République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

ECOLE NATIONALE POLYTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

DÉPARTEMENT D'ÉLECTRONIQUE

Mémoire de Master

Thème :

Apprentissage des Sons Spécifiques de l'Arabe Standard par des Apprenants Etrangers

Encadré par :

Pr GUERTI Mhania

Réalisé par :

Mlle KADRI Meriem

Promotion : Juin 2015

ملخص:

في أطروحة الماستر هذه قمنا بتطوير نظام لتعلم نطق الأصوات الخاصة بالعربية الفصحى، موجه نحو المتعلمين الأجانب. يستند هذا النظام على تحليل الأصوات و استخراج خصائصها في كافة المواضع السياقية الممكنة للحرف، لأجل هذا قمنا بتسجيل قاعدة بيانات خاصة و تجزئتها الى الوحدات المطلوبة ثم تحليلها صوتيا. لتجربة هذا النظام قمنا بمحاكاته عن طريق واجهة تم انشاءها في برنامج MATLAB.

الكلمات المفتاحية: التعلم، الأصوات الخاصة، اللغة العربية الفصحى، قاعدة البيانات، التحليل الصوتي، MATLAB

Résumé :

Dans cette mémoire de master nous avons élaboré un système d'apprentissage des sons spécifiques de la langue Arabe Standard ASSAS, destiné aux apprenants étrangers. Ce système est basé sur l'analyse des sons en différentes positions contextuelles, afin d'extraire ses caractéristiques. Pour cela une étape d'enregistrement de corpus suivie d'une segmentation et analyse acoustique ont été faites. Pour mettre en marche notre système une interface de simulation sous le logiciel MATLAB a été réalisée.

Mots clés : Apprentissage, Sons Spécifiques, Arabe Standard, corpus, analyse acoustique, MATLAB.

Abstract :

In this master thesis, we have developed a learning system of specific sounds of Standard Arabic "ASSAS", intended for foreign learners. This system is based on the analysis of sounds in different contextual positions in order to extract its characteristics. For this, a corpus recording step followed by a segmentation and acoustic analysis were made. To put in use our system a simulation interface under the MATLAB software was performed.

Keywords: Learning, Specific Sounds, Standard Arabic, corpus, acoustic analysis, MATLAB.

إهـــــــــداء

أهدي هذا العمل إلى

رسول الأنام عليه الصلاة والسلام ...

مصدر الحب الذي استمر في تشجيعي والدعاء لي... حبيبتي دومًا أمي "زهرة"

من تعلمت منه معنى الكفاح والصبر... أبي الغالي " أحمد"

زهرات بيتنا أخواتي "هناء" ، "جهينة" ، " أسماء" ، "فاطمة الزهراء" ، "راضية"

أخوالي "حيدر" و "باباعيد" وجميع أفراد عائلة "قادري".

من شاركني ساعات الأمل و أمضيت معهم أحلى الأيام صديقاتي كل بإسمها

خاصة "رشا" ، "سمراء" ، "نجلاء" ، " أمينة" ، "وفاء" ، "صوفيا"

كل زملائي بالمدرسة الوطنية المتعددة التقنيات خاصة دفعة ا لإلكترونيك 2015 و إلى كل من

مر بحياتي و ترك فيها ذكرى و أثرا طيبًا

إلى كل هؤلاء أهدي ثمرة جهدي

مريم

Remerciements

Que Dieu soit loué pour nous avoir permis d'arriver au terme de ce travail.

Je remercie et exprime ma reconnaissance à quiconque ayant allumé une bougie dans le chemin de la science et ayant occupé les tribunes du savoir pour m'éclairer.

Je tiens aussi à exprimer mes remerciements et mon gratitude à :

- Mon promotrice, Pr M. GUERTI pour avoir dirigé ce travail jusqu'à son terme, pour le temps qu'elle m'a consacré et pour ses précieux conseils ; Les Membres du jury, Mr R. ZERGUI et Mr M. MAMRI, pour l'enrichissement de cette recherche à travers leurs bénéfiques remarques et orientations conséquentes ;
- Mlle BETTAYEB pour ses précieuses aides et encouragements ;
- Mr TIDJANI pour l'aide et les conseils qu'il m'avait apporté ;
- Je remercie également tous les enseignants de l'Ecole Nationale Polytechnique, et spécialement ceux des départements des Sciences Fondamentales et de Génie Electrique, pour leur apport en savoir ;
- Je remercie enfin tous mes collègues et amis qui m'ont aidé, de près ou de loin, ne serait-ce qu'à travers leurs encouragements.

Liste des abréviations

API : **A**lphabet **P**honétique **I**nternationale.

ARMA : **A**uto **R**égressif à **M**oyenne **A**justée.

AS : **A**rabe **S**tandard.

ASSAS : **A**nalyseur des **S**ons **S**pécifiques de l'**A**rabe **S**tandard

BD : **B**ase de **D**onnées.

FFT : **F**ast **F**ourier **T**ransform.

LPC : **L**inear **P**redictive **C**oding

RAP : **R**econnaissance **A**utomatique de la **P**arole.

TAP : **T**raitement **A**utomatique de la **P**arole.

TF : **T**ransformée de **F**ourier

TFD : **T**ransformée de **F**ourier **D**iscrète

TFR : **T**ransformée de **F**ourier **R**apide

TPZ : **T**aux de **P**assage par **Z**éro

Liste des figures

Liste des tableaux

Table des matières

Chapitre 1 : Généralités sur la parole

Chapitre 2 : Techniques d'analyse de la parole

Chapitre 3 : Elaboration et Evaluation d' « ASSAS »

Introduction Générale

La parole est un moyen naturel de communication entre les hommes. Les efforts entrepris depuis une soixantaine d'années, pour mettre au point des systèmes de synthèse ou de reconnaissance automatique de la parole, ont obtenu un succès mais ils ont montré aussi la nature complexe de la communication parlée.

Les traiteurs de la parole utilisent plusieurs niveaux de traitement pour la synthèse et la reconnaissance. Le premier niveau est la détermination des caractéristiques du signal de parole lui-même, c'est à dire l'analyse acoustique.

Le but de notre travail est d'élaborer un outil d'aide en apprentissage des sons spécifiques destiné aux apprenants étrangers. Cet outil est un **A**nalyseur des **S**ons **S**pécifiques de l'**A**rabe **S**tandard « ASSAS », basé sur l'analyse de ces sons en différentes positions contextuelle (Initiale, Médiane, Finale). Le système **ASSAS** joue le rôle d'un analyseur et afficheur des caractéristiques de son choisi, via une base de données préconçue.

Pour atteindre notre objectif, nous avons structuré notre travail en trois chapitres :

- dans le premier, nous allons décrire d'une manière générale des notions sur la parole ainsi que sa production, l'appareil phonatoire de l'être humain, des spécifications du signal vocal et des notions fondamentales sur l'Arabe Standard ;
- le deuxième, nous donne une brève définition d'analyse de la parole, puis nous étudions les différentes techniques d'analyse du signal vocal ;
- dans le troisième, nous présentons le système de la « ASSAS », avec une brève explication du processus de construction de notre BD, à l'aide de l'outil d'analyse Praat, et la procédure de fonctionnement de programme ASSAS. En dernier lieu nous finissons par des conclusions générales et perspectives.

1.1 INTRODUCTION

Le but de ce chapitre est de présenter le mécanisme de la production de la parole chez l'être humain, puis de définir les notions fondamentales utilisées dans le domaine du traitement de la parole. Afin de comprendre les différents niveaux de complexité de problème, nous allons expliquer les principales caractéristiques du signal vocal.

1.2 DEFINITION DE LA PAROLE

La parole est définie comme étant l'expression concrète de la langue. C'est un mode propre à l'Homme. [1].

Sur le plan physique, la parole est le résultat d'une variation de la pression produite par l'émission d'un son par un locuteur. Il s'agit d'une onde sonore créée par le passage de l'air expulsé des poumons dans l'appareil phonatoire et articulatoire du locuteur, ce qui provoque une modification de cette onde puis elle se propage dans l'air. La production de la parole est rapide : 150-300 mots/min. 3-5 syllabes/sec. 10-15 phonèmes/sec. [2]

1.3 PRODUCTION DE LA PAROLE

Décrire le processus de production de la parole en vue de spécifier le signal ainsi produit, nécessite l'acquisition de certains nombres de connaissances liées à la complexité du processus de génération et de ses difficultés de mesure.

1.3.1 Appareil phonatoire humain

L'appareil phonatoire humain peut se présenter comme un système source/filtre avec nos poumons comme un réservoir énergétique.

Les fonctions essentielles dans l'acte de parole, ou phonation sont réalisées par trois groupes d'organes :

- **l'appareil respiratoire** : (diaphragme, poumons, trachée), soufflerie qui fournit l'énergie et la quantité d'air nécessaire à la production de sons en poussant de l'air à travers la trachée-artère ;
- **le larynx « l'organe vibrant »** : se trouve au sommet supérieure de trachée-artère, où la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles. **Les cordes vocales** sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée **glotte**. L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation

des sons non-voisés. Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales. Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales, et les force à s'ouvrir, ce qui fait tomber la pression, et permet aux cordes vocales de se refermer (figure 1.1) ;

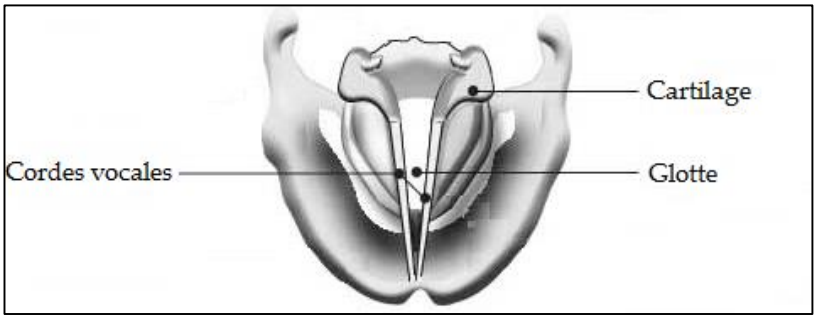


Figure 1.1 : *Le larynx*

- **Le conduit vocal**, formé des cavités résonantes supra-laryngées (pharynx, bouche, nez) où s'effectue l'articulation proprement dite par les changements de forme du conduit vocal. Ces changements résultent surtout des mouvements des lèvres, de la langue, du voile du palais (dont l'abaissement fait intervenir une cavité supplémentaire, les fosses nasales) et de la mâchoire inférieure [3] (figure 1.2).

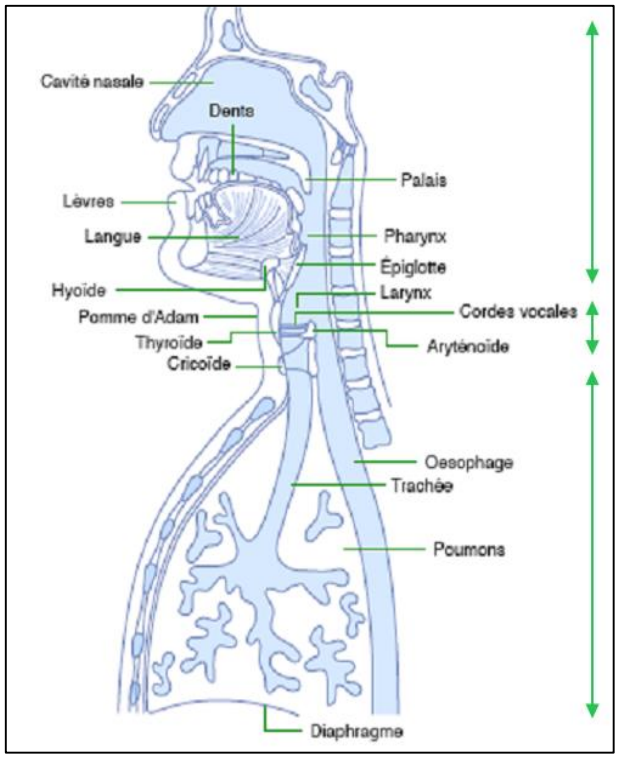


Figure 1.2 : *Appareil phonatoire humain* [4]

Le fonctionnement de cet appareil est déclenché et contrôlé par le système nerveux central du locuteur, après avoir pris la décision de parler, les muscles du diaphragme aident à gonfler et dégonfler les poumons, ces derniers envoient l'air vers le conduit vocal, les cordes vocales font vibrer l'air en provenance des poumons, le nez et la bouche font modifier l'onde sonore pour former le mot à prononcer.

1.3.2 Modélisation Source/Filtre

Le modèle source/filtre, considère le signal de parole s(n) comme le résultat de la convolution du signal glottique e(n) (la source) par un filtre h(n) qui représente le comportement fréquentiel du conduit vocal soit (figure 1.3) :

$$s(n) = e(n) * h(n) \tag{1.1}$$

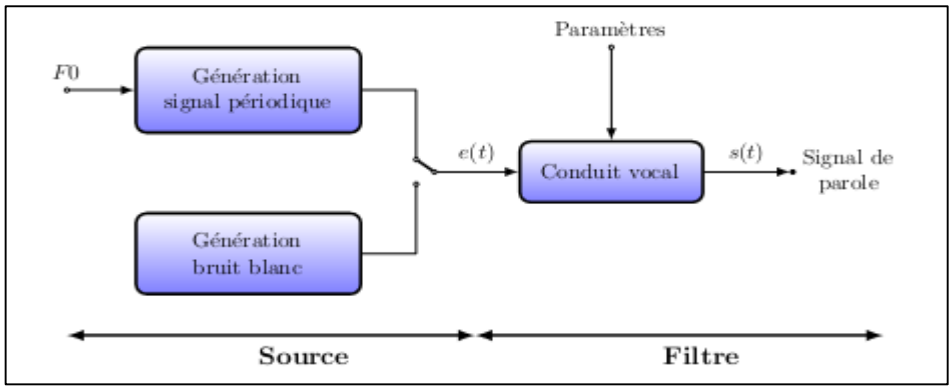


Figure1.3 : *Modélisation Source /Filtre de la parole* [6]

Dans sa représentation la plus simple, ce modèle repose sur deux contraintes fortes [6] :

- le filtre est supposé comme un système linéaire ;
- le filtre et la source sont indépendants.

1.4 CARACTERISTIQUES PHONETIQUES DU SIGNAL DE LA PAROLE

D'un point de vue linguistique, la production des sons ou d'un mot réside dans la production en série de tous les phonèmes constituant ce mot. Ces phonèmes forment les unités phonétiques qui sont classées en voyelles, consonnes et semi-voyelles.

1.4.1 Classification des sons du langage

Il est intéressant de grouper les sons de parole en classes phonétiques, en fonction de leur mode et lieu d'articulation. Dans la cavité buccale, le point d'articulation est l'endroit où se

trouve un obstacle au passage de l'air. D'une manière générale, le point d'articulation est l'endroit où vient se placer la langue pour obstruer le passage du canal d'air (figure 1.4).

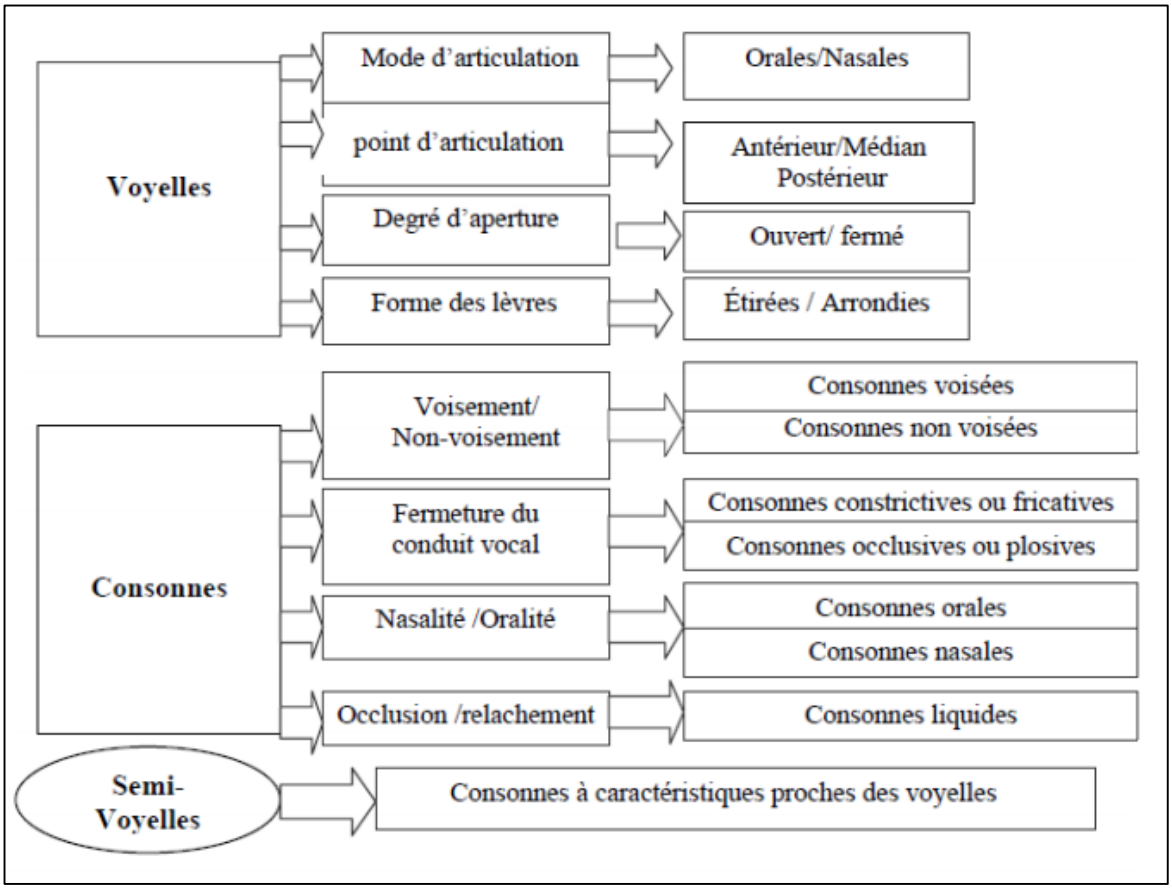


Figure 1.4 : *Classification des sons de langage* [9]

1.4.1.1 Sons voisés

Les sons voisés, tels que les voyelles, semi-voyelles et les consonnes nasales, sont produits par le passage de l'air des poumons à travers la trachée qui met en vibration les cordes vocales. Ce mode, qui représente 80% du temps de phonation, est caractérisé en général par une quasi-périodicité et une énergie élevée (figure 1.5).

1.4.1.2 Sons non voisés

Le second mode d'excitation est obtenu par divers bruits produits par le passage de l'air en un point de resserrement du canal vocal ou par des bruits d'occlusion ou de plosion, provoqués par la fermeture ou l'ouverture des lèvres, ou des chocs de la langue contre le palais. Dans cette catégorie de sons, les cordes vocales ne vibrent pas.

Les consonnes sont un exemple de son non voisé, apériodique. Ces sons sont considérés comme ayant les mêmes caractéristiques que le bruit (figure 1.5).

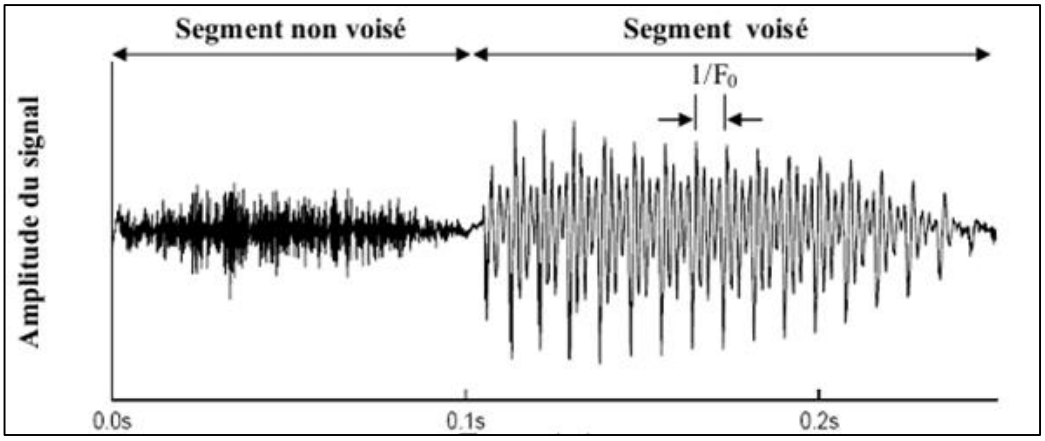


Figure 1.5 : *Représentation temporelle des segments de sons voisés et non voisés* [10]

1.4.1.3 Voyelles

Les voyelles diffèrent de tous les autres sons par le degré d'ouverture du conduit vocal. Quand ce dernier est suffisamment ouvert pour que l'air expiré par les poumons, le traverse sans obstacle, il y a production d'une voyelle. Le rôle de la cavité buccale se réduit alors à une modification du timbre vocalique.

Les voyelles se différencient principalement les unes des autres par leur lieu d'articulation (position de la langue), leur degré d'ouverture (espace compris entre la pointe de la langue et le palais), et leur nasalisation. Nous distinguons ainsi, selon la localisation de la masse de la langue, les voyelles antérieures ou avant, les médianes, les voyelles postérieures (ou arrières), l'écartement entre l'organe, le lieu d'articulation, et selon les voyelles fermées et ouvertes.

Les voyelles orales sont dues à une élévation du palais qui détermine la fermeture des fosses nasales ainsi qu'à l'écoulement de l'air expiré à travers la cavité buccale. Par contre, les voyelles nasales sont caractérisées par l'écoulement d'une partie de l'air à travers la cavité nasale [11].

1.4.1.4 Consonnes

Les consonnes se caractérisent par une fermeture partielle du conduit vocal ou constriction (constrictives ou fricatives) ou totale du conduit vocal (occlusion) : occlusives ou plosives. Nous classons principalement les consonnes en fonction de leur mode d'articulation, de leur lieu d'articulation, et de leur nasalisation. Le mode d'articulation est défini par un certain nombre de facteurs qui modifient la nature du courant d'air expiré :

- intervention ou mise en vibrations des cordes vocales : articulation sonore ;
- fermeture momentanée du passage de l'air suivie d'une ouverture brusque (explosion) : articulation occlusive ;

- rétrécissement du passage de l'air qui produit un bruit de friction : articulation fricative ;
- position abaissée du voile du palais : articulation nasale ;
- contact de la langue au milieu du canal buccal ; l'air sort des deux côtés ;
- une série d'occlusions brèves ; séparées de la luette : articulation vibrante.

La distinction du mode d'articulation conduit à deux classes : les fricatives ou constrictives et les occlusives ou plosives. Les consonnes fricatives appelées également spirantes sont créées par une constriction du conduit vocal au niveau du lieu d'articulation, qui peut être le palais, les dents ou les lèvres. Les fricatives non voisées sont caractérisées par un écoulement d'air turbulent à travers la glotte, tandis que les fricatives voisées combinent des composantes d'excitation périodique et d'autres turbulentes : les cordes vocales s'ouvrent et se ferment périodiquement, mais la fermeture n'est jamais complète. Les consonnes occlusives ou plosives sont reconnues grâce au silence provenant de la fermeture totale du conduit vocal ou occlusion. Cette dernière comporte trois phases :

- l'implosion ou fermeture ;
- l'occlusion proprement dite tenue de la fermeture ;
- l'explosion ou détente.

Les consonnes liquides combinent une occlusion et une ouverture simultanée du conduit vocal. Elles sont caractérisées par un degré de sonorité proche de celui des voyelles. Enfin, les consonnes nasales font intervenir la cavité nasale par abaissement du voile du palais. Elles sont produites par l'écoulement de l'air phonatoire dans le conduit nasal.

1.4.1.5 Les semi-voyelles

Les semi-voyelles, quant à elles, combinent certaines caractéristiques des voyelles et des consonnes. Comme les voyelles, leur position centrale est assez ouverte, mais le relâchement soudain de cette position produit une friction qui est typique des consonnes. Enfin, elles sont assez difficiles à classer [11].

1.4.2 Alphabet Phonétique International

L'**A**lphabet **P**honétique **I**nternational (**API**) associe des symboles phonétiques aux sons, de façon à permettre l'écriture compacte et universelle des prononciations, la figure suivante présente tous les symboles de l'alphabet de l'AS (figure 1.6).

Arabic letter	Buckwalter	Amended SAMPA (original)	IPA
ا	A	aa	a:
ب	b	b	b
ت	t	t	t
ث	v	v	θ
ج	j	j (Z)	ʤ
ح	H	h (X)	ħ
خ	x	x	x
د	d	d	ð
ذ	*	D	X
ر	r	r	r
ز	z	z	z
س	s	s	s
ش	SH	s^ (S)	ʃ
ص	S	S' (s.)	sˁ
ض	D	D' (d.)	dˁ
ط	T	T' (t.)	tˁ
ظ	Z	Z' (z.)	ðˁ
ع	E	E (H)	ʕ
غ	g	g (G)	ɣ
ف	f	f	f
ق	q	q	q
ك	k	k	k
ل	l	l	l
م	m	m	m
ن	n	n	n
ن	n	M (not provided)	ŋ
ن	n	c (not provided)	ɲ
ن	n	e (not provided)	ɲ
ه	h	h	h
و	w	w or uu	w or u:
ي	y	y or ii (j)	j or i:
أ	'	' (?)	ʔ
ـَ	a	a	a
ـُ	u	u	u
ـِ	i	i	i
ـً	F	an	an

Figure 1.6 : *Alphabet de l'AS en Buçkwalter, SAMPA et API* [12]

1.5 PARTICULARITE DE LA LANGUE ARABE STANDARD (AS)

L'Arabe est la langue officielle dans plus de 22 pays, elle est utilisée par 1.62 milliard de musulmans dans le monde. Nous distinguons deux types de langue Arabe : les différents dialectes Arabes utilisés dans chaque pays Arabe et l'Arabe Standard (AS) qui est la langue utilisée dans les cadres officiels, enseignée dans l'école et la langue du Saint Coran.

Le système phonétique arabe contient 40 phonèmes : 26 consonnes, trois voyelles courtes, trois voyelles longues, 2 semi-voyelles et six variantes vocaliques en contexte emphatique. Généralement ces trois voyelles courtes ne sont pas présentées dans l'écriture arabe, ils sont ajoutés avec d'autre signes diacritiques comme : la [*ʃadda*] « gémination ou dédoublement

d'une consonne) » ; le [*suku:n*] « qui désigne que la consonne n'est pas suivie d'une voyelle », pour faciliter la compréhension du contexte Arabe.

L'AS ne possède pas de voyelles nasales. Elles sont représentées sur un plan dont les axes sont les formants F_1 et F_2. Elles tracent alors un triangle dont les extrémités sont occupées par les voyelles [i, u, a]. Ce triangle représente également les positions de la langue dans la cavité buccale selon deux axes : antérieur à postérieur (avant et arrière) et de fermé à ouvert, selon que la langue est massée en avant et vers la zone dentale pour [i], basse et étalée loin du palais pour [a] (ouvert), ou massée postérieurement vers le voile pour [u] dont laquelle les voyelles soulignées sont labialisées (arrondies) [9] (figure 1.7).

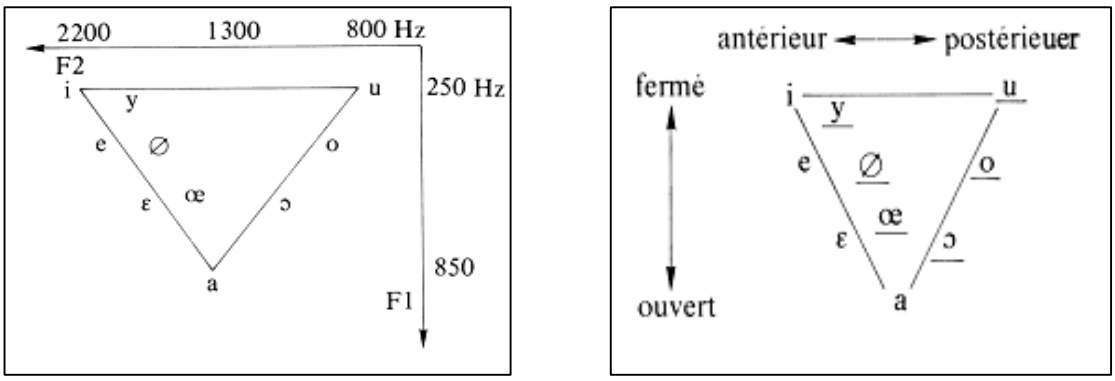


Figure 1.7 : *Triangle vocalique des voyelles « Caractéristiques acoustiques et articulatoires »* [7]

Le tableau suivant (Tableau 1.1) montre les modes et les lieux d'articulation des différentes consonnes et semi-voyelle de l'AS.

Tableau 1.1 : *Classification des consonnes et semi-voyelles de l'Arabe Standard*

Mode	**Type de phonème**		**Phonèmes Arabes**	**Lieux d'articulation**
Occlusives	Voisées		ب	Bilabiale
			د	Alvéodentale
	Non- Voisées		ق	Uvulaire
			ت	Alvéodentale
			ك	Postpalatale
			ء	Glottale
	Voisée	Emphati-	ظ	Alvéolaire
	Non- Voisée	ques	ط	Alvéodentale
Fricatives	Voisées		ز	Sifflante
			ذ	Dorsoalvéolaire
			غ	Interdentale
			ع	Uvulaire
	Non-Voisées		س	Sifflante dentale
			ث	Interdentale
			ف	Labiodentale
			ش	Chuinchante palatale
			خ	Vélaire
			ه	Glottale
			ح	Pharyngale
	Voisées	Emphati-	ص	Dorsealvéodentale sifflante
	Non-Voisées	ques	ض	Interdentale
Nasales	Voisées		م	Bilabiale
			ن	Alvéodentale
Liquide	Voisées		ل	Dentale
Affriquée	Voisées		ج	Alvéopalatale
Vibrante	Voisées		ر	Apico-alvéolaire
Semi-voyelles	Non-Voisées		و	Bilabiale
			ي	Palatale

1.5.1 Les sons spécifiques de la langue AS

L'AS contient 6 sons spécifiques, qui ne sont pas utilisés (ou prononcés) dans une autre langue, ces phonèmes sont représentés suivant le code du tableau 1.2 :

Tableau 1.1 : *Les sons spécifiques de l'Arabe Standard*

Transcription API	[ʔ]	[q]	[dˤ]	[ðˤ]	[ʕ]	[ħ]
Equivalent arabe	أ	ق	ض	ظ	ع	ح

1.6 CARACTERISTIQUES ACOUSTIQUES DU SIGNAL VOCAL

Précédemment, On a défini le signal de la parole comme étant le résultat d'une variation de la pression produite par l'émission d'un son par un système articulatoire.

La phonétique acoustique étudie ce signal en le transformant dans un premier temps en un signal électrique. De nos jours, ce signal électrique résultant, est le plus souvent numérisé, l'opération de numérisation, requiert successivement (figure1.8) :

- **une transduction ;**
- **une préamplification ;**
- **un filtrage de garde** à une fréquence de coupure $\boldsymbol{f_c}$ **;**
- **un échantillonnage** à une fréquence $\boldsymbol{f_e}$ **;**
- **une quantification** avec un nombre de bits **b** et le pas de quantification **q.**

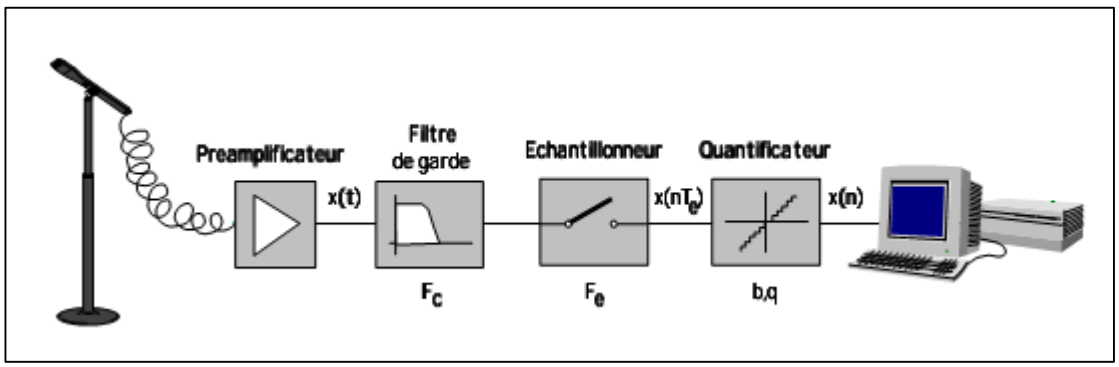


Figure 1.8 : *Enregistrement numérique d'un signal acoustique* [8]

Le signal numérisé peut être alors soumis à un ensemble de traitements statistiques qui visent à en mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre.

1.6.1 Fréquence fondamentale

La fréquence fondamentale $\boldsymbol{F_0}$ est le nombre de vibrations des cordes vocales par seconde au cours de la prononciation d'un son voisé. La gamme de variation moyenne de la fréquence

fondamentale varie d'une personne à une autre en fonction de la longueur et de la masse des cordes vocales de chaque personne, donc elle dépend, essentiellement, de l'âge, de l'état et du sexe du locuteur. Elle peut varier de :

- 70 à 250 Hz chez l'homme ;
- 150 à 400 Hz chez la femme ;
- 200 à 600 Hz chez l'enfant [13].
- 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 $\boldsymbol{F_0}$.

1.6.2 Intensité sonore

L'intensité exprime le volume sonore d'un phonème et dans le cas d'un voisement elle représente l'amplitude des vibrations des cordes vocales, elle résulte de la pression sous glottique. Pour rendre compte de l'intensité d'un son, on utilise une unité de mesure relative, le décibel (dB).

Elle est exprimée pour un signal échantillonné s_n par :

$$\mathbf{E} = \frac{\mathbf{1}}{\mathbf{T}} \sum_{N=1}^{T} \boldsymbol{s_n}^{\mathbf{2}} \tag{1.3}$$

$$\boldsymbol{E_{dB}} = \mathbf{10} * \boldsymbol{log_{10}}(\frac{\mathbf{1}}{\mathbf{T}} \sum_{N=1}^{T} \boldsymbol{s_n}^{\mathbf{2}}) \tag{1.4}$$

1.6.3 Durée phonémique

La durée représente le temps de la prononciation d'un phonème. Elle est le paramètre acoustique le plus délicat à évaluer. La difficulté de mesure réside dans sa grande variabilité qui est due au contrôle quasi impossible du système phonatoire. Chaque phonème se caractérise par ses propres durées intrinsèques et extrinsèques.

1.6.4 Formants

Les formants sont des zones fréquentielles de forte énergie, correspondent à une résonance dans le conduit vocal de la fréquence fondamentale produite par les cordes vocales. Ces formants représentent les maxima de la courbe de réponse en fréquences du conduit vocal. Chaque son à ses formants caractéristiques (figure 1.9).

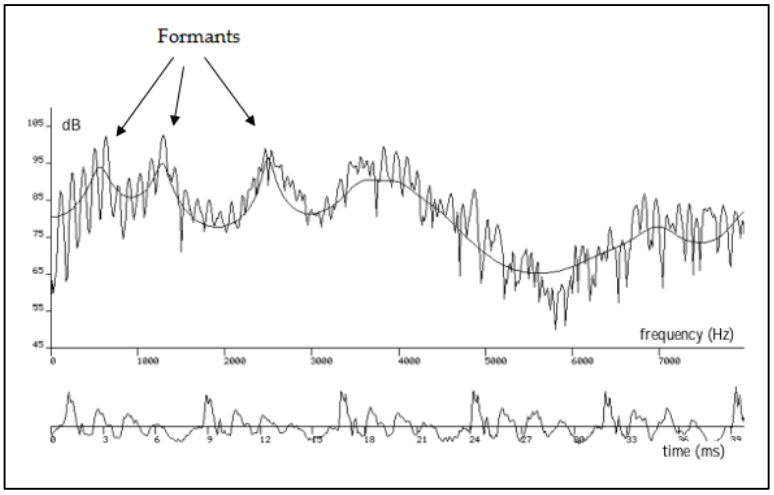


Figure 1.9 : *Représentation spectrale et temporelle d'un signal du son voisé*

1.7 COMPLEXITE DU SIGNAL DE LA PAROLE

La parole est un signal continu d'énergie finie, non stationnaire. Sa structure est complexe et variable dans le temps ; périodique ou plus exactement pseudo périodique pour les sons voisés, aléatoires pour les sons fricatifs et impulsionnels pour les sons occlusifs.

1.7.1 Continuité

Le langage oral est une suite continue de sons sans séparation entre les mots. Les silences correspondent en général à des pauses de respiration dont l'occurrence est aléatoire. Il peut très bien y avoir des intervalles de silence au milieu d'un mot et aucun intervalle entre deux mots successifs. Par conséquent, il est très difficile de déterminer le début et la fin des mots composant la phrase [9].

1.7.2 Variabilités

La parole présente une très grande variabilité qui résulte de plusieurs facteurs et ceci que ce soit pour un même ou plusieurs locuteurs. Parmi ces facteurs, les perturbations apportées par le microphone (selon le type, la distance et l'orientation) et l'environnement (bruit et réverbération). De telles variations ne donnent pas naissance à de nouveaux phonèmes, puisqu'elles ne portent aucune information sémantique. Ainsi, les phonèmes apparaissent sous une multitude de formes articulatoires, appelées allophones ou variantes [9].

1.7.3 Coarticulation

Le signal de parole est constitué d'une succession d'unités différentes. Cependant, contrairement à ce qu'on pourrait croire, ces unités ne sont pas indépendantes les unes des

autres mais s'influencent mutuellement : c'est le phénomène de coarticulation. En effet, quand on produit de la parole, on ne produit pas des segments individuels les uns après les autres : la parole n'est pas de l'épellation. Au contraire, la parole est produite par les gestes des différents articulateurs du conduit vocal (larynx, langue, lèvres, mâchoire, velum) qui se chevauchent en partie au cours du temps car ils subissent des influences diverses.

1.7.4 Redondance

Le signal de la parole est très redondant. Son traitement automatique nécessite, de réduire au maximum cette redondance afin de diminuer l'encombrement en mémoire et de limiter les durées du traitement, lequel doit se faire en temps réel. A l'inverse, le débit ne doit pas être trop faible pour conserver un bon rapport signal/bruit. En effet, Il existe une grande disproportion entre le débit du signal enregistré et la quantité utile pour une tâche de reconnaissance [9].

1.8 TRAITEMENT ATOMATIQUE DE LA PAROLE (TAP)

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage (traitement de données symboliques), cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications.

L'importance particulière du traitement de la parole dans ce cadre plus général s'explique par la position privilégiée de la parole comme vecteur d'information dans notre société humaine. L'extraordinaire singularité de cette science, qui la différencie fondamentalement des autres composantes du traitement de l'information, tient sans aucun doute au rôle fascinant que joue le cerveau humain à la fois dans la production et dans la compréhension de la parole et à l'étendue des fonctions qu'il met, inconsciemment, en œuvre pour y parvenir de façon pratiquement instantanée.

Les techniques modernes de traitement de la parole tendent cependant à produire des systèmes automatiques qui se substituent à l'une ou l'autre de ces fonctions :

1.8.1 L'analyse

Cherche à mettre en évidence les caractéristiques du signal vocal tel qu'il est produit, ou parfois tel qu'il est perçu, mais jamais tel qu'il est compris, ce rôle étant réservé aux

reconnaisseurs. Les analyseurs sont utilisés soit comme composant de base de systèmes de codage, de reconnaissance ou de synthèse, soit en tant que tels pour des applications spécialisées, comme l'aide au diagnostic médical (pour les pathologies du larynx, par analyse du signal vocal) ou l'étude des langues [8].

1.8.2 Reconnaissance Automatique de la Parole (RAP)

La RAP Sert à décoder l'information portée par le signal vocal à partir des données fournies par l'analyse. On distingue fondamentalement deux types de reconnaissance, en fonction de l'information que l'on cherche à extraire du signal vocal.

1.8.2.1 **Reconnaissance de locuteur**

L'objectif de la reconnaissance de locuteur est de reconnaître la personne qui parle. On classe également les reconnaisseurs en fonction des hypothèses simplificatrices sous lesquelles ils sont appelés à fonctionner :

- **l'identification :** le problème est de déterminer qui, parmi un nombre fini et préétabli de locuteurs, a produit le signal analysé ;
- **la vérification :** le problème est de vérifier que la voix analysée correspond bien à la personne qui est sensée la produire ;
- **dépendante de texte (avec texte dicté) :** la phrase à prononcer pour être reconnue est fixée dès la conception du système ;
- **indépendante du texte :** la phrase à prononcer pour être reconnue est fixée lors du test, donc la reconnaissance doit être assurée pour n'importe quelle phrase.

1.8.2.2 **Reconnaissance de la parole**

L'objectif de la reconnaissance de la parole est de reconnaître ce qui est dit, on classe également les reconnaisseurs de parole comme suit :

- **monolocuteur :** capable de reconnaître que la parole prononcée par la voix d'une seule personne ;
- **multilocuteur :** capable de reconnaître la parole prononcée par la voix d'un nombre fini de personnes ;
- **indépendante de locuteur :** capable de reconnaître la parole de n'importe qui ;
- **reconnaisseur de mots isolés :** le locuteur sépare chaque mot par un silence ;

- **reconnaisseur de mots connectés :** le locuteur prononce de façon continue une suite de mots prédéfinis ;
- **reconnaisseur de parole continue :** le locuteur prononce n'importe quelle suite de mots de façon continue.

1.8.3 Synthèse de la parole

La synthèse est la production de la parole artificielle. On distingue fondamentalement deux types de synthétiseurs à partir d'une représentation :

- **numérique :** inverses des analyseurs, dont la mission est de produire de la parole à partir des caractéristiques numériques d'un signal vocal telles que celles obtenues par analyse ;
- **symbolique :** inverse des reconnaisseurs de parole et capables en principe de prononcer n'importe quelle phrase sans qu'il soit nécessaire de la faire prononcer par un locuteur humain au préalable.

Dans cette seconde catégorie, on classe également les synthétiseurs en fonction de leur mode opératoire, synthétiseurs à partir :

- **du texte,** reçoivent en entrée un texte orthographique et doivent en donner lecture ;
- **de concepts**, appelés à être insérés dans des systèmes de Dialogue Homme-Machine, reçoivent le texte à prononcer et sa structure linguistique, telle que celle produite par le système de dialogue.

1.8.4 Codage

Le codage permet la transmission ou le stockage de parole avec un débit réduit, ce qui passe tout naturellement par une prise en compte judicieuse des propriétés de production et de perception de la parole.

1.9 CONCLUSION

Dans ce chapitre, nous avons présenté brièvement le phénomène de la parole « physiologique, phonétique et acoustique ». Nous avons pu ainsi voir qu'il s'agit d'un phénomène complexe qui repose sur de nombreux mécanismes physiologiques et cognitifs. En présentant le modèle source/filtre, nous avons pu introduire la description de signal de parole. Des généralités phonétiques, certaines propriétés spécifiques et une description simplifiée des différents sons de l'AS ont été présentées.

2.1 INTRODUCTION

Nous avons vue dans le chapitre précédent que La parole est un signal réel, continu, d'énergie finie. Sa structure est complexe et variable dans le temps, périodique pour les sons voisés, aléatoire pour les sons fricatifs, impulsionnel dans les phases explosives des sons occlusifs.

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

2.2 ANALYSE DU SIGNAL VOCAL

Le traitement du signal vocal a pour but de fournir une représentation moins redondante de la parole que celle obtenue par codage de l'onde temporelle tout en permettant une extraction précise des paramètres pertinents du signal de parole tels que :

- Pour la source :
    - période du fondamental
    - amplitude Ao.
- Pour le conduit vocal :
    - période des formants Ti =1/Fi , i=1,2,… (Fi : fréquences des formants)
    - amplitude Ai
    - bandes passantes Bi.

L'extrême variabilité du signal vocal est due à la :

- Complexité du couplage (source/conduit) ;
- Grande dynamique et variété des voix ;
- Variation rapide de la parole.

Cette variabilité est liée directement au locuteur :

- à son âge, son sexe et à son accent géographique
- et son état physique (fatigue, maladie) et émotionnel (content, triste, nerveux)

Toutes ces propriétés complexes du signal vocal rendent ce dernier difficile à traiter d'où la multiplicité des méthodes de traitement. Ces méthodes de traitements sont très nombreuses, nous n'en citeront que celles utilisées dans cette étude.

2.3 TECHNIQUES D'ANALYSE DU SIGNAL VOCAL

On classe habituellement, les différentes méthodes de traitement du signal en trois catégories :

- Les transformées usuelles : transformée discrète de Fourier et transformée en Z.
- Les méthodes fondées sur la déconvolution ‘‘source/ conduit’’ cepstre et codage prédictif linéaire (LPC) qui s'appuient sur un modèle même simplifié de production de la parole.
- Les méthodes basées sur un modèle de perception (filtre).

2.3.1 Analyse fréquentielle

La parole est constituée de plusieurs éléments appelés phonèmes, dépendants les uns des autres. On peut caractériser ces phonèmes ainsi que leurs liens grâce à leurs aspects fréquentiels : présence ou non de fondamental laryngé, formants, transitions phonétiques ainsi que bruits d'explosion et de friction.

2.3.1.1 Fréquence fondamentale ou pitch (F0)

La fréquence fondamentale constitue une caractéristique très importante de nombreux signaux environnementaux comme les sons de la parole. Elle correspond à la fréquence de vibration des cordes vocales lors de la production des voyelles ou des consonnes voisées. Elle génère des variations prosodiques, c'est à dire de mélodie et d'intonation, qui contribue à l'identification du sexe, de l'âge et de l'identité du locuteur, ainsi qu'à la signification du message prononcé. Le fondamental se trouve dans un registre grave, et différent selon la voix.

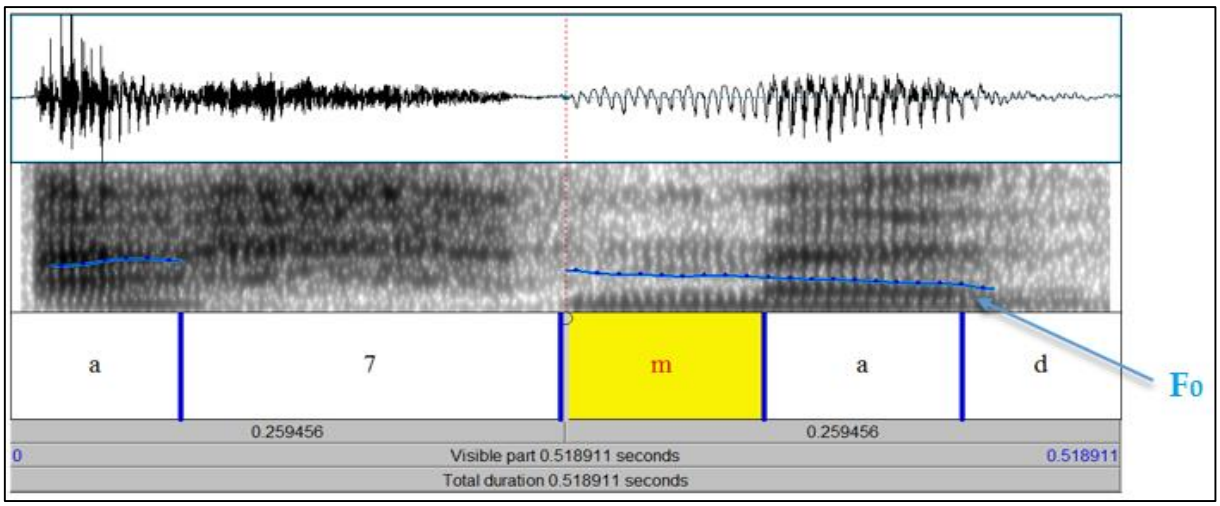


Figure 2.1 : *Visualisation de Pitch F_0*

2.3.1.2 Les formants

Les formants sont des zones fréquentielles dont l'intensité est renforcée.

Chaque voyelle est reconnaissable par l'amplification d'harmoniques déterminés du son laryngé, appelés formants. La composition formantique de chaque voyelle est indépendante de la hauteur de son fondamental. Ainsi, que l'on soit un homme, une femme ou un enfant, on prononce les mêmes voyelles.

Les formants sont caractérisés par leurs fréquences de résonance et une bande passante.

Les caractères de chaque formant sont :

• **Le 1er formant (F1) :** La zone formantique de F1 est située entre 250 et 750Hz. Le premier formant F1 correspond à l'aperture de la voyelle (ouverture de la mandibule).

• **Le 2ème formant (F2) :** La zone formantique de F2 est située entre 750 et 2500Hz. C'est surtout ce deuxième formant qui est nécessaire pour l'intelligibilité du langage, et en particulier dans la zone située autour de 2KHz. Il exprime la position plus ou moins avancée de la langue

• **Le 3ème formant (F3) :** Le troisième formant est beaucoup moins caractéristique de la voyelle que le premier et le deuxième, car sa hauteur fréquentielle varie peu pour la majorité des voyelles. Il est aussi à noter que le troisième formant donne de l'information sur l'arrondissement des lèvres.

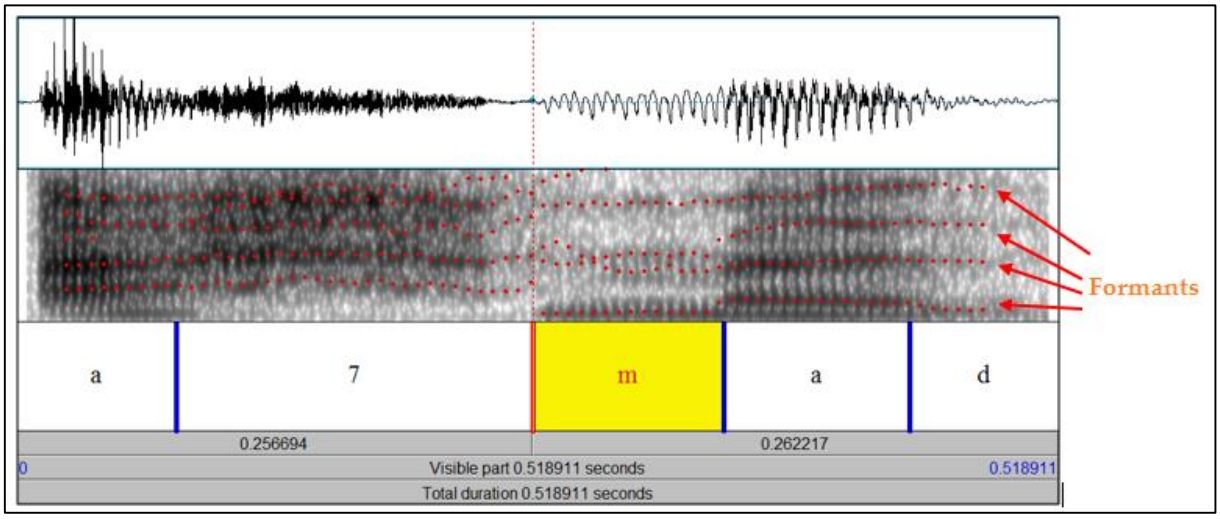


Figure 2.2 : *formants de mots [ahmad]*

Les valeurs des premiers et deuxièmes formants permettraient aux auditeurs d'identifier les voyelles orales. Leurs valeurs respectives rendent compte des propriétés du résonateur buccal et du résonateur pharyngal. Ce sont les formants les plus graves et il arrive que le premier

formant se confonde avec le fondamental, particulièrement lorsqu'il s'agit de voix de femmes ou d'enfants dont la fréquence naturelle de la voix est plus élevée.

2.3.2 Analyse spectrale :

2.3.2.1 Spectrogramme :

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un spectrogramme (figure 2.3). L'amplitude du spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux dimensions temps-fréquence. Ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants.

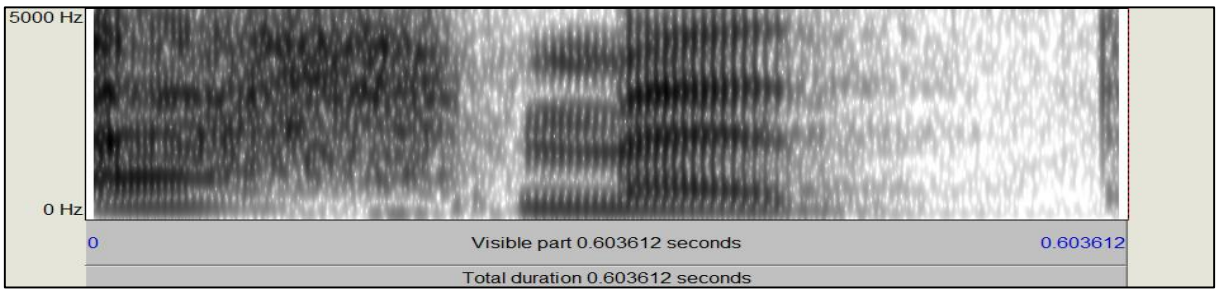


Figure 2.3 : *Spectrogramme d'un mot de 5 phonèmes [a7mad]*

2.3.2.2 Spectre obtenu par FFT

Tout son est la superposition de plusieurs ondes sinusoïdales. Grâce à la FFT, on peut isoler les différentes fréquences qui le composent. On obtient ainsi une répartition spectrale du signal (figure 2.3). Les valeurs des formants sont calculées automatiquement dans le signal de parole au moyen d'un lissage spectral.

La transformée de Fourier à court terme est obtenue en extrayant de l'audiogramme une 30aine de ms de signal vocal, en pondérant ces échantillons par une fenêtre de pondération (souvent une fenêtre de Hamming) et en effectuant un transformée de Fourier sur ces échantillons.

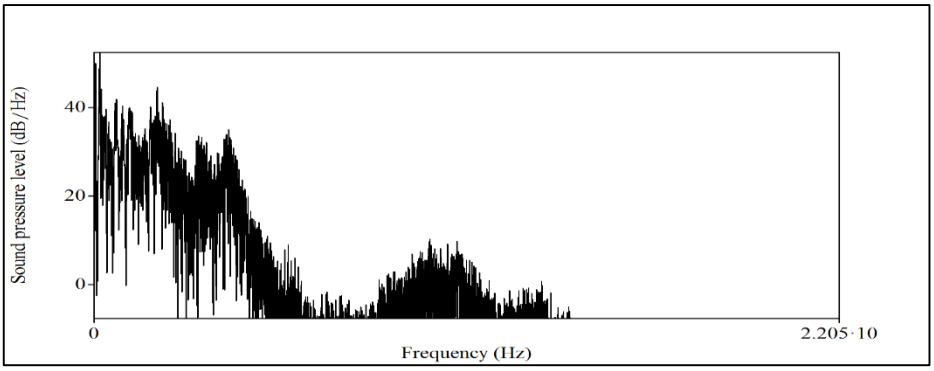


Figure 2.3 : *spectre obtenue par FFT*

2.3.3 Prétraitement du signal vocal :

Le Traitement du signal vocal numérique passe par deux principales étapes : le prétraitement et l'analyse, ces étapes permettent d'extraire les caractéristiques du signal, cela peut aider pour augmenter les performances du signal et le rendre facile à traiter.

Un échantillonnage et une préaccentuation seront appliqués sur le signal vocal. Pour les techniques de reconnaissance, d'analyse ou de synthèse de la parole, la fréquence d'échantillonnageF_s. D'après le théorème de Shannon, la perte d'information entre le signal continu et le signal discret correspondant est quasiment nulle si et seulement si la fréquence d'échantillonnage :

$$F_s \geq 2F_{max} \qquad (2.1)$$

F_{max} est la fréquence maximale du spectre du signal. F_s Peut varier de 08 jusqu'à 16 kHz. Le filtre de préaccentuation de transmittance $H(z)$ est :

$$H(z) = 1 - a.z^{-1} \qquad \text{Avec : a=0.95} \qquad (2.2)$$

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 $W(n)$ [15] (figure 2.2).

$$W(n) = \begin{cases} 0.45 + 0.46.\,\mathrm{Cos}\,(\mathrm{n}/(\mathrm{n}-1)) & \mathrm{n} \in [0, \ldots, \mathrm{n}-1] \\ 0 & \text{ailleurs} \end{cases} \quad (2.3)$$

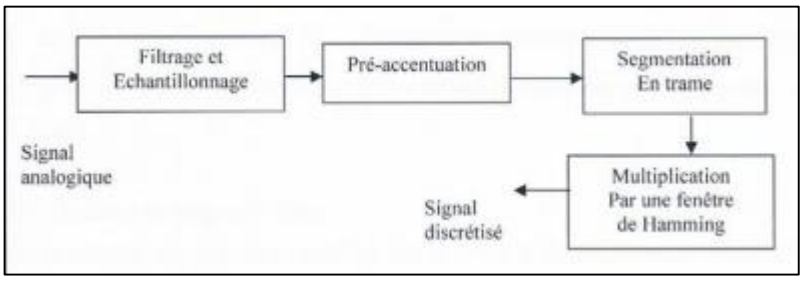


Figure 2.2 : *Prétraitement du signal vocal* [15]

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 **T**aux de **P**assage par **Z**éro (**TPZ**) et la fonction d'autocorrélation font partie des méthodes temporelles.

2.3.4 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 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 **T**ransformée de **F**ourier (**TF**) et n'a d'intérêt que si elle s'applique à une période du signal vocal, donc sur une période assez courte. Actuellement, les spectres sont obtenus numériquement par la **T**ransformée de **F**ourier **D**iscrète (**TFD**), en particulier grâce à l'algorithme de la **T**ransformée de **F**ourier **R**apide (**TFR**) ou **F**ast **F**ourier **T**ransform (**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 (figure 2.3).

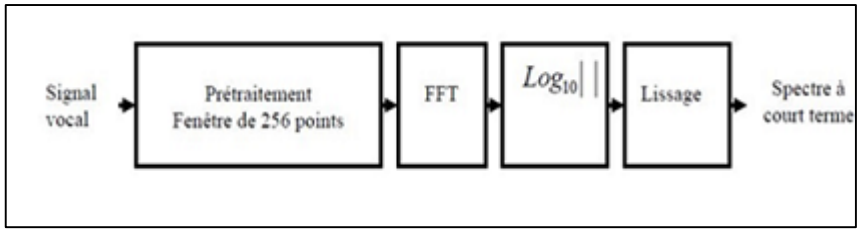


Figure 2.3 : *Analyse numérique du signal parole par FFT*

2.3.5 Méthodes paramétriques

Les méthodes paramétriques appelées aussi méthodes d'identification sont fondées sur une connaissance des mécanismes de production de la parole. Les plus utilisées sont celles basées sur l'analyse prédictive linéaire et l'analyse cepstrale. Hypothèse de base est que le conduit 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.

2.3.5.1 Codage Prédictif Linéaire

Linear **P**redictive **C**oding (**LPC**) Cette méthode se fonde sur les connaissances de la production de la parole et suppose que le modèle de production de la parole est linéaire selon le schéma (figure 2.4).

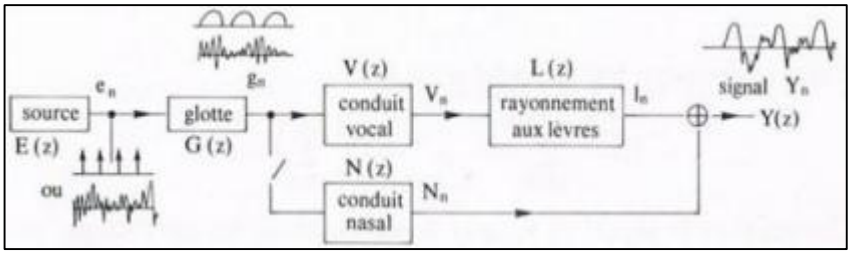


Figure 2.4 : *Modèle général de production de la parole* [7]

Globalement, ce modèle peut se décomposer en deux parties : la source active, le conduit passif de manière plus détaillée, il peut se décrire de la maniéré suivante : l'onde est modélisée comme la sortie d'un filtre passe bas à deux pôles de fréquence de coupure d'environ 100 Hz (glotte), l'entrée en de ce filtre est un train d'impulsions de période T0 pour les sons voisés ou un bruit blanc pour les sons non voisés (source). Le modèle du conduit vocal est un filtre tout pôle (**AR** : **A**uto - **R**égressif) d'ordre 2M décomposable en une cascade de résonateurs à 2 pôles en série (tuyaux résonants). Le modèle du conduit nasal est un filtre pôle zéro **ARMA** (**A**uto **R**égressif à **M**oyenne **A**justée) et le rayonnement aux lèvres peut se modéliser par un filtre tout zéro (**MA** : **M**oyenne **A**justée). L'ensemble des conduits se comporte donc comme un système linéaire ARMA [7].

Modèle glottale :

$$G(z) = \frac{1}{(1- e^{-2\pi f_g t}\, Z^{-1})^2} \qquad \text{Avec } f_g = 100 \text{ Hz} \qquad (2.4)$$

Modèle du conduit vocal :

$$V(z) = \prod_{i=1}^{M} \left(\frac{1}{1- 2e^{-2\pi B_i T}\cos(2\pi F_i T) Z^{-1} + e^{-4\pi B_i T}\, Z^{-2}}\right) \qquad (2.5)$$

F_i : Fréquence du formant n° i, Bi sa bande passante

Modèle du conduit nasal :

$$N(z) = \frac{1- 2e^{-2\pi \acute{B}_N T}\cos(2\pi \acute{F}_N T) Z^{-1} + e^{-4\pi B_n T}\, Z^{-2}}{1- 2e^{-2\pi B_N T}\cos(2\pi F_N T) Z^{-1} + e^{-4\pi B_N T}\, Z^{-2}} \qquad (2.6)$$

Avec F_n et $\acute{F_n}$ formant nasal ou anti formant nasal et respectivement, $B_n$et $\acute{B_n}$ leurs bandes passantes.

Si l'on suppose qu'une partie α du signal $g_n$est dérivée vers le conduit nasal le modèle du conduit peut se mettre sous la forme :

$$H(z) = G(z).\,[(1 - \alpha)V(z)L(z) + \alpha N(z)\,] \text{ Avec } 0 \leq \alpha \leq 1 \qquad (2.7)$$

Pour un son nasal $\alpha \cong 1$; pour un son non nasal $\alpha = 0$.

$H(z)$ Est en tout généralité un modèle ARMA d'ordre p

$$H(z) = \frac{B(z)}{A(z)} \qquad (2.8)$$

Dans le domaine temporel on aura :

$$y_n + \sum_{i=1}^{P} a_i y_{n-p} = e_n + \sum_{i=1}^{p} b_i e_{n-i} \quad (2.9)$$

Caractériser le signal $y_n$revient donc a estimé les coefficients $\{a_i;b_i\}$. Pour une source connue e_n (séquence d'impulsions ou bruit blanc). Souvent pour simplifier la résolution de ce problème, on suppose que b_i= 0, i ≥ 1 ce qui rend le modèle AR [].

2.3.5.2 Analyse cepstrale

Le défaut majeur des méthodes d'analyse, comme la FFT, pour le calcul du spectre réside dans l'intermodulation source/conduit vocal qui rend difficile la mesure du fondamental F_0 et des formants.

Le lissage cepstral est une méthode qui vise à séparer la contribution du conduit vocal de l'excitation glottique. Cette séparation est réalisée par un homomorphisme qui transforme la convolution des signaux dans le domaine temporel en une addition dans le domaine cepstral. En outre, cette méthode permet de fournir un vecteur spectral des MFCC pour des fins de la RAP et de lisser le spectre de parole pour trouver les formants.

Pour cela, nous faisons hypothèse que le signal vocal y_n est produit par le signal excitateur u_n traversant un système linéaire de réponse impulsionnelle b_n

Le but du cepstre est de séparer ces deux contributions par dé convolution. Il est fait hypothèse qu'un signal excitateur est soit une séquence d'impulsions (périodiques, de périodeT_0, pour les sons voisés), soit un bruit blanc pour les sons non voisés, conformément au modèle de production de la parole. Une transformation en Z permet de transformer la convolution en produit.

$$Y(z) = B(z).U(z) \quad (2.9)$$

Le logarithme du module uniquement (car nous ne intéressons pas à l'information de phase) transforme le produit en somme. Nous obtenons alors :

$$\log|Y(z)| = \log|B(z)| + \log|U(z)| \quad (2.10)$$

Par transformation inverse, nous obtenons le cepstre. Dans la pratique, la transformation en Z est remplacée par une TFR. L'expression du cepstre est donc :

$$C(n) = FT^{-1}\{log(FT\{y(n)\})\} \quad (2.11)$$

Le cepstre qui ne fait appel à aucune information a priori sur le signal acoustique, est basé sur une connaissance du mécanisme de production de la parole. L'espace de représentation du

cepstre ou espace quéferentiel est homogène par rapport au temps. Les premiers coefficients cepstraux contiennent l'information relative au conduit vocal. Cette contribution devient négligeable à partir d'un échantillon n_0 qui correspond à la fréquence fondamentale F_0 . Les pics périodiques visibles au-delà de n_0, reflètent les impulsions de la source.

Le spectre du cepstre pour les indices inférieurs à n_0 permet d'obtenir un spectre lissé, en éliminant les lobes secondaires dû à la contribution de la source. Ces deux contributions peuvent être séparées par une simple fenêtre temporelle notée F (liftrage) telle que le filtre adouci ou le filtre rectangulaire.

La présence d'un pic important dans le cepstre renseigne d'une part sur le caractère voisé ou non du son et d'autre part constitue une bonne indication sur la fréquence fondamentale. L'enveloppe spectrale du conduit vocal (structure formantique) est obtenue par une transformation supplémentaire (figure 2.5).

Le spectre lissé débarrassé théoriquement de la contribution de la source ne contient que des informations sur le conduit vocal et en particulier sur ses extrema (Formants) [9].

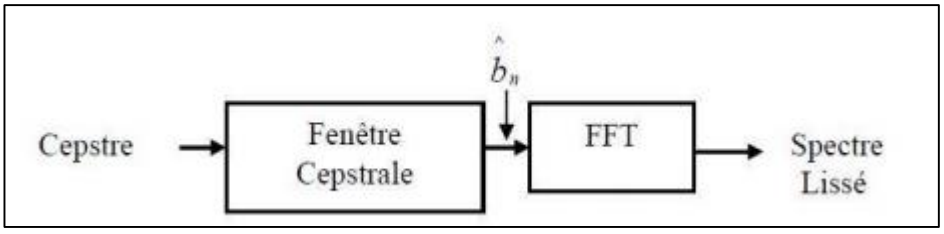


Figure 2.5 : *Obtention de la structure formantique à partir du cepstre* [9]

2.4 CONCLUSION

Dans ce chapitre nous avons introduit l'analyse du de la parole et ces paramètres pertinents, ainsi nous avons cité les principales méthodes et techniques de l'analyse de ce signal, ce qui nous aide pour introduire le chapitre suivant, qui est le sujet de notre étude : Apprentissage des Sons Spécifiques de l'Arabe Standard.

3.1 INTRODUCTION

Le but de notre travail est d'étudier les sons spécifiques de la langue AS, en vue d'élaborer un ASSAS (**A**nalyseur des **S**ons **S**pécifiques de l'**A**rabe **S**tandard), pour les apprenants étrangers.

Dans ce chapitre, nous allons présenter le système de l'«ASSAS», avec une brève explication du processus de construction de notre BD, à l'aide de l'outil d'analyse Praat. Ainsi nous allons expliquer le fonctionnement de l'interface graphique «ASSAS» qui est faite sous le logiciel MATLAB.

3.2 PRESENTATION D'« ASSAS » POUR LES APPRENANTS ETRANGERS

Avec le développement très important des moyens de calcul et de stockage, les interfaces Homme Machine sont devenues de plus en plus proches de l'interaction humaine naturelle. L' « **ASSAS** » est un outil d'aide en apprentissage des sons spécifiques de l'AS pour les apprenants étrangers.

Ce système est basé sur la construction d'une base de données contient les six sons spécifiques, prenant compte les trois positions possibles dans le mot et l'analyse de chaque phonème. Ce système joue le rôle d'un afficheur de différentes caractéristiques de son choisi, via une base de données préconçue.

3.3 METHODOLOGIE DU TRAVAIL

La méthodologie suivie dans ce travail est résumée dans une succession d'étapes, la figure 3.1 montre la démarche adoptée.

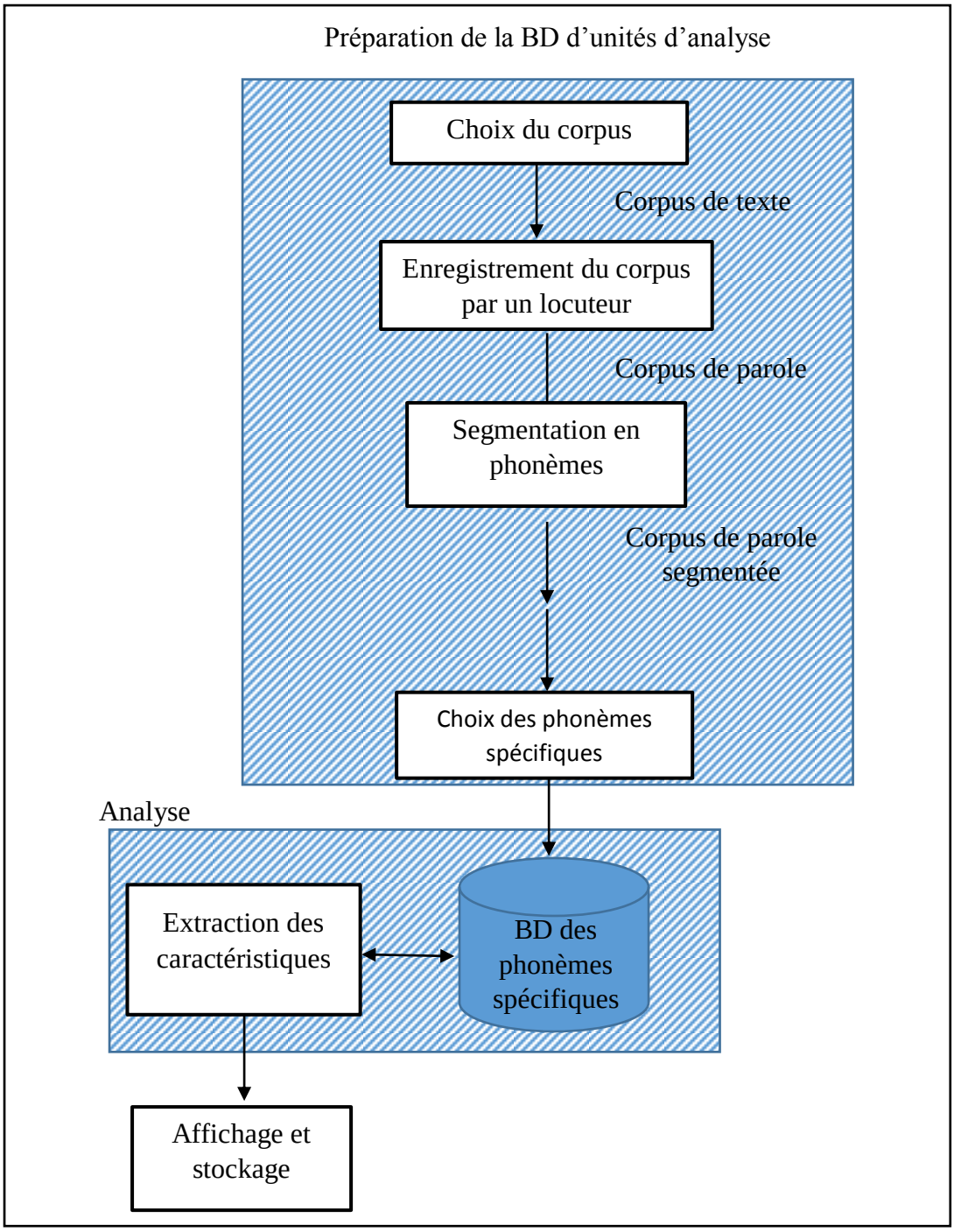


Figure 3.1 : *Schéma méthodologique du Système de l'Analyseur des Sons Spécifiques de l'Arabe Standard*

3.4 CONSTRUCTION DE LA BASE DE DONNEES

3.4.1 Choix du corpus

Le choix du corpus est un élément clé pour la qualité d'un système d'analyse. Définir le corpus revient à déterminer l'ensemble des unités à enregistrer de façon à obtenir un certain espace acoustico-prosodique meilleur.

En ce qui concerne notre travail nous avons pris un corpus constitué de mots porteurs contenant des phonèmes spécifiques à la langue Arabe pris dans différentes positions (Initiale, Initiale et Finale). Ce corpus se nomme « ASSAS ». Ce corpus est continu et composé de 18 mots (Tableau 3.1).

Tableau 3.1 : *Corpus « ASSAS »*

Mots en Arabe	Mots en Arabe
ألف	ظل
وائل	مظلة
دفء	حافظ
حساب	ضعف
أحمد	أضعاف
صالح	ماض
علم	قسم
معمل	أقلام
مرجع	صادق

3.4.2 Acquisition des données

L'acquisition des données consiste à enregistrer les phrases du corpus choisis, en utilisant un outil spécifique au traitement du signal vocal (Praat). Les enregistrements ont été effectués par une seule locutrice. La fréquence d'échantillonnage choisie est de 44100 Hz, les échantillons ont été codés sur 32 bits par échantillon.la figure 3.2 représente le audiogramme de corpus enregistré.

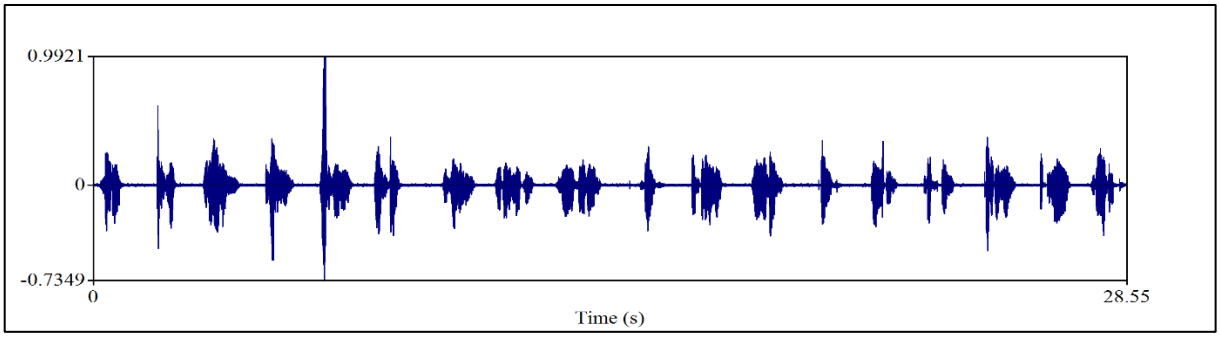


***Figure 3.2** : représentation temporelle du corpus « ASSAS »*

La figure 3.3 montre les principales informations acoustiques sur notre corpus avant segmentation.

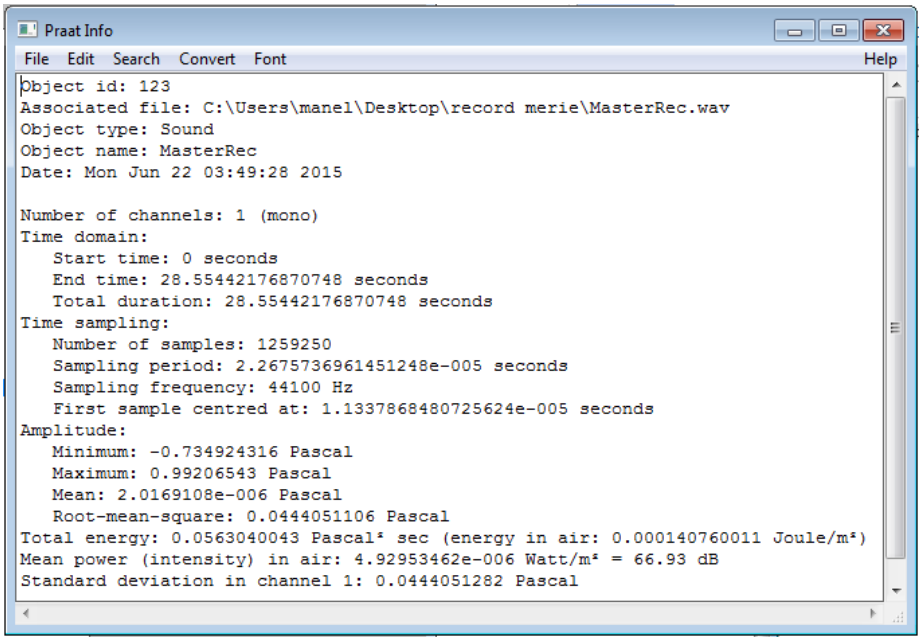


***Figure 3.3 :** informations acoustiques sur « ASSAS »*

- durée totale : 28,55 s
- une amplitude : minimum : -0,73 pascal
  maximum : 0,99 pascal
  moyenne : $2{,}01.10^{-6}$ pascal
- une énergie totale de 14,07 10^{-5} pascal².sec ;
- puissance moyenne (intensité) en air : 4.92. 10^{-5} watt/m² = 66.93 db ;

3.4.3 Segmentation

La segmentation consiste à extraire les mots puis les phonèmes du corpus enregistré. Cette segmentation a été effectuée manuellement à l'aide de l'outil **Praat.** La visualisation

d'audiogramme et de spectrogramme à la fois nous aide pour bien ajuster les limites de segmentation de chaque phonème (figure 3.4).

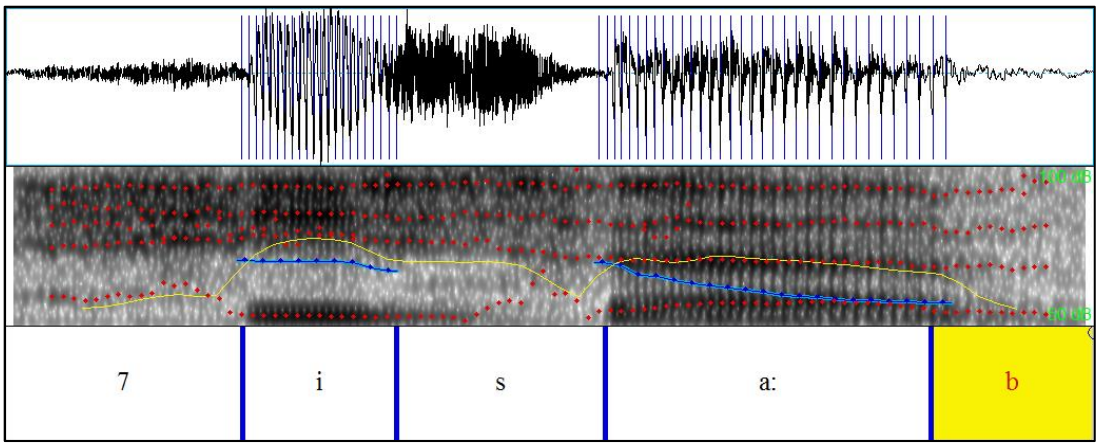


Figure 3.4 : *Segmentation en phonèmes d'un mot de corpus « ASSAS »*

La figure 3.5 montre l'audiogramme de phonème [dˤ] en contexte initial.

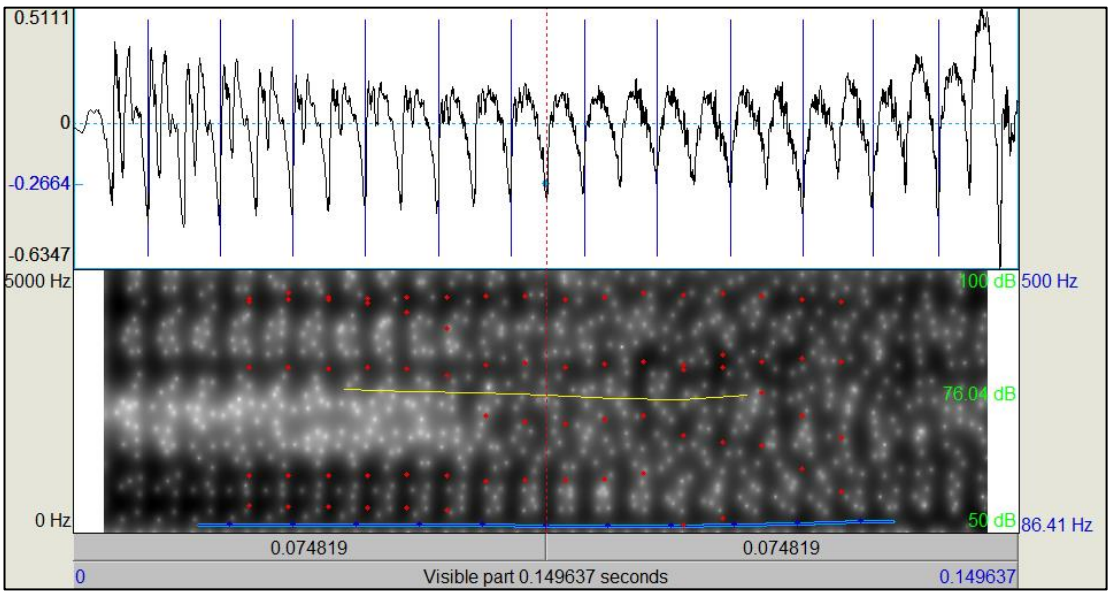


Figure3.5 : *Audiogramme de phonème [dˤ]*

3.4.4 Annotation

L'annotation est de faire décrire chaque phonème segmenté par une note descriptive (symbole, chaine de caractères, etc.). Nous avons choisis la plus simple note qui peut décrire chaque phonème, pour faciliter la manipulation des segments lors de l'analyse et l'affichage (tableau 3.2).

Tableau 3.2 : *Annotation des segments de corpus « ASSAS »*

Phonème	Annotation	Phonème	Annotation
أ	AI	ضـ	8I
ـئـ	AM	ـضـ	8M
ء	AF	ـض	8F
عـ	3I	قـ	9I
ـعـ	3M	ـقـ	9M
ع	3F	ق	9F
ظـ	6I	حـ	7I
ـظـ	6M	ـحـ	7M
ظ	6F	ح	7F

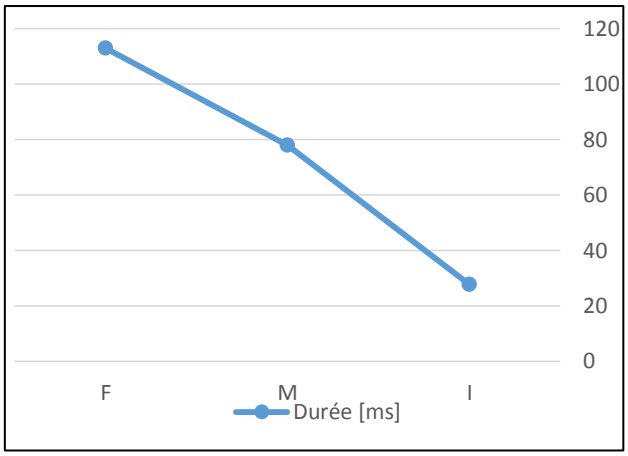
3.5 ANALYSE DES PHONEMES

Consiste à extraire la durée, l'énergie et l'intensité de phonème en différentes position (Tableau 3.3).

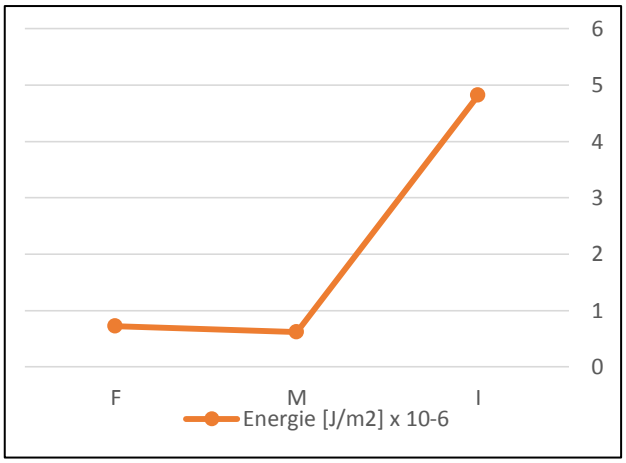
Tableau 3.3 *: Analyse générale des phonèmes étudiés*

Son en API	Contextes	Durée [ms]	Energie [J/m^2] x 10^{-6}	Intensité [dB]
[ʔ]	I	27,55	4,82	82,43
	M	77,86	0,62	69.06
	F	112,72	0,72	58.09
[q]	I	22,47	1,01	76.52
	M	93,85	2,99	75.04
	F	191,13	0,10	57.2
[dˤ]	I	128,86	20,83	82.09
	M	190,36	25,75	81.31
	F	137,93	0,15	60.37
[ðˤ]	I	149,63	9,12	77.85
	M	54,39	15,84	84.64
	F	208,77	2,05	69.93
[ʕ]	I	102,67	16,94	82.17
	M	110,56	8,31	78.76

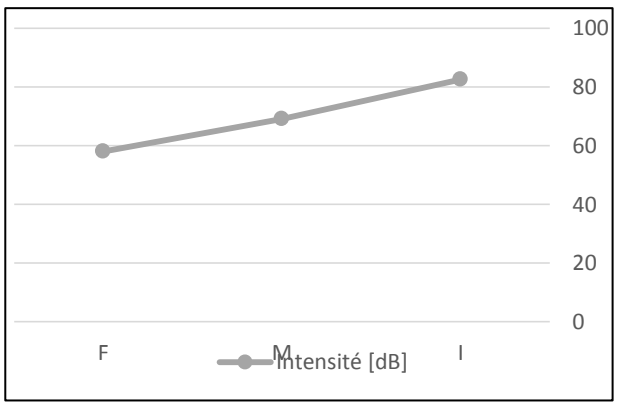
	F	98,52	1,67	72.31
[ɦ]	I	91,15	1,34	71.7
	M	126,59	3,13	73.93
	F	221,24	1,34	67.83



a)



b)



c)

Figure 3.5 : *variation de*

a) la durée,
b) l'énergie ;
c) l'intensité
en fonction de la position contextuelle de [?]

Interprétation :

Remarquons que la durée de phonème varie en fonction de sa position (durée de [?] Initiale > durée de [?] Médiane), ainsi que l'énergie et l'intensité moyenne, cela peut donner une information sur le sens et la structure syllabique du mot.

3.6 ARCHITECTURE DU PROGRAMME « ASSAS »

Avant la conception matérielle d'un outil technique, une simulation de son système doit être élaborée, pour assurer le bon fonctionnement et améliorer les performances.

La simulation de notre « *ASSAS* » est faite sous le logiciel MATLAB. Les étapes suivant décrivent l'utilisation de notre système « *ASSAS* » (Figure 3.6, Figure 3.7, Figure 3.8) :

3.6.1 Le choix du phonème à étudier

Après le lancement du programme la fenêtre ci-après (Figure 3.6) sorte et nous devrons choisir l'un des lettres présentés pour commencer l'analyse du phonème sélectionné.

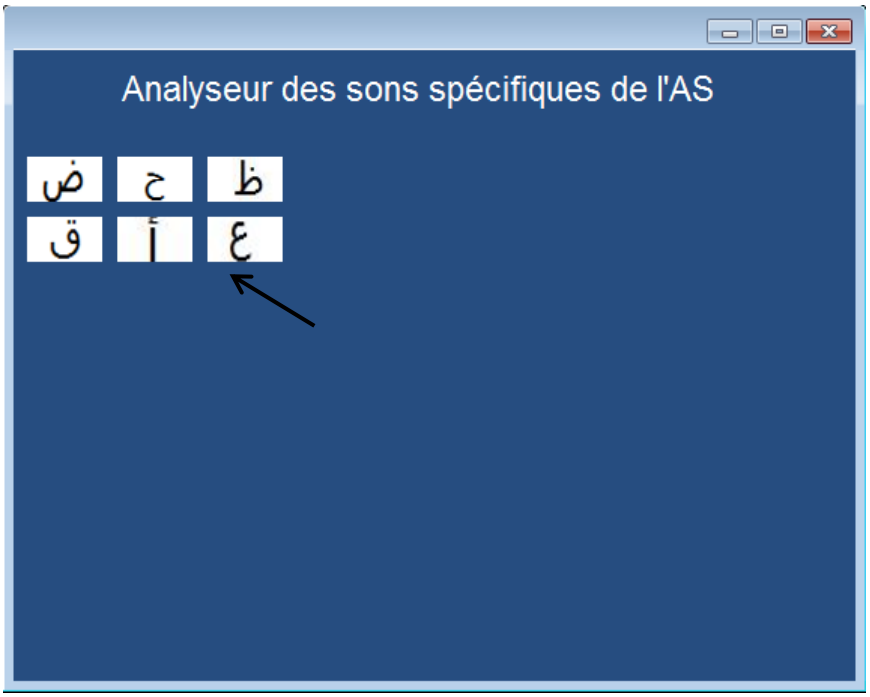


Figure 3.6 *Le choix de la lettre « ع »*

3.6.2 Le choix du contexte du phonème

Après le choix du phonème que nous voulons analyser, une nouvelle fenêtre apparait (Figure 3.7) avec trois exemples de contextes différents (position initiale, position médiane, position finale) pour chaque phonème. Dans cet état nous devons choisir le contexte du phonème que nous voulons analyser.

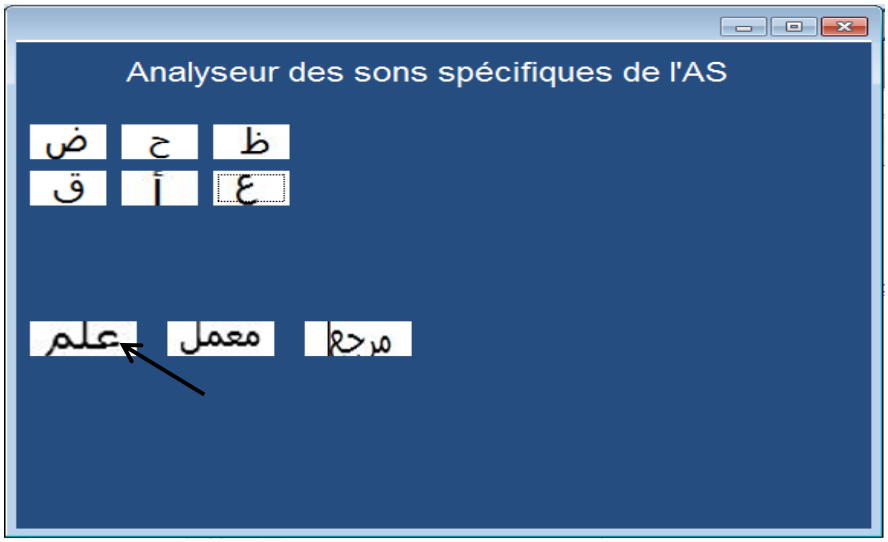


Figure 3.7 *Le choix du mot «علم» (contexte initial du phonème [ʕ])*

3.6.3 L'analyse du phonème

Après le choix du contexte nous entendons une prononciation du mot choisi et une autre fenêtre est apparait (Figure 3.8) qui contient les caractéristiques du contexte de phonème analysé (duré, intensité, audiogramme, spectrogramme).

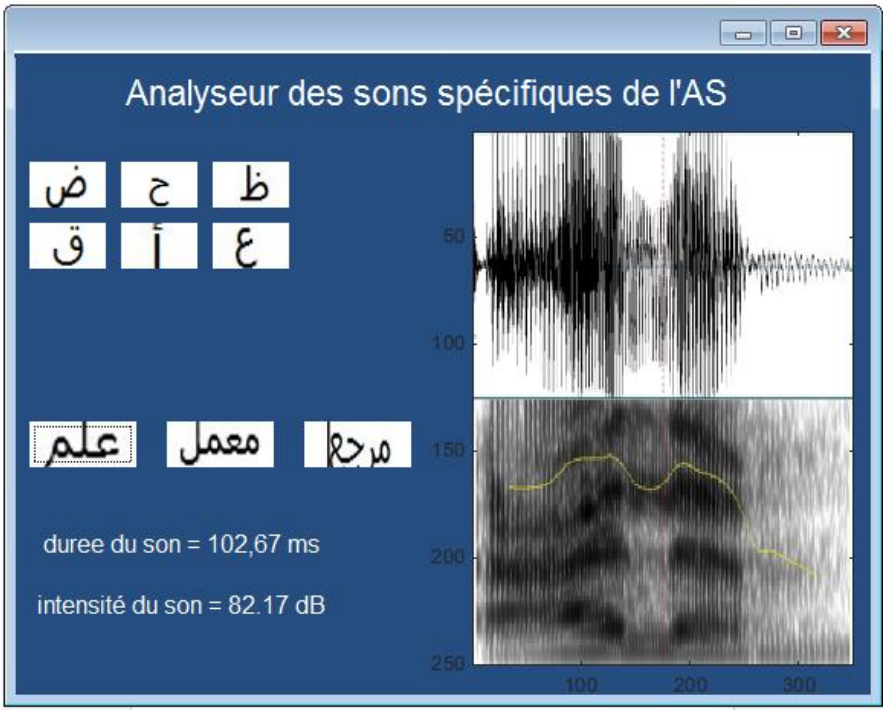


Figure 3.8 *Analyse du phonème [ʕ] en contexte initial*

3.7 CONCLUSION

Dans ce chapitre nous avons expliqué le fonctionnement du système de simulation « ASSAS » qui est fait sous le logiciel MATLAB .Et étudier les différentes caractéristiques des sons spécifiques de l'AS.

Conclusions Générales et perspectives

L'objectif de notre travail était la réalisation d'un outil d'aide en apprentissage des sons spécifiques, destiné aux apprenants étrangers. Ce travail est muni d'un programme de simulation d'une application ASSAS. Nous avons tout d'abord effectué des études générales sur les sons spécifiques de l'AS, pour cela nous avons choisi un corpus contient tous les sons spécifiques et tient en compte ses différentes positions contextuelles (Initiale, Médiane, Finale), puis nous avons fait un enregistrement, qui a été fait par une locutrice arabophone, une segmentation et annotation du corpus « ASSAS », en vue d'élaborer notre base de données. L'étude des sons est passée par plusieurs étapes d'analyse acoustique et de visualisations, qui nous ont permis une extraction des paramètres pertinents et acoustiques.

Comme perspective à ce travail, il sera très intéressant de faire une étude évaluative pour améliorer les performances de l'Analyseur, afin d'obtenir cette amélioration, nous proposons :

- un élargissement du vocabulaire du corpus « ASSAS » : plusieurs segments pour chaque position contextuelle ;
- l'amélioration de l'algorithme du programme ;
- l'amélioration de la qualité de la voix ;
- l'utilisation des caractéristiques calculées dans la reconnaissance de la parole prononcée par le locuteur étranger, et l'utilisation de la synthèse pour corriger les fautes de prononciation ;
- l'implémentation et la conception matérielle de l'ASSAS.

Références bibliographiques

[01] F. SAUSSURE, Cours de Linguistique Générale, 1975.

[02] J. LE GRAND. Parcours Traitement Automatique du Langage Naturel, Université Stendhal, Grenoble /France, 2012.

[03] http://www.claudegaberiel.be/Cineacoustique

[04] O. GODIN, Chapitre 5-Analyse de la parole IMN317, Université de Sherbrooke, Canada, Novembre 2011.

[05] S. LE MAGEUR, Thèse de doctorat : Evaluation expérimentale d'un système statistique de synthèse de la parole, HTS, pour la langue française, Université de Rennes,France Juillet 2013.

[06] CALLIOPE, La parole et son traitement automatique, Collection Techniques et Scientifiques des Télécommunications. Préface de G. FANT, CNET/ENST, Ed. Masson, 1989.

[07] T. DUTOIT, Introduction au Traitement Automatique de la Parole, Faculté Polytechnique de Mons,France,2000.

[08] M. AISSIOU, Application des Algorithmes Génétiques au Décodage Acoustico-Phonétique de la parole en Arabe Standard, Thèse de Doctorat, ENP, Alger/Algérie, 2008.

[09] M. MEDJBER, Amélioration du standard G.729 -8Kb/s par la méthode de modification de l'échelle temporelle (WSOLA), Mémoire de magister, Ecole Nationale Polytechnique, Alger/Algérie, 2007.

[10] C. D'ALESSANDRO, G. RICHARD, Synthèse de la parole à partir du texte, Technique de l'ingénieur, Institut Mines-Télécom, France, 2013.

[11] A. RAMSAY, I. ALSHURHAN, H. AHMED, Generation of a phonetic transcription for modern standard Arabic, Université de Manchester, United Kingdomb, Université de Qatar, Qatar, Février 2013.

[12] S. OUAMOUR, Indexation automatique des documents audio en vue d'une classification par locuteurs Application à l'archivage des émissions TV et Radio, thèse de doctorat, Ecole Nationale Polytechnique, Alger/Algérie, 2009.

[13] S. BALOUL, Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé, Thèse de doctorat, Université du Maine, France, mai 2003.

[14] http://www.praat.org/