N° d'ordre : 13/2001-M/EL

REPUBLIQUE ALGERIENNE DEMOCARTIQUE ET POPULAIRE

MENISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENNE

جامعة هواري بومدين
للعلوم والتكنولوجيا
U S T H B

FCULTE DE GENIE ELECTRIQUE

Laboratoire communications parlées

Thèse

présenté pour l'obtention du diplôme de magister en électronique

Thème

ETUDE DES DUREES PHONEMIQUES DES SONS SPECIFIQUES A L'ARABE STANDARD

Présenté par : M^ed Lassaâd BENZAOUI

Soutenu le : 08/12/2001

Devant le jury :

M. A. DJERADI	MC à l'USTHB (Alger)	Président
M^lle M.GUERTI	MC à l'ENP (Alger)	Directeur de thèse
M. A. HOUACINE	MC à l'USTHB (Alger)	Examinateur
M^me M. BOUDRAA	MC à l'USTHB (Alger)	Examinateur

Dédicaces

- Je dédie ce modeste travail à :
- Ma grand-mère.
- Mes parants
- Mes frères et sœurs
- La mémoire de ma tante.
- Mes amis
- Tous ceux qui me le sont chèrs.

M^{ed} Lassaâd.

Remerciements

Je remercie vivement monsieur A. DJERADI, maître de conférence à l'USTHB, pour l'honneur qu'il me fait en acceptant de présider le jury.

Je tiens particulièrement à exprimer ma profonde reconnaissance à M^{me} M. GUERTI , maître de conférence à l'ENP, pour ses nombreux conseils et suggestions qui m'ont été très enrichissant tout au cours de la réalisation de ce travail.

J'adresse mes vifs remerciements à monsieur A. HOUACINE maître de conférence à l'USTHB, d'avoir accepté d'examiner ce travail et de participer au jury.

Je remercie respectueusement Madame, M. BOUDRAA maître de conférence à l'USTHB, d'avoir bien voulu s'intéresser à ce travail et de participer au jury.

Je tiens à remercier messieurs A. AMROUCHE et M. DEBYECHE pour leurs aides et conseils. Qu'ils trouvent ici ma profonde gratitude.

Je remercie également M^{lle} H. KHELLADI pour l'aide qu'elle m'avait apporté.

Je remercie également Ms. H. SAYOUD et S. CHITROUB pour les conseils qu'ils m'avaient donnés.

Je voudrais remercier également tous ceux qui ont contribué à l'élaboration de ce travail en particulier : M^{lles} D. CHERIFI, N. LEHTIHEIT et B. CHEBINE ainsi que M. M. ABBAS.

Mes remerciements s'adressent aussi à tous mes amis et en particulier : Ms. F. BOUSSOURA, L. KACI, S. BOUKHENNOUSSE et M. M. BOUZID, qui ont toujours été disponibles pour m'apporter leur aide et leur soutien.

Enfin je tiens à exprimer toute ma reconnaissance à ma famille : mes parents, mes frères et sœurs et en particulier mes frères Hichame et Slimane pour leur aide et leur soutien.

SOMMAIRE

Chapitre 3 : CARACTERISTIQUES PROSODIQUES DE LA PAROLE ET MODELISATION DE LA DUREE.

Chapitre 4 : PREDICTION DES DUREES PHONEMIQUES.

الملخص

إن لمعرفة العناصر العروضية أهمية كبرى عند تركيب الصوت أو التعرف عليه [1, 2] . فلهذا السبب أردنا دراسة أحد هذه العناصر ألا وهي المدة الزمنية للصوت و هذا بالاعتماد على النموذج التنبئي لـ: KLATT [28 , 4] و الذي يعطي المدة الزمنية للصوت معتمدا على قيمة أساسية تسمى القيمة الذاتية.

في البدء قمنا بدراسة القيمة الذاتية للأصوات الخاصة باللغة العربية باستعمال كلمات مصطنعة بهدف أخذ كل التركيبات الصوتية الممكنة. ثم استخرجنا قواعد التنبؤ لهذه الأصوات بهدف استعمالها في تركيب الصوت من الأصوات الثنائية [9 , 10] .

بعد ذلك قمنا بإنجاز برنامج لتركيب هذه الأصوات لتجربة النتائج المحصل عليها والتي كانت حسنة.

الكلمات الخاصة : العروض المدة الصوتية المدة الذاتية التنبؤ للمدة الأصوات الثنائية.

Abstract

The knowledge of phoneme prosodic parameters has a very big importance in the speech recognition and synthesis [1, 2]. So we are interested to make a survey of phonemic duration which is one of the prosodic parameters, and this by inspiring it of a predictive model of KLATT [4, 28], which proposes a predicted duration from duration of basis called intrinsic duration.

We first measured the intrinsic duration of the Standard Arabic specific sounds, using artificial words in order to take all possible combinations of phonemes, after this we extract some prediction rules of this duration in order to use it in the synthesis by diphone [9,10].

After this we have elaborated a soft to synthesise sounds studied and to verify our results which were good.

Key words : prosody, phonemic duration, intrinsic duration, predicted duration, diphone.

Résumé

La connaissance des paramètres prosodiques a une grande importance en synthèse et en reconnaissance de la parole [1,2]. De ce fait nous nous sommes intéressés à l'étude d'un de ces paramètres pertinents, à savoir la durée phonémique. En s'inspirant du model prédictif de KLATT [4, 28], qui prédit la durée du phonème à partir d'une durée de base appelée durée intrinsèque.

En premier lieu, nous avons mesuré les durées intrinsèques des sons spécifiques à l'Arabe Standard, en utilisant des mots artificiels afin de prendre toutes les combinaisons des phonèmes possibles, puis nous avons déterminé les règles de prédiction de ces durées pour les utiliser dans la synthèse par diphones [9,10].

Par la suite nous avons élaboré un outil de synthèse pour vérifier nos résultats, qui étaient satisfaits.

Mots clés : prosodie, durée phonémique, durée intrinsèque, durée prédite, diphones.

INTRODUCTION GENERALE

Dans le Traitement Automatique de la Parole (TAP) on s'intéresse à l'analyse, la synthèse, la reconnaissance et la perception du signal vocal. Dans tous ces domaines de traitement, des paramètres pertinents doivent être extraits à savoir les paramètres prosodiques qui sont : la mélodie ou fréquence fondamentale, l'intensité ou énergie et la durée phonémique ou rythme.

Notre travail s'insère dans le cadre du TAP et particulièrement à la prosodie. Cette dernière consiste à étudier la forme de la phrase (affirmative, interrogative ou exclamative) ou bien le style (discours, récit, oral…) ainsi qu'à la variabilité du locuteur (l'âge, le sexe) et ceci afin de prendre ces effets lors de la reconnaissance ou la synthèse de la parole.

Les effets de variabilité sont localisés sur les paramètres prosodiques par :

- l'évolution de la mélodie ou fréquence fondamentale(F0) ou pitch ;
- la variation de l'intensité ou énergie ;
- la variation de la durée phonémique ou rythme.

Ces paramètres aident suivant le domaine à :

- identifier le locuteur (en reconnaissance de la parole) ;
- rendre la parole synthétique plus naturelle et intelligible.

Pour cela nous nous sommes intéressés à l'étude d'un de ces paramètres à savoir la durée phonémique et comme application l'étude à été faite sur les sons spécifiques à l'Arabe Standard tels que :

les emphatiques ([□], [δ], [□], [δ]) ;

les glottales ([ʔ]) ;

pharyngales ([ħ], [ε]).

Après le choix et l'enregistrement du corpus nous avons mesuré les durées intrinsèques de ces sons. Après cette phase nous avons mesuré les durées suivant les différents contextes, par la suite et en s'inspirant d'un modèle additif de prédiction des durées phonémiques [7], nous avons élaboré des règles de prédiction de ces sons.

Nous présentons le travail sous forme de quatre chapitres :

- dans le premier chapitre nous rappelons quelques notions fondamentales de la parole. Les particularités des sons de l'Arabe Standard sont aussi abordées ;
- nous suivons ces notions par la présentation des techniques et méthodes d'analyse et de synthèse de la parole. dans le deuxième chapitre ;
- le troisième chapitre est consacré aux paramètres prosodiques et la modélisation des durées phonémiques ainsi qu'à la méthode que nous avons étudiée ;
- les résultats et leurs interprétations sont présentés dans un quatrième chapitre, suivi d'une proposition d'un outil de synthèse de la parole que nous avons élaboré afin d'évaluer les règles obtenues ;
- ainsi nous terminons par des conclusions générales et des perspectives.

1. INTRODUCTION

Les durées phonémiques doivent être étudiées d'une manière à donner toutes leurs variabilité suivant différent contextes, pour pouvoir traiter le signale parole et de le reconnaître dans le cas de la reconnaissance, ou de le synthétise dune manière intelligible dans le cas de la synthèse de la parole. Pour cela nous avons choisi l'étude et la caractérisation des durées phonémiques de quelques sons de l'Arabe Standard afin de proposer un modèle de prédiction des durées phonémiques de ces sons.

En premier lieu nous avons étudier les durées des voyelles. Par la suite nous avons étudié les durées de quelques consonnes.

2. PREDICTION DES DUREES DES VOYELLES

L'Arabe Standard contient trois types de voyelles (haraka) appelées voyelles brèves : [a],[u],[i] ; et des voyelles longues qui sont réalisées par un allongement des voyelles brèves (el mad) : [a:], [u:], [i:].

En présence de consonnes emphatiques toutes ces voyelles sont influencées énergiquement d'où l'obtention de variantes : voyelles emphatiques et voyelles non emphatiques [34,35].

Pour étudier les durées phonémiques de ces voyelles, nous les avons mises dans des contextes de diphones constituant le corpus d'étude.

2.1. Corpus

Pour étudier les durées phonémiques des voyelles nous avons choisis l'études du cas des variantes brève/longue, l'emphase, et la position de la voyelle dans le mot, d'où le corpus est comme suit :

2.1.1. Contexte non emphatique

La voyelle est mise dans un diphone [cv] tel que v c'est la voyelle à étudier et c la consonne non emphatique.

Nous avons donc deux cas à voir suivant la position de la voyelle dans le mot : voyelle à la position milieu du mot ou à la position finale. La variante voyelle brève et

voyelle longue est également mis en évidence nous avons eu donc douze loguatomes.

- voyelle à la position finale du mot :
  - [#$cat_{NE}at_{NE}$-v#] (ex : [katata]) ;
  - [#$cat_{NE}at_{NE}$-v:#] (ex: [katata:]).
- voyelle à la position milieu du mot :
  - [#cat_{NE}-vt_{NE}a#] (ex : [katuta]) ;
  - [#cat_{NE}-v:t_{NE}a#] (ex : [katu:ta]).

2.1.2. Contexte emphatique

De même la voyelle est mise dans un diphone [cv] tel que c la consonne emphatique.

Nous avons également deux cas à voir suivant la position de la voyelle dans le mot : voyelle à la position milieu du mot ou à la position finale. La variante voyelle brève et voyelle longue est aussi mis en évidence nous avons eu donc douze loguatomes.

- voyelle à la position finale du mot :
  - [#$cat_{NE}at_{E}$-v#] (ex : [kata□ʊ]) ;
  - [#$cat_{NE}at_{E}$-v:#] (ex : [kata□ʊ]).
- voyelle à la position milieu du mot :
  - [#cat_{E}-vt_{NE}a#] (ex : [ka□ɪ t a]) ;
  - [#cat_{E}-v:t_{NE}a#] (ex : [ka□ɪ:ta]).

2.2. Mesure des durées intrinsèques

Ce corpus nous a permet d'extraire une durée qu'on l'appelle durée intrinsèque, et des lois de variation des durées phonémiques pour proposer un modèle de prédiction des durées des voyelles de l'Arabe Standard.

Après l'enregistrement et la segmentation du corpus, constituée de 24 mots répété par dix fois pour chaque locuteur c'est à dire 960 mots, nous avons mesuré les durées des voyelles ce qui nous a permit de calculer pour chaque locuteur les

valeurs moyennes des mesures ainsi que l'écart-type pour enfin extraire les valeurs des durées intrinsèques et les autres variantes.

Nous avons donc obtenu les résultats suivant :

2.2.1. Etude de la voyelle [a] et [a:]

Les mesures faites sur le corpus nous ont donnés les résultats du tableau 4.1.

phonème		[a]				[a:]			
Contexte		Non emphatique		Emphatique		Non emphatique		Emphatique	
position		milieu	finale	milieu	finale	milieu	finale	milieu	finale
Durées en ms pour les quatre locuteurs	L1	70	115	86	105	192	180	210	195
	L2	58	90	76	85	170	154	201	156
	L3	65	103	80	102	178	165	205	180
	L4	88	143	108	132	260	250	256	253
Durées moyennes en ms		70	112	87	106	200	187	218	196
Ecart type		11	19	12	16	35	37	22	35

Tableau 4.1 : Durées phonémiques des voyelles [a] et [a:] dans différents contextes.

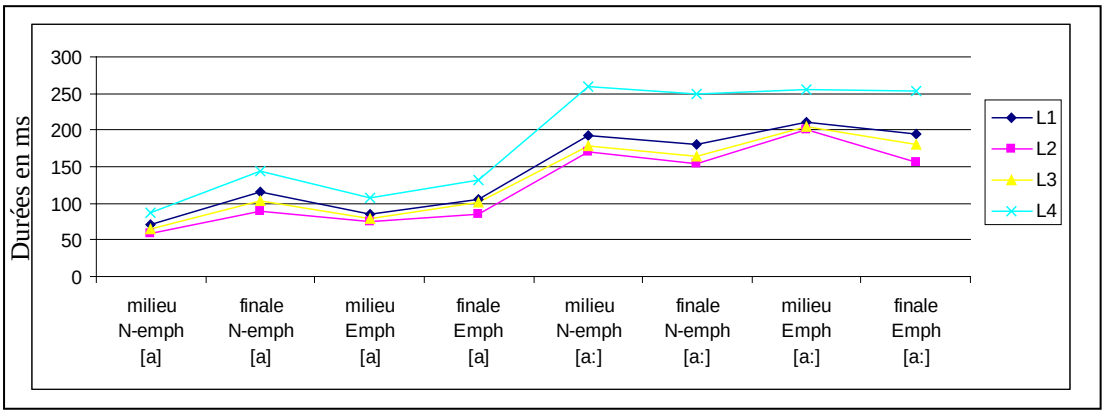

Fig.4.1. Variation des durées phonémiques des voyelles [a] et [a:].

Nous pouvons remarquer d'après les résultats du tableau 4.1. et la fig. 4.1. que la voyelle [a] obtient une durée très faible si elle se trouve dans la position du milieu ou qu'elle soit dans un contexte non emphatique par contre, s'il s'agit de la voyelle

longue [a:] cette durée devient la plus grande si la voyelle soit dans le contexte non emphatique ou dans la position du milieu.

2.2.2. Etude de la voyelle [u] et [u:]

En se qui concerne les voyelles [u] et [u:] nous résumons les résultats dans le tableau ci dessous (tableau 4.2) :

phonème		[u]				[u:]			
Contexte		Non emphatique		Emphatique		Non emphatique		Emphatique	
position		milieu	finale	milieu	finale	milieu	finale	milieu	finale
Durées en ms pour les quatre locuteurs	L1	95	110	84	117	240	197	226	217
	L2	93	103	73	98	195	170	202	155
	L3	95	110	80	120	241	178	210	167
	L4	103	122	103	128	257	240	260	254
Durées moyennes en ms		96	111	85	115	233	196	224	198
Ecart type		4	7	11	11	23	27	22	39

Tableau 4.2 : Durées phonémiques des voyelles [u] et [u:] dans différents contextes.

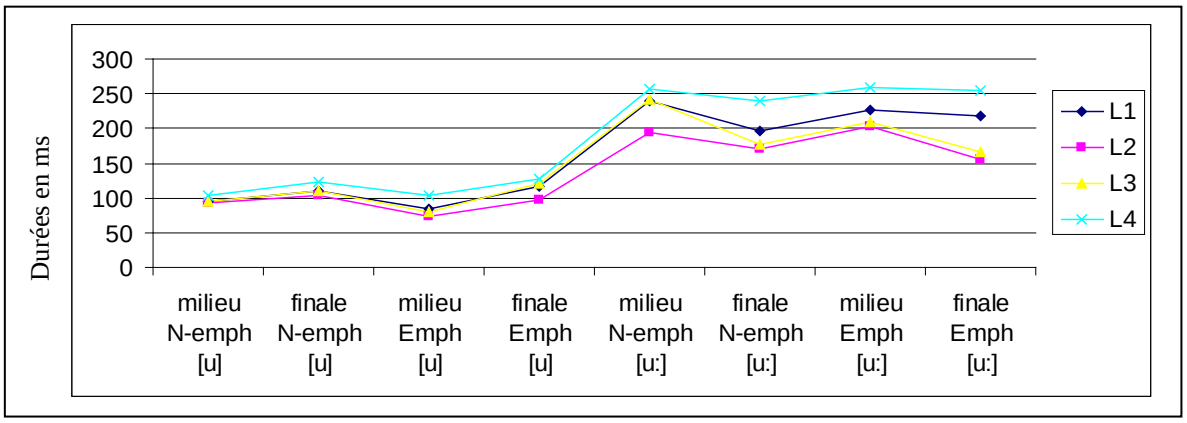

Fig.4.2. Variation des durées phonémiques des voyelles [u] et [u:].

De même pour la voyelle [u] nous pouvons remarquer d'après les résultats du tableau 4.2. et la fig. 4.2. que la voyelle [u] obtient une durée très faible si elle se trouve dans la position du milieu ou qu'elle soit dans un contexte non emphatique par

contre, s'il s'agit de la voyelle longue [u:] cette durée devient la plus grande si la voyelle soit dans le contexte non emphatique ou dans la position du milieu.

2.2.3. Etude de la voyelle [i] et [i:]

Egalement pour les voyelles [i] et [i:] les mesures faites sur le corpus nous ont donnés les résultats du tableau 4.3.

phonème		[i]				[i:]			
Contexte		Non emphatique		Emphatique		Non emphatique		Emphatique	
position		milieu	finale	milieu	finale	milieu	finale	milieu	finale
Durées en ms pour les 4 locuteurs	L1	82	105	96	110	235	210	240	226
	L2	75	95	76	98	200	160	222	197
	L3	75	97	85	102	209	162	223	210
	L4	84	121	114	128	274	252	258	247
Durées moyennes en ms		79	104	92	109	229	196	237	220
Ecart type		4	10	14	11	28	38	14	18

Tableau 4.3 : Durées phonémiques des voyelles [i] et [i:] dans différents contextes.

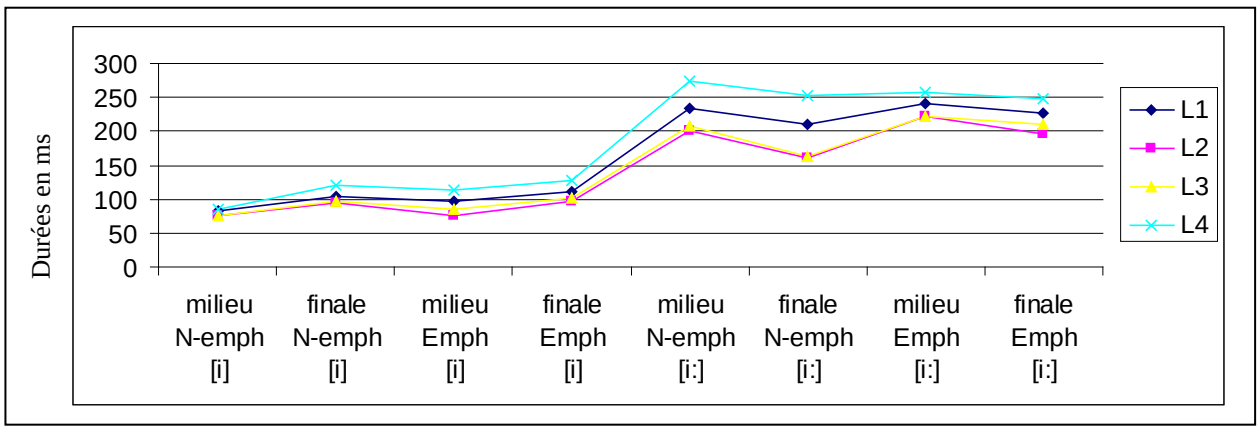

Fig.4.3. Variation des durées phonémiques des voyelles [i] et [i:].

Les remarques faites sur les voyelles précédentes restent valables pour la voyelle [i]. Les résultats du tableau 4.2. et la fig. 4.2. montrent que la voyelle [i] a une durée faible si elle se trouve au milieu ou dans un contexte non emphatique. Par contre la durée de la voyelle longue [i:] augmente si cette dernière se trouve en position milieu ou dans un contexte non emphatique.

3. MODELE DE PREDICTION DES DUREES DES VOYELLES

Afin d'extraire un modèle de prédiction des durées phonémiques des voyelles suivant différent contextes nous avons mesuré les durées intrinsèques puis calcules les rapports de durées suivant les contextes (durées des voyelles au Milieu **/** durées des voyelles à la fin), (durées des voyelles Longues **/** durées des voyelles brèves), (durées des voyelles emphatiques **/** durées des voyelles non emphatiques) [34,35]. Nous résumons ces résultats dans les tableaux 4.4 et 4.5.

phonème		[a]		[a:]		[u]		[u:]		[i]		[i:]	
contexte		Non emph	Emph	Non emph	Emph	Non emph	Emph	Non emph	Emph	Non emph	Emph	Non emph	Emph
Durées en ms	Milieu	70	87	200	218	96	85	233	224	79	92	229	237
	Fin	112	106	187	196	111	115	196	198	104	109	196	220
Rapport des durées Milieu/fin		0.62	0.83	1.07	1.13	0.87	0.73	1.20	1.16	0.75	0.84	1.19	1.07

Tableau 4.4 : Durées moyennes et leurs rapports de variation contextuelles.

Phonèmes / Contextes			[a]	[u]	[i]
Emp/N.Emp	brèves	Milieu	1.25	088	1.17
		Fin	0.94	1.04	1.05
	Longues	Milieu	1.10	0.97	1.03
		Fin	1.05	1.00	1.015
Long/brev	Emph	Milieu	2.50	2.65	2.58
		Fin	1.84	1.70	2.02
	Non Emph	Milieu	2.84	2.41	2.90
		Fin	1.66	1.76	1.86

Tableau 4.5 : Rapport des dures des voyelles suivant leurs contextes.

D'après les résultats des tableaux 4.4 et 4.5 on conclue que les voyelles dans le contexte emphatique et non emphatique ont des durées proches, mais on n'a pas pu trouver une loi liant ces durées du fait que leurs caractéristiques énergétiques défères. Donc on propose de prendre la durée intrinsèque de la voyelle dans le contexte emphatique différente par rapport à celle du contexte non emphatique.

Nous concéderons la durée de la voyelle dans la position finale du mot comme étant la durée intrinsèque du fait qu'elle ne donne pas d'influence de coarticulation à la fin de la voyelle ce qui nous donne :

Contexte		Emphatique	Non emphatique
Durées en ms.	[a]	106	112
	[u]	115	111
	[i]	109	104

Tableau 4.6 Durées intrinsèques des voyelles.

Nous proposons donc de calcules les durées intrinsèques des voyelles suivant la relation (5.1) dont on utilise une variable booléenne E pour exprimer l'emphase afin de prendre la valeur de la durée suivant le contexte emphatique D_{ie} ou non emphatique D_{ine} (tableau 4.7).

$$D_i = E.D_{ie} + \overline{E}.D_{ine} \qquad (4.1)$$

Finalement et pour prédire suivant tous les contextes les durées phonémiques des voyelles nous exploitons la relation (4.1) avec les coefficient calculée à partires des mesures des tableaux ci-dessus, nous proposons donc la relation permettant de prédire les durées phonémiques des voyelles (4.2) que nous avons élaboré en s'inspirant du modèle des représentations vectorielles après plusieurs itérations sur les modèles que nous avons étudiés et appliqués sur nos résultats.

Nous avons donc regroupé dans la relation de ce modèle la valeur intrinsèque de la durée et les cas des positions milieu et fin de la voyelle dans le mot, ainsi que le cas de la voyelle longue ou brève, le cas du contexte emphatique et non emphatique est mis dans la relation de la durée intrinsèque (4.1).

$$D_t = [\sqrt{\left((M.K_m)^2 + (L.K_l)^2\right)} + (\overline{M + L})].D_i \qquad (4.2)$$

Avec :

- D_t durée totale de la voyelle ;
- M : variable booléenne indiquant la position de la voyelle (M=1 position milieu, M =0 position finale) ;
- L : variable booléenne indiquant la variante de la voyelle (L=1 voyelle longue, M =0 voyelle brève) ;
- K_m : proportion de la position milieu tableau 4.7 ;
- K_l : proportion de la variante longue tableau 4.7.

contexte	phonèmes	[a] ou [a:]	[u] ou [u:]	[i] ou [i:]
Emphatique	D_i (en ms)	106	115	109
	K_m	0.83	0.74	0.85
	K_l	1.85	1.72	2.02
Non emphatique	D_i (en ms)	112	111	104
	K_m	0.62	0.87	0.75
	k_l	1.67	1.76	1.89

Tableau 4.7 : Paramètres de prédiction des durées phonémiques des voyelles.

Afin de mieux comprendre ces résultats nous schématisons les durées phonémiques des voyelles sur les sommets d'un cube fig. 4.6 et sur les arrêtes nous donnons les coefficients de variation ou de passage d'un contexte à un autre.

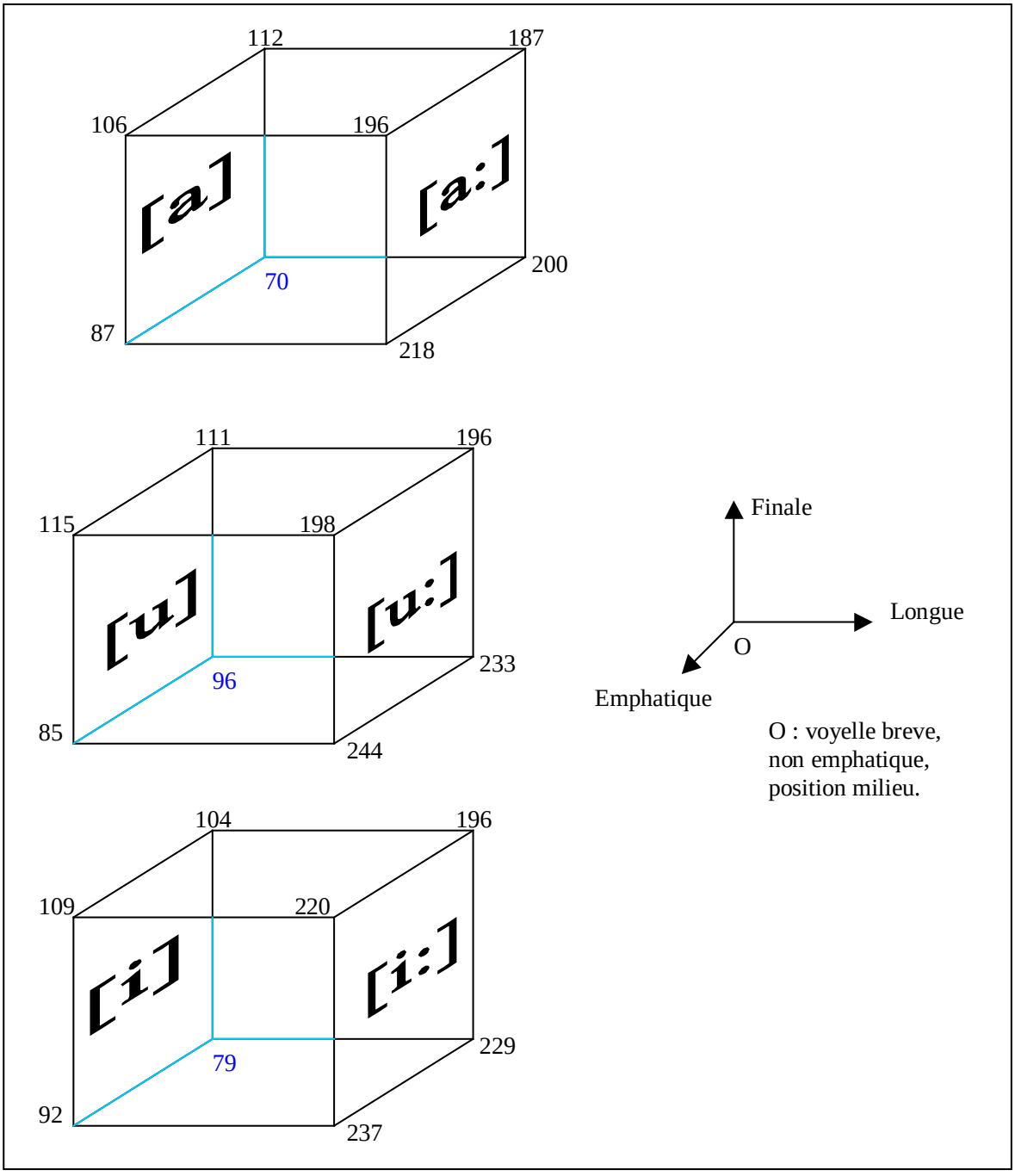

Fig. 4.6 : Représentation schématiques des durées phonémiques des voyelles.

3.1. ALGORITHME DE PREDICTION DES VOYELLES

Le modèle de prédiction des durées phonémiques des voyelles que nous avons élaboré et régie par l'algorithme suivant :

Début

Lire les diphones C_iV_i et V_iC_{i+1} (l'unité phonémique $C_iV_iC_{i+1}$)

Si C_i est emphatique alors E=1 sinon E=0

Si C_{i+1} = # alors M=0 sinon M=1

Si Vi est une voyelle longue alors L=1 sinon L=0

Lire la nature de V_i : V_i=a, u, i

Lire les paramètres D_i, K_m, K_l

Calculer

$$D_i = E.D_{ie} + \overline{E}.D_{ine}$$

$$D_t = [\sqrt{\left((M.K_m)^2 + (L.K_l)^2\right)} + (\overline{M + L})].D_i$$

Fin.

Pour mieux comprendre le modèle que nous venons de proposer, nous le schématisons par l'organigramme ci-dessous :

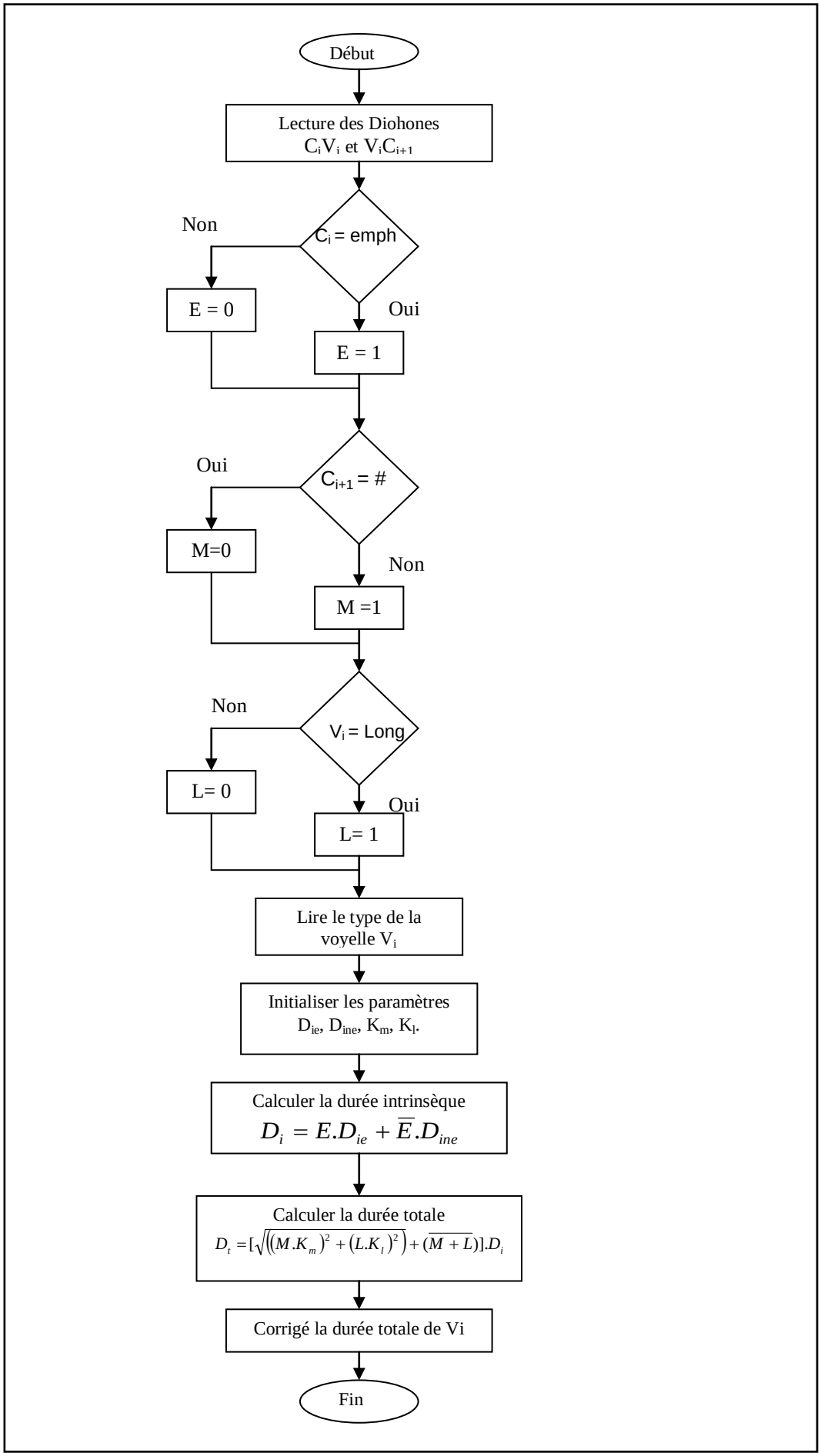

Fig. 4.7 : Organigramme de prédiction des durées des voyelles de l'A.S.

4. PREDICTION DES DUREES PHONEMIQUES DES CONSONNES SPECIFIQUES A L'AS

Comme nous avons vu, les consonnes de l'Arabe Standard ont différentes caractéristiques : occlusives, fricatives, glottales... et comme particularité de cette langue les consonnes emphatiques (son plus énergétique) qui est une caractéristique des langues sémitiques [3,12,14].

Nous avons élaboré le corpus de la manière à mètre les consonnes à étudiées dans les diphones [vc] suivant le dictionnaire de diphone de M.GUERTI [10] et nous les avons mis dans des logatomes suivant les positions début, milieu, et fin du mot. Nous avons donc à étudier 24 logatomes :

- [#c-ata #]: pour l'étude des durée intrinsèque des consonnes ;
- [#kata-c #]: afin de voir l'effet de la position finale sur la durée de la consonne.
- [#ka-cta #]: pour voir l'effet de la position milieu sur la durée de la consonne.

En premier lieu nous avons choisit les consonnes occlusifs : [ʔ],[q], [ð],[□].

En second lieu nous avons étudier les consonnes fricatives : [ð],[ε],[ħ], [□].

4.1. Durées des consonnes occlusives

Etant donnée que les sons plosifs (occlusifs) sont caractérisés par un temps de silence (tenue consonantique) et un temps de phonation appelle plosion, nons avons préféré mesurer en premier lieu les durées des plosions et de voir leur variation ; Par la suite nous avons étudier les durées des tenues consonantiques et les durées totales, finalement nous avons proposé un modèle de prédiction des durées des consonnes étudiées [35,36].

Pour étudier les durées des consonnes plosives nous avons eu 12 mots répétés dix fois pour chaque locuteur c'est à dire 480 logatomes.

4.1.1. Durées des plosions consonantiques

Après l'enregistrement et la segmentation du corpus nous avons mesuré les durées phonémiques des consonnes plosives suivant différent contextes (début, milieu et fin du mot) ainsi que la plosion et la tenue consonantique. Par la suite nous avons calculé les valeurs moyennes des mesures pour chaque locuteur ainsi que l'écart-type. Nous présentons donc les résultats des durées phonémiques des plosions des consonnes occlusives dans le tableau 4.8.

La figure fig. 4.8 montre l'évolution des plosions consonantiques suivant leurs contextes et par rapport au différent locuteurs.

Plosions des phonèmes		[d,]			[t,]			[?]			[q]		
position		début	milieu	fin	début	milieu	fin	début	milieu	fin	début	milieu	fin
Durées en ms pour les quatre locuteurs	L1	15	15	18	22	12	22	25	20	28	25	16	28
	L2	18	12	18	22	11	27	24	14	26	23	14	29
	L3	16	13	20	21	13	25	21	16	25	25	15	27
	L4	15	14	20	22	20	25	22	20	27	27	21	27
Durées moyennes en ms		16	13	19	22	14	24	23	17	26	25	16	27
Ecart type		1.22	1.15	1.01	0.43	3.53	1.78	1.58	2.59	1.11	1.41	2.62	0.82

Tableau 4.8 : Mesures des durées des plosions consonantiques.

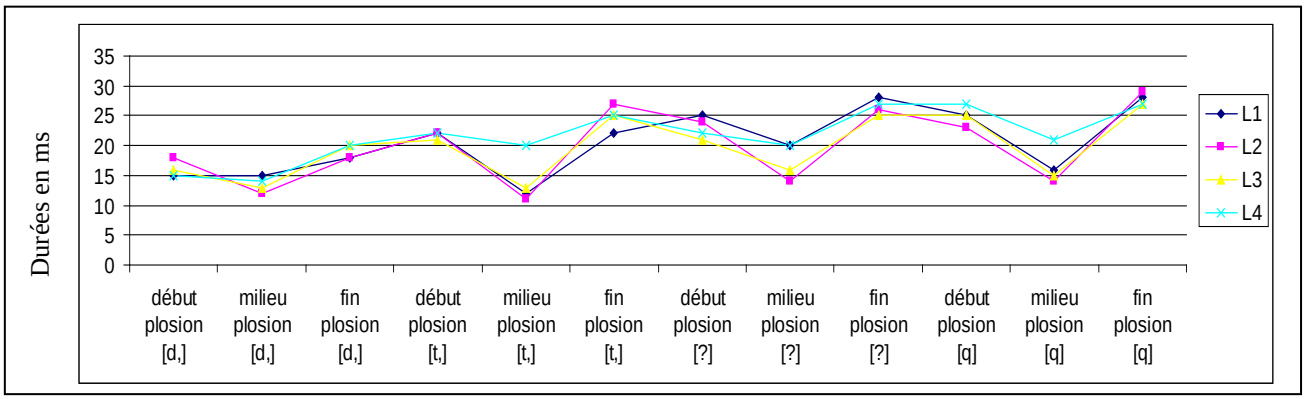

Fig. 4.8 : Variation des durée des plosions consonantiques.

4.1.2. Durées totales des consonnes occlusives

Les consonnes occlusives comme nous venons de citer se caractérisent par la juxtaposition d'une plosion et d'un silence (tenue consonantique).

Nous avons données ci-dessus les mesures des plosion puis nous donnons les mesures des durées totales (tenue +plosion) tableau 4.9 et Fig.4.9.

Plosions des phonèmes		[d,]		[t,]		[?]		[q]	
position		milieu	fin	milieu	fin	milieu	fin	milieu	fin
Durées en ms pour les quatre locuteurs	L1	103	153	106	200	150	208	106	195
	L2	95	130	102	182	114	116	98	157
	L3	101	130	110	180	130	130	108	183
	L4	112	194	120	270	156	272	118	124
Durées moyennes en ms		102	151	109	208	137	181	107	164
Ecart type		06.12	26.10	06.88	36.67	16.67	62.90	07.12	27.20

Tableau 4.9 : Durées totales des consonnes occlusives.

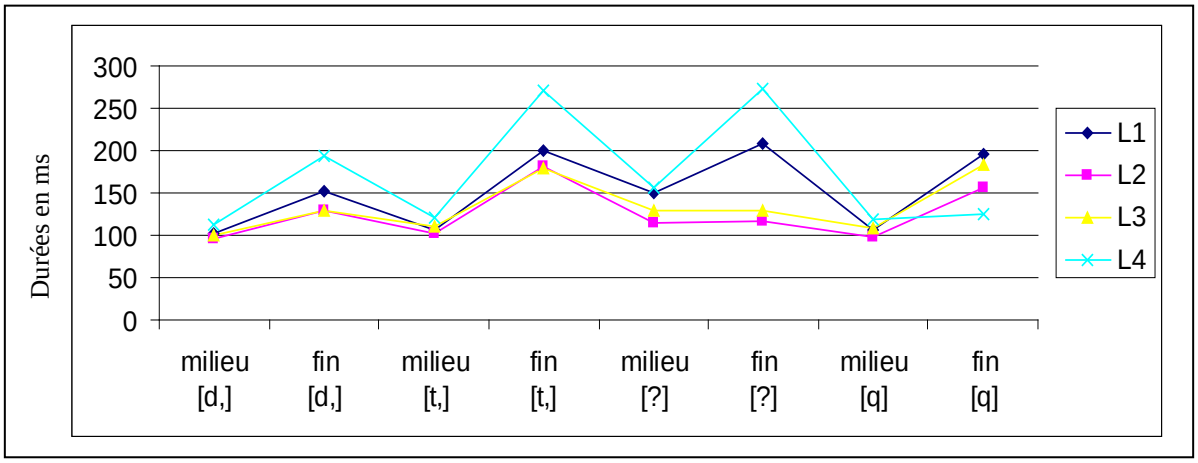

Fig. 4.9 : Variation des durée totales des consonnes occlusives.

Remarque : *nous n'avons pas mesuré les durées totales des consonnes occlusives dans le contexte début du mot du fait que nous avons choisis un corpus de mots isolés ce qui ne permet pas de connaître le début de la réalisation phonémique des consonnes occlusives (la tenue consonantique étant un silence et le début du mot est un silence aussi). Nous proposons par la suite une prédiction de ces durées.*

4.1.3. Prédiction des durée des consonnes occlusives

Afin de prédire les durées phonémiques des consonnes occlusives nous avons calcules les rapports de variation entre les durées suivant différent contextes : suivant la position (durée de la position fin /durée de la position début), (durée position de la milieu /durée de la position début),et les rapports des durées de la consonnes (durée de la plosion / durée totale de la consonne) nous regroupons les résultas dans le tableaux 4.9.

phonèmes		[d.]	[t.]	[?]	[q]
Durée de la plosion en ms	début	16	22	23	25
	milieu	13	14	17	16
	fin	19	24	26	27
Rapport	Fin/début	1.20	1.14	1.16	1.12
	Milieu/début	0.85	0.64	0.76	0.66
Durées totales en ms	milieu	102	109	137	107
	fin	151	208	181	164
Rapport (Plosion/ totale)	milieu	0.13	0.13	0.12	0.15
	fin	0.13	0.12	0.16	0.17
	moyenne	0.13	0.12	0.14	0.16
Durées Totale début en ms		123	183	164	156
Rapport	Totale fin/totale début	1.23	1.14	1.11	1.06
	Totale milieu/totale début	0.84	0.60	0.84	0.69

Tableau 4.9 : Rapports de variation des durées consonantiques.

D'après ces mesures nous proposons de prédire les durées phonémiques des consonnes occlusives de la position début du mot à partir de celle de la position finale comme suit :

$$D_{td} = D_{te} + D_p \quad (4.3).$$

$$D_{te} = K \cdot D_p \quad (4.4).$$

$$D_{td} = (1+K). D_p \quad (4.5).$$

Avec :

- D_{td} : durée totale de la position début du mot ;
- D_{te} : durée de la tenue consonantique ;
- D_p : durée de la plosion ;
- K : rapport entre la durée de la tenue et la plosion.

4.2. Durées des consonnes fricatives

Contrairement aux consonnes occlusives les consonnes fricatives ne présentent pas un temps de silence et un temps de phonation. Nous proposons donc de voir quelques cas de ces consonnes à savoir celle spécifiques à l'arabe Standard.

Nous avons choisis d'étudier les consonnes : [δ],[ε],[ħ], [□].

Pour étudier les durées de ces consonnes fricatives nous avons eu 12 mots répétés dix fois pour chaque locuteur c'est à dire 480 logatomes.

Après l'enregistrement et la segmentation du corpus nous avons mesuré les durées phonémiques des consonnes fricatives suivant différent contextes (début, milieu et fin du mot). Par la suite nous avons calculé les valeurs moyennes des mesures pour chaque locuteur ainsi que l'écart-type. Nous présentons donc les résultats des durées phonémiques des consonnes fricatives dans le tableau 4.10 et fig. 4.10. [35, 37].

phonèmes		[□]			[ħ]			[ε]			[δ]		
position		début	milieu	fin	début	milieu	fin	début	milieu	fin	début	milieu	fin
Durées en ms pour les 4 locuteurs	L1	134	85	244	125	95	196	112	68	141	115	90	158
	L2	140	102	215	87	90	168	88	52	139	89	86	112
	L3	100	80	182	128	96	160	88	53	122	86	94	105
	L4	145	86	272	115	94	150	123	70	165	94	98	124
Durées moyennes en ms		129	88	228	113	93	168	102	60	141	96	92	124
Ecarts types		17.66	08.25	33.40	16.17	02.27	17.11	15.25	08.28	15.33	11.33	04.47	20.36

Tableau 4.10 : Durées phonémiques des consonnes spécifiques à l'A.S.

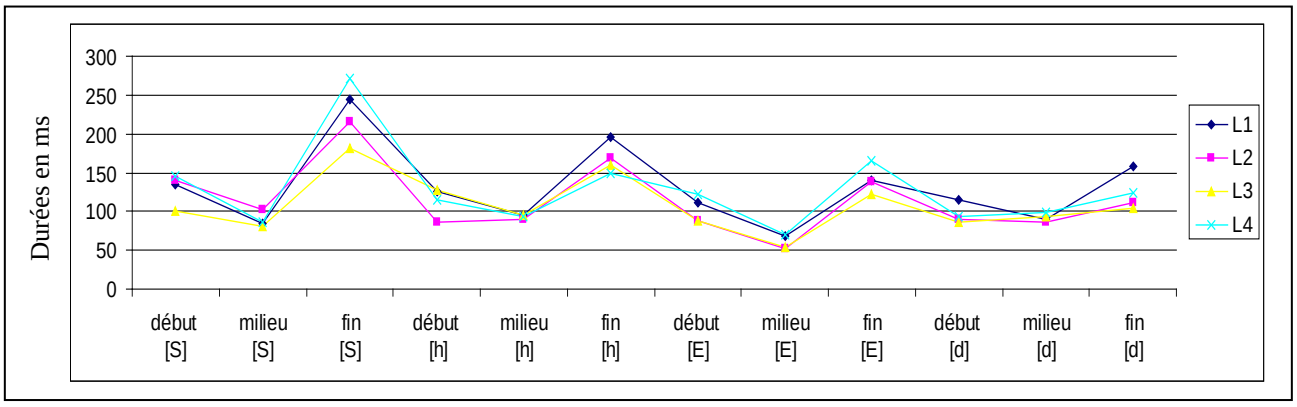

Fig. 4.10 : Variation des durées phonémiques des consonnes spécifiques à l'AS.

phonème		[□]	[ħ]	[ε]	[δ]
Durées moyennes en ms	début	129	113	102	96
	milieu	88	93	60	92
	fin	228	168	141	124
Rapport	Fin/début	1.76	1.51	1.39	1.29
	Milieu/début	0.69	0.84	0.59	0.97

Tableau45.11 : Durées phonémiques des consonnes et leurs rapports de variation.

Dans le tableau 4.11 nous présentons un résumé des mesures faites sur les durées phonémiques des consonnes, à savoir les durées dans les contextes début, milieu, et fin des mots ainsi que leurs rapports de variation (entre les différents contextes).

D'après ces résultats, nous proposons de garder le même modèle de prédiction proposé pour les consonnes plosives cf. équations : 4.3, 4.4 et 4.5 en tenant en considération le temps de silence D_{te} nul (tenue consonantique nulle) cf. équation : 4.3 c'est à dire k =0 de l'équation 4.4.

En tenant compte de l'effet de la position du diphonème dans le mot nous proposons de calculer les durées phonémiques à l'aide des relations ci-dessous.

$$D_t = D_i + D_i .(F.K_F - M.K_m)$$

(4.6).

$$D_p = K_p . D_t \qquad (4.7).$$

$$D_{te} = D_t - D_p \qquad (4.8).$$

Avec :

- D_t : durée totale de la consonne ;
- D_i : durée de totale de la consonne dans la position début du mot (considérée comme étant la durée intrinsèque) ;
- F et M : variables booléennes représentant respectivement la position finale et milieu de la consonne dans le mot ;
- K_F et km coefficients en fonction de la position fin ou milieu ;
- D_p : durée de la plosion ;

- Dte : durée de la tenue consonantique.

Ces coefficients sont donnés dans le tableau 4.12.

phonème	[δ]	[□]	[ʔ]	[q]	[□]	[ħ]	[ε]	[δ]
Durée totale début ms	123	183	164	156	129	113	102	96
K_F %	23	14	11	6	76	51	39	29
Km %	16	40	16	31	31	16	41	3
K_P %	13	12	14	16	100	100	100	100

Tableau 4.11 : paramètres de prédiction des durées phonémiques des consonnes.

Finalement nous donnons l'algorithme de prédiction des durées phonémiques des consonnes qu'elles soient plosives ou non comme suit :

Début

Lire les diphones V_iC_i et C_iV_{i+1} (l'unité phonémique $V_iC_iV_{i+1}$)

Si V_i = # alors C_i est dans la position début du mot (F=0 et M =0) sinon

Si V_{i+1} = # alors C_i est dans la position finale du mot (F=1 et M =0) sinon

C_i est dans la position milieu du mot (F=0 et M =1).

Lire la nature de C_i : C_i= ?, q, δ, □, δ, ε, ħ, □

Lire les paramètres D_i, Km, K_F, K_p

Calculer

$D_t = D_i + D_i .(F.K_F - M.K_m)$

$D_p = K_p . D_t$

$D_{te} = D_t - D_p$

Fin.

Nous donnons également l'organigramme du modèle :

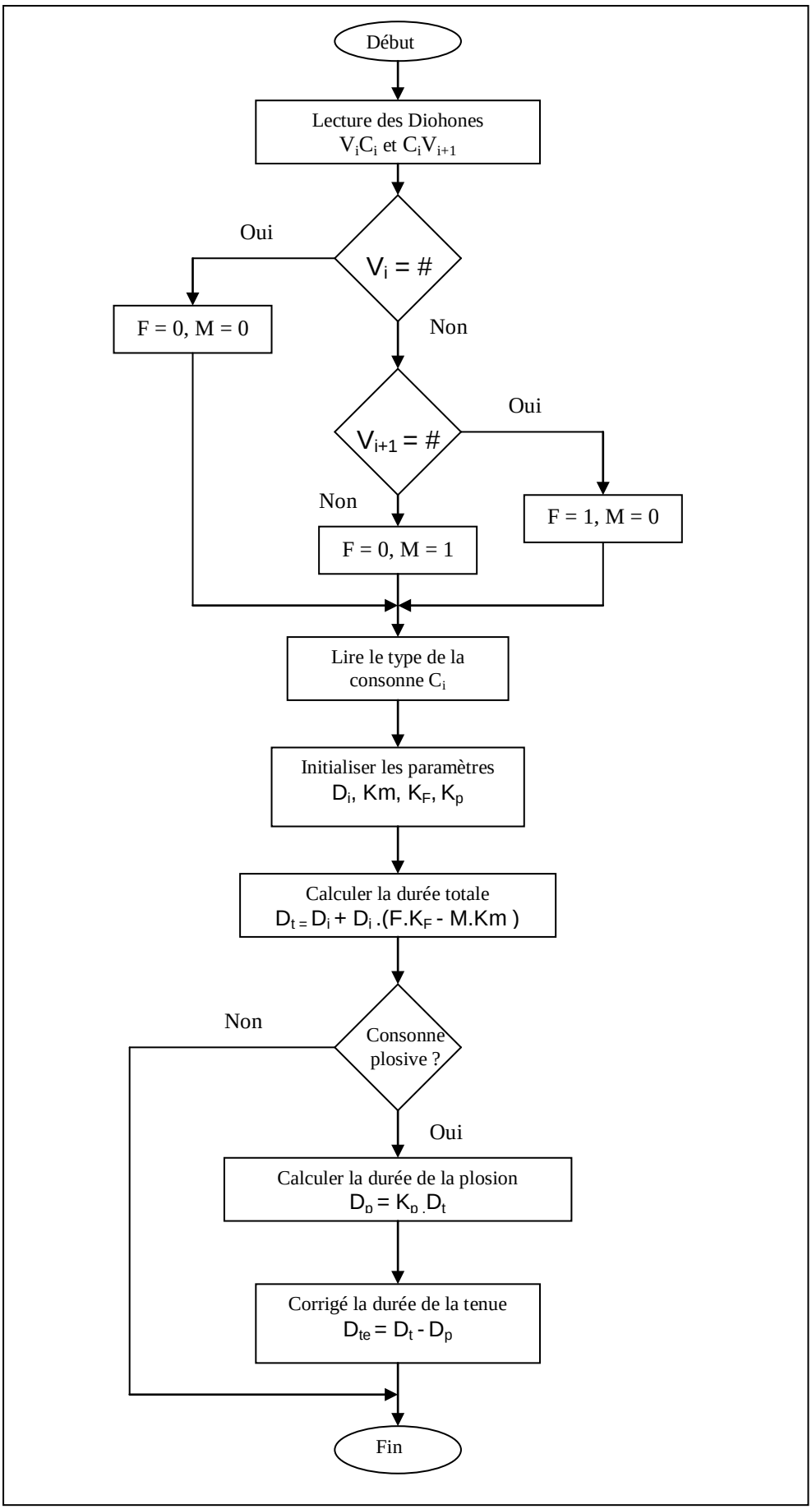

Fig. 4.11: Organigramme de prédiction des durées des consonnes de l'A.S.

5. OUTILS DE VERIFICATION PAR SYNTHESE

Après avoir élaborer les règles de prédiction des durées phonémiques des sons étudiés, nous sommes passés à la vérification et ceci par une concaténation manuelle qui a donné des résultats satisfaisantes suivant le témoignage de six personnes après une réécoute.

Puis nous avons élaboré un outil à l'aide du langage de programmation Delphi [38], afin d'assurer l'implémentation de dans nos règles et de faire une synthèse automatique de la parole.

Le logiciel élaboré a comme menu principal (Fig.4.12) :

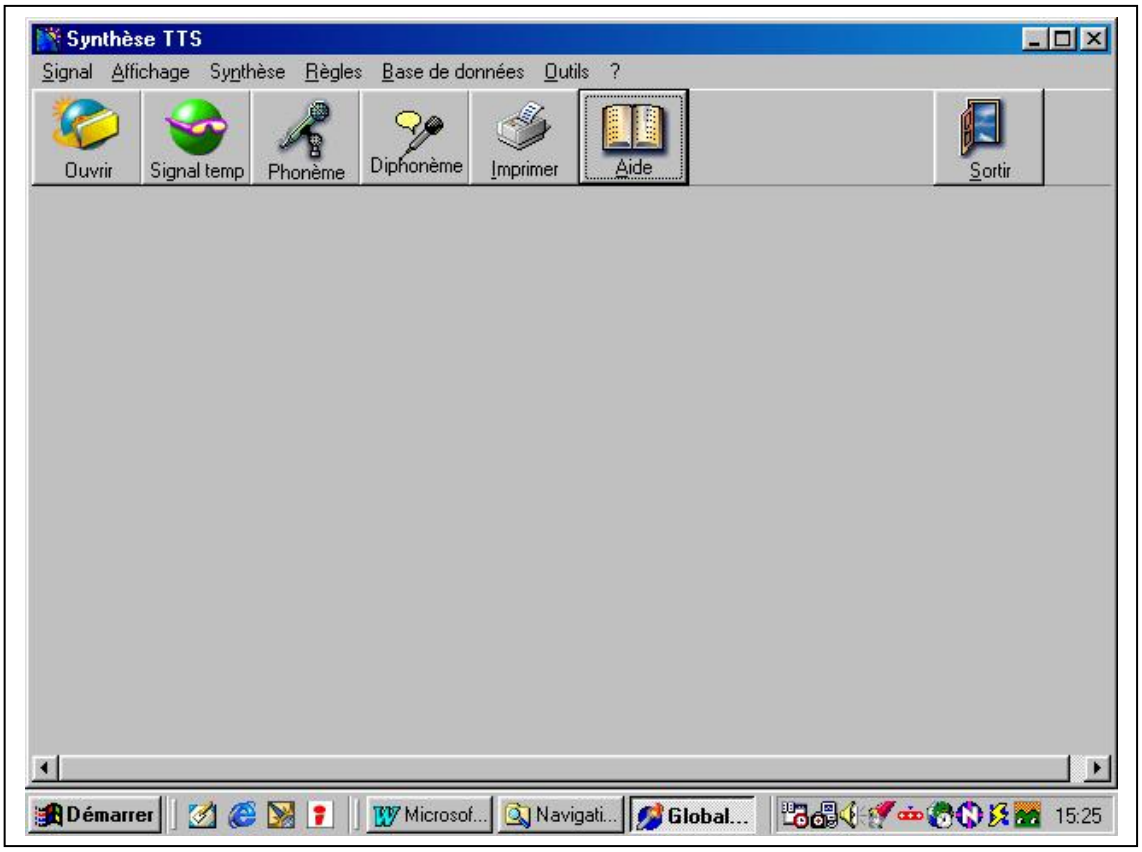

Fig.4.12 Menu principal du logiciel de synthèse de la parole TTS

Les boutons Phonème, Diphonèmes permettent d'activer une fenêtre de synthèse par phonème ou par diphonème (Fig.4.13), (Fig.4.14) :

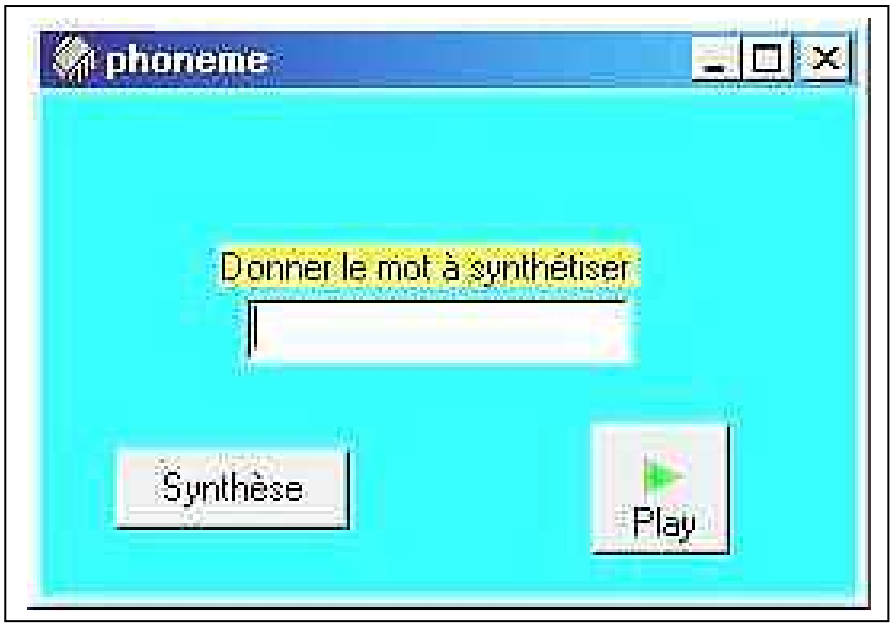

Fig.4.13 : Fenêtre de synthèse par phonèmes

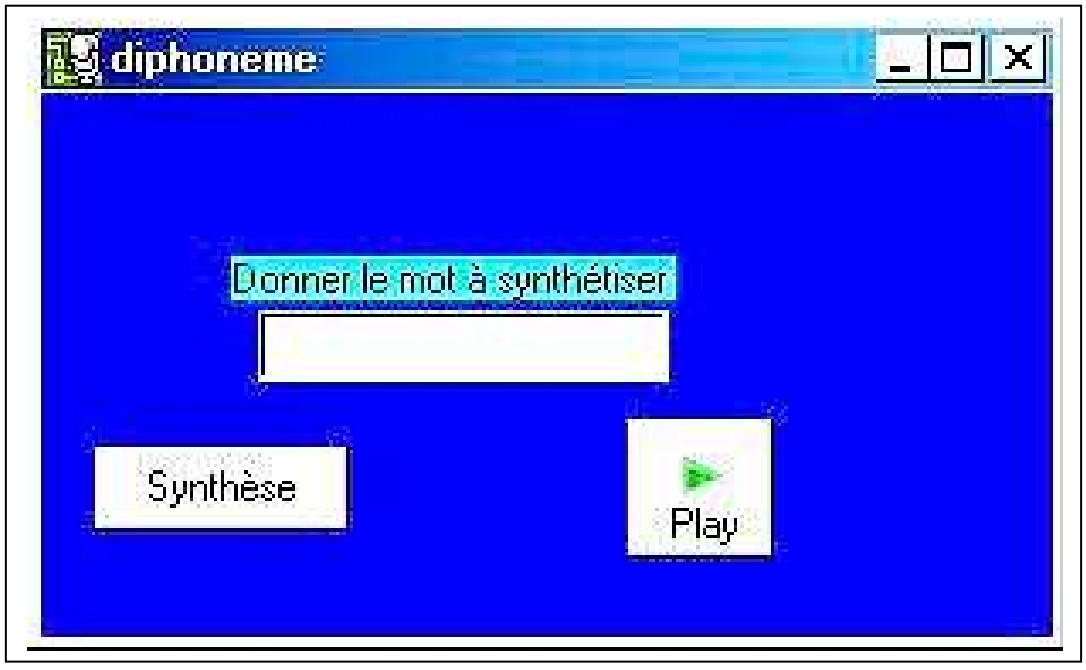

Fig.4.14 : Fenêtre de synthèse par diphonèmes

On peut également visualiser le signal obtenu ou bien un autre signal, par le bouton "signal temp".

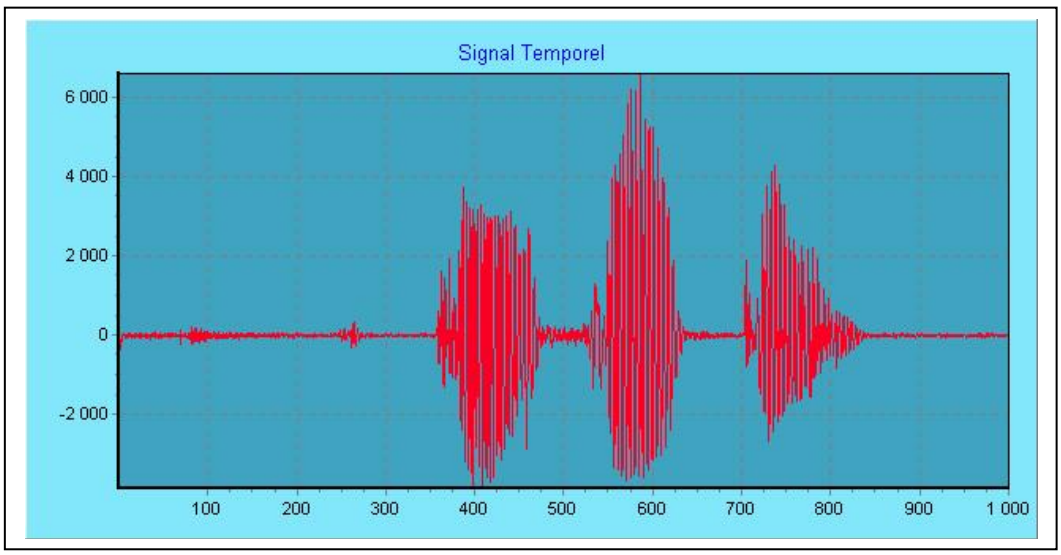

Fig. 4.15 : Fenêtre d'affichage du signal temporel.

Les courbes de variation d'énergie et de la mélodie peuvent être visualisées à l'aide des commandes affichage énergie et affichage Fréquence fondamentale Fig. 4.6.

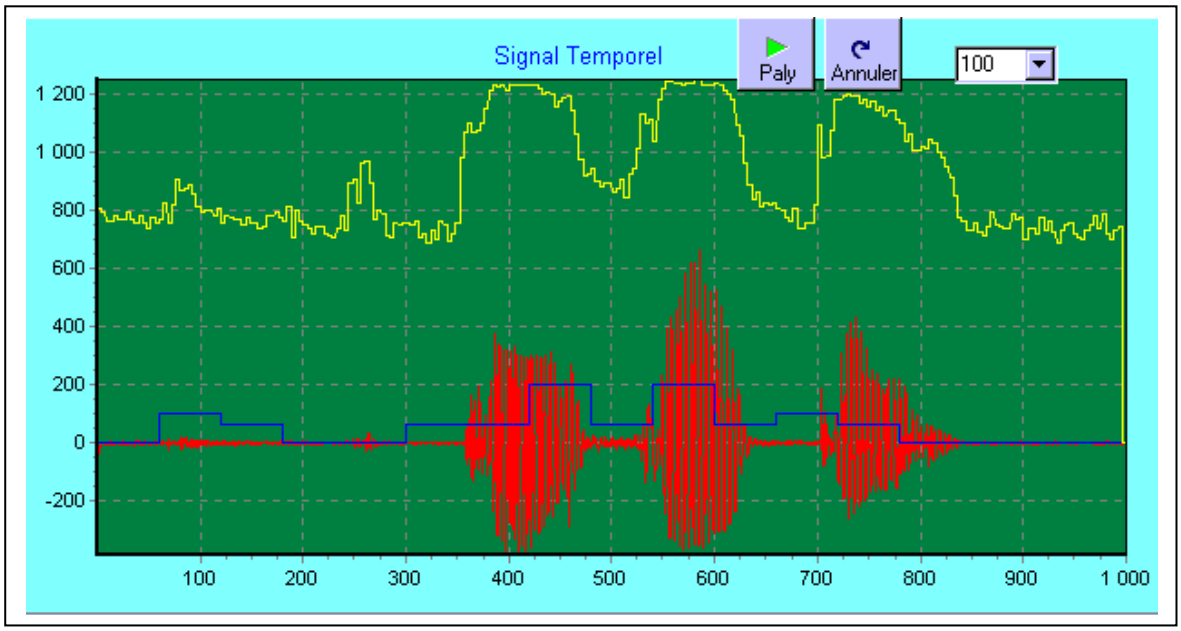

Fig. 4.16 : Fenêtre d'affichage du signal temporel, Energie et mélodie.

6. QUELQUES EXEMPLES DE SYNTHESE

Le logiciel est fait d'une manière à être évolutif, et sa base de donnée de phonèmes ou de diphonèmes peut être complétée par la prise en considération des autres effets prosodiques.

Nous donnons dans ce qui suit quelques exemples de synthèse élaborés pour les vérification de nos résultats.

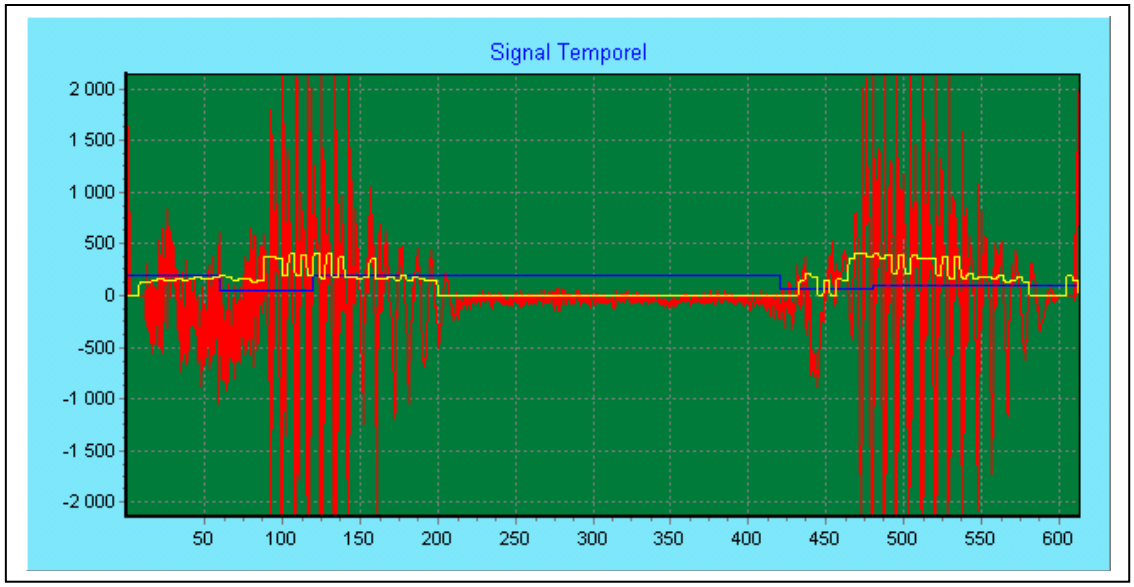

Fig. 4.17 : représentation du signal temporel, Energie et mélodie du mot [ha□a] après synthèse.

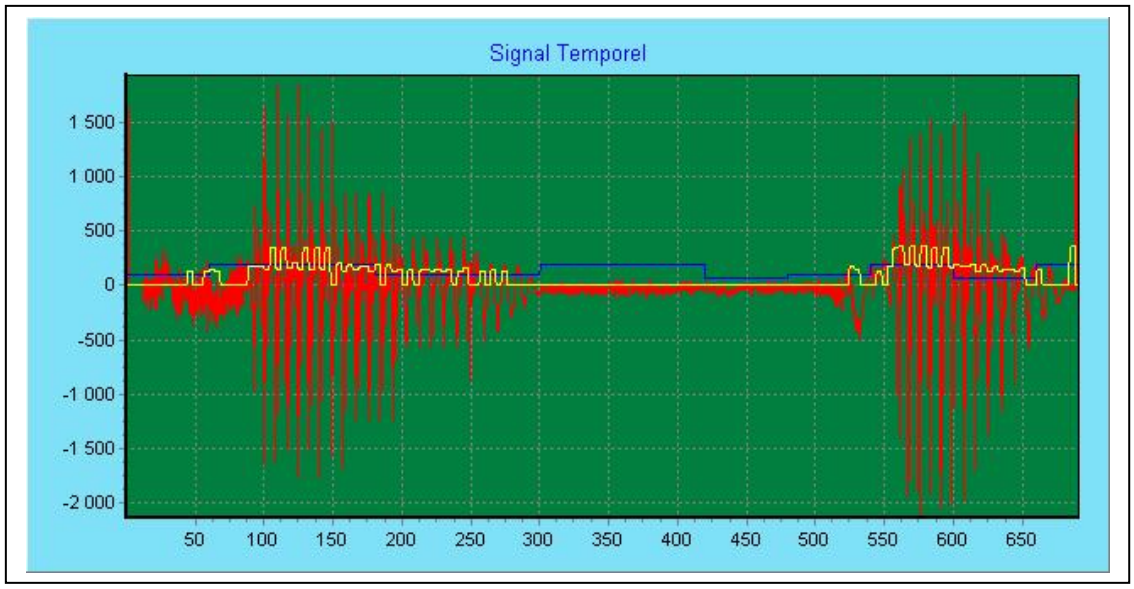

Fig. 4.18 : représentation du signal temporel, Energie et mélodie du mot [ħa:□a] après synthèse.

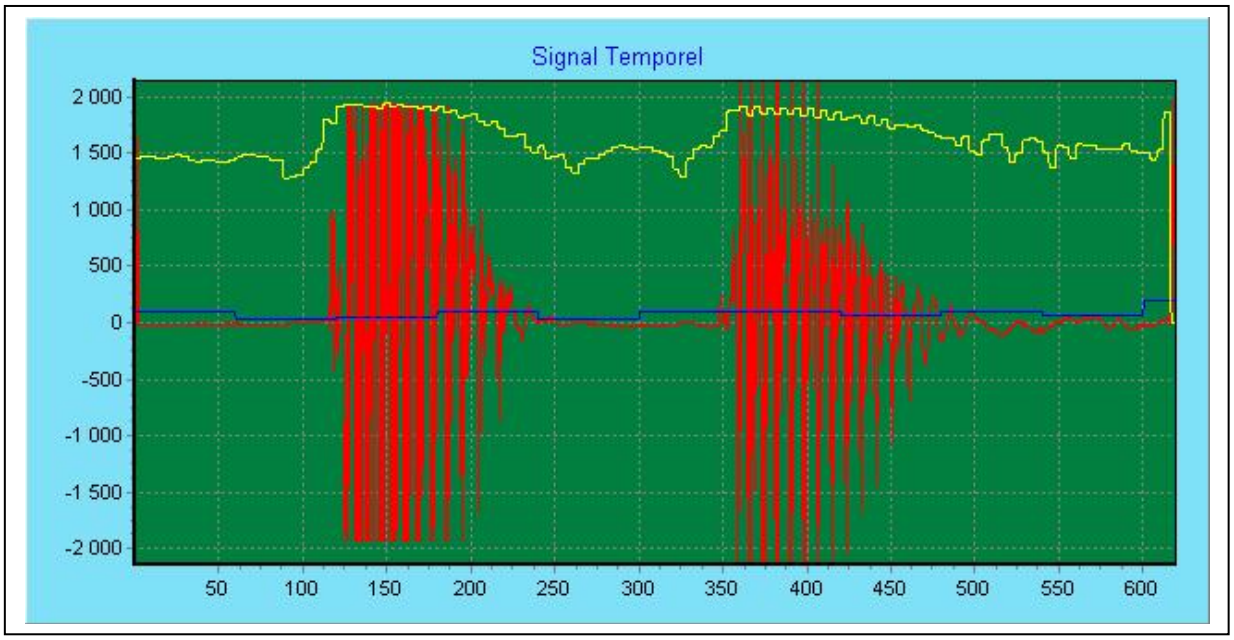

Fig. 4.19 : représentation du signal temporel, Energie et mélodie du mot [qatu] après synthèse.

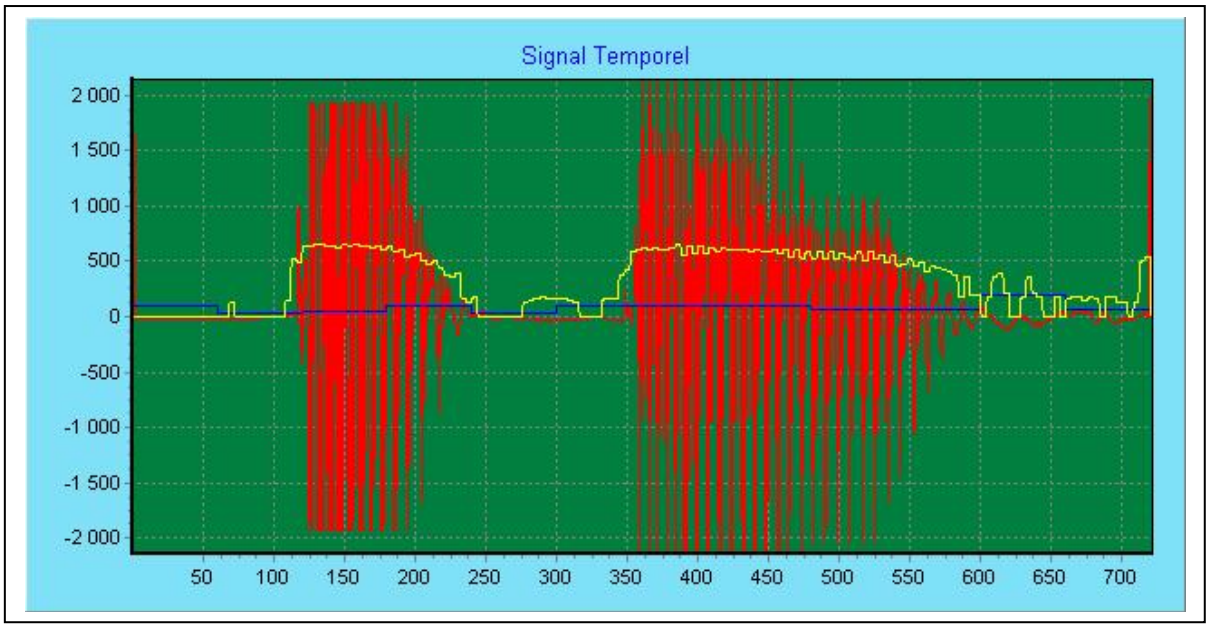

Fig. 4.20 : représentation du signal temporel, Energie et mélodie du mot [qatu:] après synthèse.

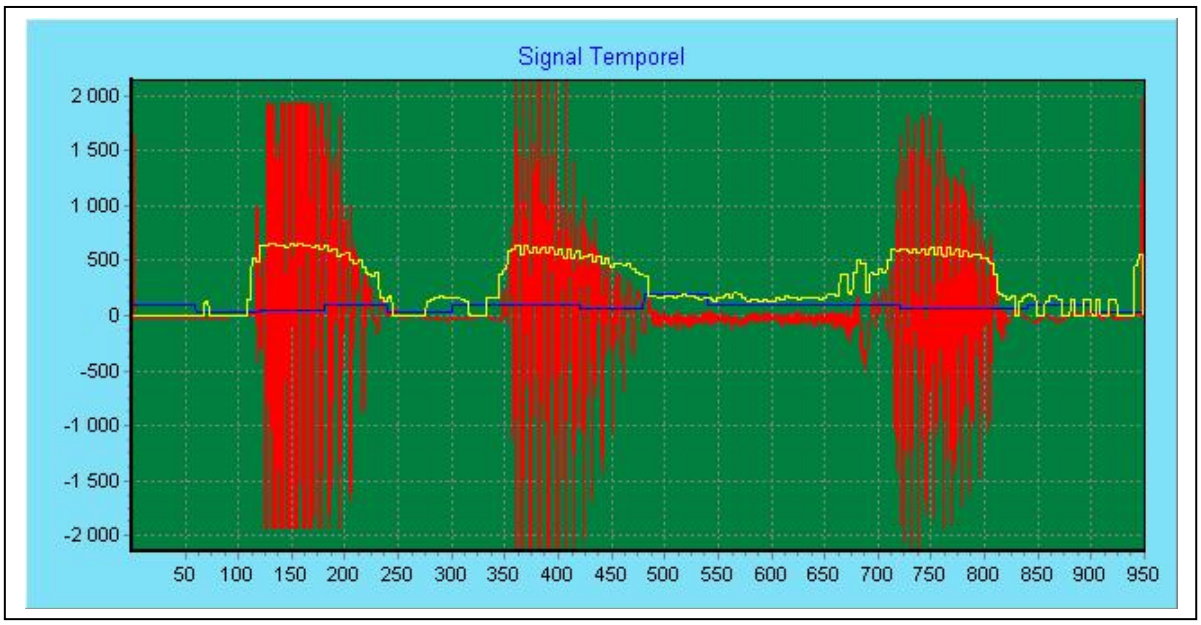

Fig. 4.21 : représentation du signal temporel, Energie et mélodie du mot [qatu□a] après synthèse.

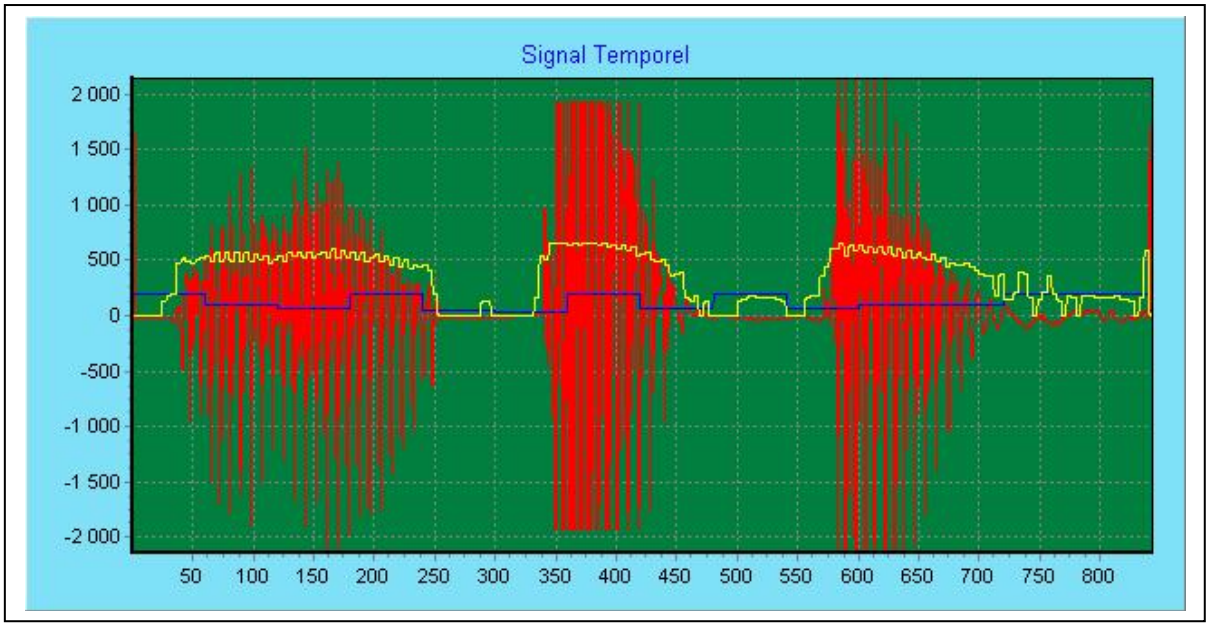

Fig. 4.22 : représentation du signal temporel, Energie et mélodie du mot [ɛaqatu] après synthèse.

7. CONCLUSION

Les mesures obtenues nous ont permet d'extraire les durées intrinsèques des sons spécifiques à l'Arabe Standard et par la suite de proposer un modèle de prédiction de ces dernières.

Nous avons essayé ce modèle par une synthèse par concaténation manuelle en premier lieu et une réécoute par 6 personnes qui ont donné des appréciations bonnes.

Par la suite nous avons essayé d'élaborer un outil pour synthétiser les sons sur lesquels on applique les règles de notre modèle. Les résultats était également satisfaisantes.

1. INTRODUCTION

Lorsque nous parlons, nous ne sommes en général pas conscients des mouvements complexes des muscles de la phonation, et il en va de même en particulier pour le contrôle de *la hauteur* et de *l'intensité* de la voix, lors des vibrations des cordes vocales. Ces deux paramètres auxquels nous joignons habituellement *les durées* successives des segments syllabiques constituent en leur évolution **la prosodie** de la phrase [1,2].

L'acquisition, la représentation et le traitement automatique des informations prosodiques dans les systèmes de reconnaissance ou de synthèse de la parole restent des questions ouvertes aussi bien sur le plan fondamental que dans le domaine des applications. Les difficultés rencontrées tiennent compte de l'extrême variabilité contextuelle de ces données ainsi que leurs connaissances, la complexité des relations qu'elles entretiennent avec tous les niveaux de la structuration linguistique et extra linguistique des messages vocaux, ainsi qu'aux problèmes liés à l'évaluation et à la pondération des paramètres qui les caractérisent.

2. QU'EST CE QUE LA PROSODIE ?

Nous pouvons trouver de nombreuses définitions de la prosodie, qui vient du mot *prose* (propre à la poésie)[19] : *forme ordinaire du discours parlé ou écrit, qui n'est pas assujettie aux règles de rythme et de musicalité*. D'un point de vue acoustique, par exemple, nous trouvons une définition comme étant : *l'étude de la durée, de la hauteur et de l'intensité des sons*, ou bien pour des aspects plus linguistiques : *partie de la phonologie qui échappe à l'analyse en phonèmes et traits distinctifs, tels que le ton, l'intonation, l'accent et la durée*. Nous trouvons également des références plus intuitives : *règles concernant l'application de la musique à des paroles ou inversement* [20, 21].

Dans le Traitement Automatique de la Parole, les paramètres prosodiques prennent une importance particulière. En synthèse, ils contribuent à conférer une meilleure intelligibilité au signal synthétique en signalant les grandes articulations de la phrase ; en reconnaissance, ils peuvent servir d'indices pour l'identification

d'éléments segmentaux déterminés et également signaler le type syntaxique de la phrase [1,2,3,4].

Les paramètres prosodiques sont :

- la mélodie (fréquence fondamentale) ou F_0 correspond à la fréquence des vibrations des cordes vocales lors de la production des phonèmes ;
- l'intensité traduit l'importance énergétique d'un phonème. Pour les sons spécifiques à l'Arabe Standard, elle permet de déterminer l'emphase (son plus énergétique), utilisée aussi pour la détection des sons voisés (périodiques) ;
- la durée ou rythme, contient un temps de phonation et un temps de silence. Elle permet la perception de l'accent, et dépend du débit d'élocution (lent, normal, rapide, etc...).

La durée phonémique, dépend des paramètres linguistiques, extralinguistiques, physio-logiques, phonologiques, interlocuteurs, extralocuteurs,...

Jusqu'à maintenant, parmi les études menées, aucune n'a pu déterminer précisément et sûrement, les relations prosodiques qu'entretiennent ces paramètres dans les stratégies d'un locuteur.

Dans le traitement prosodique du signal parole on peut s'interessé à l'études macroprosodique (macroprosodie) ou bien microprosodique (microprosodie).

2.1. LA MACROPROSODIE

Les variations temporelles des paramètres prosodiques (rythme, intonation, intensité) au cours de la production d'un énoncé contribuent, dans une large mesure, à l'identification de la structure syntaxique et discursive de cet énoncé par l'auditeur.
Si on s'intéresse à la connaissance des paramètres prosodiques d'une manière globale et de voir seulement l'allure de leurs évolutions dans la phrase parlée on dit qu'on a à traiter **la macroprosodie**. L'objectif de son traitement est la limitation du nombre des mots du lexique au moyen des seules informations macroprosodiques identifiées dans le signal de parole [22].

2.2. LA MICROPROSODIE

Si on s'intéresse à la connaissance des paramètres prosodiques d'une manière locale et de voir l'évolution des paramètres prosodiques dans le son, on dit qu'on a à traiter **la microprosodie**. Ceci correspond spécialement à la connaissance des variations prosodiques à l'intérieur du mot, phonèmes ou entre phonèmes (transitions phonémiques) (fig.3.1) [22] .

La microprosodie est le résultat de perturbations liées à la prononciation de certains phonèmes, elle est incontrôlée [10].

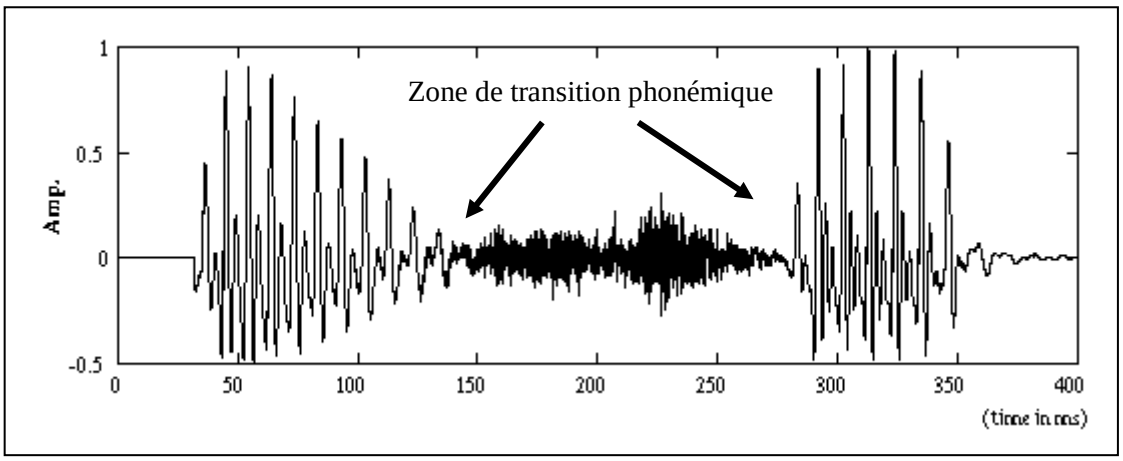

Fig.3.1 : Transition phonémique.

3. PARAMETRES ACOUSTIQUES DE LA PROSODIE

Les paramètres acoustiques qui portent les informations prosodiques (F0, intensité, durée des unités phonétiques et des pauses) doivent être ajustés contextuellement et évalués par rapport aux capacités de la perception. Par ailleurs, même si les commandes peuvent être activées indépendamment (pression sous-glottique, tension des cordes vocales, durée), certains liens fonctionnels entraînent des modifications corrélées des paramètres. L'évolution de la fréquence laryngienne porte les traces de l'ensemble des variations des commandes [21,23] ; cette particularité justifie la prééminence des travaux sur la mélodie. Les phénomènes prosodiques sont complexes et se manifestent conjointement sur plusieurs paramètres [21,23]. Cette caractéristique doit être prise en compte pour la mesure de chacun d'eux.

3.1. La fréquence fondamentale

Les variations de la fréquence laryngienne résultent de commandes conscientes ou non qui ont des effets à long terme (au delà du segment) et de perturbations incontrôlées, liées au mode de production de certaines consonnes, qui affectent les unités phonétiques (microprosodie). Elles traduisent donc des phénomènes divers qu'il convient de prendre en compte dès la phase de mesure de la fréquence fondamentale (contour global / variations locales). Les algorithmes d'extraction automatique de la F0 seront adaptés à la nature des informations que l'on doit traiter (phénomènes macroprosodiques et/ou phénomènes micro-prosodiques).

Indépendamment du contenu des énoncés, l'amplitude et la vitesse des variations de la courbe mélodique relèvent d'un grand nombre de facteurs (locuteur, type de parole, contexte phonétique, environnement...). La mesure de la fréquence fondamentale et l'interprétation de ses modifications temporelles doivent être rapportées aux contraintes particulières du contexte avant d'effectuer d'autres ajustements plus complexes qui prennent en compte des données et des connaissances diverses (notamment des corrections en fonction de la perception et de la nature des sons concernés) [24].

Du point de vue perceptif, nous parlons de hauteur. Il a été démontré qu'une échelle logarithmique traduit bien mieux qu'une échelle linéaire la perception des hauteurs que nous pouvons avoir. Pour calculer les variations entre deux valeurs de la fréquence fondamentale, nous avons la formule suivante [25] :

$$H = 6.ech. \frac{Ln(F_2/F_1)}{Ln2} \qquad (3.1)$$

H est la valeur de la hauteur pour une fréquence F2 par rapport à une fréquence F1, dans une échelle logarithmique. Cette échelle est fonction de *ech*. Si *ech* vaut 1, H sera donnée en tons, si elle vaut 2 en demi-ton, 4 en quart de ton, 8 en huitième de ton…

Le calcul de la fréquence fondamentale dépend du contexte sonore. Un milieu bruité détériore énormément les résultats des algorithmes de détection, il est donc important de soigner particulièrement les enregistrements destinés aux observations de ce paramètre.

La fréquence fondamentale n'est calculée que sur des parties voisées de la parole, c'est à dire principalement les voyelles, mais aussi quelques consonnes : ces localisations correspondent à des parties plus stables du signal, il devient possible d'en extraire des pseudo-périodes.

3.2. La Durée

Si l'on fait momentanément abstraction des difficiles problèmes liés à l'évaluation précise des durées des unités phonétiques (phonèmes, syllabes, distance de voyelle à voyelle, etc.), les valeurs objectivement mesurées doivent être corrigées en fonction du locuteur, de la vitesse locale d'élocution, du phonème et de son contexte ainsi que vis-à-vis des connaissances sur la perception de ce paramètre. Ces ajustements visent une certaine normalisation des observations de manière à mettre en évidence les unités qui se distinguent significativement de leurs voisines. Les procédures proposées pour effectuer cette normalisation sont souvent complexes et font intervenir les valeurs moyennes et les écarts types des durées . Les résultats obtenus sont parfois discutables, du point de vue de la reconnaissance automatique, à cause des limites des corpus utilisés (type de parole, nombre de locuteurs, nombre et répartition des unités, etc.), de la variété des méthodes d'étiquetage et des procédures de délimitation des segments et de mesure de leur longueur [21,24].

3.3. Les pauses

Les pauses, silencieuses ou remplies, soulignent les difficultés qu'un locuteur rencontre dans la production d'un énoncé: respiration, hésitations, temps d'accès aux informations lexicales, structuration de la phrase, mise en relief des éléments significatifs, expression de pathologie et/ou d'émotions, etc. Elles signalent des phénomènes linguistiques ou extra- linguistiques, facilitent la perception des unités et favorisent l'interprétation de toutes les informations véhiculées par le message. Les pauses sont plus fréquentes en parole spontanée qu'en lecture de textes; leur nombre et leur durée varient considérablement en fonction du contexte (locuteur, type de parole, contenu du discours, degré de productibilité des mots, etc.).

Il est nécessaire, de plus, de traiter ces phénomènes en relation étroite avec les autres paramètres de la prosodie. La participation des pauses dans la structuration des constituants de l'énoncé est reconnue par de nombreux chercheurs.

3.4. L'intensité

L'intensité est souvent négligée aussi bien dans les travaux fondamentaux que dans les systèmes de traitement automatique de la prosodie. Parmi les raisons du désintérêt pour ce paramètre on peut noter: les corrélations de ses variations avec celles de la F0, la difficulté de son évaluation, sa variabilité en fonction des sons et du contexte, les problèmes liés à sa normalisation perceptuelle, etc [26].

L'étude de l'intensité des voyelles a montré que chacun des phonèmes peut être affecté d'une valeur spécifique, mais la perception de cette intensité est liée aux configurations des variations de la fréquence fondamentale [27].

4. MODELISATION DE LA DUREE PHONEMIQUE

La durées phonémique est d'une grande variabilité. Elle dépend de plusieurs facteurs : ceux d'ordre linguistiques et ceux d'ordre extra linguistiques ; également ceux dus au locuteur. Nous devons dans le cadre du traitement de la parole connaître toutes ces variabilité afin de pouvoir modéliser ces durées si en veut faire une synthèse de la parole de haute qualité.

Dans la littérature nous pouvons citer les travaux de D. Klatt [28], C. Sorin et K. Bartkhova [29] ;
et concernant les travaux sur les sons de l'arabe : A. Alioui [30], A. Amrouche [3,31].

Du point de vue de la production, la phonation est limitée par des contraintes physiologiques, notamment en ce qui concerne les consonnes. En effet, l'appareil vocal n'est pas capable de produire des plosives qui auraient une durée importante. L'effort produit lors de l'articulation ne peut être infini, la séquence est toujours la même : effort et relâchement. Cet effort et ses variations sont véhiculés à travers le signal, ils permettent une segmentation temporelle de le parole en unités (durées).

Avant de mesurer des durées, il faut cerner correctement les unités à mesurer. On distingue les durées des unités phonétiques, syllabes, phonèmes et les durées des pauses.

Comme les autres paramètres, les durées des unités choisies sont largement dépendantes du locuteur et du débit de parole. Toute mesure ne peut donner un modèle absolu pour une application brute des relevés. La considération des résultats des observations devra plutôt s'orienter vers un modèle relatif qui pourra s'exprimer en terme d'allongements ou de réductions (raccourcissements).

Chaque phonème a une durée intrinsèque et co-intrinsèque. Ces durées sont des caractéristiques des phonèmes. On se rend compte aisément que le phonème [a], pris seul, est plus long que le phonème [b], par exemple.

La durée des différentes unités est le phénomène central pour la prosodie. En effet chaque variation de fréquence fondamentale ou d'intensité s'établit sur un certain laps de temps, durée mesurable. Etudier l'organisation temporelle de la parole est incontournable. Etudier la durée c'est observer et modéliser les durées d'unités bien déterminées. Pour pouvoir comparer ou tirer des enseignements d'études passées, il importe de bien cerner à chaque fois l'unité choisie.

Nous donnons dans ce qui suit un aperçu des principaux modèles de prédiction des durées :

- Le phonème

  Les travaux de Klatt [28] sont à la base de beaucoup de modèles actuels. Ils s'appuient sur la connaissance des durées intrinsèques (D_{Intr}) de chaque phonème ainsi que sur leur durée minimum (D_{min}). Calculer la durée d'un phonème consiste dans ce modèle à ajouter à la durée minimum, une durée qui dépend du contexte dans lequel se trouve ce phonème. La relation suivie est :

$$D_T = D_{\min} + \frac{((D_{\text{int}\,r} - D_{\min}) * k\%)}{100} \tag{3.2}$$

  avec k% : exprime un rétrécissement qui dépend du contexte phonétique et syntaxique.

  Les durées ne peuvent pas descendre en dessous d'un seuil D_{min}, propre à chaque phonème. Cependant, Klatt distingue l'état accentué ou non d'un phonème ; si on prend pour référence D_{min} d'un phonème p non accentué, la valeur minimale pour ce même phonème accentué est $2*D_{min}$.

Bartkova et Sorin [29] proposent une modélisation pour la synthèse par diphones. Elles s'appuient, tout comme Klatt, sur les durées intrinsèques des phonèmes. Des coefficients, dont les valeurs sont dépendantes des contextes phonémiques, syntaxiques et prosodiques, viennent fixer les durées intrinsèques. Dans leur modèle, le calcul des pauses se fait avant le calcul des durées phonémiques.

- La syllabe

Grosjean et Monnin [25] émettent l'hypothèse de l'existence d'une structure de performance. Cette structure de performance ne correspond pas à la structure de surface (syntaxique), mais plutôt à une structure prosodique effective. A la suite de Duez, ils mesurent les durées syllabiques (en fait la durée des voyelles) en leur incluant la pause qui suit éventuellement. Le principe d'équilibre syllabique est largement exploité dans leur étude.

Le modèle de Campbell [25] repose sur l'hypothèse que l'organisation temporelle d'un énoncé se fait à un niveau supérieur au niveau phonémique. Deux étapes se distinguent, dans la mise en œuvre de ce modèle, la première est la prédiction des durées syllabiques et la seconde est la prédiction à l'intérieur de chaque syllabe des durées phonémiques. Un processus d'apprentissage automatique permet la prédiction des durées syllabiques. En ce qui concerne les durées segmentales, leur distribution est donnée par le calcul d'un coefficient d'allongement (déviation par rapport à la moyenne). Il propose que tous les phonèmes d'une même syllabe aient le même facteur d'allongement z : le *z-score*. Le z-score de chaque réalisation phonémique des corpus d'étude est calculé :

$$Z = \frac{(D_{ob} - D_{moy})}{\sigma_p} \tag{3.3}$$

avec :

Z : z-score ;

D_{ob} : durée observée du phonème ;

Dmoy : durée moyenne du phonème ;

σ_p : ecart type du phonème.

Et la durée syllabique est donnée par la fonction suivante :

$$D_s = \sum_{i=1}^{Np} \exp(D_{moyi} + Z.\sigma_p) \qquad (3.4)$$

avec :

D_s : durée de la syllabe ;

D_{moyi} : durée moyenne du $i^{ème}$ phonème.

En fait, la durée d'une syllabe est logiquement la somme des durées de chaque phonème qui compose cette syllabe. Le modèle sera repris pour prendre en compte les positions des syllabes (notamment pour les frontières prosodiques).

Pour les études faites sur l'arabe Alioua, Ahmed dans ça thèse de doctorat (1995) [28], a étudié l'effet des consonnes d'arrière et des emphatiques sur la nature acoustique des voyelles longues de l'arabe littéral marocain. Les phonèmes n'étant pas des sons isolés, mais plutôt des segments produits dans la chaîne parlée, ils entretiennent entre eux des relations de contiguïté qui les lient dans des unités de rang supérieur. De ce fait, il est évident que les sons favorisés par leurs distributions, ou possédant des propriétés d'agir intrinsèques, imprègnent des sons voisins ou modifient leurs qualités. Tel est, en arabe, le cas des voyelles au voisinage des consonnes d'arrière et des emphatiques. Ces consonnes sont des segments typiques de l'arabe, tout comme du sémitique. Leur influence sur les voyelles voisines est régulièrement mentionnée. Aussi, Alioui c'investi dans l'analyse de leur effet sur les voyelles longues [a:], [i:], [u:] qui n'ont fait l'objet d'aucune étude spécifique selon nos recherches bibliographiques. Il a examiné cet impact sur la fréquence formantique, et aussi sur la durée, l'intensité et la fréquence fondamentale. Si nous faisons l'hypothèse que les consonnes d'arrière d'une part et les emphatiques d'autre part exercent une influence appréciable sur les voyelles adjacentes, on suppose que les consonnes non arrière et non emphatiques ne pratiquent pas ce même effet sur les mêmes voyelles. Et il a comparé les voyelles longues dans des contextes parfaitement analogues dans l'entourage des consonnes d'arrière par rapport aux non-arrières et des emphatiques comparées aux non-emphatiques.

Dans notre étude nous avion à étudier les durées phonémiques des consonnes spécifiques à l'Arabe Standard. Nous avons donc eu à extraire les paramètres prosodiques de ces consonnes ce qui nous a permet pour le cas des durées phonémique d'utiliser un model de leurs prédiction. Par la suite nous nous sommes inspiré du modèle de KLATT pour trouver des règles de prédiction des sons étudiées. Une vérification à la fin a été faite pour justifier le modèle adopté.

5. METHODOLOGIE

Afin d'extraire les paramètres prosodiques nous avons élaboré un corpus de mots contenant les diphones à étudier et suivant différent contextes.

Puis nous avons choisi quatre locuteurs professionnels : Deux masculins et deux féminins, pour enregistrer le corpus. L'enregistrement a été fait dans le Laboratoire d'Acoustique à l'Institut d'Electronique (l'USTHB d'Alger) à une vitesse d'élocution normale entre 4 et 7 syllabes par seconde en utilisant une carte « OROS-AU21 » spécifique au traitement de signal vocal, avec une fréquence d'échantillonnage de 10 kHz. Les échantillons ont été codés sur 16 bits par échantillon, dans des fichiers format de données.

Le corpus choisi est pris suivant le dictionnaire de diphones en Arabe Standard élaboré par M.GUERTI [10].

Les notations que nous avons prises pour représenter :

le silence : [#] ;

la consonne : [c] ;

une voyelle brève : [v] ;

une voyelle longue : [v:] ;

une consonne emphatique : [c_e] ;

une consonne non-emphatique : [c_{ne}] ;

le début du phonème : [-] suivi d'un phonème.

Les phonèmes sont représentés suivant le code du tableau 1.

Transcription API	[ʔ]	[q]	[ð]	[□]	[ð]	[ɛ]	[ħ]	[□]
Equivalent arabe	أ	ق	ض	ط	ظ	ع	ح	ص

Tableau 3.1 : représentation des consonnes de l'Arabe Standard suivant le code API. (Alphabets Phonétique internationale).

Les locuteurs qui ont fait l'enregistrement du corpus ont à peu près le même âge et ils n'ont pas d'influence dialectale dans leurs prononciations de l'Arabe Standard.

5.1. OUTILS

Les mesures des durées se font après une segmentation manuelle des phonèmes en utilisant le logiciel Parole1.0 [11], (fig. 3.2 de 3.2.a jusqu'à 3.2.d)

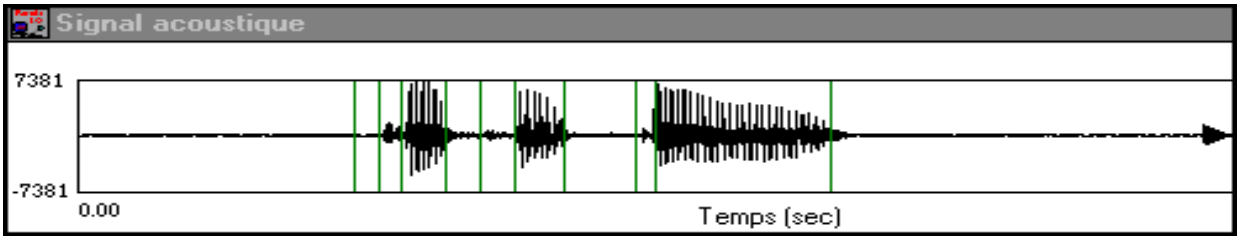

Fig. 3.2. a : forme temporelle.

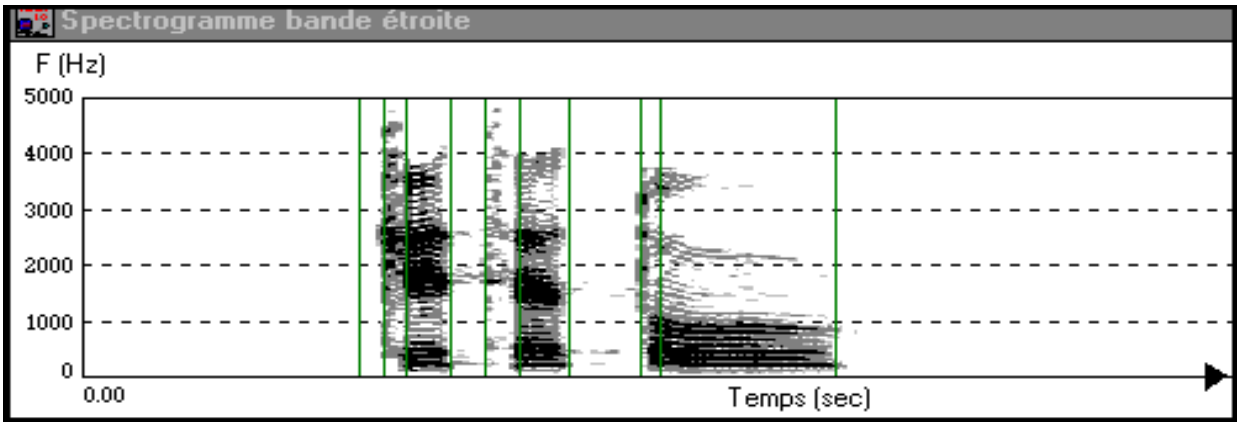

Fig.3.2. b : sonagramme du mot.

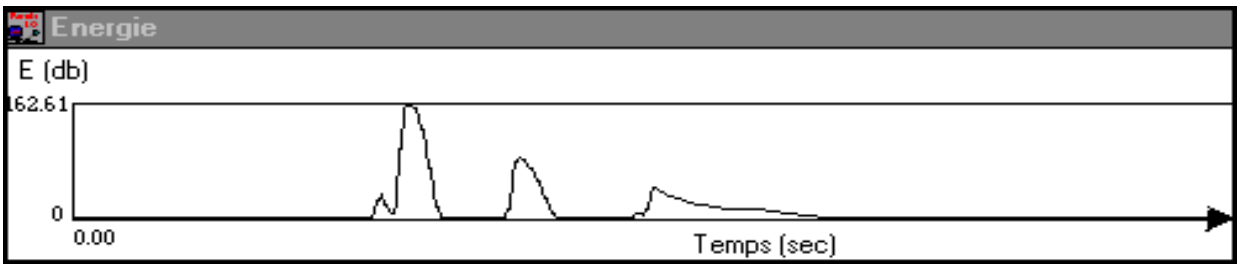

Fig.3.2.c : Energie.

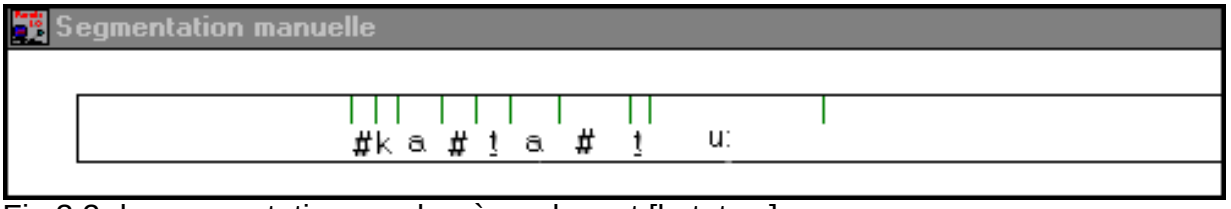

Fig.3.2.d : segmentation en phonème du mot [katat$_e$u:].

Fig.3.2 Exemple d'analyse sonagraphique du mot [katat$_e$u:] et sa segmentation en phonèmes.

Ce logiciel nous a permit d'enregistrer notre corpus à l'aide de la carte OROS-AU21 et d'analyser le signal par le sonagramme, l'énergie, le taux de passage par zéro, et de faire par la suite une segmentation manuelle.

Les figures précédentes obtenues par le logiciel Parole1.0 présentent successivement :

- le signal temporel numérique ;
- le sonagramme numérique appliqué sur le signal temporel du mot enregistré. Pour voir les fréquences formantiques et déterminer les zones de parole à segmenter ;
- l'énergie utilisée en même temps que le sonagramme pour déterminer les sons voisés et non voisés. Dans le cas de la segmentation, l'énergie permet de limiter le début et la fin d'un segment. Avec toutefois une vérification par réécoute pour s'assurer que la segmentation a été bien faite [2,3,4].

Nous avons utilisé également les logiciels Goldwave [32], (Fig.3.3), et Winsnorri [33], (Fig.3.4), pour une autre vérification des résultats de la segmentation puisque ces logiciels permettent d'améliorer la vision des zones à segmenter et de les réécouter.

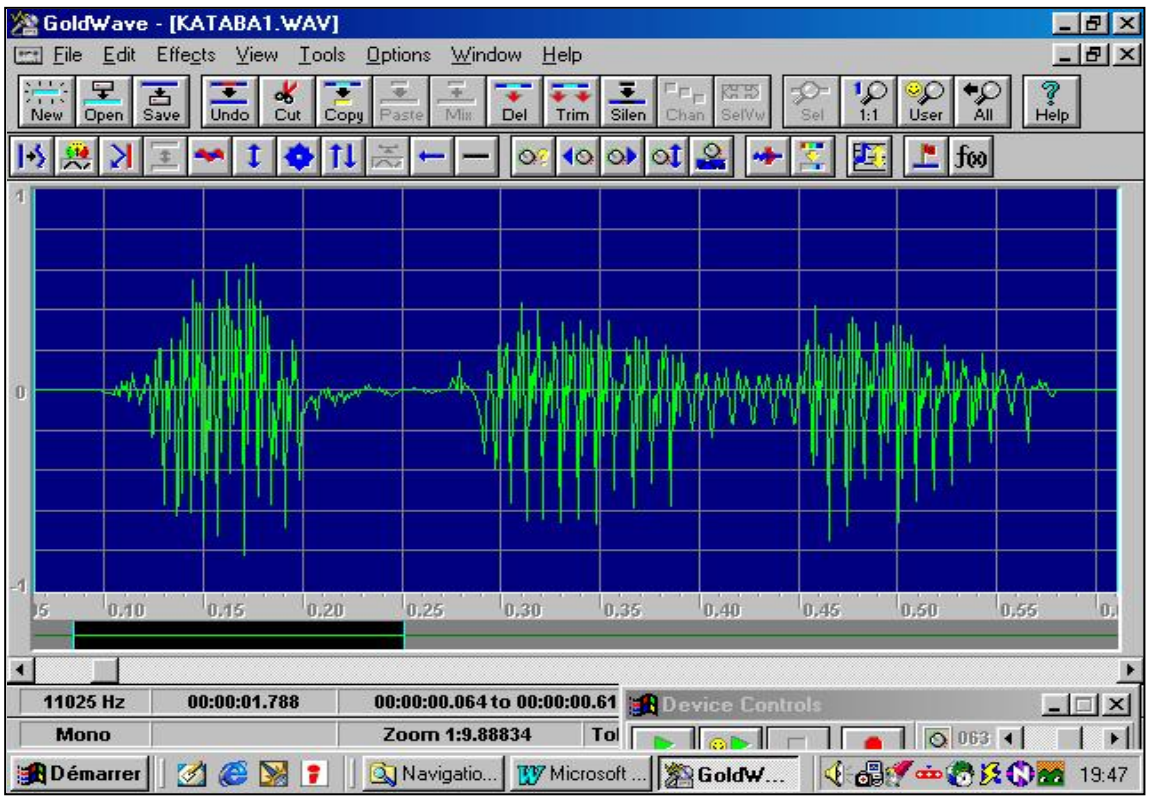

Fig.3.3 signal temporel du mot [kataba] à l'aide de Goldwave.

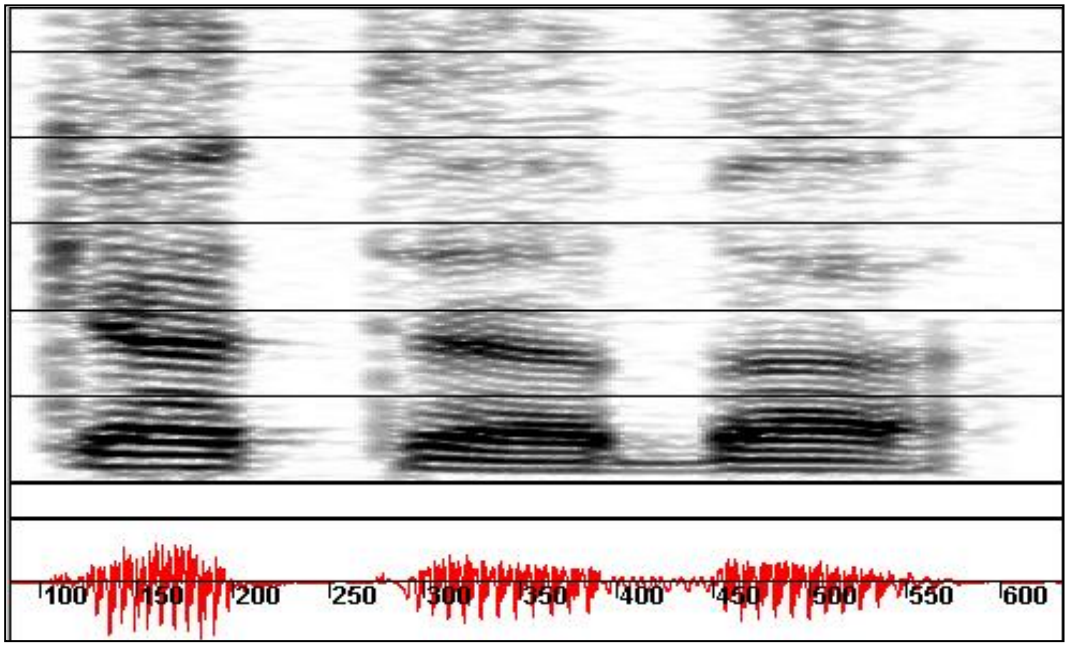

Fig.3.4 représentation temporel (en bas) et sonagraphique (en haut) du mot [kataba] à l'aide de Winsnorri.

5.2. CORPUS D'ETUDES

Nous avons élaboré le corpus d'étude d'une manière à pouvoir extraire :

- les durées phonémiques des voyelles ;
- les durées phonémiques des consonnes ;
- les variations des durées suivant le contexte.

Par conséquent nous avons choisis les sons suivants pour construire le corpus :

- les voyelles :
  - brèves [a], [u], [i] ; fatha, damma et kassra.
  - longues [a:], [u:], [i:]. Hurouf el mad.

- les consonnes :

- occlusifs non voisées : [□], [ʔ], [q] ;
- occlusifs voisées : [δ] ;
- fricatifs emphatiques : [□], [δ] ;
- fricatifs pharyngales : [ε], [ħ].

Pour former le corpus nous avons regroupé ces phonèmes en diphones suivant le dictionnaire de diphone élaboré par M. GUERTI [10] puis nous les avons placés dans des logatomes (ou mots artificiels) afin de prendre toutes les combinaisons possibles.

Nous avons par la suite enregistré et analysé se corpus à l'aides des outils numériques de traitement de signal cités ci dessus.

6. CONCLUSION

Les difficultés de l'extraction des paramètres prosodiques doivent stimuler les travaux des chercheurs car l'ouverture nécessaire des systèmes de reconnaissance à la parole spontanée ou de synthèse TTS donnera une importance inévitablement plus grande aux phénomènes prosodiques.

Par ailleurs, la prosodie implique bien d'autres phénomènes que ceux liés à la nature des sons et à la structuration linguistique des énoncés. Il est donc également indispensable d'étendre nos recherches en vue d'extraire d'autres informations telles que: l'identité du locuteur, les émotions, la qualité esthétique de la production vocale, certaines pathologie et les variantes dialectales.

Les durées phonémiques des sons de l'Arabe Standard sont peu étudiées, ce qui nous a pousser à choisir ce paramètre pour en extraire ses propriétés et variabilité afin de l'exploiter dans la reconnaissance ou la synthèse de la parole.

Pour cela nous avons choisis des outils de traitement qui nous facilitent l'étude (exemple : Parole.exe, Goldwave.exe, Winsnorri.exe).

Nous avons également élaboré un corpus de mots contenant les diphones à étudier.

1. INTRODUCTION

Le développement de l'électronique, de l'enregistrement et des télécommunications font que la parole est devenue un moyen privilégié de communication entre les Hommes.

L'apparition de l'informatique a fourni à la fois un outil d'étude, et un outil de communication parlée entre l'Homme et la Machine (ou encore le dialogue Homme-Machine). La machine doit donc traiter, reconnaître ou synthétiser l'information vocale [1].

Pour cela une étude préalable du signal parole doit être faite, afin de donner à la machine toutes les particularités : linguistiques, acoustiques et prosodiques,…

Nous présentons donc dans ce chapitre les principales caractéristiques des sons en général et des sons de l'Arabe Standard en particulier [3].

2. ETUDE DE LA PHONATION

Avant de traiter le signal vocal qui est l'information d'un message parlé nous devons le produire et ceci à partir des fluctuations de la pression de l'air émises par l'appareil phonatoire [1,3,4]. Ce dernier est le système principal de production des sons et de la parole naturelle (fig.1.1, fig.1.2).

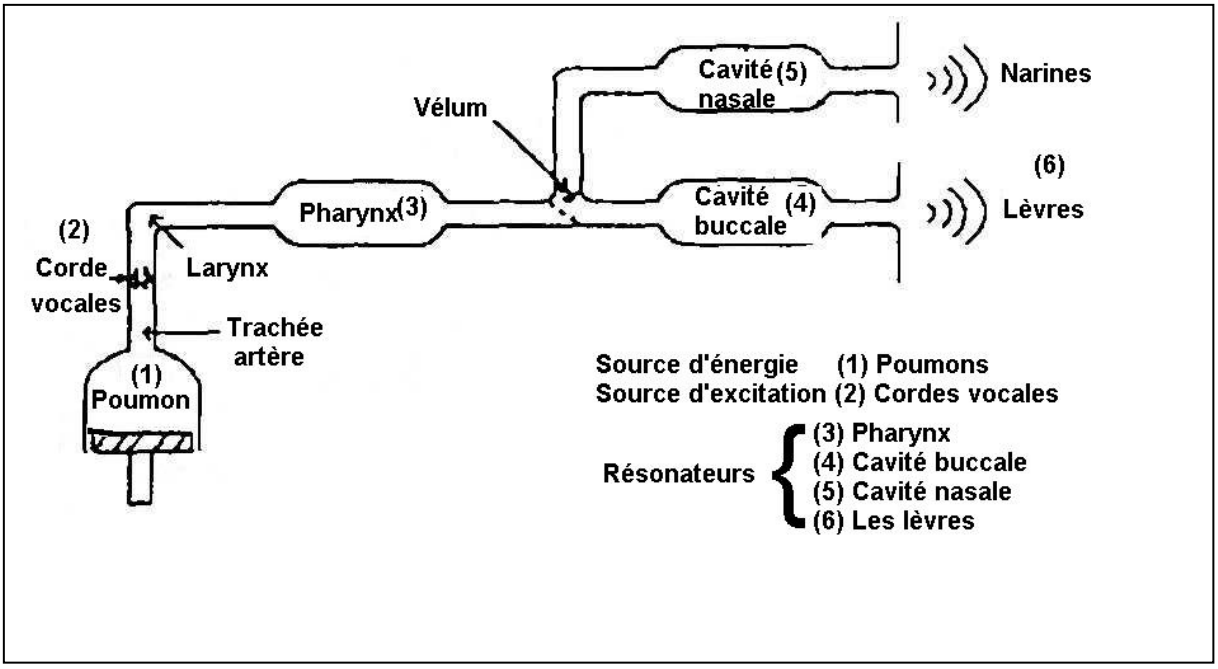

Fig.1.1: Représentation schématique du système phonatoire

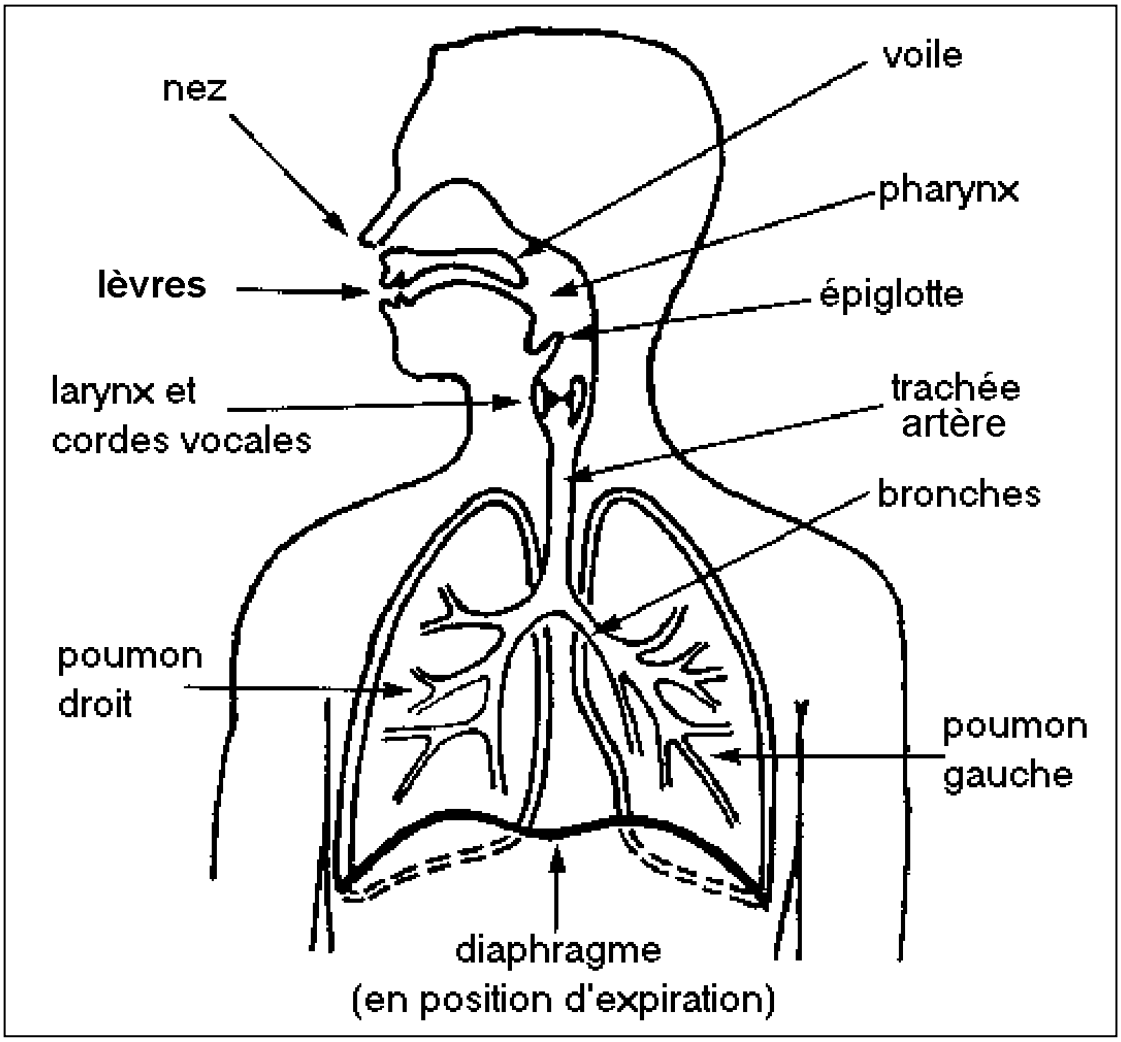

Fig.1.2: Représentation du système phonatoire

2.1. L'appareil phonatoire

L'appareil phonatoire est constitué par :

- Les poumons qui sont la source d'énergie permettant de produire les sons. En utilisant l'air dans les poumons et en changeant la pression de cet air on peut changer le débit de production des sons ;
- la trachée artère qui est un conduit assurant l'écoulement de l'air des poumons vers le larynx ;
- le larynx, localisé juste après la trachée artère et avant le conduit vocal. On trouve des replis membraneux (qu'on les appelle les cordes vocales). Ces replis

permettent de transformer l'air des poumons en un air phonatoire. Ce dernier circule librement dans le cas de la respiration et des voix chuchotées ;

- le conduit vocal, s'étend du larynx jusqu'à la cavité labiale, dans lequel l'air phonatoire s'écoule et trouve deux conduits buccal et nasal ;
- le conduit buccal est un conduit de forme variée. Il contient la langue, les mâchoires , le palais,… Il est suivi d'une cavité labiale (ou cavité des lèvres) ;
- le conduit nasal est un conduit de forme fixe. Entre ce conduit et le reste du conduit vocal, on trouve le vélum qui est un élément de couplage, permettant d'aiguiller l'air phonatoire dans le conduit buccal ou dans le conduit nasal.

3. CLASSIFICATION DES SONS

Une fois s'écoule l'air phonatoire dans les cavités nasale ou buccale, il va être utiliser comme un son phonatoire. Et en changeant le type d'articulation, on obtient un ensemble de sons afin de construire un ou plusieurs mots (phrase) [2,4]. Ces sons sont classés par différents critères :

3.1. Critères perceptifs

Le son est défini comme étant une variation de pression dans l'air, l'eau ou autres milieux élastiques quelconque pouvant être perçue par l'oreille humaine.

Dans le domaine de la phonétique, un son est une vibration provoquée par l'appareil vocal dans un milieu dans lequel elle peut se propager pour exciter notre ouïe.

Suivant la perception du son on peut avoir des sons : occlusif, fricatifs, sonores (voisées) ou sourds (non voisées)… Les sons voisés résultent des vibrations périodiques des cordes vocales. Des impulsions périodiques de pressions sont ainsi appliquées au conduit vocal.

3.2. Critères acoustiques

D'un point de vue acoustique, le son est une onde due à une vibration. Cette dernière est obtenue lorsque l'air contenu dans les poumons est contraint à passer dans le larynx (qui comprend les cordes vocales). Les cordes vocales se tendent et s'étirent au besoin. Lorsque la pression d'air s'accumule sous les cordes vocales, elles sont forcées de s'ouvrir partiellement. Leur tension naturelle les amène ensuite à se refermer. La vitesse à laquelle les cordes vocales s'ouvrent et se referment produit une vibration d'une hauteur variable (appelée la fréquence fondamentale). Selon la taille de l'appareil phonatoire de la personne.

La nature de l'excitation glottique, la forme du spectre du signal, la position des formants, l'énergie etc, sont des paramètres acoustiques permettant le classement des différents sons de la parole.

3.3. Critères articulatoires

Les sons du langage humain peuvent être étudiés sur le plan de la phonétique articulatoire (la production). Cette discipline comporte une partie sur la physiologie, consacrée à la connaissance des organes de la phonation et une autre partie descriptive portant sur le rôle des différents organes dans la production des sons du langage.

Il est possible d'opérer une classification des sons du Français [5] (par exemple) à partir de critères articulatoires. Ces critères permettent également de décrire les sons d'autres langues et résument, en quelque sorte, les possibilités et les limites de l'appareil phonatoire. Bien entendu, les modalités d'exploitation des organes articulatoires peuvent varier selon la langue ou la famille de langues considérée. Les modalités de réalisation des sons peuvent même varier, pour une même langue, à cause du problème de la variabilité.

Les critères permettant de classifier les sons du Français sont les suivants :

- Le mode articulatoire, Il est du à la qualité du passage de l'air dans le canal buccal. La réalisation des voyelles implique un passage libre de l'air le long du canal buccal. Le degré d'ouverture de la cavité buccale permet de distinguer quatre types de voyelles : les voyelles ouvertes, les mi-ouvertes, les mi-fermées et les fermées. Pour les consonnes, deux modes articulatoires sont à distinguer.

Le passage de l'air est momentanément bloqué ou obstrué lors de la production des consonnes occlusives. Il est suffisamment rétréci pour permettre l'émission d'un bruit continu lors de la réalisation des consonnes constrictives (ou fricatives) ;

- La résonance orale ou nasale, elle est en fonction de l'ouverture ou de la fermeture de l'accès vers les fosses nasales. Lors de la production des voyelles ou des consonnes nasales, le voile du palais est abaissé et permet le passage de l'air à la fois par le canal buccal et par les fosses nasales, ce qui confère aux sons une coloration particulière. Les voyelles et les consonnes produites sont alors dites " nasales ". Lorsque le voile du palais est relevé et bien accolé à la paroi pharyngale, l'air ne passe que par la cavité buccale, donnant naissance aux sons vocaliques et consonantiques dits oraux ou non nasalisés ;
- Les cordes vocales déterminent le caractère sourd ou sonore des différentes articulations. Lorsque les cordes vocales vibrent, les sons sont dits " voisés " ou " sonores" par opposition aux sons " non voisés " ou " sourds ". La réalisation des voyelles implique la mise en vibrations des cordes vocales. Pour les consonnes, l'absence ou la présence de vibrations des cordes vocales détermine leur caractère sourd ou sonore. En Français, toutes les consonnes nasales sont voisées. La distinction entre consonnes sourdes et sonores n'est faite que pour les consonnes orales ;
- Le lieu d'articulation, se situe nécessairement dans la partie supérieure du canal buccal dans une zone allant de la lèvre supérieure jusqu'à la paroi pharyngale (fig.1.3). C'est le point où l'articulateur se rapproche ou avec lequel il entre en contact avec cette zone. Les points d'articulation sont la lèvre supérieure, les incisives supérieures, les alvéoles, le palais dur et la région vélaire. La région des alvéoles se subdivise en une zone alvéolaire et une zone post-alvéolaire alors que la voûte palatine comprend les régions pré-palatale et palatale.

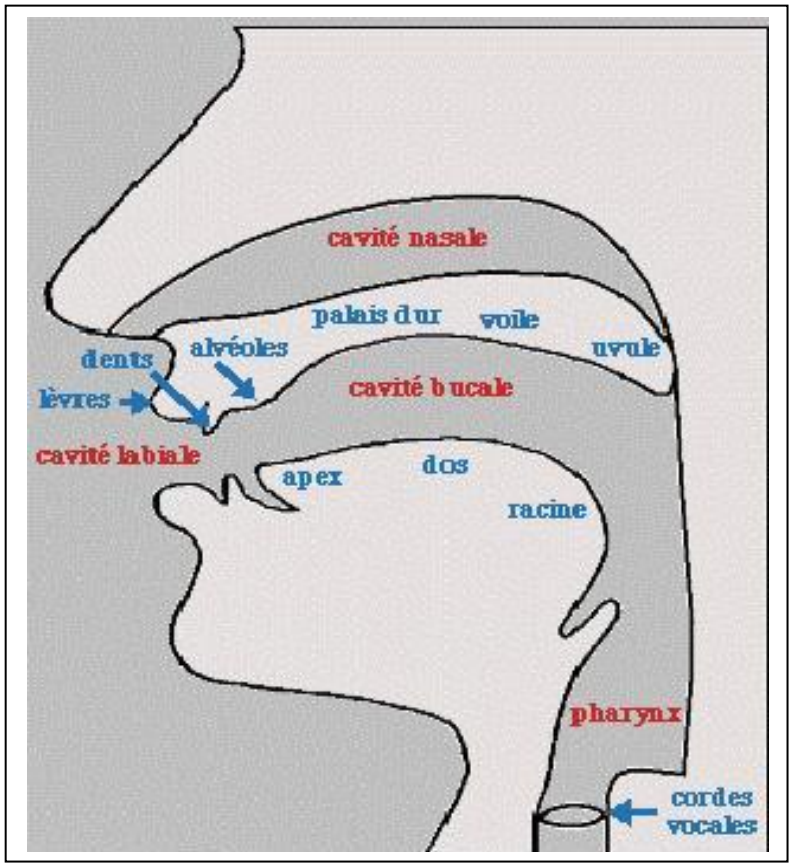

Fig.1.3: Lieux d'articulation phonémique.

- L'articulateur est constitué par la région inférieure du canal buccal. Il s'agit de la lèvre inférieure et des différentes parties de la langue. La réalisation de toute articulation implique un rapprochement plus ou moins grand ou un contact franc entre l'articulateur et le lieu d'articulation. La lèvre inférieure constitue avec les dents l'articulateur des consonnes telles que [f], [v] elles sont dites labiodentales. Les autres sons du Français ont comme articulateur le muscle lingual. La langue se subdivise en apex, région prédorsale, région dorsale et racine ;
- Les lèvres déterminent le caractère labialisé ou non labialisé d'une articulation. En effet, toute articulation peut être accompagnée ou non d'une projection des lèvres. On distingue : les voyelles arrondies ([u] ou [ϕ]), ou non arrondies ([i] et [e]) et les consonnes labialisées ([p], [b] ou [m]) ou non labialisées ([s] et [z]).

4. CLASSIFICATION PHONETIQUE DES SONS

D'un point de vue linguistique on peut dire que la production des sons ou d'un mot réside sur la production en série de toutes les lettres constituant ce mot. Ces lettres forment les unités phonétiques qui sont classées suivant les formes : voyelles, consonnes et semi-voyelles [1,3,6,7].

4.1. Les voyelles

Les voyelles (appelées « El haraka » dans l'arabe), se caractérisent par des vibrations laryngiennes périodiques en donnant des fréquences de résonances du conduit vocal. Les maxima de ces fréquences sont appelés "Formants" on dit donc qu'on a un son voisé (sonore). Les voyelles se distinguent par :

- le lieu d'articulation (traits) : antérieur/postérieur, fermé/ouvert ;
- le mode d'articulation:
- trait de labialité : étirée, arrondie ;
- trait de nasalité, ou d'oralité.

les voyelles orales sont dues à une élévation du voile du palais, qui détermine la fermeture des fosses nasales ainsi que l'écoulement de l'air expiratoire à travers la cavité buccale.

Les voyelles nasales sont caractérisées par l'écoulement d'une partie de l'air pulmonaire à travers les fosses nasales. Ce type de voyelles n'est pas disponible dans la langue Arabe.

4.2. Les consonnes

Les consonnes (appelées « Hourouf » dans l'arabe) se caractérisent par une fermeture :

- partielle du conduit vocal (constriction): constrictives ou fricatives ;
- totale du conduit vocal (occlusion): occlusives ou plosives.

Les traits correspondant aux consonnes sont: nasalité/oralité, voisement/non voisement.

Les consonnes occlusives (plosives) on les reconnaît grâce au silence provenant de la fermeture totale du conduit vocale (occlusion). Une occlusion comporte trois phases :

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

Les consonnes fricatives (constrictives) appelées également spirantes, les fricatives sont caractérisées par un bruit de friction causé par le passage rapide de l'air par un rétrécissement au point d'articulation.

Les consonnes sonnantes ou voisées présentent le degré d'obstacle le plus faible (comme [b], [v], [*f*]), est caractérisée par la vibration périodiques des cordes vocales lors de l'écoulement de l'air phonatoire dans le conduit vocal.

Les consonnes liquides ([l], [R]) combinent une occlusion et une ouverture simultanée du conduit vocal. Elles sont caractérisées par un degré de sonorité proche de celui des voyelles et, de ce fait, leurs spectres présentent des caractéristiques vocaliques, avec une structure de formants assez nette.

Les consonnes nasales ([m], [n], [ŋ]) sont produites par l'écoulement de l'air phonatoire dans le conduit nasal.

4.3. Les semi-voyelles

Les semi-voyelles ou semi-consonnes, ce sont des consonnes possédant des caractéristiques proches des voyelles et sont les sons :

- [ч] et [j] : sonnantes palatales ;
- [w] : sonnante labio-vélaire (articulation contenant une occlusion palatale avec un arrondissement des lèvres).

Pour mieux illustrer ces notions nous donnons dans les tableau1.1, 1.2, 1.3, des exemples des phonèmes de la langue française [1].

CONSONNES

Mode d'articulation		Lieu d'articulation		
		Labiales	Dentales	Vélo-palatales
Occlusives	Non voisées	[p]	[t]	[k]
	voisées	[b]	[d]	[g]
	Nasales	[m]	[n]	[ɳ]
Fricatives	Non voisées	[f]	[s]	[ζ]
	voisées	[v]	[z]	[Ʒ]
Semi-voyelles		[w]	[Ч]	[j]
Liquides			[l]	[R]

Tableau 1.1. phonèmes de la longue Française (consonnes).

VOYELLES

Mode d'articulation		Antérieures		Postérieures
		Non arrondies	Arrondies	
Orales	Fermées	[I]	[y]	[u]
		[e]	[ϕ]	[o]
		[ɛ]	[ϕ]	[Ɔ]
	Ouvertes	[a]		
Nasales	Fermées	[ɛ]		[õ]
	Ouvertes		[ã]	

Tableau 1.2. phonèmes de la longue Française (voyelles).

Exemples :

Phonèmes	Mots clés	Phonèmes	Mots clés
[’]	a<u>gn</u>eau	[o]	r<u>ô</u>le
[z]	<u>z</u>one	[ɛ]	l<u>ai</u>t
[Ʒ]	<u>j</u>oue	[œ]	b<u>œu</u>f
[w]	r<u>oi</u>	[Ɔ]	n<u>o</u>te
[Ч]	p<u>ui</u>s	[ɛ]	mat<u>in</u>
[j]	ca<u>i</u>llou	[õ]	mout<u>on</u>
[ϕ]	bl<u>eu</u>	[ã]	mam<u>an</u>

Tableau 1.3. exemples des phonèmes de la longue Française.

5. PARTICULARITE DE LA LANGUE ARABE STANDARD (AS)

L'Arabe Standard est la langue commune à tous les arabophones et est la langue des médias, de la science, de l'enseignement, de la littérature ….

Elle est structurée d'une manière différente au autres langues : les consonnes (ou hourouf) et les voyelles (ou haraka).

L'originalité de la phonétique arabe se fonde, pour une part importante, sur les consonnes emphatiques, pharyngales et laryngales, car elles donnent une valeur particulière à la langue [3,6,7].

5.1. L'emphase

L'emphase est une particularité de la langue Arabe et les langue sémitiques, et qui se caractérise par la forme plus énergétique du fait que lors de la production du son emphatique la langue se plie et s'incurve pour former un creux dont lequel le son est pressé. Ces sons sont : [□], [δ], [□], [δ]. Tous les autres sons sont dites non-emphatiques.

5.2. La gémination

C'est la production d'une consonne avec une concentration d'énergie très intense. C'est un allongement temporel qui résulte d'un dédoublement de la consonne simple correspondante. Toutes les consonnes arabes peuvent être géminées sauf la hamza.

5.3. Les voyelles

Le système vocalique de l'A.S. se compose de trois voyelles brèves : fatha [a], dhama [u] et kassra [i] et leurs correspondantes longues horouf el mad [a:], [u:], [i:] qui sont réalisées par un allongement des voyelles brèves. Cette opposition joue un rôle important dans le rythme de la langue.

Nous présentons dans le tableau ci dessous quelques caractéristiques des phonèmes de l'Arabe standard avec leurs représentations phonémique en code API (Alphabet Phonétique International).

Transcription API	Equivalent arabe	Plosive	Fricative	Nasale	vibrante	Liquide	Semi-voyelle	Voisée	Emphatique	Pharyngal	Glottale
ʔ	أ	+									+
b	ب	+						+			
t	ت	+									
θ	ث		+								
ʒ	ج		+					+			
ħ	ح		+								
χ	خ		+								
d	د	+						+			
ð	ذ		+					+			
r	ر				+			+			
Z	ز		+					+			
s	س		+								
ʃ	ش		+								
□	ص		+						+		
δ	ض	+						+	+		
□	ط	+							+		
δ	ظ		+					+	+		
ε	ع		+					+			
γ	غ		+					+			
f	ف		+								
q	ق	+							+		
k	ك	+									
l	ل					+		+			
m	م	+		+				+			
n	ن	+		+				+			
h	ه		+					+			+
w	و						+	+			
y	ى						+	+			

Tableau 1.4 : Caractéristiques des phonèmes de la langue Arabe Standard.

6. CONCLUSION

Nous avons présenté dans ce chapitre certaines notions de traitement de la parole, à savoir : la production des sons et leurs caractéristiques acoustiques et linguistiques. Ces informations changent d'une langue à une autre et d'un son à un autre. La langue Arabe Standard, qui a des particularités acoustiques comme l'emphase et la gémination, a été étudiée dans ce chapitre.

1. INTRODUCTION

L'objectif de la synthèse de la parole est de produire des sons de parole à partir d'une représentation phonémique du message, on peut dire la synthèse à partir du texte ou **TTS** (**T**ext-**T**o-**S**peech). Cela nécessite une connaissance préalable des caractéristiques du signal du phonème. Ce qui nous oblige de faire l'analyse du signal avant de procéder à sa synthèse [1,8].

Le concept de synthèse TTS de haute qualité, a de nombreuses potentialités d'applications : services de télécommunications, apprentissage des langues avec l'aide de l'ordinateur, aide aux personnes handicapées, jouets ou livres parlants, multimédia, communication Homme-Machine, ainsi que de larges extensions des recherches fondamentales et appliquées sur la parole.

2. SYNTHESE TTS

La synthèse TTS est une opération de génération automatique d'un signal de parole à partir d'un message écrit [1, 2].

Pour cela les systèmes de synthèse à partir du texte ont l'architecture de la fig.2.1 :

- l'analyse linguistique: regroupe l'analyse morphologique, syntaxique, sémantique;
- le transfert phonologique : permet de générer l'arbre prosodique ;
- l'étape de génération : cette étape est caractérisée par la méthode de synthèse employée (synthèse par unités stockées ou bien synthèse par règles) ;
- l'étape de synthèse : passage de la représentation paramétrique au signal acoustique.

Dans notre travail nous nous sommes intéressés aux méthodes de synthèse (l'étape de génération) et spécialement aux paramètres prosodiques, pour une synthèse par règles

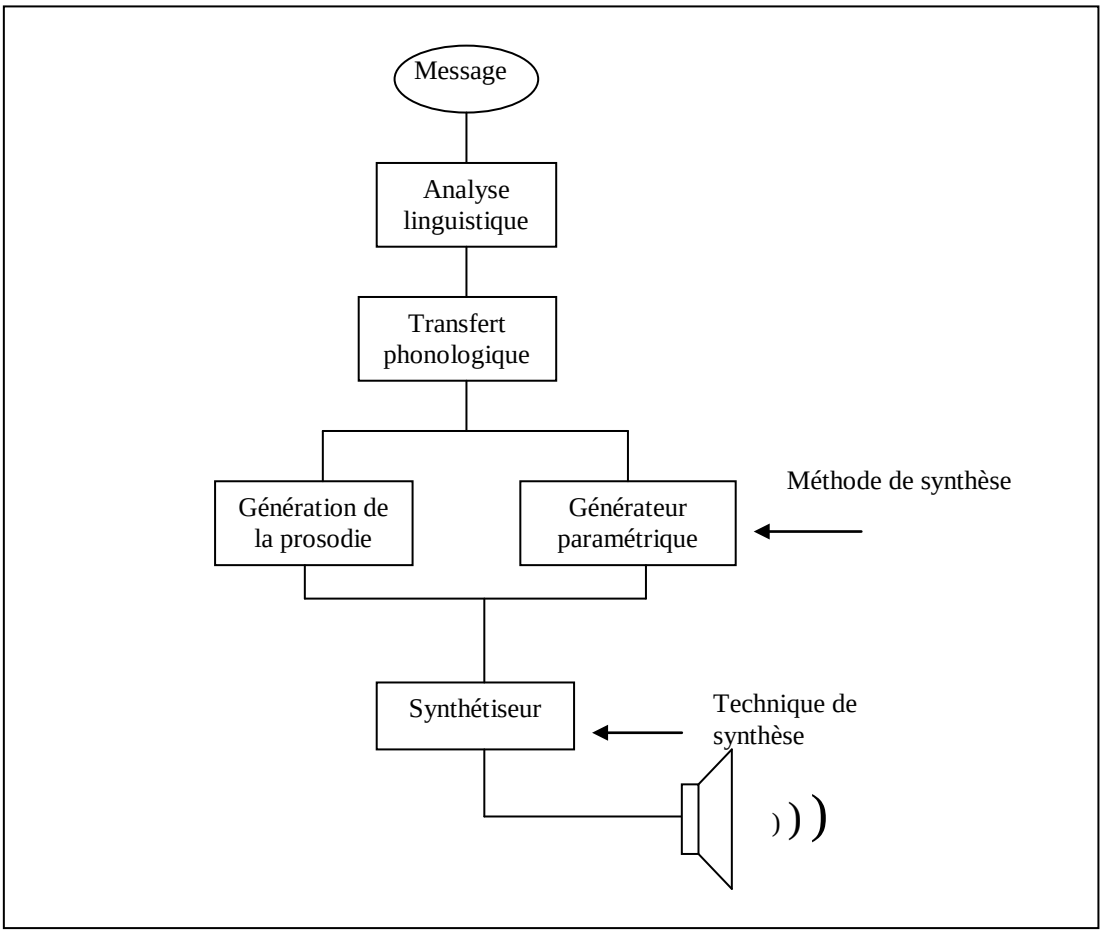

Fig.2.1 : Architecture générale d'un système de synthèse à partir du texte.

3. TECHNIQUES D'ANALYSE DE LA PAROLE

Il existe plusieurs techniques d'analyse du signal de parole : l'analyse temporelle, spectrale, **FFT** (**F**ast **F**ourier **T**ransformer), **LPC** (**L**ineair **P**redictive **C**oding), Cepstrale,...

Ces techniques peuvent être utilisées pour extraire les indices prosodiques (fréquence fondamentale ainsi que les paramètres pertinents du signal vocal), mais pour l'extraction de la durée phonémique on fait appel à d'autres techniques, par exemples : le vocodeur à canaux (Voice Coder) et le sonagraphe [2,4,9,10].

3.1. Vocodeur à canaux

Le vocodeur à canaux est un appareil destiné à transmettre la parole à un faible débit d'informations. Il consiste à représenter la fonction de transfère du conduit vocal par l'énergie du signal dans un certain nombre de canaux fréquentiels. L'excitation est représentée par une décision de voisement et la valeur de la fréquence fondamentale.

Ce vocodeur est constitué de deux grandes parties : l'analyseur et le synthétiseur (voir fig.2.4).

3.1.1. L'analyseur du vocodeur

La fonction d'analyse de l'enveloppe spectrale est effectuée à l'aide de canaux dont le nombre peut varier, suivant les réalisations, entre 10 et 20. Chaque canal traite une bande de fréquence déterminée. Le signal de parole issu d'un microphone est analysé au moyen d'un banc de filtres passe bandes couvrant l'étendue spectrale de la bande téléphonique (300 à 3400 Hz).

Le signal délivré par chacun des filtres est suivi d'un étage de détection, lui-même suivi d'un étage de filtrage passe-bas limité à 25 Hz environ. Cette limitation des variations d'énergie a pour effet de modifier ou de supprimer les transitions rapides du niveau d'énergie.

L'analyseur comporte ainsi un détecteur de voisement. Il permet de différencier les sons voisés de ceux non voisés et de donner dans le premier cas la valeur du pitch.

Les signaux issus de la sortie de l'analyseur sont codés avant d'être transmis au synthétiseur.

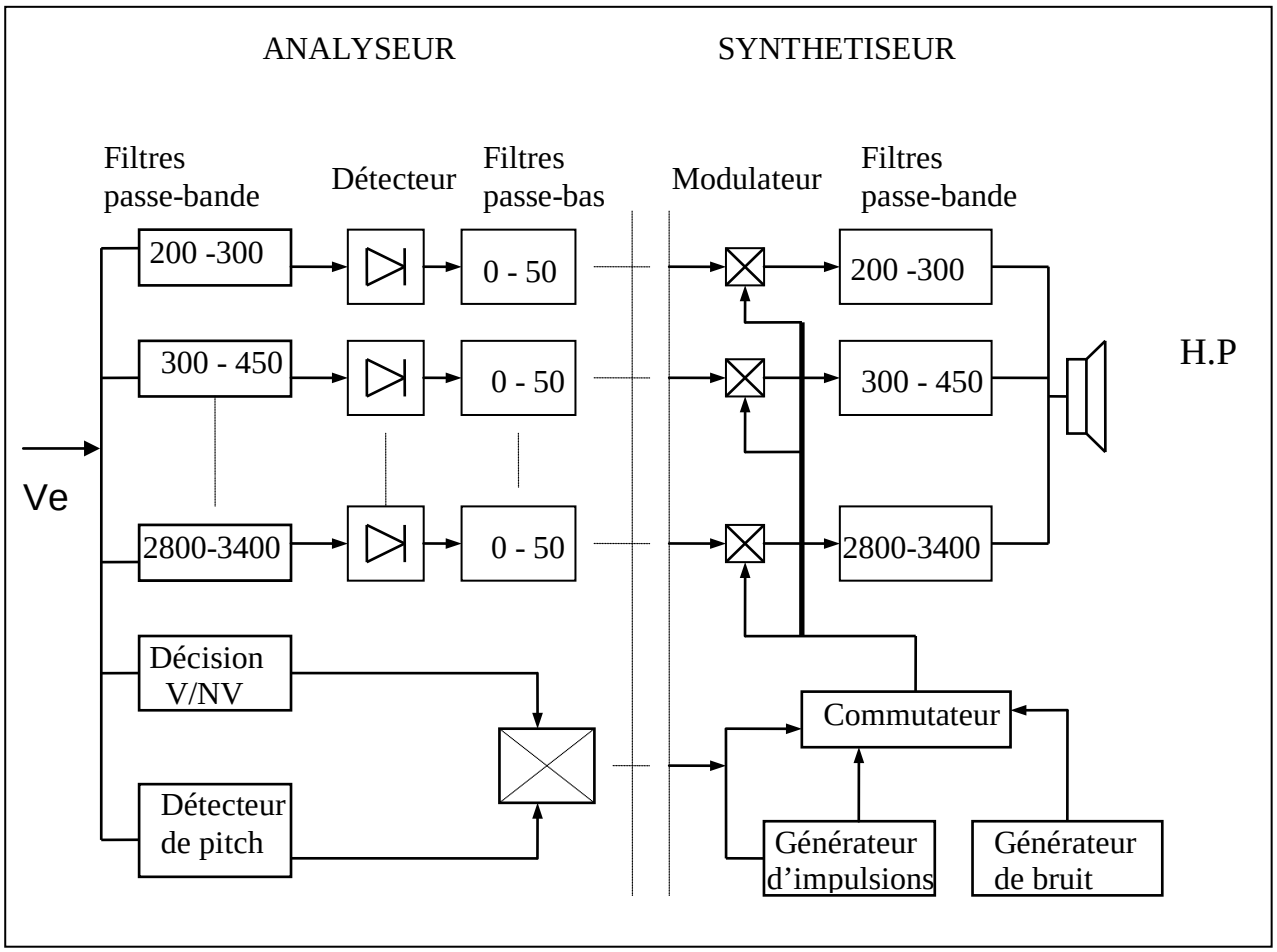

Fig.2.4 : Schéma fonctionnel d'un vocodeur à canaux.

3.1.2. Codage des signaux

Le codage désigne le passage de la représentation analogique à la représentation numérique caractérisée par son débit exprimé en bits/seconde.

Le codage permet le stockage et la transmission numérique de l'ensemble des informations relatives au spectre d'amplitude et à la structure fine de ce dernier. Cet ensemble d'information numérique est appelé : échantillon vocodeur.

Les informations numériques nécessaires pour le codage d'un échantillon vocodeur sont reparties comme suit :

- concernant le spectre : (nombre de canaux) × (4 bit/canal) ;
- concernant le pitch (période codée) on utilise 8 bits.

Pour un nombre de canaux égal à 16, et pour une fréquence f_e = 50 Hz. Le débit est de :

((16* 4)+8)*50 = 3600 bit/s (2.1)

Une fois le codage effectué, les informations numériques sont transmises et stockées dans la mémoire du calculateur.[4]

3.1.3. Le synthétiseur du vocodeur

La synthèse est effectuée à l'aide d'un banc de filtres passe-bandes identique à celui de l'étage d'analyse. Pour chaque canal, le signal basse fréquence issu du filtre d'analyse est multiplié par le signal d'excitation dans un modulateur. Selon que le son est voisé ou non voisé, le signal d'excitation attaquant les modulateurs proviendra soit d'un générateur d'impulsions soit d'un générateur de bruit.

Le signal de sortie est obtenu par addition des sorties des filtres de synthèse. On obtient ainsi la reconstitution du signal de parole [9].

3.2. Vocodeur à formants

Comme dans les vocodeurs à canaux, il s'agit d'une analyse-synthèse effectuée à l'aide de filtres passe-bandes.

La synthèse à formants se base sur un examen du mode de production de la parole par l'appareil vocal.

Le but de la synthèse à formants est de générer le signal de parole à partir des paramètres caractérisant les formants, leurs bandes passantes et éventuellement leurs amplitudes. Il s'agit donc d'une approche des caractéristiques acoustiques de la parole.

Les synthétiseurs de ce type utilisent en général des filtres tout pôle du second ordre qui sont associés en structure série, parallèle ou encore en structure hybride exemple le synthétiseur à formants de D. KLATT [4].

3.2.1. Structure série

Dans cette structure, les filtres sont placés en cascade, le second filtre ne peut renforcer que le signal délivré par le premier.

Le synthétiseur est commandé par 11 paramètres par exemple (fig.2.5) :

- les 4 formants vocaux F_1, F_2, F_3, F_4 avec F_4 fixé ;
- les 3 formants de bruit B_1, B_2, B_3 ;
- la fréquence du fondamental F_0 ;
- les 4 amplitudes A_0, A_b, A_n et A_1 de l'excitation vocale (sons voisés), du bruit, de la nasalité et du bruit injecté dans le canal vocal.

Les synthétiseurs série permettent une meilleure approximation du conduit vocal pour les voyelles, et ils présentent l'inconvénient de ne pas nécessiter des contrôles individuels pour les amplitudes des formants [9].

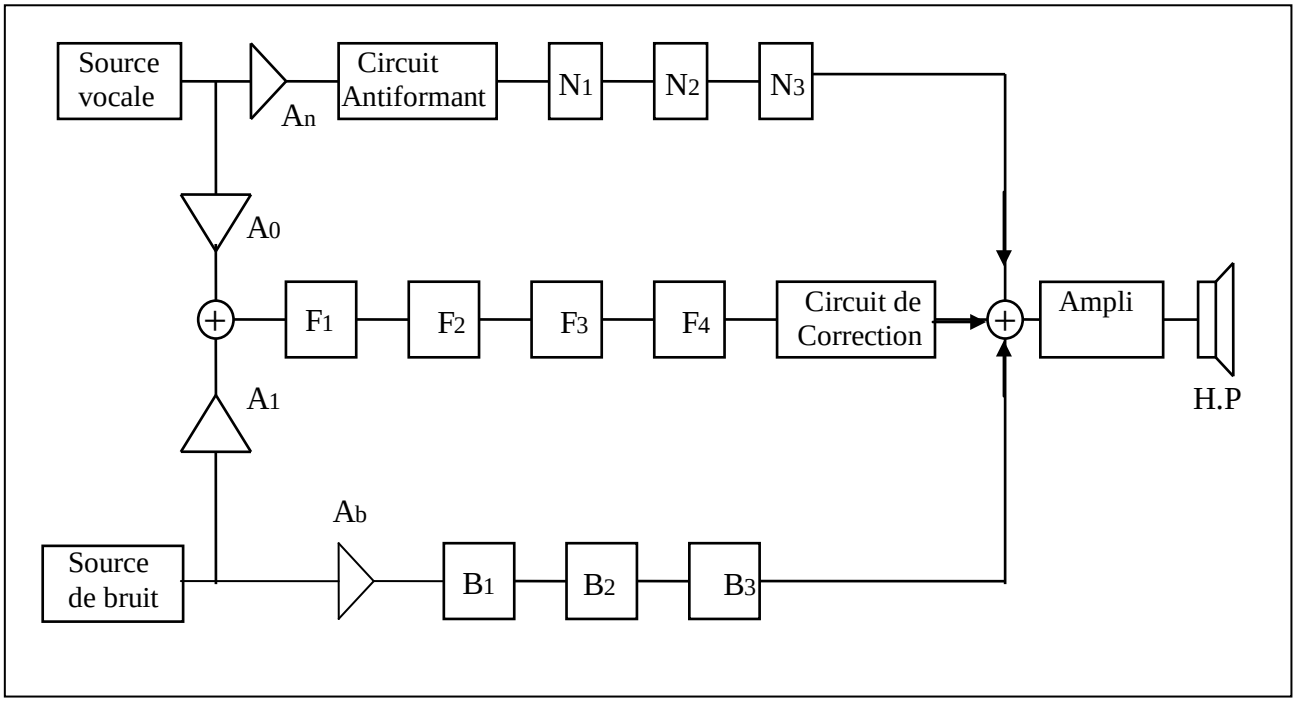

fig.2.5 : Synthétiseur à formants en structure série.

3.2.2. Structure parallèle

Dans cette structure (fig.2.6), les filtres sont tous alimentés par le même signal, qu'ils renforcent dans des bandes différentes. Chaque formant est contrôlé du point de vue de la fréquence et de l'amplitude avant d'être ajouté aux autres.

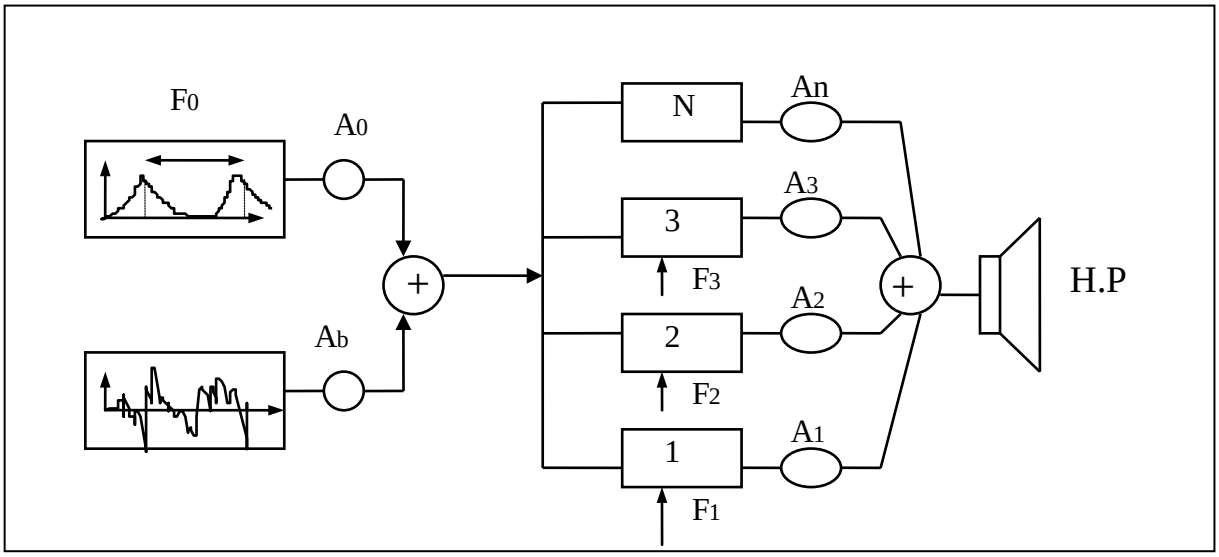

Fig.2.6 : Synthétiseur à formants en structure parallèle.

La structure parallèle permet donc de contrôler les amplitudes de chaque formant, ce qui est très important pour la synthèse des consonnes.[9]

3.3. Analyse par la prédiction linéaire

Dans la technique LPC, on effectue une analyse par codage prédictif du signal : on suppose qu'un échantillon du signal de parole peut être prédit à partir d'une combinaison linéaire d'un certain nombre d'échantillons précédents. On utilise donc des coefficients de pondération que l'on suppose constants sur une fenêtre représentant un court intervalle de temps (hypothèse de quasi stationnarité du signal de parole 10 à 30ms) .

Le modèle du conduit vocal est donc un modèle AR (Auto Régressif) fig.2.7.

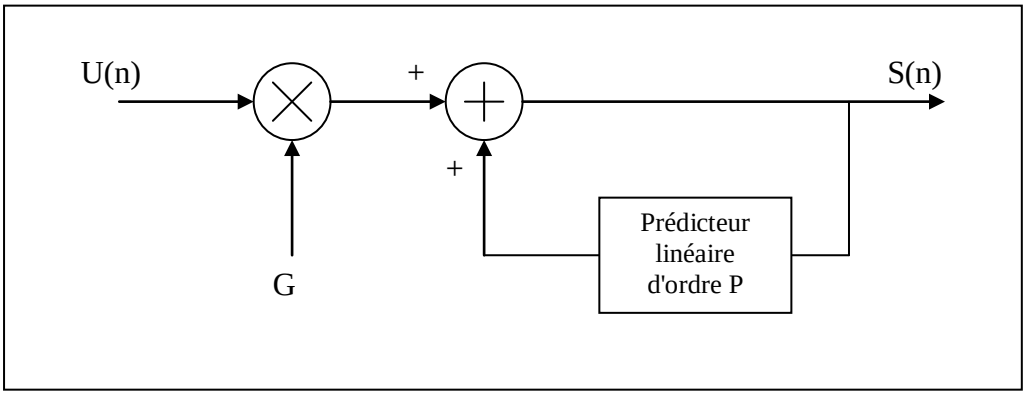

Fig. 2.7 : Modèle AR du conduit vocal.

Le model AR est régit par l'équation (2.2)

$$S(n) = \sum_{k=1}^{p} (a_k S(n-k)) + GU(n) \qquad (2.2)$$

Avec :

a_k sont les coefficients de perdition ;

U(n) échelon unité ;

G le gain du model ;

S(n) échantillon présent ;

S(n-k) $k^{ème}$ échantillon précédent.

Ce qui donne la fonction de transfert sous la forme de l'équation (2.3) :

$$A(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k . z^{-k}} \qquad (2.3)$$

Le synthétiseur à prédiction linéaire est donc réalisé suivant la forme de la figure ci-dessous (fig.2.8).

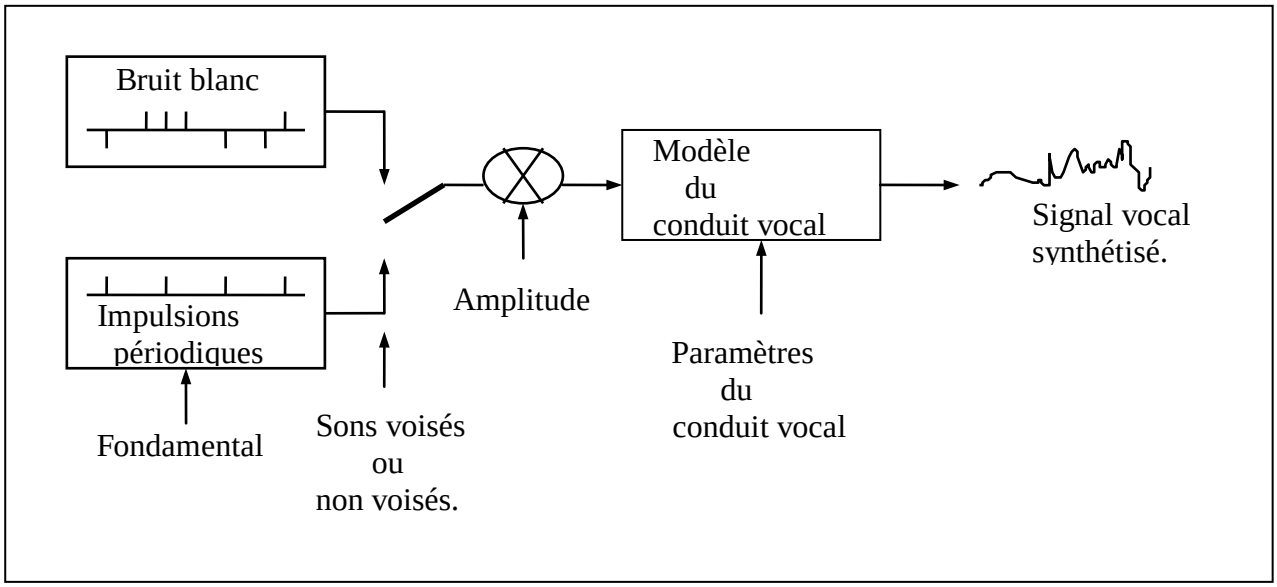

Fig. 2.8 : Modèle de synthétiseur à codage prédictif.

Les paramètres de contrôle fournis au synthétiseur sont :

- les coefficients de prédiction ;
- la décision V/NV ;
- le gain G ;
- la période T_0 du fondamental dans le cas d'un son voisé.

Le signale de sortie ou synthétisé sera prédit suivant la relation (2.4):

$$S(n) = \sum_{k=1}^{p} (a_k * S(n-k)) \qquad (2.4)$$

Ces paramètres sont renouvelés toutes les 10 à 25 ms. (la synthèse se fait au moyen d'un filtre récursif).

3.4. Simulation du conduit vocal

Dans cette technique, il s'agit de simuler le fonctionnement physique du système de production de parole.

Un modèle articulatoire reconstitue en premier lieu la forme du conduit vocal en fonction de la position des organes phonatoires (langue, mâchoire, lèvres). Le signal

vocal est en suite calculé à l'aide d'une simulation mathématique de l'écoulement de l'air dans le conduit ainsi délimité.

Les paramètres de commande d'un tel synthétiseur sont la pression subglottale, la tension des cordes vocales et la position relative des divers articulateurs.

3.5. Analyse par sonagramme numérique

Pour ce qui concerne notre étude nous avons utilisé la méthode d'analyse par sonagrammes numériques (sonagraphe) (fig. 2.9) pour une segmentation des mots du corpus en phonèmes, et l'extraction de leurs durées intrinsèques.[2,3,4,7,11].

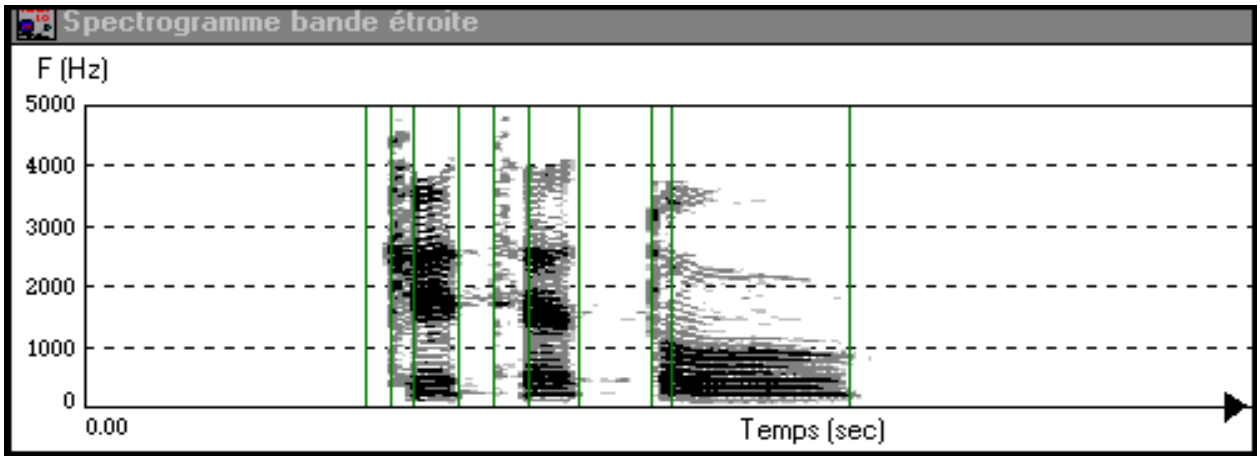

Fig.2.9 : le sonagramme numérique du mot [katat$_{e}$u:]
à l'aide du logiciel parole 1.0 [11].

L'analyse sonagraphique est bien antérieure aux possibilités de traitement numérique sur ordinateur : elle trouve ses fondements aux Etats-Unis autour de la seconde guerre mondiale à propos de l'étude sur la parole. Elle apparaît en France dans la première moitié des années cinquante. C'est essentiellement Emile Leipp qui s'en fera le défenseur et qui l'appliquera en particulier aux objets sonores musicaux. Le signal donné par un microphone était enregistré sur bande magnétique. Cette dernière, mise en boucle, est relue successivement à travers une batterie de filtres électriques passe-bande répartis régulièrement de 0 à 8000 Hz. L'énergie lue à travers chaque filtre est transcrite par un stylet sur un tambour recouvert d'un papier spécial (fig.2.10) [12]. On obtient un sonagramme développé sur une durée correspondant à la longueur de la boucle; le noircissement plus ou moins important

de chaque trace donne une indication sur l'intensité relative de la bande fréquentielle correspondante[2].

Un sonagramme [1] étant donc une représentation à trois dimensions (temps-fréquence et intensité ou énergie du son) sur laquelle se projette le spectre tridimensionnel de la parole.

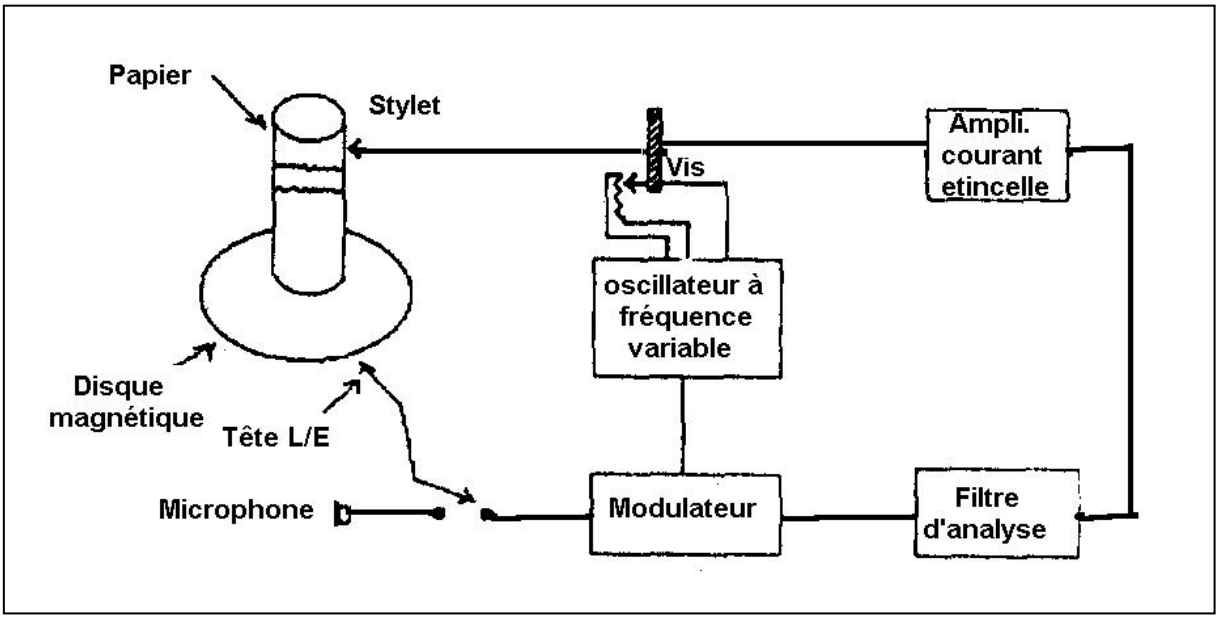

Fig.2.10 : le sonagraphe

Le sonagramme numérique est obtenu à partir du signal analogique donné par un microphone et numérisé en entrée d'un ordinateur, un logiciel adéquat découpe, pas à pas, le signal en tranches temporelles qui sont, soumises à une analyse par Transformée de Fourrier. La composition fréquentielle de chaque tranche apparaît et la concaténation de l'ensemble constitue le sonagramme de l'objet sonore initialement enregistré.

Le principe du sonagramme est fondé sur le filtrage hétérodyne :

Soit S(t) le signal de parole

et Š(t) le signal modulé par une porteuse sinusoïdale.

$$Š(t)=S(t)*\cos(2\Pi ft) \qquad (2.5)$$

Le spectre d'amplitude est décalé vers les fréquences positives (fig.2.11) et défile sous un filtre fixe de bande passante B. Selon la largeur du filtre on obtient un sonagramme large bande ou en bande étroite dans lequel l'intensité est représentée par le degré de noircissement.

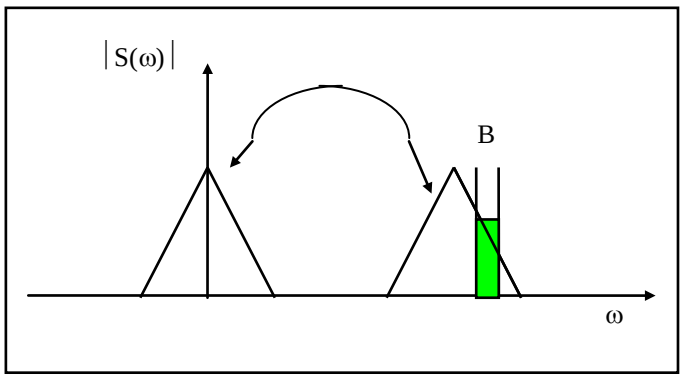

Fig.2.11 : spectre de S(t) modulé

On remarque que si B tend vers zéro, le spectre tend vers la Transformée de Fourrier de S(t).

Cette technique numérique permet aux traiteurs du signal de la parole suivant sa représentation tridimensionnelle de déterminer les segments des sons leurs fréquences ainsi que leurs énergies.

Les variations d'amplitude seront marquées par le noircissement plus ou moins grand des bandes. (On peut aussi représenter l'amplitude par des courbes de niveau tracées tous les 6 dB, utile surtout pour l'acoustique architecturale, et non pas pour l'analyse de la parole).

Au niveau des **formants**, les sorties sur papier du sonagraphe présentent parfois des défauts au niveau de la résolution, en particulier en ce qui concerne les amplitudes (marquées par le degré de noircissement) dont on obtient généralement une représentation ne dépassant pas l'échelle 1/5 [13,14].

Il faut prendre en considération pendant l'analyse d'un spectrogramme :

- le fait qu'on réduit à deux dimensions un "objet" qui en possède trois ;
- l'existence d'une relation d'incertitude entre temps et fréquence ;

- la plupart des mesures sont réalisées à l'aide d'un sonagraphe (aussi appelé spectrographe acoustique). Un sonagraphe est essentiellement un filtre passe-bande : le choix de la fréquence de filtrage joue un rôle fondamental (en général, on effectue plusieurs échantillonnages, typiquement à 10, 45, 150 et 300 Hz).

Il s'agit essentiellement d'effectuer un choix entre les paramètres qui nous intéressent :

- un spectrogramme à bandes étroites (10-45Hz) offre une bonne résolution au niveau fréquentiel, mais l'analyse temporelle est moins fine ;
- inversement, un spectrogramme à bandes larges (150-300Hz) offre une meilleure résolution temporelle et permet de dégager les formants vocaliques, mais apporte moins d'éléments pour l'étude du domaine fréquentiel.

En repérant les instants de transition, on peut faire correspondre des symboles phonétiques à diverses phases du spectrogramme. On peut étudier l'évolution mélodique d'un signal, liée, en première approximation, au fondamental, mais on se sert en pratique d'un harmonique de rang élevé (le dixième pour une voix d'homme, le cinquième pour une voix de femme ou d'enfant). On peut étudier le rythme du signal en observant l'évolution des durées des segments phonétiques qui le composent.

En général on utilise l’analyse en bande étroite si l’on s’intéresse à l’intonation ou si on veut voir la structure harmonique (structure du pitch) d’un son voisé (fig.2.12.a). Et l’analyse en bande large si l’on s’intéresse à l’évolution des formants (fig.2.12.b)

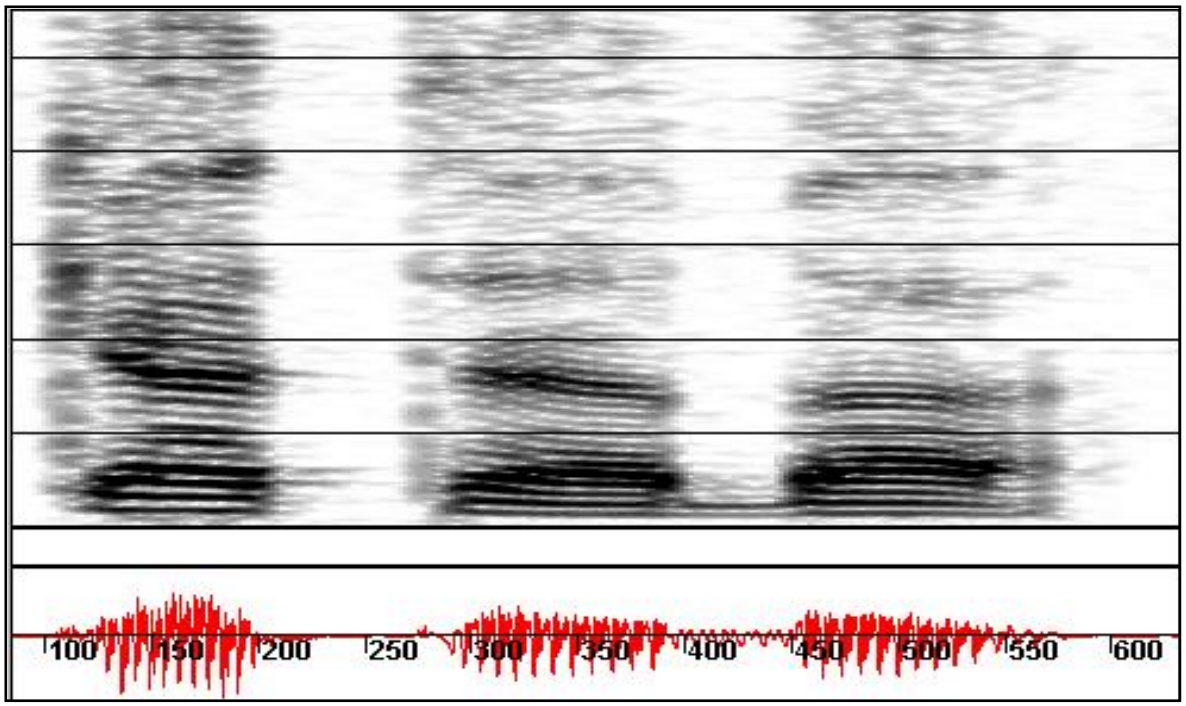

Fig.2.12.a : sonagramme à bande étroite du mot [kataba].

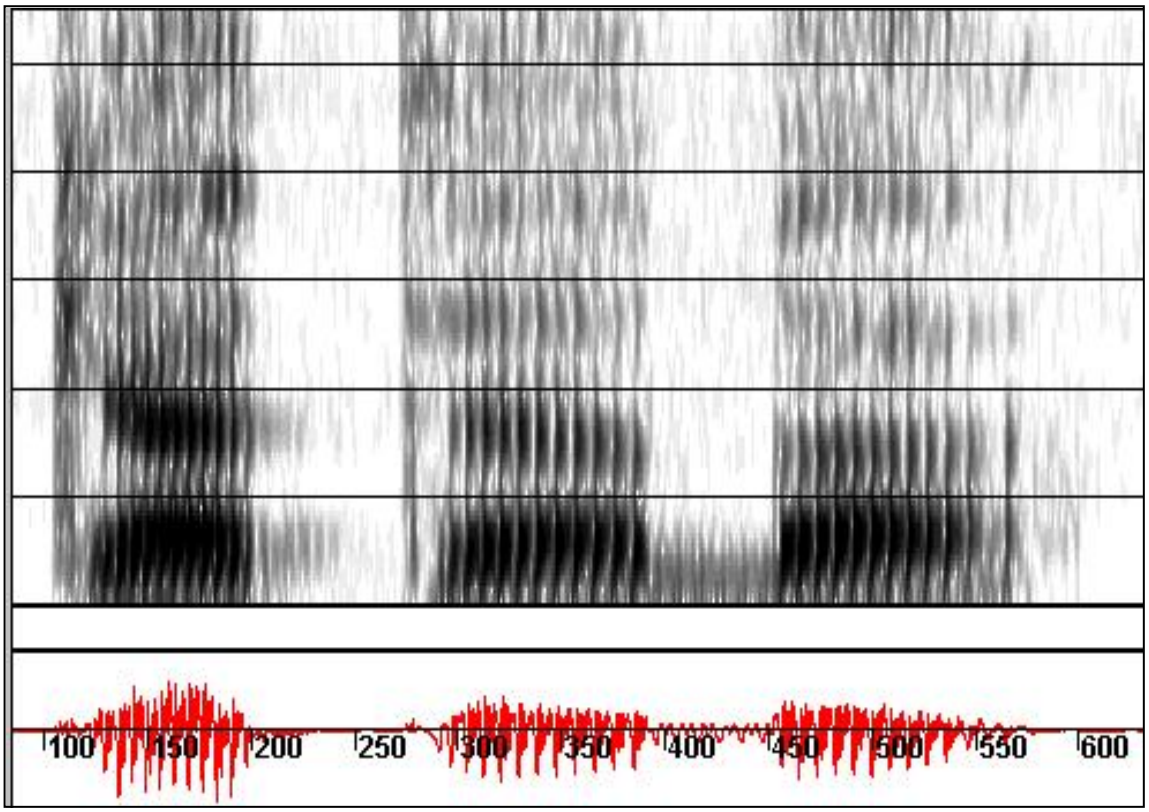

Fig.2.12.b : sonagramme à bande large du mot [kataba].

4. LES METHODES DE SYNTHESE

Les méthodes de synthèse se divisent d'abord quant à la taille du vocabulaire employé. Les systèmes dits à vocabulaire limité (cas de l'horloge parlante) ou à vocabulaire illimité [14,15].

4.1. La synthèse à vocabulaire limité

Dans cette méthode de synthèse nous pouvons élaboré la synthèse par phrase ou la synthèse par mot. La synthèse donc est obtenue par la concaténation (assembler en juxtaposer) des mots (ou phrases) préalablement enregistrés et stockés dans une mémoire formant ce qu'on appel le dictionnaire.

On note que cette méthode n'est pas une vrai synthèse mais il s'agit d'une reproduction du signal préenregistré.

4.2. La synthèse à vocabulaire illimité

Dans le cas de la synthèse à vocabulaire illimité, la concaténation d'unités comme le mot ou la phrase ne convient pas, du fait que la capacité mémoire dans la quelle ces unités soient stocker est limitée, empêchant d'agrandir le dictionnaire.

Ceci a obligé les chercheurs à trouver une méthode pour réduire la taille du dictionnaire. Cette méthode se base sur des unités plus petites "les Phonèmes" qui sont les éléments essentiels de la production de la parole.

La synthèse devient donc plus facile et le dictionnaire plus réduit ; mais l'intelligibilité reste mauvaise et ceci à cause de la zone de transition entre deux phonèmes successifs, ou encore le phénomène de coarticulation (influence d'un son sur son voisin du fait que les variations du conduit vocal lors de la production des sons est lent), dont il faut tenir compte lors des variations prosodiques dans la phrase à synthétiser [1,2,4,15].

Ce qui a permet de trouver d'autres méthodes à savoir la synthèse par règle ou par diphones.

4.2.1. La synthèse par règles

Le principe de la synthèse par règles, est la modélisation des transitions entre phonèmes sous forme de règles. A partir d'une représentation formantique du signal de parole, ces règles décrivent, par exemple, l'évolution des formants entre deux valeurs cibles [15,16,17].

Dans un tel système on stocke très peu de données (les valeurs cibles) mais, le nombre de règles nécessaires à décrire les transitions formantiques peut devenir très grand et difficile à établir si l'on souhaite une reproduction fidèle de la parole humaine.

4.2.2. La synthèse par diphones

Dans cette méthode de synthèse (appelée aussi synthèse par concaténation d'unités stockes), au lieu de modéliser la transition on la stocke dans des lexiques de signal. Ces unités minimales pour la concaténation, serons appelés "diphonème" ou "diphone" ou encore "dyade" [15,17,18].

4.2.3. Le diphonème

Le diphonème est défini comme « le segment qui s'étend de la zone stable d'une réalisation phonémique à la zone stable de la réalisation suivante et qui protège en son centre toute la zone de transition » [2,9,19,20] (fig. 2.2, fig. 2.3).

La tendance actuelle est d'augmenter de plus en plus la taille de l'unité à stocker, toujours dans le même soucis de préserver des transitions complexes, comme dans les clusters consonantiques ou dans le cas des semi-voyelles. Le terme de polyson a été proposé pour cette nouvelle unité.

Dans un tel système, la mémoire exigée peut devenir grande, mais le nombre de règles de concaténation est réduit.

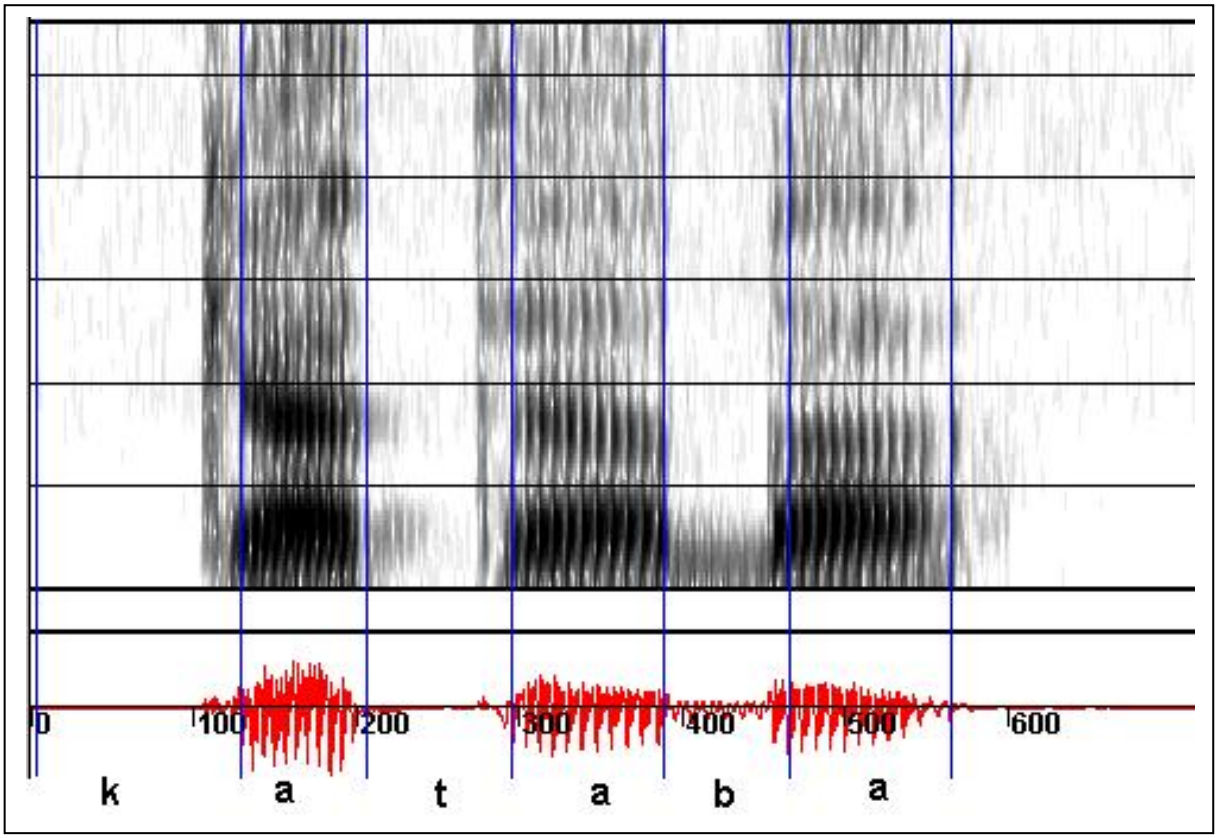

Fig.2.2 : Exemple de segmentation en phonèmes du mot [kataba].

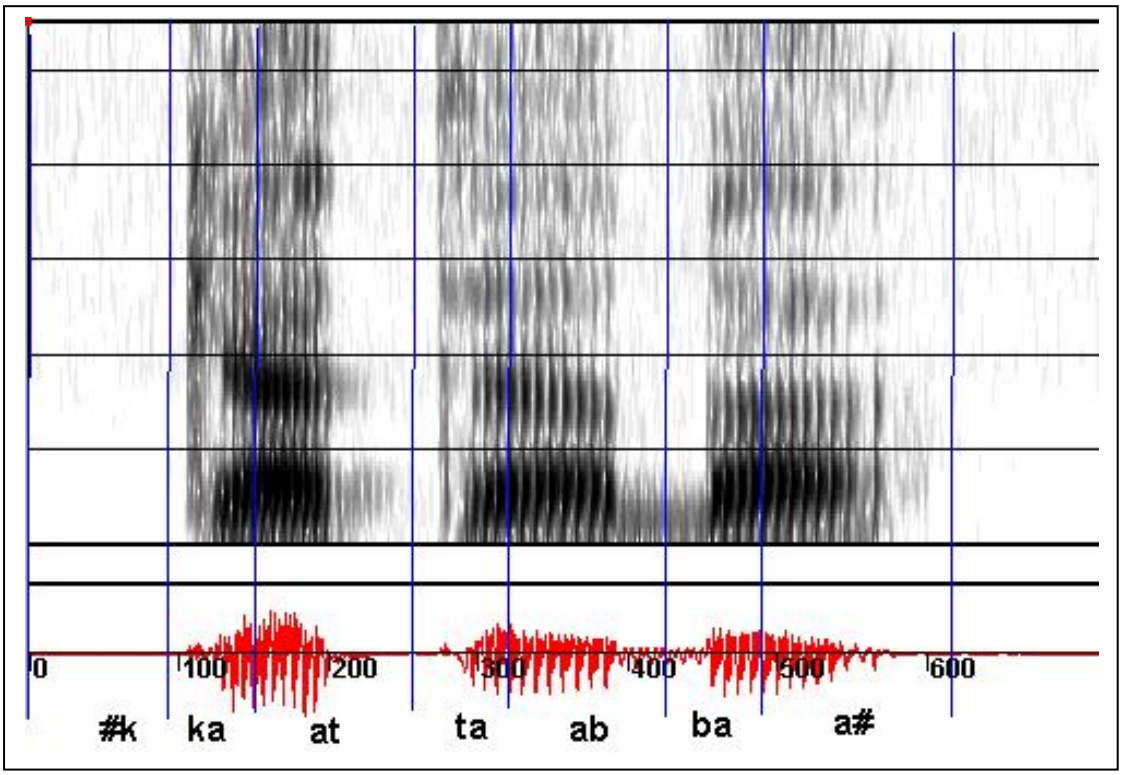

Fig.2.3 : Exemple de segmentation en diphonèmes du mot [kataba].

4. CONCLUSION

Les outils de traitement de la parole, nous aident suivant le cas à identifier, déterminer ou synthétiser la parole constituée par les différents paramètres obtenus lors de l'analyse.

L'analyse sonagraphique se prête à une grande variété de situations expérimentales. Elle donne à la fois des réponses qualitatives très pertinentes sur les objets acoustiques qui composent le monde sonore et donne également des réponses quantitatives sur leur structure. Elle révèle, à travers le sonagramme, une lecture facile des caractéristiques des sons pour les traiteurs du signal vocal.

La synthèse de la parole à partir du texte nécessite l'analyse des unités à stocker qui peuvent être des mots des phonèmes ou des diphonèmes, le but est de les reproduire en tenant en considération des zones de transitions entre phonèmes et ceci à l'aide des règles de production ou par une concaténation directe.

CONCLUSIONS GENERALES

Pour obtenir une synthèse de la parole à partir du texte (**T**exte **T**o **S**peech : TTS) de haute qualité nous sommes amenés à étudier les caractéristiques du signal de parole à savoir : la fréquence fondamentale, l'intensité et la durée phonémique. Ces paramètres prosodiques varient en fonction des effets dus au variabilité interlocuteur (âge, sexe, émotion, la fatigue, lieu géographique…) et intralocuteur et ceux dus à la nature de la phrase (interrogative, affirmative, discours, poésie, …).

Nous avons donc choisi d'étudier pour la synthèse de l'Arabe Standard, un de ces paramètres à savoir les durées phonémiques dans les sons spécifiques à l'Arabe Standard et de voir leurs variations en fonction du contexte.

Nous avons commencé par l'élaboration et l'enregistrement du corpus d'étude. Puis analysé ce corpus afin de prendre des mesures concernant les durées phonémiques.

Notre corpus est réalisé suivant le dictionnaire des diphones de l'Arabes Standard élaboré par M.GUERTI [18], et qui sont placés dans des logatomes (mots artificiels).

Les mesures que nous avons faites permettent de proposer des durées intrinsèques des phonèmes étudiés. Les résultats obtenus sont proches de ceux trouvés par d'autres chercheurs [4],[28].

Et en s'inspirant du modèle prédictif de D.KLATT [26], nous avons proposé des règles de prédiction de ces durées en fonction du contexte. Ces règles peuvent être utilisées dans la synthèse par règles appliquée aux diphones de l'Arabe Standard.

Ces résultats doivent être complétés par les méthodes de prédiction des autres paramètres prosodiques : fréquence fondamentale et intensité, afin d'obtenir une synthèse de la parole de haute qualité.

Finalement nous avons élaboré un outil de synthèse de la parole TTS dans lequel nous avons mis nos règles de prédiction des durées phonémiques des sons étudiés afin de les vérifier.

Cet outil qui nous a donné des résultats satisfaisants, dispose d'une base de donnée de phonème et de diphonème de l'Arabe Standard pouvant être complétée par l'études des autres cas et paramètres prosodiques.

Nous espérons que ce modeste travail puisse être utilisé pour des fin d'améliorer la synthèse de la parole et d'obtenir une synthèse de haute qualité en particulier pour la langue Arabe Standard.

REFERENCES BIBLIOGRAPHIQUES

[1] CALLIOPE "la parole et son traitement automatique" (Masson 1989).

[2] J. Hatton, J. M. Pierrel, G. Perrenou, J. Caelen, J. L. Gauvain, "La reconnaissance automatique de la parole". Dunod, Paris, France 1991.

[3] A. AMROUCHE "Contribution à la synthèse de la parole en Arabe Standard, modèle de prédiction des phonèmes". thèse de magistère. USTHB. Algérie. mai 1995.

[4] R. BOITE "Traitement de la parole". Presse Polytechniques Romandes (Lausanne 1987).

[5] G. BAILLY " Contribution à la détermination automatique de la prosodie du Français parlé à partir d'une analyse syntaxique, établissement d'un modèle de génération ", thèse de l'Institut National Polytechnique de Grenoble. 1983

[6] M. DEBYECHE "Reconnaissance automatique des consonnes glottales et pharyngales de l'Arabe standard en parole continue multilocuteurs". thèse de magistère. USTHB. Algérie. 1991.

[7] R. DJERADI "Logiciel de visualisation et de traitement du signal de parole : application à l'analyse des voyelles de l'Arabe". thèse de magistère. USTHB. Algérie. 1991.

[8] F. BEAUGENDRE " Une étude perceptive de l'intonation du français, développement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte ", Thèse de doctorat en sciences de l'Université de Paris XI, Notes et Documents LIMSI n° 94-25. 1994

[9] T. BARBE. "méthodologie et outils pour la mise en œuvre automatique d'une synthèse de la parole de haute qualité". Thèse de doctorat institut nationale polytechnique de GRENOBLE, France nov. 1988.

[10] M. GUERTI ''Méthodes et techniques d'analyse et de synthèse de la parole.'' S.S.A. pp 135-151. BLIDA. Algérie, 13-15 Déc. 1992.

[11] M. DEBYECHE "APHAK : un logiciel Interactif d'Analyse du signal de parole". 17ème journées Tunisiennes d'Electrotechnique et Automatique. JTEA'97, 5-6 Nov. 1997, Nabel, pp 304-309

[12]K. AMEDJKOUH et O. DJAIZ "Détection des formants en vue de la synthèse ou de la reconnaissance de la parole" mémoire d'ingénieur INI, 1994.

[13] Jean-Sylvain LIENARD "Les processus de la communication parlée" MASSON 1977.

[14] M. ROSSI " Connaissance et traitement automatique de la parole ", http://www.aupelf.fr/textinte/parole/rossi/rossi.html.

[15] M.GUERTI "Synthèse par règles à partir de diphones formantiques" conférence internationale SSA'99, Blida.

[16] Di CRISTO A., DI CRISTO P., VERONIS J. " Optimisation d'un modèle prosodique pour la synthèse par règles à partir du texte en français ", XXIIèmes Journées d'Etudes sur la Parole, Martigny, pp.135-137. 1998.

[17] F.J. CHARPENTIER " Traitement de la parole par analyse-synthèse de Fourier, application à la synthèse par diphones ", thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications. 1988.

[18] M. GUERTI "Contribution à la synthèse de la parole en Arabe Standard, synthèse par diphones et technique de prédiction Linéaire". Thèse de Magister, ILP-ALGER Algiers.

[19] Le petit Larousse "dictionnaire de français" 1983.

[20] F. EMERARD. "Synthèse par Diphones et traitement de la Prosodie". Thèse de docteur 3èmè cycle Grenoble mars 1977.

[21] D. O'SHAUGHNESSY, L. BARBEAU, D. BERNARDI, D. ARCHAMBAULT. "Diphone Speech Synthesis". Speech Communication. No 7, pp 55-65. 1988.

[22] Elizabeth COUPER-KUHLEN "An introduction to English Prosodie"

[23] Y. MORLEC "Génération multiparamétrique de la prosodie du français par apprentissage automatique" INPG, 1997.

[24] S. SELOUANI "Contribution à l'extraction des paramètres prosodiques du signal de parole cas de la fréquence fondamentale". thèse de magistère. USTHB. Algérie. 1991.

[25] Philippe LANGLAIS "Traitement de la prosodie en reconnaissance automatique de la parole". Université d'Avignon,1995.

[26] B. CHEBBINE "Extraction d'indices lies aux paramètres prosodiques de la langue Arabe". thèse de magistère. USTHB. Algérie. 1997.

[27] DI CRISTO A. " De la microprosodie à l'intonosyntaxe ", thèse de Doctorat d'état, Université de Provence, Aix-Marseille I. 1978.

[28] D .H. KLATT. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence". J.Acoust. Soc.Am.,Vol.59. No.5 pp.1208-1221. 1976

[29] K. BARTKHOVA. "Nouvelle approche dans le modèle de prédiction de la durée segmentale". Actes 14-ème J.E.P. S.E.F. PARIS. France pp188-191. 1985.

[30] A. Ahmed "L'effet des consonnes d'arrière et des emphatiques sur la nature acoustique des voyelles longues de l'arabe littéral marocain" thèse de doctorat, Université Laval 1995.

[31] A. AMROUCHE, B. BOUDRAA, J. M. ROUVAEN "Modèle de prédiction des durées des phonèmes pour la synthèse automatique de la parole en Arabe standard". COMAEI/96 vol. No2. pp12-16. Télemcen 3-5 déc 1996 Algérie.

[32] CHRIS S. CRAIG "GoldWave Version 3.02" copyright 1995 .

[33] Yves LAPRIE, Christophe RENARD "WinSnorri version 1.02.04" copyright CNRS France 1995

[34] M.L.BENZAOUI, M.GUERTI "Durées intrinsèques des voyelles de l'Arabe Standard". Pp120-125. SNAS'99. 9-10/11/99 Annaba, Algérie.

[35]] M.L.BENZAOUI, M.GUERTI " Durées des son spécifiques à l'Arabe Standard". A.J.O.T., Série B, Vol. 14, N° 1, 1999

[36] M.L.BENZAOUI, M.GUERTI "Phonemic Duration of Plosive Consonants Specifics to the Standard arabic". CATAEE'99. 19-20/10/99 Jordanie.

[37] M.L.BENZAOUI, M.GUERTI "Prédiction des Durées phonémiques des Consonnes Fricatives spécifiques à l'Arabe Standard". Pp229 - 232. SNCS'2001. 30-31/0/2001 Djelfa, Algérie.

[38] K. REISDORPH "Delphi 4, formation en 21 jours" Soms Publishing 1998.