

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
ECOLE NATIONALE POLYTECHNIQUE



DÉPARTEMENT D'ÉLECTRONIQUE

Projet de fin d'études

En vue de l'obtention du diplôme de Master en Electronique

Thème :

**Analyse d'un signal vocal par
Sonagrammes et LPC**

Réalisé par :

Mr **BOUMAZA SOUFYANE**

Proposé et Dirigé par :

M. GUERTI Professeur ENP

Promotion : Juin 2013

Dédicaces

Je dédie ce travail à :

- *Mes parents qui m'ont soutenu, orienté, aidé et encouragé le long de ma vie.*
- *mes frères et sœurs avec qui j'ai passé les plus beaux moments de ma vie.*
- *tous mes amis : Amine, Sid Ali, Mohamed El Amine, Moussaab, Raouf, Houari, Abdelkader, Ismail*
- *A tous ceux pour qui je compte et à tous ceux qui comptent pour moi.*

Soufyane

Remerciements

Tout d'abord je remercie Dieu de m'avoir donné la force et le courage d'accomplir ce travail.

*Je remercie vivement mon ma promotrice **Professeur GUERTI Mhania** pour m'avoir confié ce travail d'abord et pour son soutien constant, son rôle majeur et sa grande patience ainsi que ses encouragements durant toute la période de ce travail. Je la remercie pour ses compétences, son ouverture d'esprit et sa grande disponibilité.*

J'exprimer ma profonde gratitude à Monsieur B.BOUSSEKSOU, chargé de cour à l'ENP, pour l'honneur qu'il me fait en acceptant de présider le jury de ce PFE.

Je remercie vivement Monsieur L.Saadaoui, chargé de cours à l'ENP, pour m'avoir fait le plaisir d'accepter de participer à mon jury en tant qu'examineur.

Mes remerciements vont également à tous mes enseignants à l'Ecole Nationale Polytechnique qui ont contribué à ma formation.

Je remercie tous ceux, qui de près ou de loin, m'ont apporté leur contribution pour la réalisation de ce travail.

الملخص

الهدف من عملنا هذا هو دراسة التحليل الأكوستيكية للصوت، ومن أجل هذا قمنا بدراسة تقنيات التعرف الخطي المشفر LPC ودرسنا جملة باللغة العربية من خلال إجراء تحليل أكوستيكية وتحليل ب LPC، هذا العمل أنجز ببرنامج .PRAAT

كلمات المفاتيح : العربية الفصحى، تحليل الصوت، التعرف الخطي المشفر.

Résumé

Le but de notre travail est de faire des études sur les analyses acoustiques du signal vocal. Pour cela, nous avons étudié les techniques d'analyse par Codage Prédiction Linéaire (LPC). Pour ce faire, nous avons analysé une phrase en Arabe Standard par sonagramme et LPC. Ce travail a été réalisé a l'aide d'outil d'analyse Praat.

Mots clés : Arabe Standard, Analyse de la parole, Codage Prédiction Linéaire.

Abstract

The objective of our work is to study the acoustic analysis of speech signal. For this, we studied the technical analysis by Linear Prediction Coding (LPC). To do this, we analyzed a sentence in Standard Arabic by sonogram and LPC. This work has been carried out using praat analysis tool's.

Keywords : Standard Arabic, Speech Analysis, Linear Prediction Coding.

Liste des Abréviations

TAP	Traitement Automatique de la Parole
TOP	Transcription Orthographique de la Parole
LPC	Linear Predictive Coding
CT	Court-Terme
F0	Fréquence Fondamentale
Fe	Fréquence d'échantillonnage
FT	Fonction de Transfert
T0	Période Fondamentale
TFD	Transformée de Fourier Discrète
TF	Transformée de Fourier
FFT	Faste Fourier Transformer
AR	AutoRégressif
DFW	Dynamic Frequency Wrapping
TFCT	Transformée de Fourier à Court Terme

Liste des tableaux

Tableau 1.1 : Transcription Orthographique Phonétique de l'AS.....	10
Tableau 3.1 : les valeurs moyennes de formants de la phrase Ph.....	26
Tableau 3.2 : les valeurs de l'intensité Min-Max-Moyenne phrase Ph.....	26
Tableau 3.3 : les valeurs de pitch Min, Max, Moyenne de la phrase Ph.....	27
Tableau 3.4 : les valeurs de coefficients a_i avec la méthode d'autocorrélation (LPC)....	28

Table des Figures

Figure 1.1 : Modèle simplifié de l'appareil phonatoire.....	1
Figure 1.2 : les organes de la phonation.....	3
Figure 1.3 : Evolution de la fréquence de vibrations des cordes vocales de la phrase "Les techniques de traitement ".....	4
Figure 1.4 : Audiogramme du signal de parole du mot « Cinq » [sɛ̃k].....	6
Figure 1.5 : Audiogramme du signal de parole du mot « Parenthèse » [parãtezə].....	6
Figure 1.6 : Représentation des Formants d'un son voisé.....	7
Figure 2.1 : Prétraitement du signal vocal.....	14
Figure 2.2 : Analyse numérique du signal parole par FFT.....	15
Figure 2.3 : Modèle de production de la parole.....	18
Figure 3.1 : Présentation du logiciel Praat.....	22
Figure 3.2 a, b : Spectrogramme de la phrase / المحطة التالية / [ʔlmaħaʔ ʔa ʔtaaliʒa]	23
Figure 3.3 : Visualisation de Ph avec Praat.....	25
Figure 3.2 : diagramme de l'intensité de la Ph.....	26
Figure 3.15 : la mélodie de la phrase Ph.....	27

Sommaire

INTRODUCTION GENERALE

CHAPITRE 1 : GENERALITES SUR LA PAROLE

1.1 INTRODUCTION	1
1.2 GENERALITE SUR LA PAROLE	1
1.3 L'APPAREIL PHONATOIRE HUMAIN	1
1.3.1 Les voies aériennes inferieures	2
1.3.2 Le larynx	2
1.3.3Le conduit vocal	2
1.4 LES PARAMETRES PROSODIQUES ET ACOUSTIQUES D'UN SIGNAL VOCAL	3
1.4.1 La Fréquence Fondamentale	4
1.4.2 La durée	5
1.4.3 L'Intensité ou l'énergie	5
1.4.4 Les Formants	6
1.5 NOTIONS FONDAMENTALES SUR L'ARABE STANDARD (AS)	8
1.5.1 Système phonétique de l'Arabe Standard	9
<i>1.5.1.1 Les voyelles</i>	9
<i>1.5.1.2 Les consonnes</i>	10
1.5.2 Particularités de l'Arabe Standard	11
<i>1.5.2.1 Voyelles longues</i>	11
<i>1.5.2.2 Gémination</i>	11
<i>1.5.2.3 Emphase</i>	11
1.6 CONCLUSION	12

Chapitre 2 : TECHNIQUES D'ANALYSE DU SIGNAL VOCAL

2.1 INTRODUCTION	13
-------------------------------	----

2.2 TECHNIQUES D'ANALYSE DU SIGNAL VOCAL	13
2.2.1 Méthodes non paramétriques	14
2.2.2 Méthodes paramétriques	14
2.3 CODAGE PREDICTION LINEAIRE	15
2.3.1 Le modèle autorégressif (AR)	16
2.3.2 Modèle AR et modèle de prédiction linéaire	16
2.3.3 Pourquoi utilise-t-on le modèle autorégressif	17
2.3.4 Estimation des coefficients de prédiction linéaire	19
2.4 CONCLUSION	20

**Chapitre 3 : ANALYSE ACOUSTIQUE PAR SONAGRAMME ET CODAGE
PREDICTION LINEAIRE**

3.1 INTRODUCTION	21
3.2 L'OUTIL D'ANALYSE	21
3.2.1 Le logiciel Praat	21
3.2.2 Visualisations par spectrogramme	22
3.2.3 Analyse par spectrogrammes	23
3.2.4 Lecture de spectrogramme	24
3.3 ANALYSE ACOUSTIQUE DE NOTRE SIGNAL VOCAL	25
3.3.1 Analyse par sonagramme	26
<i>3.3.1.1 analyses formantique</i>	26
<i>3.3.1.2 Analyse de L'intensité</i>	26
<i>3.3.1.3 Analyses fréquentiels</i>	26
3.3.2 Analyse avec Codage Prédiction Linéaire (LPC)	27
3.4 CONCLUSION	28
4. CONCLUSIONS ET PERSPECTIVES	29
5. REFERENCES BIBLIOGRAPHIES	30

INTRODUCTION GENERALE

Le traitement de la parole est un vaste domaine de recherche qui demande l'intervention des experts de plusieurs spécialités. Malgré le développement remarquable des outils et les programmes informatiques, les systèmes à commandes vocales n'ont eu du succès que ces dernières années.

Le **Traitement Automatique de la Parole (TAP)** est une composante fondamentale des sciences de l'ingénieur et un domaine de recherche actif, au croisement du traitement du signal numérique et du traitement symbolique du langage. Depuis les années 60, le TAP bénéficie d'efforts de recherche très importants, liés au développement des moyens et techniques de Télécommunications et du traitement numérique de l'information. Ces efforts se sont concrétisés grâce à plusieurs applications du TAP, telles que le codage, la **Reconnaissance Automatique** et la **Synthèse de la Parole (RAP ; SP)**. Un thème important de la recherche actuelle dans le domaine du TAP, est la réalisation de véritables systèmes de dialogue oral entre l'Homme et la Machine.

La prédiction linéaire est une technique importante pour la compression d'un signal de parole en modélisant le conduit vocal par un filtre dont les coefficients sont déterminés au biais de l'analyse de la redondance du signal, cela est réalisé par la prédiction linéaire qui exploite cette redondance pour prédire un échantillon par une combinaison linéaire d'échantillons antérieurs, c'est l'idée de base du codage par prédiction linéaire LPC.

L'objectif de notre travail est de faire des études sur les analyses sonagrammes et analyses par codage prédiction linéaire. Cette étude a été effectuée par logiciel Praat.

Nous avons structuré notre travail en 3 chapitres :

- dans le premier, nous avons décrit de manière générale des notions sur le traitement de la parole, des spécifications du signal vocal et des notions fondamentales sur l'Arabe Standard ;
- dans le deuxième présente les techniques et l'outil d'analyses de signal vocal ;
- dans le troisième chapitre, nous avons appliquée les techniques d'analyses sonagrammes et les analyses par codage par prédiction linéaire (LPC).

Nous terminons notre travail par des conclusions générales et perspectives.

Chapitre I :
Généralités sur la parole

1.1 INTRODUCTION

La parole est le seul moyen qui permet de communiquer la pensée par un système de sons articulés. Les humains sont les seuls êtres vivants qui utilisent un tel type de systèmes structurés. Dans ce chapitre nous allons décrire de manière générale des notions sur le traitement automatique de la parole et de sa production, ensuite nous présentons les appareils phonatoires et auditifs humains, les spécifications du signal vocal et des notions fondamentales sur l'Arabe Standard.

1.1 GENERALITE SUR LA PAROLE

La parole est une faculté, propre à l'Homme, de communication par des sons articulés. Elle met en jeu des phénomènes de natures très différentes et peut être analysée de bien des façons. On distingue généralement plusieurs niveaux de description non exclusifs : physiologique, phonologique, phonétique, acoustique, morphologique, syntaxique, sémantique, et pragmatique [1].

1.3 ANATOMIE DE L'APPAREIL PHONATOIRE

L'appareil phonatoire est l'ensemble des organes qui permettent de produire les sons constituant la voix. Chez l'Homme, cet appareil est confondu avec l'appareil respiratoire et ses organes sont répartis entre le thorax, le cou et la tête. D'une manière générale, l'appareil phonatoire est décomposé en trois parties correspondant à trois entités fonctionnelles différentes : les voies aériennes inférieures composées des poumons et de la trachée artère, le larynx, et le conduit vocal (Fig.1.1).

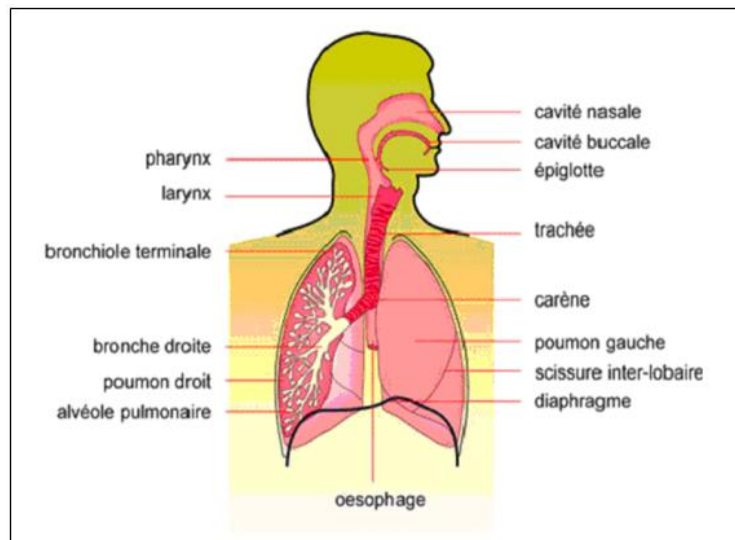


Figure 1.1 : Modèle simplifié de l'appareil phonatoire [2]

1.3.1 Les voies aériennes inférieures

Les voies aériennes inférieures correspondent à la partie de l'appareil phonatoire située dans le thorax et sont composées de deux poumons reliés à la trachée qui elle-même remonte jusqu'aux voies aériennes supérieures. La fonction première des poumons est d'assurer la fonction de respiration en permettant l'échange d'oxygène et de dioxyde de carbone entre le sang et l'air extérieur. Lors de la phonation, les poumons jouent le rôle de réservoir de pression et permettent de générer l'écoulement d'air à l'origine de la production de sons et notamment des vibrations des cordes vocales. La circulation de l'air entre les poumons et l'extérieur est réalisée grâce aux mouvements du diaphragme et aux contractions et relâchements des muscles de la cage thoracique. Cette ventilation se fait ainsi dans un mouvement de va-et-vient correspondant alternativement à l'inspiration et à l'expiration. Sauf pour des cas atypiques, la phonation intervient durant la phase d'expiration.

1.3.2 Le larynx

Le larynx est l'organe qui fait la jonction entre la trachée et le pharynx. Il se situe dans la gorge et remplit trois fonctions. La première est respiratoire puisque qu'il est le premier organe des voies aériennes supérieures. Ensuite, il joue un rôle dans la déglutition en bloquant l'accès aux poumons pour les aliments en cas de « fausse route », c'est à dire un dysfonctionnement de l'épiglotte sensée diriger ces aliments vers le tube digestif. Enfin il abrite les cordes vocales et est donc le siège de la production des sons voisés qui implique la vibration des cordes vocales. Le larynx est composé de cartilages, de muscles et de muqueuses. Une coupe coronale d'un pli vocal montre que celui-ci est constitué de plusieurs couches de tissus musculaires, ligamentaires et muqueux, dont les propriétés sont différentes (Fig.1.2).

La structure musculo-cartilagineuse autour des cordes vocales permet de faire varier leur tension, leur position et l'espace compris entre ceux-ci, appelé la glotte. Lors de la production des sons voisés de la parole (comme les voyelles par exemple), c'est la vibration des cordes vocales qui constitue la source des ondes acoustiques.

1.3.3 Le conduit vocal

Le conduit vocal est la partie des voies aériennes supérieures située au-dessus du larynx, il est localisé dans la tête et constitué du pharynx et de deux cavités résonnantes séparées par le palais : la cavité orale (ou buccale) et la cavité nasale. Lorsque les cordes vocales vibrent,

les ondes acoustiques générées se propagent dans ces cavités qui agissent comme un résonateur acoustique (Fig.1.2).

La géométrie du conduit vocal influe ainsi sur les fréquences de résonance et sur le signal acoustique de parole émis. Pour un individu, la forme de la cavité nasale est invariable. En revanche la forme de la cavité orale peut être modifiée grâce à plusieurs articulateurs comme la langue, les lèvres ou la mâchoire. Le voile du palais (ou vélum) situé à l'extrémité du palais dur est constitué d'une membrane et d'un muscle et joue le rôle de clapet entre ces deux cavités. Selon sa position, il permet soit d'obturer la cavité nasale, soit d'obturer la cavité orale ou soit de relier les deux cavités. Certains sons de la parole (comme les consonnes fricatives par exemple) sont produits dans le conduit vocal et n'implique pas la vibration des cordes vocales.

L'appareil phonatoire humain permet de générer une grande variété de sons. En contrôlant simultanément et de façon dynamique la pression pulmonaire, les propriétés des cordes vocales et la configuration géométrique du conduit vocal, un locuteur peut produire l'ensemble des voyelles et des consonnes qui constituent un langage [3].

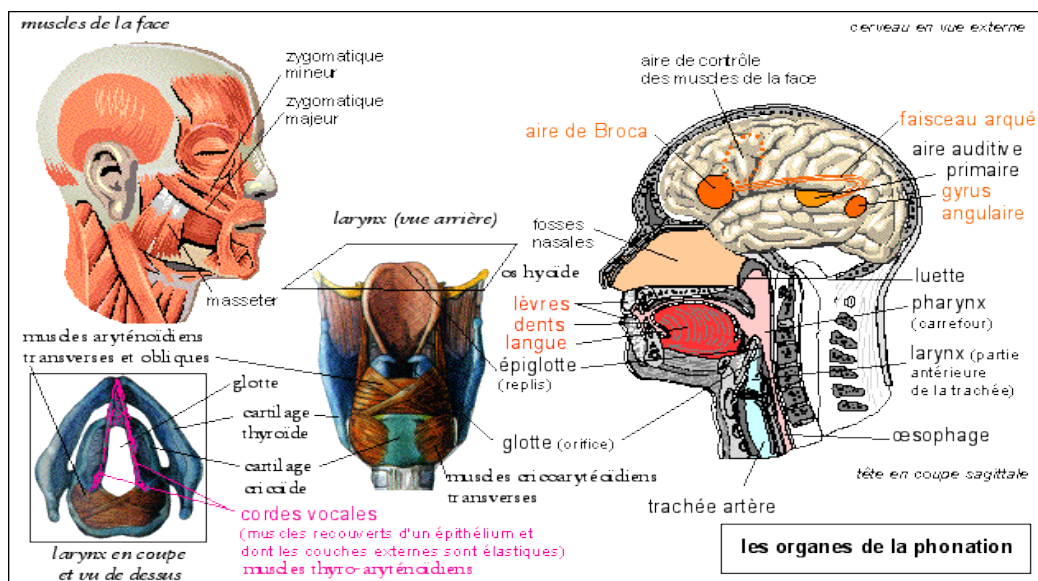


Figure 1.2 : les organes de la phonation [4]

1.4 LES PARAMETRES PROSODIQUES ET ACOUSTIQUES D'UN SIGNAL VOCAL

La prosodie est une science de la linguistique qui étudie les éléments phoniques (l'accent, l'intonation, etc.) de n'importe quelle langue, et puisque la parole est un signal réel d'énergie finie, continu, et non stationnaire ; les variations des paramètres prosodiques

physiques (La fréquence fondamentale, la durée, et l'intensité) influencent de manière directe sur ces éléments phoniques.

Les recherches en linguistique ont montré que les caractéristiques prosodiques sont des composantes indispensables à la langue et à la fonction de communication. Puisqu'elles influencent directement sur l'intelligibilité de la parole synthétique. Il existe trois manières de définir les paramètres prosodiques, selon qu'on les considère sur les plans de la production, de l'acoustique, et de la perception auditive.

1.4.1 La Fréquence Fondamentale

La Fréquence Fondamentale ou F_0 est la fréquence de vibrations des cordes vocales, elle varie d'une personne à une autre en fonction de la longueur et de la masse des cordes vocales de chaque personne (Fig.1.3).

Elle permet de diviser l'ensemble des sons de parole en trois grandes macros classes [4]:

- 70 -250 Hz pour les hommes ;
- 150 - 400 Hz pour les femmes ;
- 200 - 600 Hz pour les enfants.

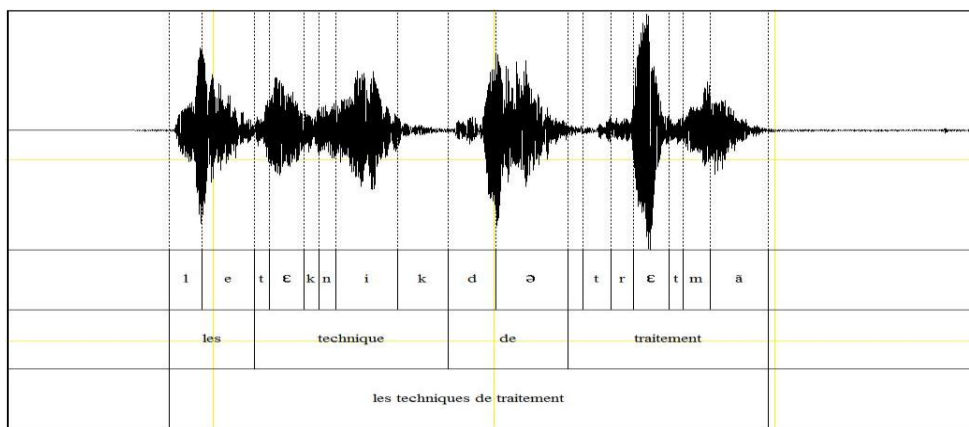


Figure 1.3 : Evolution de la fréquence de vibrations des cordes vocales de la phrase :
"Les techniques de traitement "

Les variations de la fréquence au cours de la parole constituent ce qu'on appelle la mélodie ou l'intonation. Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la F_0 .

1.4.2 La durée

La durée est une mesure très variable. Elle représente le temps de la prononciation d'un phonème. Il existe deux types :

- la durée observée, qui correspond à la mesure objective du temps de l'activation des organes de phonation ;
- la durée perçue, est liée au mécanisme de la perception et elle est fréquemment utilisée dans le cas des occlusives puisqu'elles sont caractérisées par une durée de réalisation non continue.

Généralement la durée d'une unité est mesurée par le nombre des trames qu'elle contient. Pour calculer la durée de chaque trame, il faut fixer deux événements sur le signal de parole qui délimitent les repères initial et final de cette trame.

1.4.3 L'Intensité ou l'énergie

Elle est résultante de la pression sous glottique. Généralement elle exprime le volume sonore d'un phonème et dans le cas d'un voisement elle représente l'amplitude des vibrations des cordes vocales. Elle est exprimée pour un signal échantillonné x_n par :

$$E = \frac{1}{T} \sum_{n=1}^T x_n^2 \quad (1.1)$$

$$E_{db} = 10 * \log_{10} \left(\frac{1}{T} \sum_{n=1}^T x_n^2 \right) \quad (1.2)$$

Le rythme d'élocution correspond à la vitesse du débit de parole. On peut faire varier ce paramètre de manière à ce qu'une phrase prononcée trop rapidement puisse être « ralentie » pour la rendre plus compréhensible lors de l'apprentissage d'une langue étrangère par exemple. L'intensité du son émis est liée à la pression de l'air en amont du larynx. Les figures 1.4 et 1.5 représentent l'évolution temporelle du signal vocal pour les mots cinq et parenthèse, elles donnent un exemple des parties voisées et non voisées du signal vocal.

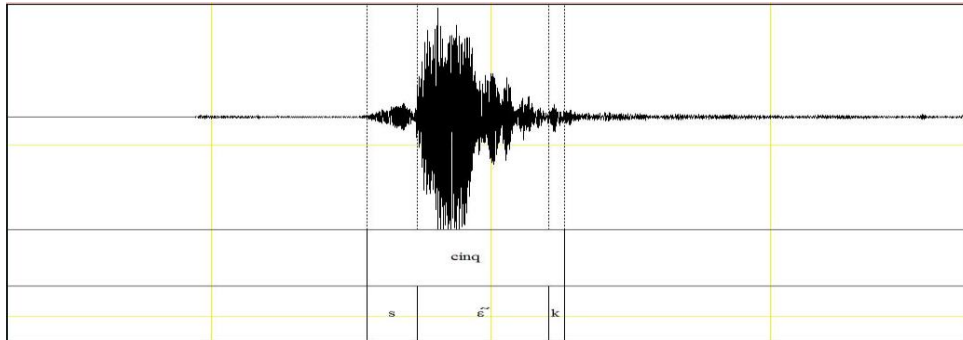


Figure 1.4 : Audiogramme du signal de parole du mot « Cinq » [sɛ̃k]

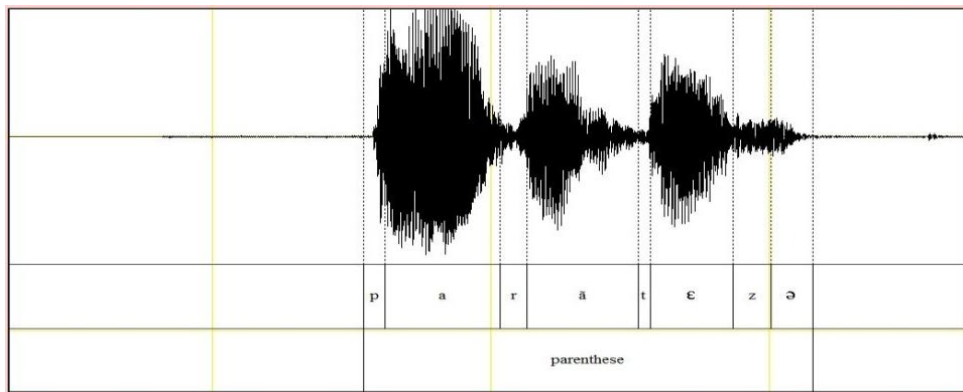


Figure 1.5 : Audiogramme du signal de parole du mot « Parenthèse » [parãtezə]

Tout l'enjeu du traitement de la parole est de modéliser l'appareil phonatoire humain de façon à créer un signal de parole synthétique aussi réaliste que l'original. Il existe plusieurs manières de le faire, notamment en utilisant un vocodeur à prédiction linéaire qui, dans un premier temps, code le signal vocal de manière à réduire le débit d'informations puis le restitue à l'aide de paramètres qui caractérisent la fonction de transfert du conduit vocal. Ces paramètres étant réactualisés toutes les 20 ms environ. En fait on part de l'hypothèse qu'un échantillon de parole peut être prédit à partir d'une pondération linéaire d'un nombre fini d'échantillons précédents.

Cette hypothèse se justifie par le fait que la forme du conduit vocal n'évolue pas rapidement. En général, on considère que l'appareil vocal est quasi stationnaire sur un intervalle de temps de l'ordre d'une vingtaine de millisecondes. On parle donc ici de statistique du signal à court terme. Cette méthode a l'avantage de donner de bons résultats au niveau du signal synthétique mais demande des capacités de calcul important pour la réalisation en temps réel [4].

1.4.4 Les Formants

Lorsqu'un excitateur entre en vibrations et fournit un signal, ce dernier passe à travers une cavité de résonance (le résonateur) qui amplifie certaines composantes. On obtient alors ce qu'on appelle les formants qui sont un facteur essentiel dans la caractérisation du timbre.

L'appareil phonatoire étant constitué de différentes cavités. Lors du passage de l'air à travers ces cavités, il est amplifié et subit différentes transformations dues aux degrés d'ouverture et de fermeture au niveau de chaque cavité à la position de la langue, des lèvres, etc. Ces cavités possèdent des fréquences de résonance qui renforcent certaines régions du spectre de sources excitatrices. Les maxima de la courbe de réponse en fréquences du conduit vocal sont appelés Formants. Chaque son a ses formants caractéristiques. Sur un spectrogramme, les formants sont représentés par des bandes noires (le degré de noirceur correspondant à l'énergie) (Fig.1.6).

La fréquence fondamentale (fréquence de vibrations des cordes vocales) est responsable de la hauteur perçue d'un son. Les fréquences d'harmonique renforcées, responsables du timbre d'un son, sont aussi numérotées. F_1 correspond à la première zone d'harmoniques renforcées, F_2 à la seconde et ainsi de suite jusqu'à F_5 .

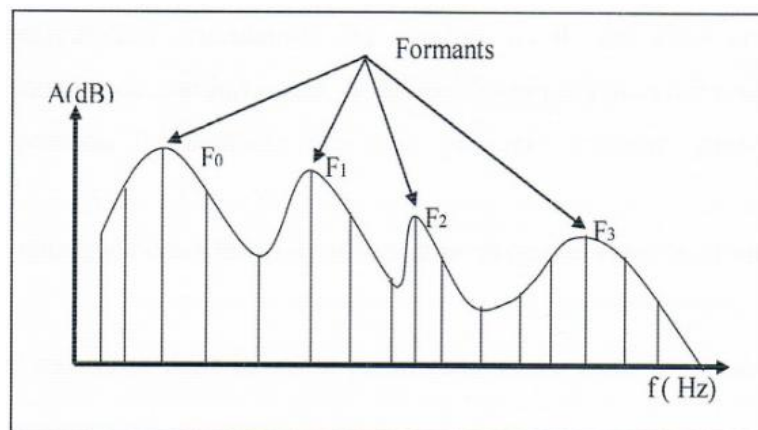


Figure 1.6 : Représentation des Formants d'un son voisé [5]

Les valeurs des formants sont très influencées par le lieu d'articulation des phonèmes. Elles donnent une image de la configuration articulaire du conduit vocal car elles correspondent aux fréquences de résonances du conduit vocal.

Le nombre des formants, selon les caractéristique du résonateur (volume, forme et ouverture), est variable : d'un seul (théoriquement) à une infinité. Néanmoins, du point de vue perceptif, seuls quelques-uns d'entre eux jouent un rôle central au niveau de la parole. Par exemple, on peut caractériser toute voyelle en ne prenant compte que ses trois premiers formants F_1 , F_2 et F_3 . (Pour une réalisation de la voyelle [i] par exemple, les trois premiers formants pourraient se situer respectivement à 300, 2200 et 3000 Hz).

Généralement, nous pouvons aller jusqu'à cinq ou six formants pour produire une parole de très haute qualité. Les formants nous permettent de décrire aussi les cibles vocalique correspondant aux zones stables ainsi que les zones de transitions (passage entre deux son consécutifs) ce qui montre leur très grande importances pour l'analyse acoustique en phonétiques au moins trois formants sont exigés pour produire les différentes voyelles généralement, on peut aller jusqu'à cinq formants pour produire une parole de haute qualité.

Les valeurs des formants sont très influencées par le lieu d'articulation, ils donnent une image de la configuration articuloire du conduit vocales, car elles correspondent aux fréquences de résonance du conduit vocale, de même des expériences qui reste à vérifier ont montré que la position fréquentielle des trois premiers formants caractérise le timbre vocalique [5]:

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

A cela on ajoute le fait que plus la cavité de résonance est large, plus que la fréquence correspondante est basse. Réciproquement plus cette cavité est petite plus la fréquence et haute.

1.5 NOTIONS FONDAMENTALES SUR L'ARABE STANDARD (AS)

L'Arabe est une langue parlée par plus de 337 millions de personnes. Elle est la langue officielle d'au moins 22 pays. C'est aussi la langue de référence pour plus de 1,3 milliard de musulmans. Comme son nom l'indique, la langue Arabe est la langue parlée à l'origine par le peuple arabe. Dans le cadre de notre travail, nous parlerons de la langue Arabe en référence à ce qui est communément appelé "l'Arabe Standard" (AS), c'est-à-dire, la langue de

communication commune à l'ensemble du Monde Arabe. Il s'agit de la langue enseignée dans les écoles, donc écrite, mais aussi parlée dans le cadre officiel [6].

1.5.1 Système phonétique de l'Arabe Standard

L'Arabe Standard(AS) comprend 40 phonèmes, dont 3 voyelles courtes, 3 voyelles longues plus 6 variantes vocaliques en contexte emphatique et 28 consonnes. Les phonèmes arabes se distinguent par la présence de deux classes qui sont appelées pharyngales et emphatiques. La graphie des lettres est différente selon leur position dans le mot. Ainsi, la lettre ب[b] est transcrite بَيْتٌ[bajtun] (une maison) en début de mot, خُبْزٌ[xubzun] (du pain) en milieu de mot, كَلْبٌ[kalbu] en fin de mot et قُرْبٌ[qurba] (à proximité de) isolé en fin de mot. Il résulte 78 formes graphiques à partir des 28 lettres. Par ailleurs, la distinction minuscules/majuscules n'existe pas [6].

Pour les besoins de la transcription les 28 consonnes arabes ont été divisées en deux groupes :

- 14 consonnes solaires qui assimilent le « ل » de l'article ;
- 14 consonnes lunaires qui n'assimilent pas le « ل » de l'article.

Les solaires se prononcent en double, comme par exemple avec le mot « soleil » شمس[chams], au lieu de prononcer الشمس, el-chams, on prononce ech-chams, car la lettre ش [chin], est une lettre solaire.

Les lettres lunaires, se prononcent normalement et simplement pour elles-mêmes, c'est-à-dire sans les doubler. Par exemple avec le mot « lune », قمر ([qamar] - lune), on prononce القمر, [el-qamar] tout à fait normalement, parce que la lettre ق [qaf] est une lettre lunaire (Tableau 1.2).

1.5.1.1 Les voyelles

On distingue trois voyelles courtes opposées à trois voyelles longues. la durée d'une voyelle longue est environ double de celle d'une voyelle courte. Ces voyelles sont caractérisées par la vibration des cordes vocales et sont réparties comme suit :

- les voyelles courtes: [i], [a], [u] sont représentées dans un texte voyellé au-dessus ou au-dessous de la consonne, (ِ . َ . ُ), exemple : تُرِكَ[turika] ;
- les voyelles longues [huruuf al madd] : [aa], [uu], [ii] sont écrites sous forme de caractères consonantiques (ي, و, ا) et sont obligatoirement représentées dans un texte écrite .exemple :

مُسَافِرُونَ [musaafiruuna].

1.5.1.2 Les consonnes

Les consonnes de l'Arabe peuvent être classées suivant plusieurs critères

Tableau 1.1: Transcription Orthographique Phonétique de l'AS [7]

Mode	Type de phonème	Phonèmes Arabes	Transcription Arabisante	Lieux d'articulation	
Occlusives	Voisées	ب د	b d	bilabiale alvéodentale	
	Non-Voisées	ق ت ك ء	q t k ,	uvulaire alvéodentale postpalatale glottale	
	Voisée	Emphatiques ظ	<u>d</u>	alvéolaire	
	Non-Voisée		t	alvéodentale	
Fricatives	Voisées	ز ذ غ ع ج	z d g , ʃ	sifflante dorsoalvéolaire interdentale uvulaire pharyngale	
	Non-Voisées	س ث ف ش خ ه ح	s t f š h h h	sifflante dentale interdentale labiodentale chuintante palatale vélaire glottale pharyngale	
	Voisée	Emphatiques ص	ʃ	dorseoalvéodentale sifflante	
	Non-Voisée		ʃ	interdentale	
	Nasales	Voisées	م ن	m n	bilabiale alvéodentale
	Liquide	Voisée	ل	l	dentale
Affriquée	Voisée	ج	ǧ	alvéopalatale	
Vibrante	Voisée	ر	r	apicoalvéolaire	
Semi-voyelles	Non-Voisées	و ي	w y	bilabiale palatale	

1.5.2 Particularités de l'Arabe Standard

Le système phonétique de la langue arabe diffère de celui des autres langues par la présence de voyelles longues (huruuf al madd), de phénomènes d'emphase et de la gémination. Ces caractéristiques donnent une valeur particulière à cette langue.

1.5.2.1 Voyelles longues

En arabe standard les voyelles longues présentent une caractéristique très importante au niveau sémantique. Par exemple, les deux mots *ḡamal* (chameau) et *ḡamāl* (beauté) ne diffèrent que par l'allongement de la voyelle finale.

Sur le plan articulatoire, il existe une similitude entre les voyelles [i] et [ī], [u] et [ū] cependant une différence existe entre les voyelles [a] et [ā] car la position de la langue est plus basse pour le [a] que pour le [ā].

1.5.2.2 Gémination

Au niveau graphique, la gémination est symbolisée par le signe de la chadda qui signifie le dédoublement de la consonne. Sur le plan phonétique, l'opposition simple/géminée peut se résumer de la manière suivante : pour une consonne non-occlusive, l'opposition se réduit essentiellement à l'opposition temporelle brève/longue ; pour une occlusive, elle réside au niveau de la durée du silence. Ce rallongement entraîne l'accentuation des propriétés de la consonne (augmentation du caractère emphatique). Une consonne géminée est un son unique pour lequel les organes de phonation ne changent pas de position (les lèvres ne se referment pas après le premier [b] dans [kab**ba**ra]). Dans beaucoup de langues, ce phénomène permet de mettre en relief un mot dans son contexte, alors qu'il s'avère être un élément distinctif sur les plans morpho-sémantiques en langue Arabe *حضر* : [hadara] "il a assisté" est différente de *حَضَّرَ* [haddara] "il a préparé" où la deuxième consonne est géminée [4].

1.5.2.3 Emphase

Le mot emphase est habituellement utilisé pour rendre compte de manifestations prosodiques liées à l'accentuation volontaire d'une syllabe. Chez les linguistes arabes, il désigne certaines qualités que possèdent les consonnes :

- l'itbaq : les consonnes qui ont cette qualité sont ص [ṣ], ض [ḍ], ط [ṭ], ظ [ẓ]. Celles-ci sont pressées et produites par la langue élevée vers le palais ;

- le tafkhiim : son contraire est le tarqiiq. Il traduit une expression acoustique grasse et épaisse de certaines consonnes ;
- l'istilaa : cette qualité décrit le mouvement articuloire que fait la langue quand elle meut vers la partie postérieure de la cavité buccale, avec ou sans tafkhiim.

Seules les consonnes ص [ṣ], ض [ḍ], ط [ṭ], ظ [ẓ], possèdent ces trois qualités et sont appelées consonnes *emphatiques* (ou consonnes pharyngalisées). Si nous comparons le français à l'arabe, nous constatons que la différence entre *patte* et *pâte* par exemple est rarement faite en français « standard ». En revanche, cette postériorisation a suscité beaucoup d'intérêt en ce qui concerne l'Arabe. Du fait de sa pertinence au niveau perceptif, la modélisation de l'emphase est primordiale en synthèse de la parole à partir du texte de l'AS. Sa prise en compte passe par l'introduction de nouvelles variantes de voyelles dans les contextes emphatiques. Néanmoins, sa mise en œuvre est directement liée à la technique de synthèse utilisée [1].

1.6 CONCLUSION

Dans ce chapitre nous avons exposé des notions de base sur le traitement de la parole, des spécifications du signal vocal et quelques caractéristiques de la langue Arabe Standard. Les objectifs de ce chapitre sont de définir les notions que nous utiliserons dans notre travail. Cette partie théorique sera complétée dans le chapitre suivant par une étude approfondie des systèmes de synthèse de la parole et ses variantes.

Chapitre II :
Analyse et Synthèse de
la parole

2.1 INTRODUCTION

Dans ce second chapitre nous essayons d'illustrer les différents techniques d'analyse du signal de la parole et pour mieux comprendre le fonctionnement de la production de la parole ainsi que la complexité de ce signal, d'où la multitude de méthodes et techniques existantes dans ce domaine.

2.2 TECHNIQUES D'ANALYSE DU SIGNAL VOCAL

Une fois que le son a été émis par le locuteur, il est capté par un microphone. Le signal vocal est ensuite numérisé à l'aide d'un Convertisseur Analogique/Numérique. Comme la voix humaine est constituée d'une multitude de sons, souvent répétitifs, le signal peut être compressé pour réduire le temps de traitement et l'encombrement en mémoire. Ainsi comme prétraitement (Fig.2.1), nous échantillons et préaccentuons le signal vocal. Pour les techniques de reconnaissance, d'analyse ou de synthèse de la parole, la fréquence d'échantillonnage peut varier de 08 jusqu'à 16 kHz. Le filtre de préaccentuation de transmittance $H(z)$

$$H(z) = 1 - a.z^{-1} \quad \text{avec : } a=0.95 \quad (2.1)$$

Qui est souvent non récursif de premier ordre, permet d'égaliser les aigus toujours plus faibles que les graves. Aussi et vu qu'il est non stationnaire, nous réalisons un fenêtrage avec une fenêtre glissante ; chaque trame couvrant une durée de 20 à 30 ms sur laquelle le signal est supposé quasi-stationnaire. Le pas d'analyse entre deux trames successives est de l'ordre de quelques dizaines de ms.

Le découpage du signal en trames produit des discontinuités aux frontières des trames, qui se manifestent par des lobes secondaires dans le spectre. Pour compenser ces effets de bord, nous multiplions en général préalablement chaque tranche d'analyse par une fenêtre de pondération de type fenêtre de Hamming notée $W(n)$ [5].

$$W(n) = \begin{cases} 0.45 + 0.46 \cdot \cos(\pi n / (n-1)) & n \in [0, \dots, n-1] \\ 0 & \text{ailleurs} \end{cases} \quad (2.2)$$

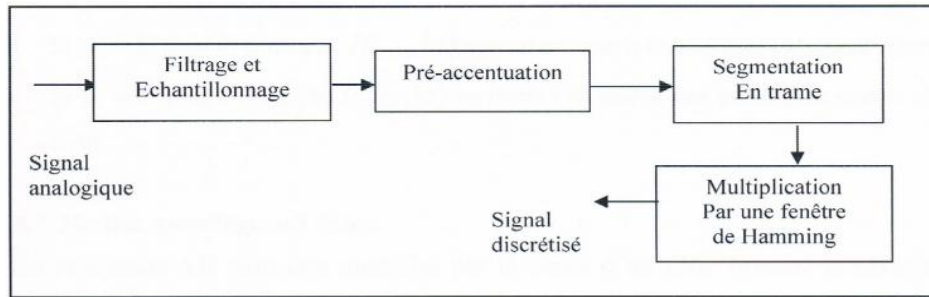


Figure 2.1 : Prétraitement du signal vocal

Le signal vocal peut être analysé soit, en tenant compte des mécanismes de production en utilisant les méthodes paramétriques, soit en utilisant les méthodes non paramétriques.

Dans la plupart des méthodes d'analyse vocale, nous supposons que le signal de parole est localement stationnaire car les propriétés de ce signal varient très doucement en fonction du temps, d'où le recours aux méthodes d'analyse à court terme. Ainsi de courts segments de la parole sont analysés, on les appelle les trames d'analyse temporelle.

Les mesures comme l'énergie, le **Taux de Passage par Zéro (TPZ)** et la fonction d'autocorrélation font partie des méthodes temporelles.

Les coefficients les plus utilisés en RAP sont les cepstres. Ils peuvent être extraits de deux façons : soit par l'analyse paramétrique, à partir du Codage Linéaire Prédictive ou **LinearPrédictive Coding(LPC)**, soit par l'analyse spectrale.

2.2.1 Méthodes non paramétriques

Le signal de parole peut être analysé dans le domaine temporel ou dans le domaine spectral par des méthodes non paramétriques, sans faire l'hypothèse d'un modèle pour rendre compte du signal observé. Les méthodes spectrales sont fondées sur la décomposition fréquentielle du signal sans connaissance a priori de sa structure fine. Une analyse spectrale du signal permet de mettre en évidence certaines caractéristiques de la production de la parole qui peuvent contribuer à l'identification phonétique. L'articulation des phonèmes a une influence directe sur la forme du conduit vocal et des cavités, et donc sur les résonances qui apparaissent dans l'enveloppe du spectre.

L'analyse fréquentielle de la parole se ramène aux opérations de la **Transformée de Fourier (TF)** et n'a d'intérêt que si elle s'applique à une période du signal vocal, donc sur une période assez courte.

En RAP, il est important de connaître l'évolution de ce spectre dans le temps. Actuellement, les spectres sont obtenus numériquement par la Transformée de Fourier Discrète (TFD), en particulier grâce à l'algorithme de la Transformée de Fourier Rapide (TFR) ou FastFourier Transform(FFT). Cependant, le nombre de paramètres spectraux calculés sur une trame par FFT reste trop élevé pour un traitement automatique ultérieur. Pour une analyse très fine de la parole, la fenêtre de Hamming est déplacée à chaque fois de 128 points environ 10 ms (Fig.2.2) [5].

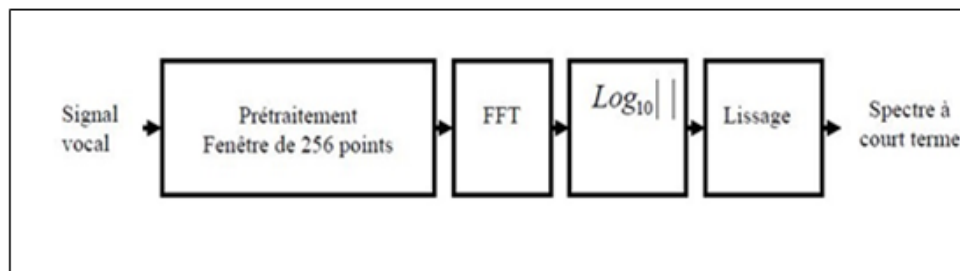


Figure 2.2 : Analyse numérique du signal parole par FFT

2.2.2 Méthodes paramétriques

Les méthodes paramétriques appelées aussi méthodes d'identification sont fondées sur une connaissance des mécanismes de production de la parole (Exemple : le conduit vocal). Les plus utilisées sont celles basées sur l'analyse prédictive linéaire et l'analyse cepstrale. L'hypothèse de base est que le canal buccal est constitué d'un tube cylindrique de section variable. L'ajustement des paramètres de ce modèle permet de déterminer à tout instant sa fonction de transfert. Cette dernière fournit une approximation de l'enveloppe du spectre du signal à l'instant d'analyse. Ces méthodes consistent à ajuster un modèle aux données observées. Les paramètres du modèle, en nombre faible, caractérisent le signal, nous pouvons ainsi injecter des connaissances a priori sur le processus physique qui a engendré ce signal .

Les avantages de cette approche sont la souplesse de l'analyse, l'introduction naturelle de l'information et les choix variés des espaces de représentations paramétriques. Dans le cas de la modélisation du signal parole, nous n'avons accès qu'à une seule sortie du système alors que l'entrée n'est pas mesurée. Il en résulte un problème d'estimation non linéaire car nous ne disposons pas d'observation de l'onde glottique d'excitation. En conséquence, nous en

sommes limités à faire quelques hypothèses relativement neutres sur l'entrée ; par exemple, bruit blanc à moyenne nulle et reporter tout l'effort de modélisation sur le système [5].

2.3 CODAGE PREDICTION LINEAIRE

La prédiction linéaire est une technique importante pour la compression d'un signal de parole en modélisant le conduit vocal par un filtre dont les coefficients sont déterminés au biais de l'analyse de la redondance du signal, cela est réalisé par la prédiction linéaire qui exploite cette redondance pour prédire un échantillon par une combinaison linéaire d'échantillons antérieurs, c'est l'idée de base du codage par prédiction linéaire LPC.

Lors de l'analyse à court terme, la redondance proche entre les échantillons du signal de parole est supprimée par un filtre d'analyse LP, représentant le conduit vocal. Ce filtre permet d'extraire la structure des formants du signal d'entrée et d'obtenir un signal de sortie de faible énergie qui correspond à l'erreur de prédiction appelée signal résiduel ou d'excitation. Le filtre inverse d'analyse est le filtre de synthèse LP, dont la fonction transfert décrit l'enveloppe spectrale du signal de la parole, il génère le signal de parole synthétisé.

Chaque trame de parole est donc modélisée en sortie du système linéaire par un signal d'excitation. Un meilleur codage de celui-ci pourrait être obtenu en utilisant un prédicteur à long terme qui prendra en compte la corrélation entre les échantillons distants du signal de parole. L'extraction de cette périodicité est obtenue par un estimateur du pitch. Cette analyse n'aura aucun effet sur les sons non voisés [8].

Dans cette présente étude nous avons utilisé la méthode par prédiction linéaire pour l'extraction des formants. Nous allons donc présenter cette méthode afin d'avoir une idée sur la façon avec laquelle nous avons calculé les formants. Le « Linear Predictive Coding » ou LPC repose sur un modèle simple décrivant le Comportement des organes vocaux lors de la synthèse d'un son. Ce modèle a été conçu à partir d'un modèle mathématique appelé « modèle autorégressif ». Nous allons donc commencer par présenter le modèle autorégressif afin de mieux comprendre le modèle LPC [7].

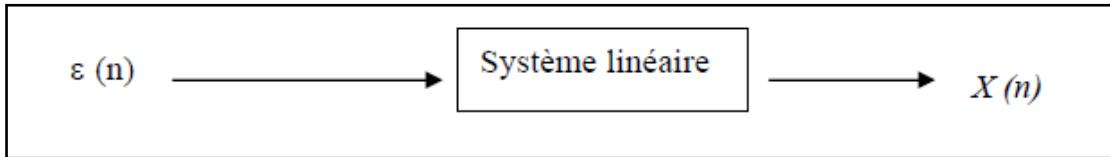
2.3.1 Le modèle autorégressif (AR)

Un processus AR peut être modélisé par la sortie d'un filtre linéaire et invariant dans le temp dont l'entrée est un bruit blanc $e(n)$ et tel que sa valeur à l'instant n ne dépend que de l'entrée $e(n)$ au même instant et des sorties aux instants précédents.

Son équation aux différences s'écrit :

$$x(n) + \sum_{i=1}^M a_i x(n-i) = e(n) \quad (2.1)$$

Sa fonction de transfert s'écrit comme suit :



$$H(Z) = \frac{X(Z)}{e(Z)} = \frac{1}{\sum_{i=1}^M a_i Z^{-i}} = \frac{1}{A(Z)} \quad (2.2)$$

Notons que $H(z)$ ne contient alors que des pôles et c'est pour cette raison que ce modèle est aussi appelé modèle tous pôles. La condition de stationnarité du signal $x(n)$ est équivalente à la condition de stabilité du filtre est que celle-ci n'est assurée que si les racines des polynômes $H(z)$ (donc les coefficients $a(i)$ du filtre), sont de modules inférieurs à 1 [7].

2.3.2 Modèle AR et modèle de prédiction linéaire

Si l'entrée d'un modèle autorégressif est inconnue alors nous pouvons estimer ce dernier par un modèle ayant les mêmes propriétés appelées modèle de prédiction linéaire. Son équation aux différences $\hat{X}(n)$ où est l'estimé de $X(n)$, s'écrit alors :

$$\hat{X}(n) = - \sum_{i=1}^M \hat{a}_i X[n-i] \quad (2.3)$$

\hat{a}_i les estimés des coefficients $a(i)$ du filtre AR ;

P : ordre de prédiction.

On note $e(n)$ l'erreur de prédiction définie comme suit :

$$e[n] = X[n] - \hat{X}[n] = - \sum_{i=1}^M \hat{a}_i X[n-i] \quad ; a(0)=1. \quad (2.4)$$

2.3.3 Pourquoi utilise-t-on le modèle autorégressif

On peut assimiler le mécanisme phonatoire à un système de transmittance :

Avec :

$$H(Z) = \frac{G}{A(Z)} \quad (2.5)$$

$$A(Z) = \sum_{i=1}^p a_i z^{-i}$$

$a(0)=1$ et $A(z)$ est un polynôme qui s'écrit comme suit:

$$x(n) + \sum_{i=1}^M a_i x(n - i) = G U(n) \tag{2.6}$$

G : le gain de ce système

Dans le domaine temporel :

$$X(z) = U(z) \cdot H(z) \tag{2.7}$$

Ce modèle de production d'un signal est appelé AR (autorégressif) avec :

Si on suppose que notre système est excité par une excitation $U(n)$ qui se présente comme :

- des **sons voisés** (ou sonores): l'excitation est un train périodique d'impulsions ;
- des **sons non voisés**: l'excitation est un bruit blanc centré (de moyenne nulle et variance nulle) (Figure 2.3).

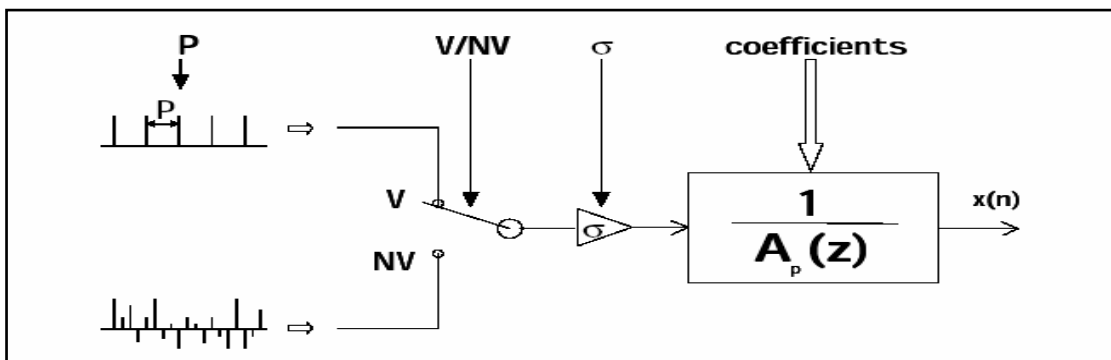


Figure 2.3 : Modèle de production de la parole [7]

La transmittance $H(z)$ est celle d'un filtre polynomial, On définit le filtre inverse dont la transmittance est définie par :

$$A(Z) = \sum_{i=1}^p a_i z^{-i} \quad a(0)=1 ; \tag{2.8}$$

Ce filtre excité par le signal original, engendre en sortie l'erreur de prédiction [7].

2.3.4 Estimation des coefficients de prédiction linéaire

Le critère usuel pour l'estimation des coefficients du modèle de prédiction est la minimisation de l'erreur quadratique de ce dernier ou de la variance. La variance est défini sous la forme suivante

$$\sigma_e^2 = R_e(0) = \sum_{i,j=0}^p a(i) a(j) \overline{X(n-i) X(n-j)} \tag{2.9}$$

$$\sigma_e^2 = \sum_{i,j=0}^p a(i) a(j) R_x(i-j) \tag{2.10}$$

La minimisation par rapport aux coefficients $a(i)$, nous mène a calculé la dérivé partielle suivante par rapport à $a(i)$:

$$\frac{\partial \sigma_e^2}{\partial a(i)} = \sum_{j=0}^p R_x(i,j) a(j) = 0 \tag{2.11}$$

$$\sum_{j=0}^p R_x(i-j) a(j) = -R_x(i) \tag{2.12}$$

L'autocorrélation est l'une des méthodes les plus utilisées des **LPC**, la variance de l'erreur de prédiction, sous forme quadratique :

$$\sigma_e^2 = [1 \quad a] \cdot R_{xx}^p \cdot \begin{bmatrix} 1 \\ a \end{bmatrix} \tag{2.13}$$

$$a = [1, a(1), a(2), \dots, a(p)] \tag{2.14}$$

La méthode d'autocorrélation assure la stabilité du modèle AR et conduit à un système de matrice de Toeplitz (symétrique et égalité des éléments diagonaux de la matrice) qui s'écrit :

$$R_{xx}(p) = \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \dots & R_{xx}(p) \\ R_{xx}(1) & R_{xx}(2) & \dots & R_{xx}(p-1) \\ \vdots & \vdots & \dots & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \dots & R_{xx}(0) \end{bmatrix} \tag{2.16}$$

$$R_x = [R_{xx}(1), R_{xx}(2), R_{xx}(3), \dots, R_{xx}(P)] \tag{2.17}$$

On écrit dans ce cas :

$$R_{xx}(p) = \begin{bmatrix} R_{xx}(0) & R_x \\ R_x & R_{xx}(p-1) \end{bmatrix} \quad (2.18)$$

On aura donc d'après la forme quadratique de la variance de l'erreur [10] :

$$\sigma_e^2 = \sigma_e^2 + 2.a.R_x + a R_{xx}^{p-1} a \quad (2.19)$$

$$R_{xx}^{p-1} a = -R_x \quad (2.20)$$

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \dots & R_{xx}(p) \\ R_{xx}(1) & R_{xx}(2) & \dots & R_{xx}(p-1) \\ \vdots & \vdots & \dots & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \dots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \quad (2.21)$$

Le calcul des coefficients de prédiction a_i se résume donc à inverser la matrice d'autocorrélation R_{xx} . c'est une matrice de Toeplitz symétrique et le système découlant de l'équation matricielle (2.20) est un système de Yule-Walker, l'inversion de la matrice R_{xx} est une opération qui nécessite un calcul fastidieux et long, un algorithme efficace peut être employé pour déterminer les coefficients LP. C'est l'algorithme de Levinson-Durbin [8].

2.4 CONCLUSION

Les généralités sur la parole citées dans le « chapitre I » nous ont permis de mieux comprendre le fonctionnement de la production de la parole ainsi que la complexité de ce signal, d'ou la multitude de méthodes et techniques existantes dans ce domaine. Notons que les différentes méthodes numériques de traitement citées dans cette partie ont toutes été utilisées dans cette étude.

Chapitre III :
Analyse acoustique du
corpus

3.1 INTRODUCTION

Dans ce chapitre, nous intéressons à l'analyse acoustique des sons pour étudier des caractéristiques de notre signal vocal (fréquence fondamentale, formant, intensité) et puis une analyse avec codage prédiction linéaire (**LPC**) en exposant les grandes lignes introduites dans les étapes de l'élaboration de notre outil d'analyse.

3.2 L'OUTIL D'ANALYSE

Nous pouvons citer quelques outils qui permettent de visualiser la forme d'onde et le spectrogramme d'un signal ou de parole, d'éditer et d'aligner des transcriptions orthographiques et phonétiques sur ce signal, tels que PRAAT, CLAN, Speech Analysis, Goldwave, Cool Edit , etc.

3.2.1 Le logiciel Praat

Le logiciel Praat a été développé par Paul Boersma et par David Weenink de l'Institut de Phonétique d'Amsterdam. Il est un logiciel d'analyse et de transcription phonétique (spectre, intonation, intensité, etc.).

Le logiciel comporte aussi des fonctionnalités importantes pour l'enregistrement, pour la manipulation et pour la synthèse de sons, pour la création d'algorithmes d'apprentissage, pour l'analyse statistique, ainsi que pour diverses expériences auditives. Praat est hautement portable, configurable et programmable. En linguistique interactionnelle, le logiciel est utilisé pour divers types de transcription alignée de données sonores (éventuellement extraits d'une vidéo), pour aligner des transcriptions déjà réalisées en texte brut, mais aussi pour l'analyse et la transcription orthographique phonétique. Avec ce logiciel, il est possible :

- d'enregistrer des fichiers audio qui pourront ensuite être analysés.
- de transcrire, d'étiqueter et de segmenter des données audio (que les enregistrements aient été effectués sous Praat ou qu'ils proviennent d'autres fichiers, au format WAV, par exemple).
- D'effectuer des analyses phonétiques et acoustiques au niveau segmental (spectrogramme, analyse de formants, sonagrammes, etc.) et au niveau suprasegmental (pitch ou Fo, intensité et durée).
- De manipuler et modifier le signal de parole (utilisation de filtres, modification des contours intonatifs et de la durée, etc.).

- de faire de la synthèse de la parole (créer des stimuli audio, synthèse articulatoire, analyse -synthèse de données modifiées, etc.).
- de faire des analyses statistiques à partir des études phonétiques (analyses de covariances, etc.). Nous pouvons résumer les fonctionnalités de ce logiciel dans la Figure 3.1 [9].

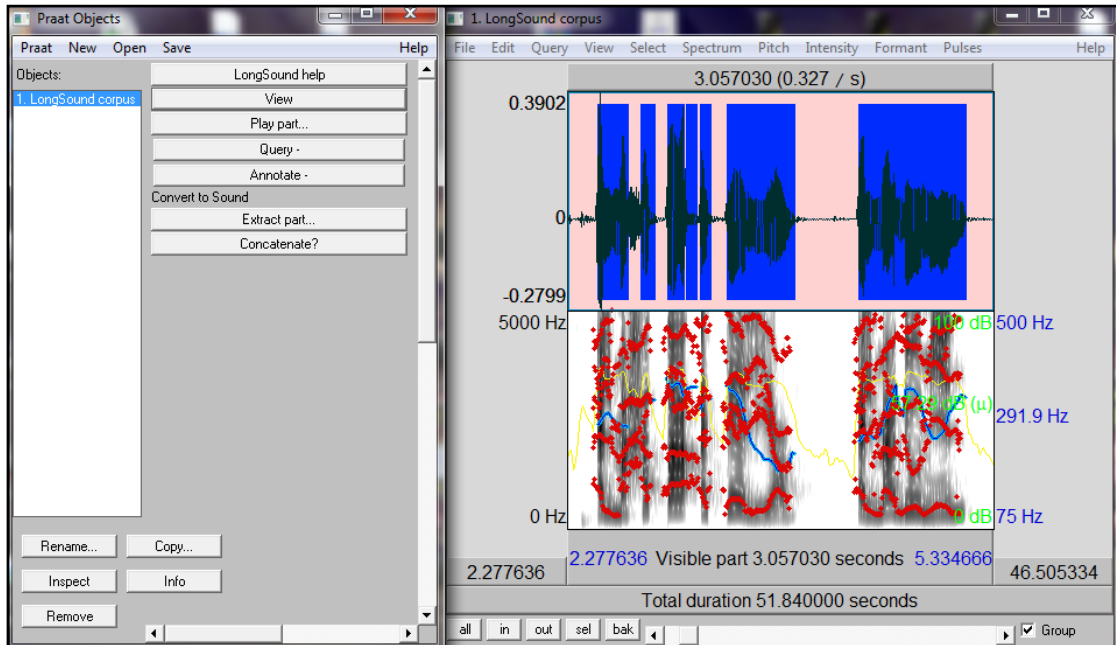


Figure 3.1 : Présentation du logiciel Praat

3.2.2 Visualisations par spectrogramme

On peut aussi visualiser différentes courbes en surimpression sur le spectrogramme :

- la fréquence fondamentale : cochez « Show pitch » dans le menu « Pitch », Sa valeur moyenne (Hz) s'affiche à droite ;
- les formants : cochez « Show formants » dans le menu « Formant », et ils apparaissent en pointillés rouges. Pour les afficher sur toute la longueur de la fenêtre, affichez la fenêtre « Formant Settings » du menu « Formant », et dans le champ « Maximum de la ration », entrez la durée de la fenêtre, en secondes, à la place de la valeur initiale,
- Les périodes du signal sonore : cochez « Show pulses » dans le menu « Pulses ». Chaque période est représentée, sur l'enveloppe, par un trait bleu vertical.

3.2.3 Analyse par spectrogrammes

Le spectrogramme est un outil de visualisation utilisant la technique de la Transformée Rapide de Fourier et par conséquent du calcul de spectres. Il a commencé à être largement utilisé en 1947, à l'apparition du sonographe, et est devenu l'outil incontournable des études en phonétique pendant de nombreuses années [10].

L'apparition de l'informatique puis d'écrans graphiques de bonne qualité a permis d'abandonner tout matériel comme le sonographe mais la technique du spectrogramme est encore aujourd'hui largement utilisée dans de nombreux domaines, du fait de sa simplicité de mise en œuvre et des résultats intéressants qu'elle procure. On parle de spectrogramme à larges bandes ou à bandes étroites selon la durée de la fenêtre de pondération. Les spectrogrammes à bandes larges sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10 ms). Ils mettent en évidence l'enveloppe spectrale du signal et permettent de visualiser l'évolution temporelle des formants. Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres. Les spectrogrammes à bandes étroites sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bandes horizontales [9]. Avec un axe des abscisses de temps en millisecondes et des ordonnées des fréquences en Hz et l'intensité est donné par le degré de noirceur de la trace. (Fig.3.2).

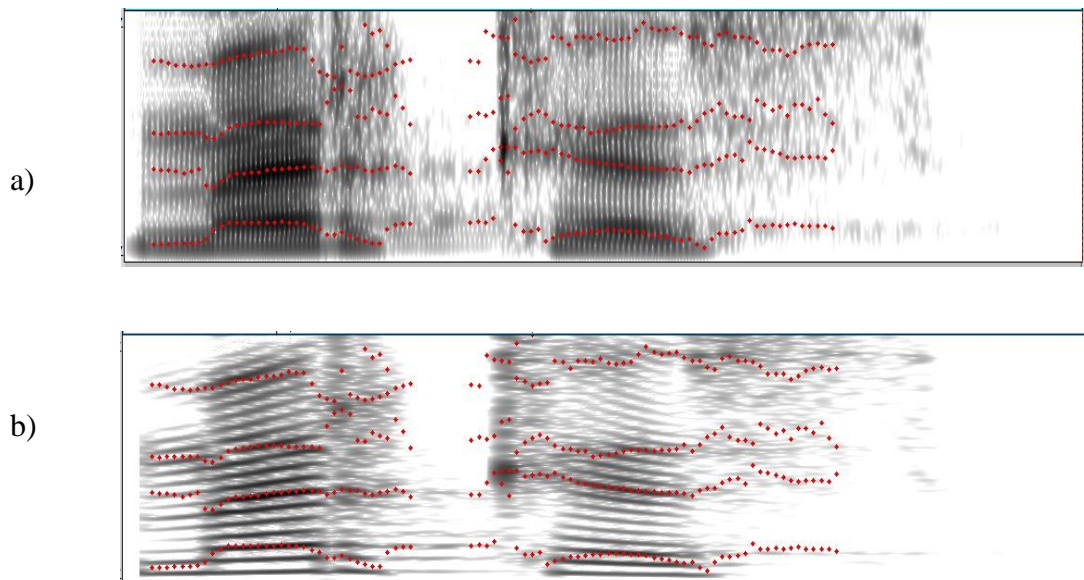


Figure 3.2 a, b: Spectrogramme de la phrase /المحطة التالية/ [ʔlmaḥaʦ ʦa ʔtaaliʒa] :

- a) en bande large avec une fenêtre de Hamming de 5ms ;
- b) en bande étroite avec une fenêtre de Hamming de 30ms.

3.2.4 Lecture de spectrogramme

La lecture de spectrogramme contient 4 étapes élémentaires :

Étape 1 : Connaître les 3 dimensions du spectrogramme. Ce sont l'énergie (l'intensité), le temps et la fréquence du spectre ;

Étape 2 : Savoir distinguer les consonnes et les voyelles :

- les consonnes sont des sons produits avec une constriction plus ou moins forte dans le conduit vocal. L'intensité du spectre est relativement faible et sur le spectrogramme sa noirceur n'est pas très forte ;
- alors que les voyelles sont des sons produits sans aucune constriction forte dans le conduit vocal, l'intensité du spectre est relativement élevée et sur le spectrogramme sa noirceur est relativement foncée.

Étape 3 : Savoir reconnaître les grandes classes de consonnes. Il y a 3 types de Consonnes, les occlusives, les fricatives et les sonantes :

- les occlusives sont produites par une occlusion complète dans le conduit vocal, donc pendant l'occlusion, l'air ne passe pas et sur le spectrogramme, il correspond à un silence (sauf le voisement pour les sonores) ;
- les fricatives sont produites avec une forte constriction (mais pas complète) dans le conduit vocal. Il y a une turbulence de l'air dans le conduit vocal et sur le spectrogramme cette turbulence correspond au bruit de friction ;
- les sonantes /m, n, l, R/ sont produites avec une constriction partielle dans le conduit nasal et vocal. L'air passe d'une façon relativement libre et sur le spectrogramme il y a des formants comme les voyelles, mais ces formants sont moins forts que ceux des voyelles ;
- il y a deux types pour les occlusives et les fricatives : sourdes et sonores. Pour les occlusives et les fricatives sonores, les cordes vocales vibrent alors sur le spectrogramme ils présentent une barre de voisement. Tandis que, les cordes vocales des occlusives et des fricatives sourdes ne vibrent pas, donc sur le spectrogramme il n'y a pas de barre de voisement.

Étape 4 : Savoir reconnaître les grandes classes de voyelles. Les voyelles se différencient les unes les autres par leurs formants. Un formant est la zone de fréquence où il y a une concentration (renforcement) d'énergie. Dans les voyelles orales, il y a en moyenne un formant par 1000 Hz (voix d'Homme). On utilise souvent le spectrogramme à bande large pour visualiser les formants et ces derniers y apparaissent sous les formes des bandes noires horizontales. Les voyelles orales sont divisées en des classes :

- les voyelles antérieures, la distance entre $F_1 - F_2$ est supérieure à la distance entre $F_2 - F_3$;
- les voyelles postérieures, la distance entre $F_1 - F_2$ est inférieure à la distance entre $F_2 - F_3$;
- les voyelles centrales, les formants sont plus (ou moins) équidistants.

3.3 ANALYSE ACOUSTIQUE DE NOTRE SIGNAL VOCAL

Nous avons fait des analyses acoustiques sur une phrase (**Ph**) en L'Arabe Standard **Ph** : /المحطة التالية/ avec leur : transcriptions orthographiques phonétiques (TOP) est : [ʔlmaħaʔ ʔa ʔtaalija].

Nous avons visualisé la phrase Ph avec Praat ;on obtient 2 parties :

Le première partie : signal audio (forme onde) de Ph ;

Le deuxième partie : un spectrogramme (les traits noirs), les formants (en rouge), l'intensité (en jaune), le pitch (en bleu) de la phrase Ph (Fig.3.3).

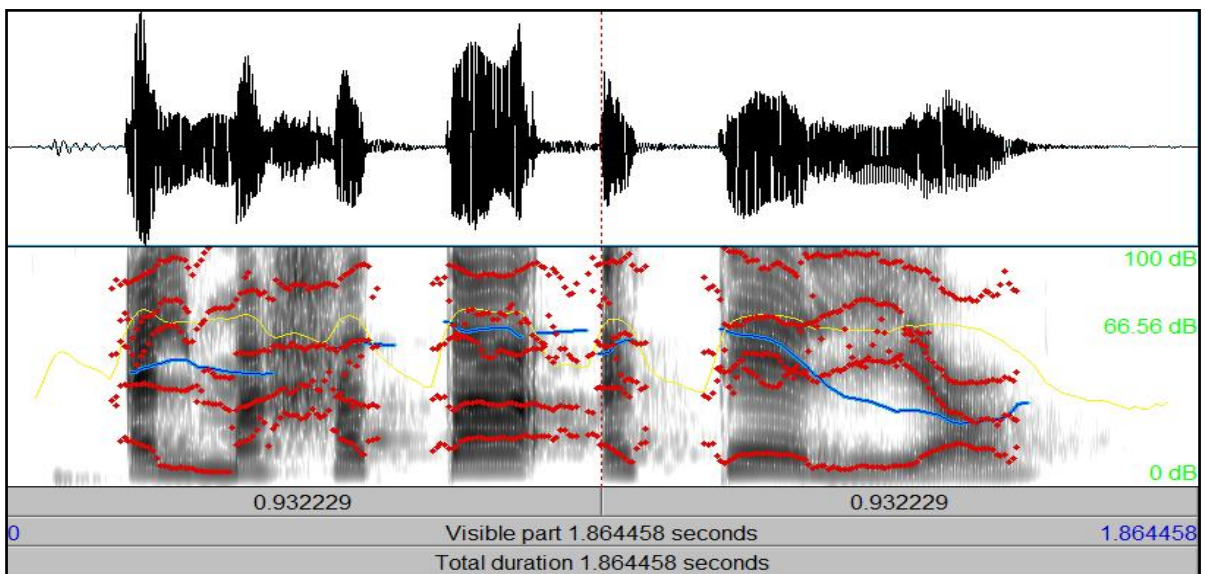


Figure 3.3 : Visualisation de Ph avec Praat

3.3.1 Analyse par sonagrammes

Nous avons fait une analyse sonagramme sur la phrase Ph, c'est-à-dire une analyse générale qui contient des analyses fréquentielles, formantiques, l'intensité.

3.3.1.1 analyses formantiques

Nous pouvons avec Praat d'extraire les valeurs moyennes des formants (F_1, F_2, \dots, F_5) pour la phrase Ph (Tableau 3.1).

Tableau 3.1 : les valeurs moyennes de formants de la phrase Ph

	F_1	F_2	F_3	F_4	F_5
Formant [Hz]	814.93	1900.26	2784.17	3819.41	4670.51

3.3.1.2 Analyse de l'intensité

Nous analysons l'intensité, puis nous extrayons les propriétés de la phrase Ph l'intensité Min-Max-Moyenne (Fig.3.4, Tableau 3.2).

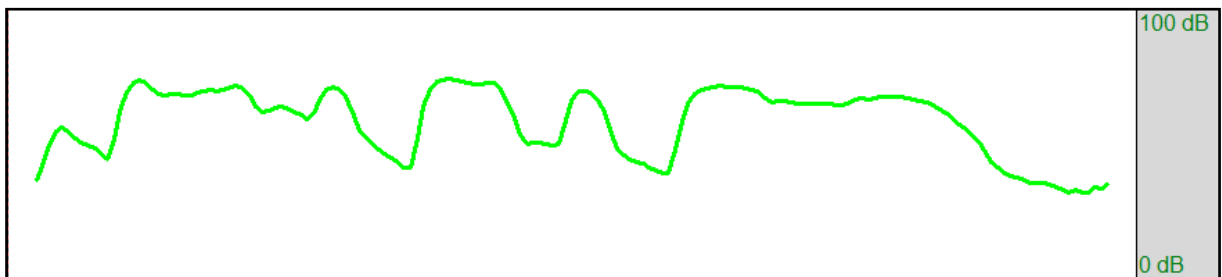


Figure 3.2 : diagramme de l'intensité de la Ph

Tableau 3.2 : les valeurs de l'intensité Min-Max-Moyenne phrase Ph

	Min	Max	Moy
Intensité en [dB]	32.68	74.45	59.89

3.3.1.3 Analyses fréquentielles

Nous avons fait l'analyse par visualisation de la variation de pitch (de mélodie) pour la phrase Ph, puis nous citons les zones des sons voisées et non voisées (Fig. 3.3).

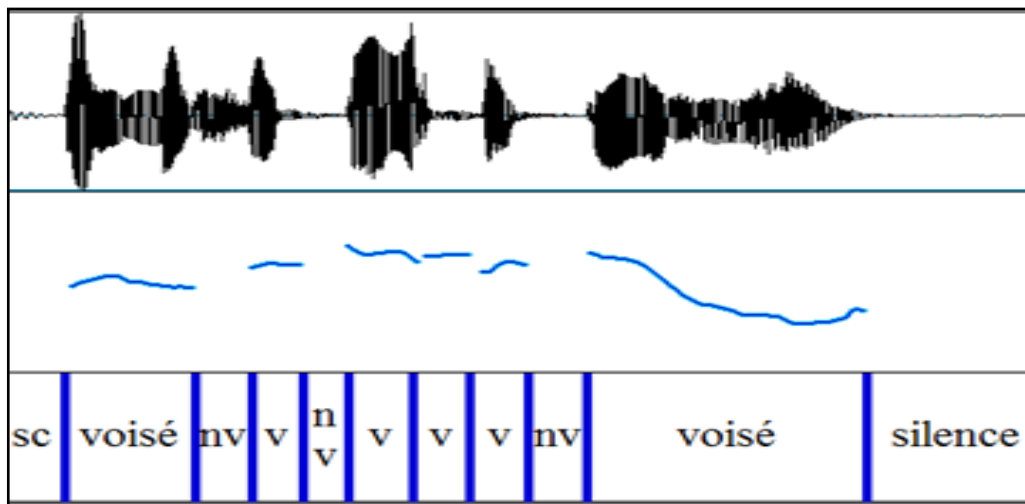


Figure 3.15 : la mélodie de la phrase Ph

Avec : V : voisé, NV : non voisé, SC : silence.

Nous avons présentées les paramètres de la variation de fréquence fondamentale de la phrase Ph dans le tableau 3.3.

Tableau 3.3 : les valeurs de pitch Min, Max, Moyenne de la phrase Ph

	F_0 MIN	F_0 MAX	F_0 moyenne
Pitch en [Hz]	188.08	368.72	288.51

3.3.2 Analyse par LPC

Nous extrayons les coefficients a_i et le Gain de la phrase Ph avec un ordre de prédiction $M=12$; plus M augmente plus l'enveloppe spectrale est semblable à celle du signal original, l'ordre est le résultat d'un compromis entre une bonne représentation de la structure formantique et la complexité de calcul. Pour satisfaire ce compromis, l'ordre est choisi généralement de 8 à 16.

A l'aide du Praat, nous pouvons extraire les coefficients a_i et le Gain du Codage Prédiction Linéaire (LPC) avec la méthode d'autocorrélation (Tableau 3.4).

CHAPITRE 3 : ANALYSE D'UN SIGNAL VOCAL PAR SONAGRAMMES ET LPC

Tableau 3.4 : les valeurs de coefficients a_i avec la méthode d'autocorrélation (LPC)

Les valeurs des coefficients et de Gain par LPC	
Gain	0.000800793
a [1]	-1.775106518
a [2]	1.711976918
a [3]	-1.53038505
a [4]	1.157291496
a [5]	-0.44257713
a [6]	0.322570196
a [7]	-0.3478578
a [8]	0.37187364
a [9]	-0.323424
a [10]	0.0907664
a [11]	0.0722829
a [12]	0.063456388

3.4 CONCLUSION

Le but de ce chapitre est de faire une analyse acoustique de notre phrase en passant par plusieurs étapes d'analyse et visualisation et qui a été terminé avec une analyse par prédiction linéaire (LPC), les résultats obtenu nous ont permis de confirmer que la phrase est bien analysée.

CONCLUSIONS GENERALES ET PERSPECTIVES

Les analyses acoustiques que nous avons réalisées tout au long de ce **PFE** nous ont permis d'apprendre des notions sur le traitement de la parole. Ils nous ont permis également de comprendre les techniques d'analyses de la parole.

Nous avons tout d'abord effectuée des études générales sur la parole puis sur l'Arabe Standard, pour cela nous avons choisi une phrase pour faire notre étude. L'étude de cette phrase passe par plusieurs étapes d'analyse acoustiques et de visualisations qui nous ont permis une extraction des paramètres pertinents et acoustiques du signal vocal, et des analyses par codage prédiction linéaire.

Comme suite à ce travail, il serait très intéressant de faire des améliorations sur l'analyse par LPC et surtout faire des applications sur codage prédiction linéaire comme synthèse de parole par LPC ;codage de parole.

6. REFERENCES BIBLIOGRAPHIES

- [1] A.Chentir , Etude de la Microprosodie en vue de la Synthèse de la parole en Arabe Standard , thèse doctorat : Ecole nationale polytechnique d'Alger , 01 Octobre 2009.
- [2] [http : //www.infovisual.info](http://www.infovisual.info).
- [3] J. Cisonni, Modélisation et inversion d'un système complexe de production de signaux acoustiques Application à la voix et aux pathologies, Thèse de Doctorat, Institut Polytechnique De Grenoble, France, Novembre 2008.
- [4] O. Amine, Synthèse de la parole en arabe standard, Mémoire de Magister, ENP, Alger, Algérie, 2011.
- [5] G. Djeghiour, thèse magister : application des réseaux de neurones a la synthèse de la parole en arabe standard, Ecole nationale supérieure des sciences humaines; 2011.
- [6] A. Mohamed, Application des Algorithmes Génétiques au Décodage Acoustico-Phonétique de la parole en Arabe Standard, Thèse de Doctorat, ENP, Alger 2008
- [7] L.Hocine ; Analyse sonographique des consonnes fricatives [s] et [š] et leurs opposées [z] et [ž] en vue de la RAP en Arabe Standard ; Projet de Fin d'Etudes ; Ecole nationale polytechnique d'Alger ; 24 Juin 2007.
- [8] D. Ducassou. Cours d'acoustique. Cours de 2ème année de médecine, Université de Nancy 1, France, 1991.
- [9] R.Gaël , synthèse de la parole a partir de texte ; Ecole nationale de télécommunication (ENST paris) , documentation de archive technique d'ingénieur , 20/10/2012.
- [10] T. Dutoit, introduction au traitement automatique de la parole, Notes de cours / DEC2 ; Faculté Polytechnique de Mons, Première édition ,2000.