie de l'université de Jijel.

D'abord, nous remercions Dieu le tout puissant. C'est grâce à lui que nous avons eu la force pour accomplir ce travail.

J'adresse mes plus vifs remerciements à Mr A. Kacha, Professeur à l'université de Jijel, mon directeur de thèse pour ses précieux conseils, ses incessants encouragements et surtout sa grande disponibilité tout au long de la réalisation de ce travail. Je le remercie pour la confiance qu'il m'a témoignée.

Mes remerciements s'adressent aussi à Mr A. SOUKKOU, Maître de conférences à l'université de Jijel, de m'avoir honoré et accepter de présider le jury de ce travail. Je tiens aussi à remercier Mr M. BENYOUCEF et Mr R. BENZID, Professeurs à l'université de Batna 2 et Mr A. DAAMOUCHE, Maître de conférences à l'université de Boumerdes, pour avoir accepté d'examiner et de juger ce travail.

J'adresse également mes remerciements à mes cher amis et collègues du centre de recherche en Technologie Industriels CRTI, en particulier Dr M. Amir, pour leur aide, leur soutien et leurs encouragements.

Mes remerciements les plus vifs, vont également à mes parents qui m'ont donné un environnement idéal durant toute mon enfance et ils m'ont enseigné l'humilité et l'honnêteté avec lesquelles j'ai essayé de mener à terme ce travail scientifique.

Enfin, mes remerciements s'adressent à tous ceux qui m'ont soutenu ou qui, d'une manière ou d'une autre, ont contribué à l'élaboration de ce modeste travail.

Résumé

L'analyse acoustique du signal de parole pour le diagnostique et l'évaluation automatique des voix pathologiques est une discipline qui nécessite encore des investigations en raison des difficultés de développer des méthodes standards pouvant aider les cliniciens et les spécialistes de la voix pour évaluer et suivre l'évolution de la voix des patients avec une bonne précision. L'analyse du signal acoustique a pour objectif d'extraire des indices pertinents permettant de déterminer les caractéristiques de la voix. L'objectif de cette thèse est le développement de méthodes d'analyse acoustique dédiées à l'évaluation automatique objective de la qualité de la voix produite par des locuteurs dysphoniques.

Dans ce travail, on s'intéresse à l'estimation des dyspériodicités vocales en utilisant la décomposition modale empirique dans le domaine log-spectral. Nous développons deux méthodes d'analyse. La première méthode est basée sur l'estimation de la source glottique via la décomposition modale empirique et la deuxième méthode est basée sur l'optimisation d'un modèle de la source glottique en utilisant la décomposition modale empirique et les algorithmes génétiques. Nous utilisons deux indices acoustiques pour quantifier les dyspériodicités vocales dans le signal de parole. Le premier indice est le rapport excitation à dyspériodicité segmental (EDRSEG). Il est inspiré de la définition de l'indice conventionnel rapport signal à dyspériodicité segmental. Le deuxième indice est le quotient d'ouverture moyen (MOQ).

Les méthodes d'analyse proposées sont appliquées à un corpus de voyelles [a] synthétiques et un corpus de la parole naturelle produite par des locuteurs normophoniques et dysphoniques et comprenant une large gamme de pathologies. Afin d'améliorer davantage les performances de la méthode d'analyse proposée, nous avons utilisé une analyse multivariée en combinant la prééminence du pic cepstral et le quotient d'ouverture moyen. La combinaison des deux indices acoustiques améliore la performance de la méthode d'estimation des dyspériodicités vocales en termes de corrélation entre l'indice acoustique et l'évaluation perceptive.

Mots Clés : Voix pathologiques, décomposition modale empirique, algorithme génétique, indices acoustiques.

Abstract

Acoustic analysis of speech for diagnosis and automatic evaluation of pathological voices is a discipline that still requires investigations because of the difficulties of developing standard methods that can help clinicians and voice specialists to evaluate and monitor the evolution of the voice of patients with good accuracy. The purpose of acoustic signal analysis is to extract pertinent markers for determining the characteristics of the voice. The objective of this thesis is the development of acoustic analysis methods devoted to the objective automatic evaluation of the quality of the voice produced by dysphonic speakers.

In this work, we are interested in the estimation of vocal dysperiodicities using the empirical mode decomposition in the log-spectral domain. We develop two methods of analysis. The first method is based on the estimation of the glottal source via the empirical modal decomposition and the second method is based on the optimization of a model of the glottal source using the empirical mode decomposition and genetic algorithms. We use two acoustic cues to quantify vocal dysperiodicities in the speech signal. The first acoustic cue is the segmental excitation-to-dysperiodicity ratio (EDRSEG). It is inspired by the definition of the conventional segmental signal-to-dysperiodicity ratio. The second acoustic cue is the mean opening quotient (MOQ).

The proposed analysis methods are applied to corpora comprising synthetic vowels [a] and natural speech produced by normophonic and dysphonic speakers with a wide range of pathologies. In order to further improve the performance of the proposed analysis method, we used a multivariate analysis combining the prominence of the cepstral peak and the mean opening quotient. The combination of the two acoustic cues improves the performance of the estimation method of voice dysperiodicities in terms of the correlation between the acoustic cue and the perceptual evaluation.

Keywords: Pathological voices, empirical mode decomposition, genetic algorithm, acoustic cues.

ملخص

يعتبر التحليل الصوتي لإشارة الكلام للتشخيص و للتقييم الآلي للأصوات المريضة فرع من المعرفة الذي لا يزال يتطلب دراسات كبيرة بسبب صعوبة تطوير طرق القياس التي يمكن أن تساعد الأطباء وخبراء النطق لتقييم ورصد تطور صوت المرضى بدقة جيدة . الهدف من تحليل الإشارات الصوتية هو استخراج المؤشرات ذات الصلة لتحديد خصائص الصوت. إن الهدف من هذه الأطروحة هو تطوير طرق التحليل الصوتي المكرسة للتقييم الآلي لنوعية الصوت الناتجة عن متحدثين مرضى.

في هذا العمل، نهتم بتقدير الخلل في الدور الصوتي باستخدام التحلل النمطي التجريبي في مجال لوغاريتم-الطيفي. لقد طورنا طريقتين للتحليل. الطريقة الأولى تركز على تقدير المصدر المزماري عن طريق التحلل النمطي التجريبي وتتركز الطريقة الثانية على البحث على القيم المثلثية لنموذج لمصدر مزماري باستخدام التحلل النمطي التجريبي والخوارزميات الجينية. اقترحنا مؤشرين صوتيين لقياس كمية الخلل في الدور الصوتي . المؤشر الأول هو النسبة الإثارة على الخلل في الدور الصوتي وهو مستوحى من تعريف المؤشر الصوتي التقليدي النسبة الإشارة على الخلل في الدور الصوتي. المؤشر الثاني هو متوسط حاصل الافتتاح. طرق التحليل المقترحة طبقت على مجموعة من أحرف العلة الاصطناعية [a] ومجموعة من الكلام الطبيعي الناتجة عن متحدثين عاديين و متحدثين مرضى حيث شملت هذه الأخيرة مجموعة واسعة من الأمراض.

من أجل زيادة تحسين طريقة التحليل المقترحة، استخدمنا تحليل متعدد المتغيرات من خلال الجمع بين المؤشرين الصوتيين بروز الذروة سيسترال ومتوسط حاصل الافتتاح. إن الجمع بين المؤشرين الصوتيين يحسن أداء طريقة تقدير الخلل في الدور الصوتي من حيث الارتباط بين المؤشر الصوتي والتقييم الإدراك الحسي.

الكلمات المفتاحية : الأصوات المريضة، التحلل النمطي التجريبي، الخوارزميات الجينية، المؤشرات الصوتية.

Table des matières

Remerciements.....	i
Résumé.....	ii
Table des matières.....	v
Liste des abréviations.....	ix
Liste des figures.....	x
Liste des tableaux.....	xii
Introduction Générale.....	01
Chapitre I : Etat de l'art de l'évaluation des troubles de la voix.....	06
I.1. Introduction.....	07
I. 2. Production de la parole.....	07
I. 2. 1. Anatomie de l'appareil phonatoire humain.....	07
I. 2. 2. Physiologie du système de production de la parole.....	10
I. 3. Modèle de production de la parole.....	10
I. 3. 1. Modèle source-filtre.....	10
I. 3. 2. Modélisation de la source glottique.....	12
I. 3. 2. 1. Modèle général de la source glottique.....	12
I. 3. 2. 2. Modèle LF.....	14
I. 4. Pathologies du larynx.....	16
I. 4. 1. Dysphonies organiques.....	16
I. 4. 2. Dysphonies fonctionnelles.....	17
I. 5. Evaluation des troubles de la voix.....	17
I. 5. 1. Evaluation perceptive.....	17
I. 5. 2. Evaluation objective.....	19
I. 5. 2. 1. Evaluation des troubles de la voix via le signal de la parole.....	20
I. 5. 2. 2. Evaluation des troubles de la voix basée sur l'estimation de la source glottique.....	26
I. 6. Conclusion.....	29
Chapitre II : La décomposition modale empirique.....	30

II. 1. Introduction	31
II. 2. Principe de la décomposition modale empirique.....	31
II. 3. Algorithme de la décomposition modale empirique.....	32
II. 4. Exemple d'illustration de l'algorithme de la décomposition modale empirique.....	33
II. 5. Implantation de la décomposition modale empirique.....	36
II. 5. 1. Interpolation des maxima	36
II. 5. 2. Critère d'arrêt du processus de tamisage	36
II. 6. Quelques caractéristiques fondamentales de la méthode EMD.....	37
II. 6. 1. Non-linéarité.....	37
II. 6. 2. Localité	38
II. 6. 3. Multirésolution	38
II. 6. 4. Auto-adaptativité	38
II. 7. Limitations de la décomposition modale empirique.....	38
II. 7. 1. Problème lié au choix du critère d'arrêt.....	38
II. 7. 2. Problème d'absence d'expression analytique	39
II. 7. 3. Problème de mixage des modes.....	39
II. 8. Autres variantes de la décomposition modale empirique.....	39
II. 9. Conclusion	39
Chapitre III : Les algorithmes génétiques	41
III. 1. Introduction	42
III. 2. Terminologie biologique	42
III. 3. Principe général de fonctionnement des algorithmes génétiques.....	44
III. 4. Description détaillée des algorithmes génétiques	46
III. 4. 1. Codage des données	46
III. 4. 1. 1. Codage binaire.....	46
III. 4. 1. 2. Codage réel.....	46
III. 4. 2. Fonction d'évaluation	46
III. 4. 3. Génération de la population initiale	47
III. 5. Opérateurs génétiques	47
III. 5. 1. Sélection	47
III. 5. 2. Croisement	48

III. 5. 3. Mutation	49
III. 6. Conclusion.....	49
Chapitre IV: Estimation des dyspériodicités vocale par la décomposition modale empirique	50
IV. 1. Introduction.....	51
IV. 2. Analyse des dyspériodicités vocales basée sur la décomposition modale empirique dans le domaine log-spectrale	51
IV. 2. 1. Séparation des composantes du signal parole par EMD.....	51
IV. 2. 2. Exemple de séparation des composantes du signal parole par la décomposition modal empirique.....	53
IV. 2. 3. Rapport harmonique sur bruit	58
IV. 3. Evaluation des troubles de la voix basée sur l'estimation de la source glottique	58
IV. 3. 1. Estimation du signal de la source glottique	58
IV. 3. 2. Estimation des dyspériodicités vocale	61
IV. 3. 3. Rapport excitation à dyspériodicité segmental	63
IV. 4. Evaluation des troubles de la voix basée sur l'optimisation d'un modèle de la source glottique.....	63
IV. 4. 1. Méthode	64
IV. 4. 2. Optimisation des paramètres du modèle de la source glottique par les algorithmes génétiques	65
IV. 4. 3. Quotient d'ouverture moyen.....	65
IV. 4. 4. Combinaison de la méthode proposée avec la méthode cepstral	67
IV. 5. Conclusion	67
Chapitre V : Résultats expérimentaux	69
V. 1. Introduction	70
V. 2. Corpus et évaluation perceptive.....	70
V. 2. 1. Voyelles synthétiques.....	70
V. 2. 2. Parole naturelle.....	71
V. 3. Résultats	71
V. 3. 1. Rapport excitation à dyspériodicité segmental.....	71
V. 3. 1. 1. Corpus de voyelles synthétiques	72

V. 3. 1. 2. Corpus de parole naturelle.....	76
V. 3. 2. Quotient d'ouverture moyen	79
V. 3. 3. Combinaison des indices acoustiques quotient d'ouverture moyen et prééminence du pic cepstral.....	87
V. 4. Conclusion.....	88
Conclusion générale	90
Références bibliographiques	93

Liste des Abréviations

CPP	Cepstral Peak Prominence
EAI	Equal-Appearing-Interval
EAV	Echelle Analogique Visuelle
ECG	Electrocardiographie
EDRSEG	Segmental Excitation-to-Dysperiodicity Ratio
EEG	Electro-Encéphalographie
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
GA	Genetic algorithm
GCI	Glottal Closure Instant
GRBAS	Grade Rough Breathy Asthenic Strained
HNR	Harmonic-to-Noise Ratio
IMF	Intrinsic Mode Function
LF	Lijlencrants-Fant
MOQ	Mean Opening Quotient
R1A	Amplitude du premier Rhamonique du cepstre
SD	Standard Deviation
SDR	Signal-to-Dysperiodicity Ratio
SDRSEG	Segmental signal-to-Dysperiodicity Ratio
SJE	Spectral Jitter Estimator
TAP	Traitement Automatique de la Parole

Liste des figures

Figure I.1. Anatomie de l'appareil phonatoire humain.....	08
Figure I.2. Anatomie du larynx.....	09
Figure I.3. Modèle général du système de production de la parole.....	11
Figure I.4. Modèle de débit glottique (a) et son dérivé (b).....	14
Figure I.5. Analyse cepstrale d'un segment extrait d'une voyelle [a] produite par un locuteur normophonique.....	22
Figure I.6. Analyse cepstrale d'un segment extrait d'une voyelle [a] produite par un locuteur dysphonique.	23
Figure I.7. Système complet de traitement de la source glottique, à partir du signal de parole jusqu'à l'intégration des informations glottiques dans une application de technologie vocale (applications biomédical).....	28
Figure II.1. Organigramme de l'algorithme de la décomposition modale empirique.....	34
Figure II.2. Exemple de décomposition d'un signal en quatre fonctions de mode intrinsèque et un résidu.....	35
Figure III.1. Niveaux d'organisation d'un algorithme génétique.....	44
Figure III.2. Organigramme général de l'algorithme génétique.....	45
Figure IV.1. Décomposition modale empirique du logarithme du spectre d'amplitude d'une trame de longueur 200 ms extraite d'une voyelle [a] soutenue produite par un locuteur dysphonique.....	55
Figure IV.2. Décomposition du logarithme du spectre d'amplitude d'une trame de longueur 200 ms extraite d'une voyelle [a] soutenue en trois composants via l'algorithme EMD.	57
Figure IV.3. Estimation du signal d'excitation.....	59
Figure IV.4. Signal d'excitation original superposé au signal de source glottique estimé pour une trame de 50 ms extraite d'une voyelle soutenue synthétique [a] obtenue par un locuteur normophonique	61
Figure IV.5. Procédure d'estimation du quotient d'ouverture moyen MOQ à partir du signal acoustique.....	64
Figure IV.6. Estimation des paramètres de la source glottique par les GAs.....	66

Figure V.1. Rapport excitation à dyspériodicité segmental en fonction des scores moyens (a) du grade, (b) du souffle et (c) de la raucité perçue pour les voyelles [a] synthétiques.....	73
Figure V.2. Proéminence du pic cepstral en fonction des scores moyens (a) du grade, (b) du souffle et (c) de la raucité perçue pour les voyelles [a] synthétiques.....	74
Figure V.3. Rapport harmonique sur bruit en fonction des scores moyens (a) du grade, (b) du souffle et (c) de la raucité perçue pour les voyelles [a] synthétiques.....	75
Figure V.4. Rapport excitation à dyspériodicité segmental en fonction du score moyen du grade perçu pour les trois corpus	78
Figure V.5. Signaux estimés de la source glottique correspondant à deux trames de 200 ms extraites de voyelles soutenues [a] produites par (a) un locuteur normophonique et (b) un locuteur dysphonique.....	80
Figure V.6. Signal de la source glottique estimé via la méthode basée sur l'EMD superposé au signal de la source glottique optimisé via le modèle LF et les GAs pour une longueur de trame égale à trois fois la période fondamentale extraite d'une voyelle [a] soutenue produite par (a) un locuteur normophonique et (b) un locuteur dysphonique.....	81
Figure V.7. Quotient d'ouverture moyen en fonction des scores moyens du grade perçu pour les trois corpus de parole naturelle	84
Figure V.8. Proéminence du pic cepstral en fonction des scores moyens du grade perçu pour les trois corpus de parole naturelle	86

Liste des tableaux

Tableau I.1. Paramètres de l'échelle GRBAS : signification et définition.....	19
Tableau III.1. Représentation des niveaux d'organisation d'un algorithme génétique.....	44
Tableau V.1. Corrélations des indices acoustiques EDRSEG, HNR et CPP avec les scores moyens du grade, du souffle et de la raucité perçus.....	76
Tableau V.2. Corrélations des indices acoustiques EDRSEG, HNR et CPP avec le score moyen du grade perçu pour les trois corpus.....	78
Tableau V.3. Moyenne et écart type de la fréquence et de la période fondamentales, valeurs optimisées des paramètres du modèle de la source glottique, MOQ et fonction d'évaluation pour les deux locuteurs normophonique et dysphonique.....	82
Tableau V.4. Coefficient de corrélation du quotient d'ouverture moyen et de la prééminence du pic cepstral avec les scores moyens du grade perçu pour les trois sous-corpus de parole naturelle.....	87
Tableau V.5. Coefficients de corrélation des MOQ et CPP avec les scores moyens du grade perçu, coefficients de régression multiple et corrélation multiple pour les trois sous-corpus de parole naturelle	88

Introduction générale

Introduction générale

La parole est l'un des moyens les plus importants en communication. En effet, la voix est considérée parmi les outils professionnels importants pour plusieurs fonctions. Par conséquent, l'utilisation intensive ou abusive de la voix peut causer des troubles de la voix et affecter la santé professionnelle et la vie sociale en général ainsi que l'activité professionnelle. Les troubles de la voix affectent diverses catégories sociales. Chez les enfants, c'est surtout à partir de 2 ans qu'on va percevoir la dysphonie, le plus souvent à la suite d'un malmenage vocal. Chez les adultes pour qui l'usage de la voix s'impose pour des raisons professionnelles ou sociales, on voit le plus des dysphonies appelées dysfonctionnelles. Dans cette catégorie, on distingue les employés des centres d'appel, les enseignants, les entraîneurs sportifs, les avocats, les chanteurs, etc. Ces patients se plaignent de fatigue vocale, de raucité, de gêne ou de douleurs pharyngées alors que l'examen du larynx ne montre pas de lésions organiques.

Les troubles de la voix 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 (anatomique ou physique) suite à une mauvaise utilisation de la voix. Ces maladies peuvent être traitées par une thérapie vocale. Une pathologie organique implique l'existence d'une lésion de l'appareil phonatoire (présence de lésion sur les cordes vocales par exemple) [1]. 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.

Généralement, les méthodes utilisées pour identifier et évaluer les voix pathologiques sont basées sur l'évaluation perceptive effectuée par des experts. Cependant, la nature subjective de l'évaluation perceptive a soulevé la question concernant la fiabilité et la validité de l'évaluation auditive-perceptive de la qualité vocale. Malgré la subjectivité et la variabilité de l'évaluation perceptive, elle est encore considérée comme une référence. En plus, les méthodes courantes d'analyse du signal de parole pour l'évaluation des voix pathologiques utilisent des signaux audio ou vidéo de la source laryngée (laryngoscopiques, vidéolaryngoscopique et stroboscopiques). Cependant, lors de l'utilisation de ces instruments, quelques locuteurs manifestent une action réflexe dans la cavité supra-glottique qui peut donner une fausse évaluation.

Face à ces limites, les spécialistes de la voix expriment la nécessité de méthodes pour évaluer objectivement la qualité de la voix. En effet, il y'a un lien étroit entre la qualité de la voix perçue et les caractéristiques du signal acoustique [2, 3]. Cette relation résulte du dysfonctionnement des cordes vocales ou des pathologies laryngales qui causent une dégradation du signal acoustique. Par conséquent, il serait vain de rechercher un accord acoustique-perceptuel en utilisant une analyse de corrélation entre le degré de la pathologie et la qualité de la voix perçue.

Une méthode objective est considérée d'autant plus performante qu'elle présente une forte corrélation avec l'évaluation perceptive. 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 du locuteur. En plus, une analyse acoustique offre plusieurs avantages tels que:

- Le coût peu élevé du système d'évaluation,
- La simplicité de mettre en œuvre,
- Elle documente de manière quantitative le degré d'enrouement perçu par le clinicien.

Dans la littérature, plusieurs méthodes d'évaluation objective ont été proposées, mais, la question de la robustesse et de la précision de ces méthodes reste posée. Parmi ces méthodes objectives existantes, on distingue celles qui utilisent le signal acoustique lui-même [4-6] et celles qui opèrent sur une estimation de l'excitation glottique [7-9]. Plus récemment, une méthode d'analyse des troubles de la voix basée sur la décomposition modale empirique (EMD : empirical mode decomposition) dans le domaine log-spectral a été proposée [10, 11]. En utilisant l'algorithme de l'EMD, le logarithme du spectre d'amplitude du signal de parole est décomposé en ces trois composantes qui sont la composante harmonique, l'enveloppe spectrale et le bruit. Le rapport harmonique sur bruit (HNR : harmonic-to-noise ratio) a été utilisé pour quantifier le degré de perturbation dans le signal de parole.

L'objectif du travail de recherche proposé dans le cadre de cette thèse est le développement de méthodes robustes et précises d'analyse acoustique dédiées à l'évaluation automatique objective de la qualité de la voix produite par des locuteurs dysphoniques. A cette fin, en se basant sur la méthode développée dans [10], nous proposons deux indices acoustiques ayant une plus forte corrélation avec l'évaluation perceptive que ceux existants dans la littérature. Ces indices acoustiques sont extraits à partir d'une estimation de la source

glottique. Le premier indice proposé est le rapport excitation à dyspériodicité segmental (EDRSEG : segmental excitation-to-dysperiodicity ratio). Il est inspiré de l'indice conventionnel rapport signal à dyspériodicité segmental. Le deuxième indice proposé est le quotient d'ouverture moyen (MOQ : mean opening quotient) défini à partir des paramètres d'un modèle de la source glottique obtenus par optimisation par les algorithmes génétiques. Les performances des deux méthodes proposées en termes de corrélation de l'indice acoustique avec l'enrouement perçu sont comparées à celles de deux méthodes de références. Le premier indice de référence est la proéminence du pic cepstral (CPP : cepstral peak prominence) utilisé conventionnellement dans le cadre de l'évaluation objective de la qualité de la voix. Le seconde indice de référence est le rapport harmonique sur bruit (HNR) calculé en utilisant la méthode d'analyse basée sur la décomposition modale empirique du signal de parole dans le domaine log-spectral. Les performances des différentes méthodes d'analyse sont évaluées en utilisant des corpus de voyelles [a] synthétiques et de parole naturelle. Afin d'améliorer davantage les performances des méthodes d'analyse proposées, nous utilisons une analyse multivariée combinant deux indices acoustiques.

Le manuscrit de cette thèse est organisé en cinq chapitres.

Le premier chapitre présente l'état de l'art de l'analyse et l'évaluation des voix pathologiques. On commence par un rappel de l'anatomie et la physiologie du système de production de la voix. Ensuite, nous présentons quelques pathologies du larynx. Un modèle de production de la parole sera également présenté. Enfin, nous aborderons les deux classes de caractérisation des troubles de la voix qui sont les méthodes d'analyse perceptive et les méthodes d'analyse objective.

Le deuxième chapitre présente le principe de la méthode de décomposition modale empirique, une technique qui sera utilisée pour décomposer le signal de parole en une composante harmonique et une enveloppe spectrale. La composante harmonique est utilisée par la suite pour calculer des indices objectifs permettant d'évaluer et de quantifier de manière automatique le degré de la dyspériodicité vocale. Après un rappel du principe de l'algorithme de l'EMD, nous présentons les étapes fondamentales de son implantation. Enfin, on présente les caractéristiques fondamentales ainsi que quelques limitations de la méthode d'EMD.

Le troisième chapitre présente le principe des algorithmes génétiques, une approche d'optimisation issue de l'intelligence artificielle qui sera utilisée pour optimiser le modèle de la source glottique.

Le quatrième chapitre présente les différentes méthodes proposées pour estimer les dyspériodicités vocales. Ce chapitre présente également les indices acoustiques utilisés pour quantifier les dyspériodicités vocales.

Le cinquième chapitre présente les résultats expérimentaux obtenus en utilisant les différentes méthodes d'analyse. Dans ce chapitre, on décrit les deux corpus utilisés dans ce travail ainsi que les différentes méthodes d'évaluation perceptive des stimuli. Les performances des différentes méthodes d'analyse sont comparées en calculant les corrélations des scores de l'évaluation perceptive avec les valeurs des indices acoustiques. Enfin, on se propose également d'utiliser l'analyse multivariée en combinant deux indices acoustiques qui sont la proéminence du pic cepstral et le quotient d'ouverture moyen afin d'améliorer la performance de l'indice acoustique en terme de corrélation avec l'enrouement perçu.

Enfin, dans la conclusion générale du manuscrit, nous résumons les principaux résultats et contributions de ce travail.

*Chapitre I: Etat de l'art de l'évaluation des
troubles de la voix*

I. 1. Introduction

La parole est un moyen de transmettre nos pensées, nos émotions et nos sentiments. Il est aussi une image sur notre santé et elle peut aussi être porteuse d'informations sur notre situation sociale. D'autre part, la voix est le support acoustique de la parole. Une altération ou une lésion de la voix peut donner lieu à des conséquences négatives sur notre vie quotidienne et nécessite la consultation d'un spécialiste de la voix.

La parole est produite suite à des phénomènes physiologiques, cognitifs et physiques propres à l'être humain. De point de vue acoustique, les sons voisés de la parole sont produits par la vibration des cordes vocales. Cette vibration peut être mesurée à l'aide d'un microphone et permet de représenter la parole sous la forme d'un signal acoustique qu'on peut analyser en utilisant les outils de traitement du signal. Donc, L'analyse du signal de parole permet de donner une image sur l'état du larynx (cordes vocales). Depuis quelques décennies, l'analyse acoustique du signal de parole pour la caractérisation et l'évaluation des troubles de la voix a fait l'objet d'un grand nombre de travaux de recherche [2, 3, 6]. 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 du locuteur.

Ce chapitre est une présentation de l'état de l'art de l'évaluation des troubles de la voix. On commence par une présentation de l'anatomie et la physiologie du système de production de la voix. Ensuite, on décrit quelques pathologies du larynx. Dans la suite du chapitre, on présente un modèle de production de la parole et on aborde le problème de caractérisation et d'évaluation des troubles de la voix en présentant les méthodes de l'analyse perceptive et les méthodes l'analyse objective.

I. 2. Production de la parole

I. 2. 1. Anatomie de l'appareil phonatoire humain

Les sons de la parole se produisent via le mouvement de nombreux muscles et organes de phonation. La coordination de ces organes est contrôlée par le système nerveux central [12]. Les organes de l'appareil phonatoire humain peuvent être classés en trois groupes principaux [13, 14] qui sont : les poumons et la trachée, le larynx et le conduit vocal. La figure I.1 montre une représentation simplifiée de l'appareil phonatoire humain.

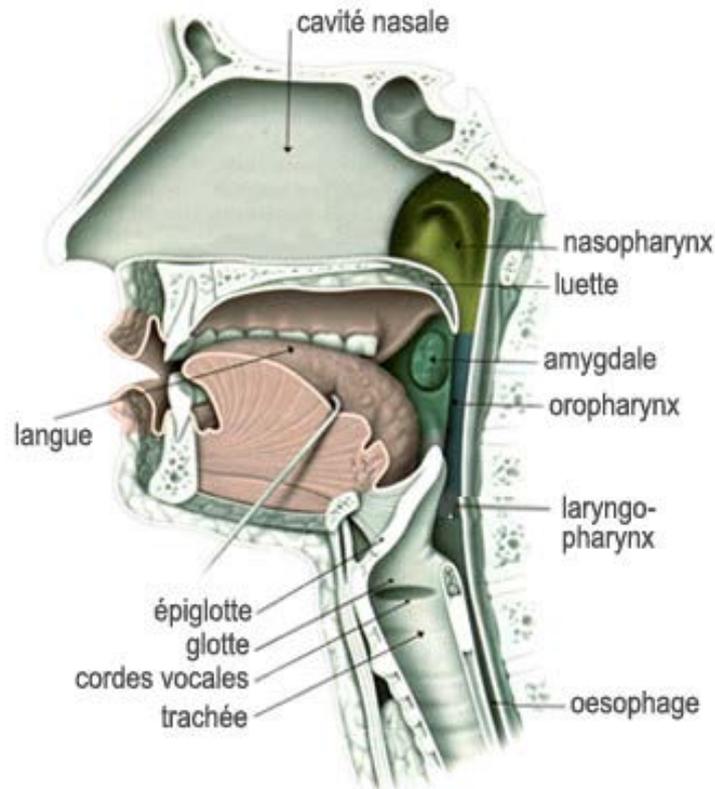


Figure I.1. Anatomie de l'appareil phonatoire humain [13].

Comme le montre la figure I.2 [15], le larynx a une structure anatomique complexe. Il a la forme d'un tube creux, situé entre le pharynx et la trachée, il est formé par un squelette ostéo cartilagineux qui contient :

- En haut, l'os hyoïde ;
- Des cartilages épais et résistant : cartilage thyroïde, cartilage cricoïde, cartilages aryénoïdes et le cartilage épiglottique qui forme le couvercle du larynx.

Ces différents éléments sont réunis entre eux et aux autres organes voisins par des ligaments et des muscles. La cavité du larynx est tapissée par la muqueuse laryngée qui est soulevée par des ligaments. Cette cavité présente :

- Un orifice supérieur communiquant avec le pharynx et fermé par l'épiglotte pendant la déglutition ;
- Un étage supérieur compris entre l'épiglotte et les cordes vocales supérieures ;
- Un étage moyen compris entre les cordes vocales supérieures et inférieures ;
- Un étage inférieur où son ouverture est dans la trachée.

Le conduit vocal contient la cavité nasale, la cavité orale et le pharynx. Sa longueur est comprise entre 13 cm et 17 cm pour les femmes et entre 17 cm et 24 cm pour les hommes [16]. L'emplacement des articulateurs (dents, langue, mâchoire, lèvres, voile de palais) permet une variation de 0 à plus de 20 cm² de l'aire de la section droite du conduit vocal.

La longueur typique du conduit nasal est de 12 cm. Le degré de couplage entre la cavité nasale et la cavité orale est déterminé en fonction de la taille de l'ouverture du voile de palais. Les caractéristiques spectrales des sons rayonnés par les lèvres peuvent être influencées par le couplage de la cavité nasale. Lorsque le voile de palais est orienté vers le bas, acoustiquement, la cavité nasale est couplée à la cavité orale et le timbre produit par ce couplage est nasal. Afin d'aboutir à un son non nasal, il faut que le voile de palais soit positionné de sorte à fermer l'entrée de la cavité nasale en la découplant ainsi de la cavité orale.

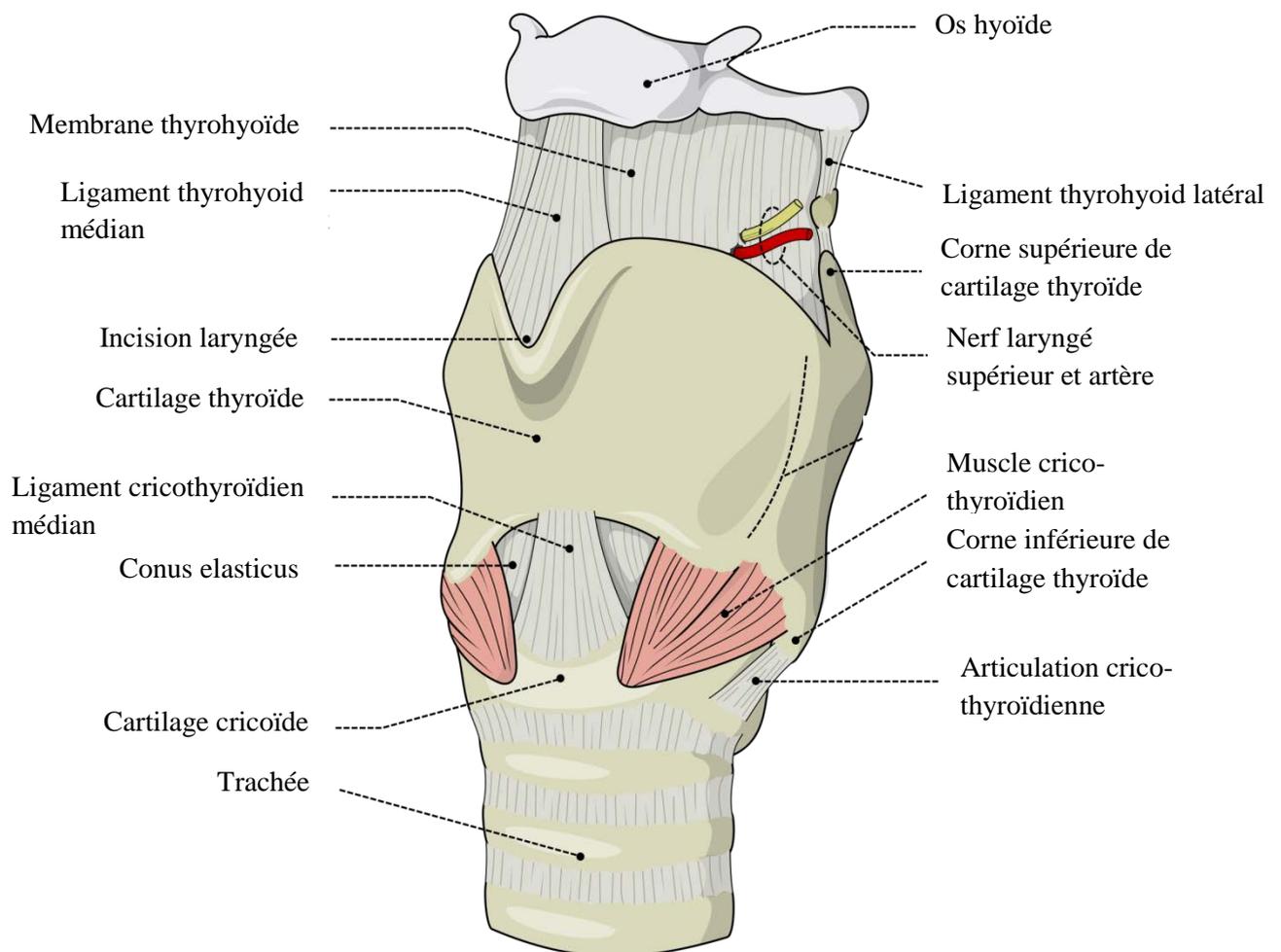


Figure I.2. Anatomie du larynx [15].

I. 2. 2. Physiologie du système de production de la parole

Les poumons fournissent l'énergie nécessaire à la production de la voix qui est l'air, avec un débit bien déterminé. L'air fourni par les poumons et acheminé par la trachée passe via le larynx et le conduit vocal. Le rôle de larynx est la production d'une excitation qui a une forme quasi-périodique dans le cas des sons voisés [17]. Dans ce cas, les cordes vocales s'écartent entre-elles formant la glotte sous l'effet de l'augmentation de la pression d'air. Le passage d'une bouffée d'air via la glotte entraîne un mouvement vibratoire des cordes vocales et la création du son (signal acoustique).

I. 3. Modèle de production de la parole

Selon la littérature, il existe deux grandes approches pour modéliser la production de la parole, qui sont les modèles physiques (modèles physiques simples [18, 19] et modèles physiques complexes [20, 21]) et les modèles de signaux (modèle source-filtre [14]). En général, les modèles de signaux offrent une implémentation à faible coût en termes de complexité de calcul par rapport aux modèles physiques [22].

I. 3. 1. Modèle source-filtre

La figure I.3 représente le modèle discret source-filtre utilisé pour la modélisation du système de production de la parole [14]. Le processus de production de la parole peut être modélisé comme une opération de filtrage linéaire. Le filtre $H(z)$ modélise le conduit vocal et les articulateurs, le filtre $R(z)$ simule le rayonnement aux lèvres.

Pendant les segments de parole voisée, une estimation de la période fondamentale T_0 est utilisée par l'excitation pour la génération d'une suite d'impulsions qui excitent le filtre d'impulsion glottique $G(z)$. Pendant les segments de parole non voisée, un bruit blanc caractérisé par un spectre plat modélise l'excitation.

La fonction de transfert du conduit vocal peut être exprimée par l'expression suivante [14]:

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

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

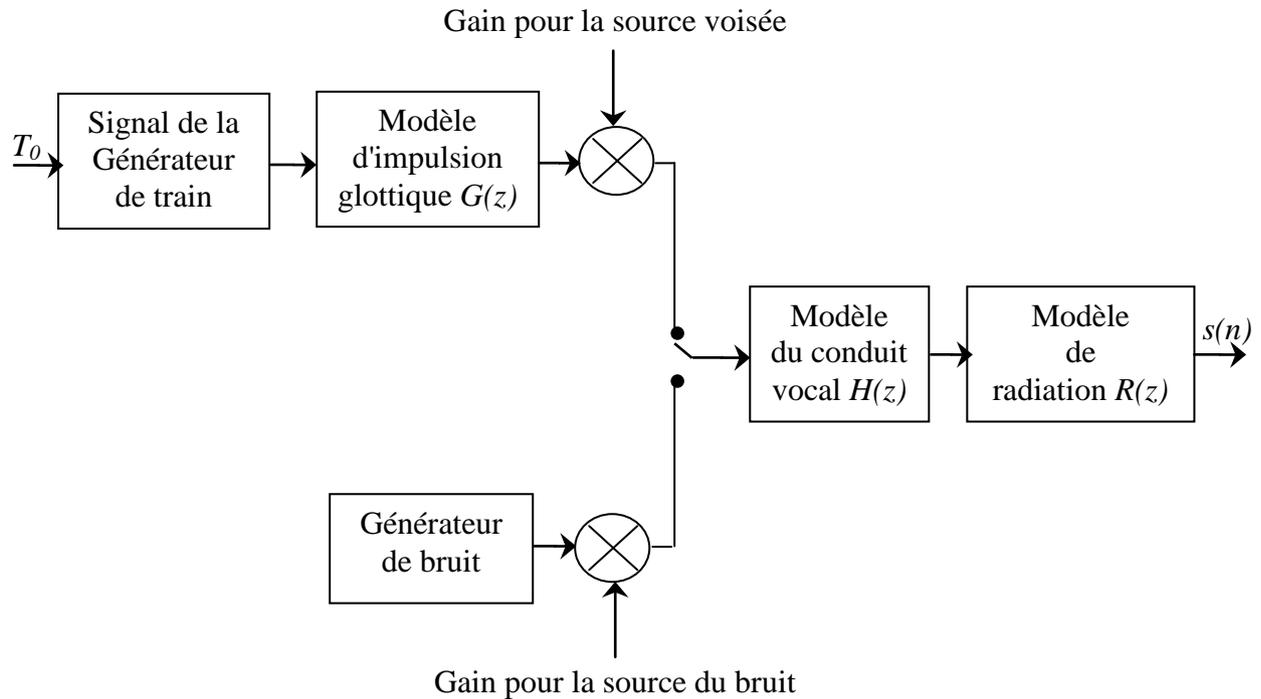


Figure I.3. Modèle général du système de production de la parole.

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. L'exemple suivant, développés dans [14], présente une source périodique modélisée par deux pôles.

$$g(n) = (\alpha^n - \beta^n) \mu(n), \quad \beta < \alpha < 1, \quad \alpha \approx 1 \quad (\text{I.2})$$

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

La fonction de transfert suivante modélise le filtre de rayonnement aux lèvres :

$$R(z) = 1 - z_0 z^{-1} \quad z_0 \approx 1, \quad z_0 < 1 \quad (\text{I.3})$$

Dans le cas de la parole non voisée, la sortie $S(z)$ du modèle est définie par:

$$S(z) = E(z)H(z)R(z) \quad (\text{I.4})$$

où $E(z)$ représente la transformée en z d'une réalisation du bruit blanc.

Dans le cas de la parole voisée, la sortie $S(z)$ est définie comme suit :

$$S(z) = E(z)G(z)H(z)R(z) \quad (\text{I.5})$$

où $E(z)$ représente la transformée en z du train d'impulsions caractérisé par une période égale à la période fondamentale T_0 du signal de parole.

La fonction de transfert du système globale $\theta(z)$ du modèle est définie comme suit :

$$\theta(z) = \begin{cases} H(z)R(z) & \text{pour les sons non voisés} \\ G(z)H(z)R(z) & \text{pour les sons voisés} \end{cases} \quad (\text{I.6})$$

I. 3. 2. Modélisation de la source glottique

L'objectif de la modélisation de la source glottique est la représentation de la forme d'onde glottique et/ou sa dérivée en utilisant certains paramètres bien définis, chacun de ces paramètres a une réalité physiologique ou physique. Dans la littérature, il existe plusieurs modèles, qui reproduisent fidèlement le comportement de la forme de la source glottique. Certains modèles visent à modéliser le débit d'air glottique comme le modèle de Lijlencrants-Fant (modèle LF) [23], le modèle de Klatt [24], le modèle de Rosenberg [25] et le modèle R++ [26]. D'autres modèles ont été proposés pour déterminer les paramètres biomécaniques des cordes vocales [27] ou pour la synthèse des troubles de la voix [28].

I. 3. 2. 1. Modèle général de la source glottique

Généralement, la source glottique (qui modélise l'air glottique) a les caractéristiques temporelles suivantes [17, 29]:

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

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

La dérivée du débit glottique est souvent considérée à la place du débit glottique elle-même [29]. Cette considération est due au fait que lors de la modélisation de la production de la voix, la fonction de transfert de l'élément de rayonnement, liant le flux acoustique au niveau des lèvres au champ acoustique devant des lèvres, peut être considéré comme une dérivée d'une première approximation.

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

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

Généralement, le modèle du débit glottique est décrit en termes de deux phases qui sont la phase ouverte et la phase fermée comme la montre la figure I.4 [29]. La phase ouverte elle-même est divisée en deux phases: la phase d'ouverture et la phase de fermeture. La phase d'ouverture est définie par l'augmentation du débit glottique de l'état initial à A_v son amplitude maximale à l'instant T_p définissant l'instant d'ouverture, comme représenté sur la figure I.4. La phase de fermeture est décrite par la diminution de la source glottique de A_v à un point au moment T_e définissant l'instant de fermeture où la dérivée atteint son minimum négatif E (E : excitation maximale).

Dans la phase de fermeture, le débit glottique retourne à l'état initial après une fermeture brutale T_a où la dérivée du débit glottique atteint 0 après excitation maximale. En pratique, il convient de mettre $T_c = T_0$, soit la totalité de la période fondamentale [23].

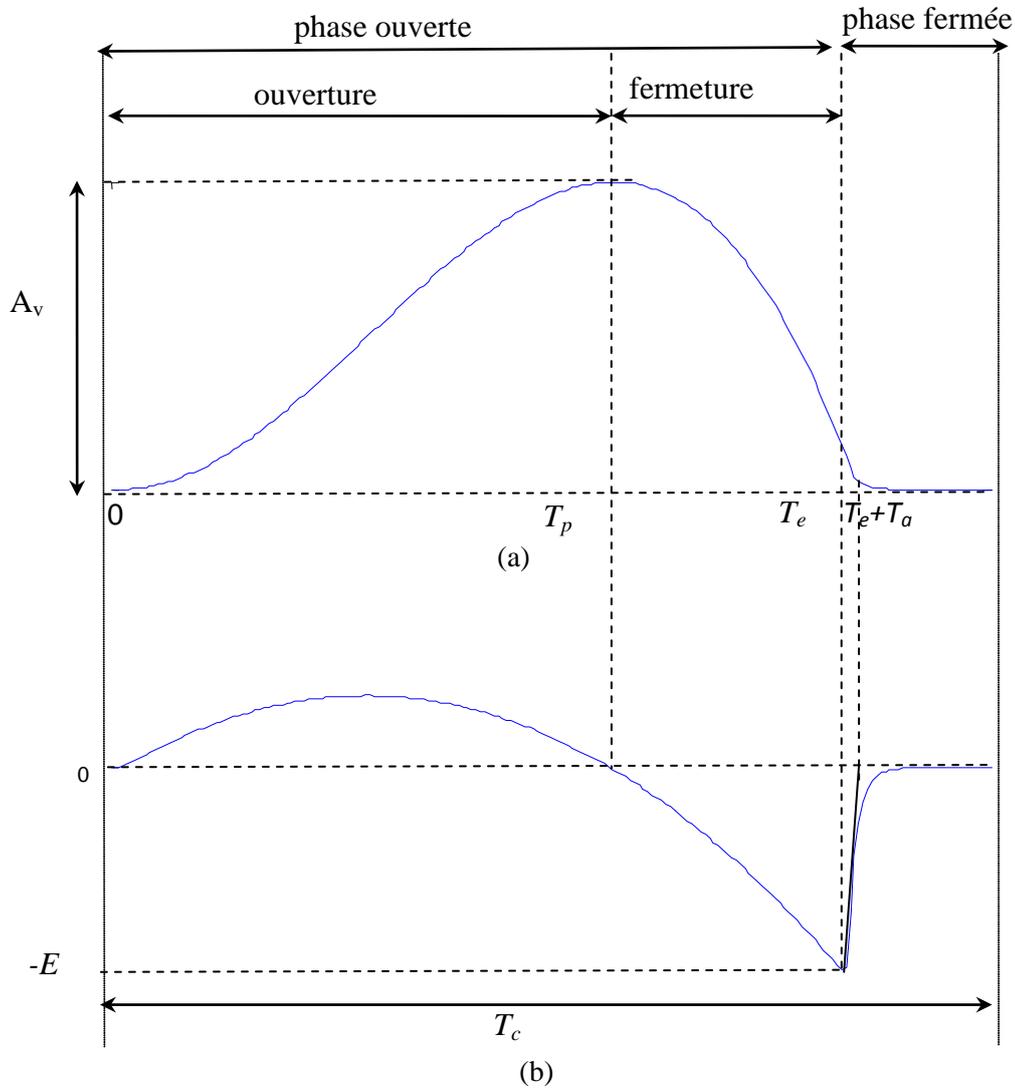


Figure I.4. Modèle de débit glottique (a) et son dérivé (b).

I. 3. 2. 2. Le modèle LF

Dans la littérature, de nombreux modèles de la source glottique sont disponibles. Le modèle proposé par Liljencrants-Fant connu comme le modèle LF [23] est l'un des modèles les plus couramment utilisés pour la synthèse et la modélisation de la parole [30, 31]. Il est décrit par les 5 paramètres suivants:

- T_0 : la période fondamentale,

- E: l'amplitude du minimum de la dérivée de la source glottique (excitation maximum),
- T_p : l'instant du maximum de la source glottique,
- T_e : l'instant de fermeture où la dérivée atteint son minimum négatif,
- T_a : Temps de retour à l'état initial.

Le modèle LF est constitué d'une partie sinusoïdale modulée par une exponentielle croissante (entre 0 et T_e), suivie d'une phase de retour exponentielle décroissante (entre T_e et T_0) comme le montre les équations génériques suivantes [17]:

$$e_g'(n) = \begin{cases} \alpha(t-T_e) \frac{\sin(\frac{\pi t}{T_p})}{\sin(\frac{\pi T_e}{T_p})} - E.e & 0 \leq t \leq T_e \\ -\frac{E}{\varepsilon T_c} (e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}) & T_e \leq t < T_0 \end{cases} \quad (\text{I.7})$$

Les paramètres ε et a sont obtenus en résolvant les deux équations suivantes [29]:

$$\varepsilon T_a = 1 - e^{-\varepsilon(T_0-T_e)} \quad (\text{I.8})$$

$$\frac{1}{a^2 + \left(\frac{\pi}{T_p}\right)^2} \left(e^{-aT_e} \left(\frac{\frac{\pi}{T_p}}{\sin\left(\frac{\pi T_e}{T_p}\right)} \right) + a - \frac{\pi}{T_p} \cotg\left(\frac{\pi T_e}{T_p}\right) \right) = \frac{T_0 - T_e}{e^{\varepsilon(T_0-T_e)} - 1} - \frac{1}{\varepsilon} \quad (\text{I.9})$$

L'intégration de $e_g'(t)$ permet l'obtention de la source glottique $e_g(t)$ suivante [29]:

$$e_g(t) = \begin{cases} \frac{-Ee^{-aT_e}}{\sin\left(\frac{\pi T_e}{T_p}\right)} \frac{1}{a^2 + \left(\frac{\pi}{T_p}\right)^2} \left(\frac{\pi}{T_p} +\right) & 0 \leq t \leq T_e \\ -\frac{E}{\varepsilon T_c} (e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}) & T_e \leq t < T_0 \end{cases} \quad (\text{I.10})$$

Dans le cadre de cette étude, le modèle LF sera utilisé par la suite pour développer une méthode d'analyse du signal de parole pour l'évaluation automatique des voix pathologiques.

I. 4. Pathologies du larynx

Il existe plusieurs types de pathologies qui peuvent affecter le larynx. Ces pathologies sont classées en deux catégories : organiques et fonctionnelles.

I. 4. 1. Dysphonies organiques

Les causes principales des dysfonctionnements organiques sont les changements morphologiques de l'anatomie du larynx, surtout au niveau de la glotte. Parmi ces dysphonies, on trouve : les nodules, les kystes et les laryngites.

- **les nodules**

Les nodules sont des petites masses blanches situées sur le front des cordes vocales au niveau de la jonction du tiers antérieur et du tiers moyen de celles-ci [28]. Les causes principales des nodules sont l'activité vocale excessive et la phonation hyperfonctionnelle.

- **Les kystes**

Les kystes des cordes vocales apparaissent dans la couche superficielle à n'importe quelle partie de celles-ci. Au moment où le kyste croît, il exerce une pression sur l'épithélium et les ligaments des cordes vocales sans affectation des couches adjacentes. Les kystes induisent une augmentation de la masse et un durcissement de la couverture. Comme conséquences des kystes, la glotte ne se ferme pas complètement pendant la phonation [28].

- **les laryngites**

Les laryngites sont des inflammation chroniques ou aiguës des cordes vocales. La cause principale de cette dysphonie est une infection virale et/ou bactérienne. Cette dysphonie augmente la masse et la dureté de la couverture ce qui entraîne une diminution de l'amplitude de vibration des cordes vocales [32]. La voix devient plus grave, rauque avec un timbre voilé, sourde et éraillée. Elle peut disparaître complètement [28].

I. 4. 2. Dysphonies fonctionnelles

Les dysphonies fonctionnelles sont dues à une mauvaise utilisation du système vocal anatomiquement et physiologiquement intact. Parmi les dysphonies fonctionnelles, on trouve : la paralysie laryngée.

- **Paralysie laryngée**

Le nerf laryngé récurrent innerve tous les muscles intrinsèques du larynx sauf le muscle cricothyroïdien. Ce dernier est innervé par le nerf laryngé supérieur. La paralysie du nerf laryngé récurrent peut aboutir à une immobilité d'une ou des deux cordes vocales. Deux types de paralysie peuvent être distingués : temporaire et dégénération des fibres nerveuses. Concernant le premier type de paralysie, après l'élimination de la cause de la paralysie, l'activité musculaire reprend. Quant au second type de paralysie, les muscles restent inactifs à moins que la régénération des fibres nerveuses soit complète. Dans le cas de la paralysie complète, toutes les unités motrices sont paralysées tandis que dans la paralysie partielle, on peut trouver des unités motrices intactes et d'autres paralysées. Donc, la phonation lors de la paralysie du nerf laryngé récurrent dépend du degré de paralysie. La paralysie laryngée se manifeste par une diminution de la dureté des muscles du larynx [32].

I. 5. Evaluation des troubles de la voix

L'évaluation des troubles de la voix se classe en deux approches : l'évaluation subjective et l'évaluation objective. La première approche est basée sur une évaluation perceptive à l'oreille de la voix et de la parole. La deuxième approche repose sur l'acquisition de mesures physiques en utilisant un capteur et application des techniques de traitement du signal sur les mesures extraites pour établir l'évaluation.

I. 5. 1. Evaluation perceptive

En pratique, les cliniciens écoutent à l'oreille pour évaluer la qualité de la voix selon une échelle d'évaluation perceptive. L'objectif de cette dernière est la description de la qualité d'une voix selon des critères quantitatifs et qualitatifs. L'échelle de mesure constitue la composante quantitative de la méthode d'analyse perceptive. Elle permet l'évaluation de la proportion de paramètres perceptifs comprise dans la voix analysée. Ces derniers constituent la composante qualitative de l'échelle d'évaluation. Le choix de la mesure et des paramètres

utilisés pour construire l'échelle est important et dépend des objectifs de l'évaluation de la qualité de la voix [33].

Généralement, Il y a trois principales échelles quantitatives qui sont [33]:

- Echelle bipolaire sémantique : son principe est oui/non (ou encore présence/absence) de deux paramètres de qualité de voix.
- Echelle de classe ou EAI (Equal-Appearing-Interval) : cette échelle propose plusieurs niveaux équidistants, généralement 4 ou 7 niveaux, pour quantifier séparément les paramètres qualitatifs.
- Echelle analogique visuelle (EAV): elle permet d'attribuer visuellement un degré de sévérité de la pathologie par le traçage d'une croix sur une ligne, habituellement de 10 cm de longueur; la normalité se localisant le plus à gauche de la ligne par contre la sévérité la plus forte se localisant le plus à droite. La distance mesurée en millimètres de la gauche à la croix tracée par l'auditeur donne une estimation du degré de sévérité.

Dans la littérature, on peut trouver plusieurs méthodes d'évaluation perceptive [34-37] qui proposent l'évaluation de la qualité de la voix selon des critères quantitatifs et qualitatifs. Dans cette étude, l'évaluation perceptive des corpus est réalisée en utilisant l'échelle GRBAS (Grade Roughness Breathiness Asthenia Strain) [38], qui est une échelle compacte et simple d'utilisation, constituée de 5 paramètres. Ces derniers sont définis dans le tableau I.1 [33]. Les trois paramètres grade global (G : Grade), raucité (R : Roughness) et souffle (B : Breathiness) sont considérés comme étant les plus fiables et les moins soumis à la variabilité d'un jury d'écoute que les deux paramètres asthénie (A : Asthenia) et forçage (S : Strain) [33].

Initiale	Terme	Signification	Définition
G	Grade	Grade globale de dysphonie	Effet globale du degré d'anormalité de la voix.
R	Roughness	Raucité	Effet d'irrégularité des vibrations des cordes vocales qui correspond aux fluctuations irrégulières de la fréquence fondamentale F_0 et/ou à l'amplitude du son glottique.
B	Breathiness	Caractère soufflé	Effet d'une fuite d'air assez importante à travers les cordes vocales, correspond à des turbulences.
A	Asthenia	Asthénie	Manque de puissance de la voix, correspond à une intensité faible du son et/ou un manque des harmoniques élevées.
S	Strain	Forçage	Effet d'un état hyperfonctionnel de phonation, correspondant à une fondamentale F_0 anormalement haute (bruit dans les hautes fréquences et/ou richesse en harmoniques dans les hautes fréquences).

Tableau I. 1. Paramètres de l'échelle GRBAS : signification et définition.

I. 5. 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. 5. 2. 1. Evaluation des troubles de la voix via le signal de parole

Les méthodes d'estimation des dyspériodicités vocales en utilisant le signal de 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 voisée 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 [32].

La méthode d'analyse développée dans [39], pour évaluer le degré d'enrouement de la voix, est fondée sur l'hypothèse qu'une onde acoustique d'une voyelle soutenue est constituée de deux composantes : une composante pseudo périodique (composante harmonique) et une composante additive (composante bruit). Le rapport harmonique à bruit (H/N) a été proposé comme indice acoustique pour évaluer et quantifier le degré d'enrouement de la voix. La composante harmonique H (H : Harmonic) est estimée comme étant l'énergie de l'onde acoustique moyenne, tandis que la composante bruit N (N : Noise) est estimée comme étant l'énergie moyenne des différences entre les périodes individuelles et l'onde moyenne. Cette méthode a été appliquée à un corpus de voyelles soutenues [a] contient 42 voix normophoniques et 41 voix dysphoniques avec différents degrés d'enrouement. Deux experts ont évalué visuellement le spectrogramme de chaque stimulus. Les résultats obtenus ont montré un accord très significatif entre le rapport harmonique à bruit et l'évaluation subjective des spectrogrammes.

L'analyse cepstrale a été utilisée pour évaluer la régularité du signal acoustique [40-44]. Le cepstre réel d'un signal donné $s(t)$ est défini, comme étant la transformée de Fourier inverse du logarithme de son spectre d'amplitude.

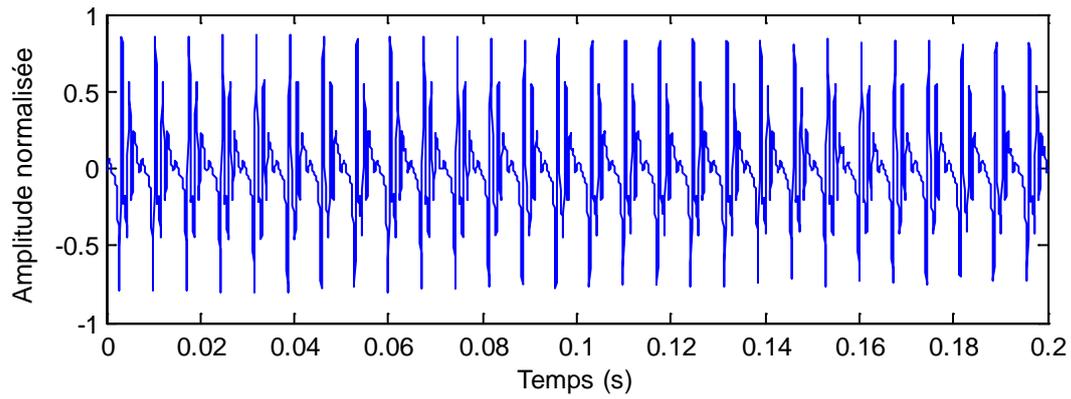
Dans l'article original [45], les auteurs ont inventé le mot cepstre (en anglais, cepstrum) en inversant la première syllabe du spectre (en anglais, spectrum), la justification étant qu'il s'agissait d'un spectre d'un spectre. De même, le mot quefrérence a été obtenu à partir de fréquence et le mot rharmonique a été obtenu à partir d'harmonique.

L'amplitude de la première rharmonique du cepstre (R1A) et la proéminence du pic cepstral (CPP : Cepstral Peak Prominence) sont des indices acoustiques proposés pour caractériser les troubles de la voix dysphoniques. Plusieurs études ont montré que ces deux indices sont corrélés de manière significative avec l'évaluation perceptive de la voix [40-44].

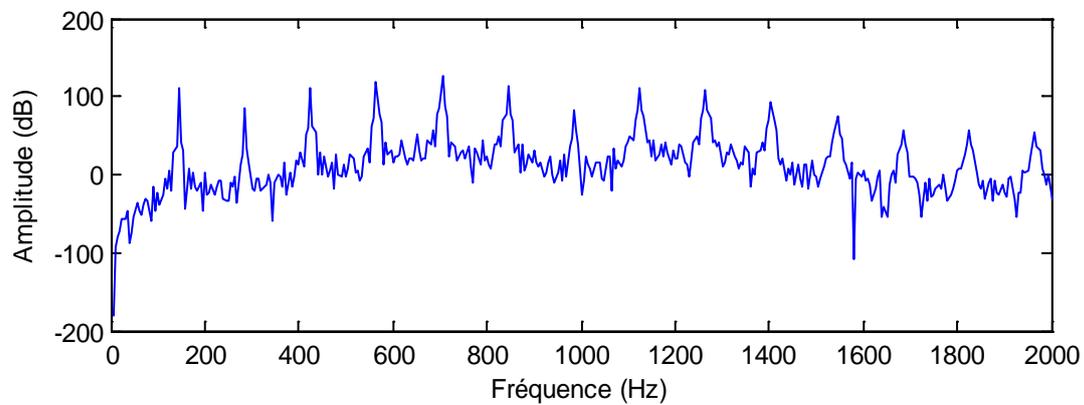
La figure I.5 illustre l'analyse cepstrale d'un segment extrait d'une voyelle [a] soutenue produite par un locuteur normophonique. Le signal acoustique correspondant à la voyelle [a] soutenue et le log-spectral sont montrés sur les figures I.5-(a) et I.5-(b), respectivement. Le cepstre, obtenu en prenant la transformée de Fourier inverse du log-spectrale, est montré sur la figure I.5-(c). La figure I.6 illustre l'analyse cepstrale d'un segment extrait d'une voyelle [a] soutenue produite par un locuteur dysphonique.

La proéminence du pic cepstral (CPP) mesure l'amplitude logarithmique de la première rharmonique du cepstre de la parole. L'indice acoustique CPP est calculé comme suit [2]:

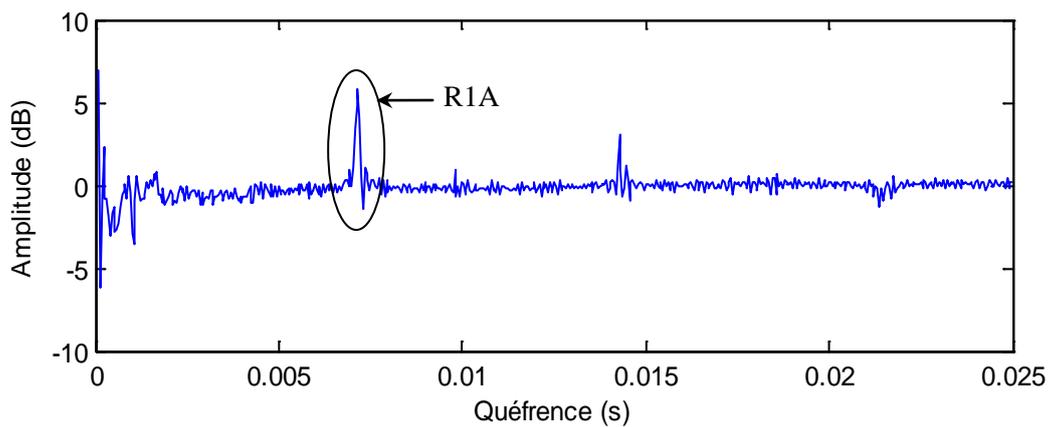
- Calculer le cepstre de chaque trame d'analyse de la parole.
- Ajuster une ligne de régression linéaire au logarithme du cepstre entre la quéfrence maximale et 1 ms.
- Obtenir, entre le minimum et le maximum attendu des quéfrences vocales, la hauteur par rapport à la ligne de régression de la plus proéminence du pic cepstral, qui est la proéminence du pic cepstral locale (par-trame).
- Obtenir la proéminence du pic cepstral global CPP en faisant la moyenne des proéminences des pics cepstraux locaux sur toutes les trames d'analyse.



(a)

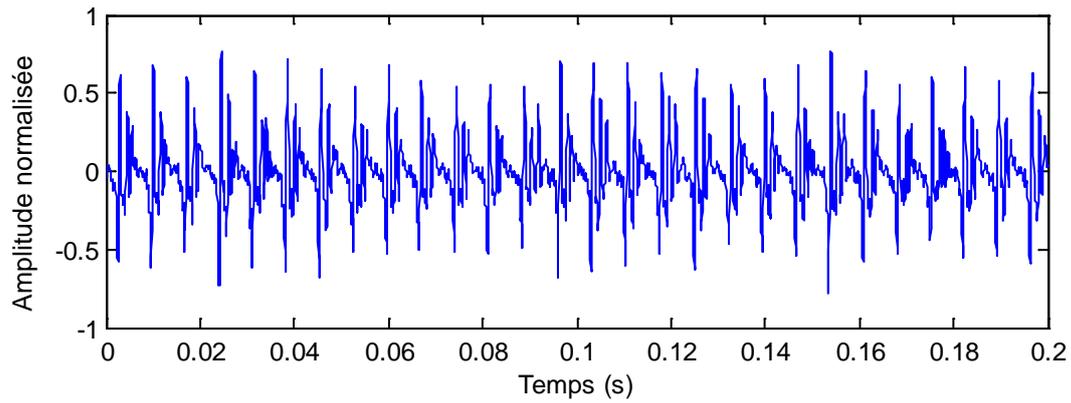


(b)

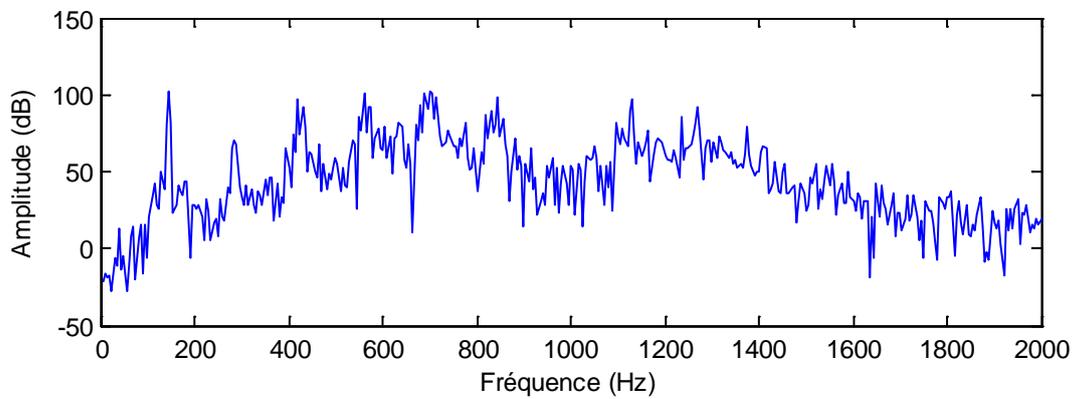


(c)

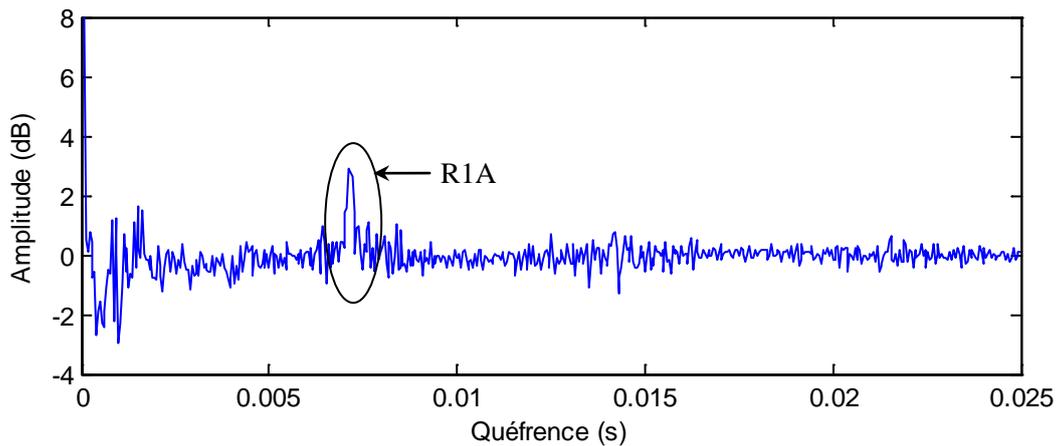
Figure I. 5. Analyse cepstrale d'un segment extrait d'une voyelle [a] produite par un locuteur normophonique. (a) Signal de parole. (b) Spectre d'amplitude. (c) Cepstre.



(a)



(b)



(c)

Figure I. 6. Analyse cepstrale d'un segment extrait d'une voyelle [a] produite par un locuteur dysphonique. (a) Signal de parole. (b) Spectre d'amplitude. (c) Cepstre.

Dans [42], une analyse cepstrale des dyspériodicités vocales dans la parole connectée a été réalisée. L'indice acoustique R1A a été utilisé pour quantifier la qualité de la voix. L'analyse a été appliquée à un corpus de parole connectée produit par 28 locuteurs normophoniques et 223 locuteurs dysphoniques. Les résultats obtenus montrent que la corrélation entre l'amplitude de la première harmonique (R1A) et l'évaluation perceptive augmente lorsque le spectre est limité en bande avant le calcul du cepstre.

Dans [43], des analyses ont été réalisées sur des échantillons préopératoires et postopératoires de 19 patients atteints d'une paralysie unilatérale du nerf laryngé récidivant qui ont subi une intervention chirurgicale. Deux orthophonistes ont été évalués perceptivement les échantillons de la parole selon le grade (G), le souffle (B), la raucité (R), l'asthénie (A) et le forçage (S). Les résultats obtenus indiquent que la proéminence du pic cepstral est le meilleur prédicteur pour quantifier la dysphonie comparée au Jitter, shimmer et bruit spectral.

Dans [44], les auteurs ont étendu l'analyse à un corpus de grande taille incluant des voyelles [a] soutenues et de la parole connectée. Ce corpus est produit par 281 locuteurs dysphoniques. Dix thérapeutes de la parole ont évalué perceptivement les stimuli de la parole connectée selon le grade (G) perçu. Les résultats obtenus confirment que la proéminence du pic cepstral est le prédicteur le plus fiable pour quantifier la dysphonie par rapport aux indices acoustiques.

Dans [4], le variogramme généralisé a été utilisé pour estimer les dyspériodicités vocales dans le signal de parole. Les rapports signal à dyspériodicité global (SDR: Signal-to-Dysperiodicity Ratio) et segmental (SDRSEG: Segmental signal-to-Dysperiodicity Ratio) ont été utilisés pour quantifier numériquement les dyspériodicités vocales.

La méthode d'analyse garantit qu'une augmentation/diminution légère des irrégularités dans le signal de parole résulte en une augmentation/diminution de la dyspériodicité vocale estimée. La méthode a été appliquée à deux corpus. Le premier corpus contient des fragments stationnaires extraits de voyelles [a] soutenues produites par 89 hommes et femmes dysphoniques ou normophoniques. Le second corpus contient quatre phrases en français ainsi que des voyelles [a] soutenues produites par 22 locuteurs dysphoniques ou normophoniques. La méthode fournit des rapports signal-à-dyspériodicité global et segmental statistiquement significativement corrélés avec le degré de l'enrouement perçu. Pour la parole connectée, il a été observé que le rapport signal-à-dyspériodicité segmental résulte en une plus forte

corrélation avec les scores de l'enrouement perçu comparé au rapport signal-à-dysperiodicité global.

Dans [5], une approche multi-bandes à base du variogramme généralisé a été utilisée pour estimer les dyspériodicités vocales. Le rapport signal à dyspériodicité a été utilisé pour quantifier le degré d'enrouement du locuteur. Cette approche a été testée sur deux corpus comprenant plus de 250 et 700 locuteurs. Les locuteurs du premier corpus ont été évalués perceptivement selon le grade (G) perçu alors que les locuteurs du deuxième corpus ont été classés en normophonique/dysphonique. Les résultats obtenus montrent que la corrélation des valeurs de l'indice acoustique avec les scores de l'évaluation perceptive est statistiquement significativement plus grande pour l'analyse multi-bandes comparée à l'analyse en bande entière. En outre, la combinaison du rapport signal-à-dyspériodicité et la proéminence du pic cepstral permet d'augmenter la corrélation avec l'évaluation perceptive.

Plus récemment, dans [10], une méthode d'évaluation des troubles de la voix basée sur la décomposition modale empirique (en anglais, empirical mode decomposition : EMD) dans le domaine log-spectrale a été proposée. En utilisant l'algorithme EMD, le logarithme du spectre d'amplitude du signal de parole est décomposé en ces trois composantes qui sont la composante harmonique, l'enveloppe spectrale et le bruit. Le rapport harmonique sur bruit (HNR) a été utilisé pour quantifier le degré de perturbation dans le signal de parole.

L'approche proposée a été testée sur des voyelles synthétiques et sur la parole naturelle. Le corpus des voyelles synthétiques comprend 48 stimuli de voyelles [a] et le corpus de la parole naturelle comprend une concaténation de la voyelle [a] avec deux phrases en Néerlandais produite par 28 locuteurs normophoniques et 223 locuteurs dysphoniques avec différents degrés d'enrouement. Cette méthode a été comparée à une méthode conventionnelle d'analyse du signal de parole qui est l'analyse cepstrale. Les résultats obtenus montrent que la méthode basée sur la décomposition modale empirique présente des meilleures performances que le SDRSEG et CPP en termes de corrélation de l'indice avec le grade (G) perçu.

Dans l'étude développée dans [10], les seuils utilisés dans l'algorithme ont été fixés de façon empirique. Par la suite, dans [11], les auteurs ont proposé une procédure appropriée qui permet d'estimer de façon automatique les seuils utilisés par l'algorithme de décomposition.

La méthode d'analyse basée sur la décomposition modale empirique sera présentée et discutée avec plus de détails dans le chapitre IV.

I. 5. 2. 2. Evaluation des troubles de la voix 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 [46], et la réparation de la voix dysphonique où la parole alaryngée est re-synthétisée en utilisant un signal glottique synthétique [47].

Le comportement de la source glottique est décrit par des caractéristiques. Ces dernières ont un intérêt considérable pour la détection des pathologies vocales. La figure I.7 illustre un système complet de traitement de la source glottique, à partir du signal de parole jusqu'à l'intégration des informations glottiques dans une application de technologie vocale (applications biomédicales).

En général, le jitter et le shimmer sont deux mesures typiques utilisées pour la caractérisation des irrégularités dans un signal quasi-périodique. Le jitter se réfère à la modification de la durée des cycles glottiques consécutifs, alors que le shimmer traduit les changements dans l'amplitude des cycles glottiques consécutifs [48].

De nombreuses méthodes de quantification des perturbations de la source glottique ont été proposées. Dans [7], le jitter à court terme obtenu par un estimateur spectral de jitter (Spectral Jitter Estimator : SJE) basé sur une description mathématique du phénomène de jitter a été proposé pour identifier les troubles de la voix dans la parole continue.

Dans [8], une estimation de la source glottique a été utilisée pour détecter les troubles de la voix. Trois ensembles de caractéristiques ont été proposés. Chaque ensemble de caractéristiques dépend du signal de parole, du signal de la source glottique ou de la prosodie. Les résultats montrent que les caractéristiques extraites à partir du signal de la parole ainsi que ceux obtenues à partir du signal de la source glottique sont relativement complémentaires,

alors qu'ils indiquent une certaine synergie avec les caractéristiques obtenues à partir de la prosodie. En effet, la combinaison des caractéristiques extraites de la source glottique et celles obtenues à partir du signal de parole donne une meilleure discrimination normophonique/dysphonique.

L'étude présentée dans [9] s'est focalisée sur une méthode de détection de la pathologie de la voix à travers une signature biométrique. Cette dernière est basée sur la densité spectrale de puissance de la source glottique du locuteur.

La caractérisation de la source glottique a été utilisée pour la classification de la dépression clinique dans la parole. Dans [49], le jitter vocal et le débit glottique ont été proposés comme indices pour estimer le risque de suicide d'un patient. Cette étude a pour objectif de distinguer entre un risque suicidaire élevé à court terme, une grande dépression et les patients non suicidaires. Il a été constaté que le jitter vocal est un discriminateur significatif seulement entre les groupes témoins suicidaires et non-suicidaires, alors que la pente spectrale du débit glottique est un discriminateur significatif entre les trois groupes.

L'étude présentée dans [50] traite le problème de la discrimination entre l'état de dépression et l'état normal à partir du signal parole. Les résultats obtenus indiquent que la combinaison des caractéristiques glottiques et prosodiques résulte en une meilleure discrimination que la combinaison des caractéristiques prosodiques et des caractéristiques du conduit vocal.

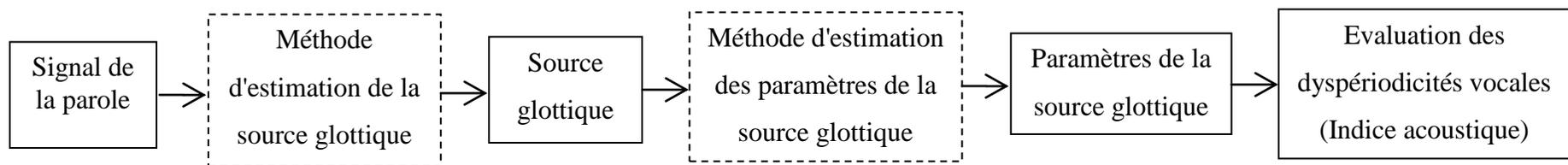


Figure I.7. Système complet de traitement de la source glottique, à partir du signal de parole jusqu'à l'intégration des informations glottiques dans une application de technologie vocale (applications biomédicales).

I. 6. Conclusion

Dans ce premier chapitre, nous avons présenté l'état de l'art de l'évaluation des troubles de la voix. La première partie de ce chapitre a été consacré à la présentation de l'anatomie et la physiologie du système de production de la voix ainsi qu'à la description de quelques pathologies du larynx.

Ensuite, nous avons présenté un modèle de production de la parole qui est le modèle source-filtre ainsi que le modèle LF de la source glottique. Ce dernier sera utilisé dans la suite pour l'évaluation automatique des troubles de la voix.

Après avoir décrit brièvement les méthodes d'évaluation perceptive, nous avons présenté un état de l'art des méthodes d'évaluation objective des troubles de la voix. Ces méthodes ont été classées en deux classes selon qu'elles utilisent le signal de parole ou une estimation de la source glottique.

*Chapitre II: La décomposition modale
empirique*

II. 1. Introduction

La décomposition modale empirique (EMD : empirical mode decomposition) est un outil d'analyse des signaux multi-composantes. Cette méthode d'analyse n'exige pas a priori des fonctions de base fixes comme les méthodes d'analyse conventionnelles (par exemple la transformée de Fourier et la transformée en ondelette). Elle a été proposée initialement dans [51] pour analyser les données issues de processus non-linéaires et/ou non stationnaires.

La décomposition modale empirique produit pour tout signal une décomposition multi-échelles pilotée par les données. Les composantes obtenues sont des formes d'onde oscillantes potentiellement non harmoniques dont les caractéristiques, forme, amplitude et fréquence peuvent varier au cours du temps [52]. Pour cela, elle a été largement utilisée dans plusieurs domaines tels que l'analyse du signal de vibration [53], le diagnostic des défauts mécaniques [54], l'analyse des signaux électrocardiographies (ECG) [55-56] et électroencéphalographies (EEG) [57], le contrôle non destructif [58], etc.

L'objectif de ce chapitre est de présenter une étude compréhensive de la méthode de décomposition modale empirique, une approche que nous allons utiliser pour analyser le signal du parole afin d'évaluer et de quantifier de manière automatique les dyspériodicités vocales.

Après un rappel du principe de l'algorithme EMD, nous présenterons un exemple d'application pour décomposer un signal synthétique. Les étapes fondamentales d'implantation de la décomposition modale empirique telle que l'interpolation des maxima et le critère d'arrêt seront présentées en détail. Enfin, nous présenterons quelques caractéristiques fondamentales ainsi que les limitations de la méthode EMD.

II. 2. Principe de la décomposition modale empirique

Le principe de la décomposition modale empirique peut être décrit par une opération de décomposition élémentaire appliquée, de manière récursive, à un signal oscillant $x(t)$. Cette opération mène à l'extraction de la partie qui oscille selon une contribution "hautes fréquences" localement $c(t)$. Donc, la différence $x(t) - c(t)$ représente la composante qui oscille selon une contribution "basses fréquences". Ensuite, la procédure est appliquée de nouveau à la différence $x(t) - c(t)$ qui est considérée comme un nouveau signal pour l'extraction de la composante qui oscille localement le plus rapidement [52].

Le processus de décomposition se répète et s'arrête si $x(t)-c(t)$ n'est plus un signal oscillant. Comme conséquence du principe précédent, le signal $x(t)$ peut être exprimé comme la somme des fonctions oscillantes $c_i(t)$ et d'un résidu $r(t)$:

$$x(t) = \sum_{i=1}^I c_i(t) + r(t) \quad (\text{II.1})$$

où Les fonctions $c_i(t)$ sont appelées les fonctions de mode intrinsèques (IMFs : Intrinsic mode functions) et I est le nombre total des fonctions $c_i(t)$.

Par définition [51, 52], une fonction de mode intrinsèque est une fonction oscillante de moyenne nulle, c-à-d, une fonction $f: [a, b] \rightarrow \mathbb{R}$ qui vérifié les conditions suivantes :

- 1- Tous les minima/maxima locaux de f sont strictement négatifs/positif.
- 2- La moyenne locale de f est nulle : la somme de l'enveloppe inférieure interpolant les minima de f et de l'enveloppe supérieure interpolant les maxima est nulle.

II. 3. Algorithme de la décomposition modale empirique

L'algorithme de la décomposition modale empirique décompose de manière adaptative un signal $x(t)$ en modes d'oscillation appelés les fonctions de mode intrinsèque (IMFs) [51]. Ces fonctions IMFs sont extraites à partir du signal lui-même. Les fonctions IMFs sont obtenues par l'intermédiaire du processus de tamisage itératif, qui implique les étapes suivantes:

1. Initialiser l'algorithme: $r_0(t) = x(t)$ (résidu initial), $i=I$ (indice des IMFs) et fixer le seuil δ .
2. Calculer la $i^{\text{ème}}$ IMF:
 - a- Initialiser $h_0(t) = r_{i-1}(t)$, $j = I$ (indice des itérations),
 - b- Extraire les maxima et les minima locaux de $r_{j-1}(t)$,
 - c- Calculer l'enveloppe supérieure $S_j(t)$ et l'enveloppe inférieure $I_j(t)$ par interpolation des maxima et des minima locaux, respectivement.
 - d- Calculer l'enveloppe moyenne.

$$M_j(t) = (S_j(t) + I_j(t)) / 2 \quad (\text{II.2})$$

- e- Calculer la $j^{\text{ème}}$ composante: $h_j(t) = r_{j-1}(t) - M_j(t)$.

f- Traiter $h_j(t)$ comme $r_{j-1}(t)$ puis calculer ensuite $h_j(t) = r_{j-1}(t) - m_j(t)$ jusqu'à :

$$SD_j = \sum_{t=0}^T \frac{|h_{j-1}(t) - h_j(t)|^2}{(h_{j-1}(t))^2} \langle \delta \rangle \quad (\text{II.3})$$

g- Calculer la $i^{\text{ème}}$ IMF comme: $\text{IMF}_j(t) = h_{j,k}(t)$.

3. Mettre à jour le résidu $r_j(t) = r_{j-1}(t) - \text{IMF}_j(t)$

4. Incrémenter l'indice de tamisage i et répéter les étapes 2-a à 2-g jusqu'à ce que le nombre des extrema locaux dans $r_j(t)$ soit inférieur ou égal à 2.

Chaque IMF peut être caractérisé par sa fréquence instantanée. Le signal peut être reconstruit exactement en additionnant tous les IMFs et le résidu

$$x(t) = \sum_{i=1}^I \text{IMF}_i(t) + r_i(t) \quad (\text{II.4})$$

I est le nombre total des IMFs.

Dans l'algorithme de la décomposition modale empirique, le processus de décomposition du signal en ensemble d'IMFs est également appelée tamisage. La figure II.1 montre l'organigramme général de l'algorithme EMD, qui comprend deux boucles:

- La première boucle principale indexée par i concerne la décomposition jusqu'au résidu final.
- La deuxième boucle est indexée par j concerne le processus de tamisage.

II. 4. Exemple d'illustration de l'algorithme de la décomposition modale empirique

La figure II.2. montre un exemple de décomposition du signal $x(t)=[5\sin(60\pi t)].[7\sin(40\pi t)]$ en quatre fonctions de mode intrinsèques et un résidu. On observe que le premier mode extrait correspond au mode le plus rapide et le dernier mode correspond aux fluctuations les plus lentes.

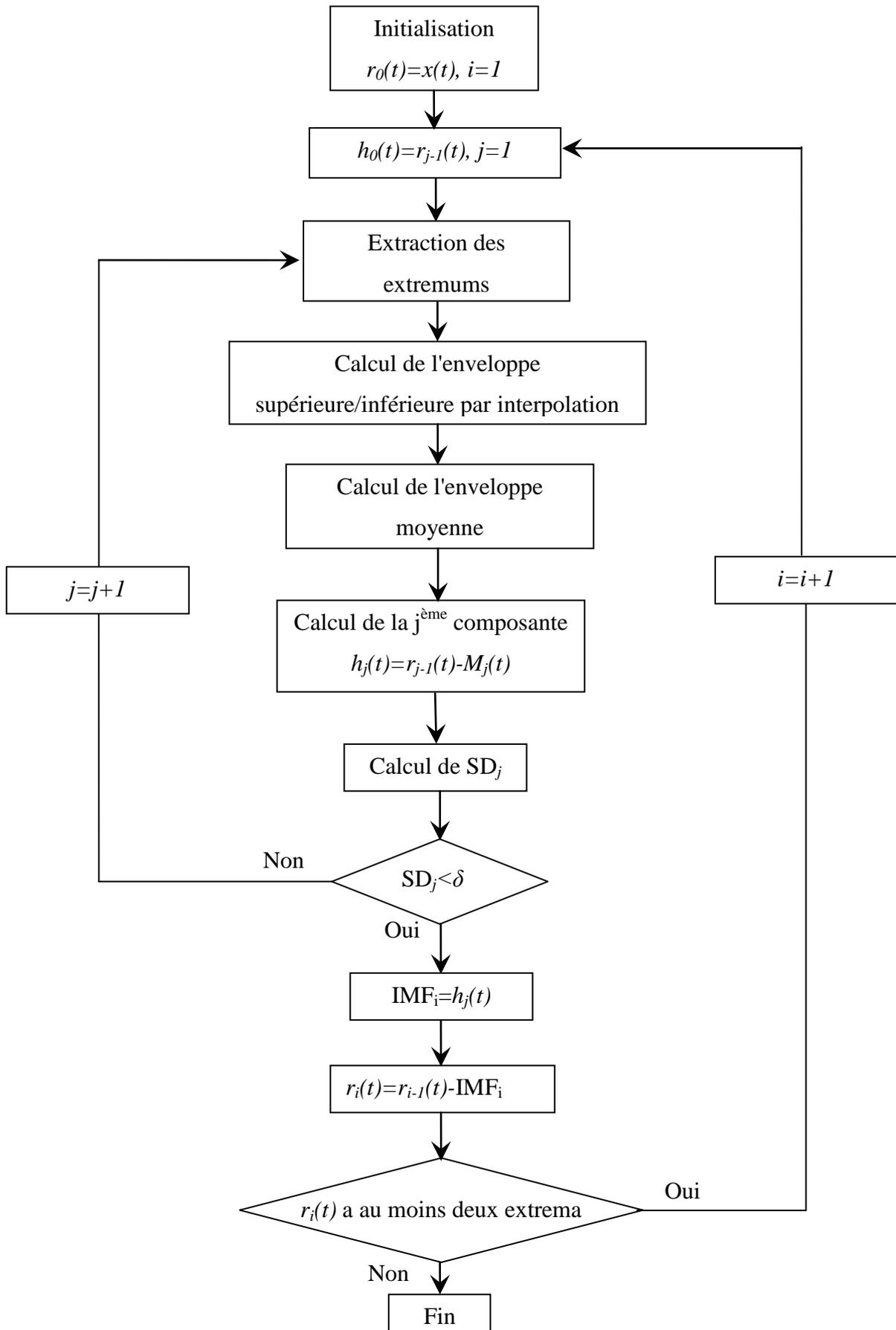


Figure II.1. Organigramme de l'algorithme de la décomposition modale empirique.

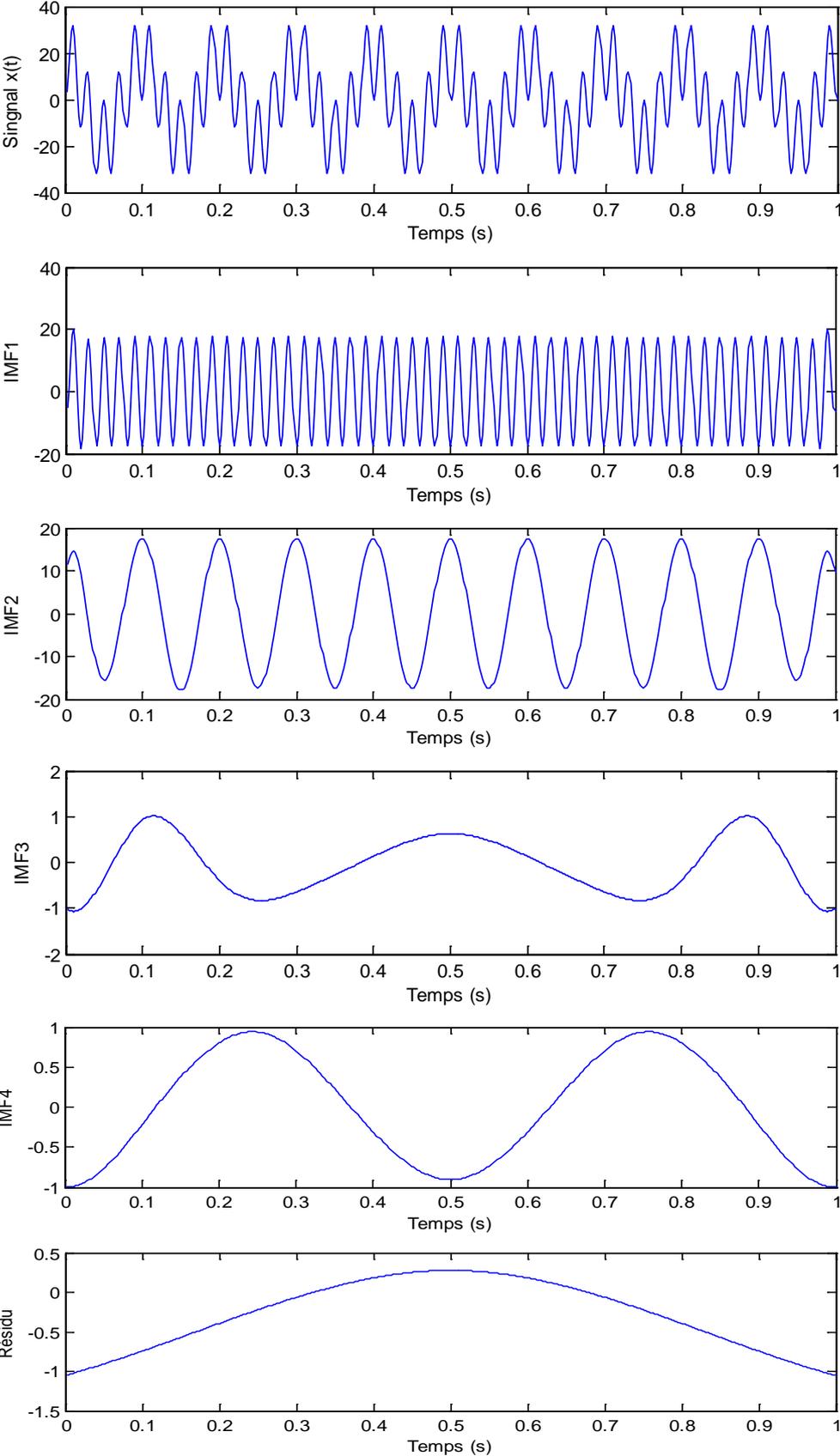


Figure II.2. Exemple de décomposition d'un signal en quatre fonctions de mode intrinsèque et un résidu.

II. 5. Implantation de la décomposition modale empirique

L'utilisation de l'algorithme de la décomposition modale empirique nécessite quelques précisions, en particulier sur le choix du type d'interpolation des maxima utilisée pour le calcul des enveloppes supérieures et inférieures ainsi que le critère d'arrêt du processus de tamisage.

II. 5. 1. Interpolation des maxima

Pendant l'étape du tamisage, l'interpolation des extrema est une étape fondamentale dans l'algorithme de la décomposition modale empirique. Cette étape permet de déterminer localement les enveloppes inférieure et supérieure de chaque mode et d'extraire l'IMF correspondante. Lors de l'extraction des fonctions de mode intrinsèques (IMFs), le choix de la méthode d'interpolation influe sur la qualité de la décomposition [59].

Dans la littérature, il existe plusieurs approches pour effectuer l'interpolation d'un signal. Il y a les méthodes simples telles que la méthode de Newton et la méthode de Lagrange et les méthodes les plus avancées, par exemple l'interpolation par splines cubiques et l'interpolation par B-splines. En générale, l'interpolation des extrema par splines cubiques produit des résultats satisfaisants alors que les autres méthodes d'interpolation ne sont pas recommandées pour l'estimation des IMFs car elles aboutissent à un nombre trop important de modes [51].

L'interpolation des enveloppes inférieure et supérieure est nécessaire pour l'estimation des IMFs et elle représente un problème non négligeable [60]. Dans le cadre de nos travaux de recherche sur l'analyse par EMD du signal de la parole, nous appliquerons l'interpolation par splines cubiques qui donne un bon compromis entre la simplicité de calcul et la qualité d'interpolation.

II. 5. 2. Critère d'arrêt du processus de tamisage.

Comme le montre l'organigramme de la figure II.1, l'algorithme de la décomposition modale empirique est constitué de deux boucles :

- La première boucle (boucle de tamisage): elle permet l'extraction de l'IMF, ce qui définit l'échelle de l'EMD (niveau de profondeur de la décomposition).
- La deuxième boucle: elle permet de confirmer que l'IMF a un comportement oscillant.

La première boucle (boucle de tamisage) est gérée par un critère d'arrêt, noté SD (SD : Standard Deviation) [51]. L'expression de SD est définie par la relation (II.4). Ce critère d'arrêt permet de calculer l'écart quadratique entre deux étapes successives de tamisage. Il est approuvé si la différence entre deux tamisages successives est inférieure à un seuil δ fixé d'avance ($SD < \delta$, étape 2-f de l'algorithme EMD).

En générale, la valeur de SD est choisie en prenant en considération les deux contraintes suivantes [51] :

- Lorsque le seuil δ est très élevé, l'EMD ne décompose pas correctement les différents modes qui se présentent dans le signal.
- Lorsque le seuil δ est très élevé, il peut y avoir un phénomène de sur-décomposition du signal avec des IMFs d'amplitudes quasi constantes qui sont modulées par une seule fréquence.

La valeur de critère d'arrêt SD est choisie de manière empirique, sa valeur typique est comprise entre 0.2 et 0.3.

II. 6. Quelques caractéristiques fondamentales de la méthode EMD

Bien que la décomposition modale empirique soit définie par un algorithme, cette méthode dispose de propriétés empiriques qui apparentent aux propriétés de décompositions plus mathématiques telles que la non-linéarité, la multirésolution, la localité et l'auto-adaptativité [52, 61-62].

II. 6. 1. Non-linéarité

De manière globale, l'algorithme de la décomposition modale empirique est non linéaire. En effet, la somme des EMD des deux signaux séparés, par exemple, est différente à l'EMD de la somme de ces deux signaux.

Le comportement non linéaire de la décomposition modale empirique est dû principalement à la dépendance de cette méthode des extrema locaux, de l'interpolation et du critère d'arrêt. La non-linéarité de l'EMD permet de mieux étudier et analyser des signaux issus de systèmes non-linéaires [63].

II. 6. 2. Localité

Dans l'algorithme de la décomposition modal empirique, l'échelle locale est définie par les extrema. En effet, l'enveloppe supérieure/inférieure, en un point donné est calculée par l'interpolation des maxima et des minima locaux. Par conséquent, l'enveloppe supérieure/inférieure est définie localement à l'échelle correspondante à l'espacement entre les maxima/minima.

II. 6. 3. Multirésolution

Une décomposition multi-échelles, ou multirésolution se résulte lors de l'application de la décomposition modale empirique à un signal, de sorte que le processus itératif de l'EMD explore séquentiellement les échelles constitutives du signal analysé de la plus fine, représentée par la première IMF, à la plus grossière, représentée par la dernière IMF (ou le résidu).

II. 6. 4. Auto-adaptativité

L'avantage de la méthode de la décomposition modale empirique est qu'elle donne une décomposition auto-adaptative du signal analysé. En effet, l'EMD est un outil qui n'exige pas a priori des fonctions de base fixes comme les méthodes d'analyse conventionnelles (par exemple la transformée de Fourier et la transformée en ondelette). Cette méthode ne dépend que des points des données du signal à analyser tels que les maxima et les minima.

II. 7. Limitations de la décomposition modale empirique

Bien que l'application de l'algorithme de la décomposition modal empirique soit assez simple et intuitive, elle comporte certains défauts. Nous donnerons par la suite un aperçu de quelques limitations de la méthode de décomposition.

II. 7. 1. Problème lié au choix du critère d'arrêt

Durant l'implémentation de l'algorithme EMD, le choix du critère d'arrêt, qui est lié directement au processus de tamisage, est libre et il se base sur des études empiriques et pas sur des relations mathématiques. Par conséquence, le faux choix du critère d'arrêt permet de limiter la légitimité de la méthode.

II. 7. 2. Problème d'absence d'expression analytique

La décomposition modale empirique comme algorithme n'a pas de fondement mathématique sous-jacent à cette décomposition tels que pour les ondelettes. En plus, il n'existe aucune garantie de la convergence de l'algorithme EMD.

II. 7. 3. Problème de mixage des modes

La méthode de la décomposition modale empirique a été souvent utilisée pour analyser des signaux non-linéaires et/ou non-stationnaires. Néanmoins, les résultats de décomposition montrent des fois un mélange de modes [64, 65]. Ce dernier consiste d'une part à la dispersion des oscillations qui ont une amplitude faible et d'autre part à l'apparition de plus d'une IMF de la même oscillation locale.

II. 8. Autres variantes de la décomposition modale empirique

Afin de résoudre le problème de mélange de modes rencontré pendant l'application de la décomposition modale empirique, des améliorations de l'algorithme EMD ont été proposées [66].

La décomposition modale empirique d'ensemble (EEMD, ensemble EMD) est une technique d'analyse de données contrôlée par un bruit blanc où les fonctions de mode intrinsèques (IMFs) sont déterminées à partir de la moyenne d'un ensemble d'essais, chaque essai est la somme du signal à analyser et d'un bruit blanc.

Malgré que la décomposition des données par l'algorithme d'EEMD permette de résoudre le problème de mélange de modes, l'inconvénient majeur de cette méthode est qu'elle a besoin d'un grand nombre d'essais afin de diminuer l'amplitude relative du bruit blanc additionné. Pour cette raison, plusieurs méthodes qui visent à améliorer l'EEMD ont été proposées [67-69].

II. 9. Conclusion

Dans ce chapitre, nous avons présenté la méthode de décomposition modale empirique, une approche capable d'analyser les données issues de processus non-linéaires et/ou non stationnaires. L'EMD est un algorithme qui peut décomposer le signal en un nombre déterminé de composantes oscillatoires (IMFs) extraites à partir du signal lui-même (elle n'exige pas a priori des fonctions de base).

L'objectif principal de l'EMD est la création d'une analyse locale multi-échelle entièrement pilotée par le signal lui-même. Les fonctions de mode intrinsèques (IMFs) extraites sont classées des hautes fréquences aux basses fréquences. L'algorithme EMD sera utilisé dans la suite pour séparer les différentes composantes du signal de parole.

Chapitre III : Les algorithmes génétiques

III. 1. Introduction

Les algorithmes génétiques (Genetic algorithms : GAs) sont des algorithmes d'optimisation qui utilisent les mêmes processus que ceux intervenant dans l'évolution naturelle telles que la mutation, le croisement et la sélection. Cette technique joue un rôle important dans la résolution de problèmes impliquant des solutions possibles dans un espace de recherche large où on ne peut pas utiliser des méthodes classiques reposant sur une exploration systématique [70]. En plus, la résolution d'un problème en utilisant un algorithme génétique ne nécessite pas la connaissance de la répartition des solutions dans l'espace [71]. Les algorithmes génétiques se sont avérés capables de résoudre de nombreux problèmes complexes où d'autres méthodes ont rencontré des difficultés. Parmi les domaines pratiques auxquels les AGs ont été appliqués avec succès, on trouve le traitement d'image [72], la technologie du laser [73], la médecine [74], l'aéronautique [75], la robotique [76], les aspects architecturaux de la conception des bâtiments [77] et le traitement du signal [78].

Généralement, les AGs ont les avantages suivants [79]:

- L'optimisation est possible pour des variables continues ou discrètes,
- Possibilité de travailler avec un grand nombre de variables,
- Ils sont bien appropriés aux ordinateurs parallèles,
- Ils fournissent une liste de variables optimales, pas seulement une solution simple,
- Ils travaillent avec des données générées numériquement, des données expérimentales ou des fonctions analytiques.

L'objectif de ce chapitre est de présenter le principe des algorithmes génétiques, une approche d'optimisation de l'intelligence artificielle. La méthode est utilisée dans la suite du travail avec la décomposition modale empirique, pour analyser le signal de la parole en vue de quantifier les dyspériodicités vocales.

III. 2. Terminologie biologique

Il est utile d'introduire une partie de la terminologie biologique qui sera utilisée dans cette étude. Dans le contexte des algorithmes génétiques, ces termes biologiques sont utilisés par analogie avec la biologie réelle [80].

Tous les organismes vivants se composent de cellules, chaque cellule contient le même ensemble d'un ou de plusieurs chromosomes (chaînes d'ADN). Conceptuellement, un

chromosome peut être divisé en gènes dont chacun a un code d'une protéine particulier. Par approximation, on peut imaginer un code caractéristique d'un gène comme la couleur des yeux. Les différents paramètres possibles pour un caractère (par exemple, bleu, brun, noisetier) sont appelés allèles. Chaque gène est situé à un locus (position) particulier sur le chromosome.

De nombreux organismes ont des chromosomes multiples dans chaque cellule. La collection complète de matériel génétique (tous les chromosomes pris ensemble) est appelée le génome de l'organisme. Le terme génotype se réfère à l'ensemble particulier de gènes contenus dans un génome. Le génotype donne lieu, sous le développement foetal et ultérieur, au phénotype de l'organisme et ses caractéristiques physiques et mentales telles que la couleur des yeux, la taille et l'intelligence.

Généralement, le terme chromosome dans les algorithmes génétiques se réfère à une solution candidate à un problème donné, souvent codé sous forme d'une chaîne de bits. Les gènes sont soit des bits simples, soit des blocs courts de bits adjacents qui codent un élément particulier de la solution candidate. Un allèle dans une chaîne de bits est soit 0 ou 1. Le croisement consiste à échanger du matériel génétique entre deux parents à un seul chromosome. La mutation consiste à changer un bit à un locus aléatoirement.

La majorité des applications des algorithmes génétiques utilisent des individus à chromosome unique. Le génotype d'un individu dans un algorithme génétique (GA) utilisant des chaînes de bits est la configuration des bits dans le chromosome de cet individu. La figure III.1 représente les cinq niveaux d'organisation d'un algorithme génétique [81] et le tableau III.1 montre ces définitions.

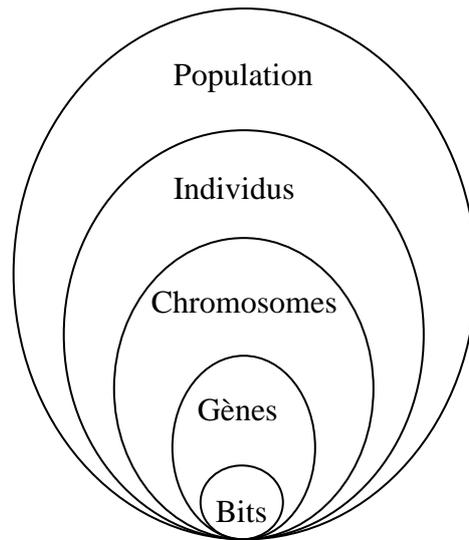


Figure III.1. Niveaux d'organisation d'un algorithme génétique.

Niveau	Représentation
Bit	Unité la plus petite qui représente une donnée.
Gène	Un gène correspond à chaque variable qu'on veut l'optimiser.
Chromosome	Suite de gènes.
Individu	Un ou ensemble de chromosomes.
Population	Ensemble des individus qui évoluent.

Tableau III.1. Représentation des niveaux d'organisation d'un algorithme génétique.

III. 3. Principe général de fonctionnement des algorithmes génétiques

L'objectif d'application d'un algorithme génétique est de trouver les valeurs des paramètres qui maximisent une fonction déterminée [82]. Par exemple, les paramètres génériques d'un modèle de la source glottique, définissent sur un espace de données. La figure III.2 montre l'organigramme général d'un algorithme génétique [83]. En effet, afin d'utiliser ce dernier, nous devons avoir les éléments suivants [84]:

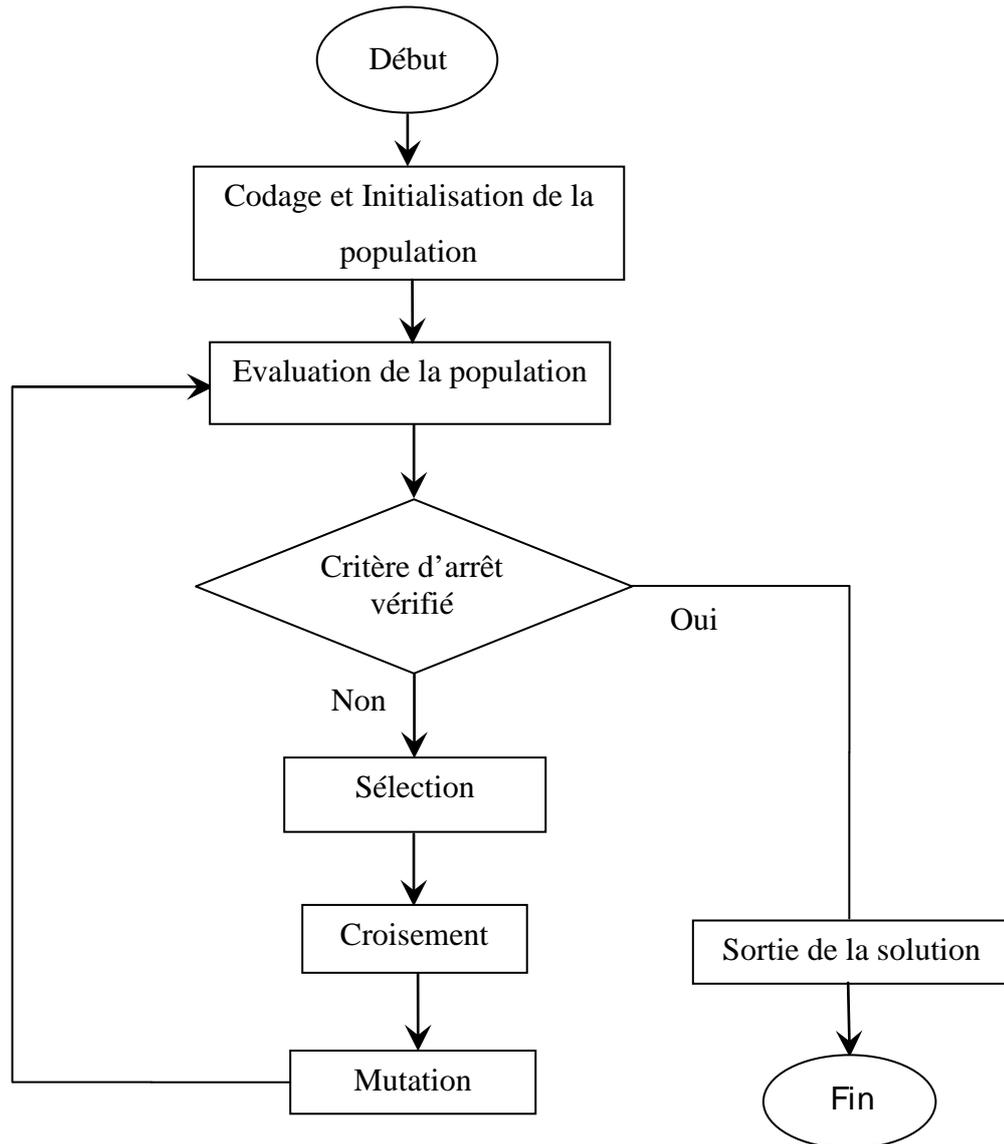


Figure III.2. Organigramme général de l'algorithme génétique.

- Le codage: est la représentation des différents états possibles de la variable dont on cherche la valeur optimale, sous une forme qu'un algorithme génétique peut l'utiliser. Généralement, le codage binaire est le plus utilisé.

- Génération de la population initiale: la population initiale est générée par un mécanisme qui devrait être capable de produire une population non homogène d'individus, ce qui devient la base aux futures générations. La taille de la population (P_s) et le nombre d'individus affectent directement la convergence de l'algorithme génétique.

- Une fonction d'évaluation: pour l'évaluation de chaque individu dans la population.

- Des opérateurs génétiques: durant la production des nouvelles générations, les opérateurs génétiques (mutation, croisement) permettent d'aboutir à une population diversifiée.

III. 4. Description détaillée des algorithmes génétiques

Les algorithmes génétiques en tant que technique pour résoudre les problèmes ont des caractéristiques essentielles qui sont: le codage des données, la fonction d'évaluation, la génération de la population initiale et les opérateurs génétiques.

III. 4. 1. Codage des données

Comme pour toutes les méthodes de recherche et d'apprentissage, le codage des solutions candidates est un facteur essentiel pour la réussite d'un algorithme génétique. La plupart des applications d'un GA utilisent des chaînes de bits de longueur fixe pour coder les solutions candidates [82]. Dans la littérature, il existe plusieurs types de codage des données, les plus utilisés sont: le codage binaire le codage réel.

III. 4. 1. 1. Codage binaire

Le codage binaire (chaînes de bits) est capable de coder facilement tous les types de paramètres du problème traité telles que les variables entières, réelles ou les chaînes de caractères [85]. Ce type de codage rend les opérations de mutation et de croisement plus faciles. Dans cette étude, ce type de codage sera utilisé pour rechercher les paramètres optimaux d'un modèle de la source glottique.

III. 4. 1. 2. Codage réel

Chaque chromosome est représenté par un vecteur dont les composantes sont les paramètres du processus à optimiser. En effet, pour l'optimisation d'une fonction $f(x_1, x_2, \dots, x_{n-1}, x_n)$ de n variables, on peut utiliser un chromosome qui contient n variables: $(x_1, x_2, \dots, x_{n-1}, x_n)$ [71].

III. 4. 2. Fonction d'évaluation

L'évaluation des individus est effectuée à partir d'une fonction d'évaluation qui s'appelle fonction coût (fitness, en anglais). Le calcul de cette fonction est indépendant de celui des autres individus. Le résultat obtenu par cette fonction permettra de sélectionner ou d'exclure un individu afin de ne garder que les individus ayant le meilleur coût dans la

population courante. L'utilisation de la fonction d'évaluation assure que les individus performants seront conservés, alors que les individus peu adaptés seront progressivement éliminés de la population [85].

Dans notre étude, la fonction d'évaluation sera une fonction d'un modèle mathématique (modèle LF, cf. § I.3.2.2) qui permet de générer la source glottique.

III. 4. 3. Génération de la population initiale

La rapidité de l'algorithme génétique est liée fortement au choix de la population initiale d'individus [85]. Habituellement, le choix de la population initiale d'individus est fait de manière aléatoire. Cependant, il existe d'autres potentialités d'initialisation comme la possibilité d'effectuer une série d'initialisations pour chaque individu, puis choisir les valeurs les plus performantes. Sinon, des estimations peuvent être faites par d'autres méthodes au bout d'une tentative pour localiser des solutions approchées, et l'algorithme peut être lancé à partir de tels points [86] et cette valeur sera utilisée comme une base pour les générations futures. Généralement, le nombre de générations est compris entre 50 et 100 [70].

III. 5. Opérateurs génétiques

Durant l'application des algorithmes génétiques, les individus ayant une meilleure performance sont choisis comme parents pour la prochaine génération. En utilisant les algorithmes génétiques, de nouveaux individus seront créés via des opérateurs randomisés simples qui sont la mutation et le croisement dans des organismes naturels [87]. Les nouvelles solutions sont évaluées avec la fonction d'évaluation et le cycle de sélection, de croisement et de mutation est répété jusqu'à ce qu'un critère de terminaison défini par l'utilisateur soit satisfait.

III. 5. 1. Sélection

L'objectif de l'opérateur de sélection est d'éliminer les mauvais individus et d'identifier les meilleurs individus de la population (enfant) à l'état présent qui seront autorisés à se reproduire (parents). La sélection est basée sur la qualité des individus, calculée à l'aide de la fonction d'évaluation. Cette opération est très importante parce qu'elle permet aux individus d'une population de survivre ou de mourir. Généralement, l'individu qui a la probabilité de survie est celui qui a une efficacité relative au sein de la population. Dans la littérature, il existe plusieurs méthodes de sélection, on citera à titre d'exemple [70, 86, 88]:

- La sélection par roulette ou proportionnelle.
- La sélection uniforme ou aléatoire.

Dans notre étude, nous avons utilisé la dernière méthode (sélection uniforme) puisque elle est très simple du fait qu'un individu de la population est sélectionné de manière aléatoire. Un individu sera sélectionné à une probabilité P_i définie comme suit:

$$P_i = \frac{1}{N_p} \quad (\text{III.1})$$

où N_p est la taille de la population.

III. 5. 2. Croisement

L'opération de croisement consiste à la création des nouveaux chromosomes enfants, différents de ceux déjà existants, à partir des chromosomes parents sélectionnés. La façon la plus simple d'effectuer le croisement est de choisir aléatoirement un point de croisement et copier tout ce qui avant ce point à partir du premier parent, puis copier tout ce qui après le point de croisement de l'autre parent [83]. Par exemple le chromosome A = 10000100 et le chromosome B = 11111111 pourraient être croisées après le troisième locus pour créer les deux nouveaux chromosomes C et D tel que le chromosome C = 10011111 et le chromosome D=11100100.

Dans la littérature, il existe plusieurs types de croisement, on citera à titre d'exemple [85]:

- Le croisement à un point.
- Le croisement à deux points.
- Le croisement uniforme.

Dans cette étude nous avons utilisé le croisement uniforme, puisque ce dernier permet de mieux naturaliser les gènes par rapport aux autres types de croisement [89].

L'opération de croisement est appliquée à la population avec une probabilité constante (P_c). En général, le choix du P_c est très important et dépend de l'application. Habituellement, P_c est compris entre 0,5 et 0,95 [89].

III. 5. 3. Mutation

L'opération de mutation consiste à la création des nouveaux chromosomes où certains bits de ces derniers sont changés. Pour le codage binaire, nous pouvons changer quelques bits sélectionnés aléatoirement de 0 à 1 ou de 1 à 0. Par exemple le chromosome 1= 00000100 pourrait être changé en seconde position pour créer le chromosome 2=01000100.

Généralement, pour l'opérateur de mutation, il est préférable d'utiliser une faible probabilité de mutation (P_m). Habituellement, P_m est compris entre 0,01 et 0,3 [90, 91].

III. 6. Conclusion

Dans ce chapitre, une étude descriptive des algorithmes génétiques a été présentée, nous avons introduit une partie des termes biologiques qui seront utilisés au cours de cette thèse.

Nous avons décrit brièvement le principe de fonctionnement des algorithmes génétiques ainsi que ces caractéristiques essentielles tel quels le codage des données, la fonction d'évaluation, la génération de la population initiale et les opérateurs génétiques.

Dans les chapitres suivants, une méthode d'analyse du signal de parole pour l'évaluation automatique des troubles de la voix basée sur les algorithmes génétiques et la décomposition modale empirique sera proposée. Les GAs seront utilisés pour rechercher les paramètres génériques optimaux du modèle LF de la source glottique.

*Chapitre IV : Estimation des dyspériodicités
vocales par décomposition modale empirique*

IV. 1. Introduction

L'évaluation clinique objective de la voix présente une grande importance pour les cliniciens les spécialistes de la voix du fait qu'elle permet d'évaluer et suivre l'évolution de la voix des patients de manière fiable.

Dans ce chapitre, on propose une approche d'évaluation des troubles de la voix basée sur une estimation de la source glottique par décomposition modale empirique. Deux indices acoustiques pour quantifier le degré de perturbation dans le signal de parole sont proposés. Le premier indice fournit une mesure directe des dyspériodicités vocales à partir du signal de la source glottique alors que le second indice donne une mesure indirecte des dyspériodictés vocales par l'intermédiaire des paramètres d'un modèle de génération de la source glottique adapté au signal de la source glottique estimé.

IV. 2. Analyse des dyspériodicités vocales basée sur la décomposition modale empirique dans le domaine log-spectrale

Une méthode proposée récemment pour l'évaluation des troubles de la voix est basée sur la décomposition modale empirique dans le domaine log-spectrale [10, 11]. En utilisant l'algorithme EMD, le logarithme du spectre du signal de parole est décomposé en ces trois composantes : la composante harmonique, l'enveloppe spectrale et le bruit. L'indice acoustique appelé rapport harmonique sur bruit (HNR : harmonic-to-noise ratio) est utilisé pour quantifier le degré de perturbation de la parole produite par les locuteurs dysphoniques.

IV. 2. 1. Séparation des composantes du signal parole par la décomposition modale empirique

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 $v(t)$ [40]:

$$x(t) = e(t) * v(t) \quad (\text{IV.1})$$

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 :

$$|X_w(f)| = |E_w(f) \times V(f)| \quad (\text{IV.2})$$

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 (IV.2), l'opération de multiplication devient une addition:

$$\log |X_w(f)| = \log |E_w(f)| + \log |V(f)| \quad (\text{IV.3})$$

L'équation (IV.3) 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, le spectre du signal d'excitation lui-même peut être considéré comme composé de deux parties: la première partie est une série régulièrement espacée d'harmoniques ayant une amplitude décroissante avec la fréquence et la deuxième partie est un bruit irrégulièrement distribué.

Le logarithme du spectre d'amplitude du signal de parole peut être considéré comme composé d'un contour à variation lente (par rapport à la fréquence) en raison de la contribution du conduit vocal, d'une série d'harmoniques caractérisées par une structure périodique et d'une partie irrégulière à variation rapide due au bruit d'aspiration à la glotte. L'algorithme de la décomposition modale empirique donne un outil qui permet de séparer les trois composantes du logarithme du spectre d'amplitude du signal de parole. En effet, l'algorithme EMD agit comme un banc de filtres [92], de sorte que la décomposition du logarithme du spectre d'amplitude du signal de parole en utilisant l'algorithme EMD fournit plusieurs composantes oscillantes (IMFs) qui peuvent être regroupées en trois classes par une opération de seuillage simple et chaque classe de composantes est affectée à une certaine partie du logarithme du spectre d'amplitude du signal de parole. Soit f_j la fréquence moyenne de la composante IMF_j du logarithme du spectre d'amplitude du signal de parole obtenu par l'algorithme EMD. Les différentes IMFs sont regroupées en fonction de leurs fréquences moyennes comme suit [10]:

Classe 1: $f_j < s_1$: IMF_j appartient à la partie enveloppe.

Classe 2: $s_1 < f_j < s_2$: IMF_j appartient à la partie harmonique.

Classe 3: $f_j > s_2$: IMF_j appartient à la partie bruit.

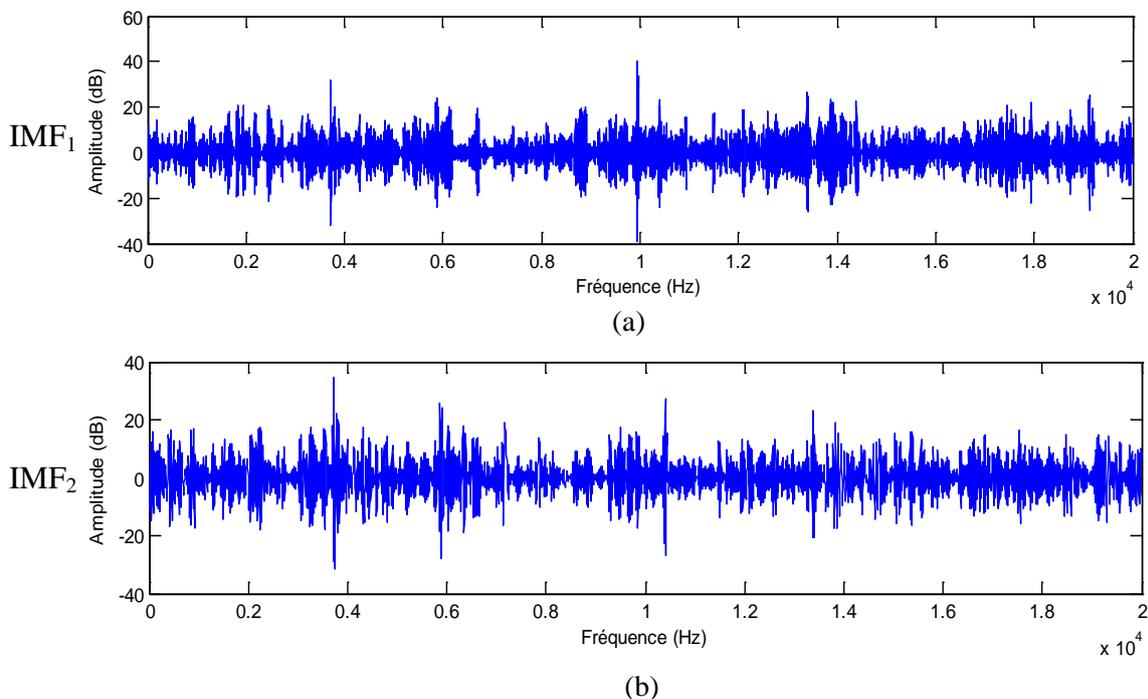
où s_j , $j = 1, 2$ sont des seuils qui dépendent de la fréquence fondamentale moyenne f_0 du signal de parole et elles sont fixées de façon empirique.

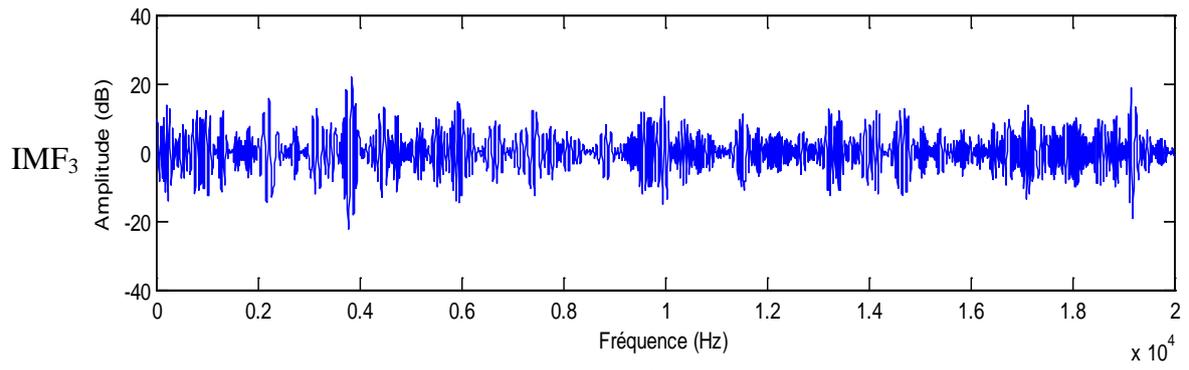
Chaque partie du logarithme du spectre d'amplitude du signal de parole est estimée en faisant la somme des IMFs appartenant à la classe correspondante. Dans [10], il a été montré que les seuils optimaux sont : $s_1=0.3/f_0$ et $s_2=4/f_0$.

IV. 2. 2. Exemple de séparation des composantes du signal parole par la décomposition modale empirique

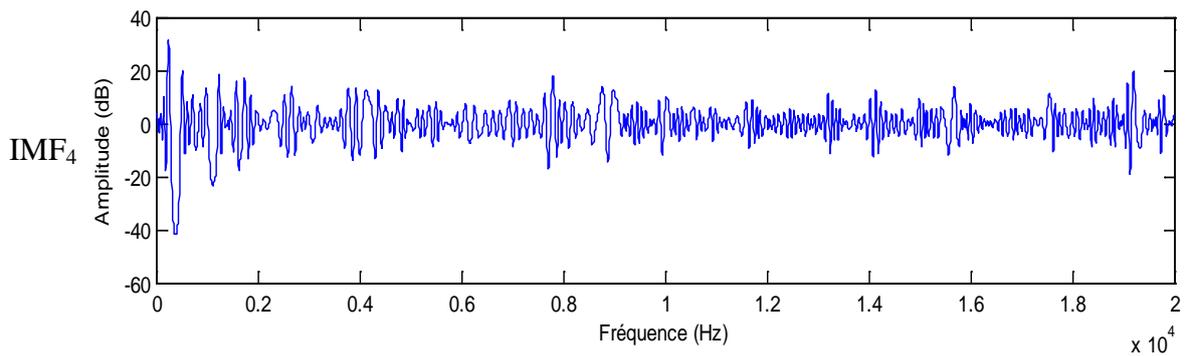
Comme illustration de l'efficacité de l'algorithme EMD pour la décomposition des composante du signal parole, la figure IV.1 et la figure IV.2 montrent les IMFs et les différentes composantes estimées [93] du logarithme d'amplitude de spectre d'une trame de longueur de 200 ms d'un signal de parole extraite d'une voyelle [a] soutenue produite par un locuteur dysphonique successivement.

Dans cet exemple, La décomposition du logarithme d'amplitude de spectre de la trame analysée via un algorithme EMD produit 10 IMFs. La composante enveloppe spectrale est constituée des quartes dernier IMFs, la composante harmonique est constituée des IMF_3, \dots, IMF_6 et la composante bruit est constituée des IMF_1 et IMF_2 .

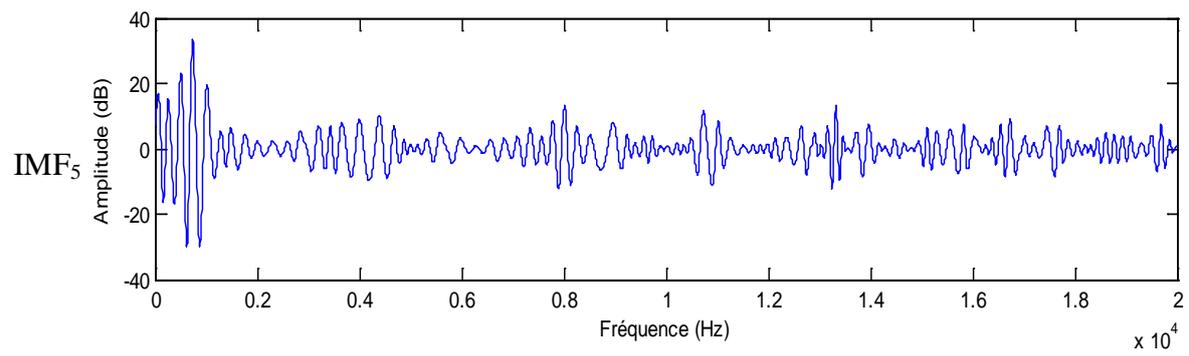




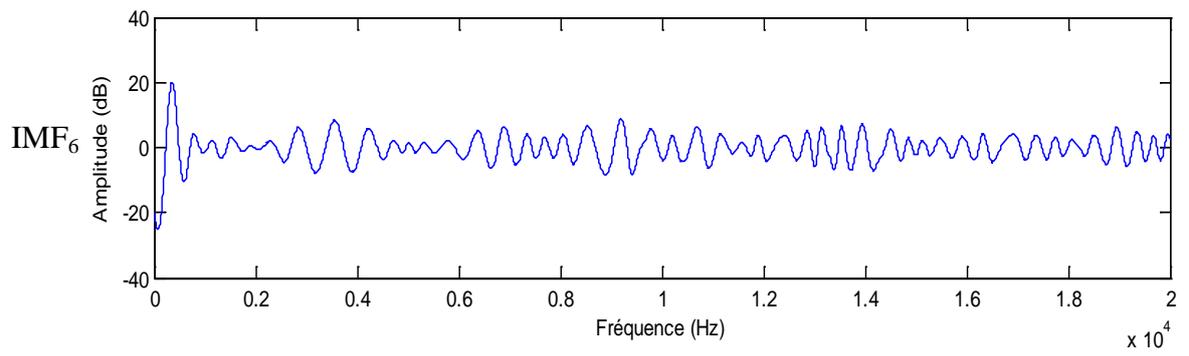
(c)



(d)



(e)



(f)

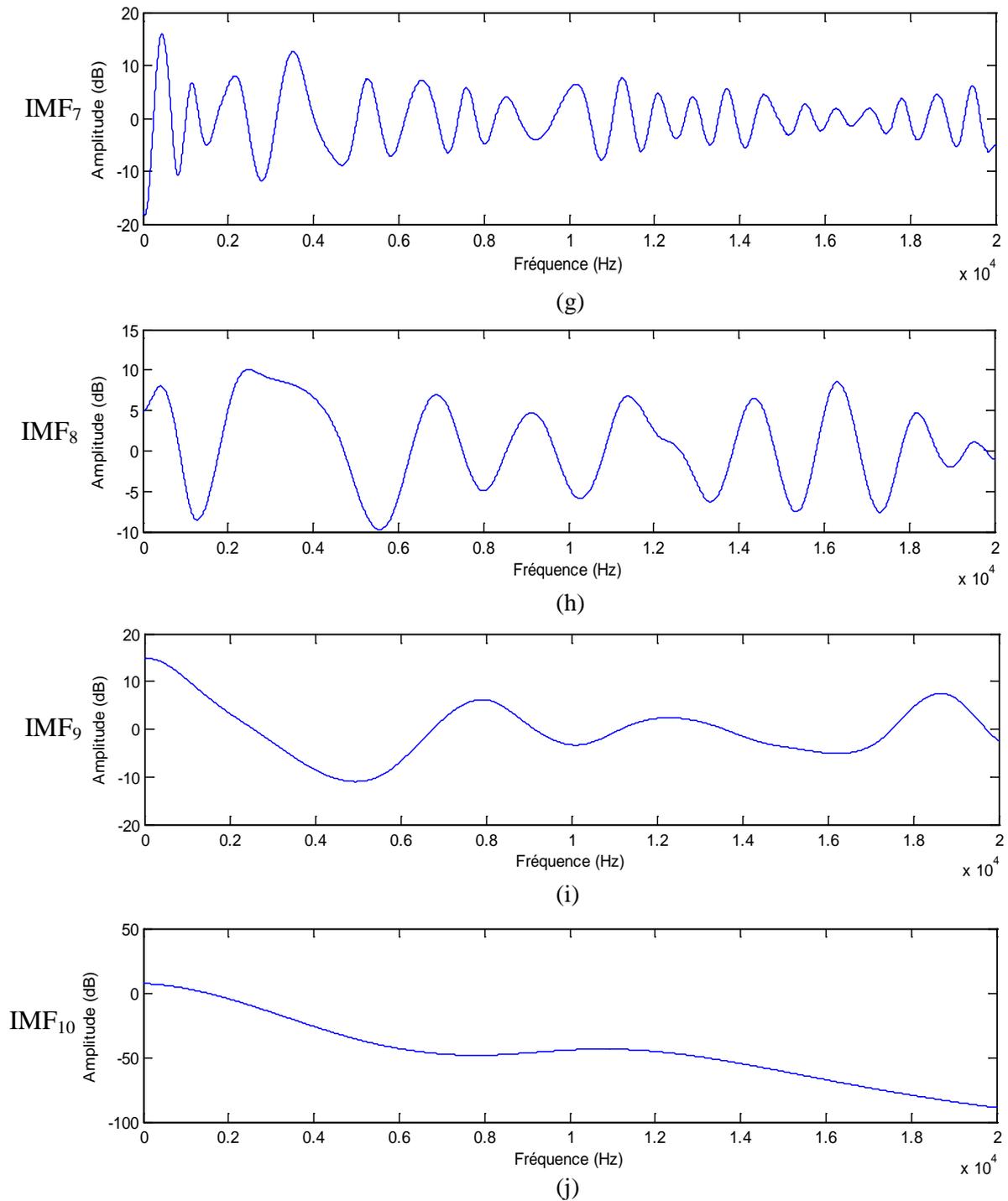
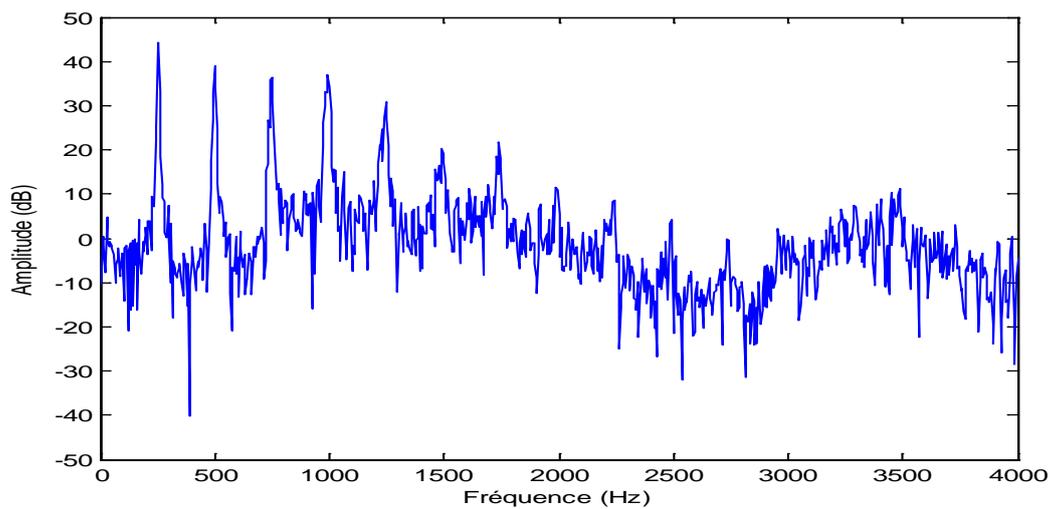
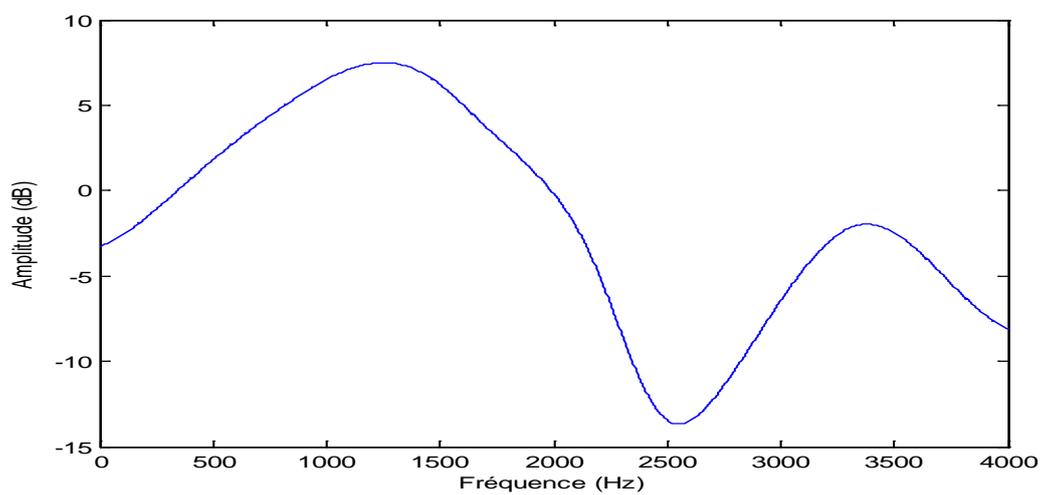


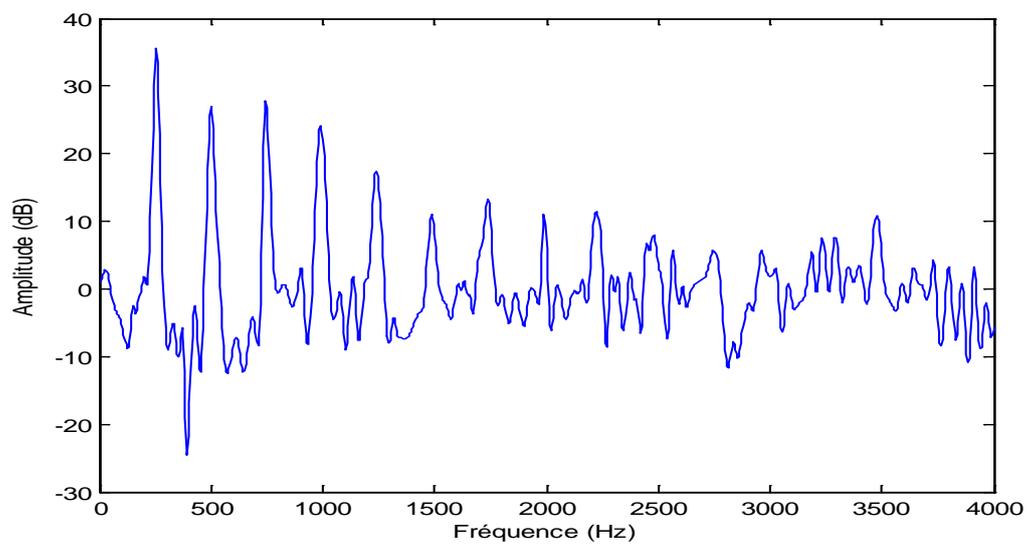
Figure IV.1. Décomposition modale empirique du logarithme du spectre d'amplitude d'une trame de longueur 200 ms extraite d'une voyelle [a] soutenue produite par un locuteur dysphonique.



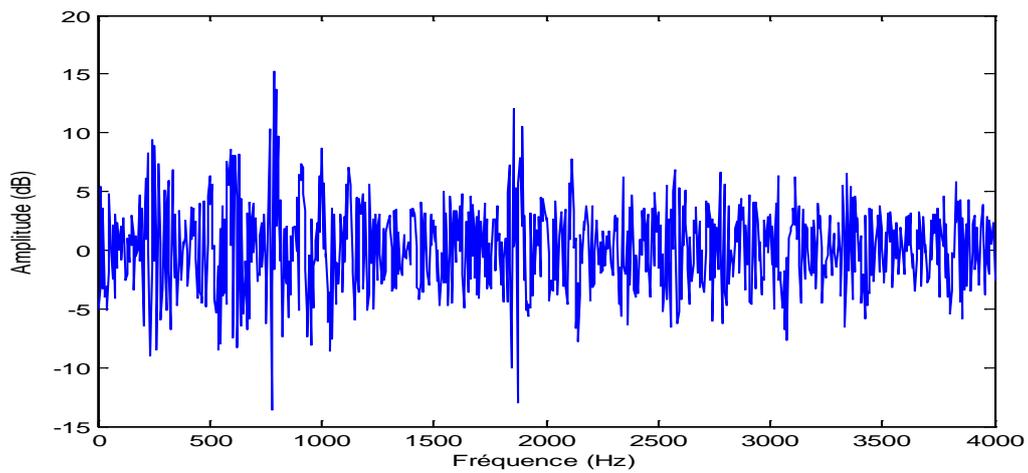
(a)



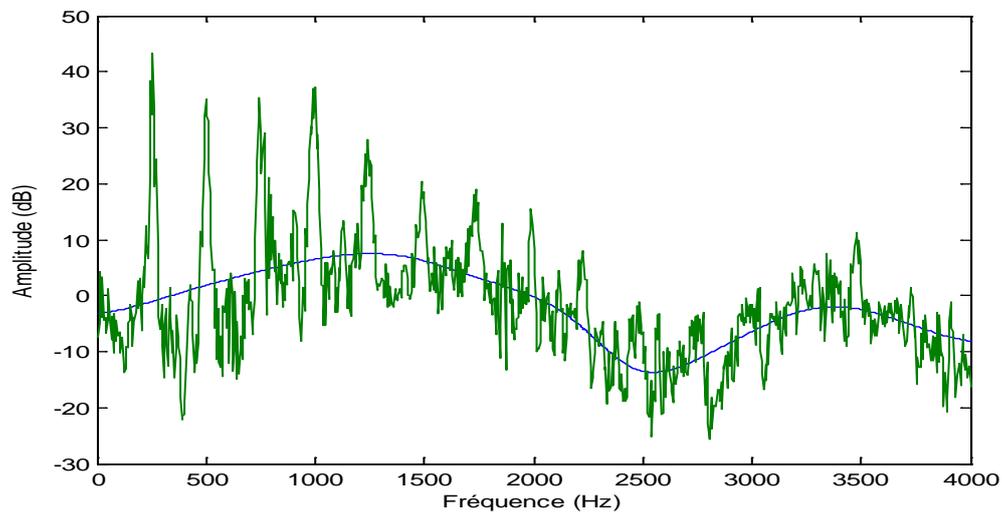
(b)



(c)



(d)



(e)

Figure IV.2. Décomposition du logarithme du spectre d'amplitude d'une trame de longueur 200 ms extraite d'une voyelle [a] soutenue produite par un locuteur dysphonique en trois composants via l'algorithme EMD. (a) log du spectre d'amplitude la trame du signal de parole. (b) enveloppe spectrale. (c) partie harmonique. (d) partie bruit. (e) somme des trois composants superposée à l'enveloppe estimée.

IV. 2. 3. Rapport harmonique sur bruit

L'indice acoustique utilisé pour quantifier le degré de perturbation dans le signal de parole est le rapport harmonique sur bruit (HNR). Pour un signal parole donné, l'intervalle d'analyse est divisé en L trames et l'indice acoustique global HNR est calculé comme la moyenne des HNR_i ($i = 1, \dots, L$) des L trames [10] :

$$HNR = \frac{1}{L} \sum_1^L HNR_i \quad (IV.4)$$

où

$$HNR_i = 10 \log \left[\frac{\sum_{k=0}^{M-1} H^2(k)}{\sum_{k=0}^{M-1} N^2(k)} \right], i = 1, \dots, L \quad (IV.5)$$

Et $H(k)$ désigne le spectre d'amplitude de la composante harmonique, $N(k)$ le spectre d'amplitude de la composante bruit et M dénote le nombre de points fréquentiels.

IV. 3. Evaluation des troubles de la voix basée sur l'estimation de la source glottique

Cette méthode d'estimation des dyspériodicités vocales utilise l'algorithme de décomposition en mode empirique pour estimer le signal d'excitation du signal de parole. L'indice acoustique nommé rapport excitation à dyspériodicité segmental (EDRSEG) est proposé pour quantifier le degré de perturbation dans le signal de parole.

IV. 3. 1. Estimation du signal de la source glottique

En utilisant l'algorithme de la décomposition modale empirique, le logarithme du spectre d'amplitude du signal acoustique est décomposé en modes oscillatoires nommés fonctions de mode intrinsèque qui sont regroupées en deux classes: l'enveloppe spectrale et la composante harmonique par un simple seuillage. L'exploitation de l'information de phase conjointement avec la composante harmonique estimée permet d'estimer le signal d'excitation. L'organigramme montré à la figure IV.3 illustre les différentes étapes de la méthode [94].

Pour illustrer la méthode d'estimation de la source glottique, une voyelle [a] synthétique a été simulée en utilisant le modèle source-filtre en utilisant le modèle de Klatt

pour modéliser la source glottique [24]. La figure IV.4-(a) montre le signal d'excitation original superposé au signal d'excitation estimé pour une trame de 50 ms extraite d'une voyelle [a] synthétique produite par un locuteur normophonique. Comme on l'observe, une composante haute fréquence est présentée dans le signal d'excitation estimé. Cette composante est interprétée comme un artefact lié à la méthode de décomposition. Elle peut être atténuée en utilisant un filtre passe-bas. Les figures IV.4-(b) et IV.4-(c) montrent le signal d'excitation estimé filtré par un filtre passe-bas de Butterworth d'ordre deux avec une fréquence de coupure de 2000 Hz et 1000 Hz, respectivement. Dans cette étude, un filtre passe-bas de Butterworth d'ordre deux avec une fréquence de coupure de 1000 Hz a été appliqué.

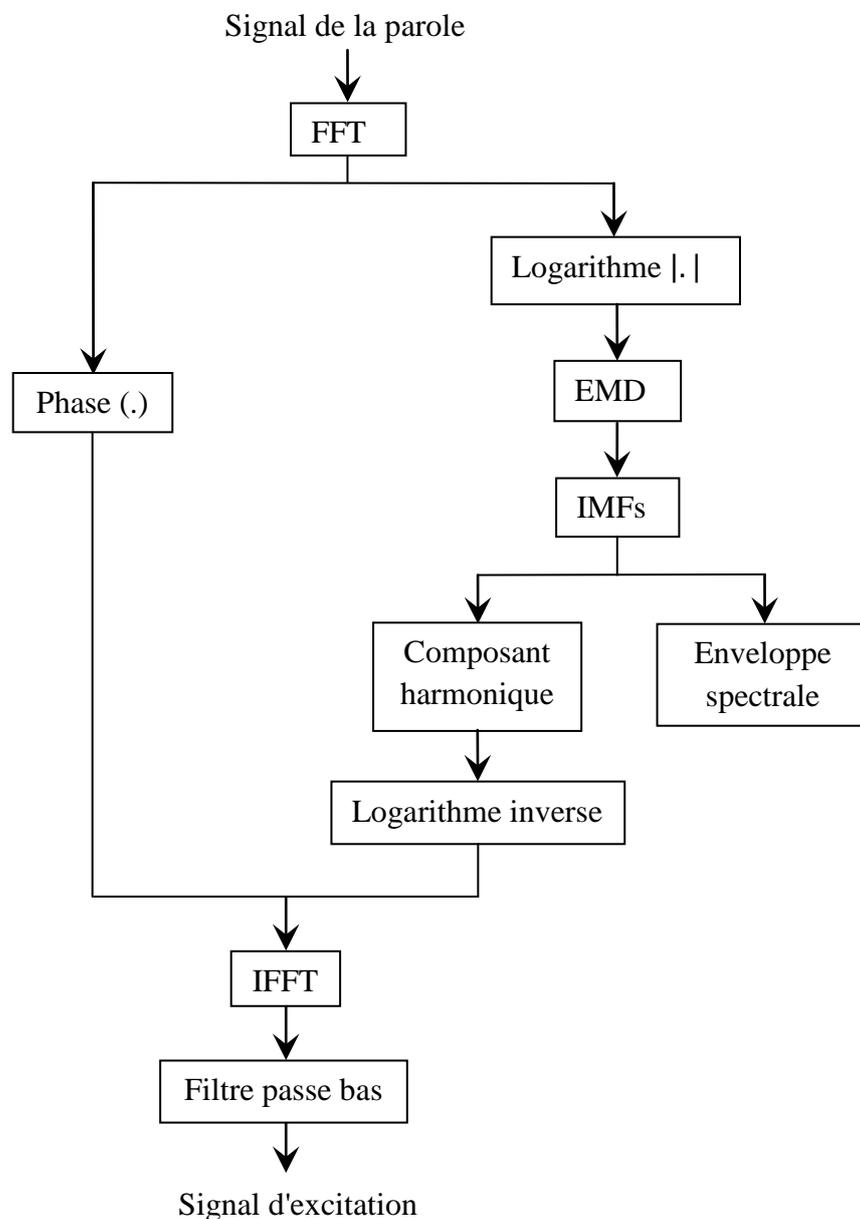
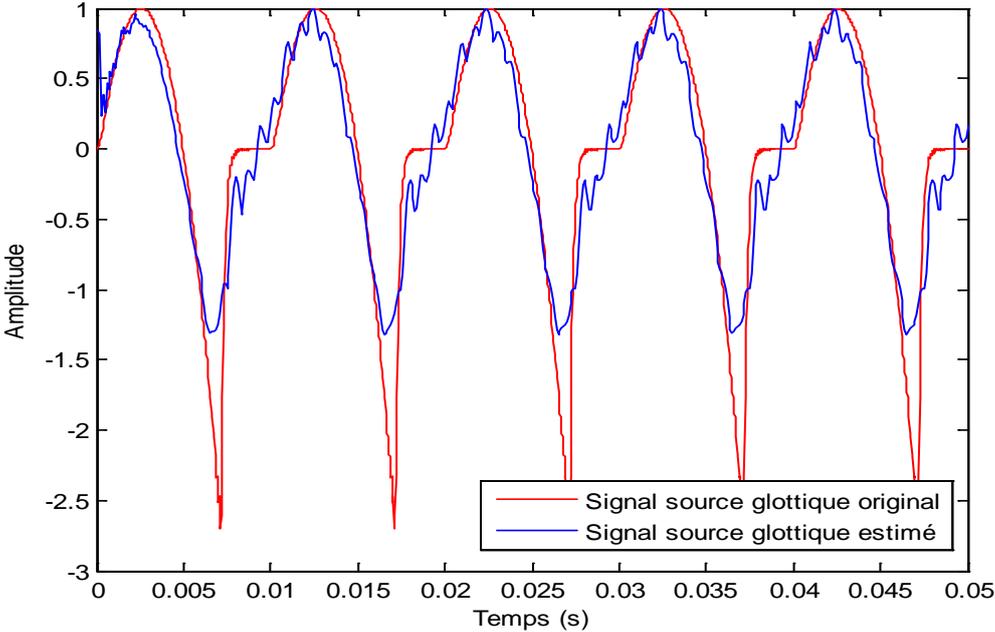
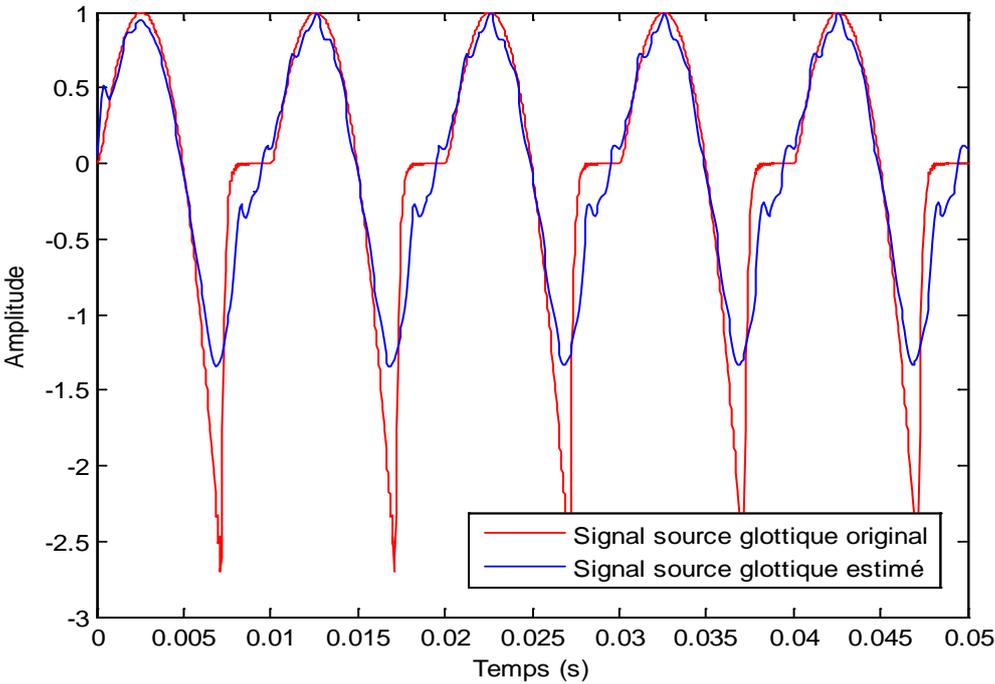


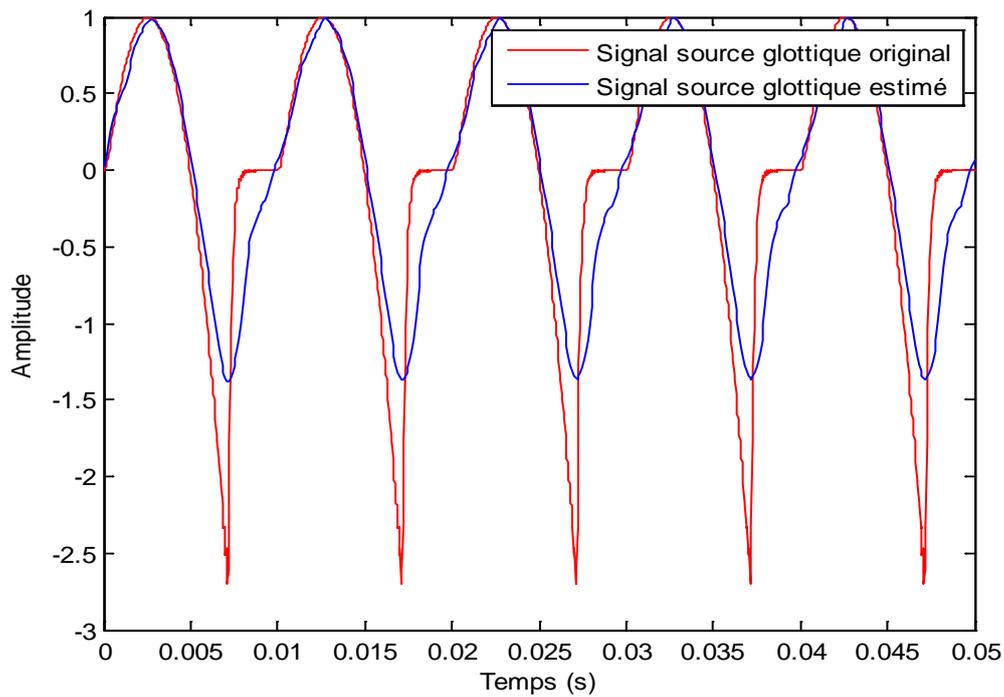
Figure IV.3. Estimation du signal d'excitation.



(a)



(b)



(c)

Figure IV.4. Signal d'excitation original superposé au signal de la source glottique estimé pour une trame de 50 ms extraite d'une voyelle [a] synthétique produite par un locuteur normophonique (a) sans filtrage, (b) avec filtrage passe-bas à 2000 Hz et (c) avec filtrage passe-bas à 1000 Hz.

IV. 3. 2. Estimation des dyspériodicités vocales

Les dyspériodicités vocales sont estimées à partir du signal d'excitation en utilisant le variogramme généralisé. Pour un signal d'excitation $e(n)$ périodique de période T_0 , on peut écrire

$$e(n) = e(n - T_0) \quad (\text{IV.6})$$

Une mesure de la déviation du signal d'excitation par rapport à la périodicité, calculée sur un intervalle de longueur N , permet de quantifier numériquement le degré d'irrégularité du signal. L'énergie de la dyspériodicité dans le signal d'excitation peut être évaluée comme suit [95]:

$$\hat{\gamma} = \min_T \left\{ \sum_{n=0}^{N-1} [e(n) - e(n-T)]^2 \right\} \quad (\text{IV.7})$$

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

où T_{min} et T_{max} sont, respectivement la longueur et le décalage de la trame d'analyse.

L'expression entre accolades dans (IV.7) est appelée variogramme du signal $e(n)$. Formellement, l'expression entre crochets 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 d'excitation à l'intérieur de la trame d'analyse.

Les signaux vocaux sont censés d'être localement stationnaire au mieux. L'amplitude du signal évolue d'une trame à la suivante en raison des attaques et des déclins, de l'intensité des segments et de l'accentuation. Afin de tenir compte de ces changements lents de l'amplitude du signal, un facteur de pondération (un gain) est introduit. La relation (IV.6) devient:

$$e(n) = a e(n - T_{opt}), \quad 0 \leq n \leq N-1. \quad (\text{IV.8})$$

En conséquence, le variogramme empirique généralisé peut être écrit comme suit [96]:

$$\hat{\gamma} = \min_T \left\{ \sum_{n=0}^{N-1} [e(n) - a e(n-T)]^2 \right\} \quad (\text{IV.9})$$

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

Le gain a est forcé d'être positif. Il est défini de manière à égaliser les énergies du signal dans les trames d'analyse courante et décalée:

$$a = \sqrt{\frac{E_0}{E_T}}, \quad E_0 = \sum_{n=0}^{N-1} e^2(n), \quad E_T = \sum_{n=0}^{N-1} e^2(n-T). \quad (\text{IV.10})$$

Où E_0 et E_T sont, respectivement, les énergies des trames d'analyse courante et décalée. L'estimation de la valeur instantanée de la dyspériodicité est donnée comme suit:

$$d(n) = e(n) - a e(n - T_{opt}), \quad 0 \leq n \leq N-1, \quad (\text{IV.11})$$

où T_{op} est le délai qui minimise le variogramme généralisé pour la position courante de la trame. La valeur de T_{op} peut être positive ou négative. L'énergie de la dyspériodicité par trame est définie comme suit:

$$E = \sum_{n=0}^{N-1} \left[e(n) - ae(n - T_{opt}) \right]^2 \quad (IV.12)$$

IV. 3. 3. Rapport excitation à dyspériodicité segmental

Les indices acoustiques utilisés pour quantifier les dyspériodicités vocales dans le signal de parole sont le rapport signal à dyspériodicité global et le rapport signal à dyspériodicité segmental. Dans le cadre de cette étude, l'indice acoustique nommé rapport excitation à dyspériodicité segmental (EDRSEG : segmental excitation-to-dyspériodicity ratio) est proposé comme alternative pour quantifier les dyspériodicités vocales dans le signal d'excitation [94].

Le rapport excitation à dyspériodicité segmental est inspiré de la définition du rapport signal à dyspériodicité segmental en substituant l'énergie du signal de la parole par l'énergie du signal d'excitation. Par conséquent, pour un signal d'excitation donné, l'intervalle d'analyse est divisé en K trames de longueur N et le rapport excitation à dyspériodicité segmental est calculé comme suit:

$$EDRSEG = \frac{10}{K} \sum_{k=0}^{K-1} \log \frac{\sum_{n=kN}^{kN+N-1} e^2(n)}{\sum_{n=kN}^{kN+N-1} d^2(n)} \quad (IV.13)$$

où $e(n)$ désigne le signal d'excitation et $d(n)$ est la dyspériodicité obtenue par l'expression (IV.11).

IV. 4. Evaluation des troubles de la voix basée sur l'optimisation d'un modèle de la source glottique

Dans cette partie, on propose une méthode d'évaluation des troubles de la voix basée sur l'optimisation des paramètres d'un modèle de la source glottique en utilisant les algorithmes génétiques.

IV. 4. 1. Méthode

La procédure générale de l'approche proposée pour l'évaluation des troubles de la voix est décrite par l'organigramme de la figure IV.5. Cette méthode d'analyse utilise l'algorithme de l'EMD pour estimer le signal de source glottique à partir du signal de parole [10, 11]. Au moyen de l'algorithme EMD, le logarithme du spectre d'amplitude du signal de parole est décomposé en modes d'oscillations (IMFs) qui sont regroupés en deux classes: l'enveloppe spectrale et la composante harmonique par un simple seuillage. L'exploitation de l'information de phase conjointement avec la composante harmonique estimée permet l'estimation du signal de source glottique. Un modèle paramétrique de génération de la source glottique est adapté au signal de source glottique estimée, et en utilisant les algorithmes génétiques (AGs), on obtient les paramètres optimaux du modèle générique du signal de la source glottique.

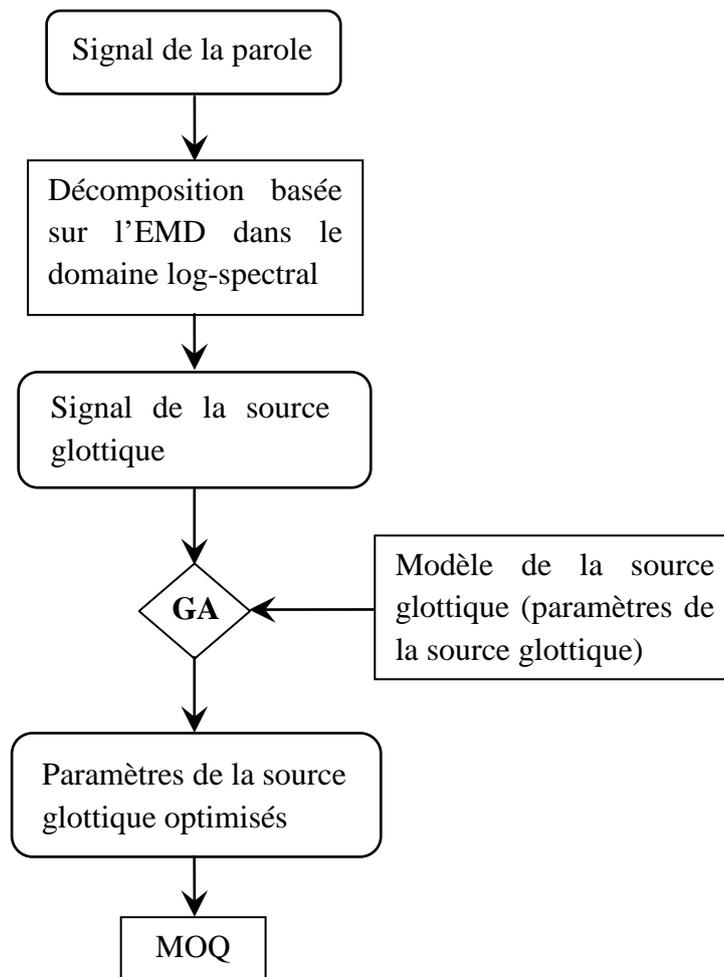


Figure IV.5. Procédure d'estimation du quotient d'ouverture moyen MOQ à partir du signal acoustique.

L'indice acoustique nommé quotient d'ouverture moyen (MOQ : mean opening quotient) est proposé pour quantifier le degré de perturbation dans le signal de parole. Les différentes étapes de la méthode proposée sont détaillées dans les sections suivantes [97].

IV. 4. 2. Optimisation des paramètres du modèle de la source glottique par les algorithmes génétiques

La méthode d'optimisation des paramètres du modèle de la source glottique en utilisant les algorithmes génétiques est présentée à la figure IV.6. Les paramètres du modèle à optimiser sont la phase d'ouverture T_p et la phase ouverte ($T_e - T_p$) du modèle de la source glottique définies par Liljencrants-Fant [23]. Les valeurs des différents paramètres de l'AG utilisées pour l'optimisation des phases d'ouverture et la phase ouverte sont les suivantes:

- L'intervalle des variables à optimisées (T_p et T_e): $[0 - T_0]$,
- La taille de la population: 20,
- Croisement: 0.75,
- Mutation: 0.1,
- Génération: 100.

Les valeurs initiales de la population sont générées aléatoirement, chaque individu de cette population est évalué en utilisant la fonction d'évaluation δ définie comme étant la différence absolue entre le modèle du signal de la source glottique $e_g(t)$ et le signal de la source glottique $e(t)$ estimé par la méthode EMD.

$$\delta = \sum_{n=0}^N |e_g(n) - e(n)| \quad (\text{IV.14})$$

IV. 4. 3. Quotient d'ouverture moyen

L'indice acoustique que nous proposons pour quantifier le degré de perturbation dans le signal de parole est le quotient d'ouverture moyen (MOQ) défini comme étant la moyenne du rapport de la longueur de phase d'ouverture à une période fondamentale. Pour un signal donné, l'intervalle d'analyse est divisé en L trames. Le quotient d'ouverture moyen est calculé comme [97]:

$$MOQ = \frac{1}{L} \left[\frac{\sum_{i=1}^L (T_p)_i}{T_0} \right], i = 1, \dots, L \quad (\text{IV.15})$$

où T_0 et $(T_p)_i$ sont, respectivement, la période fondamentale et la longueur de phase d'ouverture optimisée de la trame i .

L'indice acoustique MOQ reflète l'augmentation du débit glottique de l'état initial à son amplitude maximale qui correspond à la fermeture des cordes vocales. Si la fermeture des cordes vocales est altérée en raison d'un dysfonctionnement, cela affecte le quotient d'ouverture moyen.

L'indice acoustique MOQ fournit une indication sur la fermeture des cordes vocales et peut être utilisé comme descripteur acoustique pour l'évaluation des troubles de la voix. Il est prévu que les locuteurs normophoniques et dysphoniques présentent des mesures différentes du quotient d'ouverture moyen qui reflètent le degré de sévérité du trouble de la voix.

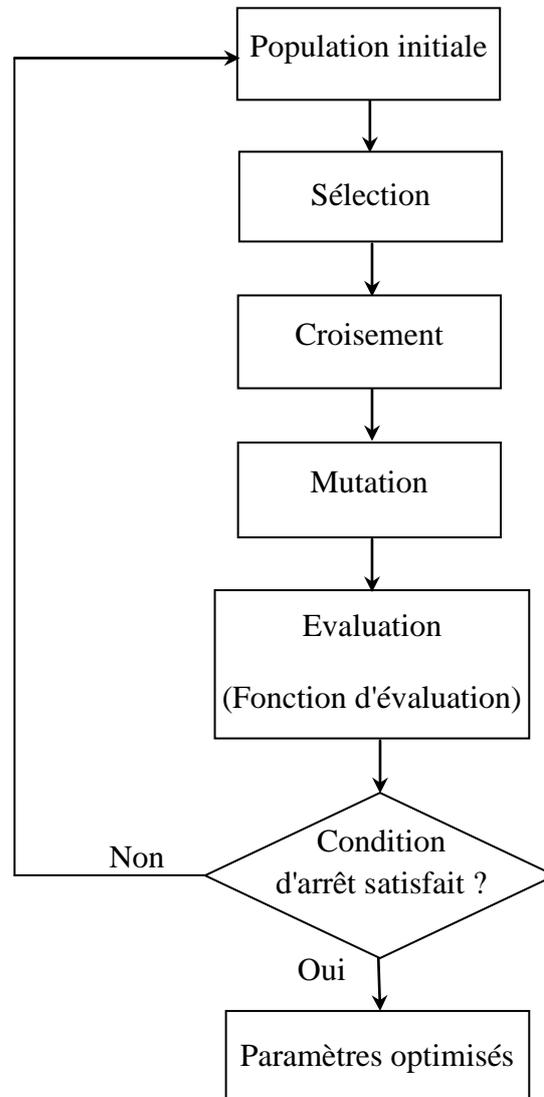


Figure IV.6. Estimation des paramètres de la source glottique par les GAs.

IV. 4. 4. Combinaison de la méthode proposée avec l'analyse cepstrale

Dans le but d'améliorer la performance de l'indice acoustique en terme de corrélation avec l'enrouement perçu, on se propose d'utiliser la proéminence du pic cepstral (CPP) et le quotient d'ouverture moyen (MOQ) comme variables prédicteurs pour prédire le score de l'enrouement perçu.

Le score de l'enrouement est exprimé comme suit:

$$Score = a_0 + a_1 MOQ_i + a_2 CPP_i \quad (IV.17)$$

Les paramètres a_i sont les coefficients de régression obtenus en minimisant l'erreur quadratique moyenne de la prédiction du score de l'enrouement perçu comme une combinaison linéaire des deux indices acoustiques MOQ et CPP pour les différentes stimuli.

IV. 5. Conclusion

Dans ce chapitre, trois méthodes d'analyse du signal de parole pour l'évaluation automatique des voix dysphoniques ont été présentées. Ces méthodes sont basées sur la séparation des composantes du signal parole en utilisant la méthode EMD :

La première méthode est une approche d'analyse des voix désordonnées basée sur la décomposition modale empirique (EMD) dans le domaine log spectrale. En utilisant l'algorithme EMD, le logarithme du spectre du signal de parole est décomposé en ces trois composantes : la composante harmonique, l'enveloppe spectrale et le bruit. Le rapport harmonique sur bruit (HNR : harmonic-to-noise ratio) a été utilisé pour quantifier le degré de perturbation de la parole produite par les locuteurs dysphoniques.

La deuxième méthode d'estimation de la dyspériodicité vocale utilise l'algorithme de décomposition en mode empirique pour estimer le signal d'excitation du signal de parole. Le marqueur acoustique nommé rapport excitation à dyspériodicité segmental (EDRSEG) a été proposé pour quantifier le degré de perturbation dans le signal de parole.

La troisième méthode d'estimation de la dyspériodicité vocale a été basée sur l'optimisation des paramètres du modèle LF de la source glottique. L'indice acoustique nommé quotient d'ouverture moyen (MOQ) a été proposé pour quantifier le degré de perturbation dans le signal de parole.

Dans le chapitre prochain, nous allons appliquer les performances des différentes méthodes d'analyse et des indices acoustiques en les testant sur un corpus de voyelles synthétiques et un corpus de la parole naturelle. Ce dernier comprend des stimuli qui sont produits par des locuteurs normophoniques et dysphoniques et comprennent une large gamme de pathologies.

Chapitre V: Résultats expérimentaux

V. 1. Introduction

Dans ce chapitre, on présente les résultats expérimentaux. On commence par une description des corpus utilisés dans ce travail et de l'évaluation perceptive. Ensuite, la méthode d'analyse des dyspériodicités vocales proposée est appliquée aux corpus. Les indices acoustiques rapport excitation-à-dyspériodicité segmental (EDRSEG) et quotient d'ouverture moyen (MOQ) sont utilisés pour quantifier numériquement les dyspériodicités vocales. Les performances de la méthode proposée en termes de corrélation de l'indice acoustique avec l'enrouement perçu sont comparées à celles de méthodes de références. Enfin, une analyse multivariée combinant la proéminence du pic cepstral et le quotient d'ouverture moyen pour prédire les scores de l'enrouement perçu est réalisée.

V. 2. Corpus et évaluation perceptive

Les corpus utilisés dans ce travail incluent des voyelles [a] synthétiques et de la parole naturelle.

V. 2. 1. Voyelles [a] synthétiques

Les voyelles [a] synthétiques sont simulées avec un synthétiseur utilisant des modèles de l'aire glottique et du débit d'air à travers la glotte pour modéliser la source et un modèle de tubes cylindriques pour modéliser la propagation de l'onde acoustique dans la trachée et le conduit vocal [28]. La capacité de ce synthétiseur à simuler les troubles de la voix avec différents timbres vocaux a été démontrée par plusieurs expériences perceptives [98-100].

Ce premier corpus comprend 48 stimuli de sons synthétiques [a] de 1-seconde combinant trois valeurs de la fréquence fondamentale, quatre niveaux de la perturbation cycle à cycle (jitter) et quatre niveaux du bruit additif. Les fréquences fondamentales sont 100 Hz, 120 Hz et 140 Hz.

Huit thérapeutes et un phonéticien ont évalué perceptivement les voyelles synthétiques selon le grade (G) perçu, le souffle (B) et la raucité (R) de la voix perçue en quatre degrés : 0 (normal), 1 (faible), 2 (modéré) et 3 (sévère).

V. 2. 2. Parole naturelle

Le corpus comprend des voyelle [a] et deux phrases néerlandaises ("Papa en Marloes staan op het station. Zewachten op de trein") produites par 22 locuteurs normophoniques âgés de 19 à 40 ans et 229 dysphoniques âgés de 8 à 85 ans avec différents degrés de dysphonie [101]. Les diagnostics sont les suivants : nodules (42), dysphonie fonctionnelle (81), polype (11), œdème (29), parésie/paralysie (18), laryngite aigu (5), kyste (8) et autres (34).

Les segments voisés des deux phrases ont été extraits et concaténés. Trois corpus de stimuli artificiels ont été formés. Le premier sous-corpus (corpus [S1]) comprend la voyelle [a]. Le second sous-corpus (corpus [S2]) comprend des concaténations des phrases complètes et de la voyelle [a]. Le troisième sous-corpus (corpus [S3]) comprend des concaténations de segments voisés des phrases et de la voyelle [a]. Tous les stimuli ont été échantillonnés à 44100 Hz.

Cinq juges ont évalué perceptivement le sous-corpus incluant la concaténation des phrases complètes et de la voyelle [a]. 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 de la voix [102]. Chaque stimulus a été assigné comme score global la moyenne des cinq scores. L'enregistrement, la segmentation et l'évaluation perceptives ont été effectués à l'hôpital général de Sint-Jan, à Bruges, en Belgique.

V. 3. Résultats

V. 3. 1. Rapport excitation à dyspériodicité segmental

Afin de montrer l'efficacité de la méthode basée sur l'estimation de la source glottique pour la caractérisation des voix pathologiques, l'indice acoustique rapport excitation à dyspériodicité segmental (EDRSEG) a été appliqué à l'analyse des voyelles [a] synthétiques et de la parole naturelle. La performance de l'indice acoustique EDRSEG en termes de corrélation avec l'évaluation perceptives est comparée à celle de la proéminence du pic cepstral (CPP) et du rapport harmonique sur bruit (HNR). Le CPP est calculé en utilisant le logiciel exécutable de Hillenbrand's CPPs [103].

L'effet de la longueur de la trame sur la décomposition du logarithme du spectre d'amplitude du signal vocal en ces deux composantes: harmonique et enveloppe spectrale a

été étudié pour différentes longueurs de la trame en réalisant des expériences sur la voyelle [a] synthétique. Il a été constaté que la corrélation la plus forte est obtenue pour une longueur de trame de 200 ms [10-11, 93]. Par conséquent, la longueur de la trame a été fixée à cette valeur.

D'autre part, l'effet de la longueur de la trame sur la corrélation entre le rapport excitation à dysperiodicité segmental et le score du grade, de la raucité et du souffle perçus a été étudié pour différentes longueurs de la trame en réalisant des expériences sur la voyelle [a] synthétique. La longueur de la trame a été fixée à l'inverse de la fréquence fondamentale ($1/f_0$), valeur qui donne la plus forte corrélation entre l'indice acoustique EDRSEG et les scores moyens de l'évaluation perceptive [94].

V. 3. 1. 1. Corpus de voyelles synthétiques

La figure V.1 représente graphiquement le rapport excitation à dyspériodicité segmental estimé en fonction des scores globaux moyens du grade, du souffle et de la raucité perçus pour les voyelles [a] synthétiques. Comme on l'observe, les valeurs de l'indice acoustique EDRSEG a tendance de décroître linéairement lorsque les scores moyens perçus augmentent. Les valeurs du rapport excitation à dyspériodicité segmental varient de 8.6 dB à 24.6 dB [94].

Pour comparaison, les figures V.2 et V.3 montrent, respectivement, les indices acoustiques CPP et HNR en fonction des scores globaux moyens du grade, du souffle et de la raucité perçus. Comme on l'observe, lorsque le degré de sévérité de la pathologie de la voix augmente/diminue, les deux indices acoustiques CPP et HNR ont tendance de diminuer/augmenter.

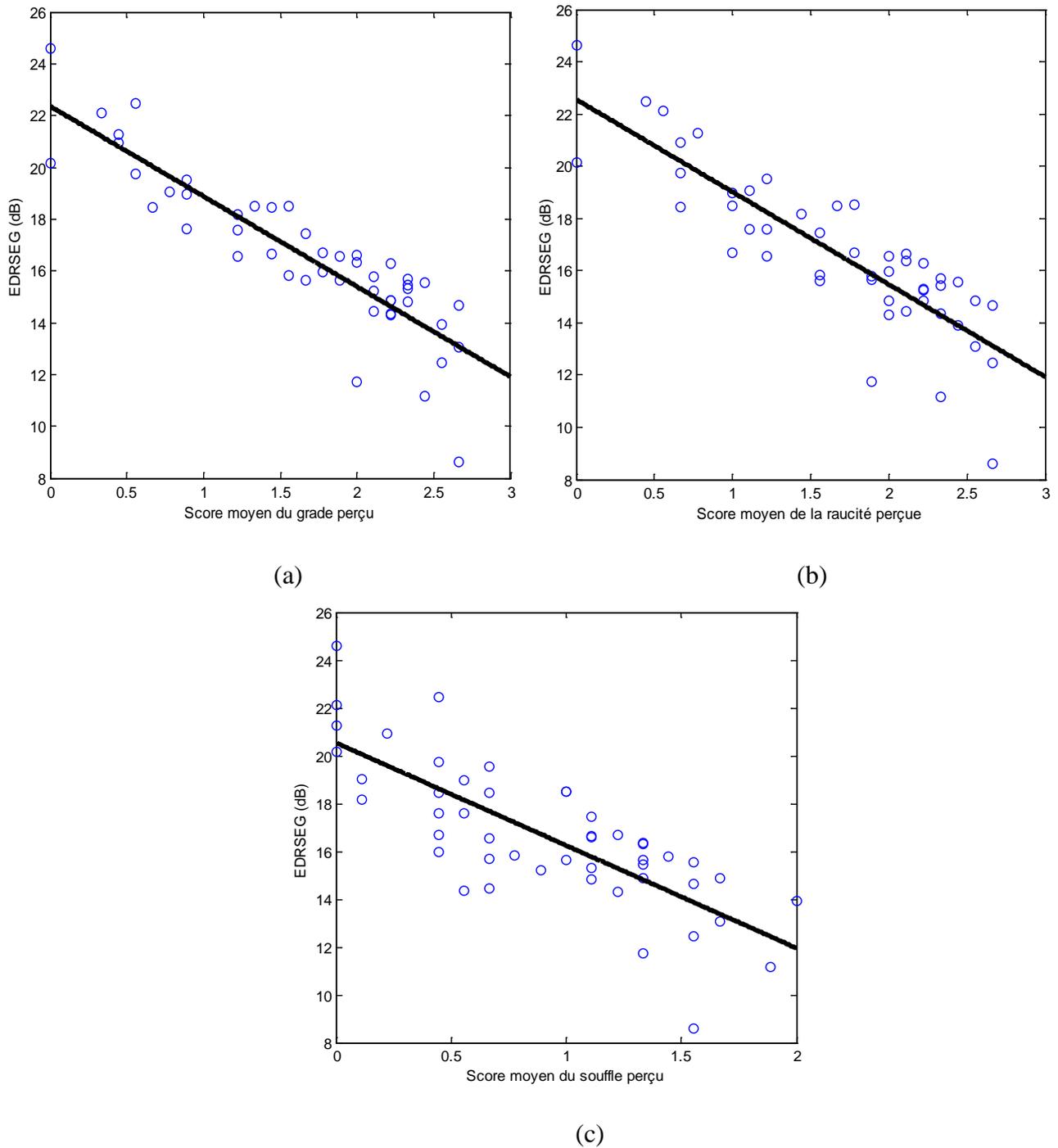


Figure V.1. Rapport excitation à dyspériodicité segmental en fonction des scores moyens (a) du grade, (b) de la raucité et (c) du souffle perçus pour les voyelles [a] synthétiques.

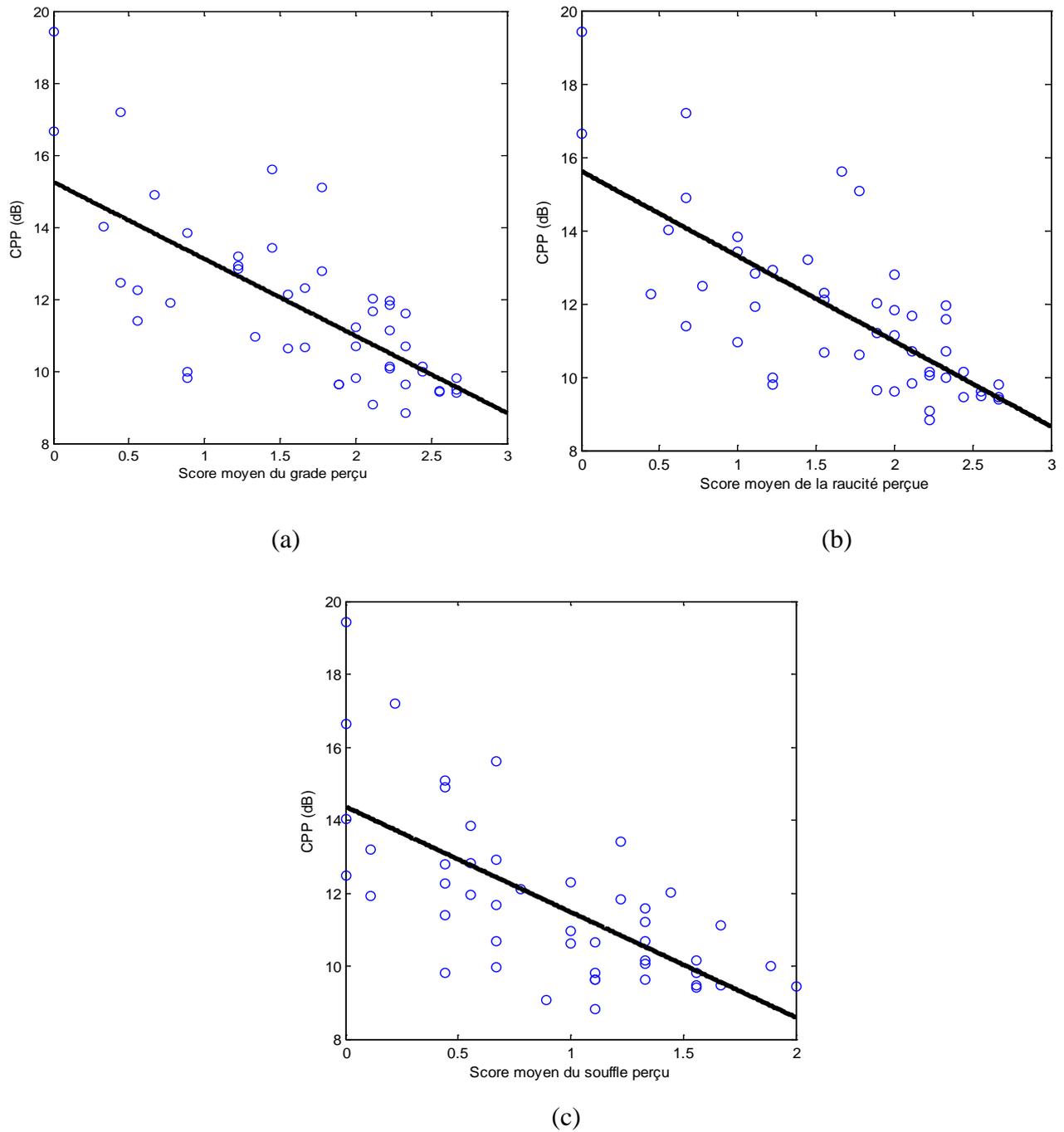


Figure V.2. Proéminence du pic cepstral en fonction des scores moyens (a) du grade, (b) de la raucité et (c) du souffle perçus pour les voyelles [a] synthétiques.

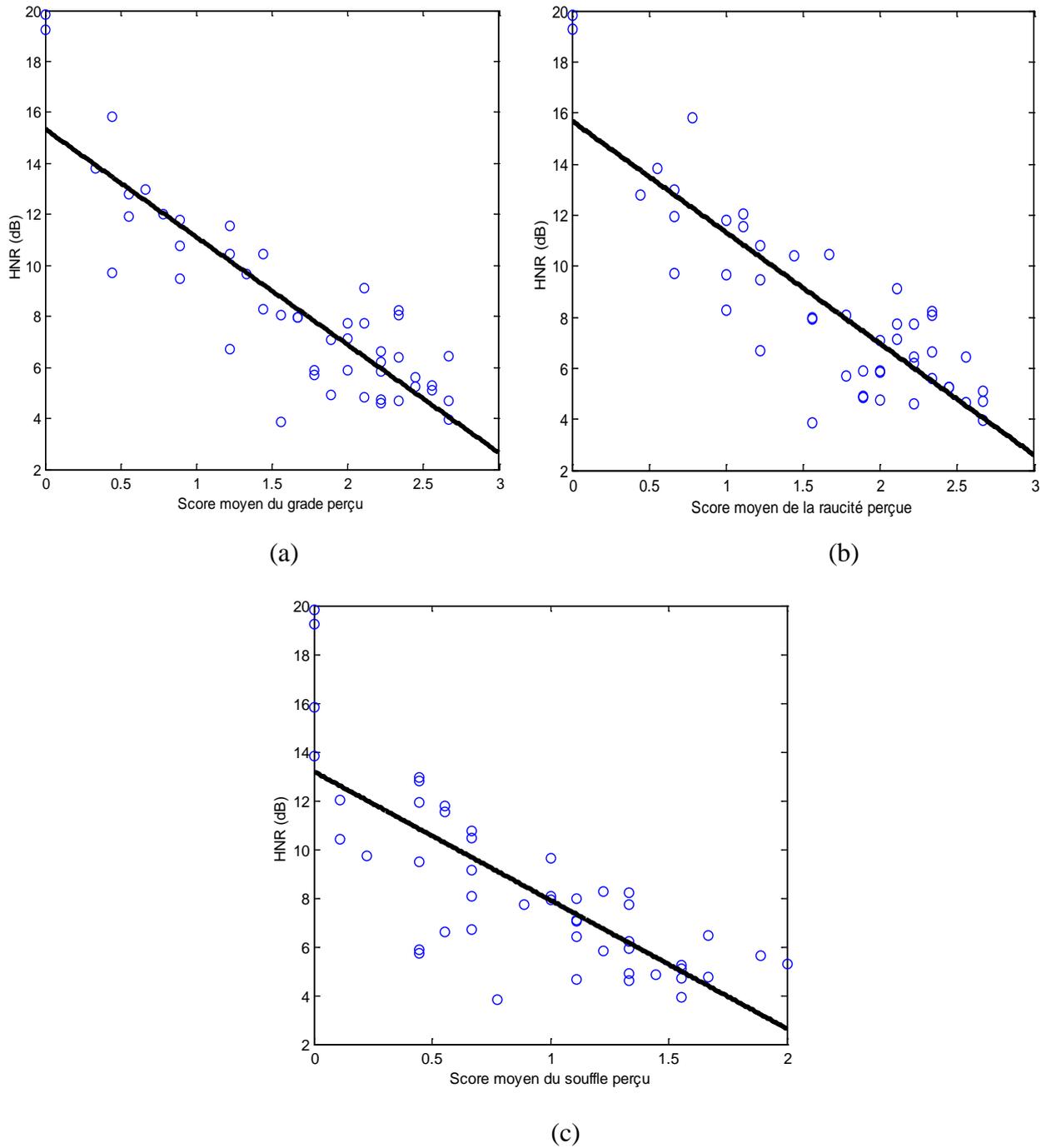


Figure V.3. Rapport harmonique sur bruit en fonction des scores moyens (a) du grade, (b) de la raucité et (c) du souffle perçus pour les voyelles [a] synthétiques.

Le tableau V.1 présente le coefficient de corrélation (ρ) des indices acoustiques EDRSEG, CPP et HNR avec les scores moyens du grade (G), du souffle (B) et de la raucité (R). Nous observons que le rapport excitation à dyspériodicité segmental et le rapport harmonique sur bruit sont fortement corrélés avec les scores moyens du grade et de la raucité et moins corrélés avec les scores moyens du souffle. Les scores moyens du grade, du souffle et de la raucité sont plus fortement corrélés avec les indices acoustiques EDRSEG et HNR estimés via la méthode de décomposition modale empirique que l'indice acoustique CPP estimés via la méthode via la méthode cepstrale.

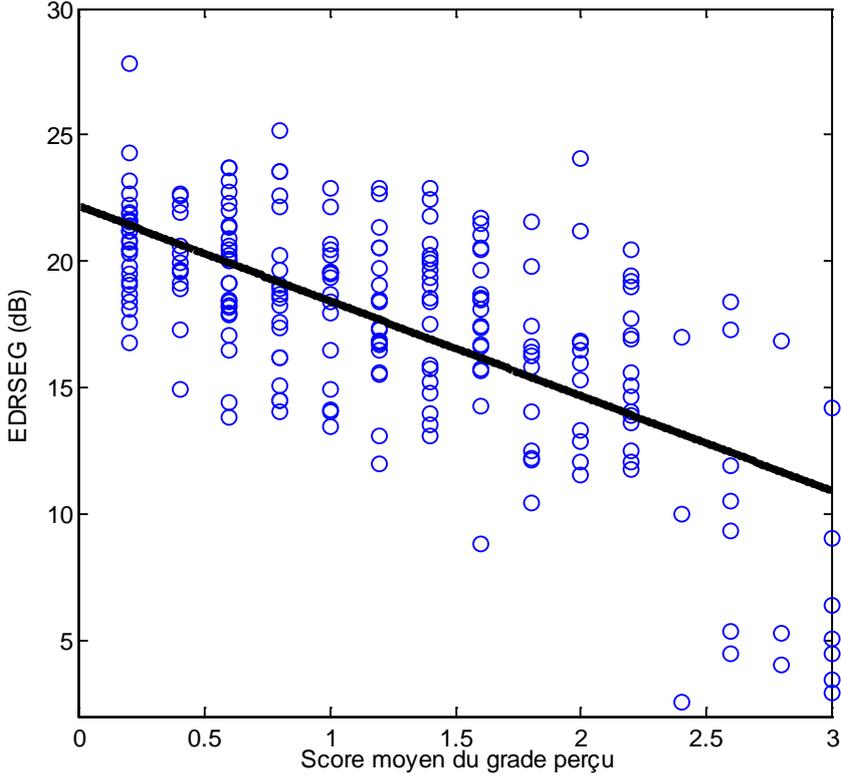
Corrélation	G	R	B
ρ_{EDRSEG}	-0.88	-0.84	-0.78
ρ_{CPP}	-0.71	-0.73	-0.68
ρ_{HNR}	-0.87	-0.85	-0.76

Tableau V.1. Corrélations des indices acoustiques EDRSEG, HNR et CPP avec les scores moyens du grade, du souffle et de la raucité perçus.

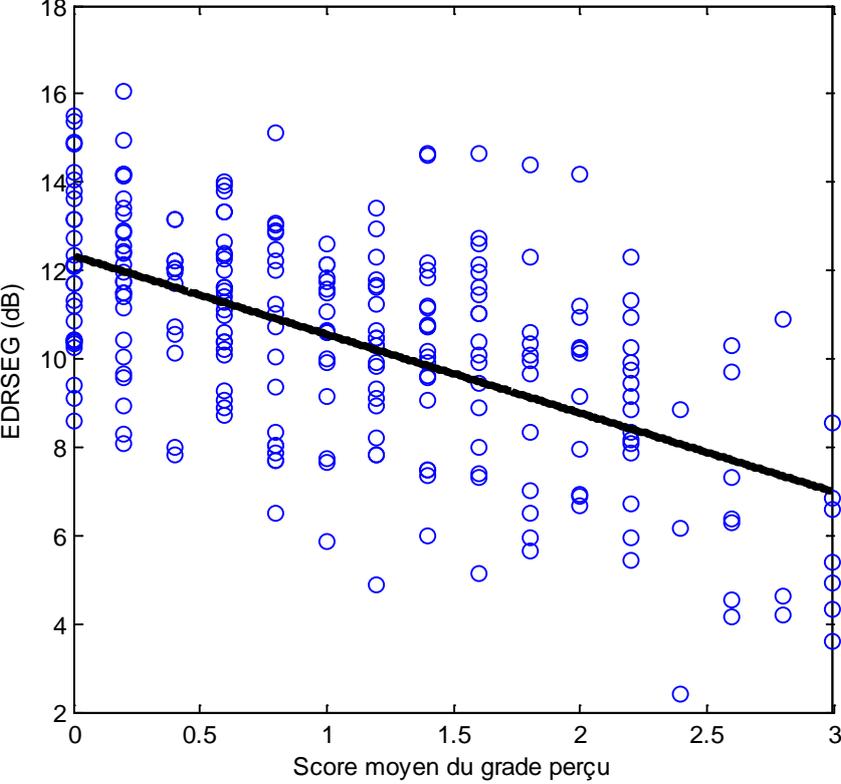
V. 3. 1. 2. Corpus de parole naturelle

La figure V.4 montre graphiquement les valeurs de l'indice acoustique EDRSEG en fonction des scores moyens du grade perçu pour les trois sous corpus de la parole naturelle. On observe que lorsque le degré d'enrouement perçu augmente/diminue, le rapport excitation à dyspériodicité segmental a tendance de diminuer/augmenter pour les trois sous corpus de la parole naturelle.

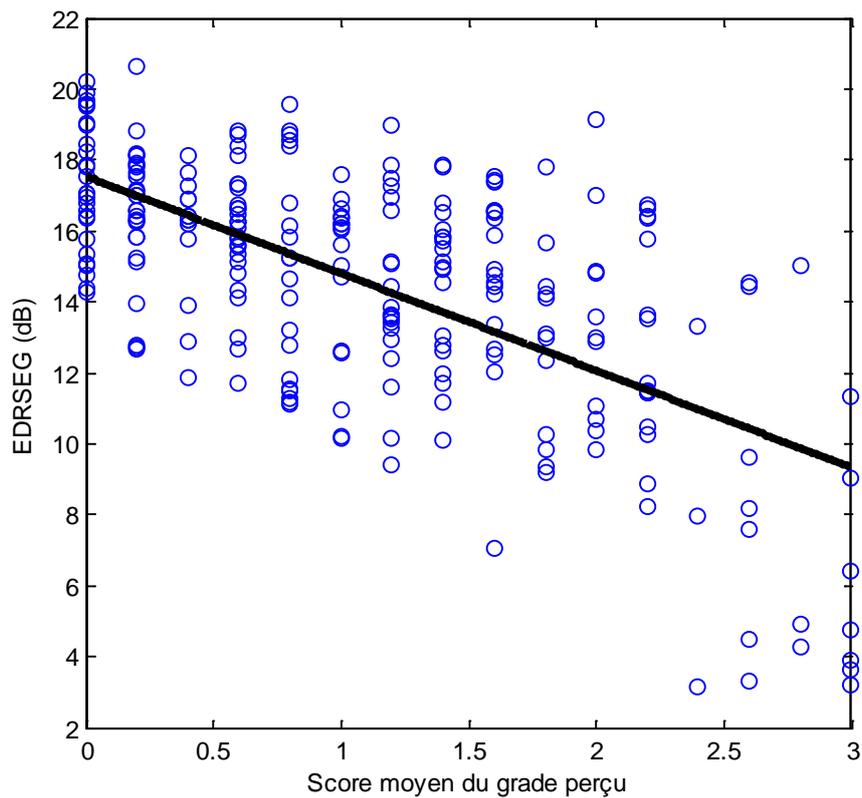
Les corrélations des indices acoustiques EDRSEG, HNR et CPP avec les scores moyens du grade perçu ont été calculées pour les trois sous corpus de la parole naturelle (Tableau V.2). On observe que le rapport excitation à dyspériodicité segmental est plus fortement corrélé avec les scores moyens du grade que les autres indices acoustiques pour le corpus [S1] comprenant les voyelles [a] plus corrélé que le CPP pour le corpus [S3] comprenant des concaténations de segments voisés des phrases et de la voyelle [a] alors qu'il présente une corrélation plus faible que le HNR et le CPP pour le corpus [S2] comprenant des concaténations des phrases complètes et de la voyelle [a].



(a)



(b)



(c)

Figure V.4. Rapport excitation à dyspériodicité segmental en fonction du score moyen du grade perçu pour les trois corpus (a) corpus [S1], (b) corpus [S2] et (c) corpus [S3].

	Corpus [S1]	Corpus [S2]	Corpus [S3]
ρ_{EDRSEG}	-0.68	-0.57	-0.65
ρ_{HNR}	-0.65	-0.71	-0.67
ρ_{CPP}	-0.59	-0.66	-0.63

Tableau V.2. Corrélation des indices acoustiques EDRSEG, HNR et CPP avec le score moyen du grade perçu pour les trois corpus.

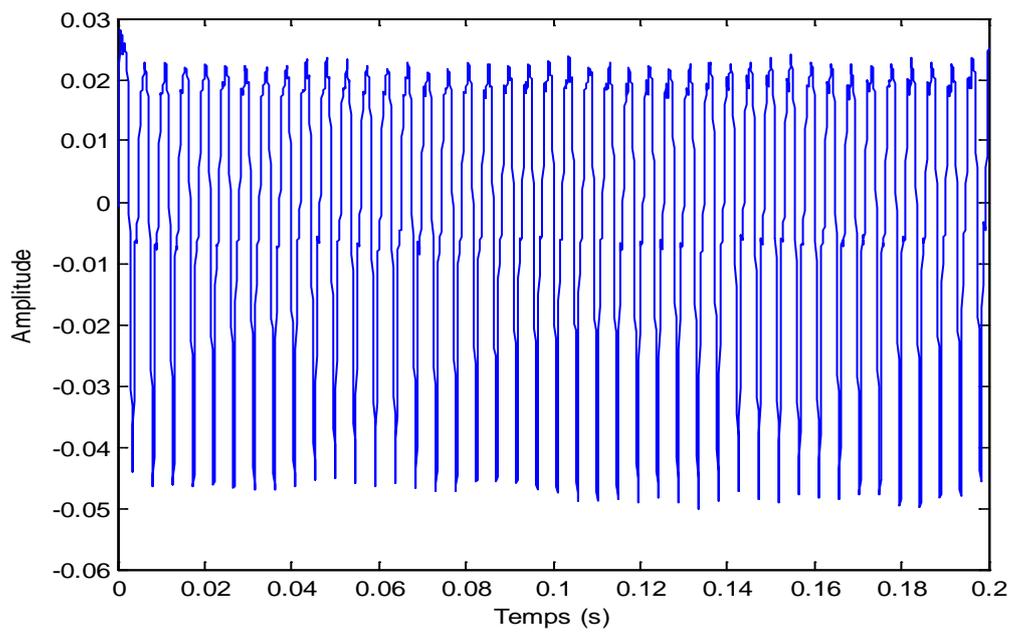
V. 3. 2. Quotient d'ouverture moyen

Les études précédentes ont montré que la meilleure longueur de la trame qui fournit une décomposition précise du signal de parole en une composante harmonique et une enveloppe spectrale est de 200 ms [10-11, 93]. La longueur de la trame a été fixée sur cette valeur.

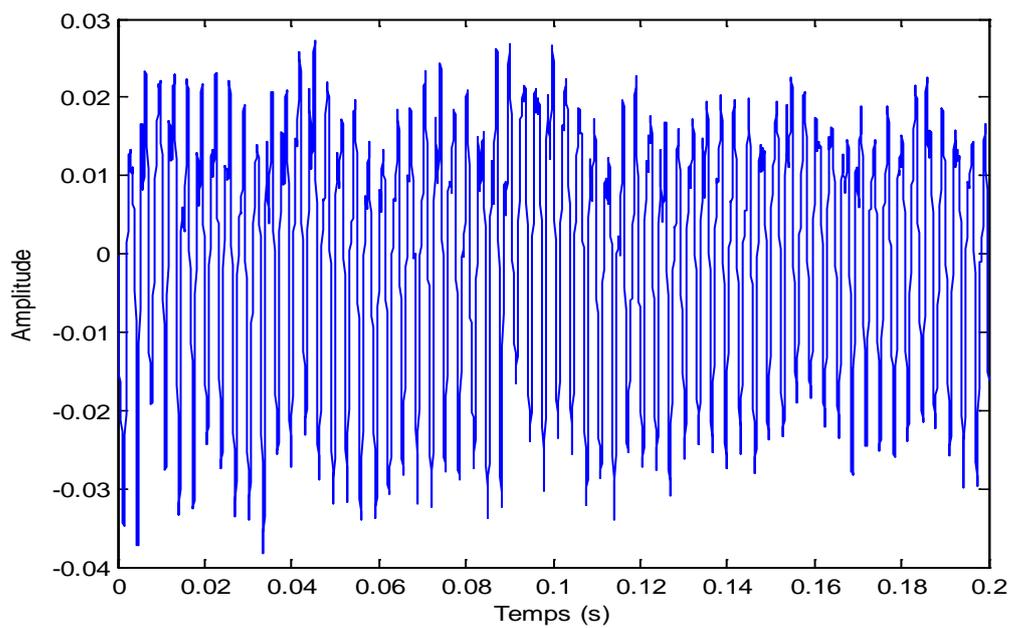
La figure V.5 montre, comme exemple, les signaux de la source glottique estimés d'une trame de 200 ms extraite d'une voyelle [a] soutenue produite par un locuteur normophonique assigné un score moyen de 0 et un locuteur dysphonique assigné un score moyen de 3. Les valeurs de la fréquence fondamentale moyenne F_0 des deux trames sont respectivement 215 Hz et 155 Hz [97]. Comme on peut l'observer, le signal de la source glottique correspondant au locuteur normophonique (figure V.5-(a)) a une forme quasi-périodique tandis que le signal de la source glottique correspondant au locuteur dysphonique a une forme irrégulière. Ces résultats sont en bon accord avec ceux publiés dans la littérature [104].

Lors de l'analyse de la dyspériodicité vocal basée sur l'optimisation des paramètres d'un modèle de la source glottique, le signal de la source glottique est modélisé par le modèle paramétrique LF dont les paramètres optimaux sont obtenus par un algorithme génétique (GA). Pendant la procédure d'optimisation, la longueur de la trame est fixée à environ trois fois la période fondamentale maximale possible. Pour une période fondamentale maximale de 12.5 ms, soit une fréquence fondamentale possible minimale de 80 Hz, ceci correspond à une longueur de trame de 37.5 ms. Par conséquent, la longueur de la trame utilisée dans la procédure d'optimisation est fixée à 40 ms. L'optimisation est consacrée à la phase d'ouverture T_p et à la phase ouverte ($T_e - T_p$). La fermeture brutale est fixée à $N_a = 10$ et 3 périodes fondamentales sont employées pour générer le modèle LF de la source.

La figure V.6 montre le signal de la source glottique estimé en utilisant la méthode basée sur la décomposition modale empirique superposé au signal de la source glottique optimisé en utilisant le modèle LF et les algorithmes génétiques pour une longueur de trame égale à trois la période fondamentale, extraite d'une voyelle [a] soutenue produite par un locuteur normophonique et un locuteur dysphonique [97]. On peut observer que l'optimisation via le modèle LF et les algorithmes génétiques fournit une estimation avec une bonne précision de la source glottique.

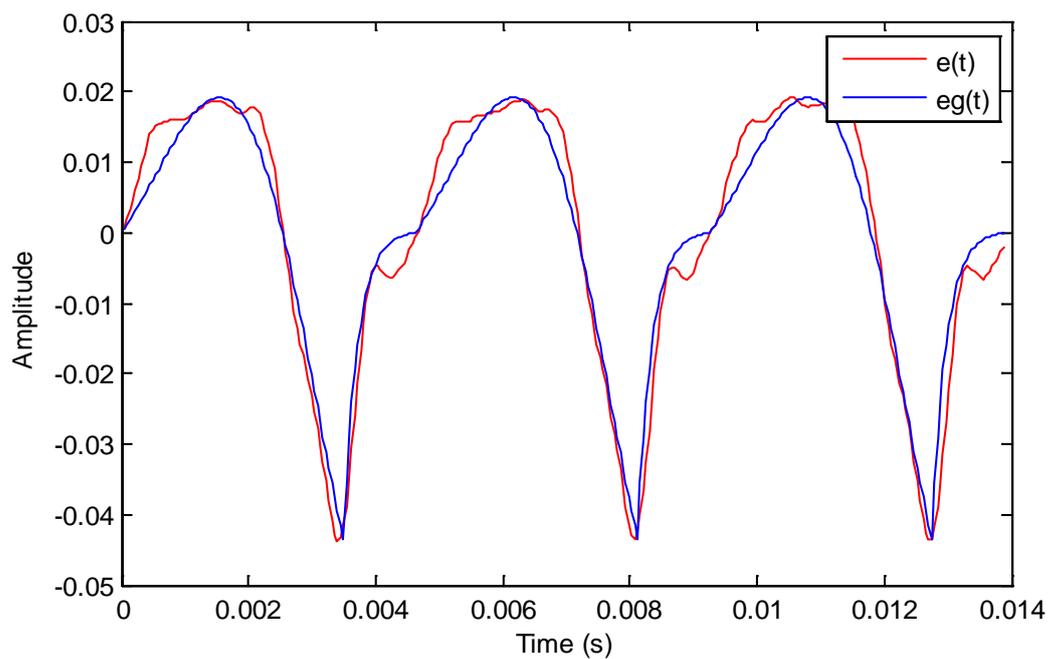


(a)

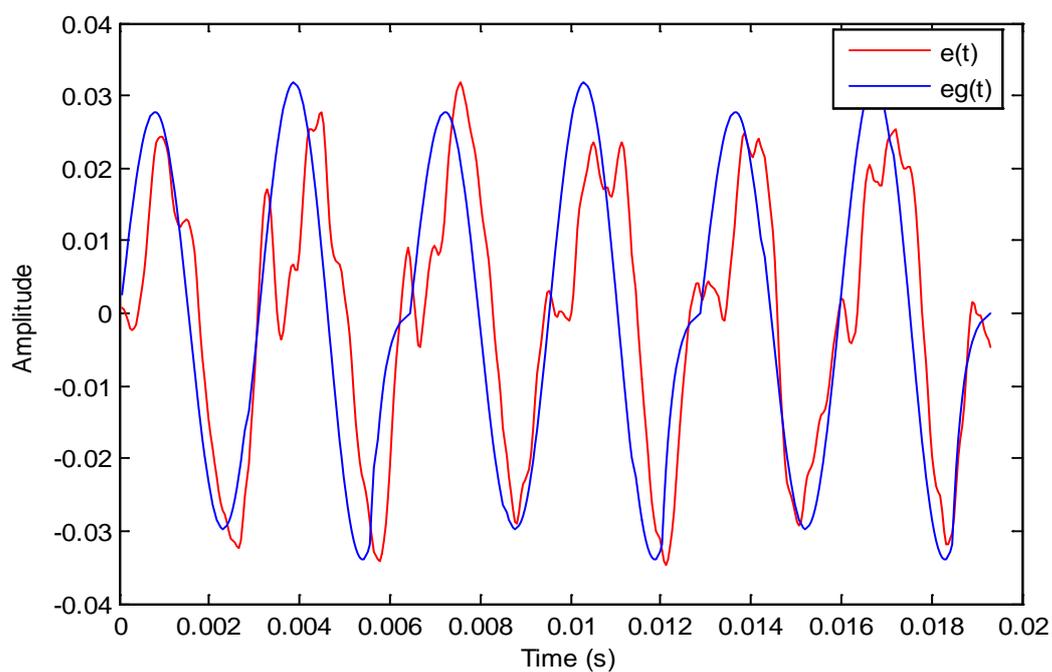


(b)

Figure V.5. Signaux estimés de la source glottique correspondant à deux trames de 200 ms extraites de voyelles soutenues [a] produites par (a) un locuteur normophonique et (b) un locuteur dysphonique.



(a)



(b)

Figure V.6. Signal de la source glottique estimé via la méthode basée sur l'EMD superposé au signal de la source glottique optimisé via le modèle LF et les GAs pour une longueur de trame égale à trois fois la période fondamentale extraite d'une voyelle [a] soutenue produite par (a) un locuteur normophonique et (b) un locuteur dysphonique.

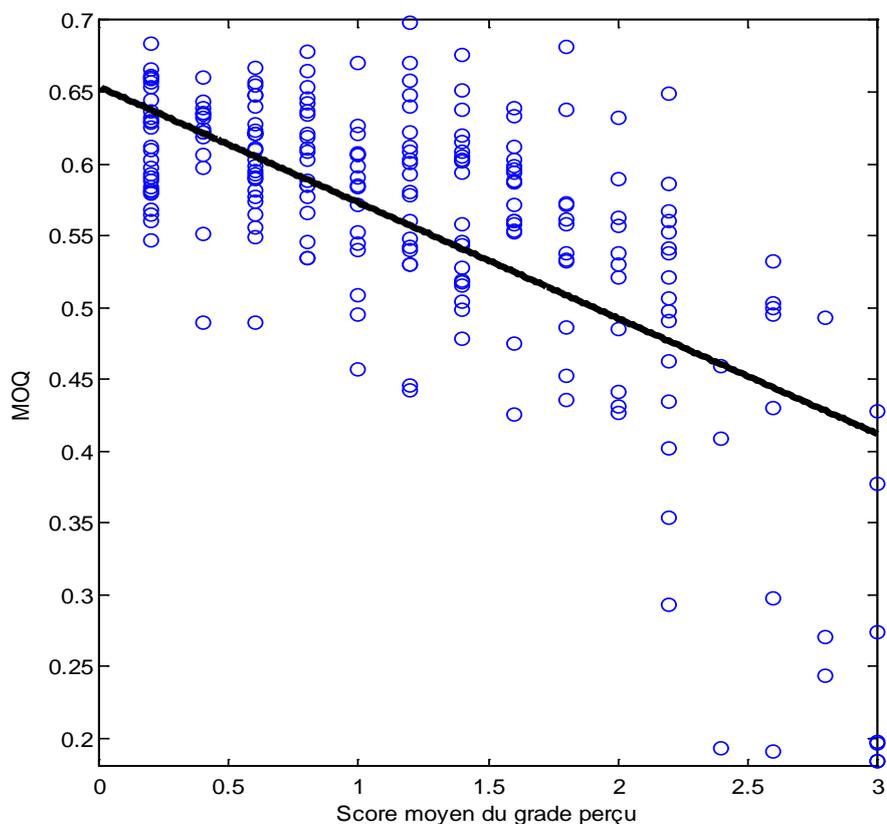
Le tableau V.3 donne l'écart type et la moyenne de la fréquence et de la période fondamentales ainsi que les valeurs des paramètres T_p et T_e , de l'indice acoustique MOQ et de la fonction d'évaluation pour les deux locuteurs normophoniques et dysphonique [97]. La dysperiodicité pour les deux locuteurs est mesurée en estimant le quotient d'ouverture moyen MOQ en combinant l'EMD et les algorithmes génétiques. Le quotient d'ouverture moyen est de 0.56 pour le locuteur normophonique et de 0.24 pour le locuteur dysphonique. Le quotient d'ouverture moyen associé à la voix pathologique est petit comparé au quotient d'ouverture moyen de la voix modale.

		Normophonique	Dysphonique
F_0 (Hz)	Moyenne	215	155
	Ecart type	0.9	6.6
T_0 (ms)	Moyenne	4.6	6.4
	Ecart type	0.02	0.39
T_p (ms)		2.57	1.57
T_e (ms)		3.49	5.59
MOQ		0.56	0.24
σ (%)		0.30	1.07

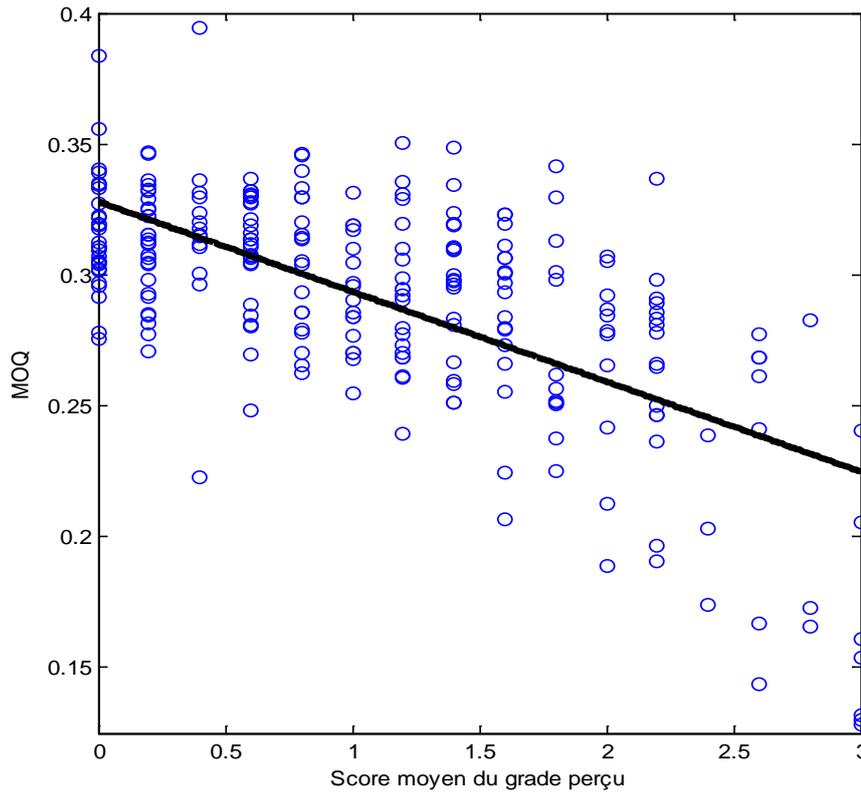
Tableau V.3. Moyenne et écart type de la fréquence et de la période fondamentales, valeurs optimisées des paramètres du modèle de la source glottique, MOQ et fonction d'évaluation pour les deux locuteurs normophonique et dysphonique.

La figure V.7 montre graphiquement le quotient d'ouverture moyen en fonction des scores moyens du grade perçu pour les trois sous-corpus de la parole naturelle. Les figures montrent que lorsque le degré de sévérité de la pathologie de la voix augmente/diminue, le quotient d'ouverture moyen a tendance de diminuer/augmenter pour les trois sous-corpus de la parole naturelle.

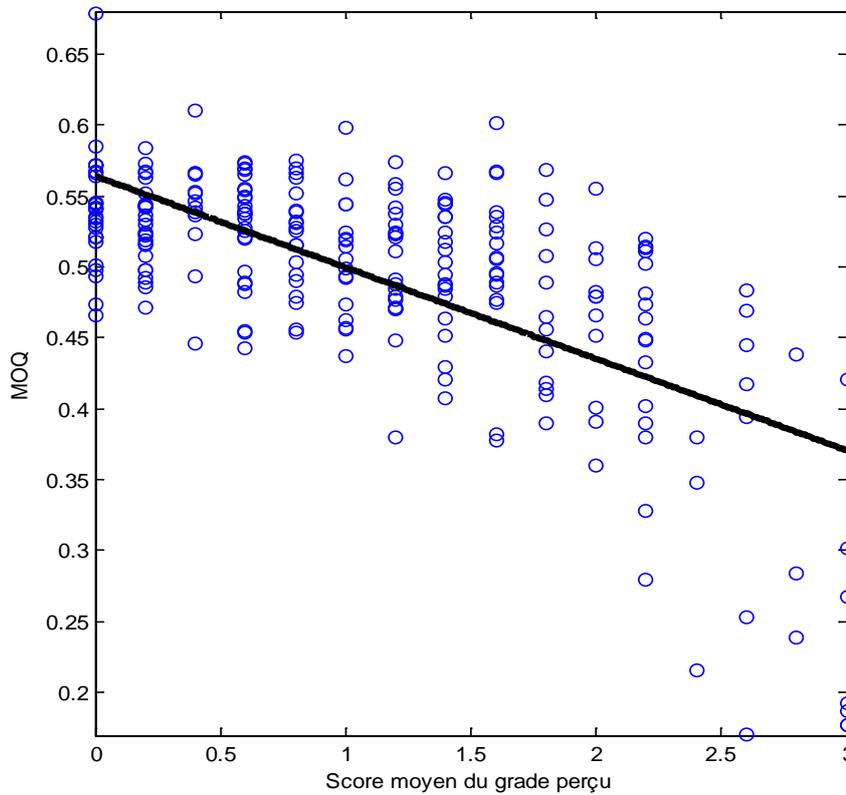
Pour comparaison, la figure V.8 montre l'indice acoustique CPP en fonction des scores globaux moyens du grade perçu. Comme on l'observe, lorsque le degré de sévérité de la pathologie de la voix augmente/diminue, l'indice acoustique CPP a tendance de diminuer/augmenter pour les trois sous-corpus de la parole naturelle.



(a)

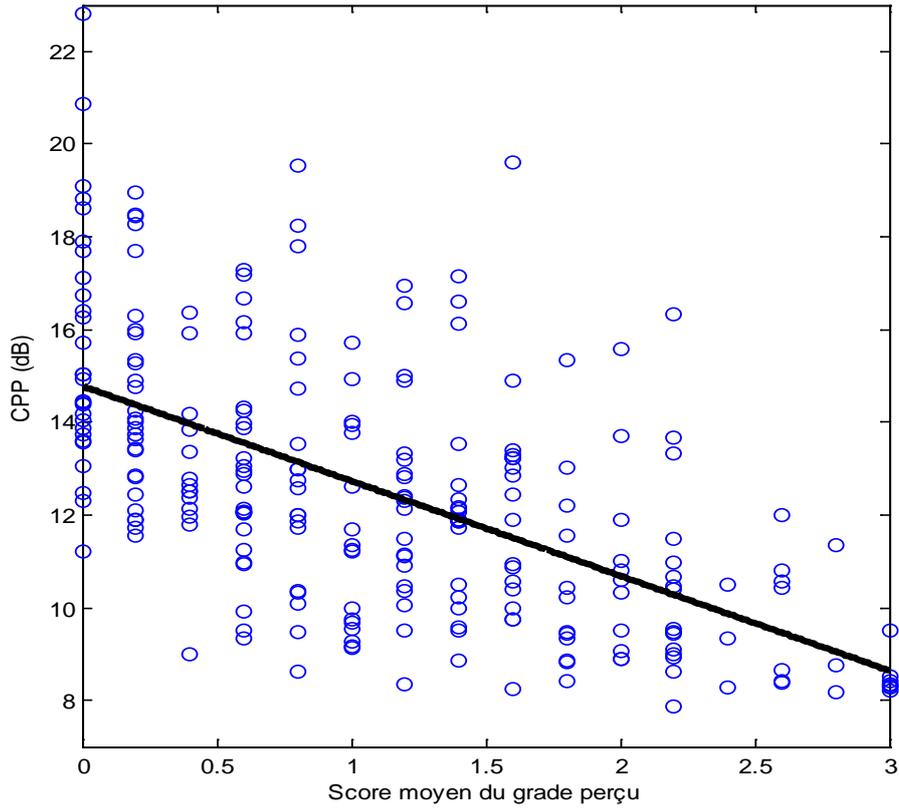


(b)

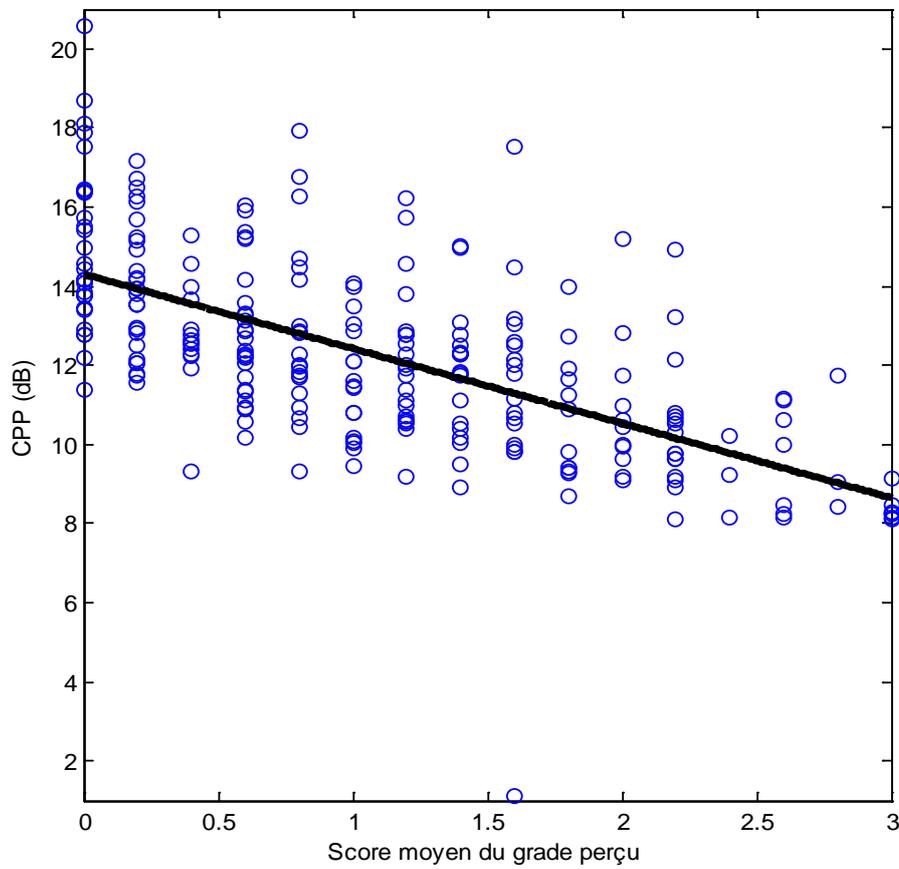


(c)

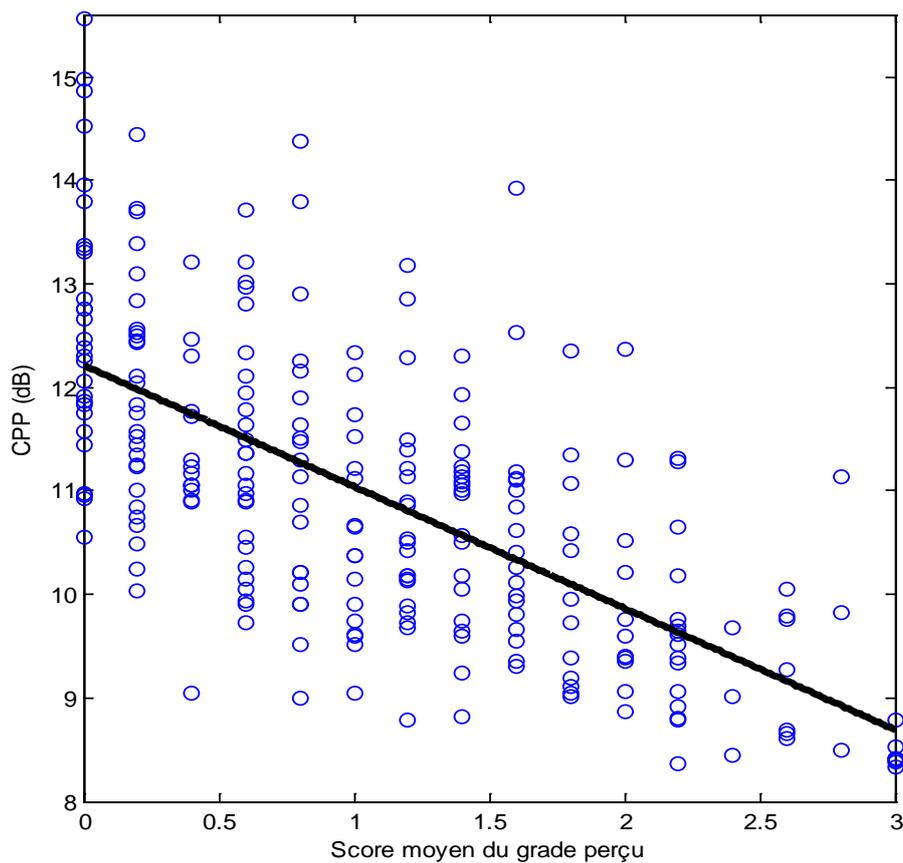
Figure V.7. Quotient d'ouverture moyen en fonction des scores moyens du grade perçu pour les trois corpus de parole naturelle (a) corpus [S1], (b) corpus [S2] et (c) corpus [S3].



(a)



(b)



(c)

Figure V.8. Proéminence du pic cepstral en fonction des scores moyens du grade perçu pour les trois corpus de parole naturelle (a) corpus [S1], (b) corpus [S2] et (c) corpus [S3].

La corrélation entre le quotient d'ouverture moyen et les scores moyens du grade perçu a été calculée pour les trois corpus de la parole naturelle. Le tableau V.4 montre le coefficient de corrélation (ρ) des indices acoustiques MOQ et CPP avec le score moyen du grade (G). Nous observons que la meilleure corrélation $\rho = -0.68$ entre l'indice acoustique MOQ et les scores moyens du grade perçu est obtenue pour le corpus [S1] comprenant les voyelles [a] pour les deux autres sous-corpus. On observe une plus forte corrélation entre l'indice acoustique MOQ estimé via la méthode EMD-GA et les scores moyens du grade perçu comparés pour les corpus [S1] et [S3] comparée à l'indice CPP alors que les deux indices acoustiques montrent des corrélations peu différentes pour le corpus [S2].

	Corpus [S1]	Corpus [S2]	Corpus [S3]
ρ_{MOQ}	-0.68	-0.65	-0.66
ρ_{CPP}	-0.59	-0.66	-0.63

Tableau V.4. Coefficient de corrélation du Quotient d'ouverture moyen et de la proéminence du pic cepstral avec les scores moyens du grade perçu pour les trois sous-corpus de parole naturelle.

V. 3. 3. Combinaison des indices acoustiques quotient d'ouverture moyen et proéminence du pic cepstral

Afin d'améliorer la corrélation entre les indices acoustiques et l'évaluation perceptive, nous avons combiné le quotient d'ouverture moyen et la proéminence du pic cepstral via l'analyse de régression linéaire multiple. La combinaison est faite pour les trois sous-corpus de voyelles naturelles. Les résultats obtenus sont montrés dans le tableau V.5. Les deux premières colonnes donnent les coefficients de corrélations des indices acoustiques MOQ et CPP avec les scores moyens du grade perçu pour les trois sous-corpus et colonnes suivantes donnent les valeurs des coefficients a_1 et a_2 de la régression linéaire multiple et la corrélation multiple R.

En se référant au tableau V.5, on observe que la combinaison des indices acoustiques MOQ et CPP permet d'augmenter la corrélation multiple à $R=0.73$ pour le corpus [S1] et [S3] et à $R= 0.75$ pour le corpus [S2].

	ρ_{CPP}	ρ_{MOQ}	a_0	a_1	a_2	R
Corpus [S1]	-0.59	-0.68	1.5559	-0.6706	-1.0158	0.73
Corpus [S2]	-0.66	-0.65	2.0145	-1.2715	-1.0268	0.75
Corpus [S3]	-0.63	-0.67	1.6110	-0.7415	-1.1519	0.73

Tableau V.5. Coefficients de corrélation des MOQ et CPP avec les scores moyens du grade perçu, coefficients de régression multiple et corrélation multiple pour les trois sous-corpus de parole naturelle (a) corpus [S1], (b) corpus [S2] et (c) corpus [S3].

V. 4. Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de l'estimation des dyspériodicités vocales. Les performances des méthodes proposées en termes de corrélation de l'indice acoustique avec l'enrouement perçu ont été comparées à celles de méthodes de références.

Les indices acoustiques utilisés pour quantifier le degré d'enrouement perçu sont la proéminence du pic cepstral, le rapport harmonique sur bruit, le rapport excitation à dyspériodicité segmental et le quotient d'ouverture moyen. Nous avons testé la performance des différentes méthodes est analysée en calculant la corrélation des scores de l'évaluation perceptive avec les valeurs des indices acoustiques.

L'analyse de corrélation entre l'évaluation perceptive et les différents indices acoustiques utilisés montre que, pour la même production, les indices acoustiques sont corrélées avec les scores moyens du grade perçu, mais les corrélations sont différentes.

Dans le but d'améliorer la performance de l'indice acoustique en terme de corrélation avec l'enrouement perçu, nous avons utilisé l'analyse multivariée, en combinant la proéminence du pic cepstral et le quotient d'ouverture moyen. La combinaison des deux indices acoustiques améliore la performance de la méthode d'estimation des dyspériodicités vocales en termes de corrélation entre l'indice acoustique et l'évaluation perceptive.

Conclusion générale

Conclusion générale

Le travail réalisé s'inscrit dans le cadre de l'analyse du signal de parole pour l'évaluation automatique des voix pathologiques. L'objectif de l'analyse du signal acoustique est d'extraire des indices pertinents permettant de déterminer les caractéristiques de la voix. La plupart des indices acoustiques sont habituellement obtenus à partir de fragments stables extraits de voyelles soutenues. Cette thèse a été consacrée au développement de nouvelles méthodes d'analyses acoustiques dédiées à l'évaluation automatique objective de la qualité de la voix produite par des locuteurs dysphoniques. Les méthodes sont basées sur la séparation des composantes du signal parole en utilisant la décomposition modale empirique.

La première méthode d'estimation des dyspériodicités vocales opère sur une estimation de l'excitation glottique. L'algorithme de la décomposition modale empirique a été utilisé pour décomposer le logarithme du spectre d'amplitude du signal acoustique en fonctions de modes intrinsèques qui ont été regroupées en deux classes: L'enveloppe spectrale et la composante harmonique par un simple seuillage. L'exploitation de l'information de phase conjointement avec la composante harmonique estimée permet d'estimer le signal d'excitation. L'indice acoustique nommé rapport excitation à dyspériodicité segmental (EDRSEG) a été proposé pour quantifier le degré de perturbation dans le signal de parole.

La deuxième méthode d'estimation des dyspériodicités vocales est basée sur l'optimisation des paramètres du modèle LF de la source glottique en utilisant les algorithmes génétiques. Cette approche est réalisée en deux étapes. La première étape consiste à estimer le signal d'excitation en utilisant la décomposition modale empirique. La seconde étape a pour objectif l'optimisation du modèle LF de la source glottique estimée via les algorithmes génétiques. L'indice acoustique nommé quotient d'ouverture moyen (MOQ) a été proposé pour quantifier le degré de perturbation dans le signal de parole.

Les différentes méthodes développées ont été appliquées à l'analyse de deux corpus pour lesquels l'évaluation perceptive a été réalisée sur l'échelle GRBAS. Le premier corpus est composé de voyelles [a] synthétiques imitant la parole naturelle pathologique. Ce corpus synthétique a été évalué perceptivement en fonction du grade (G), de la raucité (R) et du souffle (B) perçus par neuf experts. Le second corpus est constitué de la parole naturelle comprenant des voyelles [a] soutenues et de la parole connectée produites par 251 locuteurs normophoniques et dysphoniques. L'évaluation perceptive de ce deuxième corpus a été

réalisée par cinq thérapeutes professionnels de la voix selon le degré du grade perçu. Les performances des méthodes d'analyse ont été évaluées en calculant la corrélation entre les valeurs des indices acoustiques et les scores de l'évaluation perceptive.

L'analyse de la corrélation entre l'évaluation perceptive et les différents indices acoustiques a montré que, pour la même production, les valeurs des différents indices acoustiques sont fortement corrélées avec les scores moyens du grade perçu mais avec différentes valeurs.

Afin d'améliorer davantage les performances de la méthode d'analyse proposée, nous avons utilisé une analyse multivariée en combinant la proéminence du pic cepstral et le quotient d'ouverture moyen. La combinaison des deux indices acoustiques améliore la performance de la méthode d'estimation des dyspériodicités vocales en termes de corrélation entre l'indice acoustique et l'évaluation perceptive.

Comme perspectives à ce travail de recherche, les points suivants peuvent être suggérés:

- Utiliser d'autres modèles alternatifs, tels que le modèle de Klatt et le modèle de Rosenberg, pour modéliser la source glottique.
- Utiliser d'autres méthodes d'optimisation.
- Exploiter les différents indices acoustiques (CPP, HNR, EDRSEG et MOQ) dans un système de classification automatique multi-classes des voix pathologiques en utilisant les méthodes de l'intelligence artificielle.

Références bibliographiques

Références bibliographiques

- [1] F. Le Huche and A. Allali, “La voix: Pathologies vocales d'origine organique”, Tome 3, 2^e édition, Masson, Paris, 2012.
- [2] J. Hillenbrand and R. A. Houde, “Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech”, *J. Speech Lang. Hear. R.*, vol. 39, no. 2, pp. 311–321, 1996.
- [3] V. Parsa and D. G. Jamieson, “Identification of pathological voices using glottal noise measures”, *J. Speech, Language, Hearing Res.*, vol. 43, no. 2, pp. 469–485, Apr. 2000.
- [4] A. Kacha, F. Grenez and J. Schoentgen, “Estimation of dysperiodicities in disordered speech”, *Speech Comm.*, vol. 48, no. 10, pp. 1365–1378, 2006.
- [5] A. Alpan, Y. Maryn, A. Kacha, F. Grenez and J. Schoentgen, “Multi-band dysperiodicity analyses of disordered connected speech”, *Speech Comm.*, vol. 53, no. 1, pp. 131–141, 2011.
- [6] Y. Qi, R. E. Hillman and C. Milstein, “The estimation of signal-to-noise ratio in continuous speech of disordered voices”, *J. Acoust. Soc. Amer.*, vol. 105, no 4, pp. 2532–2535, 1999.
- [7] M. Vasilakis and Y. Stylianou, “Voice pathology detection based on short-term jitter estimations in running speech”, *Folia Phoniatr. Logo.*, vol. 61, no. 3, pp. 153–170, Jul. 2009.
- [8] T. Drugman, T. Dubuisson and T. Dutoit, “On the mutual information between source and filter contributions for voice pathology detection”, In: *Interspeech 2009*, pp. 1463–1466, Brighton, United Kingdom, 6-10 September 2009.
- [9] P. Gómez, R. Fernández, V. Rodellar, V. Nieto, A. Álvarez, L. M. Mazaira, R. Martínez and J. I. Godino, “Glottal source biometrical signature for voice pathology detection”, *Speech Comm.*, vol. 51, 2009, pp. 759–781.
- [10] A. Kacha, F. Grenez and J. Schoentgen, “Assessment of disordered voices using empirical mode decomposition in the log-spectral domain”, In: *Interspeech 2012*, pp. 66-69, Portland, USA, 9-13 September 2012.

- [11] A. Kacha, F. Grenez and J. Schoentgen, “Multiband vocal dysperiodicities analysis using empirical mode decomposition in the log-spectral”, *Biomed. Signal Process. Control*, vol. 17, pp. 11–20, 2015.
- [12] C. Mertens, “Analysis of vocal tremor in normophonic and dysphonic speakers,” PhD thesis, Université Libre de Bruxelles, 2015.
- [13] D. O’shaughnessy, “Speech communication human and machine,” Addison-Wesley, 1987.
- [14] J. H. Deller, J. G. Proakis and J. H. L. Hansen, “Discrete-time processing of speech signals,” Prentice-Hall, 1993.
- [15] H. Gray, “Anatomy of the human body,” Philadelphia: Lea & Febiger; Bartleby.com, 2000. www.bartleby.com/107/.
- [16] D. G. Childers, “Speech processing and synthesis toolbox,” John Wiley & Sons, 2000.
- [17] T. Dubuisson, “Glottal source estimation and automatic detection of dysphonic speakers,” PhD thesis, University of Mons, Belgium, 2012.
- [18] K. Ishizaka and J. L. Flanagan, “Synthesis of voiced sounds from a two-mass model of the vocal chords,” *Bell Syst. Tech. J.*, vol. 51, no. 6, pp. 1233–1368, July-August 1972.
- [19] S. Maeda, “A digital simulation method of the vocal-tract system,” *Speech Commun.*, vol. 1, no. 3-4, pp. 199–229, Dec. 1982.
- [20] F. Alipour, D. A. Berry and I. R. Titze, “A finite-element model of vocal-fold vibration,” *J. Acous. Soc. Am.*, vol. 108, no. 6, pp. 3003–3012, Dec. 2000.
- [21] H. Luo, R. Mittal, X. Zheng, S. A. Bielamowicz, R. J. Walsh and J. K. Hahn, “An immersed-boundary method for flow–structure interaction in biological systems with application to phonation,” *J. Comput. Phys.*, vol. 227, no. 22, pp. 9303–9332, May 2008.
- [22] T. Hézard, “Production de la voix : exploration, modèles et analyse/synthèse,” Thèse de Doctorat, Université Pierre et Marie Curie - Paris VI, 2013.

- [23] G. Fant, J. Liljencrants and Q. Lin, “A four-parameter model of glottal flow,” *STL-QPSR*, vol. 4, pp. 1–13, 1985.
- [24] D. Klatt and L. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” *J. Acous. Soc. Am.*, vol. 87, no. 2, pp. 820–857, Feb. 1990.
- [25] A. E. Rosenberg, “Effect of glottal pulse shape on the quality of natural vowels,” *J. Acous. Soc. Am.*, vol. 49, no. 2, pp. 583–590, 1971.
- [26] R. Veldhuis, “A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation,” *J. Acous. Soc. Am.*, vol. 103, no. 1, pp. 566–571, Jan. 1990.
- [27] P. Gómez-Vilda, R. Martínez, F. Díaz, C. Lázaro, A. Alvarez, V. Rodellar and V. Nieto, “Estimation of vocal cord biomechanical parameters by non linear inverse filtering of voice,” In: *Proc. of the 3rd Int. Conf. on Non-Linear Speech Processing (NOLISP’05)*, pp. 174–183, Barcelona, Spain, 19-22 April 2005.
- [28] S. Fraj, “Synthèse des voix pathologiques,” Thèse de Doctorat, Université Libre de Bruxelles, 2010.
- [29] B. Doval, C. D’Alessandro and N. Henrich, “The spectrum of glottal flow models,” *Acta Acustica United With Acust.*, vol. 92, no. 92, pp. 1026–1046, 2006.
- [30] S. Sahoo and A. Routray, “A novel method of glottal inverse filtering,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 7, pp. 1230–1241, July 2016.
- [31] J. B. Alonso, M. A. Ferrer, P. Henríquez, K. López-de-Ipina, J. Cabrera and C. M. Travieso, “A study of glottal excitation synthesizers for different voice qualities,” *Neurocomputing*, vol. 150, pp. 367–376, 2015.
- [32] A. Kacha, “Analyse de la parole continue en vue de la caractérisation des troubles de la voix,” Thèse de Doctorat, Université Libre de Bruxelles, 2006.
- [33] G. Pouchoulin, “Approche statistique pour l’analyse objective et la caractérisation de la voix dysphonique”, Thèse de doctorat, Université d’Avignon et des Pays de Vaucluse, 2008.

- [34] G. L. Holmgren, “Physical and psychological correlates of speaker recognition”, *J. Speech Hear. Res.*, vol. 10, pp. 57–66, 1967.
- [35] N. Isshiki, H. Okamura, M. Tanabe and M. Morimoto, “Differential diagnosis of hoarseness”, *Clin. Linguist. Phonet.*, vol. 21, pp. 9–19, 1969.
- [36] B. Hammarberg, B. Tritzell, J. Gauffin, J. Sundberg and L. Wedin, “Perceptual and acoustic correlates of abnormal voice qualities”, *Acta Oto-Laryngol.*, vol. 190, pp. 441–451 1980.
- [37] P. Dejonckere, C. Obbens, G. De Moor and G. Wieneke, “Perceptual evaluation of dysphonia: reliability and relevance”, *Clin. Linguist. Phonet.*, vol. 45, pp. 76–83, 1993.
- [38] M. S. De Bodt, F. L. Wuyts, P. H. Van de Heyning and C. Croux, “Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual ratings of voice quality”, *J. Voice*, vol. 11, no. 1, pp. 74–80, 1997.
- [39] E. Yumoto, W. J. Gould and T. Baer, “Harmonics-to-noise ratio as an index of the degree of hoarseness”, *J. Acous. Soc. Am.*, vol. 71, no. 6, pp. 1544–1550, 1982.
- [40] G. De Krom, “A Cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals”, *J. Speech Hear. Res.*, vol. 36, no. 2, pp. 254–266, 1993.
- [41] A. Alpan, “Objective assessment of disordered connected speech”, PhD thesis, Université Libre de Bruxelles, 2011.
- [42] A. Alpan, J. Schoentgen, Y. Maryn, F. Grenez and P. Murphy, “Cepstral analysis of vocal dysperiodicities in disordered connected speech”, In: *Interspeech 2009*, pp. 959-962, Brighton, United Kingdom, 6-10 September 2009.
- [43] Y. D. Heman-Ackah, D. D. Michael and G. S. Goding, “The relationship between cepstral peak prominence and selected parameters of dysphonia”, *J. Voice*, vol. 16, no. 1, pp. 20–27, 2002.
- [44] Y. D. Heman-Ackah, R. J. Heuer, D. D. Michael, R. Ostrowski, M. Horman, M. M. Baroody, J. Hillenbrand and R. T. Sataloff, “Cepstral peak prominence: A more

- reliable measure of dysphonia,” *Ann Otol., Rhinol., Laryngol.*, vol. 112, no. 4, pp. 324–329, Apr. 2003.
- [45] B. P. Bogert, M. J. R. Healy and J. W. Tukey, “The quefreny alanalysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking,” In: M. Rosenblatt (ed.): *Proc. Symposium Time Series Analysis*, Wiley, New York, ch. 15, pp. 209–243, 1963.
- [46] T. Drugman, P. Alku, B. Yegnanarayana and A. Alwan, “Glottal source processing: From analysis to applications,” *Comput. Speech Lang.*, vol. 28, no. 5, pp. 1117–1138, 2014.
- [47] A. del Pozo and S. Young, “Continuous tracheoesophageal speech repair,” In: *Proc. European Signal Processing Conf. (ICSP’96)*, Florence, Italy, 4–8 September, 2006.
- [48] P. Lieberman, “Some acoustic measures of the fundamental periodicity of normal and pathologic larynges”, *J. Acous. Soc. Am.*, vol. 35, no. 3, pp. 344–353, Mar. 1963.
- [49] A. Ozdas, R. G. Shiavi, S. E. Silverman, M. K. Silverman and D. M. Wilkes, “Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 9, pp. 1530–1540, Sep. 2004.
- [50] R. Sun, E. Moore and J. F. Torres, “Investigating glottal parameters for differentiating emotional categories with similar prosodics,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 4509–4512, Taipei, Taiwan, 19-24 April 2009.
- [51] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis”, *Proc. R. Soc. London Ser. A*, 454: 903-995, 1998.
- [52] G. Rilling, “Décompositions modales empiriques: Contributions à la théorie, l’algorithme et l’analyse de performances”, *Thèse de doctorat*, Université de Lyon-Ecole Normale Supérieure de Lyon, Dec. 2007.

- [53] J. Weidong, “A method for time-frequency feature extraction from vibration signal based on Hilbert-Huang Transform,” In: Proc. IEEE, Symp. Intelligent Control and Automation (WCICA 08), pp. 8460-8464, Chongqing, China, 25-27 June 2008.
- [54] B. Liu, S. Riemenschneider and Y. Xu, “Gearbox fault diagnosis using empirical mode decomposition and hilbert spectrum,” *Mech. Syst. Signal Process.*, vol. 20, pp. 718–734, 2006.
- [55] H. Li, L. Yang and D. Huang, “Application of Hilbert-Huang transform to heart rate variability analysis”, In: 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 08), IEEE International Symposium on, pp. 648-651, 16-18 May 2008.
- [56] M. S. Manikandan and K. P. Soman, “A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed. Signal Process. Control*, vol. 7, no. 2, pp. 118-128, 2012.
- [57] A. K. Tafreshi, A. M. Nasrabadi and A. H. Omidvarnia, “Epileptic seizure detection using empirical mode decomposition”, In: *Signal Proc. and Information Technology (ISSPIT 08)*, IEEE International Symposium on, pp. 238-242, 16-19 December 2008.
- [58] L. Chen, X. Li, X. Li and Z. Huang, “Signal extraction using ensemble empirical mode decomposition and sparsity in pipeline magnetic flux leakage nondestructive evaluation”, *Rev. Sci. Instrum.*, vol. 80, no. 2, pp. 025105 (1-6), Feb. 2009.
- [59] 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, Mai. 2014.
- [60] J. Cexus, “Analyse de signaux non-stationnaires par transformation de Huang, Opérateur de Teager-Kaiser, et Transformation de Huang-Teager (THT) ”, Thèse de doctorat, Université de Rennes I, 2005.
- [61] A. Thioune, “Décomposition modale empirique et décomposition spectrale intrinsèque: applications en traitement du signal et de l'image”, Thèse de doctorat, Université de Bourgogne, 2015.

- [62] T. Oberlin, “Analyse de signaux multicomposantes: contributions à la décomposition modale Empirique, aux représentations temps-fréquence et au Synchrosqueezing”, Thèse de doctorat, Université de Grenoble, 2013.
- [63] N. E. Huang and S. S. Shen, “Hilbert-huang transform and its applications”, volume 5. World Scientific Publishing Com. Inc., 2005.
- [64] N. E. Huang, M. L. C. Wu, S. R. Long, S. S. Shen, W. Qu, P. Gloersen and K. L. Fan, “A confidence limit for the empirical mode decomposition and Hilbert spectral analysis”, In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, The Royal Society, vol. 459, no. 2037, pp. 2317–2345, Sep. 2003.
- [65] H. Li, L. Yang and D. Huang, “The study of the intermittency test filtering character of Hilbert–Huang transform”, *Math. Comput. Simulat.*, vol. 70, no. 1, pp. 22–32, 2005.
- [66] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method”, *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.
- [67] N. E. Huang, J. R. Yeh and J. S. Shieh, “Complementary ensemble empirical mode decomposition a novel noise enhanced data analysis method”, *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 135–156, Apr. 2010.
- [68] J. Zhang, R. Yan, R. X. Gao and Z. Feng, “Performance enhancement of ensemble empirical mode decomposition”, *Mech. Syst. Signal Process.*, vol. 24, no.7, pp. 2104–2123, Oct. 2010.
- [69] M. E. Torres, M. A. Colominas, G. Schlotthauer and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptative noise”, In : Proc. IEEE Int. Conf. acoustics, Speech and Signal Processing (ICASSP’11), pp. 4144–4147, Prague, Czech Republic, 22-27 May 2011.
- [70] R. Lowen and A. Verschoren, “Foundations of Generic Optimization: Volume 2: Applications of Fuzzy Control, Genetic Algorithms and Neural Networks,” Springer, 2007.

- [71] D. Mokeddem, “Contrôle flou des processus biotechnologiques à base d’algorithmes génétiques”, Thèse de doctorat, Université Ferhat Abbas de Sétif, Algérie, 2014.
- [72] C. K. Chow and S. Y. Yuen, “A solution to illumination direction estimation of a shaded image: Genetic algorithm”, *Image Vision Comput.*, vol. 28, no. 12, pp. 1717–1730, 2010.
- [73] J. Wu, N. Hu, H. Setiawan, X. Huang, T. O. Raubenheimer, Y. Jiao , et al., “Multi-dimensional optimization of a terawatt seeded tapered free electron laser with a multi-objective genetic algorithm”, *Nucl. Instrum. Meth. A*, vol. 846, pp. 56-53, 2017.
- [74] D. Paul, R. Su, M. Romain, V. Sébastien, V. Pierre and G. Isabelle, “Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier”, *Comput. Med. Imag. Grap.*, Oct. 2016.
- [75] X. L. Lu, L. I. U. Hu, G. L. Wang and W. U. Zhe, “Helicopter sizing based on genetic algorithm optimized neural network”, *Chinese Journal of Aeronautics*, vol. 19, no. 3, pp. 212–218, Aug. 2006.
- [76] C. West, A. Montazeri, S. D. Monk and C. J. Taylor, “A genetic algorithm approach for parameter optimization of a 7-DOF robotic manipulator”, *International Federation of Automatic Control (IFAC) Papers OnLine*, vol. 49, no. 12, pp. 1261–1266, 2016.
- [77] C. A. C. António, J. B. Monteiro and C. F. Afonso, “Optimal topology of urban buildings for maximization of annual solar irradiation availability using a genetic algorithm”, *Appl. Therm. Eng.*, vol. 73, no. 1, pp. 424–437, 2014.
- [78] L. Lu, J. Yan and C. W. de Silva, “Feature selection for ECG signal processing using improved genetic algorithm and empirical mode decomposition”, *Measurement*, vol. 94, pp. 372–381, 2016.
- [79] R. L. Haupt and S. E. Haupt, “Practical genetic algorithms”, John Wiley & Sons, 2004.
- [80] M. Mitchell, “An introduction to genetic algorithms”, Massachusetts Institute of Technology (MIT) press, Cambridge, 1998.

- [81] K. Bendine, “Contribution à l'étude du contrôle actif des structures composites par incorporation de matériaux intelligents”, Thèse de doctorat, Université El Djillali Liabès de Sidi Bel Abbès, Algérie, 2016.
- [82] M. Melanie, “An introduction to genetic algorithms”, Massachusetts Institute of Technology (MIT) press, Cambridge, 1999.
- [83] Z.-Y. Li and M.-S. Zhang, “Pin assignment optimization for large-scale high-pin-count BGA packages using genetic algorithm,” *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 5, no. 2, pp. 232–244, Feb. 2015.
- [84] N. Durand, “Algorithmes génétiques et autres méthodes d'optimisation appliqués à la gestion de trafic aérien”, Thèse de doctorat, Institut National Polytechnique de Toulouse (INPT), 2004.
- [85] M. Amir, “Amélioration des performances des antennes imprimées basée sur des méthodes de modélisation et d'optimisation sophistiquées”, Thèse de doctorat, Université El Hadj Lakhdar de Batna, Algérie, 2014.
- [86] D. A. Coley, “An introduction to genetic algorithms for scientists and engineers”, World scientific, 1999.
- [87] E. Cantu-Paz, “Efficient and accurate parallel genetic algorithms”, Springer Science & Business Media, New York, 2001.
- [88] A. Soukou, “Modélisation et commande des systèmes industriels complexes par les techniques Intelligentes”, Thèse de doctorat, Université Ferhat Abbas de Sétif, Algérie, 2008.
- [89] D. L. Carroll, “Genetic algorithms and optimizing chemical oxygen-iodine lasers”, *Devel. Theoretical Appl. Mechanics*, vol. 18, no. 3, pp. 411-424, 1996.
- [90] H. Chiroma, S. Abdulkareem, A. Abubakar, A. Zeki, A. Y. Gital and M. J. Usman, “Correlation study of genetic algorithm operators: crossover and mutation probabilities,” In: *International Symposium on Mathematical Sciences and Computing Research 2013 (iSMSC 2013)*, pp. 39–43, Perak, Malaysia, 6–7 December 2013.

- [91] V. P. Patil and D. D. Pawar, “The optimal crossover or mutation rates in genetic algorithm: a review,” *Int. J. App. Eng. Tech.*, vol. 5, no. 3, pp. 38–41, Jul.-Sep. 2015.
- [92] P. Flandrin, G. Rilling and P. Conçalvès, “Empirical mode decomposition as a filter bank,” *IEEE Signal Proc. Letters*, vol. 11, no. 2, pp. 112–114, 2004.
- [93] M. Boudjerda and A. Kacha, “Analyse par décomposition modale empirique de stimuli synthétiques de parole pathologique”, In : Proc. 2nd International conference on biomedical engineering (BIOMEIC’14), Tlemcen, Algeria, 15–16 October 2014.
- [94] M. Boudjerda and A. Kacha, “Estimation of vocal dysperiodicities in synthetic disordered speech by empirical mode decomposition”, In: 9^{ème} journée sur le Génée Electric (CGE’09), Bordj El Bahri, Alger, Algerie, 14–15 Avril 2015.
- [95] J. Haslett, “On the sample variogram and sample autocovariance for non-stationary time series”, *J. Roy. Stat. Soc. D-Sta.*, vol. 46, no. 4, pp. 475–485, 1997.
- [96] A. Kacha, F. Grenez and J. Schoentgen, “Generalized variogram analysis of vocal dysperiodicities in connected speech”, In: Proc. of the Fourth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2005), pp. 155-158, Florence, Italy, 29–31 October 2005.
- [97] M. Boudjerda and A. Kacha, “Assessment of disordered voices based on an optimized glottal source model”, *Turk. J. Elec. Eng. & Comp. Sci.*, vol. 25, no. 4, pp. 3201–3214, 2017.
- [98] S. Fraj, F. Grenez and J. Schoentgen, “Evaluation of a synthesizer of disordered voices,” In: 3rd Advanced Voice Function Assessment International Workshop, pp. 69–72, Madrid, Spain, 18–19 May 2009.
- [99] S. Fraj, F. Grenez and J. Schoentgen, “Synthetic hoarse voices: a perceptual evaluation,” In Proc. Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), pp. 95–98, Florence, Italy, 29–31 October 2009.

- [100] S. Fraj, J. Schoentgen and F. Grenez, “Development and perceptual assessment of a synthesizer of disordered voices,” *J. Acous. Soc. Am.*, vol. 132, no. 4, pp. 2603–2615, 2012.
- [101] 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.
- [102] M. Hirano, “The laryngeal examination,” in *Speech Evaluation in Medicine*, J. Darby, Ed. New York: Grune and Stratton, 1981.
- [103] <http://homepages.wmich.edu/~hillenbr/cpps.exe>.
- [104] T. Drugman, T. Dubuisson and T. Dutoit, “Phase-based information for voice pathology detection”, In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4612–4615, Prague, Czech Republic, 22–27 May 2011.