Analyse non linéaire de la parole pour la détection des voix pathologiques

Safaa Mrad

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Tunis El Manar

المدرسة الوطنية للمهندسين بتونس

Ecole Nationale d'Ingénieurs de Tunis

Département Technologies de l'Information et de la Communication

Projet de Fin d'Etudes

présenté par

Safa MRAD

pour obtenir

le Diplôme National d'Ingénieur en Télécommunications

Analyse non linéaire de la parole pour la détection des voix pathologiques

réalisé au sein de

L'équipe Géostat

Institut National de Recherche en Informatique et en Automatique

Centre Bordeaux Sud-ouest

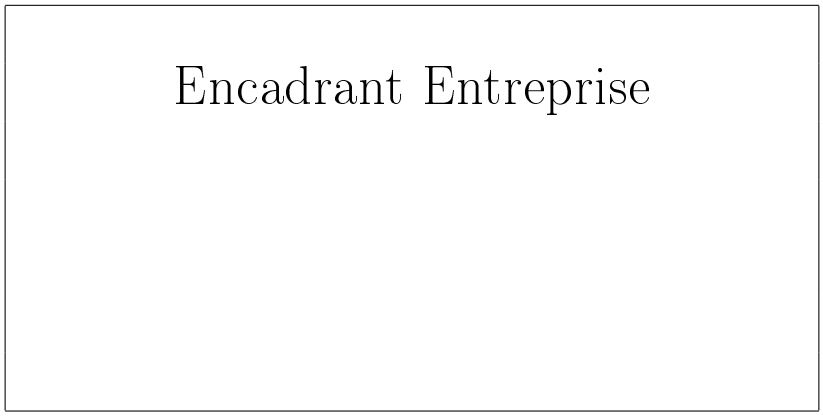
(période : 26/03/2013 - 25/09/2013)

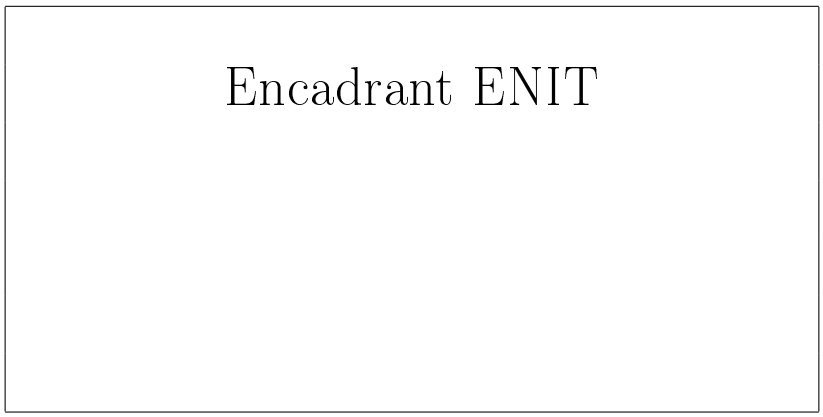
Encadrant INRIA BSO : Khalid Daoudi, Chargé de recherches, INRIA Bordeaux

Vis-à-vis ENIT : Monia Turki, Maître assistante, ENIT-Tunisie

Année universitaire : 2012-2013

Signatures





Dédicaces

Merci Allah de m'avoir donné la capacité d'écrire et de réfléchir, la force d'y croire et la patience d'aller jusqu'au bout.

Je dédie ce mémoire de projet de fin d'étude à celle qui m'a donné la vie, symbole de tendresse qui s'est sacrifiée pour mon bonheur et ma réussite, à ma mère Monia.

A mon père Fredj, école de mon enfance, ombre de mes années d'études et celui qui a veillé tout au long de ma vie à m'encourager, à me donner l'aide et à me protéger.

Puisse Dieu vous procurer santé et longue vie.

A mon seul et unique frère Samir, le rayon de soleil qui égaye ma vie avec son interminable sens d'humour.

A mes amis, mes cousins et cousines.

A tous mes professeurs qui m'ont soutenu durant mes années d'études.

A tous ceux qui m'aiment.

Je dédie ce travail.

Remerciements

Avant tout développement sur cette expérience professionnelle, il apparaît opportun de commencer ce présent mémoire par des remerciements à ceux qui m'ont beaucoup appris au cours de ce stage, et même à ceux qui ont eu la gentillesse de faire de ce stage un moment profitable.

Merci à Monsieur Khalid Daoudi, mon maître de stage de m'avoir ouvert les portes de l'équipe. Je le remercie également pour ses conseils et sa contribution à alimenter ma réflexion.

Je remercie aussi Monsieur Hussein Yahia, chef de l'équipe Géostat pour son accueil.

Je désir remercier mon encadrante à l'ENIT, Madame Monia Turki pour sa disponibilité.

J'aimerais enfin adresser un remerciement particulier à mes collègues Hicham et Suman qui m'ont apporté leurs supports moral et intellectuel tout au long de ces 6 mois.

Résumé

Le présent travail s'inscrit dans le cadre de mon projet de fin d'études en vu de l'obtention du diplôme national d'ingénieur en télécommunications.

Ce projet porte sur l'analyse et le traitement non-linéaire du signal pour mettre en place des descripteurs de voix permettant de séparer les voix pathologiques des voix normales.
Afin de contribuer à établir des mesures objectives permettant de refléter au mieux la qualité de la voix, on se propose en premier lieu d'utiliser l'algorithme de détection des instants de fermeture de la glotte (IFG) développé dans l'équipe Géostat. Tous les travaux dans les étapes suivantes reposeront sur cet algorithme dont les performances ont dépassé celles des algorithmes de l'état de l'art pour la détection des IFG.
Ensuite, on donne une nouvelle définition du pitch ainsi qu'une nouvelle approche de son calcul. Le dernier stade consiste à exploiter cette nouvelle définition du pitch dans le calcul de certains descripteurs audio standards : Jitter, Shimmer et HNR. On propose aussi de nouveaux descripteurs audio. Enfin pour évaluer la pertinence des descripteurs, des exercices de classification ont été effectués.
L'exploitation de la nouvelle définition du pitch a donné des résultats avantageux en terme de classification. Ceci est prouvé en les comparant aux résultats obtenus par l'outil de référence Praat.

Mots-clés : Voix pathologiques, pitch, mesures objectives de la qualité de la voix, jitter, shimmer, classification, Instants de Fermeture de la Glotte (IFG).

Abstract

This report is presented in partial fulfilment of the requirements for the National Diploma in Telecommunication Engineering. The work presented in this report is my graduation research project carried out in INRIA Bordeaux, France in the form of an internship (based on a merit scholarship from Campus France).

The research focuses on the application of non-linear signal processing and analysis techniques for implementing voice descriptors that can distinguish between healthy and pathological voices.
To help establish objective measures that reflect the voice quality in a better way than the existing measures, we first suggest to use an algorithm for detecting the glottal closure instants (GCI) which was developed by Geostat. Further work will then be based on this algorithm which performances exceeded those of the state of the art algorithms for detecting GCI.
On the second step, we give a new definition to the pitch and a new approach for its calculation. The last stage is about exploiting the new measure of the pitch to calculate some of the standard audio descriptors : Jitter, Shimmer and HNR. We also suggest new voice descriptors. Finally, to assess the relevance of these descriptors, classification trials were conducted.
The new definition of the pitch gave favorable results in terms of classification. This was proved by comparison to the results obtained by the reference tool Praat.

Keywords : Pathological voices, pitch, objective measures of voice quality, jitter, shimmer, classification, Glottal Closure Instants (GCI).

Table des matières

Liste des figures

Liste des tableaux

Liste des acronymes

DFA Detrended Fluctuation Analysis

DIANA Dispositif Informatisé d'ANalyse Acoustique

EGG ElectroGlottoGraphie

ELS European Laryngological Society

HNR Hormonic to Noise Ratio

IFG Instant de Fermeture Glottale

Oq quotient Ouvert

MP Marque de Pitch

PIO Pression Intra Orale

PPE Pitch Period Entropy

PSG Pression Sous Glottique

RPDE Recurrence Probability Density Entropy

SNR Signal to Noise Ratio

VHI Voice Handiap Index

VN Voix Normale

VP Voix Pathologique

Introduction Générale

La communication prend une place de plus en plus importante dans notre société. L'art de transmettre un message ou une information n'est plus aussi simple que l'on croit car aujourd'hui, la communication ne fait plus intervenir l'homme uniquement, mais on parle aussi de communication et interaction homme-machine.
Malgré cette évolution mettant en œuvre l'expression corporelle, les images et le texte, la parole reste toujours le principal support de communication. La parole est universelle et l'apprentissage de la langue parlée est plus rapide que tout autre moyen. C'est pour ceci que depuis plusieurs années différentes disciplines d'études sont apparues dans le cadre des sciences de la parole. Il y a alors la linguistique qui porte sur l'étude des langages, la phonétique qui étudie la production et la transmission des sons employés dans la communication verbale, la phonologie qui est axée sur la formation des mots et des phrases etc. Les sciences de la parole se sont développées jusqu'à avoir attiré l'attention des spécialistes en médecine, ils ont ainsi traité les troubles de la voix qui apparaissent suite à des maladies, résultats d'interventions médicales ou accidents.

La voix et la parole ont aussi suscité l'attention des spécialistes dans l'ingénierie des télécommunications. Le développement des technologies de transmission, stockage, analyse et synthèse de la voix est à l'origine de tous les progrès médicaux portant sur la parole. Derrière toutes ces technologies, il existe plusieurs théories et modèles mathématiques dont l'usage est courant dans l'ingénierie du traitement numérique du signal. On fait aussi appel aux techniques d'analyses statistiques des séries temporelles ainsi qu'à la théorie de l'information.
Ainsi, les mathématiques ont mis en œuvre récemment de nouvelles approches qui permettent de faire évoluer le domaine médical en terme de diagnostic des pathologies de la

voix. On cite l'exemple des études de recherches menées par Max Little [1] qui ont mis en évidence les limites de l'application du traitement linéaire du signal à la parole en le substituant par de nouvelles approches non linéaires ayant permis jusqu'à présent de donner de bons résultats en terme d'identification des voix pathologiques en les distinguant des voix normales [1].
En effet, plusieurs études faisant appel aux différents outils et approches mathématiques cités ci-dessus ont pour objectif commun de développer des mesures acoustiques, ce qu'on peut aussi appler descripteurs audio permettant de décrire au mieux les voix pathologiques et optimisant ainsi la discrimination entre ces dernières et les voix normales. Ces descripteurs présenteront après pour les spécialistes en médecine un outil d'aide au diagnostic.

C'est dans ce cadre que se situe ce projet. On vise à renforcer l'apport des technologies par rapport au domaine médical s'intéressant aux voix pathologiques. Ceci sera réalisé en travaillant sur les modèles mathématiques et les méthodes de traitement du signal de parole afin de mettre en place de nouvelles mesures acoustiques. Ces dernières auront pour objectif de reflèter une réalité physiologique de la production de la voix et qui pourront atteindre de bonnes performances en terme de classification en classe de voix normale (VN) et classe de voix pathologique (VP). Ces mesures acoustiques présentent des outils d'aide au diagnostic médical des pathologies de la voix.
Pour ce faire, notre travail s'articule sur 5 chapitres. Le chapitre 1 donne une présentation de l'institut d'accueil. Le chapitre 2 est dédié à l'exposition de la problématique, l'objectif ainsi que la méthodologie. Ensuite, le chapitre 3 se rapporte à une étude de l'état de l'art des travaux en traitement de parole pour la détection des voix pathologiques. On a consacré le chapitre 4 à la présentation des outils employés pour la réalisation de ce projet. On termine enfin par la partie réalisation dans laquelle on met l'accent sur les descripteurs utilisés et développés et par la suite on analyse leurs résultats en terme de pouvoir discriminant.

1. http ://www.maxlittle.net/home/index.php

Chapitre 1

L'entreprise d'accueil : INRIA Bordeaux Sud-ouest

1.1 Introduction

Dans ce premier chapitre, on donne un aperçu global sur l'Institut National de Recherche en Informatique et en Automatique -INRIA-, ensuite on présente plus particulièrement le centre d'accueil qui est le centre de Bordeaux Sud-ouest et l'équipe d'accueil.

1.2 Présentation d'INRIA

L'Institut National de Recherche en Infomatique et en Automatique (INRIA) est un établissement public à caractère scientifique et technologique qui est sous la tutelle du ministère de recherche et du ministère de l'économie, des finances et de l'industrie.
Il a été créé en 1967 à Roquencourt visant à mener des recherches fondamentales et appliquées dans les domaines des sciences et technologies de l'information et de la communication (STIC).

Les équipes-projets INRIA dont le nombre a atteint 179 rassemblant 1 800 chercheurs de l'institut et environ de 1 600 universitaires et chercheurs d'autres organismes. Ensemble, elles inventent les technologies numériques de demain en partenariats étroits avec les acteurs de la recherche publique et privée en France et à l'étranger. Elles ont publié près

de 5 000 articles en 2011 et sont à l'origine de la création de 110 start-ups.
Le budget primitif d'Inria s'élève en 2013 à 233 millions d'euros dont près de 27 % de ressources propres.

Aujourd'hui, plus de la moitié des équipes-projets INRIA sont impliquées dans les programmes cadres de recherche et de développement européens (PCRDT). Dans le cadre du 7e PCRDT, l'INRIA a contribué à identifier 2 défis scientifiques majeurs : l'Internet du futur et le patient numérique.

INRIA dispose de 8 centres de recherches qui sont implantés un peu partout en France : Recquencourt, Rennes, Sophia Antipolos, Grenoble, Nancy, Lille, Salay et Bordeaux où ce sont déroulés 6 mois de satge. Dans le paragraphe suivant on présente ce centre.

1.3 Le centre d'accueil : Bordeaux Sud-ouest

1.3.1 INRIA BSO

C'est en janvier 2008 que fut la création du centre de recherche INRIA Bordeaux Sud-Ouest avec le soutien du Conseil Régional d'Aquitaine.

Le centre compte aujourd'hui 20 équipes de recherche accueillant différentes nationalités et qui produisent plus de 500 publications par an.
Les thématiques de recherches de ces équipes sont axées essentiellement sur : Algorithmes, Programmation des réseaux et des systèmes distribués et sûrs ; Modèles et Simulations ; Perception, Cognition, Interaction.
Il s'agit alors de répondre à des questions qu'on pose à travers la modélisation informatique et mathématiques ainsi que la programmation des systèmes complexes et les interactions entre agents humains et artificiels.

Les équipes de recherche de INRIA BSO travaillent en grande majorité avec des établissements d'enseignement supérieur et de recherche.
Dans le cadre de ces partenariats, INRIA BSO mène des activités de recherche en Sciences et Technologies du Numérique riches dans un environnement particulièrement dynamique. Ces équipes sont également associées à des partenaires industriels comme Total, Thales,

France Telecom, Airbus, etc. Ces engagements permettent d'envisager une croissance importante au niveau des travaux des équipes dans les années à venir.

1.3.2 L'équipe Géostat

L'équipe Géostat pour Géométrie et statistiques dans les données d'acquisition. Ses travaux portent plutôt sur l'optimisation, apprentissage et méthodes statistiques.
Ses axes de recherche englobe ainsi les domaines suivants :
– prédictabilité dans les systèmes complexes,
– ondelettes optimales,
– méthodes multiéchelles issues de la physique pour l'analyse des systèmes complexes,
– analyse, classification et détection.

Ainsi, au sein de Géostat on s'intéresse essentiellement aux domaines applicatifs suivants :
– signaux complexes en astronomie, optique adaptative
– analyses des signaux turbulents issus des observations satellitaires.
– analyse du signal parole

1.4 Conclusion

Toutes les activités de l'INRIA s'appuient sur le transfert technologique selon des stratégies bien étudiées. C'est ce qui fait sa grande notériété aujourd'hui dans le monde de la recherche et le développement scientifique.
Toutes les informations de ce chapitre se trouve sur le site web de l'INRIA [1].

1. http ://www.inria.fr/

Chapitre 2

Cadre du projet

2.1 Introduction

Dans ce présent chapitre, on présente le projet de fin d'étude dans son cadre d'application en mettant au point la stratégie à suivre pour parvenir à la solution.

2.2 Problématique et objectif

Il est vrai qu'un trouble de la voix peut avoir un impact néfaste sur la vie professionnelle et sociale, de ce fait, les études portant sur les dysfonctionnements de la voix et de la parole se sont beaucoup développées au cours des 15 dernières années. Ces dernières portent sur différents types d'analyse permettant de mettre en évidence les troubles de la voix.

Ici on considère l'étude des pathologies de la voix du point de vue traitement du signal de parole. Ce point de vue implique forcément la considération des approches mathématiques utilisées dans ce domaine.

En effet, dans le cadre de ce projet de fin d'étude d'intitulé : "Analyse non linéaire de la parole pour la détection des voix pathologiques", on cherche en premier lieu à étudier les mesures acoustiques existantes et les modifier dans le but qu'elles deviennent plus pertinentes en terme de discrimination entre VN et VP. Ensuite on vise à mettre en

place de nouvelles mesures acoustiques correspondant à une réalité physiologique et qui conviennent à l'exercice de la séparation des deux classes de voix.

En effet, jusqu'à présent deux types de mesures ont été adoptées :

1. les mesures linéaires :
   - basées sur des transformations linéaires du signal.
2. les mesures non linéaires :
   - basées sur l'analyse des systèmes déterministes et aléatoires non linéaires.

Il est admis que les mesures linéaires ne sont pas suffisamment adaptées pour l'analyse des voix pathologiques, c'est pour cela qu'on fait appel aux mesures non linéaires [1]. Cependant, le défaut majeur de ces mesures est qu'elles ne correspondent pas à une réalité physiologique de la production de la voix.

Notre objectif consiste alors à mettre en place des mesures acoustiques basées sur des approches non linéaires en remédiant à la faiblesse cité ci-dessus, c'est-à-dire en les faisant correspondre à une réalité physiologique de la production de la voix et en obtenant des taux prometteurs en terme de classification des voix normales et voix pathologiques.

Pour ce faire, tout d'abord une compréhension du modèle de la production vocale s'impose. C'est ce qu'on introduit dans le paragraphe suivant.

2.3 Production de la voix

La production de la voix s'avère être l'un des mécanismes les plus complexes dans le corps humain. Entre la bouche et les poumons, il y a tout le nécessaire pour créer des sons riches et variés.
Ainsi le signal de parole produit par l'homme est constitué de zones de sons quasi-périodiques (voyelles et consonnes voisées), des zones apériodiques (fricatives) et des zones de sons impulsionnels (plosives).
Le signal de parole comporte aussi des informations sous des aspects très diversifiés dont les caractéristiques changent au cours du temps, d'où le statut de signal non stationnaire.
La figure 2.1 donne la composition de l'appareil phonatoire humain qui comporte 3 parties essentielles : la structure sub-glottique (poumons, branches et trachée), le larynx (cordes

vocales et glotte) et le conduit vocal (pharynx, cavité buccale, nasale, joues et langue).

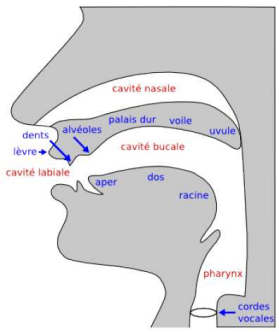


FIGURE 2.1 – Coupe de profil des organes phonatoires

L'air passant à travers la trachée met la glotte en vibration ce qui met les cordes vocales en tension et il y a par la suite une production de son correspondant à un signal quasi-périodique voisé. La fréquence de vibration correspond alors à la fréquence fondamentale du son émis, elle définit le nombre de cycles produits par seconde. C'est ce qu'on appelle le pitch (F0). La fréquence fondamentale est un critère caractéristique à chaque personne. C'est une fonction du sexe et de l'âge aussi ; chez l'homme, elle varie de 80 Hz à 100 Hz. Pour une femme, elle est entre 175 Hz et 300 Hz, et quant à un enfant, sa fréquence fondamentale est très variables allant de 200 Hz à 600 Hz.

2.4 Modélisation des signaux vocaux

Le signal de parole, comme l'on a expliqué dans la section2.3 est en gros le résultat de l'excitaion du conduit vocal par un train d'impulsions ou un bruit donnant lieu respectivement aux sons voisés ou non voisés[2]. De ce fait, cette production peut être approximée par un modèle source-filtre comme l'indique la figure 2.2.

La sortie du filtre linéaire est ainsi donnée par l'équation suivante où s est le signal de parole produit en résultat :

$$s(n) = g(n) * h(n) = \sum_{m} h(m).g(n-m)$$

Le modèle source-filtre est employé dans une large panoplie d'études vu sa simplicité. Il permet de décrire le processus de production de la parole comme étant le résultat de

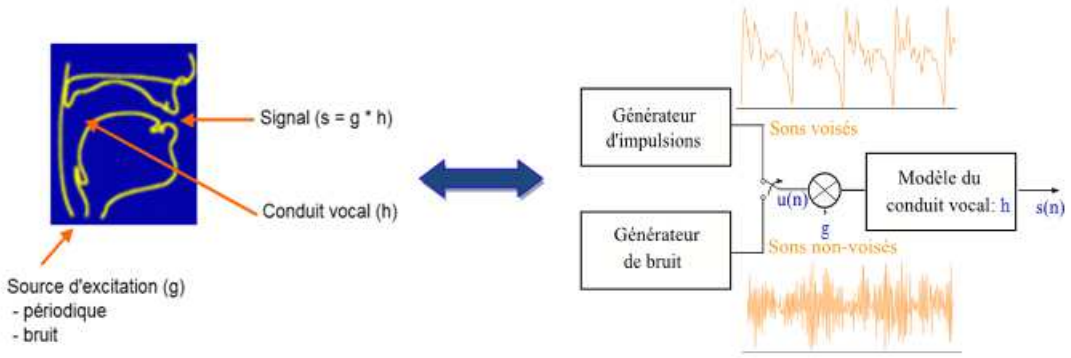


FIGURE 2.2 – Modèle de production de la parole

deux contributions indépendantes de la source et du conduit vocal.

Le défaut de ce modèle est qu'il ne tient pas compte des aspects non linéaires dans la production de la voix[3]. En effet, les études portant sur l'analyse des voyelles tenues ont montré que la non linéarité est une caractéristique de la génération de la parole. Ceci étant vrai pour les voix pathologiques et les voix normales également.

En effet, certaines pathologies de la voix se manifestent par un arrêt total des oscillations. Le cas contraire est aussi possible : on note des oscillations interminables, ainsi le signal de parole est dominé par un bruit provenant de l'écoulement d'air à travers le conduit vocal. Dans ces cas, le traitement de signal non linéaire prend place pour représenter au mieux la production vocale en se rapprochant de la réalité et en surmontant les inconvénients de la modélisation linéaire.

Il est vrai que la modélisation non linéaire avec tout ce qu'elle ramène d'outils et de mesures a réussi à résoudre certains problèmes, mais elle présente des points faibles qu'on doit prendre en considération et parmi lesquels on cite la description sophistiquée de certains problèmes qui s'avère injustifiée surtout dans le cas où le modèle linéaire peut résoudre ces problèmes d'une manière optimale.

Dans ce projet on fait appel à un algorithme fondé sur une méthode de traitement non linéaire de signal dont la fonction principale est la détection des moments de fermeture de la glotte. Ce sera la base de toutes les mesures et les descripteurs employés plus trad.

2.5 Méthodologie

S'agissant d'un projet qui s'inscrit dans un cadre de recherche et développement, il paraît évident que la première étape nécessaire pour entamer ce travail consiste en une recherche bibliographique approfondie. Alors en débutant le travail on s'est proposé d'élaborer la veille technologique dans le domaine de détection des voix pathologiques. Cette étude a porté sur les méthodes appliquées par les spécialistes en médecine ainsi que les méthodes découlant des approches de traitement de signal.
On s'est intéréssé particulièrement aux mesures acoustiques standards ayant toujours permis de discriminer entre VN et VP.

Ayant acquis des connaissances plus approfondies du domaine, la deuxième étape consiste à suggérer des modifications au niveau de ces mesures afin qu'elles caractérisent les voix d'une manière plus pertinente.
En effet, notre philosophie de travail se base sur l'exploitation de l'algorithme IFG. Il s'agit du point fondamental de tout ce qui sera développé. C'est un algorithme non linéaire développé dans l'équipe Géostat qui permet de détecter toutes les transitions importantes dans un signal donné. Il a alors été adapté pour la détection des instants de fermeture de la glotte. Il a été démontré que l'algorithme des IFG surpasse tous les algorithmes de l'état de l'art surtout en présence du bruit. D'ailleurs les algorithmes de l'état de l'art ne travaillent que sur des signaux voisés, alors que celui qu'on emploie dans le cadre de ce projet ne nécessite forcément pas des signaux voisés pour avoir de bonnes performances, la preuve est qu'il est capable de détecter les transistions importantes même pour les voix perturbées (pathologiques dans notre cas).
Ainsi, grâce à l'algorithme IFG on se propose de redéfinir le pitch. Ceci sera expliqué en détail dans le chapitre 5.
Ensuite dans la troisième étape on se propose d'étudier les statistiques des distributions des séquences du pitch obtenues par l'algorithme IFG. Ceci mènera à la quatrième étape où on introduit nos nouvelles mesures.
Une partie du travail sera consacrée au choix de la méthode de classification car l'objectif final est de parvenir à mettre en place de nouvelles mesures non linéaires qui ne correspondent seulement pas à une réalité physiologique, mais qui sont en plus capables de différencier d'une manière pertinente entre les deux classes de voix qu'on vient de citer

ci-dessus. Donc un bon choix de classifieur s'impose. Ce choix a été fait en se référant à des travaux antérieurs ayant utilisé les mêmes données que celles exploités dans ce projet[1, 8].

Afin d'évaluer la pertinence de ces travaux, une étude comparative est réalisée au fur et à mesure qu'on avance sur nos résultats. On évaluera ainsi dans ce présent rapport les performances de nos mesures en les comparant avec celles obtenus par d'autres algorithmes connus pour ce genre d'application.

2.6 Conclusion

Cerner la problématique du projet est une étape curciale. Elle permet d'éclaircir l'objectif et exposer ainsi la démarche suivie. C'était l'objectif de ce chapitre qu'on résume en 2 points :

- le but final se compose en deux parties : mieux décrire les voix normales et pathologiques par des descripteurs audio basés sur des principes non linéaires du signal en faisant correspondre ces descripteurs à une réalité physiologique dans la production de la voix. Ensuite mettre en place de nouvelles mesures acoustiques. Avec toutes ces mesures, on vise à avoir des taux de classification considérables.
- pour ce faire, une étape d'identification des approches existantes doit être faite. A partir de cette étude on sélectionnera les mesures à considérer pour commencer notre propre travail. Les résultats obtenus avec ces mesures là seront la base sur laquelle vont reposer les étapes qui suivront.

Chapitre 3

Veille technologique dans le domaine de traitement de parole pour la détection des voix pathologiques

3.1 Introduction

Ce chapitre est consacré à l'étude de l'état de l'art. On présentera ainsi au travers ses sous sections les principales pathologies de la voix ainsi que toutes les techniques mises en œuvre pour la mise en évidence de ces dérnières. L'objectif final de ce chapitre est de parvenir à identifier et maîtriser les différentes mesures acoustiques mis en place pour le jugement et l'évaluation de la qualité vocale.

3.2 Principales pathologies de la voix

Quand la voix est altérée plusieurs composantes impliquées dans la production de la parole se trouvent perturbées à des degrés différents. Les paramètres de la voix qui sont au nombre de 3 sont alors affectés. Ces paramètres sont définis d'un point de vue acoustique de la manière suivante :

– Hauteur : c'est le paramètre de la voix qui définit le caractère grave ou aigu. Il correspond au mécanisme de vibration des cordes vocales, plus les vibrations sont nombreuses

et plus le son est aigu. L'hauteur est aussi appelée fréquence fondamentale et caractérisée par sa large variabilité.

- Intensité : c'est le paramètre gâce auquel on peut distinguer un son fort d'un son faible. L'intensité est fortement corrélée avec l'énergie du signal vocal et à l'amplitude des vibration.
- Timbre : c'est le paramètre à travers lequel on peut identifier une personne juste en écoutant sa voix. Il dépend des caractéristiques anatomiques des cavités de résonance et des conditions d'accolement des cordes vocales ainsi que de leur épaisseur.

Comme les troubles de la voix agissent sur ces paramètres acoustiques, une gamme de pathologies de la voix doit être examinée. Ainsi on donne sur la figure 3.1 les origines des dysfonctionnements de la voix ainsi que leurs natures :

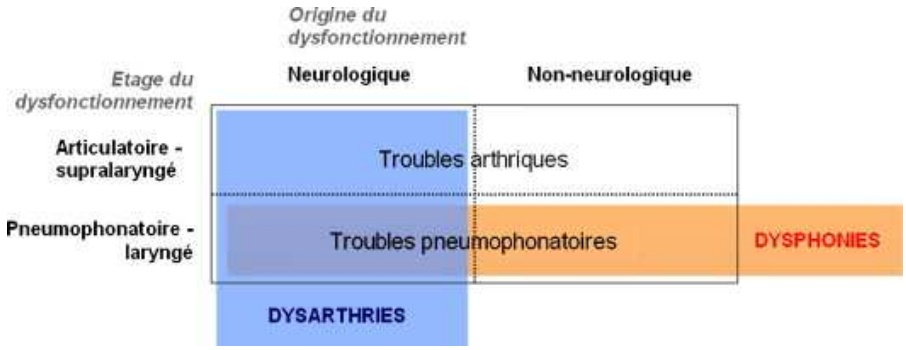


FIGURE 3.1 – Origines des troubles de la voix

On introduit alors les principales pathologies de la voix comme suit :

- La dysphonie : elle se manifeste souvent par l'enrouement, dont l'origine est soit neurologique entraînant des troubles de la mobilité des cordes vocales, ou bien organique et dans ce cas on constate des lésions des cordes vocales ou du larynx.
- La dysarthrie : il s'agit d'un trouble au niveau articulatoire où les muscles ne sont plus aptes à exécuter des mouvements rapides et précis.
- La dysprosodie : il s'agit d'un désordre au niveau de l'élocution, atteignant la mélodie du discours. La personne atteinte présente une expression verbale lente, une intonation monotone avec des pauses inappropriées et des accélérations brèves.

On note alors que ces pathologies soit elles proviennent d'une altération du mouvement des cordes vocales soit elles les affectent. Les cordes vocales sont caractérisées par deux positions principales qu'on présente sur la figure 3.2.

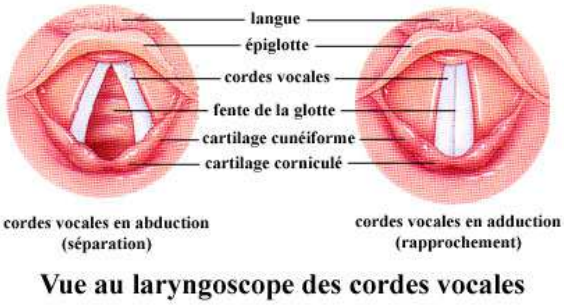


FIGURE 3.2 – Positions des cordes vocales

Lorsque les cordes vocales sont affectées, leurs fermeture peut devenir incomplète ce qui serait à l'origine du bruit causé par l'air dissipé du conduit vocal. Afin de mettre en évidence ces troubles de la voix et de mesurer leur degré de sévérité, un examen de la voix se trouve indispensable. C'est ce qu'on introduit dans la section suivante.

3.3 Les examens de la voix

La voix est une partie du corps qui reflète le vécu d'une personne, c'est pour ceci que dans son diagnostic il est préférable de focaliser l'attention sur l'ensemble de l'individu. Pour ce faire, plusieurs types d'examens de la voix ont été proposés par les spécialistes[9] :

3.3.1 L'analyse perceptive

Elle repose sur le jugement humain et la capacité de l'auditeur à évaluer la qualité de la voix du patient. C'est une évaluation qui met en jeu beaucoup de subjectivité. De ce fait, il est nécessaire de faire appel à un jury d'experts et ce dans le but d'augmenter la fiabilité des réponses aux tests. Pour qualifier ce qu'ils entendent, les médecins utilisent le protocole d'Hirano : c'est l'échelle GRBAS recommandée par la Société Française de Phoniatrie. Elle est formée de 5 items qu'on liste :

- G - Grade : pour l'appréciation globale de la qualité de la voix
- R - Roughness : il s'agit de l'impression audible de fluctuations anormales de la fréquence fondamentale.
- B - Breathiness : concerne l'impression audible de fuite d'air en phonation
- A - Asthenicity : jugement de la fatigue vocale ou de la voix hypotonique

– S - Strain : permet la distinction de voix forcée et voix hypertonique

A chaque item une note de 0 à 3 doit être attribuée : 0 normale, 1 peu altérée, 2 modérément altérée, 3 très altérée. En pratique, ce sont les items G, B et R qui sont retenus car ils ont fait preuve de fiabilité et de pertinence clinique. Certes il existe d'autres échelles, mais cette dernière est la plus utilisée.
Notons que l'analyse perceptive appelée "Gold Standard" est le moyen de référence pour valider toutes les autres méthodes d'évaluation de la qualité vocale.

3.3.2 Imagerie glottique

Cette approche est réalisée par deux méthodes :

– La vidéokymographie et la cinématographie digitale ultrarapide qui donnent une analyse détaillée des oscillations apériodiques des cordes vocales. Le nombre d'images par seconde peut atteindre 2000.
– La vidéolaryngostroboscopie renseigne sur les caractéristiques vibratoires des plis vocaux telles que l'occlusion glottique et sa symétrie en donnant des appréciations sur une échelle de 0 à 3.

3.3.3 Aérodynamique phonatoire

C'est l'approche évaluant l'utilisation laryngée de l'air des poumons pour la production de la voix. Il y a essentiellement 3 paramètres à extraire à partir de cette famille d'examen :

– temps maximum de phonation (TMP) : il s'agit d'un test simple permettant de quantifier l'utilisation de l'air pulmonaire dans la phonation. Le patient doit exécuter une inspiration profonde, ensuite émettre une voyelle tenue un (/a/) par exemple en phonation confortable le plus longtemps possible.
– débit aérien phonatoire : mesuré à l'aide d'un pneumotachographe lors de l'émission d'une voyelle tenue.
– pression sous glottique (PSG) : donnée en mesurant la pression intraorale (PIO). Le patient prononce d'une manière répétitive une syllabe comportant une voyelle et une consonne occlusive non voisée, un /papapa/ par exemple.

3.3.4 Analyse acoustique : approche qualitative /approche quantitative

Bien que l'analyse acoustique peut se diviser en deux familles d'approches, elles reposent toutes sur le même principe qui est la décomposition du signal sonore selon ses composantes fréquentielles : fréquence fondamentale, les harmoniques et les formants. Ce type d'analyse permet le calcul de la fréquence fondamentale et de sa variabilité (approche quantitave). On obtient alors le phonétogramme qui est un tracé de F_0 en fonction de l'intensité vocale. C'est une mesure de la performance laryngée extrême car elle renseigne sur l'étendue complète de la voix du point de vue de la hauteur et de l'intensité sonore.

3.3.5 Techniques adjuvantes

Elles sont au nombre de 2 :

– L'électroglottographie (EGG) : c'est une mesure non invasive permettant d'obtenir une image de l'accolement et de l'ouverture des cordes vocales. Elle permet également une très bonne représentation du cycle d'oscillation des cordes vocales. Les mesures en électroglottographie sont dénuées de bruits aérodynamiques, c'est pour celà qu'elles donnent une bonne estimation de la fréquence fondamentale. Lors du test, les éléctrodes sont placées de part et d'autre du cartilage thyroïde comme l'indique la figure 3.3 :

FIGURE 3.3 – Electrodes de l'EGG

– L'électromyographie des muscles intrinsèques du larynx (EMG) : elle concerne le rôle des muscles intrinsèques du larynx dans le contrôle de la fréquence fondamentale de la voix. Les muscles sont étudiés au repos, en respiration et en phonation.

3.3.6 L'auto évaluation par le patient

L'instrument classique utilisé dans cet examen est le test VHI pour Voice Handicap Index. Il permet de quantifier l'impact des troubles vocaux sur la qualité de la vie. Son

utilisation donne des résultats qui satisfont les cliniciens et les chercheurs. Ce test contient dix questions dont la réponse est une parmi 5 choix multiples gradués. Ce test est traduit en plusieurs langues. Dans l'annexeA on donne un exemple en Anglais.

3.3.7 Approche multidimensionnelle : Protocole de l'ELS (European Laryngological society)

Comme la voix est un phénomène extrêmement complexe, son exploration impose l'analyse simultannée de plusieurs paramètres. C'est pour cela que la société européenne de laryngologie a mis en place un protocole basique d'analyse suivant cinq dimensions :

1. Dimension perceptive réalisée en se basant sur 3 items de l'échelle GRBAS d'Hirano qui sont : G, R et B.
2. Dimension biomécanique pour la mesure de l'occlusion glottique et de la régularité de son mouvement.
3. Dimension aérodynamique afin de mesurer le quotient de phonation.
4. Dimension acoustique dont l'objectif est l'étude de la fréquence fondamentale et ses variations ainsi que l'amplitude et ses perturbations.
5. Dimension de l'auto-évaluation par le patient lui même en jugeant la qualité de sa voix ainsi que son impact sur son environnement.

3.4 Protocol expérimental de l'évaluation de la voix

L'évaluation de la voix met en jeu plusieurs éléments. Il y a alors l'exercice vocal à demander du patient, le matériel d'enregistrement, le cadre de l'enregistrement, etc. Toutes ces composantes doivent être séléctionnées d'une manière adéquate au type du test à réaliser.

3.4.1 Les enregistrements

Qu'il s'agisse d'analyse perceptive ou instrumentale, ces deux approches nécessitent qu'on enregistre la voix du patient conecrené. Il est à noter que plusieurs conditions s'imposent lors de l'enregistrement afin d'éviter des mesures et des jugements érronés.

3.4.1.1 Support vocal

Appelé aussi matériau phonétique, le support vocal à enregistrer doit être standardisé et représentatif de la voix du sujet. On rencontre généralement deux types de support utilisé : la parole et la voyelle tenue.

- La parole : c'est le matériau clinique par excellence. Elle représente le matériau le plus intuitif et le plus facile à évaluer. De plus, elle contient partiquement un nombre considérable d'information dont le clinicien aurait besoin, telles que : la production des phonèmes voisés qui témoignent de la vibration des cordes vocales, les transitions (passage entre phonème non-voisé / voisé), hauteur, intensité et intonation. On retrouve fréquemment le premier paragraphe de "La chèvre de Monsieur Seguin" d'Alphonse Daudet. L'utilisation systématique d'un même texte permet la standardisation et la comparaison des données.
- La voyelle tenue : on utilise les voyelles tenues car en les comparant aux autres lettres elles sont les moins affectées par le mouvement articulatoire et les contextes prosodiques (accents régionaux par exemple), et par la suite elles permettent à l'auditeur de se concentrer sur les caractéristiques laryngées.La voyelle /a/ est très connue pour son utilisation dans ce domaine. C'est une voyelle commune à de nombreuses langues, ce qui simplifie la question de l'accent. Elle est aussi peu complexe à produire et facilement reproductible. Le /a/ tenu est intéressant également parce qu'il est peu perturbé par l'émotion du sujet.

Dans la littérature, une diversité de points de vue s'est présenté quant à l'utilisation de la parole et du /a/ tenu. Mais à la fin, les arguments favorisent l'utilistaion de la voyelle tenue pour l'évaluation instrumentale, et la parole est plutôt orientée pour l'analyse perceptive qui est le standard de référence. Dans le tableau suivant, on présente les différents exercices demandés au patient le jour de l'examen où il se présente pour fournir des échantillons de sa voix. Les tâches utilisées dans l'analyse de la voix consistent à produire des voyelles isolées, lire des phrases courtes et produire un monologue spontanné sur une figure donnée.

L'enregistrement de chaque patient s'obtient au cours d'une seule séance. Dans le cas où le patient ou le spécialiste n'est pas satisfait de la qualité de production, la répétition d'une ou plusieurs tâches est alors effectuée. L'enregistrement est précédé d'une discussion de quelques minutes avec le patient pour l'explication du déroulement de l'examen. L'examen proprement dit dure de 30 à 45 minutes.

Exercice	Information vocale
Exercice 1	La voyelle tenue /a/ pendant au moins 5 secondes sur un ton identique
Exercice 2	La voyelle tenue /a/ le plus longtemps possible sur un ton identique
Exercice 3	Les séquences /papapa/
Exercice 4	Prononciation de /pataka/ de 3 à 5 fois
Exercice 5	La phrase "Papa ne m'a pas parlé de beau-papa"
Exercice 6	Les premières lignes du texte "La chèvre de monsieur Seguin" d'Alphonse Daudet pendant 1 minute sans consignes de lecture
Exercice 7	Les premières lignes du texte "La chèvre de monsieur Seguin" d'Alphonse Daudet pendant 1 minute à voix haute
Exercice 8	Description de l'image de "Voleur de biscuit" (pour le monologue spontanné)

TABLE 3.1 – Liste des exercices vocaux

3.4.1.2 Cadre de l'enregistrement

Les enregistrements vocaux se réalisent théoriquement dans une pièce insonorisée, mais compte tenu des conditions réelles dans les hopitaux et dans les cabinets de consultation, une telle disposition est difficile à réaliser. Pour cela, il est préférable de se placer dans des salles un peu isolées, calmes avec une rumeur de fond inférieure à 40 dB afin d'éviter au maximum tout bruit préjudiciable au diagnostic.
Pour l'utilisation des microphones, il faut éviter de disposer le microphone dans une salle trop réverbérante ou près d'une source de bruit (telle que le ventilateur du PC ou une fenêtre) ou sur une table qui transmet fortement les bruits (pendulette etc..)
Lors de l'examen, le patient doit être assis. Pour les mesures aérodynamiques, la distance entre la bouche et le micro est fixée à 5 cm. Pour les mesures acoustiques, le micro est placé à 30 cm de la bouche.

3.4.1.3 Matériels

Au fil du temps, divers outils ont été employés pour réaliser les tests de la voix et de la parole. Dans cette étude, on cite 3 types de matériel récents : l'enregistreur numérique portatif, le dispositif DIANA et le dispositif EVA-2. En parallèle à ces enregistrements, il y a aussi la mesure du signal électroglottographique.

Enregistreur numérique portatif de haute qualité La qualité du son enregistré dépend de la qualité du micro utilisé. De ce fait, le micro à utiliser doit être associé à

un faible bruit et capter le son avec le moins de distorsion possible.
Il faut faire attention aux microphones éléctrodynamiques, même ceux de qualité, ils sont à rejeter car ils perturbent trop la voix par leurs distorsion et manque de dynamique engendrant ainsi un bruit de fond excessif. Aussi, tout microphone économique livré parfois avec des PC est à prohiber, car ils sont de mauvaise qualité.
Un point crucial pour les microphones, c'est la manière dont on les utilise : il est nécessaire de maintenir le micro à une distance à peu près constante du sujet. Pour cela, il ne faut pas demander au locuteur de tenir lui même le microphone mais fixer ce dernier sur un pied étudié pour cet usage. Ainsi le microphone est toujours à une distance constante de sa bouche. On peut aussi le disposer latéralement à quelques centimètres de la commissure des lèvres pour éviter qu'il soit sur la trajectoire du débit d'air oral. Cette disposition privilégie le signal oral en atténuant les bruits parasites de l'environnement.

Le dispositif DIANA C'est le Dispositif Informatisé d'ANalyse Acoustique conçu pour l'étude de nombreux paramètres objectifs de la phonation tels que : le son, la hauteur, l'intensité, la stabilité... Il fonctionne avec un micro-ordinateur PC sous Windows. Les logiciels d'investigation clinique dont il dispose permettent d'appliquer des protocoles d'analyse physiologique de la voix. Il permet de synthétiser les données collectées pour chaque patient afin d'en faciliter l'interprétation. Un gestionnaire de patients permet de restituer, classer, comparer les données et de formuler des commentaires insérés dans les enregistrements.

Le dispositif EVA-2 C'est le système d'Evaluation Vocale Assistée, une évolution de DIANA permettant d'effectuer des mesures supplémentaires regroupant ainsi la plupart des paramètres de la production de parole : son, hauteur, intensité de la voix, débits d'air, pressions...
EVA-2 est doté de nombreux capteurs permettant ces prises de mesure permettant ainsi au praticien d'affiner le diagnostic, de faire le suivi d'une intervention chirurgicale, d'un traitement pharmaceutique ou d'une rééducation vocale. Cet appareillage permet d'enregistrer simultanément des paramètres acoustiques et aérodynamiques par l'utilisation d'une pièce à main. Cette pièce contient un microphone, un sonomètre, un pneumotachographe mesurant le débit d'air oral et nasal ainsi que deux capteurs de pression.

EVA-2 permet aussi de synchroniser d'autres appareils tel que l'éléctroglottographe qui permet d'obtenir une image de l'accolement et de l'ouverture des cordes vocales.

L'éléctroglottographie (EGG) L'appareil pour réaliser l'électroglottographie est constitué de deux électrodes que l'on place sur le cou du patient, au niveau du cartilage thyroïdien et qui sont reliées à un générateur de courant constant et de très faible intensité. L'EGG permet d'obtenir une image de l'accolement et de l'ouverture des cordes vocales, il donne une très bonne représentation du cycle d'oscillation de ces dernières et permet une bonne mesure de la fréquence fondamentale car il est dénué de bruits aérodynamiques.

3.4.2 Description des dispositifs DIANA et EVA-2

DIANA et EVA-2 offrent des logiciels d'investigation clinique permettant d'appliquer des protocoles d'analyse physiologique de la voix et de la parole. Tous les programmes fonctionnent dans l'environnement Windows. Ils permettent d'organiser les données de chaque patient afin de faciliter les manipulations telles que l'ajout de commentaires. Les deux appareillages offrent l'interface donnée sur la figure 3.4 pour l'enregistrement des patients :

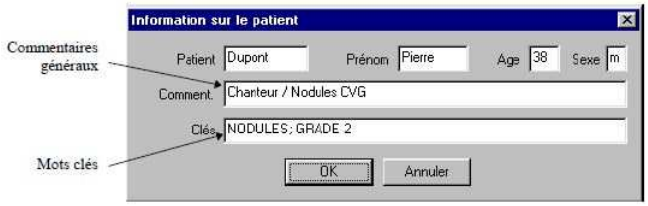


FIGURE 3.4 – Interface d'enregistrement des patients avec DIANA et EVA-2

Le gestionnaire des enregistrements des patients offre la possibilité de classer les données par patient, par type d'examen effectué, par date et même par commentaire.
Les protocoles de mesure disponibles pour DIANA et EVA-2 permettent d'effectuer des mesures acoustiques. EVA-2 offre, en plus des mesures acoustiques, des mesures aérodynamiques.

3.4.2.1 Mesures acoustiques fournies par DIANA et EVA-2

Ces deux dispositifs ont des logiciels d'investigation clinique pour appliquer des protocoles d'évaluation acoustique de la voix et de la parole qu'on présente ci-dessous :

- Profil vocal : les indices proposés par ce protocole sont : jitter ratio, jitter factor, relative average perturbation (RAP), shimmer factor, amplitude perturbation quotient (APQ), signal ratio, normalized noise energy (NNE), harmonic to noise ratio (HNR). Les différentes mesures sont présentées sous forme de tableaux, de diagrammes, et sous forme de représentation radiale où chaque axe explore une dimension de la voix (permet de visualiser de façon synthétique l'ensemble des paramètres de la voix par rapport à des valeurs seuils).
- Phonétogramme : permet d'évaluer l'étendue vocale d'un sujet en enregistrant l'intensité minimale et maximale en fonction de la hauteur de la voix.
- Spectrographie : le Spectrogramme (ou sonagramme) permet de visualiser les variations spectrales du signal sonore au cours du temps.
- Feedback temps réel : permet d'évaluer la phonation dans le cas des dysarthries en visualisant en temps réel les formants des voyelles produites.
- Prosodie : permet l'analyse objective des troubles prosodiques, particulièrement ceux qui sont d'origine neurologique. DIANA permet d'étudier les variations mélodiques, les variations d'intensité, la durée des énoncés et du débit syllabique ainsi que la répartition des pauses silencieuses dans le discours.
- Temps maximal de phonation : spécialement destiné à évaluer l'endurance du patient en phonation. On obtient alors les résultats suivants : la durée de la phonation (temps maximal de phonation), la fréquence fondamentale moyenne, l'intensité SPL moyenne qui est l'équivalent d'une pression acoustique et le débit d'air moyen.
- Investigation EGG : permet d'obtenir une image de l'accolement et de l'ouverture des cordes vocales. On a alors comme résultats : la fréquence fondamentale, les paramètres d'instabilités (jitter shimmer) et le quotient de fermeture (Closed Quotient).

3.4.2.2 Mesures aérodynamiques fournies par EVA-2

EVA-2 propose des logiciels d'investigation de paramètres aérodynamiques permettant de donner une indication sur le contrôle respiratoire, d'analyser les phénomènes liés

à la source de la phonation (fuite glottique, estimation du forçage) et d'explorer les phénomènes liés à la nasalité. On présente ci-dessous les mesures fournies par EVA-2 :

- Spirométrie : c'est un protocole permettant de détecter d'éventuelles insuffisances respiratoires.
- Nasalité : permet de mettre en correspondance le signal de parole et les débits d'air oral et nasal.
- Rhinomanométrie : permet le calcul de la résistance nasale au cours de la respiration.
- Efficacité glottique : permet d'évaluer le rendement glottique en mettant en place un rapport entre la quantité d'énergie acoustique émise et la quantité d'énergie aérodynamique nécessaire à cette émission. Ces indices explorent la notion de forçage vocal. Voici un exemple de test pour mieux expliquer cette notion de forçage vocal :
  1. Le sujet prononce une série de "pa pa pa pa pa".
  2. Sur la tenue de /p/, les lèvres sont fermées et la glotte est ouverte. Un équilibre des pressions s'établit dans le conduit vocal, ainsi la pression sous-glottique est estimée en mesurant la pression intra-orale.
  3. Sur l'émission de /a/, il y a la mesure de l'intensité acoustique et du débit d'air oral.

3.4.3 Mesures standards de la qualité de la voix

On vient de voir dans les paragraphes précédents que les différents types d'analyse permettent d'établir des mesures, dégager des indices et cerner les tendances de la voix, et ce d'une d'un point de vue objectif ainsi que quantitatif. Ici on présente les mesures standards, celles qui sont prises en compte en premier lieu.
On précise qu'il existe une liste exhaustive de ces mesures, ainsi on ne présente que certaines variantes :

3.4.3.1 Le Jitter

C'est la mesure la plus connue en terme d'évaluation de l'instabilité à court terme de la fréquence fondamentale F_0. On entend dire par "court terme" le passage d'un cycle au cycle suivant. La figure3.5 donne une représentation de cycles consécutifs stables en

vibration. Le jitter caractérise la dispersion temporelle du signal vocal. En effet, quelque soit la voix, normale ou pathologique la F_0 est caractérisée par des variations. Dans le cas normale ces variations sont légères et reste toujours autour du pitch. Dans le cas pathologique, ces variations ne sont plus cernables et c'est pour cela que pour les voix pathologiques le pitch ne varie pas dans un seul sens, c'est-à-dire qu'il peut augmenter et diminuer également.

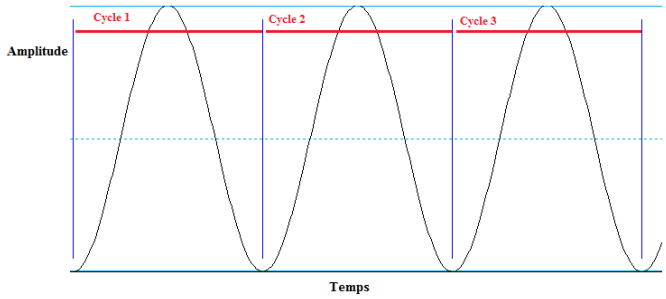


FIGURE 3.5 – Exemple de cycles en vibration stable en fréquence

Le jitter existe en plusieurs variantes [15] parmi lesquelles on cite le jitter absolu et le jitter factor :

Mesures	Formules
Jitter absolu (Hz)	$\frac{1}{N-1}\sum_{i=1}^{N-1}\lvert F_{0\,i} - F_{0\,i+1}\rvert$
Jitter factor(%)	$100.\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\lvert F_{0\,i} - F_{0\,i+1}\rvert}{F_0}$

TABLE 3.2 – Les variantes du Jitter

Certains outils avancés d'analyse de la voix fixent une valeure limite au jitter définissant ainsi le seuil pathologique. On donne l'exemple du MDVP qui fixe le seuil du jitter factor(%) à 1.04%.

3.4.3.2 Le Shimmer

C'est la mesure équivalente au jitter mais en terme d'amplitude. Le shimmer reflète le taux d'instabilité à court terme de l'amplitude des vibrations. Le shimmer est donné aussi sous plusieurs variantes. Dans le tableau 3.3 on en donne 2.

Mesures	Formules
Shimmer (dB)	$\frac{1}{N-1}\sum_{i=1}^{N-1} \|20.\log\left(\frac{A_{i+1}}{A_i}\right)\|$
Shimmer(%)	$100.\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\|A_i - A_{i+1}\|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$

TABLE 3.3 – Les variantes du Shimmer

Le A_i donne l'amplitude maximale sur l'intervalle i. Certains outils considère cette amplitude en la normalisant, d'autre non. Aussi, certains considère l'amplitude crête à crête c'est-à-dire la différence entre l'amplitude maximale et l'amplitude minimale sur l'intervalle donnée, alors que d'autres ne prennent que l'amplitude maximale sur l'intervalle.

3.4.3.3 Le HNR - Harmonic to Noise Ratio

Le HNR, une mesure explorant la présence du bruit au cours de la phonation peut être calculé selon plusieurs méthodes. Quelque soit la source de provenance du bruit dans le signal de la parole, le HNR permet de l'évaluer. En effet, plus le HNR est élevé et plus la voix est jugée comme étant une voix normale. Dans le cas inverse, c'est-à-dire en considérant le NHR, lorsque ce dernier tend vers zéro la voix est considérée normale.
Tout comme le jitter et le shimmer le calcul du HNR diffère d'un outil à un autre[15, 16, 17]. Une des méthodes les plus connues est celle élaborée par Yumoto et Gould (1982) [21]. Cette méthode propose de considérer de 25 à 50 cycles consécutifs du signal traité. En effet à travers ces cycles là on parvient à établir une forme moyenne du signal, ainsi en faisant la soustraction entre le signal original et la forme moyenne on obtient le résidu qui est considéré comme étant le bruit associé à ce signal. En attribuant H à l'énergie de la forme d'onde moyenne et N à l'énergie du bruit, on obtient le HNR tel que

$$HNR = \frac{H}{N}$$

Ainsi en considérant tous les cycles inclus dans la formation du signal on obtient la valeur final du HNR en la convertissant en dB.

3.4.4 Les nouvelles tendances de mesure

Ayant déjà exposé les paramètres classiques décrivant le comportement de la voix. On se propose dans ce présent paragraphe d'examiner les nouvelles techniques mises en place par les acteurs du domaine.
En effet, les spécialistes jugent qu'il y a une sorte de manque d'éfficacité clinique dans les résultats obtenus par les mesures classiques, c'est ainsi qu'on a pensé à soit les renforcer ou les remplacer par d'autres mesures évaluant la voix d'une manière plus détaillée et plus pertinente.

3.4.4.1 Les mesures non linéaires

Parmi l'une des nouvelles approches de calculs de paramètres acoustiques et de développement de descripteurs audio, on cite celles qui se basent sur des méthodes non linéaires afin d'aboutir à un bilan acoustique résumant au mieux l'état de la voix d'une personne donnée.
En effet, deux principes sont à la base des mesures objectives classiques : le traitement numérique linéaire de signal et les concepts classiques d'analyse de forme d'onde. Ces principes présument certaines hypothèses mathématiques sur le signal qui sont parfois loin d'être vrais en réalité. C'est ce point là qui peut limiter les performances des algorithmes classiques se référant à ces principes.
Il y a presque 30 ans qu'on s'est rendu compte de l'importance de la mesure non linéaire dans la modélisation du mouvement des cordes vocales. Après la considération de celle-ci, on a découvert une multitude de phénomènes typiquement non linéaires produits par le système vocal, et à partir de là on a commencé à se référer à l'analyse des systèmes aléatoires non linéaires dans l'étude des phénomènes observés dans les voix pathologiques.
Parmi les mesures non linéaires traitées récemment, on évoque :

3.4.4.1.1 Correlation Dimension (D2) D2 est une mesure géométrique qui décrit le degré de corrélation entre deux points représentés dans l'espace des phases [6]. Elle

est ainsi très utile pour décrire des phénomènes irréguliers d'un signal donné. Ce paramètre a été largement utilisé par les chercheurs en raison de sa simplicité et sa convergence rapide dans les calculs numériques.

Considérant une série temporelle $(Y(n))_{n\in N}$, la $D2$ est obtenue comme suit :

1. Pour chaque échatilllon, on réalise le plongement vers l'espace des phases en créant le vecteur de retard correspondant (time delayed vector)

$$Y(k) = [Y(k),\, Y(k+2\tau),\, Y(k+3\tau) \ldots Y(k+(M-1)\tau]$$

   où M est la dimension de plongement déterminée selon le "embedding theorem" et τ est le délai de plongement.

2. Dans ce nouvel espace, on calcule l'intégrale de corrélation $C(r)$ donnée par la formule suivante :

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(r - \|Y(i) - Y(j)\|)$$

   Avec Θ la fonction échelon tel que : $\Theta = 0\ si\ x \leq 0,\ 1\ si\ non$

   En effet, l'intégrale de corrélation calcule le nombre de points dont la distance de séparation est inférieure à r qui est le rayon du voisinage circulaire autour de $Y(i)$.

3. On définit ainsi le paramètre $D2$ de la manière suivante :

$$D2 = \lim_{r\to 0} \lim_{N\to\infty} \left(\frac{\partial \ln(C(r,N))}{\partial \ln(r)} \right)$$

3.4.4.1.2 Detrended Fluctuation Analysis (DFA) Mathématiquement ce nouveau paramètre acoustique est défini pour mettre en évidence les processus d'auto-similarité dans les séries temporelles. Pour le signal vocal, la DFA mesure le degré d'auto-similarité du bruit dans le signal de la parole[10].

Dans la parole, le bruit est souvent généré par un flux d'air dissipé à travers les cordes vocales, un processus pareil est caractérisé par un exposant statistique α identifiant les éléments de la dynamique dans un signal. Dans certains troubles de la voix, la fermeture incomplète des cordes vocales entraîne une modification du bruit de respiration et le caractère d'auto-similarité du bruit dans le signal de la parole est un indicateur de dysphonie.

L'algorithme DFA calcule la variation de l'amplitude du signal de parole sur une plage d'échelles de temps, ainsi l'auto-similarité du signal est quantifiée par α la pente de la ligne droite obtenue à partir de la représentation des variations de l'amplitude en fonction de l'échelle de temps.

Soit un signal $(Y(n))_{n\in N}$ représenté par le vecteur $(Y(1), Y(2) \ldots Y(N))$:

1. La première étape consiste à intégrer ce signal pour obtenir la série intégrée $(X(1), X(2) \ldots X(N))$ où
$$X(k) = \sum_{i=1}^{k} Y(i) \; Pour \; k \in \{1, \ldots N\}$$
2. Chaque échantillon de la série intégré est divisé en $\frac{N}{L} = N_{max}$ fenêtres indépendantes de longueur L. Pour chaque fenêtre, une droite des moindres carrées est estimée pour représenter la tendance de cette fenêtre
$$X_L(k) = a.k + b$$
3. Chaque échantillon est ensuite redressé en retranchant la tendance locale dans chaque fenêtre $X(k) - X_L(k) \;\; k \in \{1, \ldots N\}$
4. La DFA représente alors l'écart-type des résidus de cette regression pour toute la série. Ainsi, on peut obtenir la grandeur caractéristique des fluctuations telle que :
$$F(L) = \sqrt{\frac{1}{L.N_{max}} \sum_{k=1}^{L.N_{max}} \left(X(k) - X_L(k)\right)^2}$$
5. Cette étape consiste à répéter ce processus en commençant par la deuxième étape pour différentes longueurs de fenêtre $(L_1 \ldots L_N)$pour obtenir le graphe $\log\left(F(L_i)\right)$ en fonction de $\log\left(L_i\right)$.

   La pente de la droite de régression des moindres carrés de ce graphique α donne une estimation du paramètre d'auto-similarité du processus $(X(k))_{k\in N}$.

Ainsi, on obtient le paramètre d'auto-similarité ou l'exposant statistique normalisé

$$\alpha_{norm} = \frac{1}{1 + \exp(-\alpha)}$$

3.4.4.1.3 Recurrence Probability Density Entropy (RPDE) Il s'agit d'une nouvelle méthode mise en place afin de mesurer la périodicité d'un signal après sa recons-

truction dans un nouvel espace[11]. En effet un signal représenté dans l'espace des phases retourne toujours au même point après une certaine période T qu'on appelle "recurrence period".

Mathématiquement, ce paramètre est utile pour mesurer à quel degré une série temporelle répète la même séquence, c'est en effet très semblabe à l'autocorrélation linéaire sauf qu'on mesure la répétitivité dans l'espace des phases du système. La RPDE est par la suite jugée comme étant un paramètre plus fiable basé sur la dynamique du système qui a généré le signal. Ce paramètre a été longuement étudié, et il a monté un succès dans la détection des anomalies dans différents contextes biomédicaux, et plus particulièrement les anomalies de la parole.

En disposant d'un signal $(Y(n))_{n\in N}$ représenté par le vecteur $(Y(1), Y(2) \ldots Y(N))$:

1. Pour chaque échantilllon on réalise le plongement vers l'espace des phases en créant le vecteur de retard correspondant (time delayed vector)

$$Y(k) = [Y(k),\, Y(k+2\tau),\, Y(k+3\tau) \ldots Y(k+(M-1)\tau]$$

   M est la dimension de plongement et τ est le délai de plongement.

2. Dans l'espace des phases, autour de chaque point il y a formation d'un voisinage circulaire de rayon ϵ, à chaque fois où la série temporelle retourne à ce cercle après l'avoir quittée, les différences entre les temps successifs de retour sont enregistrés dans un histogramme. On voit les étapes de calcul de la RPDE sur la figure 3.6.

Ce processus est réalisé pour différentes périodes T_i (recurrence period), ainsi on obtient l'entropie normalisée qui est à la fin le paramètre RPDE de la manière suivante :

$$H_{norm} = -\frac{\sum_{i=1}^{T_{max}} P(T_i).\ln(P(T_i))}{\ln(T_{max})}$$

avec T_{max} défini comme étant la période de récurrence maximale.

3.4.4.1.4 Pitch Period Entropy (PPE)

Il s'agit d'un paramètre permettant d'évaluer la capacité d'une personne donnée à maintenir une fréquence fondamentale stable lors de l'examen de la voyelle tenue. En effet, toutes les personnes présentent une variation du pitch, même les personnes saines ne présentant aucune pathologie de la voix, c'est ce qu'on appelle une variation naturelle carcatérisée par des vibrations lisses et des

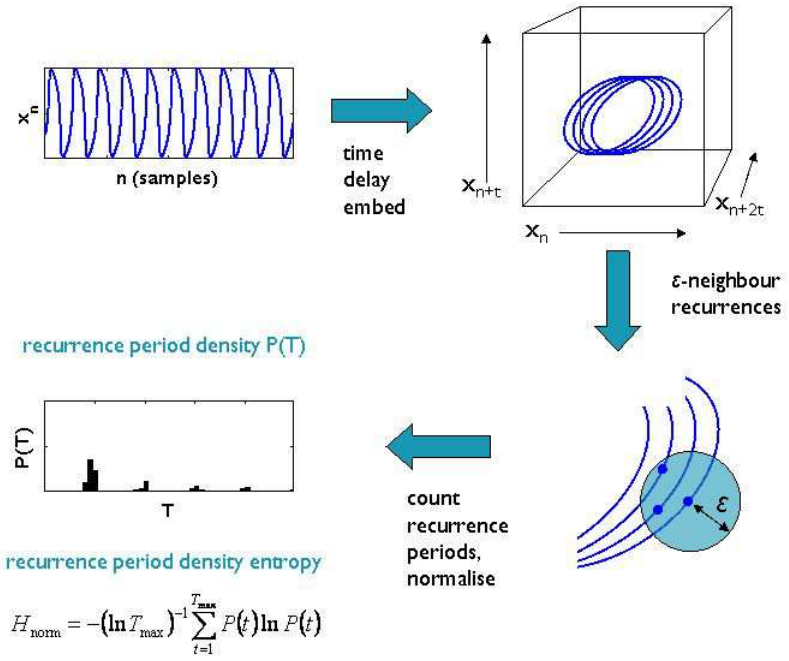


FIGURE 3.6 – Calcul RPDE

tremblements très fins. Ces variations peuvent être détectées par le jitter par exemple. Pour les personnes souffrant de dysphonie, on a repéré un symptôme lié au pitch qui est la difficulté de maintenir ce dernier stable durant les voyelles tenues.
La difficulté s'avère au niveau de discrimination entre la variation naturelle du pitch et la variation liée à la dysphonie. De ce fait, une nouvelle mesure est introduite prenant en considération la différence avec la variation naturelle du pitch et utilisant une échelle logarithmique pour le pitch[8].

1. La première étape à appliquer dans l'algorithme de calcul de la PPE est l'obtention de la séquence de fréquence fondamentale F_0 en utilisant un algorithme de détection de pitch (PDA).
2. Ensuite on convertit cette séquence à l'échelle logarithmique exprimée en demi tons. On obtient ainsi $F_{0,eldt}$ telle que :

$$F_{0,eldt} = 12.\log_2\left(\frac{F_0}{127}\right)$$

   On obtient alors la séquence $p(t)$ où p est le pitch en demi ton à l'instant t.
3. Afin d'obtenir la série qui caractérise les variations de demi tons qu'on note $r(t)$, on procède au filtrage permettant d'éliminer les corrélations temporelles linéaires et

d'aplatir le spectre.

On peut ainsi construire la distribution de probabilité discrète décrivant les variations relatives des demi-tons $P(r)$.

4. La dernière étape consiste à obtenir la PPE en calculant l'entropie de la distribution de probabilité $P(r)$.

$$PPE = -\frac{\sum_i^{L_{PPE}} p(r_i).\ln(p(r_i))}{\ln(L_{PPE})}$$

3.4.4.2 Les approches temps-fréquences

Ces méthodes reposent sur le principe d'application aux signaux de parole une décomposition adaptée en temps et en fréquences pour extraire des caractéristiques du signal de la parole à partir des paramètres dégagés de cette décomposition.
Dans des méthodes pareilles, le signal est approximé par une combinaison de fonctions qui sont le résultat de différentes translations, modulations et mises à l'échelle d'une fonction de base, bien définie dans le domaine temporel et fréquentiel[12]. De telles approches permettent de générer une version du signal d'origine en minimisant les dimensions du nouvel espace de représentation, ce qui serait de très grande utilité pour les tâches de classification. Ce type de représentation de signaux est appelé représentation parcimonieuse. En effet, un vecteur est défini comme étant parcimonieux si la majorité de ses coefficients sont nuls.
En gros, la technique utilisée dans les approches Temps-Fréquence consiste à chercher dans un ensemble de signaux ou fonctions élémentaires appelés atomes, les éléments dont la simple combinaison linéaire représenterait parcimonieusement le signal étudié. Les atomes choisis sont séléctionnés à partir d'un grand ensemble de fonctions que l'on nomme dictionnaire.
Il existe plusieurs techniques de décomposition Temps-Fréquences telles que les ondelettes, les paquets d'ondelettes, les algorithmes de matching prsuit, etc. Dans plusieurs applications liées à la voix, l'algorithme de matching pursuit a donné de bons résultats, particulièrement en employant les atomes de Gabor[13]. Cette combinaison a été sélectionnée en raison de sa résolution supérieure en temps et en fréquence ainsi que pour son adéquation pour les applications de reconnaissance de formes.
En appliquant la technique du matching pursuit avec les atomes de Gabor, chaque signal $x(t)$ est représenté par une combinaison linéaire de fonctions Temps-Fréquence $g(t)$

sélectionnées à partir du dictionnaire. Le signal $x(t)$ s'écrit alors sous la forme suivante :

$$x(t) = \sum_{n=0}^{\infty} a_n g_{\gamma_n}(t)$$

où $(g_{\gamma_n}(t))$ sont les fonctions temps-fréquences extraites du dictionnaire. Elles sont définies de cette manière :

$$g_{\gamma_n}(t) = \frac{1}{\sqrt{s_n}}.g(\frac{t-p_n}{s_n}).\exp\{j(2\pi f_n t + \phi_n)\}$$

où s_n est le facteur d'échelle contrôlant la largeur de la fenêtre. p_n est le paramètre gérant l'emplacement de la fonction dans le temps. f_n et ϕ_n sont respectivement la fréquence et la phase de la fonction exponentielle.

Le principe consiste alors à sélectionner pas à pas à partir d'un dictionnaire donné les atomes qui sont les plus corrélés avec le signal $x(t)$. On définit l'atome le plus corrélé au signal par l'atome qui attire le maximum d'énergie du signal d'origine afin de minimiser l'erreur, ce qu'on appelle résidu.

A la fin, après M itérations le signal $x(t)$ est exprimé de la manière suivante :

$$x(t) = \sum_{n=0}^{M-1} < R^n x\,,\, g_{\gamma_n} > g_{\gamma_n}(t) + R^M x(t)$$

Etant donnée R le résidu (l'erreur) après chaque itération qui, à la base était initialisée comme tel : $R^0 = x$.

Donc ici, on a présenté l'approche de représentation des signaux de la parole en temps et fréquence en se référant à la technique du matching pursuit appelée en Français poursuite de vecteurs en vu de sa simplicité et de l'éfficacité de ses résultats, mais il ne faut pas oublier de signaler l'inconvénient de cette technique qui réside dans l'énorme nombre d'itérations à parcourir afin de converger vers une solution.

3.4.4.3 Les signaux électroglottographiques

L'exploitation des signaux électroglottographiques extraits à partir de l'électroglottographie est une nouvelle approche dans la mise en évidence des troubles de la voix. L'électroglottographie est une technique utilisée pour enregistrer le comportement du la-

rynx en mesurant la variation de l'impédance électrique pendant qu'une personne parle. Le dispositif en question appelé l'électroglottographe ou EGG donne des siganux qui correspondent à l'aire de contact entre les cordes vocales. Notons alors que les signaux extraits de l'EGG ne correspondent pas à un signal acoustique, mais ils reflètent plutôt une image de l'accolement et l'ouverture des cordes vocales.

Les signaux EGG permettent d'estimer des paramètres relatifs à la source glottique tels que la fréquence fondamentale. Lors de la phase ouverte de la glotte, ces signaux varient très peu. Les informations qu'ils apportent concernet la phase fermée [20] qui peut être caractérisée par le calcul de divers quotients : quotient fermé et quotient ouvert. Parmi les paramètres les plus exploités à partir de ces signaux on cite la fréquence fondamentale et le quotient ouvert (Open quotient, Oq)[14]. L'EGG donne des mesures fiables de ces paramètres car le dispositif est lié directement au tissu qui est un meilleur conducteur que l'air.

Le Oq consiste à mesurer le rapport entre la phase ouverte du mouvement des cordes vocales et le cycle complet. Dans cette technique, on fixe un seuil H qui détermine la frontière entre phases fermées et phases ouvertes.Le seuil ayant été validé par des obserations stroboscopiques est entre 25% et 35 %. La figure3.7 donne un aperçu sur le signal EGG.

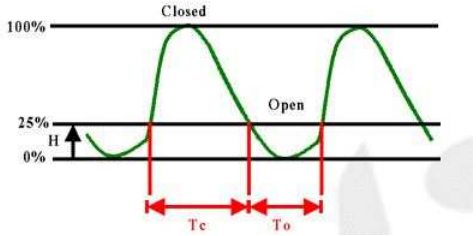


FIGURE 3.7 – Signal EGG et quotient d'ouverture

Ainsi le Oq peut être calculé de la manière suivante :

$$Oq = \frac{T_O}{T_O + T_C}$$

L'apport de cette mesure par rapport à la mise en évidence de pathologie de la voix est qu'elle permet de dégager l'existence d'éventuels phénomènes d'hyper-adduction ou hypo-adduction des cordes vocales.

Lorsque $Oq > 0.6$ les cordes vocales sont en hypo-adduction. Dans ce cas, on remarque

une fermeture incomplète de la glotte c'est à dire que les cordes vocales ne viennent pas ensemble lors de la production vocale. D'où l'origine de la voix soufflée. Comme la dysphonie entraîne des troubles de la mobilité des cordes vocales, l'$Oq > 0.6$ met en évidence la dysphonie en hypo-adduction.

Pour des valeurs de $Oq < 0.4$ les cordes vocales sont en hyper-adduction. Ceci se manifeste par un timbre serré essentiellement lors de la production de la parole. Le serrage des cordes vocales entraîne souvent une émission vocale difficile gênant l'intelligibilité. L'hyper-adduction provoque aussi un effet d'arrêt inapproprié dans la parole.

En se référant au descriptif des 3 troubles de la voix qu'on traite dans cette étude, on remarque qu'un $Oq < 0.4$ caractérise la dysarthrie, la dysprosodie et la dysphonie en hyper-adduction. La détermination de la fréquence fondamentale à partir du signal EGG est calculé à partir des instants de fermeture et d'ouverture de la glotte (IFG et IOG). En effet, chaque cycle glottique k est caractérisé par l'instant d'ouverture de la glotte $IOG(k)$ et l'instant de fermeture de la glotte $IFG(k)$. Ainsi l'inverse de la fréquence fondamentale est donné comme suit :

$$T_0(k) = IFG(k+1) - IOG(k)$$

La fréquence fondamentale est alors : $F_0 = \frac{1}{T_{0(k)}}$

Le O_q peut être donné sous la forme suivante :

$$Oq(k) = \frac{IFG(k+1) - IOG(k)}{T_0(k)}$$

Le signal EGG est caractérisé par une variation brutale au niveau des instants d'ouverture et de fermeture de la glotte. De ce fait, la dérivée du signal EGG qu'on note DEGG présente des pics importants correspondant à ces moments. Les signaux DEGG peuvent aussi être exploités pour la détermination de la fréquence fondamentale et le quotient ouvert. C'est une autre alternative pour le calcul de ces paramètres.

3.5 Conclusion

Dans ce présent chapitre on a exposé le problème de détection des troubles de la voix sur le plan médical ainsi que scientifique. En résumé, pour diagnostiquer une pathologie de la voix, il y a l'analyse perceptive et l'analyse acoustique. L'analyse acoustique s'ajoute à la

perceptive comme étant un outil d'aide au diagnostic. Les méthodes acoustiques exploitent le traitement de signal linéaire et non linéaire. Les mesures reposant sur ces approches ont été étudiées dans ce chapitre. Les mesures fondées sur les approches non linéaires ont montré une efficacité qui dépasse les mesures linéaires en terme de caractérisation de la voix. Leur défaut majeure est qu'elles ne reflètent pas la physiologie réelle de la production vocale. La motivation derrière ce travail est de faire correspondre ces mesures non linéaire à des phénomènes tangibles de la production de la voix.

Dans la suite du travail, on ne fera appel qu'aux mesures standards présentée dans 3.4.3.

Chapitre 4

Base de données et outils

4.1 Introduction

Vu la richesse et la diversité des outils employés dans la réalisation de ce traval, une phase de description de ces derniers s'avère être nécessaire.
Pour ce faire, on commence par la base de données, ensuite on présente les outils logiciels.

4.2 Base de données

L'équipe Géostat possède la "Massachusetts Eye and Ear Infirmary (MEEI) Voice Disorders Database" qui est distribuée par Kay Elemetrics. C'est une base de données contenant environ 1400 enregistrements vocaux qui sont de deux types : une voyelle tenue (/a/) et une phrase bien spécifique. Ces échantillons sont obtenus à partir d'enregistrement de 700 personnes. Il s'agit de la seule BD qui est commercialement disponible avec ce type d'enregistrements.
Cette base de donnée est souvent utilisée pour beaucoup de travaux de recherche. Tous les enregistrements et les informations cliniques relatives aux personnes archivées dans la base ont été réalisé au sein de MEEI Voice and Speech Lab. Les informations cliniques englobent des informations de diagnostic ainsi que des informations d'identification du patient (âge, sexe, tabagisme, etc). Toutes ces informations sont stockées dans des tableaux Excel.
Les informations de diagnostic sont les résultats des mesures avec le logiciel MDVP qu'on

va présenter dans le paragraphe 4.3.1. Ces informations sont affichées dans les colonnes des fichiers Excel. Plusieurs mesures sont données : les variantes du jitter, celles du shimmer, HNR, le pitch, etc. On exploitera les mesures du jitter(%) et le shimmer(%).
Notons aussi que les pathologies présentées dans cette base sont diversifiées découlant de plusieurs origines.

Dans ce présent travail, on a considéré seulement les enregistrements de voyelle tenue. Ces échantillons sont obtenus à partir d'un total de 710 personnes dont 53 possédant des voix normales et 657 souffrant de pathologie de la voix. Chaque échantillon de voix normale a une durée égale à 3 secondes. Un échantillon d'un /a/ tenu pour une voix pathologique est de durée égale à 1 seconde. La fréquence d'échantillonnage appliquée aux VN est 50 kHz, pour les VP, il y a deux fréquences d'échantillonnage : 25 kHz et 50 kHz.
Une description plus détaillée de la BD est donnée dans l'annexe B.
On précise que les fichiers audio ont une extension particulière : "NSP". De ce fait ces fichiers ne peuvent être lus que par des logiciels spécifiques. Afin de pouvoir exploiter ces derniers sous l'environnement principal du travail qui est matlab, une fonction conçue spécialement pour ce type de fichier a été téléchargée : c'est la fonction "nspread" [1].

4.3 Logiciels, environnements et algorithmes

4.3.1 Le logiciel Multi Dimensional Voice Program (MDVP)

MDVP est le premier des logiciels qu'on a utilisé dans ce travail. C'est un logiciel de référence pour l'évaluation acoustique quantitavie de la voix permettant de mesurer plus de 22 descripteurs audio pour un seul échantillon de voix [2].

MDVP fixe des seuils pathologiques pour les descripteurs qu'il mesure. Ces seuils sont fixés en se basant sur la base de données décrite dans la section 4.2. Pour chaque échantillon les résultats sont données graphiquement et numériquement. Le résultat graphique est donné dans la figure 4.1. C'est une représentation radiale où la couleur rouge présente le dépassement du seuil pathologique.

1. http ://www.mathworks.fr/matlabcentral/fileexchange/3580-nspread
2. http ://www.kayelemetrics.com/

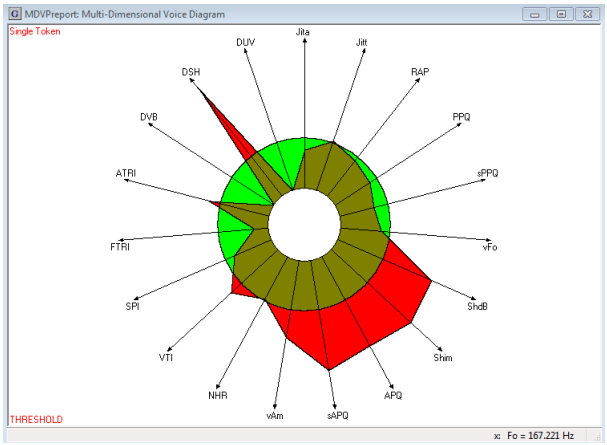


FIGURE 4.1 – Rapport d'évaluation de voix obtenu par MDVP

MDVP donne en résultat aussi un tableau de toutes les mesures qu'on voit sur la figure 4.1 en forme de tableau en précisant également les instantants des marques de pitch.

4.3.2 Le logiciel Praat

Développé par Paul Boersma et David Weenink à l'institut de sciences phonétiques de l'université d'Amsterdam en 1996, Praat est un logiciel libre et gratuit donnant la possibilité de manipuler, traiter et synthétiser les sons [3].
Praat dispose de fonctionnalités qui font de lui un outil complet pour l'étude de la parole. N'importe quelle personne pourrait se servir de ce logiciel car il détient des interfaces graphiques et des menus simplifiés qui sont utiles pour les non-experts. Pour les utilisateurs avancés, plusieurs possibilités de manipulations, scripting et mesures sont mis à disposition.

Praat permet d'établir 4 grandes familles d'analyse :
- instabilité en hauteur (jitter et ses dérivés),
- instabilité en amplitude (shimmer et dérivés),
- analyse du bruit (HNR et ses dérivés),
- évaluations de phénomènes transitoires.

Le travail avec Praat a été basé sur l'utilisation des scripts vu qu'on devait obtenir plusieurs mesures pour un large nombre d'enregistrements.
On donne l'exemple d'un script développé pour obtenir les valeurs du pitch pour les échantillons de voyelles tenues de voix pathologiques.

3. http ://www.fon.hum.uva.nl/praat/

```
form read_files
sentence source_directory C :\kaylab\DisordredVoicedatabase\PATHOL\AH\
endform

writeFileLine ("pitch_path.txt", "pitch praat ")
Create Strings as file list... list 'source_directory$'
file_count = Get number of strings

for i from 1 to file_count
select Strings list
current_file$ = Get string... i
Read from file... 'source_directory$' 'current_file$'
name$ = selected$ ("Sound")
do ("To Pitch...", 0, 75, 600)
pitch = do ("Get mean...", 0, 0, "Hertz")
appendFileLine ("pitch_path.txt",pitch)
endfor
```

L'utilisation de Praat a beaucoup servis dans le cadre d'une étude comparative élaborée au cours de notre démarche.
Les paramètres obtenus avec Praat sont essentiellement le jitter, shimmer, déviation standard par rapport à la fréquence moyenne, etc. On a aussi fait appel aux mesures des marques de pitch employés pour aboutir aux mesures données par ce logiciel.

4.3.3 L'algorithme de détection des instants de fermeture de la la glotte (IFG) de Géostat

C'est un algorithme robuste pour la détection précise des instants de fermeture de la glotte. Il est développé par Vahid Khanagha dans le cadre de sa thèse à l'équipe Géostat. Cet algorithme se base sur le Formalisme Microcanonique Multiéchelles (FMM) qui permet d'établir des analyses de la dynamique non linéaire des signaux complexes en estimant des paramètres géométriques locaux du signal qu'on appelle exposants de singularité

(EdS).
Un type particulier de ses EdS est ce qu'on appelle Variété la Plus Singulière (VPS) permet de marquer les transitions critiques du système qu'on étudie.
Pour la voix en zone voisée, les VPS sont employées pour développer un algorithme précis et robuste pour la détection des IFG.

L'algorithme de Géostat a été testé sur des bases de données connues et jusqu'à aujourd'hui, il donne les meilleurs résultats par rapport aux autres algorithmes existants pour ce genre d'application. Sa robustesse a été prouvée surtout en présence du bruit [4]. Dans la suite du travail, on appelle cette algorithme l'algorithme IFG.

4.3.4 L'environnement Matlab

C'est en exploitant les fonctionnalités de l'Environnement de Développement Intégré (EDI) matlab que les principaux travaux de ce projet ont été réalisés.
C'est la version 2012 qui a été utilisée (Matlab R2012a). Le grand avantage de Matlab consiste à offrir un grand nombre de fonctions dédiées au calcul numérique, au traitement de signal et aux statistiques.

Le choix Matlab a été fait en vu de la rapidité des calculs et la facilité de développement sous ce dernier. Notons également que Matlab présente moins de bug et de crash critiques par rapport à d'autre langages employés pour le traitement de signal.

4.4 Conclusion

Le choix de certains outils a été fait au fur et à mesure qu'on avançait dans le travail. D'autres choix ont été fixés dès le début pour avoir les idées claires et pouvoir construire notre plan d'étude.

Chapitre 5

Développements algorithmiques et résultats expérimentaux

5.1 Introduction

Ayant présenté dans le chapitre 4 tous les outils mis en œuvre pour la réalisation du projet, on présente dans ce chapitre les étapes de réalisation et les résultats.
On précise que le nombre d'enregistrements employés dans la pratique n'inclut pas tous les échantillons de voyelles tenues. On a travaillé avec 50 enregistrements de VN et 652 de VP, soit au total 702 enregistrements.

5.2 Méthodes de classification et validation

5.2.1 Méthodes de classification

Afin d'évaluer la signification de nos paramètres en pratique, on se propose d'employer les exercices de classification afin de séparer la classe des VN de celle des VP en utilisant les mesures décrites dans 5.4. On commence tout d'abord par la clarification du contexte de la classification, ensuite on proposera les méthodes employées.
En effet dans le cadre de ce projet on fait appel à la classification supervisée qui est l'opération qui consiste à discriminer les données à l'aide préalable d'un large nombre de

données déjà séparées en groupes ou en classes. Ici on ne considère que le cas binaire, c'est-à-dire on ne dispose que de deux groupes : VN et VP. La séparation est faite selon un critère où les données d'une même classe sont regroupées de manière que ce critère soit le même.

La classification se fait en deux étapes :

1. L'apprentissage : à partir de données dont les classes sont connues il y a apprentissage de la règle de décision.
2. Le test : une fois la règle de décision connue, le test se fait pour prédire les classes des nouvelles données.

Donc pour résumer, étant donné un ensemble de données déjà classées et un ou plusieurs nouvelles données, il s'agit de trouver la bonne classe à laquelle appartient ou appartiennent les nouvelles données.

Pour l'exercice de classification, on présente les données sous la forme d'un tableau rectangulaire où les lignes correspondent à l'ensemble individus I et les colonnes correspondent aux caractéristiques J. Donc la matrice de données sera de taille $I * J$ où $I > J$.

$$A = \overbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & .. & .. \\ a_{21} & a_{22} & \dots & .. & & \\ a_{31} & & & & & \\ a_{41} & .. & .. & a_{ij} & .. & .. \\ \dots & & & .. & & \\ \dots & & & & & \end{bmatrix}}^{J} \Bigg\} I$$

Dans notre cas on dispose de 702 individus ($I = 702$), à chaque individu on attribut des caractéristiques : ce sont les mesures des descripeurs qu'on présentera dans le paragraphe 5.4. Quant à J déterminant le nombre de caractéristiques pris en compte pour la discrimination, il est variable tel que $J = 1, 2$, ceci veut dire que les individus sont décrits par un seul critère $j = 1$ où deux critères $j = 2$.

Un élément nécessaire dans le cadre de ce type de classification est l'asssociation des labels ou ce qu'on appelle aussi étiquettes à chaque individu. En effet dans la phase d'apprentissage ces derniers sont censés représenter la nature de l'individu c'est-à-dire sa classe

(puisque les classes sont connues dans l'apprentissage). Après avoir mis en place la règle de décision, il suffit d'attribuer la bonne étiquette à l'individu test qu'on cherche à classer. Les méthodes de classification et l'approche de l'obtention de la base d'apprentissage et celle du test seront explicitées plus tard mais sur la figure 5.1 on donne le schéma global de l'exercice de classification.

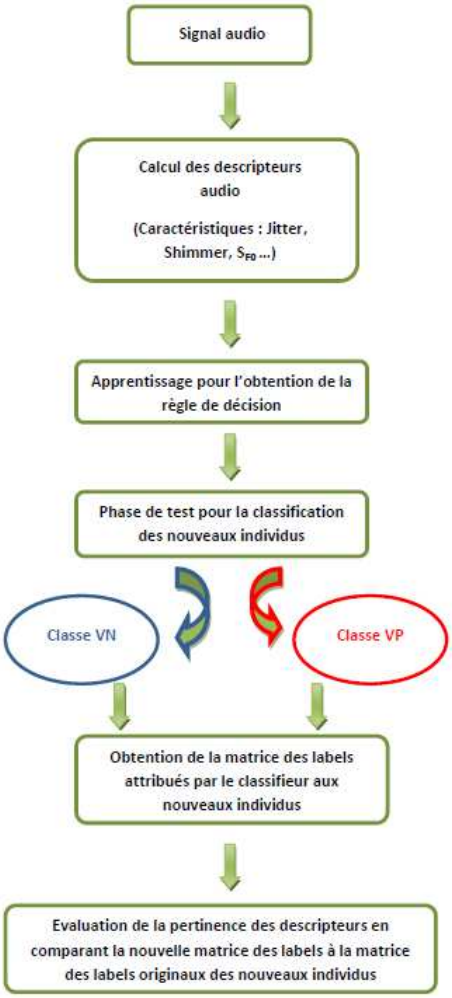


FIGURE 5.1 – Schéma global de l'exercice de classification

Machines à vecteurs de support (SVM)

SVM est un classifieur binaire discriminatif[1]. Ce type de classifieur a été employé car il est adapté aux données de grandes dimensions, il est en plus connu pour sa garantie théorique et ses bons résultats en pratique qu'on peut voir dans de nombreuses études. Lors de l'apprentissage, SVM transforme les observations dans un espace à dimensions

1. Prend en considération la séparation entre les 2 classes, contrairement au type génératif qui traite chaque classe indépendamment de l'autre

élevées dans le but de les traiter plus facilement et par conséquent rendre possible leur séparation. Ensuite SVM obtient l'hyperplan séparant les 2 classes et ayant une marge maximale. Ceci se fait en résolvant un problème d'optimisation spécifique. Quant à la marge maximale, elle est définie comme étant la distance entre cet hyperplan et les échantillons les plus proches de chaque classe. Ces échantillons sont appelés vecteurs supports. Dans la phase de classification, pour chaque entrée x, une fonction est calculée et dont le signe définira la classe de l'entrée x.

5.2.2 Méthode de validation

En apprentissage supervisé, il est nécessaire d'évaluer le performance en prédiction des classifieurs. Dans ce présent travail on a utilisé un indicateur qui est souvent le plus utilisé pour ce genre d'évalution, c'est le taux de bon classement. Alors pour les voix pathologiques que classifieur a bien classé comme étant pathologique on note cet indicateur (PP). Pour les voix normales classées normales l'indicateur est (NN).

Comme l'on a indiqué dans le paragraphe 5.2.1, les données, c'est-à-dire les mesures des descripteurs audio qu'on obtient pour les 702 enregistrements de voix sont toujours divisés en deux parties ; une pour l'apprentissage et une pour le test. La question à laquelle on veut trouver une réponse à travers ce paragraphe est : Comment faire cette partition ? En effet, ce sont les techniques de ré-echantillonnage qui permettent de répondre à ce genre de question. On entend dire par ré-echantillonnage le fait de subdiviser l'échantillon principal.

Il existe plusieurs approches de ré-echantillonner un échatillon N de taille mais ici, on a choisi la validation croisée "cross validation". Dans cette approche il y a 3 techniques :

- holdout method : on divise l'échantillon de taille N en un échantillon d'apprentissage et un échatillon de test. Ici il faut faire attention car on doit réserver la plus grande partie pour l'apprentissage. Généralement cet échantillon contient plus que 60% de l'échantillon principal.
- k-fold cross validation : l'échantillon de base est divisé en k échantillons. A chaque fois, on utilise $(k-1)$ échantillons pour l'apprentissage et la validation ou le test se fait sur l'échatillon restant.
- leave one out : c'est un cas particulier du k-fold cross validation où $k = N$.

Les résultats de ce projet sont obtenus en appliquant la dernière méthode : leave one out. Ainsi le nombre des itérations est égal à N. A chaque fois on apprend sur $(N-1)$ observations et on fait le test sur l'observation restante. Cette approche a été adaptée car notre échantillon principal présente un déséquilibre. On dispose en tout de 50 obseravtions normales et 652 pathologiques. Le problème avec les autres méthodes de validation c'est qu'en fixant au préalable et d'une manière aléatoire le nombre d'observations à considérer dans l'apprentissage, on risque d'avoir toutes ces observations dans la classe des VP (vu qu'ils sont plus nombreux). Ainsi, en employant cette méthode non seulement on évite ce risque là mais en plus on est sûr qu'on a exploité toutes les observations disponibles vu que chacune des ces dernières a été empoyée dans l'apprentissage ainsi que dans la validation[5].

5.3 Nouvelle définition du pitch

Le pitch ou la fréquence fondamentale définit le nombre d'oscillations des cordes vocales pendant une seconde. C'est plus précisément le nombre de cycles d'ouverture-fermeture des cordes vocales.
Il existe plusieurs algorithmes dont l'objectif est le calcul du pitch. L'algorithme IFG est conçu à la base pour cette tâche. En effet, les IFG présentent les marques du pitch (MP) [19]. Une fois on a ces marques, on a alors la séquence du pitch qui est l'ensemble des F_{0i} dont la moyenne donne la fréquence fondamentale.

$$F_0 = \frac{1}{N}\sum_{i=1}^{N} F_{0i}$$

En effet, les $F0_i$ correspondent à l'inverse des $T0_i$:

$$F_{0i} = \frac{1}{T_{0i}}$$

On appelle $T0_i$ la distance entre deux MP consécutifs :

$$T_{0i} = MP_{i+1} - MP_i$$

Toutes les mesures développées dans ce travail sont basées sur les MP et les distances qui les séparent. Dans la figure 5.2, on donne un exemple de distance entre les MP. La séquence du pitch est formée à partir des distances qui séparent les MP.

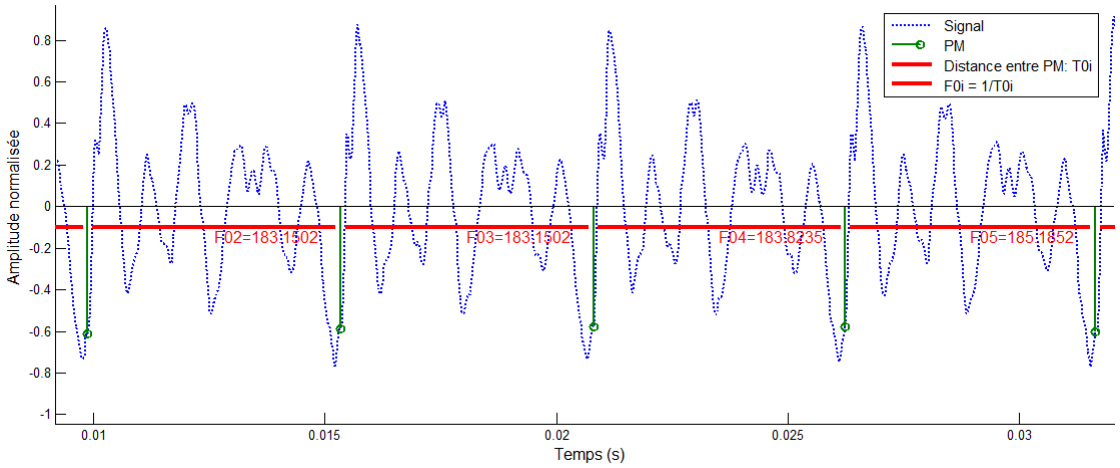


FIGURE 5.2 – Distances entre les marques du pitch

Comme on l'a décrit dans l'étude bibliographique pour la majorité des pathologies de la voix il y a des altérations au niveau des cordes vocales, et la fermeture de ces dernières devient incomplète. Ainsi il serait déraisonnable de parler de fermeture glottique pour les voix pathologiques et de chercher le nombre de cycles d'ouverture et de fermeture des cordes vocales. C'est cette motivation qui est derrière la re-définition du pitch. On cherchait à définir un pitch qui a un sens pour les voix pathologiques tout en essayant de rester proche de la définition originale du pitch pour les voix normales. La nouvelle définition consiste alors à substituer la formule de calcul du pitch par une autre qui serait plus proche de la réalité surtout pour les voix pathologiques.

En effet au fil de ce travail l'idée qui s'est développée autour du pitch est qu'il serait inapproprié d'associer une fréquence fondamentale à une voix pathologique qui est égale à la moyenne des F_{0i}, et ce à cause des fluctuations notées dans les séquences du pitch chez ces dernières. C'est-à-dire que la moyenne des F_{0i} donne une estimation erronée qui serait à la limite insensée pour ce type de voix.

Comme pour les voix normales les cordes vocales oscillent sous une fréquence presque constante, alors la moyenne des fréquences des tous les cycles ne serait pas très différente de la fréquence modale, c'est à dire la fréquence la plus répandue. C'est de cette manière qu'on définit le nouveau pitch comme étant la fréquence ayant le maximum de nombre d'observation.

$$F_0 = \{F_{0i}^*, telle\ que\ P(F_{0i}^*) = \max(P(F_{0i}))\}$$

Afin de s'assurer qu'on est sur la bonne voix en re-définissant le pitch, on a mis en place une opération simple visant à examiner la différence entre le pitch donné par Praat et le pitch donné par l'algorithme IFG.

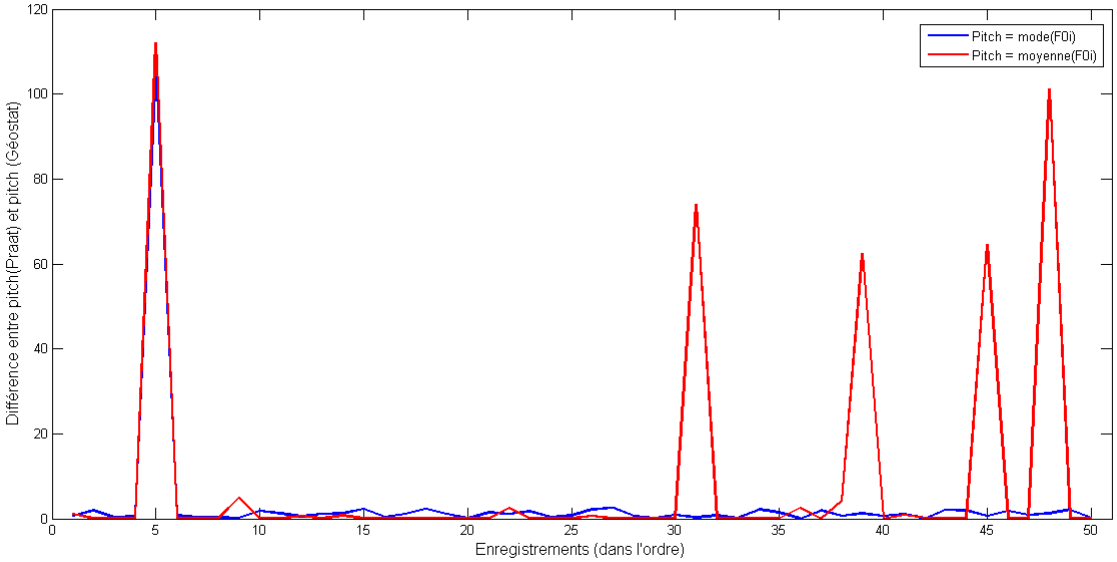


FIGURE 5.3 – Différence du Pitch entre Praat et l'algorithme IFG

Sur la figure 5.3, on calcul $|pitch(Praat) - pitch(Géostat)|$ pour les 50 enregistrements normaux. La courbe bleue représente cette différence lorsque le pitch suit la nouvelle définition. La courbe rouge donne la différence pour la mesure classique du pitch.
Ainsi avec la nouvelle définition on se rapproche encore plus de Praat et les différences avec ce dernier diminuent. En effet, lorsque $F_0 = \frac{1}{N}\sum_{i=1}^{N} F_{0i}$, on remarque la présence de 5 enregistrements pour lesquels les mesures du pitch avec Praat et Géostat sont largement différentes. En passant à la différence avec la nouvelle mesure de pitch, il n'en reste qu'un seul enregistrement. Donc en conclusion l'estimation du pitch s'est améliorée en restant en harmonie avec l'outil de référence.

5.4 Descripteurs audio et résultats

Dans l'étude bibliographique on a évoqué plusieurs mesures acoustiques mais notre travail a porté essentiellement sur les paramètres standards : Jitter, Shimmer et HNR.

5.4.1 Les descripteurs standards

5.4.1.1 Jitter

Présenté dans la section 3.4.3, le jitter permet d'évaluer les fluctuations d'un cycle à l'autre de la fréquence fondamentale. La formule qu'on adopte dans ce projet est la suivante :

$$Jitter\ (\%) = 100.\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|F_{0\,i} - F_{0\,i+1}|}{F_0}$$

et on ne fixe pas de seuil pathologique.

5.4.1.2 Shimmer

Donné également dans la section 3.4.3, le shimmer est l'équivalent du jitter mais en terme d'amplitude.

$$Shimmer(\%) = 100.\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i}$$

où A_i représente l'amplitude crête à crête sur chaque cycle. On précise que dans notre travail on a traité les amplitudes normalisées.
Le calcul du shimmer est fortement lié au calcul du jitter. En fixant un seuil de F_{0i} par cycle, pour tous les autres F_{0i} si elles sont au delà de cette valeur seuil on élimine ce cycle dans le calcul du jitter et on l'élimine par la suite dans le calcul du shimmer.

5.4.1.3 Le HNR

Sachant que dans ce travail on s'est beaucoup référé aux outils avancés d'analyse de la voix, pour le HNR on a utilisé le même algorithme que Praat [17]. Au début, on avait l'intention d'employer cet algorithme directement en modifiant la section du code qui définit la largeur des fenêtres à considérer dans les calculs. Sachant que Praat est développé en C++, ce n'était pas évident d'éffectuer le changement. On s'est référé alors au travail de thèse de A.Tsanas [10] où il a développé le même algorithme que Praat mais sous matlab. Ce dernier nous a alors donné sa fonction de calcul du HNR. Pour nos mesures, on a intervenu au niveau de cet algorithme en changeant l'emplacement temporel des fenêtres considérées dans l'estimation du HNR moyen qui est exprimé en dB. Dans le tableau en annexe C on donne les étapes de calcul du HNR.

5.4.1.4 Résultats et interprétations pour les descripteurs standards

Afin de juger la pertinence des mesures du jitter et shimmer avec les nouveaux MP et la nouvelle définition du pitch, la première étape à laquelle on a songé est de comparer leurs valeurs à celle de Praat et à celle de la base de données de MDVP. Ensuite, on passera à l'évaluation du taux de classification.

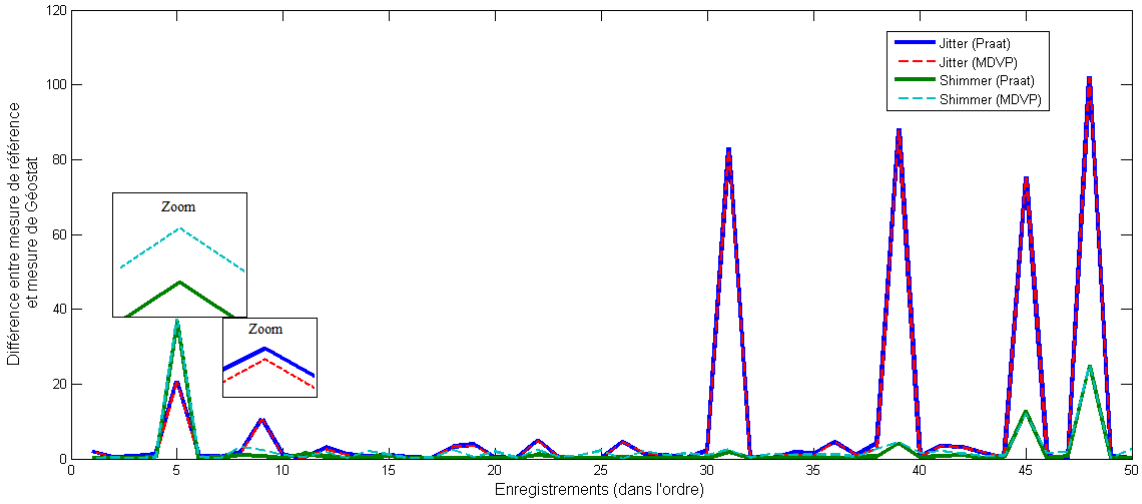


FIGURE 5.4 – Comparaison des mesures de l'algorithme IFG et des outils avancés de mesure pour le jitter et le shimmer

La figure 5.4 donne les résultats de comparaison. En pointillé on a les mesures de différences avec MDVP, le rouge pour le jitter et le bleu clair pour le shimmer. La ligne continue donne les mesures de différences avec Praat, le bleu foncé pour le jitter et le vert pour le shimmer.

Au niveau du jitter, on remarque que nos mesures sont très proches de celles de Praat et de MDVP. Seulement 5 enregistrements normaux présentent de grandes différences. Ces remarques s'appliquent aussi au niveau du shimmer. De plus la différence de mesure du shimmer avec Praat tend plus vers zéro qu'avec MDVP.

Quant au HNR, afin d'évaluer la pertinence de la nouvelle méthode de calcul on se propose tout d'abord de visualiser sa distribution pour les deux classes. Les taux de classification seront donnés plus tard. La figure 5.5 donne en bleu la distribution du HNR pour les voix normales, en rouge pour les voix pathologiques. On voit que chaque classe a son propore pic. En effet la valeur moyenne du HNR obtenue des voix normales est égale à 24,26 dB, celle des voix pathologiques est égale à 21.4 dB. Ces valeurs sont en harmonie

avec les valeurs moyennes du HNR qu'on retrouve dans plusieurs études [22]. En effet, les mesures courantes du HNR sont autour de 20 dB, quand la voix est altérée, elles baissent d'environ 2 dB. Donc déjà, les mesures qu'on obtient correspondent parfaitement aux valeurs typiques, et de plus, les distributions qu'on voit semblent être avantageuses pour les exercices de classification.

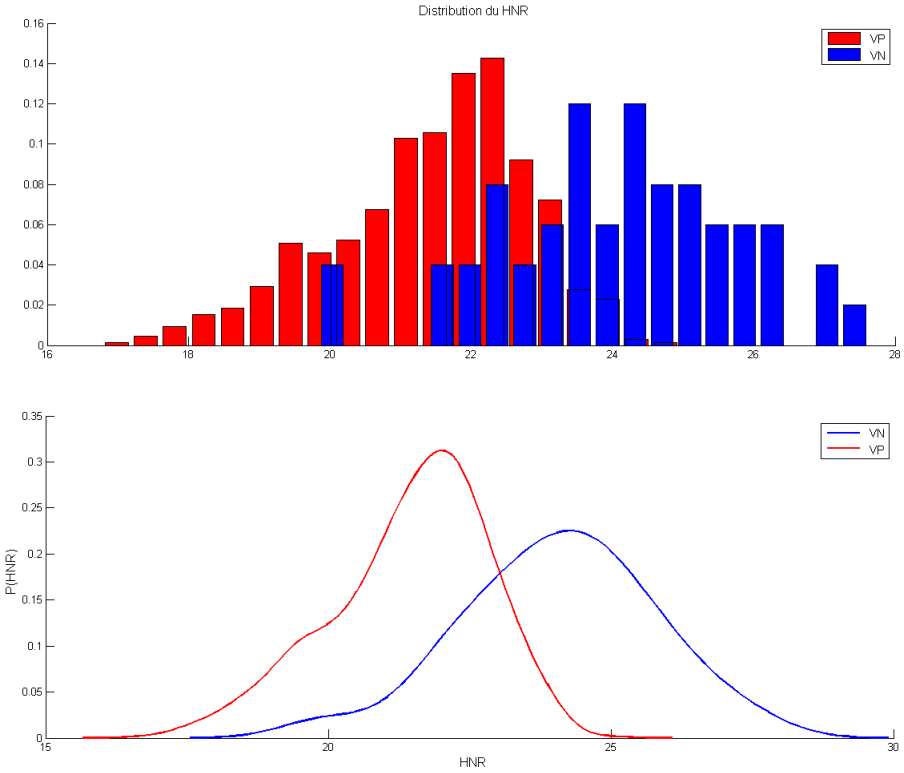


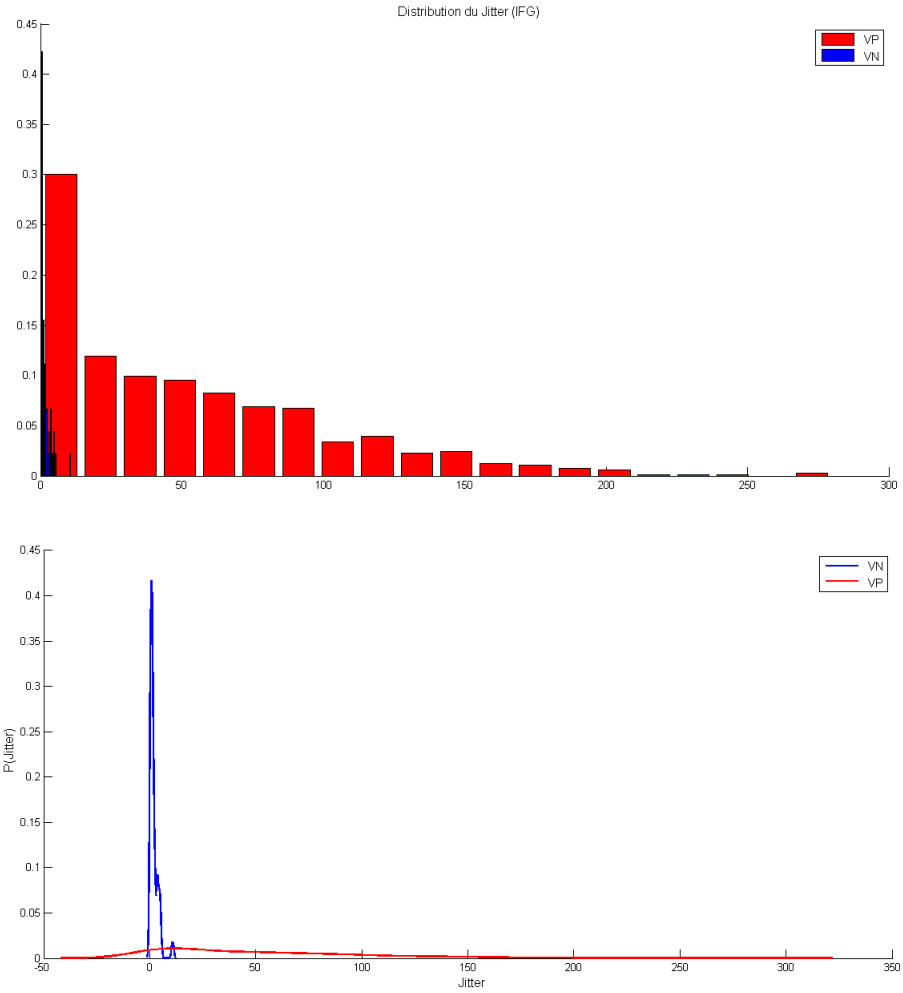
FIGURE 5.5 – Distribution du HNR

Afin de passer à la procédure de classification deux décisions ont été prises :

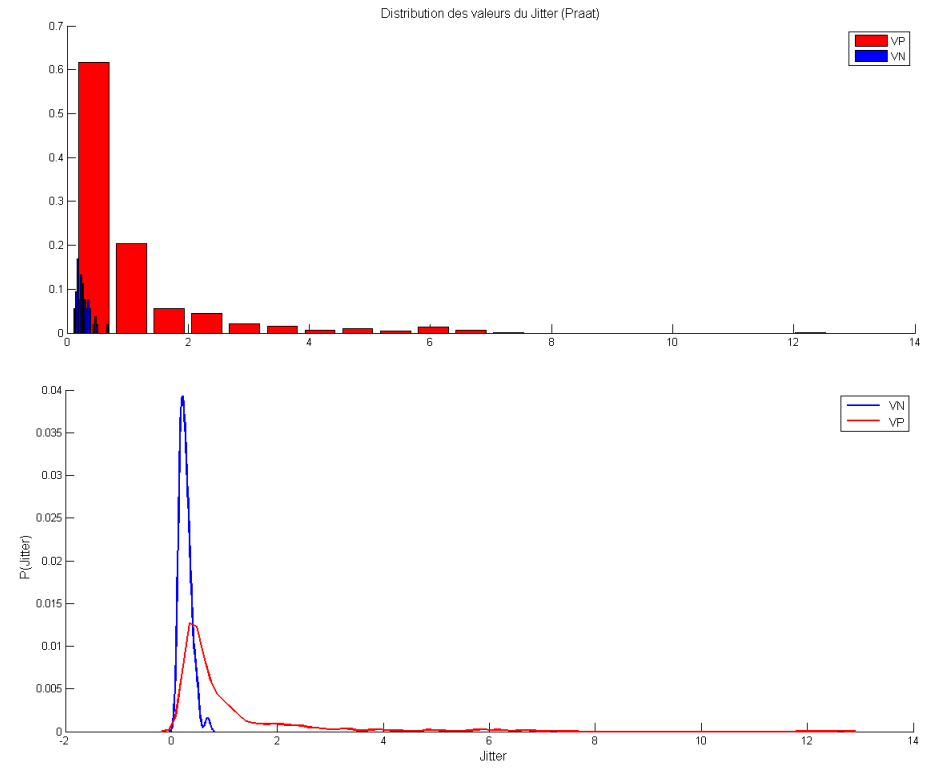
- comme au niveau du shimmer on est plus proche de Praat, alors les comparaisons des taux de classification seront faites par rapport aux taux de classification obtenus par les mesures de ce logiciel.
- éliminer les 5 enregistrements dont le jitter et le shimmer sont très différents de celles de Praat car elles présentent des valeurs aberrantes engendrant plus d'erreur dans l'estimation des classes.

Afin de mieux voir l'effet de l'élimination de ces 5 enregistrements sur l'aire de chevauchement, on donne les distributions du jitter et du shimmer après avoir écarté ces 5 fichiers et on donne parallèlement les distributions de ces descripteurs calculés par Praat.

La figure 5.6 montre la réduction de l'aire de chevauchement avec l'algorithme IFG.

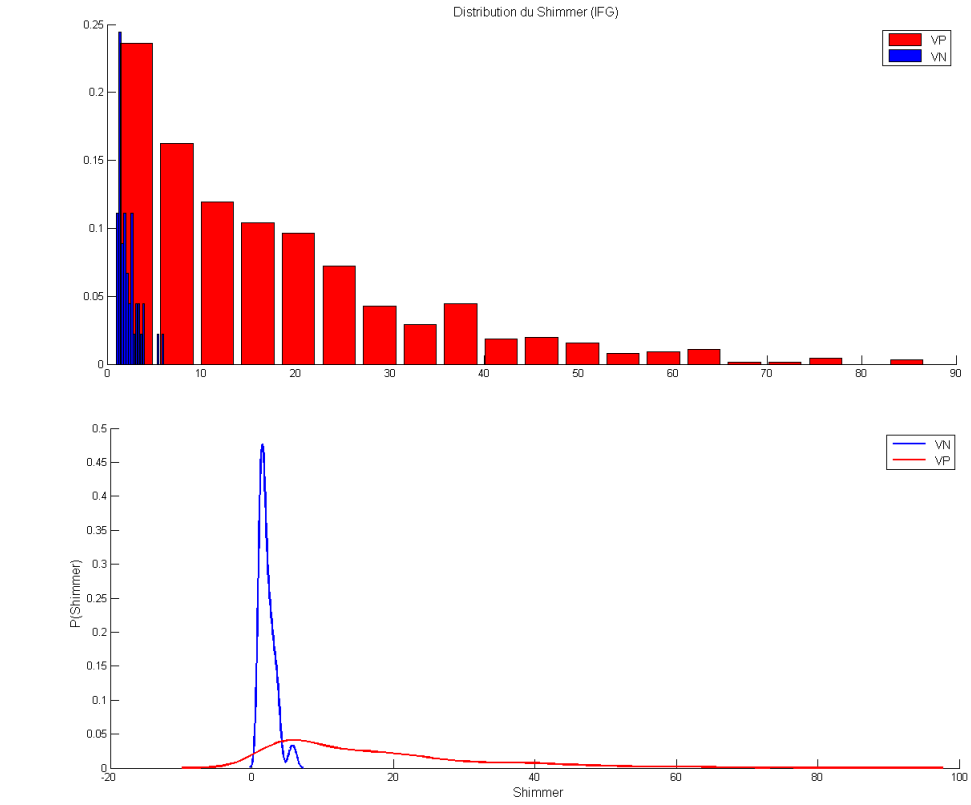


(a)

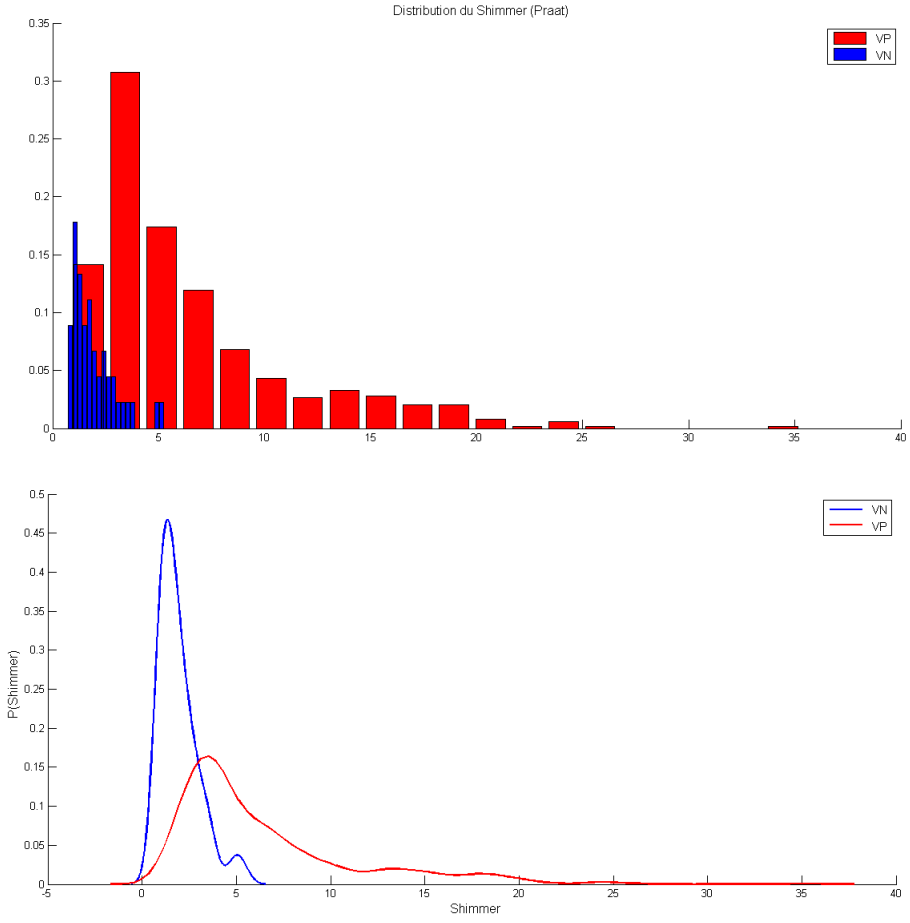


(b)

FIGURE 5.6 – Distributions du Jitter, (a) Alg. IFG, (b) Praat



(a)



(b)

FIGURE 5.7 – Distributions du Shimmer, (a) Alg. IFG, (b) Praat

On remarque de plus l'augmentation de la plage des valeurs du jitter pour les VP. Ceci permettra de mieux destinguer cette classe de voix. L'apport de notre approche de calcul serait plus visible en passant à la classification.

Quant au shimmer dont les distributions sont données dans la figure 5.7, on arrive à réduire l'aire de chevauchement. La distribution du shimmer des voix pathologiques devient aplatie avec l'algorithme IFG ce qui renseigne sur la variabilité des changements de ce descripteur pour les VP. La figure 5.7 montre aussi qu'avec les mesures de l'algorithme IFG, on est capable de mieux classer les voix normales.

Les figures 5.7 et 5.6 forment une première approche d'évaluation des nouvelles mesures données par l'algorithme IFG. On montre par la suite le tableau de classification en éliminant les 5 enregistrements : dans la table 5.1 on donne mes résultats obtenus par l'algorithme IFG et on le compare au tableau de classification obtenu par les mesures de Praat : table 5.2.

Descripteur	PP (%)	NN (%)	Overall (%)
Jitter	74.54	97.78	76.04
Shimmer	78.22	95.57	79.34
HNR	90.95	71.11	89.67

TABLE 5.1 – Résultats de classification avec Jitter, Shimmer et HNR (Géostat)

Descripteur	PP (%)	NN (%)	Overall (%)
Jitter	67.55	88.88	68.94
Shimmer	79.19	86.67	79.68
HNR	77.95	94.55	79.1

TABLE 5.2 – Résultats de classification avec Jitter, Shimmer et HNR (Praat)

C'est ainsi que les nouvelles mesures obtenues dans Géostat permettent d'avoir de meilleurs taux de classification au niveau du jitter. Les PP et NN de Géostat pour le jitter dépassent celles de Praat. Quant au shimmer, au niveau du PP, on a presque le même taux. En terme de bonne classification de voix normales donnée par NN, on est bien meilleur. Ces résultats correspondent aux premières conjectures efféctuées grâce à l'analyse des distributions.

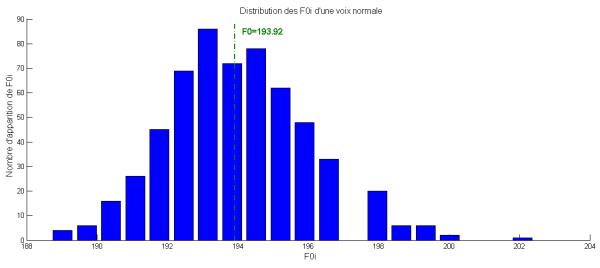
Au niveau du HNR, il est vrai que la performance de Praat pour le bon classement des voix saines dépasse celle de l'algorithme IFG, mais en s'intéressant aux taux de classification total *overall* et à la bonne classification des voix pathologiques PP, on est non seulement

meilleur, mais en plus le $PP = 90.95\%$ est un résultat très intéressant. Son exploitation pour augmenter les performances de la bonne classification des voix pathologiques serait probablement d'un grand apport.

5.4.2 Les descripteurs de distributions

La visualisation des distributions des séquences du pitch laisse réfléchir à l'étude de la dispersion des F_{0i} autour de la fréquence fondamentale. En effet, les distributions des F_{0i} ont aussi joué un rôle curcial dans les calculs et l'ajustement des paramètres acoustiques. A chaque échantillon de voix une unique distribution de F_{0i} est associée. Il est certainement évident que la répartition des F_{0i} d'une VN est différente d'une VP car généralement une personne souffrant d'une pathologie de la voix présente de grandes fluctuations dans le pitch par conséquent la séquence du pitch occupe une large plage de valeurs. Contrairement à une personne dont la voix est saine, la séquence du pitch devrait être presque constante variant entre des valeurs très proches. La figure 5.8 donne en (a) un exemple de distribution de la séquence du pitch pour une voix féminine normale,on remarque alors que les F_{0i} varient aux alentours de F_0. Sur la figure (b) correspondant également à une voix féminine mais pathologique on note un $F_0 = 260.79Hz$ alors que les F_{0i} varient de 100 Hz allant jusqu'à 500 Hz.

(a)



(b)

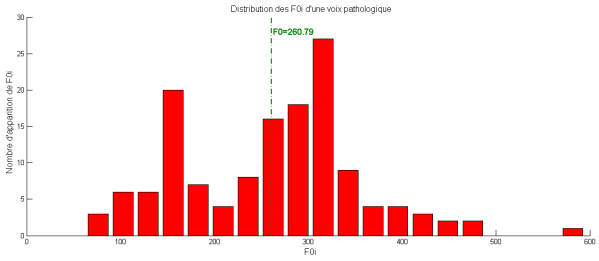


FIGURE 5.8 – (a) Distribution des F0i d'une VN (CEB1NAL) (b) Distribution des F0i d'une VP (BRT18AN)

C'est ainsi qu'on introduit les mesures des moments statistiques centrés d'ordre k tel que

$k = 1, 2, 3, 4$. Leurs mesures se font en substituant la moyenne des F_{0i} par le mode des F_{0i} c'est-à-dire en appliquant la nouvelle définition du pitch.

5.4.2.1 L'écart moyen

En manipulant les distribution des F_{0i}, le premier paramètre auquel on pense est le plus simple des paramètres de mesure de dispersion donnant la moyenne des écarts des F_{0i} par rapport au pitch.

$$EM_{F_0} = \frac{1}{N}\sum_{i=1}^{N}(F_{0i} - F_0)$$

Considérer ce paramètre est un premier pas pour évaluer la manière dont les F_{0i} s'écartent du pitch.

5.4.2.2 La déviation standard

On passe à la déviation standard car elle présente l'avantage d'être toujours positive ce qui permettra une meilleure comparaison entre VN et VP. Etant donné que la déviation standard est donnée par l'équation suivante :

$$\sigma_{F_0} = \sqrt{\frac{1}{N}.\sum_{i=1}^{N}(F_{0i} - F_0)^2}$$

et en tenant compte du fait que pour les VN la séquence des pitch est presque constante, alors le résultat auquel on s'attend en évaluant la dispersion autour du pitch est d'avoir des déviations standard par rapport à la moyenne qui sont relativement plus petites que celles des VP.

5.4.2.3 Le skewness

Premier paramètre de forme en statistique et appelé aussi coefficient de dissymétrie, le skeweness était notre moyen pour évaluer le comportement des F_{0i} en terme de décalage ou étalement par rapport à la valeur moyenne.
Le skewness donné comme de cette manière :

$$S_{F_0} = \frac{\frac{1}{N}.\sum_{i=1}^{N}(F_{0i} - F_0)^3}{\sigma_{F_0}^3}$$

Il peut être négatif pour une distribution décalée à gauche c'est-à-dire que les F_{0i} sont inférieures au pitch, positif pour des F_{0i} supérieures à ce dernier et nul ou dans notre cas on dit qu'il est autour de zéro pour des F_{0i} proches du mode, c'est bien à ce résultat qu'on s'attend pour les voix normales.

5.4.2.4 Le kurtosis

Deuxième paramètre d'évaluation de forme et appelé aussi coefficient d'applatissement. Sa formule est donnée comme suit :

$$K_{F_{0i}} = \frac{\frac{1}{N} \cdot \sum_{i=1}^{N} (F_{0i} - F_0)^4}{\sigma_{F_0}^4}$$

On a tenu compte de cette mesure afin d'évaluer le regroupement des F_{0i} autour de F_0. Les $K_{F_{0i}}$ très élevés correspondent plutôt à des échantillons de VN car de telles valeurs indiquent que la distribution des F_{0i} est pointue donnant ainsi une séquence de pitch dont les valeurs sont autour de F_0. Le cas contraire caractérisera les VP

5.4.2.5 Résultats et interprétations pour les descripteurs de distributions

L'évaluation du pouvoir descriptif des voix pathologiques de ces mesures se fait en termes de taux de classification. Comme le calcul de ces paramètres statistiques est fortement corrélé à la présentation des histogrammes des données (dans notre cas : l'histogramme des F_{0i} par enregistrement), au fil de l'avancement du travail on s'est proposé de calculer ces paramètres de deux manières :

1. la première méthode consiste à représenter l'histogramme de chaque enregistrement en allant de sa F_{0i} minimale à sa F_{0i} maximale. Ceci se fait indépendamment des autres enregistrements. Ensuite on fixe le nombre de bars de l'histogramme à une valeur par défaut pour tous les enregistrements. Ici on s'est proposé le nombre $N_c = 20$, c'est-à-dire que l'histogramme de chaque enregistrement sera formé de 20 bars.

   On commence par évaluer les distributions des descripteurs pour estimer leur pouvoir discriminant.

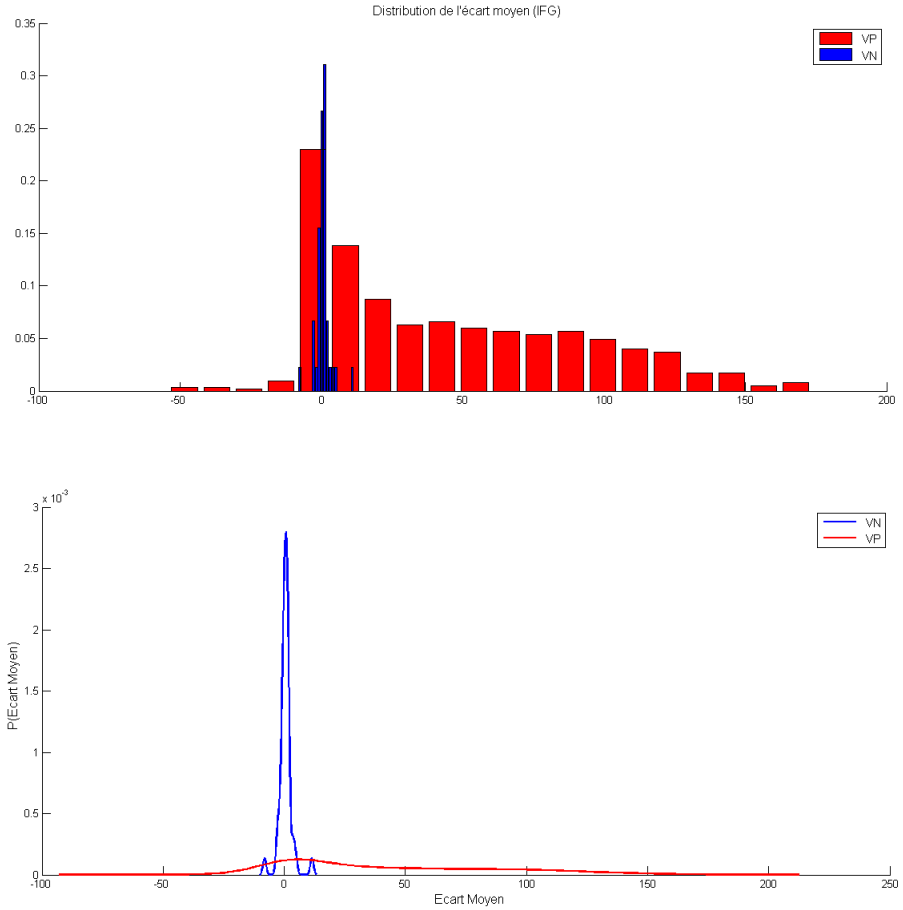


FIGURE 5.9 – Distribution de l'écart moyen par rapport au mode (nouveau pitch) avec choix de N_c

La figure 5.9 donne alors la distribution de l'écart moyen des F_{0i} par rapport au pitch. En effet l'allure obtenue montre un faible pouvoir discriminant de l'écart moyen. On peut même conclure à partir de ce stade que l'utilisation de ce descripteur ne peut pas être retenue pour les prochaines étapes. Ceci sera validé à partir des résultats de la classification.

Quant à la déviation standard dont les distributions sont données sur la figure 5.10, malgré l'élimination des fichiers engendrant des valeurs aberrantes il y a toujours la large surface de chevauchement. Dans ce cas la visualisation de la distribution n'est pas suffisante pour évaluer le caractère discriminant de ce descripteur. Le tableau 5.3 donne les résultats de classification qui permettront une évaluation pertinente. Ensuite, sur la figure 5.11 il y a la distribution du skewness qui semble être la plus prometteuse parmi toutes les distributions déjà analysées.

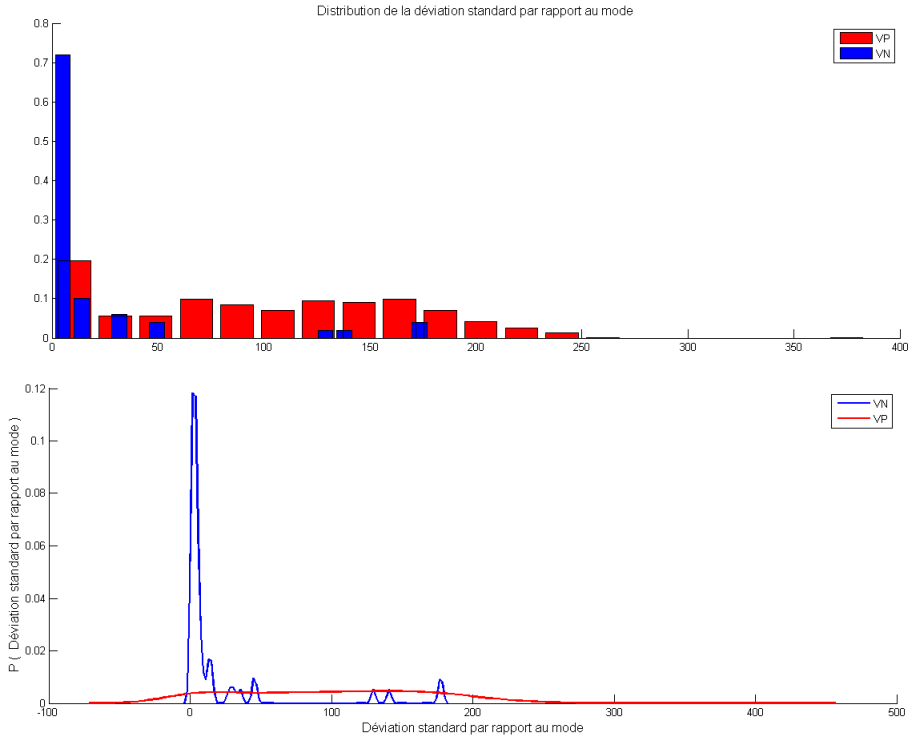


FIGURE 5.10 – Distribution de la déviation standard par rapport au mode (nouveau pitch) avec choix de N_c

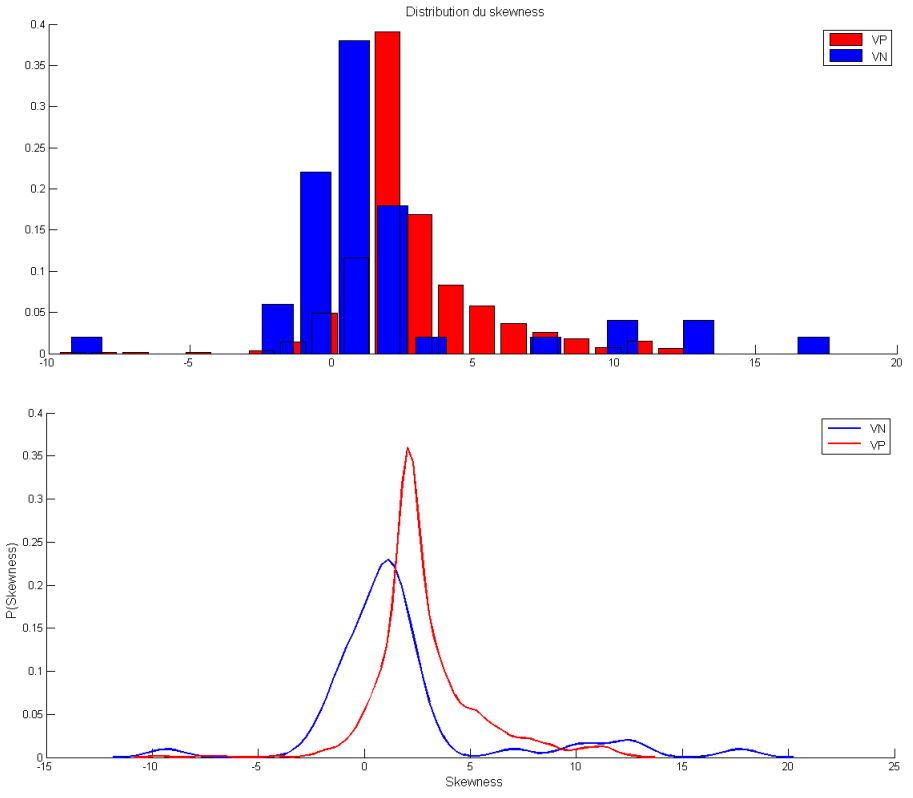


FIGURE 5.11 – Distribution du Skewness avec choix de N_c

Ce qui est intéressant au niveau de ce descripteur c'est qu'on voit clairement la différence des pics des deux classes. Les voix normales sont plus centrées autour d'un skewness nul. C'est le résultat qu'on voulait atteindre dès qu'on a commencé à étudier ce descripteur.

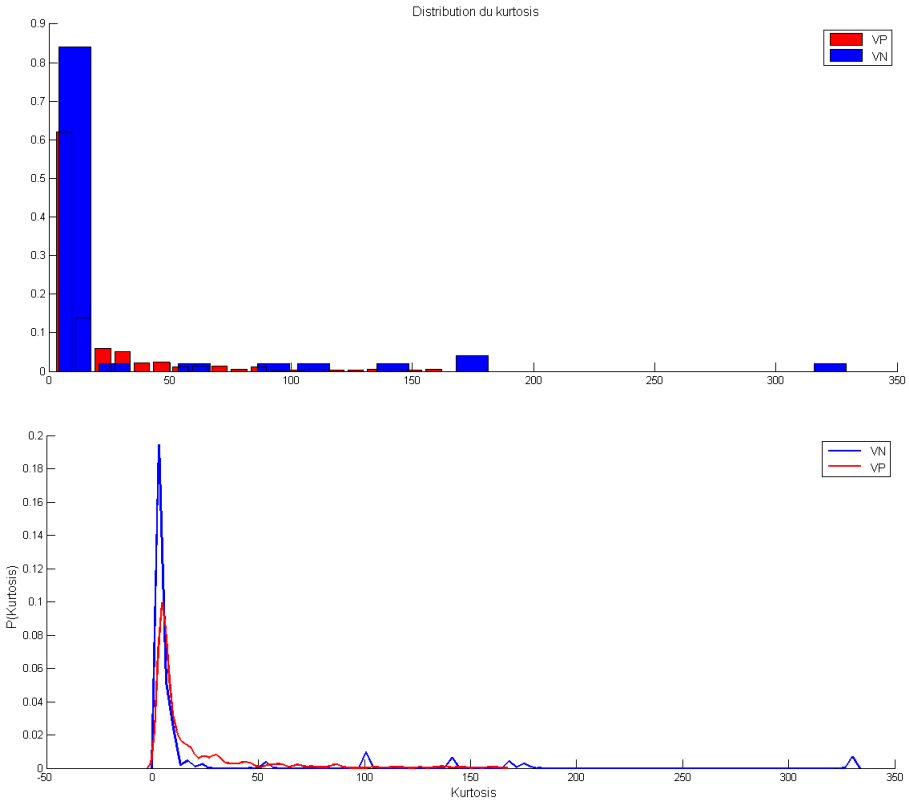


FIGURE 5.12 – Distribution du Kurtosis avec choix de N_c

Le dernier descripteur évalué est le kurtosis. Sa distribution sur la figure 5.12 met en évidence des pics des deux distributions pour des valeurs de kurtosis très rapprochées avec une large surface de chevauchement. A priori le kurtosis présente de faibles performances en terme de caractérisation des voix. On continue à évaluer ceci par la classification.

On expose alors sur tableau 5.3 les résultats de classification : On a alors des résultats qui sont en harmonie avec les analyses des distributions. On remarque les faibles taux de classification obtenus par l'écart moyen et le kurtosis, c'est pour cela qu'on décide d'éliminer ces descripteurs dans la suite du travail.

2. la deuxième alternative de mesure consiste à fixer une valeur de F_{0i} minimale et de F_{0i} maximale pour tous les enregistrements. Pour ce faire, on parcourt tous les

Descripteur	PP (%)	NN (%)	Overall (%)
EM_{F_0}	46.47	64.44	47.63
σ_{F_0}	75.46	95.55	76.76
S_{F_0}	75.15	82.22	75.6
$K_{F_{0i}}$	36.04	82.22	39.02

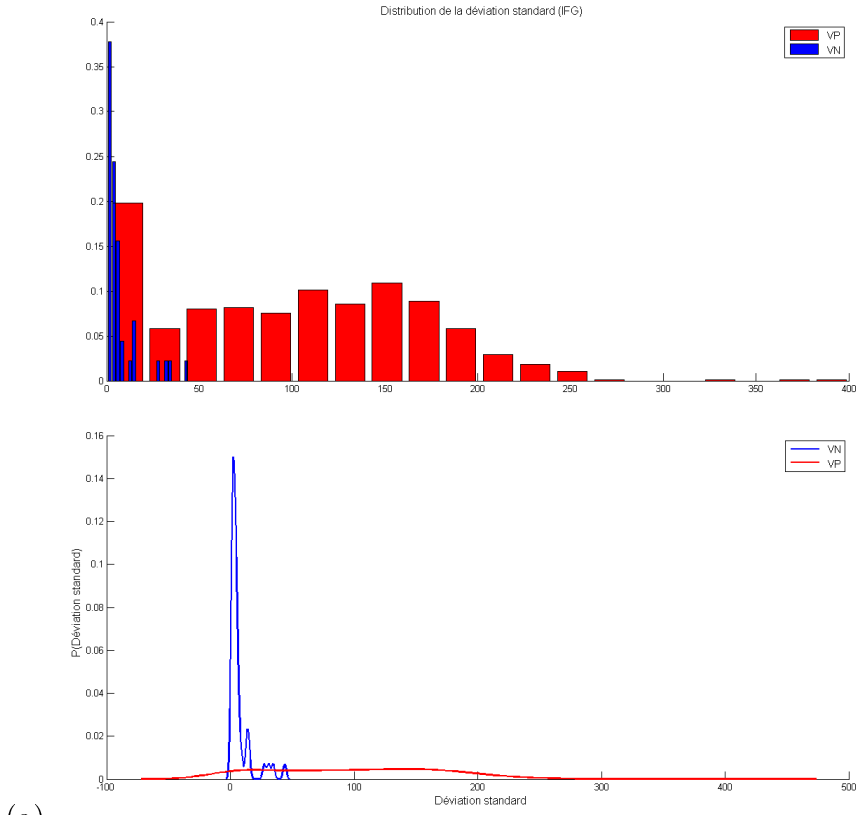
TABLE 5.3 – Résultats de classification avec les descripteurs de distribution

enregistrements en vérifiant pour chacun la plus petite F_{0i} et la plus grande, ensuite on compare ces dernières entre elles. Ainsi la représentation des histogrammes s'est faite entre 40 Hz et 5000 Hz. Sachant que le fait d'avoir $F_{0i} = 5000Hz$ est un évènement très rare qui peut nous induire en erreur, on décide alors de lancer une batterie de test afin de fixer la bonne F_{0i} supérieure qu'on appelle F_t. Ainsi en parcourant une plage de F_t sous laquelle on parcourt une plage du nombre de bars de l'histogramme et en testant à chaque fois les taux de classifcation, on aboutit aux meilleurs taux avec les paramètres suivants : $F_t = 1000Hz$, $N_c = 380$.

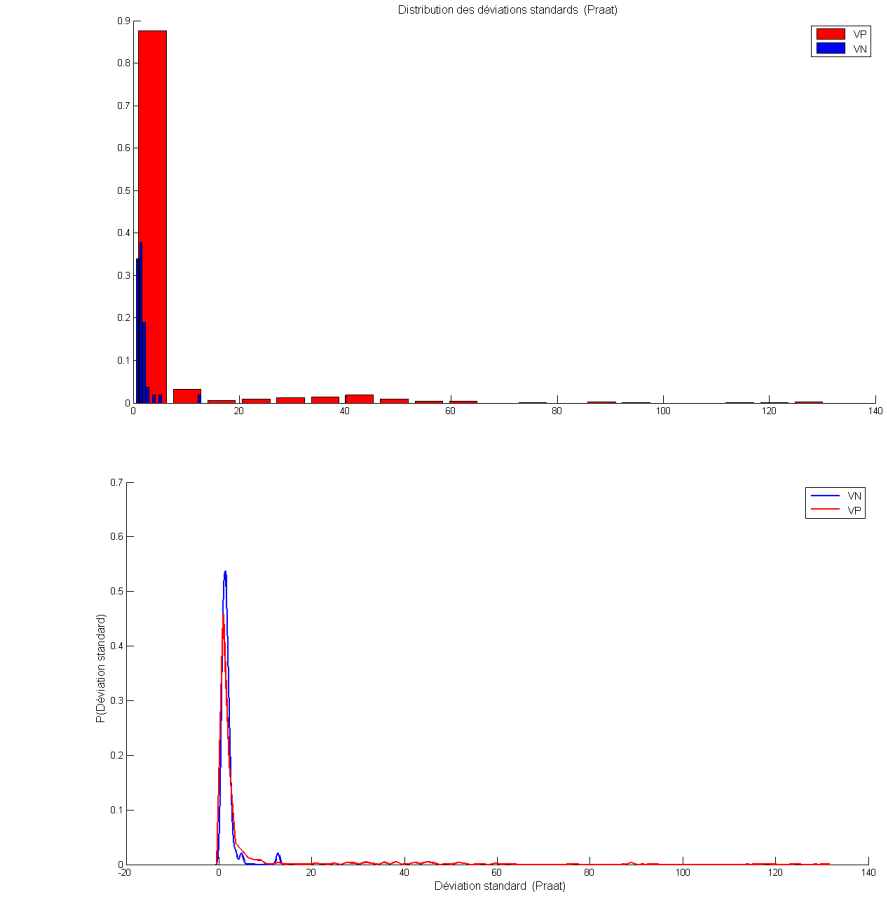
On montre alors les nouvelles distributions pour les descripteurs qu'on a décidé de garder : la déviation standard et le skewness. La figure 5.13 (a) donne la distribution de la nouvelle mesure de la déviation standard, la figure 5.13 (b) donne celle obtenue par praat. La représentation des densités de probabilités permet de dégager des remarques plus pertinentes comme on l'a déjà dit.

La nouvelle mesure de la déviation standard donne des des distributions qui sont presque identiques aux distributions obtenues avec l'ancienne méthode (celle présentée en premier). L'apport de cette nouvelle mesure se voit en comparant la distribution à celle de Praat. Avec Praat, les VN et les VP ont des pics très rapprochés qui seront probablement à l'origine d'un grand taux d'erreur dans l'exercice de classification. On remarque que ce n'est pas le cas dans la figure 5.13 (a) où on donne la densité de probabilité et où la distribution des VP devient plate. Ceci veut dire que la dispersion des F_{0i} des voix pathologiques autour de la fréquence fondamentale varie dans tous les sens. Ce qui est bien réel car chez ces derniers, les variations des F_{0i} sont instables.

On fait pareil pour le skewness. On a alors la figure 5.14 sous forme comparative donnant en (a) le skewness de l'algorithme IFG et en (b) le skewness de Praat.

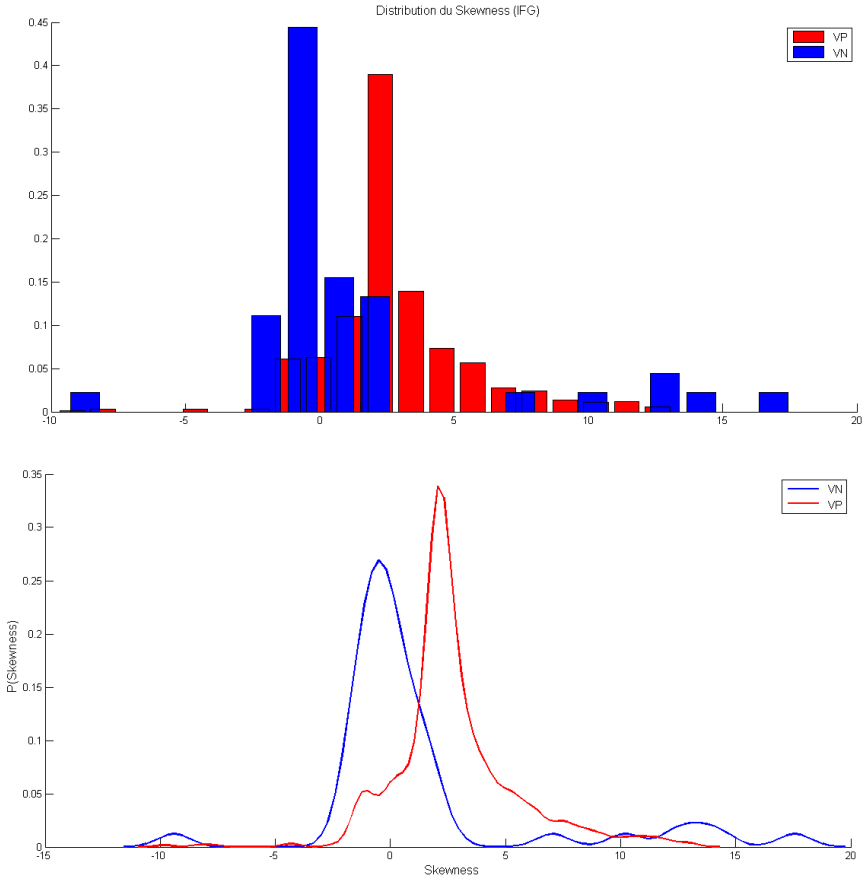


(a)

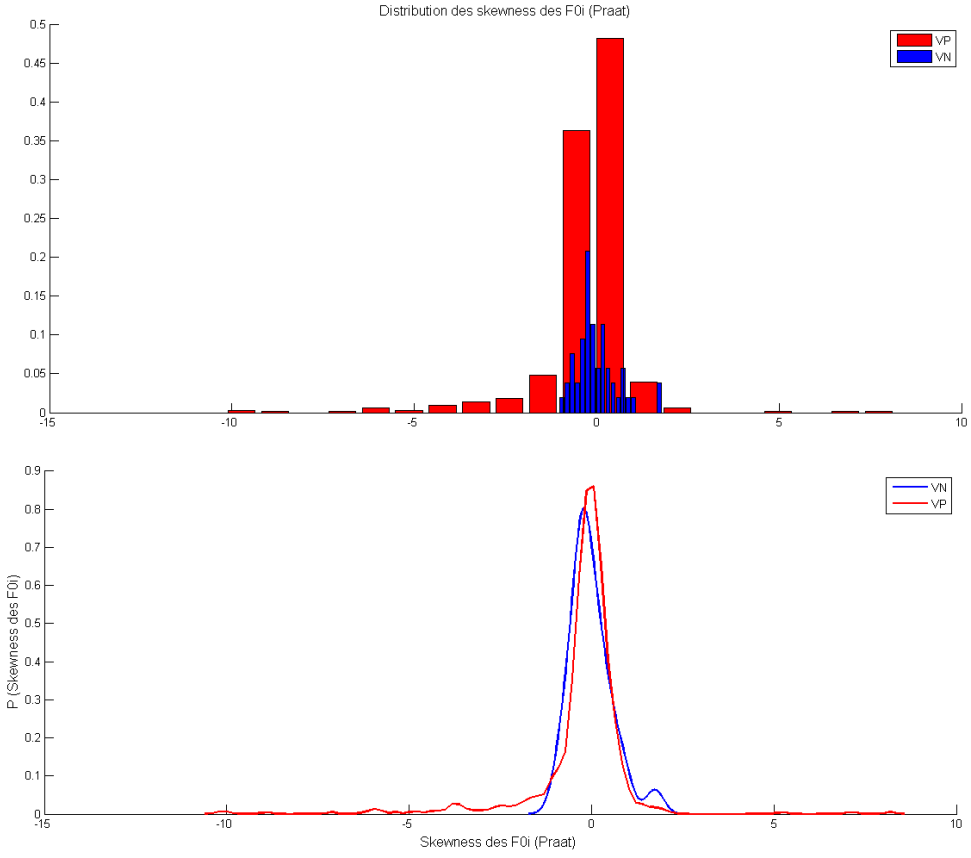


(b)

FIGURE 5.13 – Distributions des mesures de la déviation standard avec fixation de F_t et N_c, (a) Alg. IFG, (b) Praat



(a)



(b)

FIGURE 5.14 – Distributions des mesures du Skewness avec fixation de F_t et N_c, (a) Alg. IFG, (b) Praat

Il est vrai que les distributions obtenues avec l'algorithme IFG présentent une surface de chevauchement entre VN et VP plus large que celle obtenue avec Praat. Notre mesure est avantageuse au niveau de la séparation des pics des deux classes. Avec notre nouvelle mesure, on est loin de confondre les pics des VN et VP. Ces interprétations seront mieux évaluées en testantnt le pouvoir discriminant des mesures. On présente alors dans le tableau 5.4 les résultats de classification pour les mesures de l'algorithme IFG.

Descripteur	PP (%)	NN (%)	Overall (%)
σ_{F_0}	75.62	95.56	76.9
S_{F_0}	81.44	80	81.35

TABLE 5.4 – Résultats de classification avec les descripteurs de distribution avec fixation de F_t et N_c

Comme on a comparé les distributions de ces nouvelles mesures à celles de praat, on compare aussi les résultats de classification à Praat.

Descripteur	PP (%)	NN (%)	Overall (%)
σ_{F_0}	6.52	97.78	12.48
S_{F_0}	16.77	86.67	21.34

TABLE 5.5 – Résultats de classification avec les descripteurs de distribution de Praat

En conséquence, l'introduction des paramètres d'évaluation des distributions a permis de gagner sur deux plans :

- on obtient des taux de classification qui dépassent ceux de Praat pour les mêmes descripteurs.
- on atteint le meilleur taux de classification en terme de PP avec le skewness. Son PP devance le PP du jitter et du shimmer. Ceci permet de conclure que ce descripteur caractérise les voix avec plus de pertinence rendant possible une meilleure séparation entre VN et VP.

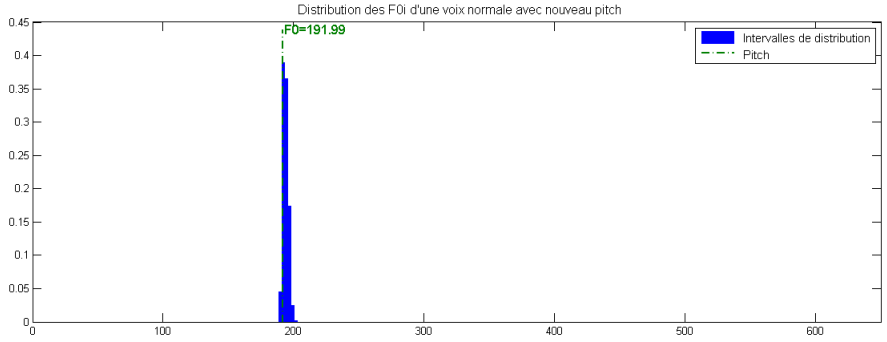
Ainsi, remarquant que le skewness donne de bon résultat et sachant que ce dernier est à l'origine de l'analyse de l'asymétrie des distributions autour d'une valeur donnée, on se propose alors d'aller plus loin que le skewness en entrant plus dans les détails des distributions et leurs proportions. C'est ce qu'on introduit dans la section suivante.

5.4.3 Les nouveaux paramètres de proportions

L'analyse du skewness et la visualisation des distributions des séquences du pitch laisse réfléchir à la prise en considération des proportions des F_{0i} aux alentours du mode. La figure 5.15 montre clairement la différence des distributions entre voix normale et voix pathologique.

En effet pour une voix normale toutes les classes [2] rassemblant les F_{0i} sont autour du pitch et un autre point très important qu'on signale est que le nombre de classes obtenu pour donner la distribution des F_{0i} d'une voix normale est inférieur à celui obtenu pour les voix pathologiques. Ainsi, pour les voix normales, plus on s'éloigne de la classe modale et plus la proportion (nombre d'éléments par classe) tend vers 0. Nos paramètres qu'on définit dans cette section repose sur ce principe. On donne l'exemple sur la figure 5.15 où dans la figure (a) on remarque que la proportion de la classe de centre 100 est carrément nulle alors que pour la voix pathologique donnée sur la figure (b) la proportion de cette même classe est égale à 0.0214. On précise que ces proportions sont normalisées par rapport au nombre total des F_{0i}.

(a)



(b)

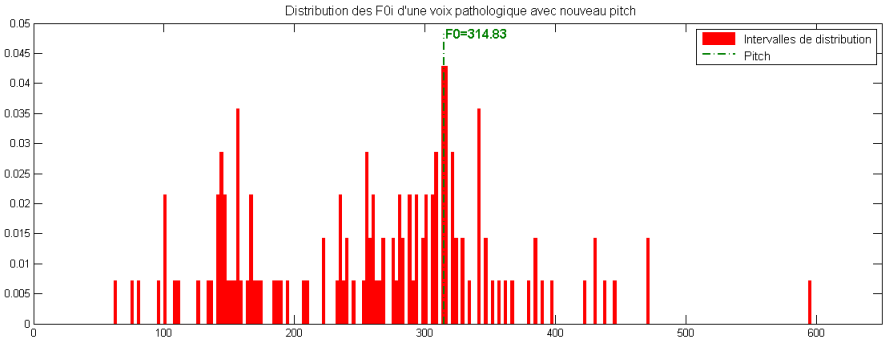


FIGURE 5.15 – Distributions des F_{0i} pour l'analyse des proportions

2. Dans ce paragraphe on définit une classe comme étant une bar de l'histogramme.

On donne alors les 3 nouveaux paramètres de proportion en définissant tout d'abord le paramètre α comme étant le nombre de classes à considérer autour de la classe modale. En tenant compte de la symétrie des distributions, ces paramètres dépenderont de $-\alpha$ et α. P est la proportion normalisée :

$$P = \frac{card(classe)}{nombre\,total\,des\,F_{0i}}$$

On mentionne que l'analyse du skewness était à l'origine de la considération de la symétrie dans nos calculs.

– Différence de proportion de part et d'autre de la classe modale :

$$D = |P_{-\alpha} - P_{\alpha}|$$

– Rapport des proportions :

$$R = \frac{P_{-\alpha}}{P_{\alpha}}$$

– Rapport entre la proportion maximale et la proportion minimale :

$$Rmm = \frac{\max(P_{-\alpha}\,,\,P_{\alpha})}{\min(P_{-\alpha}\,,\,P_{\alpha})}$$

Ayant obtenus ces mesures on passe par la suite à l'évaluation de leurs pouvoirs discriminant. Comme on l'a présenté pour des résultats antécédants l'évaluation se fait en termes de PP, NN et $Overall$.

Descripteur	PP (%)	NN (%)	Overall (%)
D	74.69	71.11	74.46
R	91.26	28.89	87.23
Rmm	81.59	46.67	79.34

TABLE 5.6 – Résultats de classification avec les descripteurs de proportions

En analysant le tableau 5.6 on remarque que le paramètre D a un potentiel de description des caractéristiques de la voix presque similaire aux descripteurs standards.
En effet ce descripteur correspond à une caractérisation similaire à l'étendue vocale mais sur une plage de fréquence définie autour du pitch et non pas sur toute la séquence du pitch. Quant aux descripteurs R et Rmm ils ont permis de bien classer les voix pathologiques, mais leur défaut majeur consiste à mal classer les voix normales en donnant des

mesures pareilles aux mesures affectées aux voix pathologiques entrainant ainsi le classifieur à traiter les voix normales comme étant pathologiques.
Malgré ce déséquilibre dans les taux de classification, on ne rejette pas ces mesures. Au contraire, il faudrait exploiter les PP élevés. C'est pour cela qu'on a pensé à faire la classification en combinant 2 descripteurs.

5.4.4 Résultats avec combinaisons binaires de descripteurs

Afin de sélectionner les descripteurs audio à combiner, il serait intéréssant de cultiver le choix des combinaisons par la signification physiologique de ces descripteurs. On s'est fixé un deuxième critère de choix de combinaison de descripteurs : c'est le taux de bon classement des voix pathologiques. En effet l'un des buts qu'on voudlait atteindre au fur et à mesure qu'on avancait dans le travail est l'amélioration de PP. Jusqu'à maintenant le meilleur PP obtenu est au alentour de 81%. On a décidé alors d'assigner un seuil de pourcentage en PP pour les combinaisons, c'est-à-dire toute combinaison ayant un taux de $PP < 80\%$ est à éliminer. C'est ainsi qu'on présente le tableau 5.7 donnant des combinaisons avec des bons taux de discrimination entre VN et VP.

Descripteurs	PP (%)	NN (%)	Overall (%)
Jitter, Shimmer	80.38	95.56	81.35
Shimmer, S_{F_0}	80.38	93.33	81.2
Shimmer, D	80.36	95.55	81.34
Shimmer, R	80.06	91.1	80.77
Shimmer, Rmm	81.44	91.11	82.06
S_{F_0}, R	80.21	77.78	80.06
Rmm, R	84.82	35.55	81.64
HNR, Jitter	81.14	88.89	81.64
HNR, Shimmer	82.36	82.22	82.35
HNR, S_{F_0}	84.81	80.01	84.05
HNR, D	86.8	77.78	86.23
HNR, R	91.26	73.33	90.1
HNR, Rmm	91.87	71.11	90.53

TABLE 5.7 – Résultats de classification avec combinaisons binaires de descripteurs

Le choix de la combinaison du jitter avec le shimmer est une association classique mettant en œuvre la variation de l'amplitude et les fluctuations de la fréquence fondamentale. Il s'agit alors d'une combinaison faisant appel à des descripteurs complémentaires. C'est sur principe là que les autres combinaisons ont été choisies. La fusion du shimmer avec

respectivement le S_{F_0}, D, R et Rmm évalue les fluctuations de l'amplitude avec un aspect de fluctuations fréquentielles car S_{F_0}, D, R et Rmm reflètent le comportement des distributions des F_{0i} autour du pitch.
Quant à la combinanison du skewness avec le paramètre R, et le Rmm avec R, elles exploitent des mesures très corrélées, c'est-à-dire que toutes les deux peuvent évaluer le même aspect d'une voix, cela dit, on les teste quand même mais on voit que le NN de la fusion du S_{F_0} et R est relativement faible par rapport aux autres NN obtenus, aussi on remarque ce résultat au niveau de la combinaison de Rmm et R.
Au niveau des combinaisons avec le HNR, on commence par les combinaisons classiques, HNR/Jitter et HNR/Shimmer. Elles ont donné des résultats équilibrés en PP et NN. Ensuite, on remarque que plus on emploie les paramètres explorant les distributions des F_{0i} avec le HNR, plus les résultats s'améliorent, spécialement en terme de PP.
Donc les voix pathologiques sont plus pertinemment mises en évidence en les caractérisant par la richesse harmonique et la richesse fréquentielle autour du pitch.

5.5 Conclusion

Au travers ce chapitre on a montré non seulement l'éfficacité des MP détectées par l'algorithme de Géostat mais aussi la pertinence de la nouvelle définition du pitch.
Les résultats de classification obtenus avec les paramètres standards dépassent ceux obtenus par le logiciel avancé d'analyse de la voix Praat. L'exploitation des mesures évaluant les distributions des F_{0i} a donné des résultats avantageux qui ont mené à mettre en place de nouvelles mesures. Ces mesures donnent une nouvelle estimation de l'étendue vocale selon les proportions des F_{0i}. Il est vrai que ces mesures, seules, ont donné des résultats de classification déséquilibrés entre estimation de classe de VN et classe de VP mais elles ont donné des résultats prometteurs en les combinant avec les autres descripteurs audio. On a aussi montré que l'évaluation de l'harmonicité des voix est importante pour caractériser les voix pathologiques.

Conclusion Générale

Le but du projet consistait à modifier les descripteurs audio standards des voix pathologiques de manière à ce qu'ils permettent de mieux décrire les voix pour discriminer entre voix pathologique et voix normale. Ensuite, le deuxième objectif visait à mettre en place de nouvelles mesures acoustiques dont la performance en terme de classification apporte un plus, c'est-à-dire permet d'augmenter les taux de bonne classification.
Pour ce faire, on a commencé par une étude de l'état de l'art dans un contexte médical afin d'identifier les principaux symptômes d'un trouble de la voix. Dans la même étape, une étude de l'état de l'art dans un contexte scientifique mathématique a été faite. Son objectif est d'identifier les principales approches suivies dans le domaine de l'identification des pathologies de la voix par le traitement de la parole.
Sachant qu'il existe deux manières de traitement de signal pour parvenir à l'identification des voix pathologiques, qui incluent le traitement linéaire et non linéaire, l'objectif fixé au fur et à mesure de l'avancement du travail est d'améliorer les taux de classification obtenus par les descripteurs standards.
Alors on a choisi de considérer 3 descripteurs standards : le jitter, shimmer et HNR. Les mesures ont été basées sur l'estimation des marques du pitch (MP) qui correspondent aux instants de fermeture glottale (IFG). Le second pas élaboré et qui a été d'un grand apport est la re-définition du pitch, correspondant par défaut à la moyenne des F_{0i} de la séquence du pitch, on s'est proposé de le définir comme étant égal à la F_{0i} qui a la plus grande fréquence d'apparition dans une séquence du pitch. Ainsi cette nouvelle définiton a permis de donner des valeurs du pitch pour les voix normales qui sont presque pareilles à notre outil de référence qui est le logiciel Praat : outil couramment employé pour l'analyse objective de la voix.
Au fur et à mesure qu'on avancait dans le travail, de nouvelles idées se sont développées et nous ont fait réfléchir à considérér les distributions des F_{0i} et leurs proportions.

En effet, dans les études antérieures, toutes les mesures qui ont été utilisées exploitaient les F_{0i} en eux mêmes. L'innovation apportée par ce projet est le fait de chercher à dégager des descripteurs pertinents mettant en œuvre les distributions de la séquence du pitch de part et d'autre de la fréquence fondamentale qu'on définit de notre propre manière.
Ainsi nos nouvelles mesures ont permis de donner des résultats considérables, surtout en terme de classification des voix pathologiques comme étant pathologiques. Il est vrai que ces nouveaux descripteurs ont un rendement faible en classant les voix normales comme étant normales, mais lorsqu'on s'intréresse au taux de bonne classification sur les deux classes "*Overall*", on remarque qu'on a un taux élevé ce qui renseigne encore plus sur l'éfficacité de ces nouvelles mesures pour décrire les voix pathologiques. Ceci dit, il ne fallait pas rejeter ces mesures à cause du déséquilibre des résultats qu'elles procurent, au contraire on les a exploité en les fusionnant avec les descripteurs décrivant au mieux les voix normales.
Les résultats du pouvoir discriminant des descripteurs a été testé sur une base de donnés largement connue et exploitée dans les travaux de recherches, ce qui renforce encore plus leurs pertinences.

Dans les travaux de recherches la fin n'est jamais atteinte, et il y aura toujours un chemin pour optimiser les solutions et les améliorer. C'est dans cette perspective qu'on propose, pour des futurs travaux, d'évaluer les résultats de classification avec une approche autre que le leave-one-out. On propose par exemple la technique de ré-échantillonnage de la base de données : bootstap.
En continuité de l'exploitation des histogrammes, on propose aussi de lancer des batteries de tests ayant comme entrées les sorties des histogrammes.
Afin d'aller plus loin, on suggère de subdiviser la base de données par catégorie de maladies, de cette manière, des études caractéristiques de groupes de maladies peuvent être réaliser afin d'ajuster les descripteurs à chaque classe. On pourrait aussi évaluer des sous groupes selon un critère donné, le sexe par exemple, tabagisme, etc.

Certes que les 6 mois passés à l'équipe Géostat ont été bénfiques sur le plan technique, mais aussi sur le plan personnel et professionnel. Ce stage m'a sensibilisée au besoin d'un projet de recherche appliqué à la médecine. J'ai pu ainsi comprendre et cerner ce que c'est une problématique scientifique ce qui m'a donnée une envie encore plus grande de m'impliquer dans mon travail.

Ce stage m'a permis d'enrichir mon savoir-être en développant la qualité de la patience car les travaux de recherche n'aboutissent toujours pas dès les premiers essais. J'ai appris aussi que grâce aux rencontres et la communication avec les membres de l'équipe qu'on apprend le plus de choses. Côtoyer les membres de l'équipe de Géostat quotidiennement m' a apportée une maturité immédiate qui permet de mieux cibler le problème et augmenter par conséquent la rapidité des traitements.

Bibliographie

[1] M.Little, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection", BioMedical Engineering OnLine, 2007.

[2] C.Barras, "Classification et reconnaissance de la parole", LIMSI-CNRS, pp. 5, 2004.

[3] M.Little, "NONLINEAR, BIOPHYSICALLY-INFORMED SPEECH PATHOLOGY DETECTION", IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings.

[4] V.Khanagha, "Nouvelles méthodes multi-échelles pour l'analyse non linéaire de la parole", Thèse de doctorat en Sciences, 2013.

[5] C.Cawley et L.C Talbot, "Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers", 36, pp. 2585-2592, 2003.

[6] JJ.Jiang, Y.Zhang, "Chaos in voice, From modeling to Measurement", Journal of Voice, Vol 20, pp. 2-17, 2006.

[7] C.Bouveyron, S.Girard, C.Schmid, "Analyse Discriminante de Haute Dimension", Rapport de recherche nř5470 INRIA Rhône-Alpes, 2005.

[8] M.Little, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease", IEEE Transactions on BioMedical Engeneering, pp. 1015-1022, 2008.

[9] P.Auzou S.Pinto, "Les Dysarthries", SOLAL Editeur, 2007.

[10] A.Tsanas, "Accurate telemonitoring of Parkinson's disease symptom severity using non linear speech signal processing and statistical machine learning", Thèse de doctorat en Philosophie, 2012.

[11] M.Little, "Biomechanically Informed Nonlinear Speech Signal Processing", Thèse de doctorat en Philosophie, 2006.

[12] S.Mallat, Z.Zhang, "Matching Pursuits With Time-Frequency Dictionaries", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL 41, NO 12, Décembre 1993.

[13] K.Umpathy, "Discrimination of Pathological Voices Using a Time-Frequency Approach", IEEE TRANSACTIONS ON BIOMEDICAL ENGENEERING, VOL 52, NO 3, Mars 2005.

[14] A.Belhaj, "Evolution des paramètres de source pour différentes qualités vocales, déterminés sur le signal EGG", e-STA, VOL 7, NO 1, pp. 23-29, 2010.

[15] Division of PENTAX Medical Company, "Multi-Dimensional Voice Program (MDVP) Software Instruction Manual", Juin 2008.

[16] A.Ghio, "L'évaluation acoustique", 2007.

[17] P.Boersma, "ACCURATE SHORT-TERM ANALYSIS OF THE FUNDAMENTAL FREQUENCY AND THE HARMONICS-TO-RATIO OF A SAMPLED SOUND", Proceedings 17, pp 97-110, 1993.

[18] O.Amir, M.Wolf, "A clinical comparasion between two acoustic analysis softwares : MDVP and Praat", Biomedical Signal Processing and Control, vol.4, pp. 202-205, 2009.

[19] D.G.Silva et M.Andrea, "Jitter Estimation Algorithms for Detection of Pathological Voices", EURASIP Journal on Avances in Signal Processing, Juin 2009.

[20] N. Henrich et C.d'Alessandro, "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation", Acoustical Society of America, pp. 1321-1332, 2004.

[21] E.Yumoto et J.Gould,"Harmonics-to-noise ratio as an index of the degree of hoarsness", Journal of the Acoustical Society of America, Vol 71, pp.1544-1550, 1982.

[22] E.SICARD et A.MENIN-SICARD, "Implémentation dans VOCALAB d'indicateurs objectifs de la qualité de la voix dans le cadre de l'évaluation de la voix", Rééducation orthophonique, 254, pp. 43-60, Juin 2013.

Annexe A

Test VHI en Anglais

VOICE HANDICAP INDEX

Instructions: These are statements that many people have used to describe their voices and the effects of their voices on their lives. Circle the response that indicates how frequently you have the same experience.

0 = Never 1= Almost never 2 = Sometimes 3 = Almost always 4 = Always

1. My voice makes it difficult for people to hear me
   0 1 2 3 4

2. People have difficulty understanding me in a noisy room
   0 1 2 3 4

3. My voice difficulties restrict me in personal and social life
   0 1 2 3 4

4. I feel left out of conversations because of my voice
   0 1 2 3 4

5. My voice problem causes me to lose income
   0 1 2 3 4

6. I feel as though I have to strain to produce voice
   0 1 2 3 4

7. The clarity of my voice is unpredictable
   0 1 2 3 4

8. My voice problem upsets me
   0 1 2 3 4

9. My voice makes me feel handicapped
   0 1 2 3 4

10. People ask, "What's wrong with your voice?"
    0 1 2 3 4

Annexe B

Descriptif de la base de données

Description selon le sexe

	Normale	Pathologique
Masculin	21	200
Féminin	32	276
Sexe non précisé (snp)	0	247
Total	53	723

Description selon la nature des enregistrements

	Normal	Pathologique	Total
Voyelle	53	657	710
Phrases	53	662	715

Les fréquences d'échatillonnage

	10 kHz	25 kHz	50 kHz	Total
Voyelle normale	0	0	53	53
Voyelle pathologique	0	580	77	657
Phrase normale	17	36	0	53
Phrase pathologique	13	648	1	662

Remarque 1 : Il n' y a pas d'enregistrements pour toutes les personnes (dont la voix est pathologique) décrites dans la base. C'est pour cela qu'au niveau de la table personnes il y a 723 personnes dont la voix est pathologique alors qu'au niveau de la table Enregistrements il y a 657 voyelles pathologiques et 662 phrases pathologiques.

Remarque 2 : Il y a des enregistrements pour lesquels le sexe n'est pas précisé.

Annexe C

Algorithme de calcul du HNR

Remarque : τ_{max} correspond à l'échantillon qui fourni le maximum global de l'autocorrélation sur le segment x_i.

Entrée	**Application**	**Sortie**
Segment du signal formé de 6 cycles x_i	Soustraction de la moyenne locale	$x_i - \bar{x}_i$
$x_i - \bar{x}_i$	Multiplication par la fenêtre de Hanning afin d'éliminer les effets de bords	$a_i(t) = (x_i - \bar{x}_i).w(t)$
$a_i(t)$	Calcul de l'autocorrélation du segment multipilié par la fenêtre de Hanning $\int a(t).a^*(t+\tau)$	$r_{ai}(\tau)$
$w(t)$	Calcul de l'autocorrélation de la fenêtre de Hanning $\int w(t).w^*(t+\tau)$	$r_w(\tau)$

$r_w(\tau)$ et $r_{ai}(\tau)$	Obtention de l'autocorrélation du segment original x_i $\frac{r_{ai}}{r_w}$	$r_{xi}(\tau)$
$r_{xi}(\tau)$	Calcul de l'autocorrélation normalisée du segment original x_i $\frac{r_{xi}(\tau_{max})}{r_x(0)}$	$r'_{xi}(\tau_{max})$
$r'_{xi}(\tau_{max})$	Définition du HNR logarithmique $10.\log_{10}\left(\frac{r'_x(\tau_{max})}{1-r'_x(\tau_{max})}\right)$	HNR(dB)

Annexe D

Manuel de codes

Dans cet annexe, on donne un descriptif des fonctions les plus importantes.

Fonction	**Fonctions internes**	**Description**
Getpitch		Retourne la valeur du pitch [pitch] = GetPitch (ps, N_c, F_t) Entrées : ps : séquence du pitch f_{0i} N_c : nombre de bars de l'histogramme F_t : valeur de f_{0i} maximale à considérer Sorties : pitch : la fréquence fondamentale (f_{0i} ayant la plus grande probabilité d'apparition)
GetShimmer		Retourne la valeur du shimmer [shimmer] = GetShimmer(s, y, F_t) Entrées : s : amplitude normalisée y : vecteur des IFG F_t : valeur de f_{0i} maximale à considérer Sorties : shimmer : la valeur du shimmer en (%)
GetJitter	GetPitch	Retourne la valeur du jitter [Jitter] = GetJitter(ps, N_c, F_t) Entrées : ps : séquence du pitch formée des f_{0i} N_c : nombre de bars de l'histogramme F_t : valeur de f_{0i} maximale à considérer Sorties : Jitter : la valeur du jitter en (%)

GetStatistics	GetPitch	[dspr dev sk kr] = GetStatistics (ps, N_c, F_t) Entrées : ps : séquence du pitch f_{0i} N_c : nombre de bars de l'histogramme F_t : valeur de f_{0i} maximale à considérer Sorties : dspr : EM_{F_0} l'écart moyen par rapport au pitch dev : σ_{F_0} la déviation standard par rapport au pitch sk : S_{F_0} le skewness ou coefficient de dyssimétrie par rapport au pitch kr : $K_{F_{0i}}$ le kurtosis ou coefficient d'applatissement
GetStatProportions		Donne les mesures des 3 paramètres de proportions [D, R, R_{mm}] = GetStatProportions(ps, N_c, pas, F_t) Entrées : ps : séquence du pitch f_{0i} N_c : nombre de bars de l'histogramme pas : nombre de bars de l'histogramme à considérer de part et d'autre du mode F_t : valeur de f_{0i} maximale à considérer Sorties : D : différence entre la proportion à droite du mode et celle à gauche R : rapport entre la proportion à droite du mode et celle à gauche R_{mm} : rapport de la valeur maximale entre la proportion à droite et la proportion à gauche, divisé par la valeur minimale entre elles
HNR_function_modified		Retourne la valeur du HNR [HNR] = HNR_function_modified(s, f_s, gciloc) s : amplitude normalisée du signal f_s : la fréquence d'échantillonnage gciloc : vecteur donnant la localisation des IFG Sorties : HNR : valeur du HNR en dB