N° d'ordre: 50/2016-C/ELN

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et d'Informatique

THÈSE

Présentée pour l'obtention du **diplôme** de **DOCTORAT 3ème Cycle (LMD)**

En: ELECTRONIQUE

Spécialité: Télécommunications

Par: IMEDJDOUBEN Fayçal

Thème

Apport à la synthèse concaténative de la parole à partir du texte

Soutenue publiquement le 17/10/2016, devant le jury composé de:

M	FERGANI Belkacem	Professeur	à l'USTHB	Président
M	HOUACINE Amrane	Professeur	à l'USTHB	Directeur de thèse
Mme	FALEK Leila	Professeur	à l'USTHB	Examinatrice
Mme	KOURGLI Assia	Professeur	à l'USTHB	Examinatrice
M	BENSLAMA Zoubir	Maître de Conférence/A	à l'U.Blida	Examinateur

Résumé:

Le travail de cette thèse de doctorat s'inscrit dans le cadre de la réalisation d'une synthèse de la parole à partir du texte Arabe par concaténation. Cette dernière est partagée en deux grandes parties de traitements séquentiels. La première partie englobe tous les traitements linguistiques qui permettent de produire de la prononciation correspondante au texte d'entrée afin d'utiliser ces symboles phonétiques pour produire de la parole synthétique. Cette partie est subdivisée en deux phases de traitement. La première phase concerne la transcription graphème-phonème en utilisant deux méthodes différentes pour transcrire le texte Arabe. La première méthode est basée sur un lexique des exceptions, alors que la deuxième méthode utilise une base de règles de transcription graphème-phonème. La deuxième phase concerne la génération des allophones en utilisant les phonèmes issus de la transcription graphème-phonème, la génération des allophones se fait en utilisant une base de règles de transcription phonème-allophone. L'ensemble aboutit à notre système de phonétisation automatique pour la synthèse de la parole à partir du texte Arabe. La deuxième partie englobe tous les traitements acoustiques du synthétiseur de la parole qui permettent de générer de la voix synthétique. Cette partie consiste à la sélection des unités acoustiques préenregistrées à concaténer, stockées dans une base de données acoustique. Ensuite ces unités acoustiques subissent des traitements spécifiques au point de concaténation selon la nature des sons à concaténer (voisés, non voisés) afin de générer un signal de parole synthétique le plus naturel et intelligible possible.

Mots clés: Langue Arabe, synthèse de la parole à partir du texte, transcription phonétique, allophone, traitement de la parole, synthèse concaténative, diphone, marqueur de pitch, TD-PSOLA.

Abstract:

The work of this doctoral thesis is part of the realization of a speech synthesis from Arabic text by concatenation. This latter is divided into two large parts of sequential treatments. The first part includes all the linguistic processing which allow to produce the pronunciation corresponding to the input text in order to use these phonetic symbols to produce synthetic speech. This part is subdivided into two processing phases. The first phase concerns the grapheme-phoneme transcription by using two different methods to transcribe the Arabic text. The first method is based on a lexicon of exceptions, whereas the second method uses a base of rules of grapheme-phoneme transcription. The second phase concerns the generation of allophones by applying the rules of phoneme-allophone transcription to the phonemes obtained from the grapheme-phoneme transcription. The whole system performs an automatic phonetization for Arabic text-to-speech synthesis. The second part includes all the acoustic processing of the speech synthesizer that allow to generate the synthetic speech. This part consists in selection of prerecorded acoustic units to be concatenated from in an acoustic database. Then these acoustic units undergo specific processing at the point of concatenation according to the nature of the sounds involved (voiced, unvoiced) in order to generate an Arabic synthetic speech as natural and intelligible as possible.

Key words: Arabic language, text to speech, phonetic transcription, allophone, speech processing, concatenative synthesis, diphone, pitch mark, TD-PSOLA.

REMERCIEMENTS

Je remercie le bon Dieu de la foi qu'il nous a donnée pour aller toujours de l'avant.

Ma gratitude se dirige tout d'abord à mes parents qui ont toujours témoigné de leur présence et surtout de leur patience depuis le début de mes études et en particulier durant la période du déroulement de ma thèse de Doctorat.

Je tiens à remercier toute personne ayant contribué de près ou de loin afin que ce modeste travail voit le jour.

Mes remerciements se dirigent en premier lieu à mon Directeur de thèse Prof. HOUACINE Amrane qui a su me diriger et suivre mon travail de recherche et pour ses précieux conseils. Je tiens à remercier Mr FERGANI Belkacem pour avoir accepté de présider le jury de soutenance. Comme je remercie les membres du jury Mr BENSLAMA Zoubir, Mme FALEK Leila et Mme KOURGLI Assia, pour avoir bien voulu participer au jury et examiner ce travail. Je tiens aussi à remercier Mr REMRAM Youcef qui a mis à ma disposition son laboratoire (Laboratoire d'Instrumentations (LINS)) qui dispose d'une chambre sourde afin d'effectuer les enregistrements sonores.

J'exprime mes plus profonds respects à tous mes enseignants qui m'ont guidé depuis le début jusqu'à la fin de notre cursus universitaire ainsi que pour le savoir et la connaissance qu'ils m'ont transmis.

Je n'oublie pas de remercier tous mes amis qui ont su apporter compassion et présence dans les moments difficiles; Rafik, Riad, Ramzi, Tahar … et tous les autres.

MERCI A TOUS

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION GENERALE

Notre thème de recherche porte sur la conception et la réalisation d'un système de synthèse de la parole par concaténation à partir du texte Arabe. Cet axe de recherche est un champ vaste et difficile du fait de la nécessité d'acquérir des connaissances sur divers plans pour la mise en œuvre d'un système fonctionnel. Ces connaissances touchent plusieurs axes de recherche telle que la linguistique, le traitement de la parole et l'informatique. Le travail de cette thèse de doctorat admet des contributions à différents niveaux pour la réalisation d'un synthétiseur de la parole à partir du texte arabe, ceci que ce soit sur les traitements linguistiques effectués sur le texte arabe en utilisant une structure variable des règles de transcription de la langue arabe afin de générer la transcription phonétique adéquate, que ce soit sur les traitements acoustiques effectués sur les unités acoustiques (diphones) au point de concaténation dans le but de réduire les discontinuités perceptibles au niveau de cette région pour améliorer la qualité de la parole synthétisée.

La parole est un moyen de communication très important chez l'être humain. Ce dernier a envisagé d'intégrer la parole dans des systèmes qui sont capable d'interagir avec lui, dans le but de rendre la machine plus vivante et plus proche de l'être humain. Ces énormes progrès qui ont été fournis au cours de ces trente dernières années ont été récompensés dans divers domaines d'application de la synthèse de la parole et ont même permis de remplacer parfois l'homme par des machines parlantes. Mais d'autres efforts restent à fournir dans ce champ de recherche afin d'assurer la perfection de tels systèmes.

L'objectif de la synthèse de la parole à partir du texte est de produire un signal acoustique de parole à partir de n'importe quel texte (introduit directement à l'aide d'un clavier, ou scanné et reconnu par un système de reconnaissance optique des caractères, ou générer automatiquement par un système homme-machine). Ce signal de parole généré artificiellement doit être de bonne qualité, soit sur le plan de l'intelligibilité ou sur le plan du naturel.

La synthèse de la parole à partir du texte est un domaine qui rassemble deux disciplines différentes: traitement de la langue et traitement de la parole, ces deux disciplines sont deux étapes essentielles, complémentaires et indissociables afin de réaliser un synthétiseur de la parole à partir du texte.

Le traitement de la langue concerne tous les traitements linguistiques permettant le passage de la représentation orthographique du texte en entrée, à une représentation phonétique en sortie. Cette discipline fait intervenir les différentes branches de la linguistique telle que la phonétique (étude des sons d'une langue), la phonologie (étude de l'organisation des sons d'une langue), la syntaxe (étude de l'arrangement des mots d'une langue), la grammaire (étude des éléments qui constitue une langue). Ces dernières sont combinées afin

de produire de la prononciation correspondante d'un texte donné et fournir les informations prosodiques de chaque son (fréquence fondamentale, durée, intensité,…) dans le but de les utiliser dans le bloc des traitements acoustiques.

Le traitement de la parole englobe tous les traitements acoustiques du synthétiseur de la parole qui permettent la génération du signal de la parole synthétique le plus intelligible et naturel possible. Cette discipline utilise les outils de traitement du signal dans le but de modifier les unités acoustiques de parole utilisées lors de la synthèse de la parole de sorte à avoir en sortie du synthétiseur de la parole, un signal de parole synthétique le plus proche possible par rapport au signal de parole naturelle.

Il existe plusieurs techniques de synthèse de la parole. La technique la plus utilisée et qui donne de bons résultats, c'est la synthèse de la parole par concaténation qui consiste à générer le signal de parole synthétique en concaténant des unités acoustiques (phonèmes, diphones, triphones, syllabes,…) obtenues par segmentation du signal de parole naturelle. Ces segments de parole sont stockés dans une base de données acoustique.

Cette thèse de Doctorat est partagée en cinq chapitres. Le premier chapitre est consacré aux généralités sur la synthèse de la parole à partir du texte. Ce chapitre donne un aperçu général sur les différentes techniques de la synthèse de la parole ainsi que les domaines d'application de cette dernière.

Le deuxième chapitre traite de la transcription phonétique du texte Arabe. En spécifiant toutes les particularités de la langue Arabe et en détaillant l'ensemble des règles établies pour la transcription du texte Arabe ainsi que les différentes étapes que nous avons suivies afin d'aboutir à notre système de phonétisation automatique.

Dans le troisième chapitre nous abordons l'étape de la génération des allophones. Celui-ci contient en détail la démarche que nous avons suivie pour la génération des différentes réalisations sonores (allophones) du parlé Arabe. Il contient également les différentes règles établies de transformation des phonèmes en allophones.

Par contre, le quatrième chapitre est consacré aux traitements acoustiques du synthétiseur de la parole. Ce chapitre comporte tous les traitements spécifiques effectués sur les portions de parole à concaténer ainsi que l'algorithme développé afin de produire un signal de parole synthétique de bonne qualité.

Enfin, le cinquième chapitre est consacré à l'évaluation de notre synthétiseur de la parole à partir du texte. Cette phase d'évaluation est partagée en deux grandes parties. La première partie est consacrée à l'évaluation de la transcription phonétique générée, alors que la deuxième partie est dédiée à l'évaluation de la qualité de la synthèse de la parole produite par notre système de synthèse de la parole à partir du texte Arabe.

CHAPITRE I: GENERALITES SUR LA SYNTHESE DE LA PAROLE A PARTIR DU TEXTE

I.1 Introduction:

La synthèse de la parole à partir du texte est une application qui permet de lire un texte orthographique en utilisant une voix humaine synthétique issue de la segmentation des enregistrements du signal de parole naturelle. Ce type d'application a vu une croissance rapide dans ces dernières années du fait que la parole est un langage de communication, le plus facile et le plus efficace entre les personnes. D'autre part la nécessité d'intégrer la parole dans divers domaines d'application (télécommunications, automobile,…) est apparue du fait de l'évolution rapide de la technologie et des exigences des consommateurs.

Le développement d'un système qui produit de la parole à partir du texte (Text-To-Speech) prend énormément du temps à cause de la complexité de cette tâche qui est très difficile à réaliser puisque cet axe de recherche fait intervenir plusieurs domaines à la fois tels que le traitement de la langue, le traitement acoustique du synthétiseur de la parole et la partie programmation de l'ensemble [Eke02], [Rou06].

La qualité du signal de parole généré par les synthétiseurs de la parole à partir du texte est un paramètre crucial afin de juger si la synthèse est de bonne qualité ou non [Lem99]. Les deux critères d'évaluation les plus utilisés sont l'intelligibilité (la parole produite doit être compréhensible) et le naturel (le degré de similitude avec la voix humaine). La valeur de ces deux critères d'évaluation varie selon la technique de la synthèse utilisée pour produire de la parole synthétique et selon les besoins de l'application à mettre en œuvre.

I.2 Structure de la synthèse de la parole à partir du texte:

L'architecture générale de la synthèse de la parole à partir du texte est représentée dans la Figure1. Cette dernière est composée de deux blocs de traitements séquentiels.

Le premier bloc est consacré au traitement de la langue du synthétiseur de la parole dans le but de faire correspondre le texte d'entrée à une séquence phonétique, ainsi que la génération des informations prosodiques du texte à synthétiser.

Le deuxième bloc entame la phase de la génération de la parole artificielle en faisant correspondre les séquences phonétiques issues des traitements linguistiques aux segments de parole préenregistrés afin de sélectionner les bons segments de parole. Ces derniers subissent

des modifications à l'aide des outils de traitement du signal en intégrant les informations prosodiques.

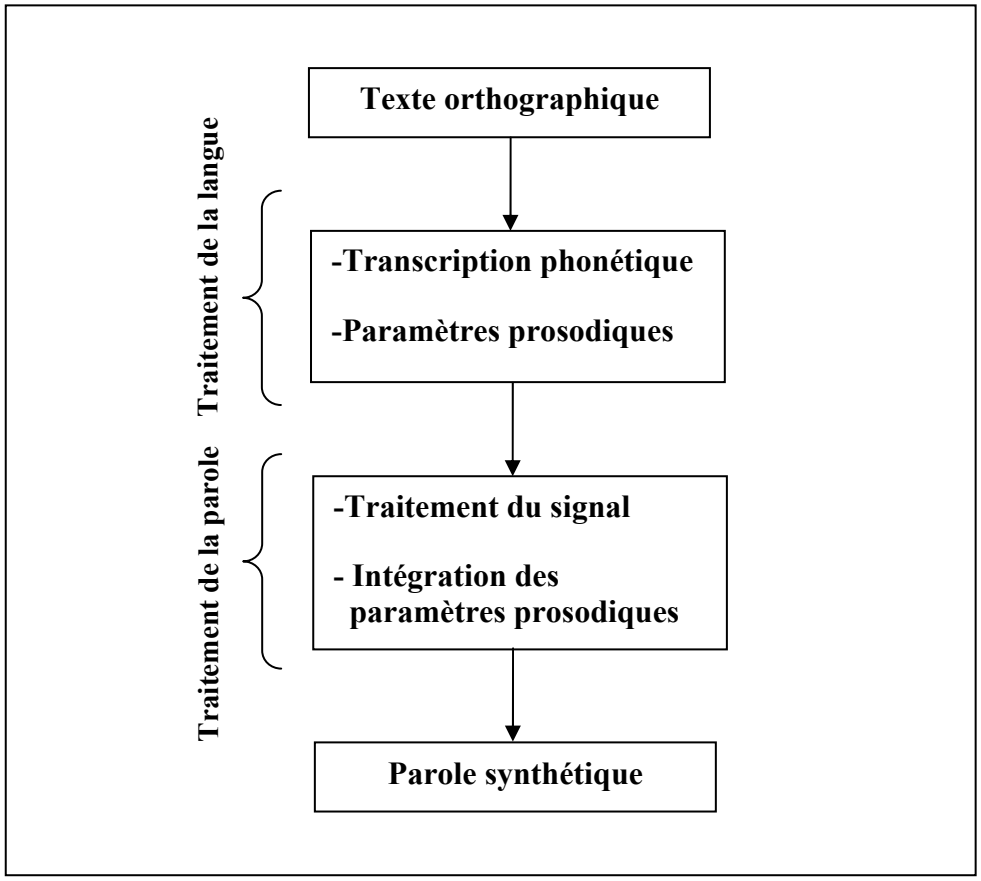


Figure1. Architecture générale de la synthèse de la parole à partir du texte.

I.2.1 Traitement automatique de la langue:

Le traitement de la langue est un domaine très vaste. Ce dernier regroupe plusieurs disciplines à la fois telles que la linguistique, l'informatique et l'intelligence artificielle dans le but de réaliser des programmes informatiques qui traitent la langue automatiquement.

Le rôle principal du traitement de la langue est de traiter automatiquement des données linguistiques afin d'exploiter ces résultats de traitement dans divers domaines en relation avec:

- Traitement du signal (traitement de la parole, synthèse de la parole et reconnaissance automatique de la parole).
- Production ou modification du texte (traduction automatique, correction orthographique, résumé automatique, génération automatique,…).
- Extraction d'information (classification de documents, recherche d'information,…).

Dans le cadre de cette thèse de doctorat nous n'allons nous intéresser qu'aux traitements linguistiques destinés à la synthèse de la parole à partir du texte.

I.2.1.1 Transcription orthographique-phonétique:

La transcription orthographique-phonétique a pour but de faire correspondre à chaque graphème (la plus petite unité distinctive de l'écriture) un ou plusieurs phonèmes (la plus petite unité distinctive de la chaîne parlée).

Avant toute transcription phonétique, il faut toujours commencer par mettre en forme le texte d'entrée (normalisation du texte). Cette étape s'appelle les prétraitements du texte qui a pour but de préparer le texte pour le bloc de traitement qui le suit afin que la transcription phonétique se fasse d'une manière juste. En général les prétraitements du texte se distinguent d'une langue à l'autre. Dans ce qui suit nous allons citer les cas les plus généraux des prétraitements du texte, à savoir:

- Traitement des espaces et des ponctuations.
- Découpage du texte sous forme de phrases et en mots.
- Traitement des sigles et des abréviations.
- Traitement des chiffres, dates et des symboles.

Ces prétraitements sont effectués à l'aide d'un dictionnaire contenant l'ensemble de ces anomalies. Ce dernier à pour but de réécriture des sigles, abréviations, chiffres, dates,…etc, sous forme de lettre afin de les transcrire phonétiquement par le bloc de phonétisation automatique du texte [Lev93].

La phase de la transcription phonétique (transcription graphème-phonème) est réalisée en général en combinant deux méthodes de transcription phonétique (voir Figure2). La première méthode se base sur un dictionnaire de phonétisation, alors que la deuxième méthode utilise des règles de transcription graphème-phonème qui se trouvent dans une base de règles de transcription phonétique [Bou01], [Eli00]. Ces règles de transcription phonétique sont organisées d'une façon hiérarchique.

La phonétisation à base d'un dictionnaire de phonétisation est une méthode très simple à mettre en œuvre mais la constitution du dictionnaire prend énormément du temps afin d'intégrer tous les mots d'exception.

Ce dictionnaire de phonétisation contient la liste des mots d'exception dont les prononciations sont données explicitement, plutôt que déterminées par les règles de la prononciation. La démarche employée pour transcrire cette liste de mots d'exception est simple, il suffit de faire une comparaison entre le mot testé avec la liste de mots d'exception

contenus dans le dictionnaire de phonétisation. En cas de succès le dictionnaire génère directement la transcription phonétique correspondante au mot testé sinon ce dernier sera transcrit par le biais des règles de transcription phonétique.

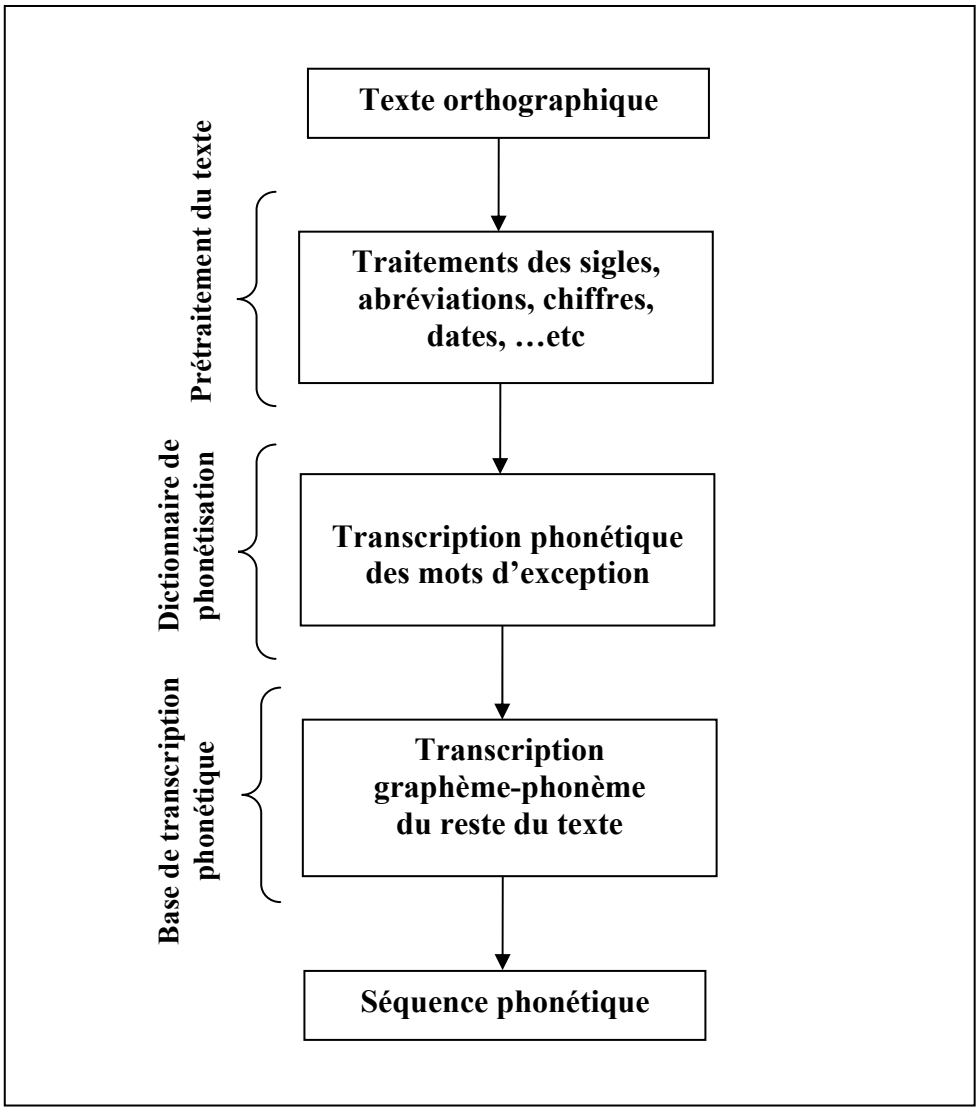


Figure2. Schéma général de la transcription orthographique-phonétique.

La deuxième méthode utilise des règles de prononciation afin de transcrire le texte orthographique (Figure3). Ces règles sont sauvegardées dans une base de règles de transcription phonétique. La démarche suivie pour faire la conversion graphème-phonème s'appuit sur un balayage des graphèmes du premier jusqu'au dernier en appliquant la règle qui le correspond, où chaque graphème est remplacé par un ou plusieurs phonèmes selon son contexte gauche et droit [Bal03], [Saï06].

La structure générale des règles établies suit la forme suivante:

Ph= CG + C + CD

Avec:

Ph: Résultat phonétique.
CG: Contexte gauche du caractère testé.
C: Caractère testé.
CD: Contexte droit du caractère testé.

Cette règle signifie que lorsqu'un caractère C, est précédé par le caractère CG et suivi par le caractère CD, aura comme résultat phonétique le phonème Ph. Cette structure de règles reste inchangée pour tous les graphèmes, donc chaque graphème du texte à transcrire phonétiquement utilise la même forme de règle d'où l'avantage d'uniformiser les règles de transcription graphème-phonème.

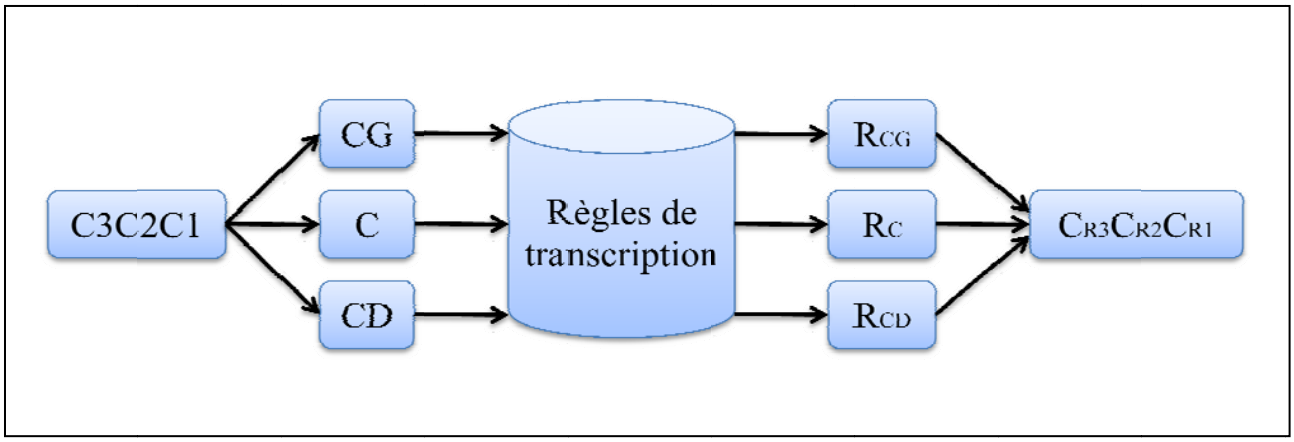


Figure3. Schéma du module de transcription par des règles de prononciation.

Les séquences phonétiques issues de la transcription graphème-phonème sont représentées généralement par des symboles issus de l'alphabet phonétique international (API), ou par des symboles issus de la norme SAMPA (Speech Assessment Methods Phonetic Alphabet).

L'alphabet phonétique international (API) est un système de représentation utilisé par les linguistes afin de représenter l'ensemble des sons du langage parlé des langues. Ces symboles phonétiques sont inspirés de l'alphabet latin et grec.

La norme SAMPA est un codage de la représentation phonétique des prononciations des langues utilisant le jeu de caractères ASCII (American Standard Code for Information Interchange) pour la représentation des symboles phonétiques utilisable sur ordinateur.

Afin de bien expliquer l'alphabet phonétique international, nous allons donner un exemple de la représentation phonétique de la langue Française en utilisant l'alphabet phonétique international (API) voir Table1.

API	Exemple	Type de la lettre	API	Exemple	Type de la lettre
b	**b**al, ro**b**e	consonne	œ	fl**eu**r	voyelle
s	**s**ouris, piè**c**e	consonne	e	**é**t**é**, n**ez**	voyelle
k	**c**arpe, **k**iwi, **qu**i	consonne	ɛ	m**e**r, j'aim**ais**	voyelle
d	**d**ate	consonne	o	s**o**t, s**eau**, s**au**t	voyelle
f	**f**ace, **ph**are	consonne	ɔ	p**o**rte, p**o**rt, m**o**rt	voyelle
g	**g**are, ba**gu**e	consonne	i	**i**lle, am**i**	voyelle
ʒ	**j**ournal, gor**g**e	consonne	u	c**ou**p, a**oû**t	voyelle
l	**l**a, a**l**ors	consonne	y	n**u**, j'ai **eu**	voyelle
m	**m**a**m**an	consonne	ɑ̃	r**an**g, av**an**t	voyelle
n	**n**on	consonne	ɛ̃	r**ein**, br**in**, p**ain**	voyelle
ɲ	**gn**ôle, a**gn**eau	consonne	ɔ̃	b**on**, t**on**	voyelle
p	**p**etit	consonne	œ̃	br**un**, **un**	voyelle
ʁ	**r**a**r**e	consonne	j	**y**eux, a**il**	semi-voyelle
t	**t**ordu	consonne	w	f**ou**et, v**oi**r	semi-voyelle
v	**v**oir, **w**agon	consonne	ɥ	f**u**ite , l**u**i	semi-voyelle
z	**z**èbre, o**s**er	consonne			
ʃ	**ch**at, **sh**ort	consonne			
a	p**a**tte, p**a**p**a**	voyelle			
ɑ	p**â**te, t**a**s	voyelle			
ə	f**e**nêtre	voyelle			
ø	j**eu**, f**eu**	voyelle			

Table1. Exemple de l'alphabet phonétique international (API) pour la langue Française.

I.2.1.2 Traitement prosodique:

Avant tout traitement prosodique, il faut toujours commencer par l'étape de la syllabation des séquences phonétiques issues de la transcription graphème-phonème. La syllabation est une étape essentielle dans les traitements prosodiques. Elle consiste en la décomposition des mots en syllabes (prononciation d'une seule émission de voix) [Cro00], dans le but d'avoir une séquence optimale de représentants (Figure4).

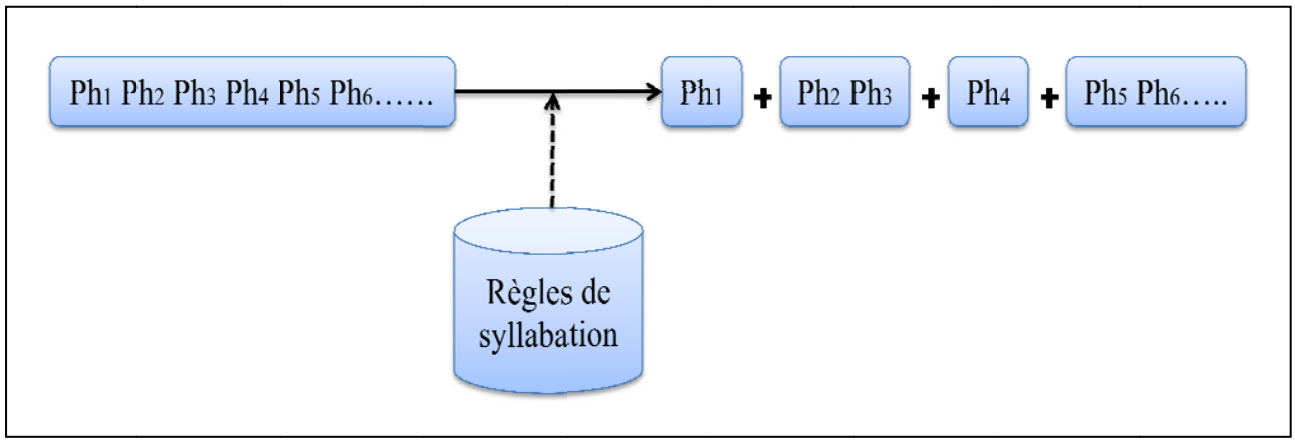


Figure4. Schéma de la syllabation.

La syllabe est une combinaison d'un ou plusieurs phonèmes (consonnes et voyelles). Il existe en général deux types de syllabes:

- Syllabe ouverte (une syllabe qui se termine par une voyelle prononcée).
  Ce type de syllabe peut être structuré par les schémas suivants:
  - ✓ Composée seulement d'une voyelle brève (V).
  - ✓ Enchainement d'une consonne et voyelle brève (CV).
  - ✓ Enchainement d'une consonne et voyelle longue (CVV).

- Syllabe fermée (une syllabe qui se termine par une consonne prononcée).
  Ce type de syllabe peut être structuré par les schémas suivants:
  - ✓ Enchainement d'une consonne, voyelle brève et une autre consonne (CVC).
  - ✓ Enchainement d'une consonne, voyelle longue et une autre consonne (CVVC).
  - ✓ Enchainement d'une consonne, voyelle brève et deux autres consonnes (CVCC).

- ✓ Enchainement de deux consonnes, voyelle brève et une autre consonne (CCVC).

La prosodie est un domaine de la phonétique qui a pour but de décrire les sons de la parole au niveau du langage parlé [Nes06]. La prosodie traite les phénomènes prosodiques suivants:

- l'intonation (correspond au ton de la voix).
- l'accentuation (correspond au marquage des accents sur les mots).
- le rythme (correspond à la succession de durées des sons de parole dans le temps).
- le débit (correspond à la vitesse d'élocution).
- la pause (correspond à un silence).

Ces phénomènes prosodiques peuvent se caractériser par des modifications aux niveaux de la fréquence fondamentale (pitch), l'intensité et la durée [Tho07], [Qua07], (voir Figure5).

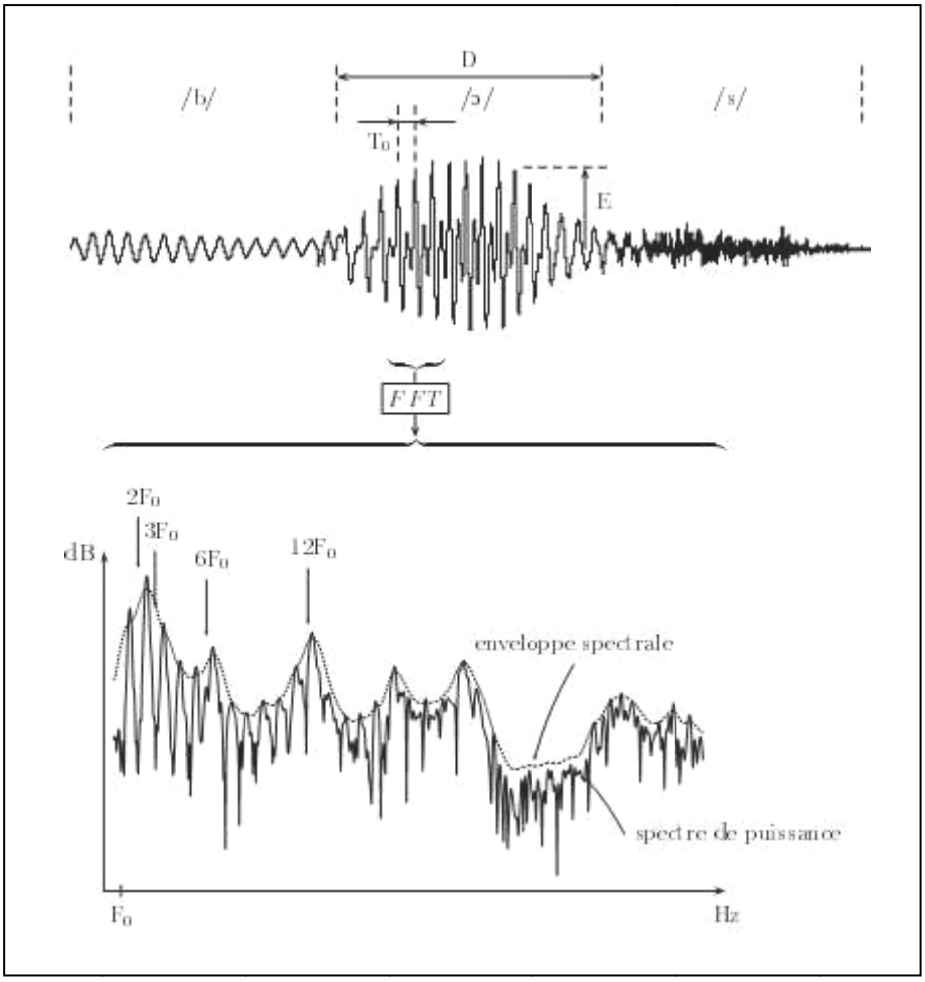


Figure5. Les paramètres prosodiques [Qua07].

- **La fréquence fondamentale (pitch):**

La fréquence fondamentale est définie comme la plus basse fréquence d'une forme d'onde périodique du signal de parole (l'inverse de la période du signal de parole). Lorsque la fréquence fondamentale est basse, le son aperçu est un son grave, à l'inverse lorsque la fréquence fondamentale est haute on obtient un son aigu (Figure6).

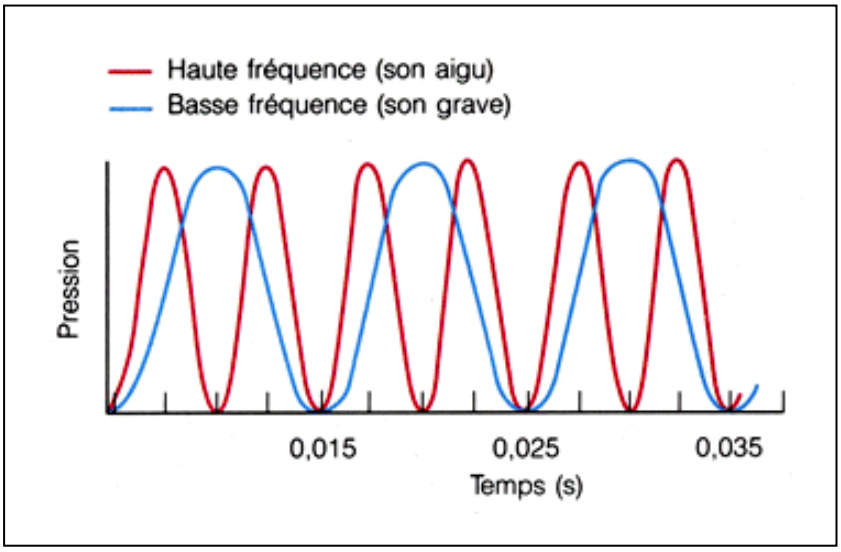


Figure6. Représentation d'un son aigu et d'un son grave.

Le calcul de la fréquence fondamentale s'effectue uniquement dans les portions de parole voisées, alors que les sons non voisés ont une fréquence fondamentale nulle du fait de la non vibration des cordes vocales dans ce cas (voir Figure7).

La fréquence fondamentale chez les êtres humains varie selon l'âge, la nature du sexe, et l'état émotionnel de la personne (heureux, nerveux,…) [Dut00], pour exemple:

- De 200 à 600 Hz chez les enfants.
- De 150 à 450 Hz chez les femmes.
- De 80 à 160 Hz chez les hommes.

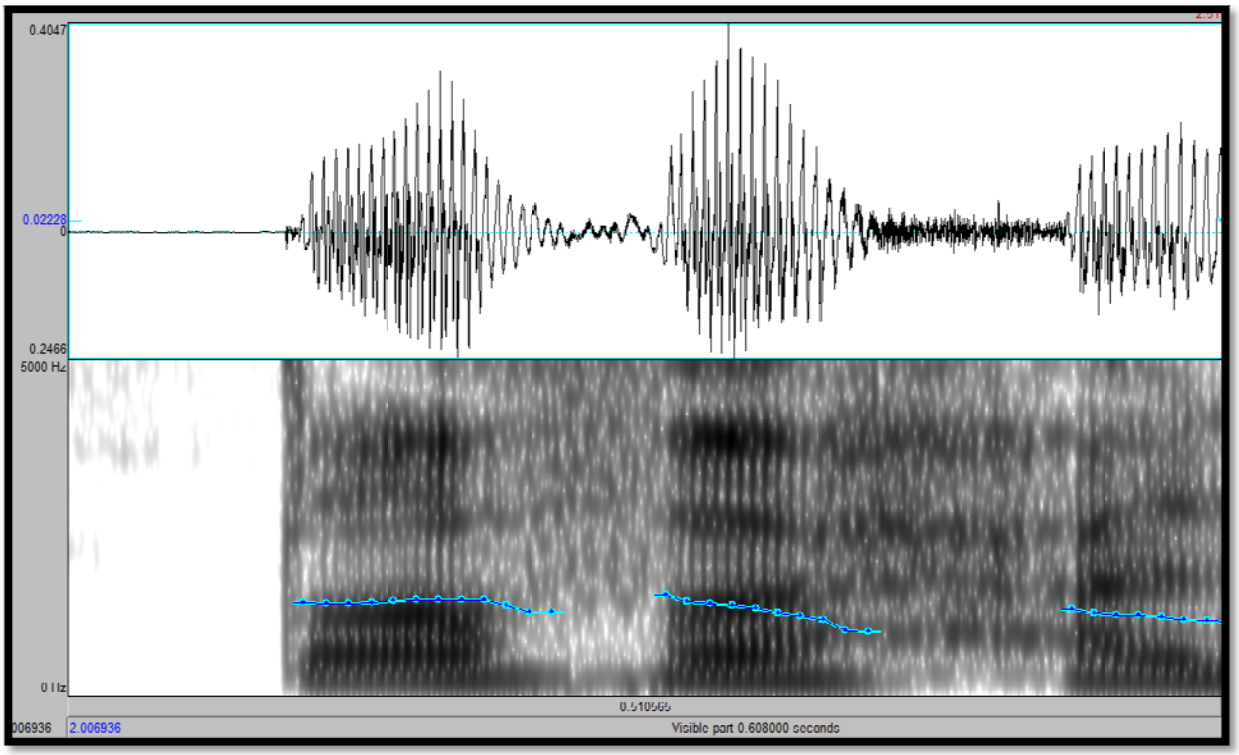


Figure7. Variation de la fréquence fondamentale dans le mot ‘طَفَحَ’.

Plusieurs techniques de détection de la fréquence fondamentale sont développées dans divers domaines, soit dans le domaine temporel, ou spectral, ou par l’utilisation des deux domaines à la fois [Che01].

Les techniques de la détection de la fréquence fondamentale les plus utilisées dans le domaine temporel sont les suivants:

- Méthode d’autocorrélation.
- Méthode AMDF (Average Magnitude Difference Function).
- Méthode DARD (Data Reduction).

Parmi les techniques de détection du pitch dans le domaine spectral on peut citer:

- Méthode CEP (Cepstral).
- Méthode HPS (Harmonic Product Spectrum).
- Intercorrélation avec le Peigne Spectrale.

- **L'intensité:**

L'intensité d'un signal de la parole est l'énergie de ce signal de parole. Cette dernière est calculée en prenant des trames de parole stationnaires (allant de 10 ms à 20 ms). L'intensité d'un signal de la parole permet de différencier entre deux sons de la parole (son fort, son faible).

L'intensité d'un signal de la parole est proportionnelle à l'amplitude (l'écart entre la valeur maximale et la valeur minimale) de ce signal. Donc si l'amplitude est d'une grande valeur, l'intensité est aussi grande, ce qui implique que le son produit est un son fort. L'inverse donne un son faible (voir Figure8).

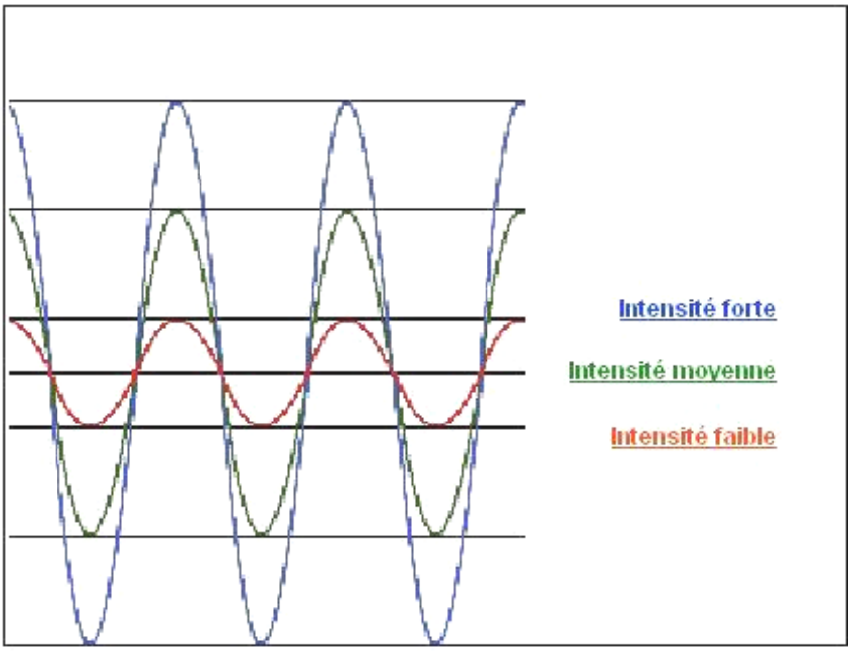


Figure8. Différentes intensités pour la même onde.

L'énergie (E) d'un signal x(n) est calculée par la formule suivante:

$$\mathrm{E} = \sum_{n=0}^{N} \mathrm{x(n)}^2$$

Cette énergie s'exprime aussi en décibels (dB) par la formule:

$$\mathrm{E(dB)} = 10 \times log_{10}\left(\sum_{n=0}^{N} \mathrm{x(n)}^2\right)$$

Dans la Table2 nous allons donner quelques exemples de niveau sonore des différentes sources acoustiques représentées à l'échelle des décibels (dB). L'oreille humaine ne perçoit pas les sons en dessous de 0 dB, alors que les sons au dessus de 120 dB peuvent détruire le système auditif humain.

Niveau sonore (dB)	**Sources acoustiques**
0	Seuil d'audition
10	Chambre sourde
20	Bruit de fond naturel dans le silence
30	Studio d'enregistrement
40	Campagne calme
60	Voix normale
70	Bureau calme
80	Orchestre symphonique
90	Voix criée
110	Sirène des pompiers
120	Seuil de douleur
140	Formule 1
150	Avion à réaction
170	Explosion

Table2. Exemple de niveau sonore des différentes sources acoustiques.

- **<u>La durée:</u>**

La fréquence d'échantillonnage étant définie, la durée d'un son de parole correspond à la longueur du signal ou le nombre d'échantillons de la séquence considérée.

La durée d'un phonème ou d'une syllabe varie selon différents critères (Table3). Ces critères peuvent être comme suit:

- Le type du son (voyelle, consonne).
- Le type de la syllabe (ouverte, fermé).
- La position de la syllabe (au début, au milieu, à la fin).
- Le type de la phrase prononcée (affirmative, interrogative,…).

Voyelles (brève, longue)	Durée moyenne (ms)	Syllabe ouverte / durée moyenne (ms)	Syllabe fermée / durée moyenne (ms)
/a/	88,78	94,84	78,73
/i/	87,41	93,02	80,66
/u/	87,06	91,09	83,46
/a:/	199,9	200,7	174,55
/i:/	181,36	-	-
/u:/	185,25	-	-

Table3. Exemple de différentes durées des voyelles (brève, longue) de la langue Arabe [Bou05].

I.2.2 Traitement de la parole:

Le traitement de la parole est l'étude des signaux de parole et les méthodes de traitement de ces signaux. Les signaux sont habituellement traités en une représentation numérique, de sorte que le traitement de la parole peut être considéré comme un cas particulier du traitement numérique du signal, appliqué à un signal de parole. Les aspects de traitement de la parole comprennent: l'acquisition, la manipulation, le stockage, le transfert et la production de signaux de la parole.

Les traitements de la parole destinés à la synthèse de la parole à partir du texte (Text-To-Speech) sont des traitements effectués à base des outils de traitement du signal dans le but de synthétiser de la parole artificielle d'une qualité acceptable. Cela est possible en intégrant les informations prosodiques générées déjà par les traitements linguistiques.

Le signal de parole peut subir des modifications dans le domaine temporel ou dans le domaine fréquentiel ou l'utilisation des deux domaines en même temps.

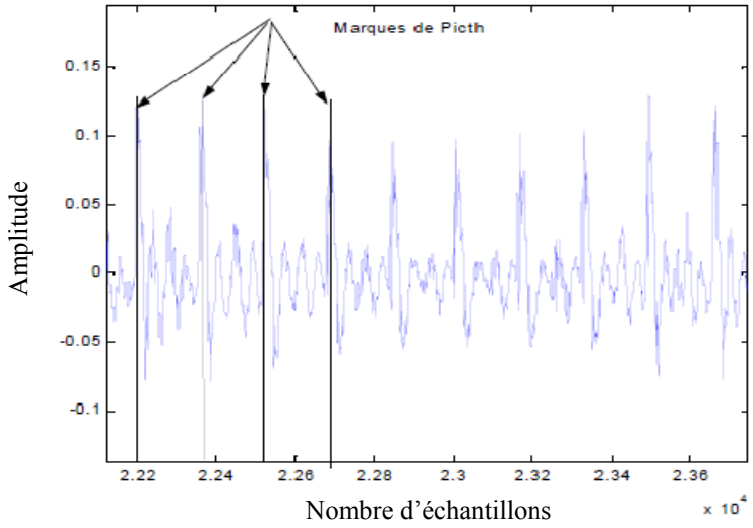
I.2.2.1 Domaine temporel:

Les outils de traitement du signal les plus utilisés dans le domaine temporel afin d'analyser et de modifier les portions de la parole sont les suivants:

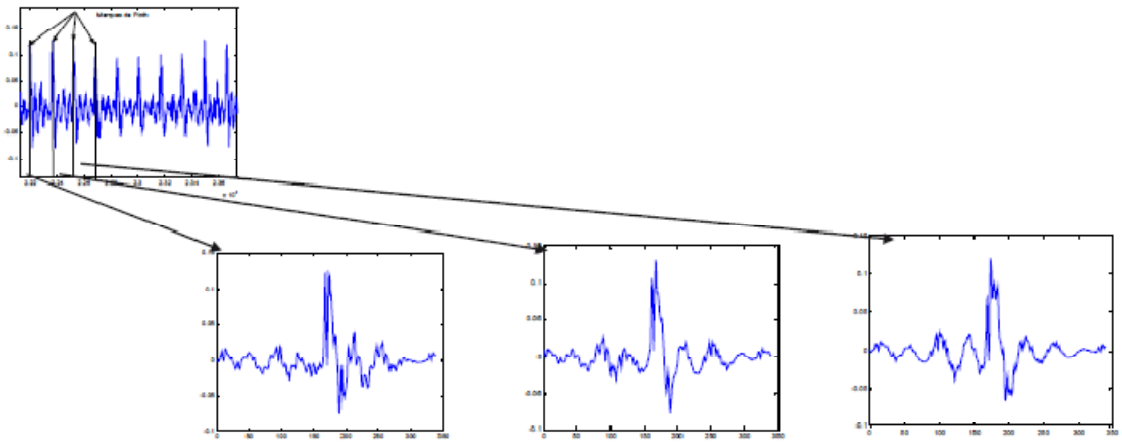
- Le calcul de l'énergie du signal.
- Le taux de passage par zéro.
- Le fenêtrage du signal.
- L'autocorrélation.

La modification des paramètres prosodiques dans le domaine temporel se fait généralement en utilisant la méthode TD-PSOLA (Time Domain Pitch Synchronous Overlap and Add) [Cha10], [She12]. Cette méthode permet de modifier à la fois la fréquence fondamentale du signal ainsi que la durée de ce dernier (Figure9). Le principe de cette méthode est simple et peut se résumer dans les étapes suivantes:

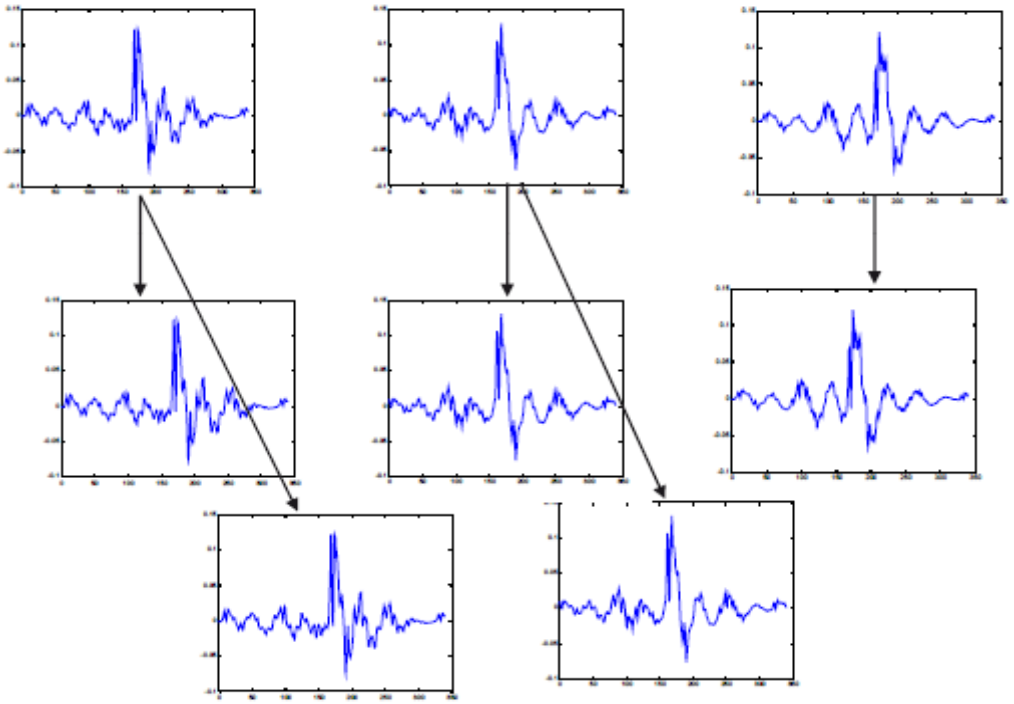
- Trouver les marques de pitch du signal.
- Appliquer la fenêtre de Hanning centrée sur la marque de pitch et s'étendant de la marque de Pitch précédente à la suivante.
- Addition des trames de parole avec un recouvrement, pour modifier la durée il faut soit dupliquer ou supprimer les trames. Concernant la modification de la fréquence fondamentale il faut tout simplement augmenter ou diminuer l'écart entre deux marques de pitch.



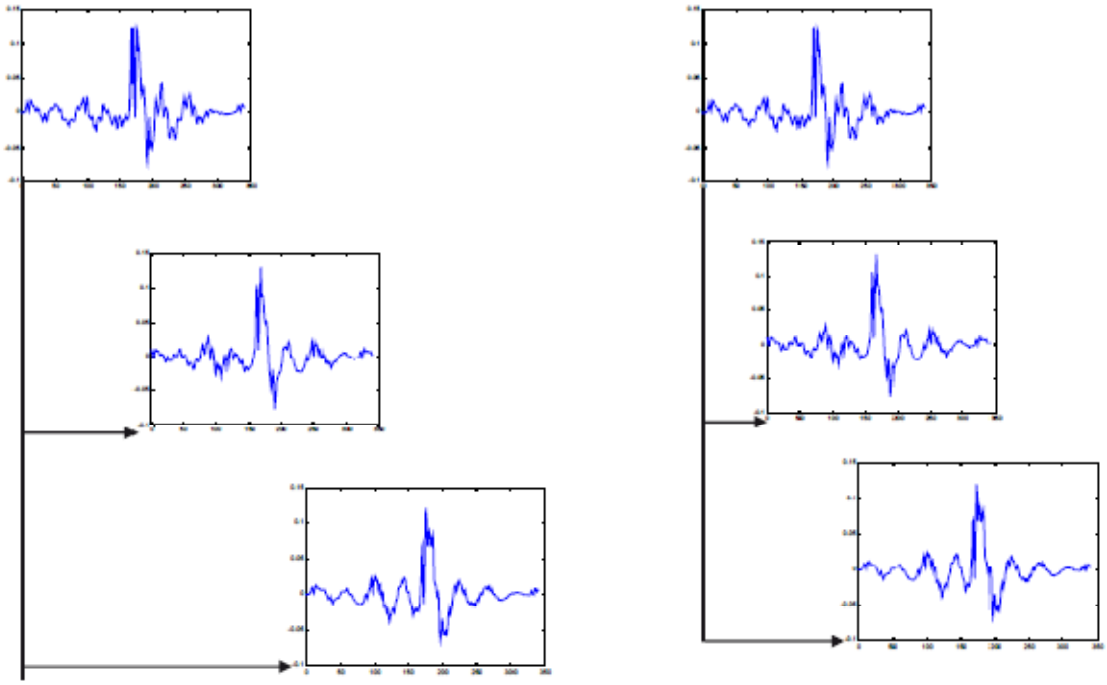
(a)



(b)



(c)



(d)

Figure9. Les étapes de la méthode TD-PSOLA, (a) détection des marques de pitch, (b) découpage du signal sous forme de petites trames de parole, (c) modification de la durée du signal, (d) modification de la valeur de la fréquence fondamentale [Ric10].

I.2.2.1 Domaine fréquentiel:

L'analyse dans le domaine fréquentiel à pour but de représenter le signal étudié par son spectre de fréquence. Cette transformation minimise le nombre énorme d'information redondante dans le domaine temporel, donc une seule composante fréquentielle peut représenter une trame d'information temporelle (compression d'information).

Quelques outils de traitement du signal utilisé dans le domaine fréquentiel pour l'extraction d'information pertinente du son de la parole, à savoir:

- La transformée de Fourier.
- La transformée de Fourier à court terme.
- Le spectrogramme.
- La densité spectrale de puissance.

Les méthodes qui permettent de synthétiser de la parole dans le domaine fréquentiel sont très nombreuses on peut citer:

- La synthèse par la FD-PSOLA (Fourier Domain Pitch Synchronous Overlap and Add).
- La synthèse par la modélisation sinusoïdale et recouvrement addition (Figure10).
- Synthèse par LPC (Linear Predictive Coding).

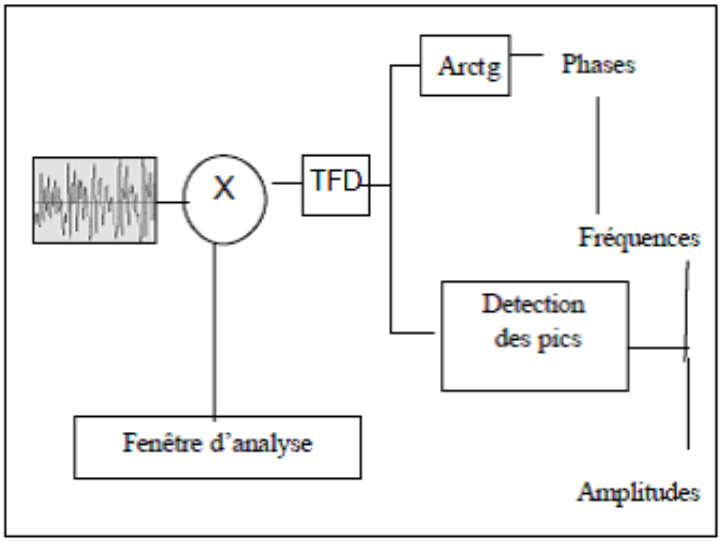


Figure10. Principe de la synthèse par la modélisation sinusoïdale et recouvrement addition [Mna05].

I.3 Signal de la parole:

La parole est la forme vocalisée de la communication humaine. Chaque mot prononcé est créé à partir de la combinaison phonétique d'un ensemble limité de voyelles et de consonnes (unités sonores de la parole).

Le signal de la parole est un ensemble d'ondes acoustiques circulant dans l'air avec une forme aléatoire. Cette dernière est constituée d'un ensemble de fréquences (fréquence fondamentale, harmoniques) portant de l'information utile telle que l'identité du locuteur, l'accent, l'intention, l'âge, l'expression, style de discours, l'émotion et l'état de santé de l'orateur.

Les sons de la parole sont produits par les vibrations de la pression d'air générées en poussant l'air inspiré par les poumons à travers les cordes vocales vibrantes et le conduit vocal et à travers les voies respiratoires (la bouche et le nez). L'air expiré est modulé et mis en forme par les vibrations des cordes glottiques, la résonance du conduit vocal et des cavités nasales, la position de la langue et les ouvertures et fermetures de la bouche (Figure11).

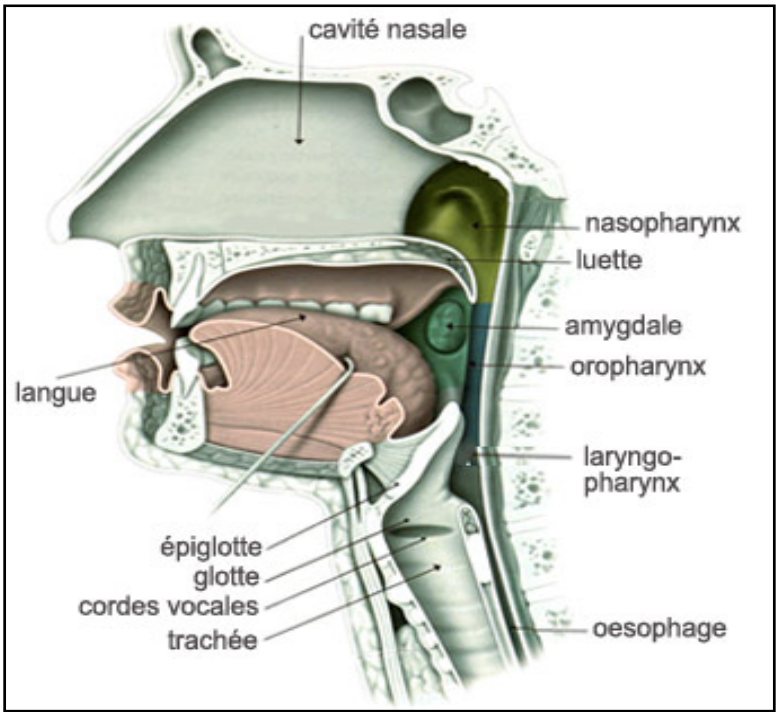


Figure11. L'appareil phonatoire humain.

Le signal de la parole est caractérisé par les paramètres suivants:

- L'intensité qui correspond à l'amplitude du son de la parole. Cette dernière permet de classifier un son soit faible ou fort.
- Le timbre, c'est le spectre de fréquences d'un signal de la parole (son riche, son profond,…).
- La hauteur correspond à la fréquence du signal. Cette dernière permet de distinguer les différentes notes d'un son (liée à la vitesse de vibration de l'air).
- La durée correspond au temps que dure une réalisation acoustique (courte durée, longue durée).

I.3.1 Son voisé:

Le son voisé est un signal quasi-périodique, caractérisé par une excitation périodique de la glotte. Le son voisé est produit par les vibrations périodiques des cordes vocales sur le conduit vocal. Les formants sont les fréquences de résonance des cavités vocales. La hauteur du son produit est relative à la fréquence de vibration des cordes vocales (la fréquence fondamentale).

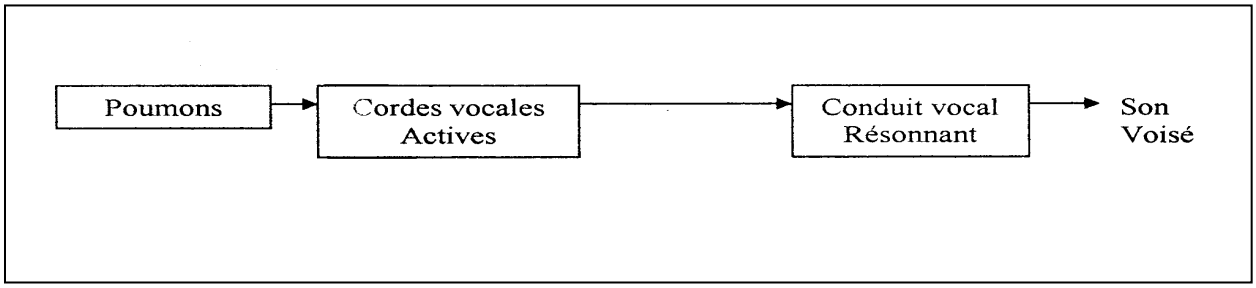


Figure12. Production d'un son voisé.

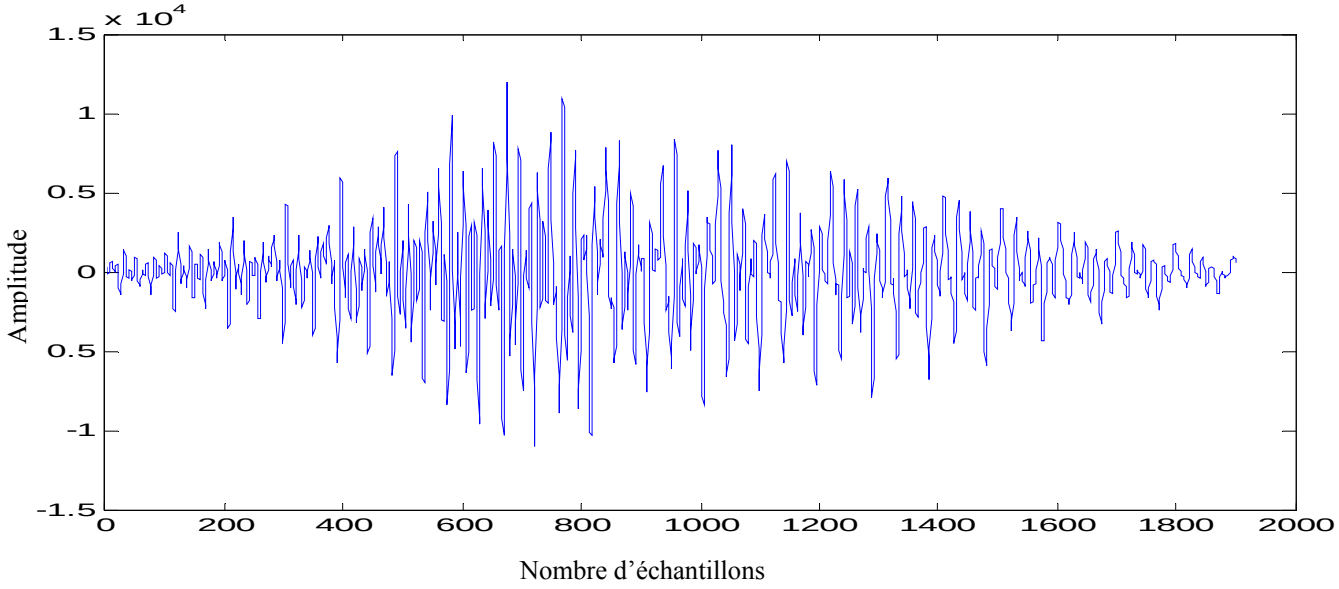


Figure13. Représentation temporelle d'un son voisé.

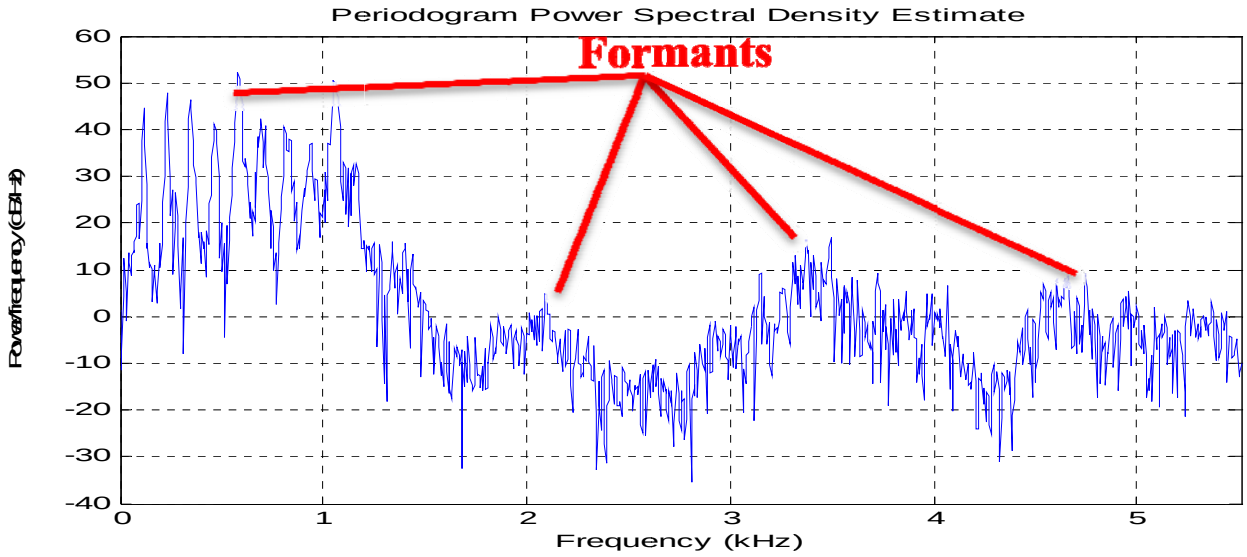


Figure14. Représentation fréquentielle d'un son voisé.

I.3.2 Son non voisé:

Un son non voisé est un signal qui ne présente pas de structure périodique (les cordes vocales sont relâchées). Ce dernier est généré en faisant passer le flux d'air à travers les différents organes de l'appareil phonatoire humain. Les perturbations de pression dues à ces mécanismes d'excitation fournissent une onde acoustique qui se propage le long du conduit vocal vers les lèvres.

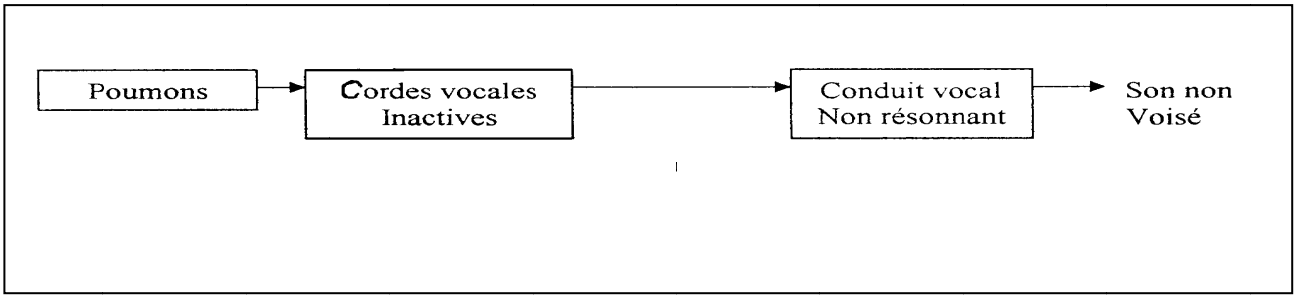


Figure15. Production d'un son non voisé.

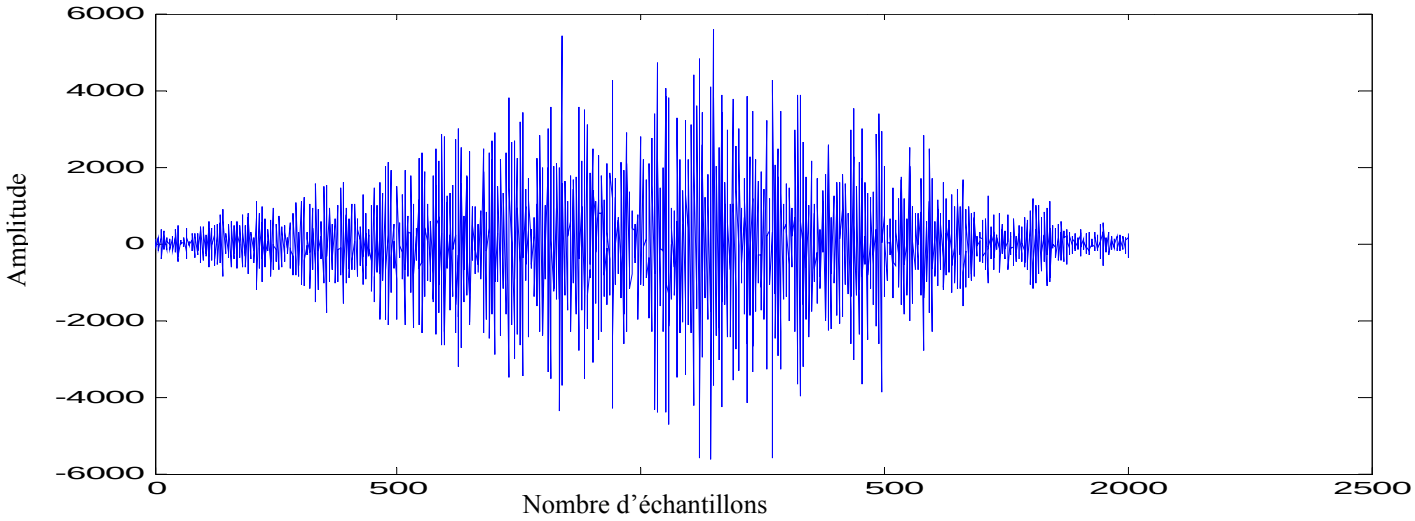


Figure16. Représentation temporelle d'un son non voisé.

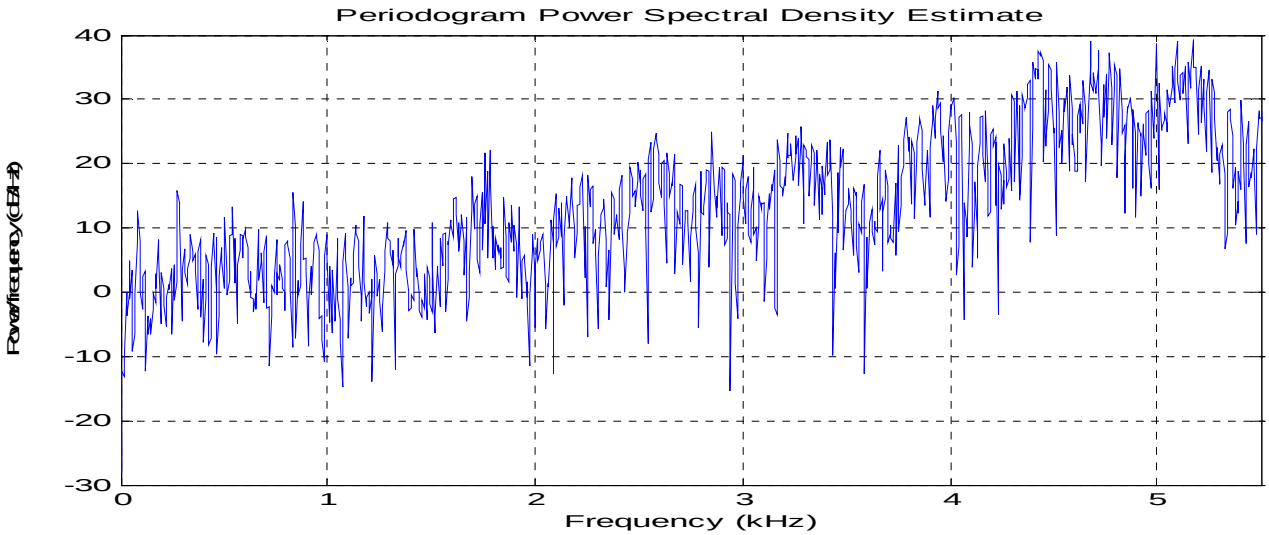


Figure17. Représentation fréquentielle d'un son non voisé (pas de structure formantique).

I.4 Les différentes techniques de la synthèse de la parole:

Il existe différentes techniques pour synthétiser de la parole artificielle. Ces techniques de synthèse ont chacune des avantages et des inconvénients. Le choix de l'utilisation d'une technique par rapport à l'autre repose sur les besoins des utilisateurs et le domaine d'application.

Les techniques de la synthèse de la parole sont les suivants:

- Synthèse par formants.
- Synthèse articulatoire.
- Synthèse par concaténation.
  - Synthèse par diphones.
  - Synthèse par sélection d'unités.
- Synthèse par modèle de Markov caché HMM**.**
- Synthèse par réseaux de neurones.

I.4.1 Synthèse par formants:

La synthèse vocale par formant (Figure18) appelée aussi synthèse par règles est basée essentiellement sur la modélisation paramétrique du spectre de la parole, exactement la modélisation des formants (mesurés en Hz) du signal de la parole [Div08]. La synthèse par formants est la plus ancienne méthode pour la synthèse de la parole. Cette dernière a dominé les implémentations de synthèse pendant une longue période.

La synthèse par formants consiste donc en la reconstruction artificielle des caractéristiques de formants à produire. Ceci est effectué en excitant un ensemble de résonateurs par une source de voisement ou par un générateur de bruit pour obtenir le spectre de parole désiré et en contrôlant la source d'excitation pour simuler soit le voisement ou soit le non voisement. L'ajout d'un ensemble d'anti-résonateurs permet en outre la simulation des effets de la voie nasale.

L'avantage de cette technique est que ces paramètres sont fortement corrélés avec la production et la propagation du son dans la voie orale. Le principal inconvénient de cette approche est que les techniques automatiques de spécification des paramètres de formants sont encore largement non satisfaisantes, et par conséquent, la majorité des paramètres doivent encore être optimisées manuellement.

La synthèse de la parole par formants n'utilise pas les sons de la parole humaine, mais s'appuie sur des règles établies par les linguistes pour générer les paramètres qui permettront la synthèse de la parole. Afin d'élaborer les règles, les linguistes ont étudié les spectrogrammes des sons de la parole et ont dérivé les règles d'évolution des formants.

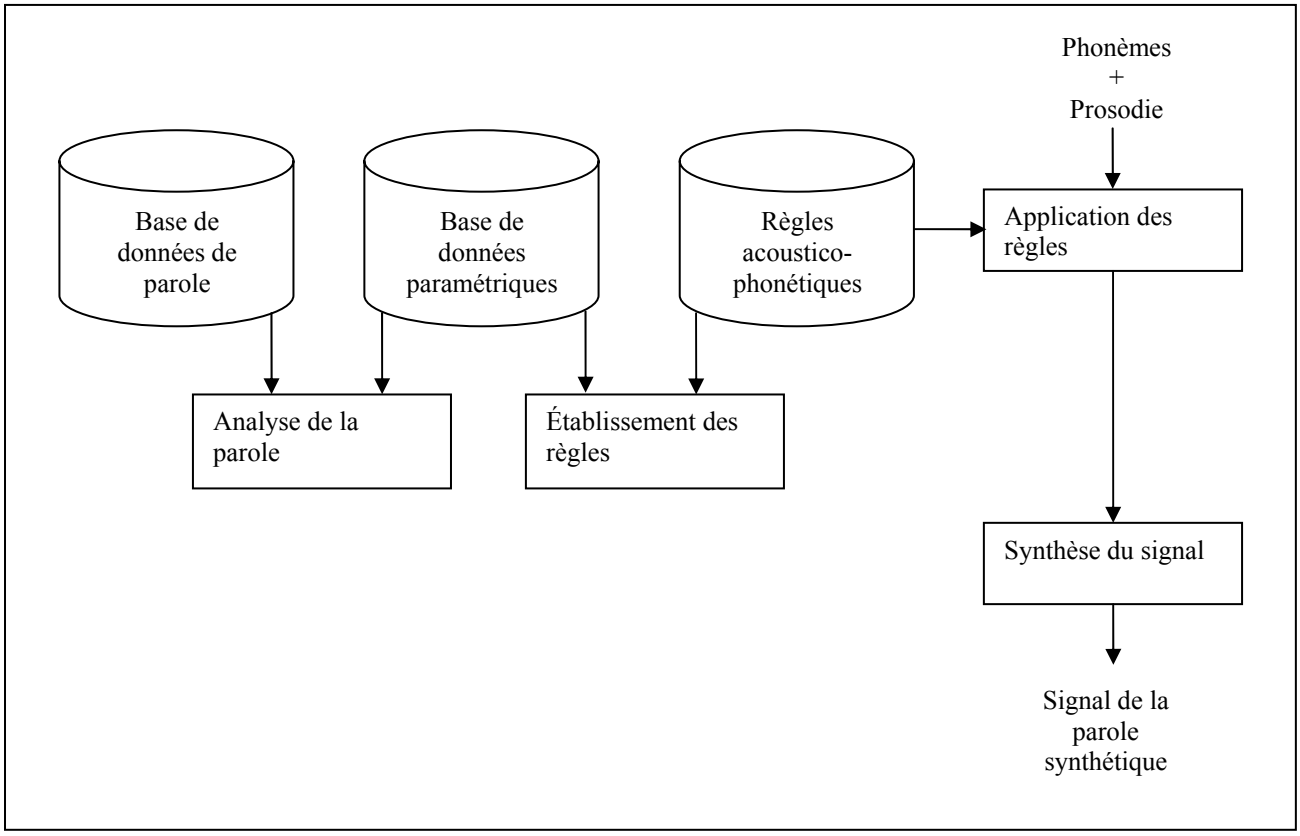


Figure18. Schéma représentatif d'un système de synthèse de la parole par formants.

Cette technique de synthèse de la parole conduit inévitablement à un résultat sonore peu naturel. Cependant, la parole synthétisée à base des règles est très intelligible. En outre, lorsque la mémoire et les coûts de traitement sont limités, comme dans les systèmes embarqués, ces synthétiseurs sont plus intéressants parce qu'ils n'ont pas besoin d'utiliser une base de données acoustique.

I.4.2 Synthèse articulatoire:

En comparaison avec les autres techniques de la synthèse de la parole, la synthèse articulatoire est la méthode la plus compliquée en ce qui concerne la structure du modèle et de la charge de calcul. L'idée de la synthèse articulatoire est de modéliser les mécanismes de production de la parole humaine le plus parfaitement possible [Kro92], (Figure19).

La mise en œuvre d'une telle technique de synthèse de la parole est très difficile, par conséquent cette méthode n'est pas encore largement utilisée. Les expériences avec des systèmes de synthèse articulatoire n'ont pas eu autant de succès que d'autres systèmes de synthèse, mais en théorie, la synthèse articulatoire donne le meilleur potentiel pour la parole synthétique de haute qualité. Par exemple, il est impossible d'utiliser la synthèse articulatoire pour produire des sons que les humains ne peuvent pas produire (en raison de la physiologie humaine). Dans d'autres techniques de synthèse de la parole, il est possible de produire de tels sons, et le problème est que ces sons sont habituellement perçus comme des effets secondaires indésirables.

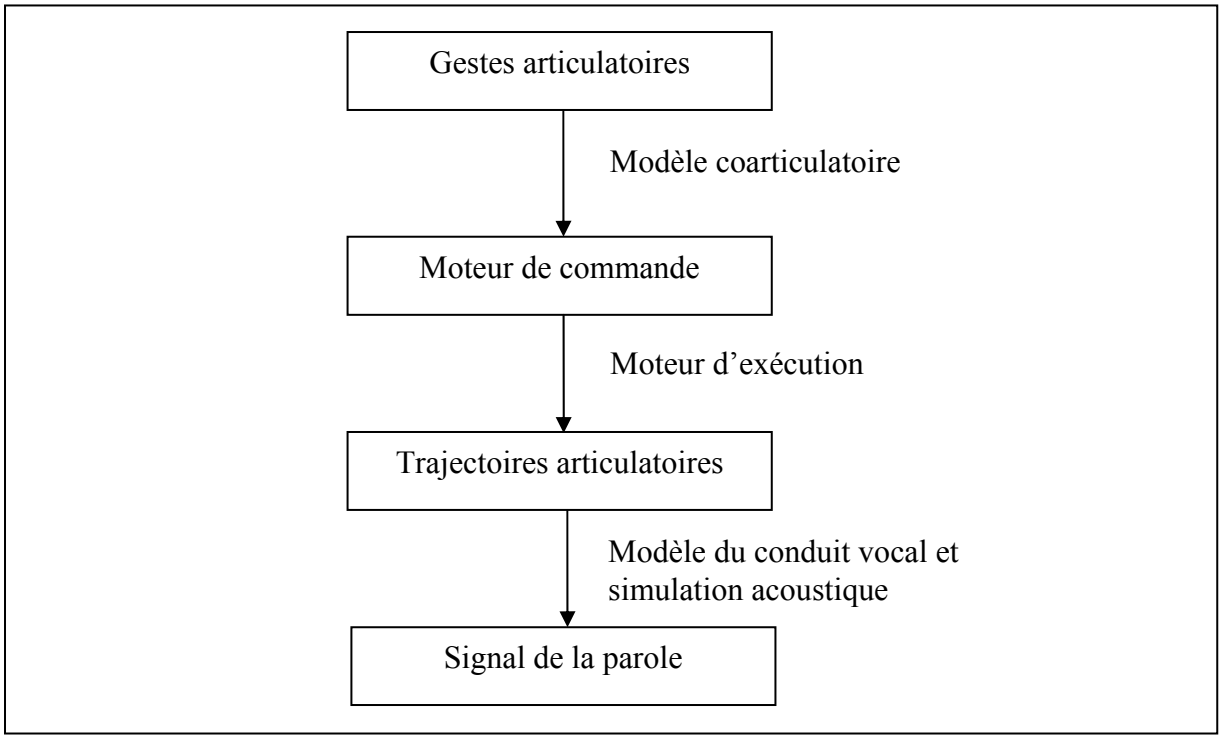


Figure19. Schéma de principe de la synthèse articulatoire.

Le système de la synthèse articulatoire contient à la fois les modèles physiques de l'appareil vocal humain et la physiologie des cordes vocales. Il est courant d'utiliser un ensemble de fonctions de la région pour modéliser la variation de la surface de section transversale du conduit vocal entre le larynx et les lèvres. Le modèle articulatoire comporte un grand nombre de paramètres de commande qui sont utilisés pour l'ajustement très détaillé de la position des lèvres et de la langue, de la pression des poumons, et de la tension des cordes vocales. Les données qui sont utilisées comme base de la modélisation sont habituellement obtenues à travers l'analyse aux rayons X de la parole naturelle.

I.4.3 Synthèse par concaténation:

Relier des segments de la parole naturelle préenregistrée est probablement la manière la plus facile de produire de la parole synthétique avec une intelligibilité et un naturel acceptable [Sud15]. Cependant, les synthétiseurs par concaténation sont généralement limités à une seule voix et nécessitent plus de capacité de mémoire que d'autres méthodes.

Un des aspects les plus importants dans la synthèse par concaténation est de trouver la longueur correcte de l'unité acoustique à concaténer. La sélection est habituellement un compromis entre des unités plus longues et plus courtes.

Avec des unités plus longues on aura plus de naturel dans la voix produite artificiellement, moins de points de concaténation et un bon contrôle de la coarticulation est atteint, mais la quantité d'unités acoustiques et la mémoire nécessaire sont plus importantes.

Avec des unités acoustiques plus courtes, moins de mémoire est nécessaire, mais la collecte des échantillons et les procédures d'étiquetage deviennent plus difficiles et complexes. Dans les systèmes actuels, les unités acoustiques utilisées sont généralement des mots, des syllabes, demi-syllabes, phonèmes, diphones, et parfois même triphones.

Il existe plusieurs problèmes dans la synthèse vocale par concaténation par rapport à d'autres méthodes, à savoir:

- Distorsion due aux discontinuités dans les points de concaténation, qui peuvent être réduits en utilisant les diphones ou certaines méthodes spéciales pour lisser le signal.
- Les besoins en mémoire sont généralement très élevés, surtout si les unités acoustiques utilisées lors de la concaténation sont longues, tels que des mots ou des syllabes.
- La collecte de données et l'étiquetage des échantillons de la parole prennent généralement beaucoup de temps, mais le compromis entre la qualité et le nombre d'échantillons doivent être effectués.

I.4.3.1 Synthèse par diphones:

La synthèse de la parole par diphones (Figure20) [Elb11] est l'une des méthodes les plus populaires, utilisée pour créer une voix synthétique à partir des enregistrements sonores ou des échantillons d'une voix humaine.

Les diphones sont définis pour s'étendre du point central de la partie stable du premier phonème jusqu'au point central de la partie stable du phonème suivant, de sorte qu'ils contiennent les transitions entre les phonèmes adjacents. Cela signifie que le point de concaténation sera dans la région de la zone la plus stable du signal, ce qui réduit la distorsion du signal de la parole aux points de concaténation. Un autre avantage des diphones est que l'effet de coarticulation n'a plus besoin d'être formulé sous forme de règles.

Dans la synthèse par concaténation de diphones, le plus grand défi est d'avoir la continuité dans le signal de la parole synthétisée. Pour éviter les distorsions audibles causées par les différences existantes entre les segments successifs, au moins la fréquence fondamentale et l'intensité des segments de paroles doivent être contrôlables. Enfin, la synthèse vocale par concaténation de diphones est affectée par le processus difficile de créer la base de données acoustique à partir de laquelle les unités de la parole seront sélectionnées.

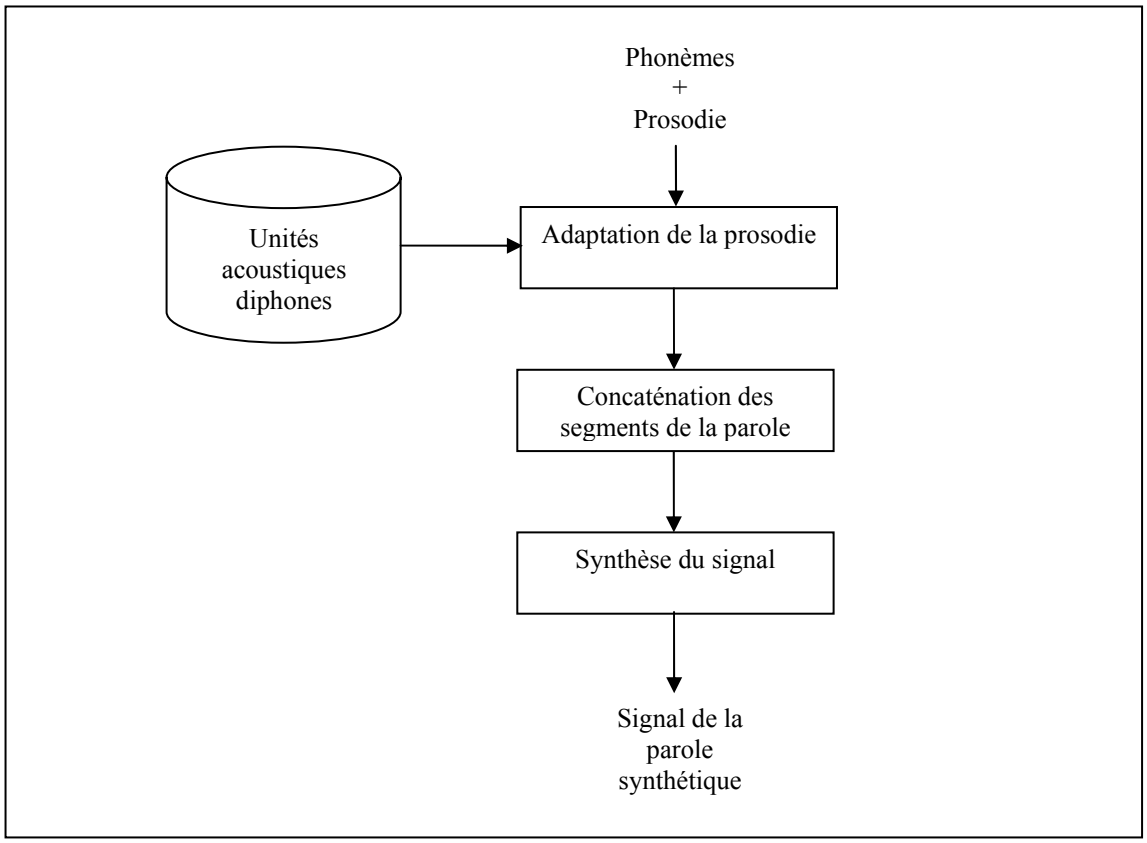


Figure20. Schéma de principe de la synthèse par diphones.

I.4.3.2 Synthèse par sélection d'unités:

La synthèse par sélection d'unités acoustiques (diphones, triphones, syllabes,…) est une technique de synthèse qui s'appuie sur l'utilisation d'une base de données acoustique contenant pour chaque unité acoustique stockée, plusieurs réalisations afin d'avoir différentes prosodies (Figure21). L'avantage de cette méthode réside dans la minimisation des discontinuités aperçues lors de la synthèse de la parole ainsi que les distorsions provoquées par le traitement du signal, mais cette amélioration à son coût qui est la taille de la base de données acoustique qui peut avoir une taille de plusieurs heures de parole.

Lors de la concaténation, la sélection des unités acoustiques se fait de la manière suivante: L'unité qui correspond à la prosodie, la plus proche de la cible est sélectionnée et concaténée, de sorte que les modifications prosodiques nécessaires sur l'unité sélectionnée soient minimisées. Afin de réaliser cette tâche de sélection, un algorithme de sélection d'unités est nécessaire pour choisir les unités qui répondent le mieux à la spécification de la cible. Cette sélection est basée sur la minimisation des deux types de fonctions de coûts, qui sont le coût de cible et le coût de joignement.

La fonction du coût de cible est une mesure des différences entre les caractéristiques de l'unité candidate et l'unité cible. La fonction du coût de joignement est une mesure des différences entre les caractéristiques de l'unité candidate et son unité voisine.

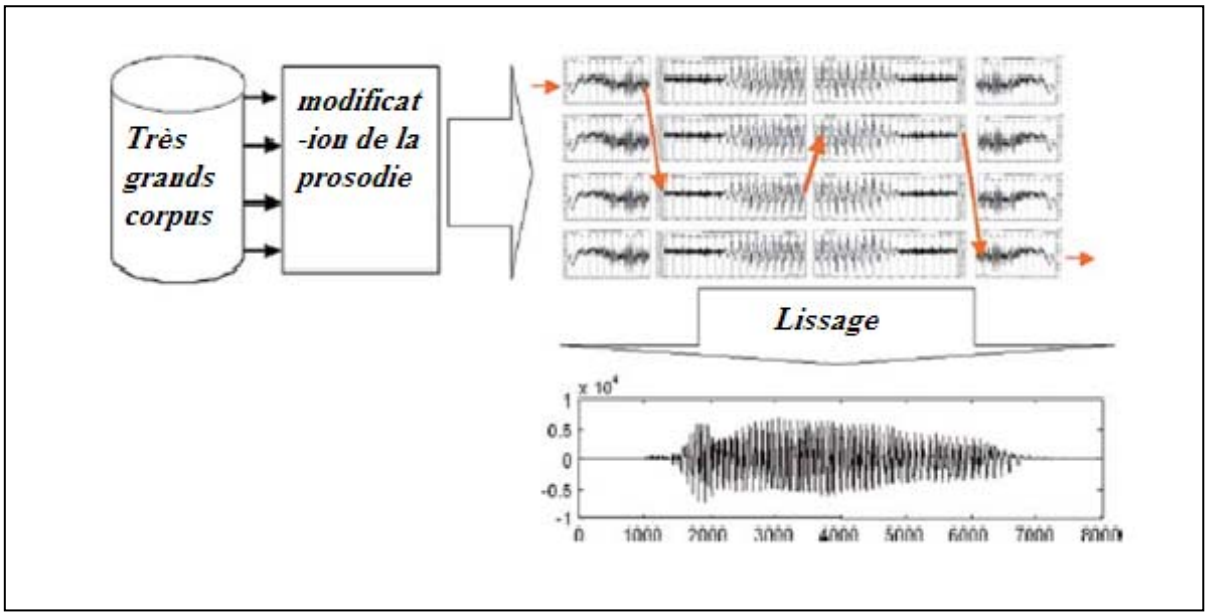


Figure21. Principe de la synthèse par sélection d'unités [Dut00*].

I.4.4 Synthèse par HMM:

Les Modèles de Markov cachés (HMM) ont été utilisés depuis les années 70 comme une approche réussie pour la reconnaissance de la parole. L'idée est que les trames de temps constants du signal acoustique codé en PCM (Pulse Code Modulation) sont converties en vecteurs caractéristiques représentant des caractéristiques spectrales les plus importantes du signal. Les séquences des vecteurs caractéristiques sont corrélées à des séquences de phonèmes mappés au texte sous-jacent par le biais d'un lexique de prononciation.

La synthèse par Modèle de Markov caché (HMM) [Tok02] est composée de deux phases principales: la phase d'apprentissage et la phase de synthèse. La phase d'apprentissage consiste à décider quelles caractéristiques des modèles devraient être formées pour les MFCC (Mel frequency cepstral coefficients) et leurs dérivées premières et secondes qui sont les types les plus communs des caractéristiques utilisées. Les caractéristiques sont extraites par trame et mis dans un vecteur caractéristique. L'algorithme de Baum-Welch est utilisé avec les vecteurs caractéristiques pour produire des modèles pour chaque phonème. Un modèle se compose généralement de trois états qui représentent le début, le milieu et la fin du phonème.

La phase de synthèse se compose de deux étapes: tout d'abord, les vecteurs caractéristiques pour une séquence donnée de phonème doivent être estimés. D'autre part, un filtre est mis en œuvre pour transformer les vecteurs caractéristiques en signaux audio.

La synthèse par Modèle de Markov caché (HMM) a le potentiel de produire un discours de haute qualité (Figure22). Ceci est principalement dû aux HMM qui produisent des contours spectraux continus, par opposition à la synthèse par concaténation où nous pourrions faire face à des incohérences au niveau des points de concaténation résultant des artefacts de discours audible.

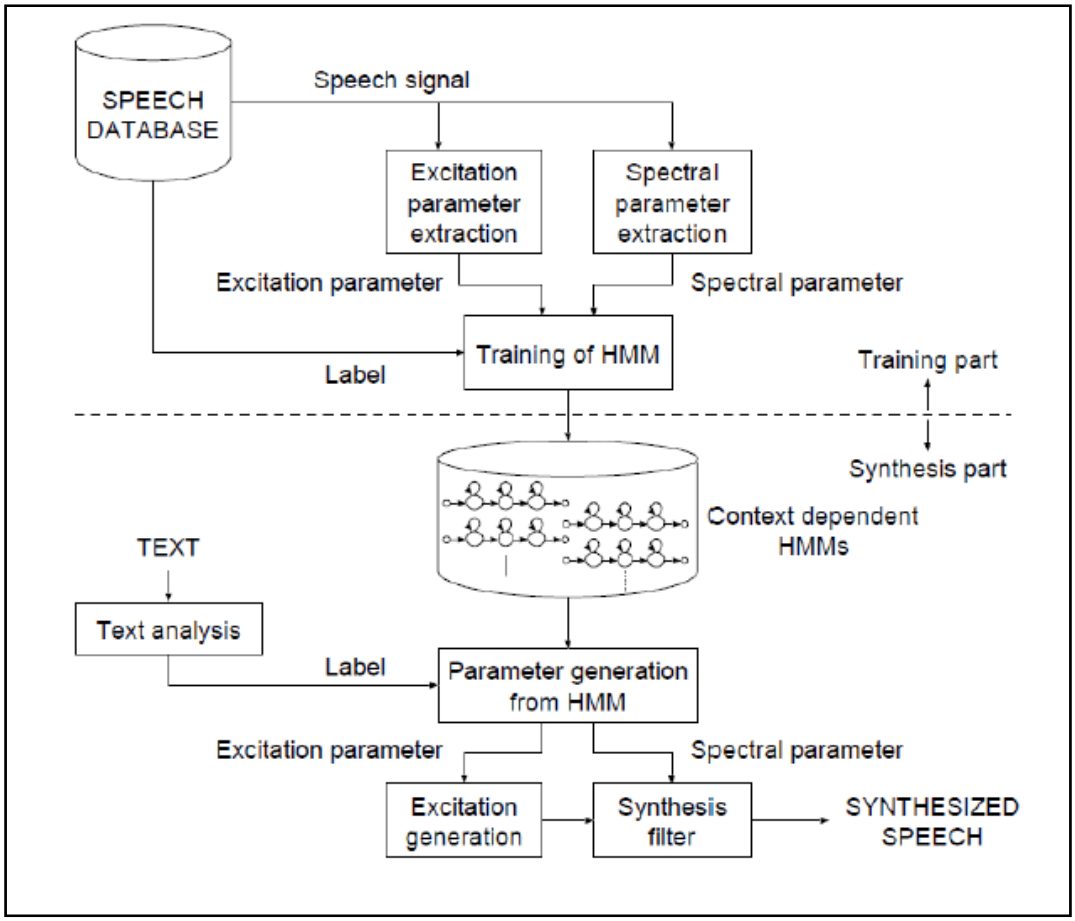


Figure22. Schéma représentatif de la synthèse par Modèle de Markov caché (HMM) [Tok02].

I.4.5 Synthèse par réseaux de neurones:

Les réseaux de neurones permettent à un synthétiseur de la parole de modéliser les effets de coarticulation dans une grande gamme de contextes, en utilisant un nombre limité de données [Als09]. Les modèles des réseaux de neurones artificiels (RNA) se composent de nœuds de traitement interconnectés, où chaque nœud représente un modèle de neurone artificiel, et l'interconnexion entre deux nœuds à un poids (représente l'efficacité d'une connexion synaptique) qui lui est associée.

Les modèles des réseaux de neurones artificiels avec différentes topologies effectuent différentes tâches de reconnaissance de formes.

Dans le cadre de la synthèse de la parole (Figure23), un mappage est nécessaire du texte (espace linguistique) vers la parole (espace acoustique). Ainsi, l'exploitation des capacités du mappage de motif du modèle RNA effectue le mappage complexe et non linéaire de l'espace linguistique vers l'espace acoustique pour générer de la parole synthétique.

L'approche des réseaux neurones employée à la synthèse de la parole offre les avantages de la portabilité du langage, un son proche de la parole naturelle, et respecte l'exigence de stockage faible. Les résultats des expériences d'acceptabilité indiquent que les systèmes de synthèse de la parole basé sur les réseaux de neurones ont le potentiel de fournir une meilleure qualité de voix que les approches traditionnelles, mais une certaine amélioration dans le système est toujours souhaitable.

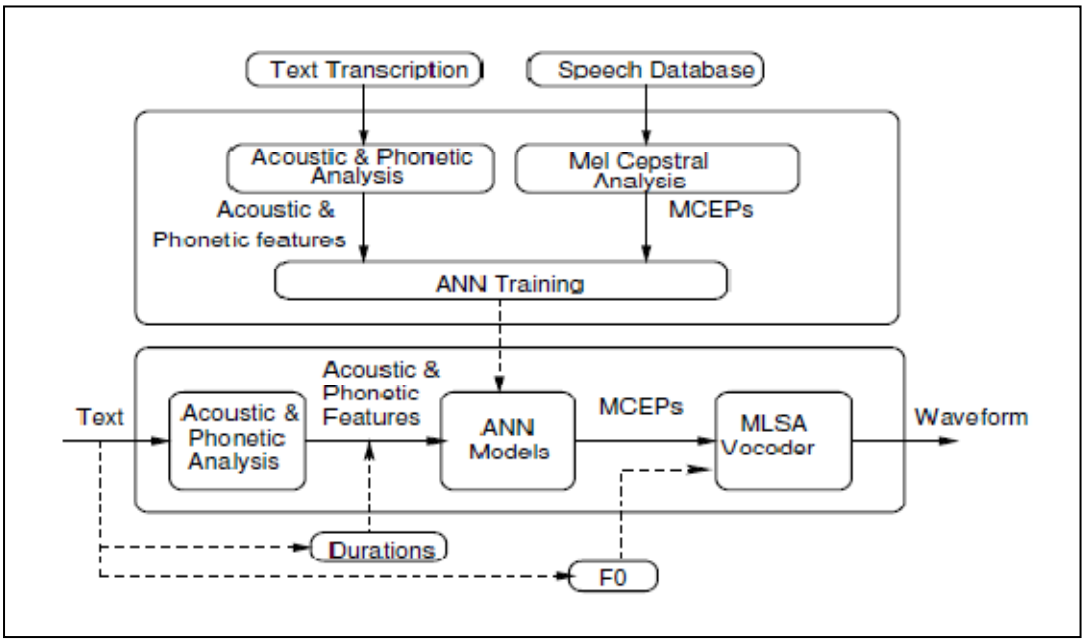


Figure23. Principe de la synthèse par réseaux de neurones [Vee10].

I.5 Domaines d'application:

Les domaines d'application de la synthèse de la parole sont très nombreux tels que les applications destinées au domaine des télécommunications, aide aux personnes handicapées, des applications pour l'apprentissage des langues, ainsi que la communication homme-machine,…etc.

La voix artificielle est intégrée dans plusieurs domaines d'application mais la méthode de la mise en œuvre dépend essentiellement de l'application utilisée, Dans certains cas, tels que les systèmes d'annonce ou d'avertissement, le vocabulaire illimité n'est pas nécessaire et le meilleur résultat est généralement obtenu avec un système de messagerie simple. D'autre part, certaines applications, telles que les machines de lecture pour les personnes non-voyantes ou la lecture des courriers électroniques, exigent un vocabulaire plus riche et un système de synthèse de la parole à partir du texte (TTS) est nécessaire.

I.5.1 Services de télécommunications et multimédias:

Les nouvelles applications en synthèse de la parole sont dans le domaine des télécommunications. Le signal de la parole synthétisée a été utilisé pendant des décennies dans tous les types de systèmes de renseignements téléphoniques, mais la qualité était loin d'être satisfaisante pour les clients. Aujourd'hui, la qualité a atteint un niveau tel que les clients adoptent cette solution pour un usage quotidien.

Avec la parole synthétique les messages des e-mails peuvent être écoutés via la ligne téléphonique normale. La parole synthétisée peut également être utilisé pour prononcer le texte du service de messages courts (SMS) dans les téléphones mobiles.

Pour des applications multimédias totalement interactives, un système de reconnaissance automatique de la parole est également nécessaire. La reconnaissance automatique du langage parlé est encore loin, mais la qualité des systèmes actuels est assez bonne pour qu'ils puissent être utilisés pour donner quelques instructions de commande.

I.5.2 Aides aux personnes sourdes et muettes:

Les gens qui sont nés sourds ne peuvent pas apprendre à parler correctement et les personnes ayant des difficultés d'audition ont également des difficultés pour parler. La parole synthétisée donne aux gens sourds et muets la possibilité de communiquer avec des gens qui ne comprennent pas la langue des signes.

Avec le clavier, il est généralement beaucoup plus lent de communiquer qu'avec la parole. Une façon d'accélérer ceci est d'utiliser le système de saisie prédictive qui affiche toujours le mot le plus fréquent pour tout fragment de mot saisi, et l'utilisateur peut alors appuyer sur une touche spéciale pour accepter la prédiction. De même différentes expressions pré-composées, telles que des salutations, peuvent être employées.

I.5.3 Aides aux personnes non-voyantes:

Probablement le domaine d'application le plus important et utile dans la synthèse de la parole est l'application destinée au aide à la lecture et de la communication pour les personnes non-voyantes.

La première application commerciale des synthétiseurs de la parole à partir du texte était la machine de lecture pour les aveugles introduite par Raymond Kurzweil vers la fin des années 70. Elle se composait d'un scanner optique et d'un logiciel de reconnaissance du texte. Cette machine était capable de produire un discours tout à fait intelligible à partir du texte.

Les prix des premières machines de lecture étaient beaucoup trop élevés pour l'utilisateur ordinaire, et ces machines ont été utilisées principalement dans les bibliothèques ou les lieux liés à cette application. Aujourd'hui, la qualité des machines de lecture a atteint le niveau acceptable et les prix sont devenus abordables pour les individus, donc le synthétiseur de la parole sera très utile et un dispositif essentiel chez les personnes non-voyantes à l'avenir. Peu importe à quelle vitesse le développement des machines destinées aux aides à la lecture et de la communication, il ya toujours des améliorations à faire.

I.5.4 Apprentissage des langues:

Le signal de parole synthétique peut être utilisé dans de nombreuses situations d'enseignement. Un ordinateur muni d'un synthétiseur de la parole peut être utilisé pour enseigner sans arrêt (24 heures par jour et 360 jours par année). Il peut être programmé pour des tâches spéciales comme l'orthographe et l'enseignement de la prononciation pour différentes langues. Elle peut également être utilisée avec des applications éducatives interactives.

Particulièrement avec des personnes ayant une déficience à lire (les dyslexiques), la synthèse de la parole peut être très utile, car en particulier certains enfants peuvent se sentir très embarrassés quand ils doivent être aidés par un professeur. Il est également presque impossible d'apprendre l'écriture et la lecture sans le son audio. Avec un logiciel approprié, un apprentissage non supervisé peut résoudre ces problèmes (facile et peu coûteux).

Un synthétiseur de la parole relié à l'unité de traitement de texte est également une aide utile à l'épreuve de la lecture. Beaucoup d'utilisateurs trouvent plus facile de détecter les problèmes grammaticaux et stylistiques lors de l'écoute de la lecture.

I.5.5 Communication homme-machine:

En principe, la synthèse de la parole peut être utilisée dans toutes sortes d'interactions homme-machine. Par exemple, les systèmes d'avertissement et d'alarme qui utilisent de la parole synthétisée, peuvent donner une information plus précise sur la situation actuelle.

Le système de communication homme-machine est un domaine d'application très intéressant pour la synthèse de la parole, vu l'avancement technologique et la nécessité d'intégrer de la parole humaine dans les systèmes qui interagissent avec l'être humain. Aujourd'hui la synthèse de la parole est utilisée dans divers secteur d'industrie surtout dans le secteur d'automobile et dans le développement des robots parlants.

I.5.6 Livres et jouets parlants:

Les livres et jouets parlants sont des domaines qui ont beaucoup bénéficié de la technologie de la synthèse de la parole. Puisque de nombreux livres et jouets pour enfants actuellement sont munis d'une sortie vocale à cause de la voix humaine, qui est le moyen de communication le plus simple chez les enfants.

I.6 Logiciels de la synthèse de la parole (TTS):

Les logiciels développés pour la synthèse de la parole à partir du texte (TTS) sont très nombreux vu le nombre de langues utilisées dans le monde entier. Chaque logiciel a ses propres caractéristiques, selon la méthode de la synthèse de la parole et selon les spécificités de la langue traitée.

La qualité du signal de la parole synthétisée varie d'un système à un autre selon les exigences de l'application à réaliser, soit du coté intelligibilité ou soit du côté naturel de la voix synthétisée. Dans ce qui suit nous allons citer quelques logiciels développés pour les langues étrangères ainsi que spécifiquement pour la langue Arabe.

I.6.1 Pour les langues étrangères:

En ce qui concerne les logiciels de la synthèse de la parole à partir du texte traitant les langues étrangères sont données en détails dans ce qui suit:

- **AT&T Natural Voices™:** C'est un logiciel développé par le laboratoire de recherche AT&T consacré à faire avancer la science et la technologie des communications et de l'information pour créer des services innovants fondés sur ces progrès. Ce logiciel inclut quatre langues: l'Anglais, Français, Allemand et Espagnol. La base de données acoustique utilisée dans cette application TTS est une base de données de la parole de haute qualité enregistrée dans des conditions optimales avec un équipement d'enregistrement de haute qualité. Les plus petites unités acoustiques dans la parole (appelées phonèmes) sont soigneusement étiquetées de sorte que lorsque un nouveau mot ou une phrase est nécessaire, les algorithmes peuvent sélectionner le meilleur jeu de sons à extraire de la base de données, afin de les raccorder pour produire de la parole synthétique. Pour savoir comment le faire efficacement c'est difficile. Beaucoup de leur recherche est consacrée à l'amélioration de ces algorithmes pour réaliser encore plus à l'avenir des synthétiseurs de la parole à partir du texte avec une voix plus proche de la voix naturelle.

- **Acapela group:** Le système TTS réaliser par Acapela group englobe une trentaine de langues à la fois tel que l'anglais, le Français, l'Italien, …etc. Afin de reproduire un son naturel de chaque langue, un narrateur enregistre une série de textes (la poésie, actualités politiques, résultats sportifs, des mises à jour de la bourse, …etc), ces enregistrements contiennent tous les sons possibles dans la langue choisie. Ces enregistrements sont alors découpés et organisés dans le but de créer la base de données acoustique. Pendant la création de la base de données acoustique, toute la parole enregistrée est segmentée sous forme de: diphones, syllabes, morphèmes, mots, expressions, et phrases. Pour générer de la parole artificielle des mots d'un texte donné, le système de la synthèse de la parole commence par effectuer une analyse linguistique sophistiquée qui transpose le texte écrit vers le texte phonétique. Une analyse grammaticale et syntaxique permet alors au système de définir comment prononcer chaque mot afin de reconstruire le sens. Avec l'intégration de la prosodie (elle donne le rythme et l'intonation d'une phrase). En conclusion, le système produit des informations associant l'écriture phonétique avec le ton et la longueur requise de la prononciation. Le son est généré en sélectionnant les meilleures unités stockées dans la base de données acoustique.

- **eSpeak:** C'est un logiciel open source de la synthèse de la parole à partir du texte, développé pour la langue Anglaise et d'autres langues, ce logiciel peut fonctionner dans les deux plateformes Linux et Windows. eSpeak utilise la méthode de la synthèse de la parole par formants. Ceci permet à beaucoup de langues d'être fournies dans une petite taille. La parole produite par ce système de TTS est claire, et peut être utilisé à des vitesses élevées, mais n'est pas aussi naturel ou lisse que les grands synthétiseurs qui sont basées sur des enregistrements de la parole humaine. Le logiciel eSpeak comporte plusieurs caractéristiques, à savoir:

    - Comprend des voix différentes, dont les caractéristiques peuvent être modifiées.
    - Peut produire une sortie de la parole comme un fichier WAV.
    - Le programme et ses données, y compris de nombreuses langues, s'élève à environ 2 Mo.
    - Peut traduire le texte en séquences de phonèmes, de sorte qu'il pourrait être utilisé comme une entrée pour une autre machine de synthèse de la parole.
    - Les outils de développement sont disponibles pour la production et le réglage des données de phonèmes.
    - Le langage de programmation utilisé est le langage C.

- **Festival:** Le système de synthèse de la parole Festival est un système multilingue de synthèse vocale développé par Alan W. Black au centre de recherche de technologie de la parole (CSTR) à l'Université d'Edimbourg. Le centre de recherche CSTR s'intéresse à la recherche dans tous les domaines de la technologie de la parole, y compris la reconnaissance de la parole, synthèse de la parole, traitement du signal de la parole, accès à l'information, interfaces multimodales et systèmes de dialogue. Ce laboratoire de recherche a beaucoup de collaborations avec une large communauté de chercheurs en sciences de la parole, du langage, de la cognition et l'apprentissage de la machine pour laquelle Edimbourg est célèbre. Le système de synthèse de la parole Festival est un logiciel libre, écrit en C ++. Ce logiciel utilise la bibliothèque Edinburgh outils vocale (Edinburgh Speech Tools Library) pour l'architecture de bas niveau.

- **MBROLA:** Le projet MBROLA est initié par le laboratoire de théorie des circuits et traitement du signal (TCTS) de la Faculté Polytechnique de Mons (Belgique). L'objectif de ce projet est d'obtenir un ensemble de synthétiseurs vocaux pour autant de langues que possible, et de les fournir gratuitement pour des applications non-commerciales et non militaires seulement. Le but principal est de renforcer la recherche universitaire sur la synthèse de la parole, et plus particulièrement sur la génération de la prosodie, connu comme l'un des plus grands défis relevés par les systèmes de synthèse de la parole à partir du texte pour les années à venir. MBROLA est un synthétiseur de la parole basé sur la concaténation des diphones. Il prend une liste de phonèmes en entrée, ainsi que l'information prosodique (durée des phonèmes et la variation de la fréquence fondamentale), afin de produire de la parole synthétique à la fréquence d'échantillonnage de la base de données de diphones utilisé. Le synthétiseur de la parole MBROLA n'est pas considéré comme un système de synthèse de la parole à partir du texte (TTS), car il n'accepte pas du texte orthographique en entrée. Donc il faut le fusionner avec un bloc des traitements linguistiques qui génère les séquences phonétiques afin que le synthétiseur de la parole MBROLA devient un système TTS. La base de données acoustique de diphones conçue pour une langue donnée devrait être adaptée au format MBROLA. Cette étape est nécessaire pour exécuter le synthétiseur de la parole. Le projet MBROLA lui-même a été organisé afin d'inciter d'autres laboratoires de recherche ou les entreprises à partager leurs bases de données de diphones.

I.6.2 Pour la langue Arabe:

Les synthétiseurs de la parole à partir du texte développé spécifiquement que pour la langue Arabe sont très peu et rares. Dans ce qui suit nous allons citer quelques systèmes TTS pour la langue Arabe:

- **SAKHR TTS:** La société SAKHR a été initialement créée comme une division d'Al Alamiah Electronique en 1982 avec la vision ambitieuse pour apporter un soutien à la langue Arabe dans le domaine de la technologie de l'information. La compagnie de logiciels SAKHR est devenue rapidement le leader du marché des technologies et des solutions avancé de la langue Arabe. Avec plus de 28 années de recherche et de développement en informatique pour des applications destinées à la langue Arabe. La compagnie de logiciels SAKHR fournit plusieurs solutions pour la langue Arabe, y-compris: Traduction automatique, Reconnaissance optique de caractères, Technologie de la parole, Gestion des connaissances, Services de recherche avancée, Traduction et localisation professionnelle. SAKHR TTS convertit le texte Arabe en voix humaine naturelle. Les caractéristiques du logiciel SAKHR TTS sont les suivantes:

  - La voix Arabe la plus naturelle et la plus intelligible sur le marché.
  - Système TTS intégré pour les appareils mobiles.
  - Diacritisation du texte Arabe puissante basée sur des règles avec 97% de précision.
  - Normalise le texte ambigu tel que: les dates, heures, devises, abréviations.
  - Vocabulaire illimité, taille du texte illimité, entrée phonétique et prosodique.

- **ArabTalk:** C'est un logiciel de la synthèse de la parole à partir du texte développé à RDI (Recherche et développement international). Ce synthétiseur utilise la synthèse hybride concaténative/paramétrique pour produire de la voix artificielle. Les modèles basés sur les réseaux de neurones artificiels (RNA) sont employés pour le traitement de la durée, l'énergie, et la prédiction de la fréquence fondamentale. En outre, il y a une synthèse en temps réel par un algorithme de sélection qui explore un large corpus de parole. Un bloc de traitement phonologique a été développé et de nombreux modèles à base de règles ont été utilisé dans la procédure de la conversion lettre vers le son. La recherche effectuée au sein du RDI vise à faire progresser le processus de développement de la synthèse de la parole à partir du texte pour la langue Arabe de haute qualité, ce qui donne un son de voix Arabe naturel et humain.

- **kacst_atts2:** La version disponible sur le web est une version démo et elle est téléchargeable gratuitement. Ce logiciel de la synthèse de la parole à partir du texte est développé au sein de l'université du Roi Fahd pour le pétrole et les minéraux (Cité du Roi Abdulaziz pour la science et de la technologie). Le logiciel kacst_atts2 est le premier synthétiseur de la parole à partir du texte destiné à la langue Arabe, qui est disponible pour les chercheurs et prêt pour une application dans différents domaines tels que la communication homme-machine et la lecture de divers textes de la langue Arabe. Le système est basé sur une synthèse de la parole par concaténation. Deux types d'unités de parole ont été utilisés de façon indépendante. La première est basée sur des unités de parole de type diphones, alors que la deuxième est basée sur les allophones. Un modèle paramétrique a été construit pour synthétiser de la parole et

donner à l'utilisateur un contrôle sur la fréquence fondamentale et le tempo de la parole synthétisée. Ce système TTS nécessite en entrée un texte entièrement voyellé, alors que presque tous les textes de la langue Arabe sont aujourd'hui non voyellé, un système de vocalisation automatique a été développé afin de résoudre ce problème. Le signal de la parole synthétisée est intelligible. Cependant, le système nécessite encore des recherches et des développements dans de nombreuses directions, y compris la manipulation de divers textes de la langue Arabe tels que les dates, les abréviations, les noms étrangers, les numéros et de la manipulation prosodique d'une manière naturelle.

I.7 Conclusion:

Les généralités sur la synthèse de la parole à partir du texte présentées dans ce chapitre nous ont permis de mieux situer et comprendre le processus de fonctionnement d'un synthétiseur de la parole à partir du texte à divers plans soit sur l'aspect linguistique ou soit sur l'aspect acoustique.

Nous avons donné dans ce chapitre une description générale des différentes étapes suivies pour aboutir à un système qui transforme du texte orthographique à un signal de la parole intelligible. Un bref aperçu sur les différentes techniques de la synthèse de la parole a été fourni dans ce chapitre ainsi que le domaine d'application et les différents logiciels de la synthèse de la parole à partir du texte (TTS) développés pour les langues étrangères et spécifiquement pour la langue Arabe.

Chapitre II: La transcription phonétique du texte Arabe

II.1 Introduction:

La phonétisation automatique du texte est un élément indispensable pour la réalisation d'un synthétiseur de la parole à partir du texte. Celle-ci est intimement liée aux caractéristiques de la langue considérée et permet d'établir une correspondance entre les graphèmes constituant le texte et une séquence de phonèmes qui servira à produire la prononciation du texte [Zek10].

Dans ce chapitre nous allons traiter la partie linguistique du synthétiseur de la parole qui a pour but de générer les phonèmes correspondant au texte édité, cette dernière n'est pas une tache facile à réaliser. La difficulté des traitements linguistiques varie selon la langue à traiter, par exemple la langue Arabe est une langue moins difficile à transcrire phonétiquement que d'autres langues telles que les langues européennes [Bra06], [Ste09], du fait que la forme orthographique étant proche de la forme phonétique, et du fait qu'on peut mettre des règles de prononciation pour chaque graphème de la langue à l'exception de quelques mots.

La connaissance de la langue Arabe constitue une grande partie du travail pour la réalisation d'un système de phonétisation automatique dédié à la synthèse de la parole à partir du texte. En effet, la phonétisation ne peut se faire sans un travail d'analyse, de compréhension et de modélisation de la langue appropriée.

La phonétisation du texte est une étape essentielle et fondamentale dans un système de synthèse de la parole à partir du texte. Plusieurs travaux ont été réalisés dans le domaine de la phonétisation automatique du texte Arabe, nous citons : les travaux réalisés par Salem Ghazali [Gha90], M. Elshafei Ahmed [Els91], Mansour M. Al-ghamdi [Alg02], [Alg04], Yousif A. El-imam [Eli89], [Eli04], Tahar Saidane [Saï04], [Saï05], Zouhir Zemirli [Zem96], [Zem06], Othman. O. Khalifa [Kha11] .

Ces derniers utilisent un ensemble de règles de transcription graphème-phonème, qui permettent d'associer à chaque graphème du texte édité, un ou plusieurs phonèmes en prenant en compte le contexte gauche (caractère précédant le graphème à transcrire) et le contexte droit (caractère suivant le graphème à transcrire) systématiquement. Ces règles sont organisées de façon hiérarchique. Ainsi que l'utilisation d'un dictionnaire des exceptions afin de transcrire les mots qui ne sont pas pris en charge par les règles de transcription graphème-phonème.

II.2 Etude la langue Arabe:

La langue Arabe est la langue maternelle de plus de 200 millions de personnes dans le monde Arabe, et elle est la langue officielle dans 22 pays. Avec la migration de ressortissants Arabes vers les pays hors du monde Arabe, la langue Arabe s'est étendue à presque tous les coins de la terre. Étant la langue du Coran, la langue Arabe est très respectée dans le monde musulman (plus d'un milliard de musulmans dans le monde entier). Beaucoup d'enfants musulmans non-Arabes commencent à apprendre l'Arabe à un âge précoce, pour leur permettre de lire et comprendre le Coran.

L'alphabet de la langue Arabe se compose de 28 lettres qui sont toutes des consonnes (Figure24), bien que trois d'entre elles s'emploient aussi comme des voyelles longues (ا و ي). Nous considérons pour notre part que l'alphabet Arabe compte 28 consonnes et 6 voyelles (3 courtes « ـَ ـُ ـِ » et 3 longues « ـَا ـُو ـِي ») et quelques signes diacritiques (ـً ـٌ ـٍ ـّ ـْ).

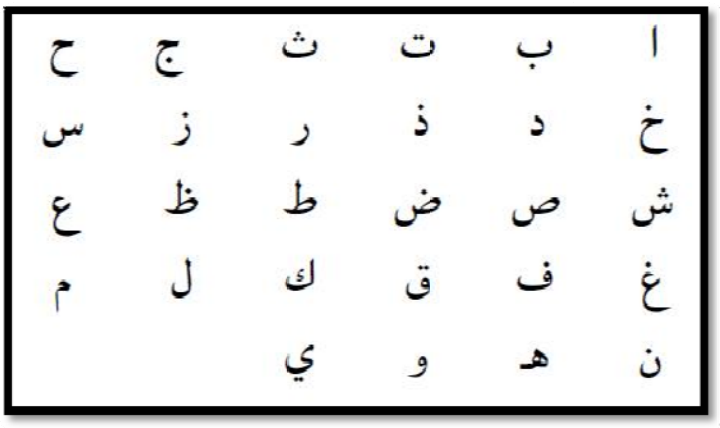


Figure24. L'alphabet de la langue Arabe.

II.2.1 Caractéristiques de la langue Arabe:

La langue Arabe s'écrit et se lit de droite à gauche. Les lettres Arabes changent de forme de présentation selon leur position dans le mot (Table4).

A la fin du mot, d'une lettre non joignable	A la fin du mot, d'une lettre joignable	Au milieu du mot	Au début du mot
غ	غ	ـغـ	غـ

Table4. Les différentes variations de la lettre غ dans un mot.

Un mot Arabe s'écrit avec des consonnes et des voyelles. Les voyelles sont ajoutées au dessus ou au dessous des consonnes (Figure25). L'absence des voyelles génère une certaine ambiguïté à deux niveaux:

– Sens du mot.
– Difficulté à identifier la fonction du mot dans la phrase.

Sept des lettres Arabes s'attachent uniquement aux lettres précédentes, mais pas aux lettres suivantes. Ces lettres sont les suivantes: ا د ذ ر ز و لا, pour exemple le mot 'نزل'.

Les voyelles longues correspondent seulement à une prolongation de la durée de la voyelle courte, une fois prononcées, et qui sont les suivantes:

– La voyelle courte fatha « ـَ » /a/ est prolongée lorsque elle est suivie par la consonne alif « ا », comme dans 'رَايَةٌ' /ra:jatun/ (drapeau).

– La voyelle courte damma « ـُ » /u/ est prolongée lorsque elle est suivie par la consonne waw « و » /w/, comme dans 'رُوحٌ' /ru:Xun/ (esprit).

– La voyelle courte kasra « ـِ » /i/ est prolongée lorsque elle est suivie par la consonne ya «ي» /j/, comme dans 'كَرِيمٌ' /kari:mun/ (généreux).

Les autres signes diacritiques sont les suivants :

– Le sekun « ـْ » indique que la consonne n'est pas munie de voyelle (exemple : كُنْ).

– Le chadda « ـّ » indique le redoublement de la consonne, bien qu'elle soit écrite seulement une fois (exemple : مَدَّ). Elle s'emploi uniquement dans les voyelles « ـَ ـُ ـِ » mais jamais avec le sekun « ـْ ».

– Le doublement de voyelle s'appelle tanwin: Fathatan « ـً » prononcé: /an/, Dammatan « ـٌ » prononcé: /un/, Kasratan « ـٍ » prononcé: /in/.

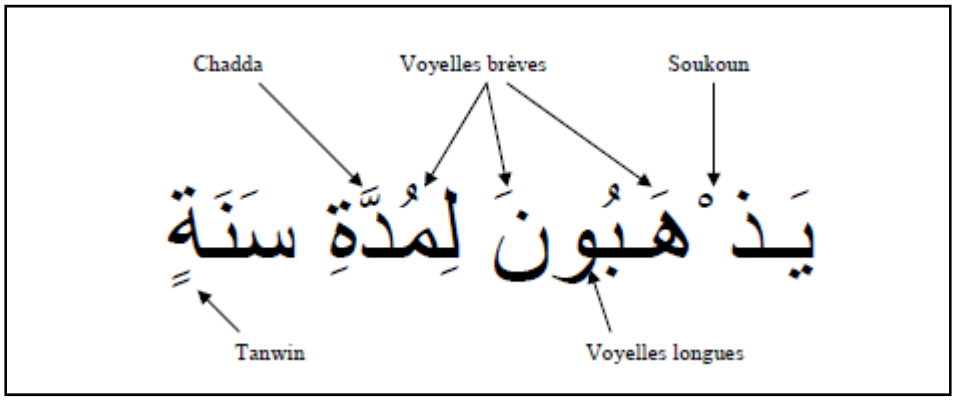


Figure25. Les différents signes diacritiques de la langue Arabe.

Lorsque un mot commence par l'article «ال» /al/, la consonne qui le suit immédiatement peut être classifiée comme lunaire ou solaire:

– Les lettres lunaires initiales d'un nom n'assimilent pas l'article qui les précède et par conséquent ne reçoivent pas le chadda. Ce sont: ا ب ج ح خ ع غ ف ق ك م ه و ي

Exemple: اَلْقَمَرُ ⟶ **la lettre lam « ل » /l/ est prononcée.**

– Les lettres solaires initiales d'un nom assimilent l'article qui les précède et reçoivent ainsi le chadda. Ce sont: ت ث د ذ ر ز س ش ص ض ط ظ ل ن

Exemple: اَلشَّمْسُ ⟶ **la lettre lam « ل » /l/ est muette.**

II.2.2 Prononciation et description des consonnes Arabes:

Dans ce qui suit nous allons décrire les différentes consonnes de la langue Arabe selon leur prononciation et selon leur forme d'écriture (voir Table5).

- **Hamza « ء » :** Le signe de la hamza « ء » /?/ a été ajouté à l'écriture de la langue Arabe à un stade assez tard. Par conséquent la Hamza n'a pas de forme réelle indépendante comparable aux autres consonnes. La hamza est une lettre, mais généralement elle n'est pas considérée comme une consonne de l'alphabet de la langue Arabe. Le son de la Hamza existe dans les langues européennes dans la parole, mais il n'est pas représenté par l'écrit. En Arabe, il est à la fois entendu et écrit. Phonétiquement c'est un coup de glotte, prononcé comme une prise dans la gorge en tenant son souffle et en le libérant soudainement. La hamza est utilisée fréquemment, mais les règles d'écriture sont assez compliquées. La hamza dans sa forme écrite, peut apparaître sur la même ligne que les autres lettres comme dans 'مَسَاءٌ' /masa:?un/ (la soirée), ou en combinaison avec une lettre de support. Une lettre de support est une lettre qui a une hamza au-dessus ou au-dessous. Voir les exemples ci-dessous:

  - La hamza au-dessus de la consonne alif « ا ». Elle est représentée par la forme suivante « أ ». Cette hamza est suivie par les voyelles courtes fatha « ـَ » /a/ ou damma « ـُ » /u/.
  - La hamza au-dessous de la consonne alif « ا ». Elle est représentée par la forme suivante « إ ». Cette hamza est suivie par la voyelle courte kasra « ـِ » /i/.
  - La hamza sur la consonne waw « و ». Elle est représentée par la forme suivante « ؤ ».
  - La hamza sur la consonne ya «ي». Elle est représentée par la forme suivante « ئ ».
  - La hamza est combinée avec un alif long. Elle est représentée par la

forme suivante « آ ». Cette dernière est appelée alif madda, et se prononce par le son /?a:/.

- **Alif « ا » :** La consonne alif « ا » /?/ est la première lettre de l'alphabet de la langue Arabe. La lettre alif n'a pas de prononciation (pas de son), mais lorsque elle est munie des voyelles courtes, elle est prononcée comme la hamza « ء » /?/. L'une de ses principales fonctions est d'agir comme un support pour le signe hamza « ء ». La consonne Alif est également utilisée comme une voyelle longue lorsque elle est précédée par la voyelle courte fatha « ـَ » /a/.

- **Ba « ب » :** La consonne ba « ب » /b/ est prononcée par un arrêt bilabial vocalisé. La lettre ba « ب » se prononce comme la lettre « b » de l'Anglais.

- **Ta « ت » :** La consonne ta « ت » /t/ est produite par un arrêt dental sans vocalisation. La lettre ta « ت » se prononce comme la lettre « t » de l'Anglais.

- **Tha « ث » :** La consonne tha « ث » /T/ est un son fricative interdental sans vocalisation. La lettre tha « ث » se prononce comme on prononce le « th » du mot Anglais ‘teeth’ (les dents).

- **Jim « ج » :** La consonne jim « ج » /Z/ est produite par un son palato-alvéolaire vocalisé. En réalité, cette lettre a trois prononciations différentes selon l'origine dialectale du locuteur. Pour notre part en considère la prononciation utilisée que pour la langue Arabe standard. Donc, la lettre jim « ج » se prononce comme la lettre « j » du mot Anglais ‘job’ (profession).

- **Hha « ح » :** Cette consonne n'a pas de son équivalent dans les langues Européennes. Elle est prononcée dans le pharynx par une forte expiration. Le son produit par la consonne hha « ح » /X/ est un son fricatif pharyngal sans vocalisation.

- **Kha « خ » :** Le son de cette consonne se produit dans beaucoup de langues, tel que l’Allemand, l’Espagnole, …etc. La consonne kha « خ » /x/ est produite par un son fricatif uvulaire sans vocalisation. La lettre kha « خ » se prononce comme le « j » du mot Espagnol ‘mujer’ (femme).

- **Dal « د » :** La consonne dal « د » /d/ est prononcée par un arrêt dental vocalisé. La lettre dal « د » se prononce comme la lettre « d » de l'Anglais.

- **Thal « ذ » :** La consonne thal « ذ » /D/ est produite par un son fricatif interdental vocalisé. La lettre thal « ذ » se prononce comme le « th » du mot Anglais ‘this’ (ceci).

- **Ra « ر » :** La consonne ra « ر » /r/ est un son roulé (trille alvéolaire vocalisé), prononcée comme une succession rapide du rabat de la langue. Un bon exemple pour la prononciation de la lettre ra « ر », c'est de prononcer la lettre « r » du mot Anglais ‘very’ (très).

- **Zin « ز » :** La consonne zin « ز » /z/ est générée par un son alvéolaire sifflant vocalisé. La lettre zin « ز » est prononcée comme la lettre « z » du mot Anglais ‘zoom’ (zoom).

- **Sin « س » :** La consonne sin « س » /s / est produite par un son alvéolaire sifflant sans vocalisation. La lettre sin « س » se prononce comme la lettre « s » du mot Anglais ‘state’ (état).

- **Shin « ش » :** La consonne shin « ش » /S/ est générée par un son palato-alvéolaire sifflant sans vocalisation. La lettre shin « ش » se prononce comme on prononce le « sh » dans le mot Anglais ‘push’ (pousser).

- **Sad « ص » :** Lors de la prononciation de la consonne sad « ص » / s`/ le corps et la racine de la langue sont simultanément tirés en arrière vers la paroi arrière de la gorge (pharynx). Le son produit est fricatif alvéolaire sans vocalisation.

- **Dad « ض » :** La consonne dad « ض » / d`/ est prononcée par un arrêt alvéolaire vocalisé, avec une prononciation plus loin dans la bouche (une langue soulevée et tendue).

- **Tta « ط » :** La consonne tta « ط » / t`/ est classée comme un arrêt alvéolaire sans vocalisation (prononcée comme la consonne ta « ت » mais avec la langue pressée contre la gencive supérieure).

- **Zha « ظ » :** La consonne zha « ظ » / D`/ est produite par un son fricatif interdental vocalisé (prononcée comme la consonne zin « ز » mais avec la langue pressée contre la gencive supérieure). Le son de la lettre zha « ظ » est quelque peu semblable à celui de la lettre thal « ذ ».

- **Ayn « ع » :** Cette consonne n'a pas de son équivalent dans les langues Européennes. La consonne ayn « ع » /H/ est produite par un son fricatif laryngé vocalisé. Cette lettre est prononcée en appuyant sur la racine de la langue contre la paroi arrière du pharynx (partie supérieure de la gorge).

- **Ghayn « غ » :** La consonne ghayn « غ » /G/ est produite par un son fricatif uvulaire vocalisé. Ce son est un peu similaire au son de la lettre « r » de langue française.

- **Fa « ف » :** La consonne fa « ف » /f/ est produite par un son fricatif labiodental sans vocalisation. La lettre fa « ف » se prononce comme la lettre « f » de l'Anglais.

- **Qaf « ق » :** Cette consonne n'a pas de son équivalent dans les langues Européennes. La consonne qaf « ق » /q/ est générée par un arrêt uvulaire sans vocalisation, prononcée par la fermeture de l'arrière de la langue contre la luette.

- **Kaf « ك » :** La consonne kaf « ك » /k/ est produite par un arrêt vélaire sans vocalisation. La lettre kaf « ك » se prononce comme la lettre « k » de l'Anglais.

- **Lam « ل » :** La consonne lam « ل » /l/ est produite par un son alvéolaire latéral vocalisé. La lettre lam « ل » se prononce comme la lettre « l » du mot Anglais ‘let’ (laisser).

- **Mim « م » :** La consonne mim « م » /m/ est produite par un son nasal bilabial vocalisé. La lettre mim « م » se prononce comme la lettre « m » de l'Anglais.

- **Nun « ن » :** La consonne nun « ن » /n/ est produite par un son nasal alvéolaire vocalisé. La lettre nun « ن » se prononce comme la lettre « n » du Français.

- **Ha « ه » :** La consonne ha « ه » /h/ est produite par un son fricatif glottal sans vocalisation. La lettre ha « ه » se prononce comme la lettre « h » de l'Anglais.

- **Waw « و » :** La consonne waw « و » /w/ est produite par un son bilabial vocalisé. La lettre waw « و » est également utilisée comme une voyelle longue lorsque elle est précédée par la voyelle courte damma « ـُ » /u/. Elle est prononcée comme la lettre « w » du mot Anglais ‘well’ (bien).

- **Ya « ي » :** La consonne ya « ي » /j/ est produite par un son alvéo-palatal vocalisé. La lettre ya « ي » est également utilisée comme une voyelle longue lorsque elle est précédée par la voyelle courte kasra « ـِ » /i/. Elle est prononcée comme la lettre « y » du mot Anglais ‘yes’ (oui).

N°	Consonne	Nom	isolée	A la fin du mot	Au milieu du mot	Au début du mot
1	ا	Alif	ا	ـا	-	-
2	ب	Ba	ب	ـب	ـبـ	بـ
3	ت	Ta	ت	ـت	ـتـ	تـ
4	ث	Tha	ث	ـث	ـثـ	ثـ
5	ج	Jim	ج	ـج	ـجـ	جـ
6	ح	Hha	ح	ـح	ـحـ	حـ
7	خ	Kha	خ	ـخ	ـخـ	خـ
8	د	Dal	د	ـد	-	-
9	ذ	Thal	ذ	ـذ	-	-
10	ر	Ra	ر	ـر	-	-
11	ز	Zin	ز	ـز	-	-
12	س	Sin	س	ـس	ـسـ	سـ
13	ش	Shin	ش	ـش	ـشـ	شـ
14	ص	Sad	ص	ـص	ـصـ	صـ
15	ض	Dad	ض	ـض	ـضـ	ضـ
16	ط	Ta	ط	ـط	ـطـ	طـ
17	ظ	Zha	ظ	ـظ	ـظـ	ظـ
18	ع	Ayn	ع	ـع	ـعـ	عـ
19	غ	Ghayn	غ	ـغ	ـغـ	غـ
20	ف	Fa	ف	ـف	ـفـ	فـ

21	ق	Qaf	ق	ـق	ـقـ	قـ
22	ك	Kaf	ك	ـك	ـكـ	كـ
23	ل	Lam	ل	ـل	ـلـ	لـ
24	م	Mim	م	ـم	ـمـ	مـ
25	ن	Nun	ن	ـن	ـنـ	نـ
26	ه	Ha	ه	ـه	ـهـ	هـ
27	و	Waw	و	ـو		
28	ي	Ya	ي	ـي	ـيـ	يـ

Table5. Les différentes formes d'écriture des consonnes de la langue Arabe dans un mot.

II.2.3 Prononciation et description des voyelles Arabes:

La langue Arabe dispose de six voyelles (trois courtes, et trois longues). Le son aperçu (son voisé) est le même pour les deux types de voyelles (courtes et longues), mais avec une légère différence dans la durée de la voyelle longue qui est plus longue que la durée de la voyelle courte. Dans la langue Arabe la durée des voyelles longues est à peut près deux fois la durée des voyelles courtes.

- **Voyelle courte fatha « ـَ » :** La voyelle courte fatha est représentée par un petit trait diagonal sous la forme suivante « ـَ ». Cette dernière est toujours écrite au-dessus de la consonne quelque soit sa position dans le mot (au début, au milieu, à la fin). La voyelle courte fatha « ـَ » /a/ est prononcée comme la voyelle « a » du mot Anglais 'bag' (sac).

- **Voyelle courte damma « ـُ » :** La voyelle courte damma est représentée par un signe similaire à une virgule ou par un très petit waw « و ». Cette voyelle courte damma est représentée par la forme suivante « ـُ ». Cette dernière est toujours écrite au-dessus de la consonne quelque soit sa position dans le mot (au début, au milieu, à la fin). La voyelle courte damma « ـُ » /u/ est prononcée comme la voyelle « o » du mot Anglais 'do' (faire).

- **Voyelle courte kasra « ـِ » :** La voyelle courte kasra est représentée par un petit trait diagonal sous la forme suivante « ـِ ». Cette dernière est toujours écrite au-dessous de la consonne quelque soit sa position dans le mot (au début, au milieu, à la fin). La voyelle courte kasra « ـِ » /i/ est prononcée comme on prononce la voyelle « i » du mot Anglais 'in' (dans).

- **Voyelle longue fatha :** La voyelle longue fatha n'a pas de représentation graphique spécifique dans l'écriture Arabe, mais elle est représentée par la combinaison de la voyelle courte fatha « ـَ » et la consonne alif « ا », sous les formes suivantes: « ـَا », « ـَى », « آ ». La voyelle longue fatha /a:/ est prononcée comme la voyelle « a » du mot Anglais 'ask' (demander).

- **Voyelle longue damma « ـُو » :** La voyelle longue damma n'a pas de représentation graphique spécifique dans l'écriture Arabe, mais elle est représentée par la combinaison de la voyelle courte damma « ـُ » et la consonne waw « و », sous la forme suivante « ـُو ». La voyelle longue damma /u:/ est prononcée comme la voyelle « o » du mot Anglais 'short' (court).

- **Voyelle longue kasra «ـِي » :** La voyelle longue kasra n'a pas de représentation graphique spécifique dans l'écriture Arabe, mais elle est représentée par la combinaison de la voyelle courte kasra « ـِ » et la consonne ya « ي », sous la forme suivante «ـِي ». La voyelle longue kasra /i:/ est prononcée comme on prononce le « ee » du mot Anglais 'meet' (rencontrer).

II.3 Problèmes liés à la transcription phonétique:

Les situations qui peuvent engendrer des difficultés lors de la transcription phonétique du texte Arabe, peuvent se résumer par les cas suivants:

- Des graphèmes qui sont transcrits phonétiquement par le même phonème, à titre d'exemple:

  - Les lettres alif al-hamza « أ », waw al-hamza « ؤ » et ya al-hamza « ئ » sont transcrites phonétiquement par le phonème /?/.
  - Les lettres ta « ت » et ta-marbouta « ة » sont transcrites phonétiquement par le phonème /t/.

- Un graphème qui est transcrit phonétiquement par plusieurs phonèmes, à titre d'exemple:

  - Le graphème fathatan « ـً » est transcrit phonétiquement par les phonèmes /an/. Ce son est réalisé par la combinaison des deux phonèmes fatha « ـَ » /a/ et nun « ن » /n/.

- Le graphème dammatan « ـٌ » est transcrit phonétiquement par les phonemes /un/. Ce son est réalisé par la combinaison des deux phonèmes damma « ـُ » /u/ et nun « ن » /n/.
- Le graphème kasratan « ـٍ » est transcrit phonétiquement par les phonèmes /in/. Ce son est réalisé par la combinaison des deux phonèmes kasra « ـِ » /i/ et nun « ن » /n/.
- Le graphème alif madda « آ » est transcrit phonétiquement par les phonèmes /?a:/. Ce son est produit par la combinaison des deux phonèmes hamza « ء » /?/ et voyelle longue fatha /a:/.

- Un graphème qui est transcrit phonétiquement par différents phonèmes, à titre d'exemple:

- La lettre alif « ا » est transcrite phonétiquement par le phonème /?/ lorsque elle est munie des voyelles courtes, ainsi que par le phonème de la voyelle longue /a:/ lorsque elle est précédée par le graphème fatha « ـَ ».
- La lettre waw « و » est transcrite phonétiquement par le phonème /w/, ainsi que par le phonème de la voyelle longue /u:/ lorsque elle est précédée par le graphème damma « ـُ ».
- La lettre ya « ي » est transcrite phonétiquement par le phonème /j/, ainsi que par le phonème de la voyelle longue /i:/ lorsque elle est précédée par le graphème kasra « ـِ ».
- Les voyelles longues de type (fatha, damma, et kasra) sont transcrites phonétiquement respectivement par les phonèmes /a:/, /u:/, et /i:/, ainsi que par les phonèmes /a/, /u/, et /i/ lorsque elles sont suivies par l'article «ال» (les voyelles longues se transforment en voyelles courtes).

- Un graphème qui est transcrit phonétiquement comme le graphème qui le précède, à titre d'exemple:

- Le graphème chadda « ـّ » est transcrit phonétiquement par le phonème du graphème qui le précède, pour exemple la chadda « ـّ » qui est au-dessus de la consonne dal « د » dans le mot ' مَدَّ' /madda/ (étendre), prend le phonème /d/ de la consonne dal « د » qui le précède.

- Un graphème qui est présent dans l'écriture du mot, mais il n'est pas transcrit phonétiquement par un phonème (pas de son), à titre d'exemple:

- Le graphème sekun « ـْ » n'est jamais transcrit phonétiquement par un phonème. Il ne correspond à aucun son.
- Le graphème alif « ا » qui se trouve à la fin du mot n'est pas transcrit phonétiquement par un phonème dans le mot 'نَامُوا' /na:mu:/ (dormi). Il ne correspond à aucun son.

- Un graphème qui n'est pas présent dans l'écriture du mot, mais il est transcrit phonétiquement par un phonème, à titre d'exemple:

  - La lettre alif « ا » n'est pas présente dans l'écriture du mot 'ذَٰلِكَ' /Da:lika/ (cela), mais elle est transcrite phonétiquement par le phonème de la voyelle longue fatha /a:/.

- Un graphème qui a différentes formes de représentation graphique dans l'écriture d'un mot, à titre d'exemple:

  - Le graphème ba « ب » /b/ présent dans le mot 'بَابٌ' /ba:bun/ (porte), a différentes formes de représentation graphique au début et la fin du mot précité.

- Les problèmes liés à la transcription phonétique de l'article « ال » dans un mot, à titre d'exemple:

  - L'article « ال » est transcrit phonétiquement par les phonèmes /?al/ (succession de trois phonèmes /?/, /a/, /l/), lorsque l'article « ال » se trouve au début de la phrase et il est suivi par une lettre lunaire.
  - L'article « ال » est transcrit phonétiquement par les phonèmes /?a/ (succession de deux phonèmes /?/, /a/), lorsque l'article « ال » se trouve au début de la phrase et il est suivi par une lettre solaire.

- Les problèmes liés à la transcription phonétique des abréviations, les chiffres et les dates, à titre d'exemple:

  - L'abréviation du mot 'إِنْتَهَى' (terminé) est représentée par les lettres ' اه '. Cette dernière est transcrite phonétiquement par la séquence phonétique suivante /?in-taha:/.

II.4 Phonétisation du texte Arabe:

Afin de réaliser la transcription phonétique du texte Arabe, nous avons engagé en premier lieu une étude approfondie de la langue Arabe sur divers plans (grammatical, phonologique, phonétique, et codage de la langue), par la suite nous avons englobé tous les problèmes liés à la transcription phonétique du texte Arabe afin de générer notre lexique des exceptions, ainsi que notre base de règles de transcription graphème-phonème. Cette dernière traite l'ensemble des formes suivantes:

- Les consonnes qui sont de l'ordre de 28 consonnes.
- Les voyelles courtes symbolisées par (ـَ ـُ ـِ).

- Les voyelles longues symbolisées par (ـِي ـُو ـَا).
- Le doublement de voyelle qui s'appelle tanwin symbolisé par (ـً ـٌ ـٍ).
- Sekun symbolisé par « ـْ ».
- Ya maqsurah symbolisé par « ى ».
- La gémination appelée chadda symbolisée par « ـّ ».
- L'élision de la consonne alif symbolisée par « ا ».
- Les lettres lunaires (ا ب ج ح خ ع غ ف ق ك م ه و ي), qui n'assimilent pas la consonne lam symbolisée par « ل » de l'article « ال ».
- Les lettres solaires (ت ث د ذ ر ز س ش ص ض ط ظ ل ن), qui assimilent la consonne lam « ل » de l'article « ال ».
- La suppression des voyelles courtes, ainsi que tanwin à la fin d'une phrase.
- Remplacement des voyelles longues par des voyelles courtes lorsque les voyelles longues sont suivies par l'article « ال ».
- Les différentes représentations graphiques de la consonne alif qui sont symbolisées par (آ ئ ؤ إ أ).
- La suppression de la consonne ta marbota symbolisée par « ة » à la fin d'une phrase et la remplacer par la consonne ha symbolisée par « ه ». En dehors de ça elle est remplacée par la consonne ta symbolisée par « ت ».

Notre système de transcription phonétique est totalement basé sur la norme Unicode dans le codage des caractères de la langue Arabe, ainsi que dans les règles établies de la transcription graphème-phonème. Le test des caractères de la langue Arabe ne se fait pas sur le caractère lui-même mais sur son code selon la norme Unicode.

Cette méthode de représentation de la langue Arabe donne à notre système une portabilité et une souplesse de manipulation des caractères de la langue Arabe sans que la machine ne soit configurée en langue Arabe.

L'architecture de notre système de phonétisation automatique (Figure26), se compose de deux phases de traitements séquentiels: La première étape consiste à un prétraitement du texte. Ce dernier est suivi par la seconde phase qui comprend la transcription phonétique du texte en relation avec les caractéristiques de la langue.

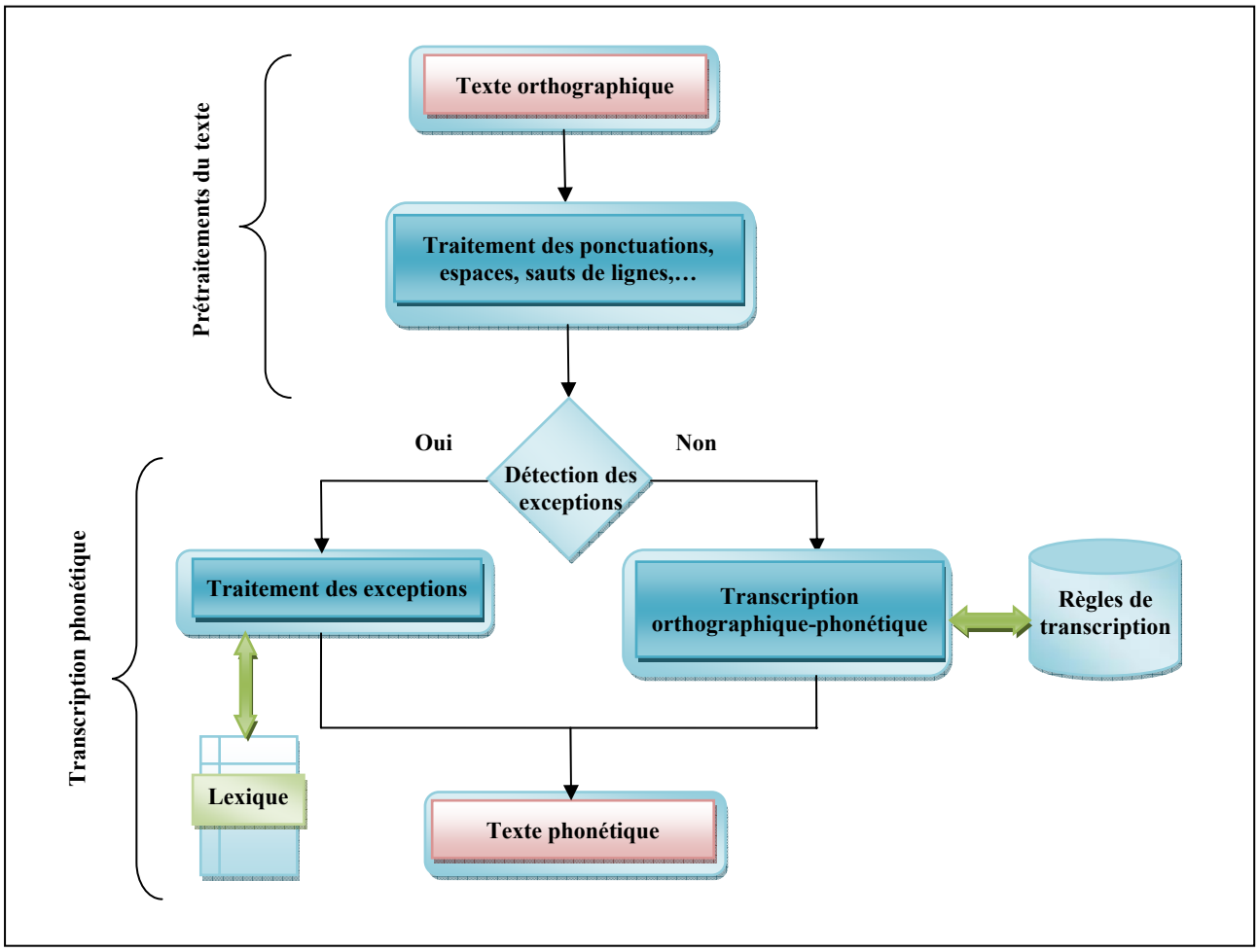


Figure26. Architecture du système de phonétisation automatique.

II.4.1 Prétraitement du texte:

Les caractères de la langue Arabe n'appartiennent évidemment pas au code ASCII, d'où la nécessité d'utiliser un autre code qui prend en charge la langue Arabe. Celui-ci est l'Unicode (Table6). Ce dernier permet de coder tous les caractères utilisés par la langue Arabe et d'échanger des données de texte entre différentes plates-formes et systèmes. L'autre avantage de la norme Unicode est d'utiliser le même code pour l'ensemble des différentes représentations graphiques d'une lettre donnée (Les lettres changent de forme de présentation selon leur position dans le mot).

Avant toute transcription phonétique du texte, il faudrait obligatoirement, la précéder par des prétraitements du texte afin d'éliminer toutes les ambiguïtés qui pourraient altérer l'exactitude de la transcription phonétique (le texte prétraité ne comporte aucune ambiguïté pour les traitements linguistiques ultérieurs).

Unicode (Hex)	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
062		ء	آ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
063	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ					
064		ف	ق	ك	ل	م	ن	ه	و	ى	ي	ً	ٌ	ٍ	َ	ُ
065	ِ	ّ	ْ													

Table6. Standard Unicode pour les caractères Arabes.

Ces prétraitements concernent les étapes suivantes:

- Lecture du texte au format 16 bits à cause de la norme Unicode.
- La transformation des données lues, en système Hexadécimal à cause de la phase de la transcription phonétique qui utilise la norme Unicode pour la phonétisation du texte.
- Traitements des ponctuations, des espaces et des sauts de lignes.
- Suppression des signes diacritiques qui sont ajouté à l'article « ال ».
- La suppression de la consonne alif symbolisée par « ا » au début d'une phrase et la remplacer par la consonne hamza symbolisée par « ء ».
- Organisation du texte d'entrée sous forme de cellules où chaque cellule contient un mot, cette opération est faite pour faciliter la phase de la transcription phonétique à base d'un lexique des exceptions.

II.4.2 Transcription graphème-phonème:

La transcription graphème-phonème constitue le cœur de notre travail, sans elle aucun système de synthèse de la parole ne pourra lire un texte donné (produire de la parole synthétique). Cette étape consiste à produire de la prononciation correspondante au texte en entrée sous la forme d'une liste de phonèmes.

La phase de la phonétisation du texte ne constitue pas une tache facile à concrétiser à cause de la non-correspondance directe entre les graphèmes de la langue Arabe et les phonèmes appropriés, ainsi que des mots qui ne sont pas modélisables par des règles de transcription graphème-phonème [Saï05*], d'où la nécessité de mettre en œuvre deux stratégies différentes pour transcrire le texte Arabe, à savoir:

- Phonétisation à base d'un lexique des exceptions.
- Phonétisation à base de règles.

II.4.2.1 Phonétisation à base d'un lexique des exceptions:

Notre lexique des exceptions (voir Table7) contient à la fois une liste de mots spéciaux et des abréviations (28 mots spéciaux et 9 abréviations), ces derniers ne sont pas pris en charge par la base de règles de transcription graphème-phonème à cause de leurs structures non modélisables par des règles. Cette méthode de transcription fait introduire directement la phonétisation correspondante aux mots traités sans passer par des règles de transcription phonétique.

Cette technique de transcription graphème-phonème est effectuée en comparant le mot traité avec une liste de mots (codés selon la norme Unicode) contenue dans un lexique des exceptions. Notons que la comparaison elle-même ne se fait pas sur les caractères qui constituent le mot mais selon leur code en norme Unicode.

Mots spéciaux	Transcription phonétique	Abréviations	Transcription phonétique
أُولَئِكَ	?ula:?ika	أنا	?an-ba?ana:
هَكَذَا	ha:kaDa:	اه	?in-taha:
الرَّحْمَنُ	?a-rraX-ma:nu	ثنا	XadaTana:
ذَلِكَ	Da:lika	الخ	?ila:_?a:xirihi
هَذَا	ha:Da:	رحه	raXimahu_lla:hu

Table7. Un échantillon de notre lexique des exceptions.

Cette stratégie de transcription assure plus de rapidité dans le traitement des mots d'exceptions mais engendre des erreurs de transcription dans des cas plus précis et afin de résoudre ces problèmes nous avons intégré une dizaine de règles qui traitent les erreurs de transcription liées aux liaisons existantes entre les mots (l'influence des mots voisins sur le mot traité). Ces règles traitent les cas suivants:

- Les lettres solaires qui suivent l'article « ال » dans un mot. Exemple: 'الرَّحْمَنُ' /?a-rraX-ma:nu/ (Rahman).
- Suppression des signes diacritiques à la fin du mot qui se situe à la fin d'une phrase. Exemple: 'رَبُ السَّمَوَاتِ' /rabbu ssama:wa:t/ (le Seigneur des cieux).
- Suppression des voyelles longues et les remplacer par des voyelles courtes lorsque les voyelles longues sont suivies par l'article « ال ». Exemple: 'هَذَا الرَّجُلُ' /ha:Da rraZul/ (cet homme).

Afin de faciliter la manipulation des mots d'exceptions (l'intégration de nouveaux mots d'exceptions ou la suppression des mots d'exceptions existants), ainsi que la génération de la

bonne transcription phonétique correspondante aux mots d'exceptions. Nous avons procédé à une génération automatique de notre lexique des exceptions (Figure27), cela est réalisable en prenant deux listes de mots l'une contient les mots d'exceptions et l'autre contient les mêmes mots d'exceptions mais réécrits de telle façon que la transcription phonétique générée soit juste.

Ces deux listes sont injectées dans notre système de génération du lexique des exceptions, la première liste est sauvegardée dans la première colonne de notre lexique des exceptions mais codée selon la norme Unicode dans le but de les utiliser dans la phase de la transcription graphème-phonème à base d'un lexique des exceptions. La deuxième liste est transcrite automatiquement en utilisant une base de règles de transcription phonétique appropriée, le résultat de la transcription sera sauvegardé dans la deuxième colonne de notre lexique des exceptions (les séquences phonétiques sont représentées selon la notation SAMPA).

Notons que chaque liste des mots spéciaux et des abréviations auront chacune leur propre lexique des exceptions, cette stratégie de séparation des deux lexiques sera bénéfique dans le temps d'exécution.

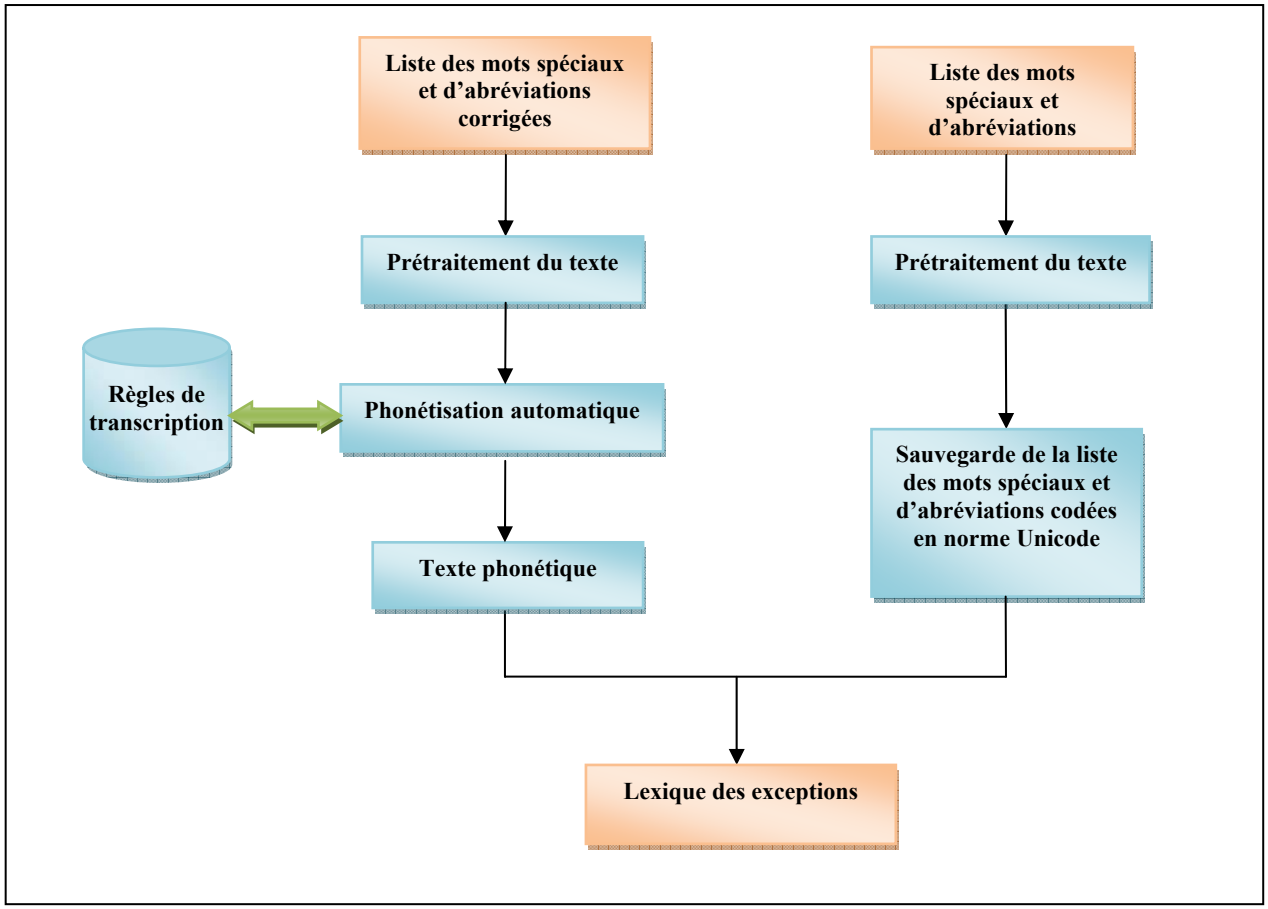


Figure27. Schéma de génération du lexique des exceptions.

II.4.2.2 Phonétisation à base de règles:

Après avoir traité les mots d'exceptions en utilisant un lexique des exceptions. Le reste du texte est transcrit à l'aide d'une base de règles de transcription graphème-phonème. Cette dernière utilise la norme Unicode pour le test des graphèmes de la langue Arabe. Les règles établies (100 règles de transcription graphème-phonème) traitent l'ensemble des réalisations graphiques de la langue Arabe qui sont au nombre de 44 graphèmes (voir Table5) pour enfin obtenir 34 phonèmes (28 consonnes, 6 voyelles).

A notre connaissance tous les systèmes qui réalisent la transcription phonétique à base de règles utilisent la même structure pour l'élaboration des règles de transcription phonétique (structure fixe pour l'ensemble des règles de transcription), où chaque graphème est remplacé par un ou plusieurs phonèmes selon son contexte gauche et droit systématiquement. Alors que notre système à base de règles de transcription phonétique utilise une structure variable pour l'élaboration de l'ensemble des règles de transcription graphème-phonème de la langue Arabe. Où chaque graphème de la langue Arabe, reçoit ces propres règles de transcription phonétique parce qu'il n'est pas nécessaire d'uniformiser les règles pour tous les graphèmes de la langue Arabe.

Autre particularité de notre système de transcription phonétique à base de règles est que nous obtenons les séquences phonétiques du texte Arabe sans passer par une table de conversion graphème-phonème (la phonétique est incorporée dans chaque règle de notre base de règle). L'avantage de cette démarche suivie pour transcrire le texte Arabe est bénéfique à divers plans (la précision du résultat phonétique, le temps d'exécution des règles).

La structure des règles élaborées pour transcrire le texte Arabe a été partagée en deux modèles:

- Le premier modèle utilise les contextes gauches et droits pour transformer les graphèmes en phonèmes, sa structure de règles est de la forme suivante: chaque graphème est remplacé par un ou plusieurs phonèmes selon ses contextes gauches, ses contextes droits, ou les deux contextes à la fois.
- Au contraire le deuxième modèle ne s'appuie pas sur les contextes gauche et droit pour transcrire le texte mais fait correspondre directement les graphèmes aux phonèmes, sa structure de règles est de la forme suivante : chaque graphème est associé à un ou plusieurs phonèmes directement sans passer par les contextes gauche et droit.

Afin de détailler les règles élaborées dans notre base de règles de transcription graphème-phonème, nous allons décrire dans ce qui suit quelques règles de transcription suivi d'un exemple approprié. Le résultat phonétique généré est représenté selon la notation SAMPA [Wel97], (voir Table8).

N°	Graphème	Nom	Phonème	N°	Graphème	Nom	Phonème
0	ء	Hamza	?	20	ف	Fa	f
1	ئ ؤ إ أ ا	Alif	?	21	ق	Qaf	q
2	ب	Ba	b	22	ك	Kaf	k
3	ت ة	Ta	t	23	ل	Lam	l
4	ث	Tha	T	24	م	Mim	m
5	ج	Jim	Z	25	ن	Nun	n
6	ح	Hha	X	26	ه	Ha	h
7	خ	Kha	x	27	و	Waw	w
8	د	Dal	d	28	ي	Ya	j
9	ذ	Thal	D	29	ـَ	Fatha	a
10	ر	Ra	r	30	ـُ	Damma	u
11	ز	Zin	z	31	ـِ	Kasra	i
12	س	Sin	s	32	آ ـَا ـَى	Longue fatha	a:
13	ش	Shin	S	33	ـُو	Longue damma	u:
14	ص	Sad	s`	34	ـِي	Longue kasra	i:
15	ض	Dad	d`	35	ـْ	Sukun	-
16	ط	Ta	t`	36	ـً	Fathatan	an
17	ظ	Zha	D`	37	ـٌ	Dammatan	un
18	ع	Ayn	H	38	ـٍ	Kasratan	in
19	غ	Ghayn	G	39		Silence	_

Table8. Correspondance graphème-phonème de la langue Arabe selon la notation SAMPA.

Règle1: **[Pho] = C**

Avec:

Pho: Phonème.
C: Caractère testé.

Cette règle indique qu'un caractère **C**, sera transcrit par le phonème **Pho**.

Exemple: **[ ] =** **'ـّـ'**

Cet exemple indique que le signe diacritique chadda **'ـّـ'** (représenté en Unicode par **'651'**), n'aura aucune transcription phonétique.

Règle2: **[Pho1] + [Pho2] =** **C**

Avec:

Pho1: Le premier phonème.
Pho2: Le deuxième phonème.

Cette règle indique qu'un caractère **C**, sera transcrit par les phonèmes **Pho1, Pho2**.

Exemple: **[a] + [n] =** **'ـًـ'**

Cet exemple indique que le signe diacritique fathatan **'ـًـ'** (représenté en Unicode par **'64B'**), sera transcrit par la succession des deux phonèmes **[a]** et **[n]**.

Règle3: **[Pho] = CG + C**

Avec:

CG: Contexte gauche du caractère testé.

Cette règle indique qu'un caractère **C**, précédé d'un caractère **CG**, sera transcrit par le phonème **Pho.**

Exemple: **[ ] =** **'ـُـ'** **+** **'و'**

Cet exemple indique que la consonne waw **'و'** (représentée en Unicode par **'648'**), précédée de la voyelle courte damma **'ـُـ'** (représentée en Unicode par **'64F'**), sera muette.

Règle4: **[Pho] =** **C + CD**

Avec:

CD: Contexte droit du caractère testé.

Cette règle indique qu'un caractère **C**, suivi par un caractère **CD**, sera transcrit par le phonème **Pho**.

Exemple: **[a:] =** **'ـَـ'** **+** **'ا'**

Cet exemple indique que la voyelle courte fatha **'ـَـ'** (représentée en Unicode par **'64E'**), suivie de la consonne alif **'ا'** (représentée en Unicode par **'648'**), sera transcrite par le phonème de la voyelle longue **[a:]**.

Règle5: **[Pho] = CG2 + CG1 + C + CD**

Avec:

CG1: Le premier contexte gauche du caractère testé.
CG2: Le deuxième contexte gauche du caractère testé.

Cette règle indique qu'un caractère **C**, précédé par des caractères **CG1, CG2** et suivi par le caractère **CD**, sera transcrit par le phonème **Pho**.

Exemple: **[ ] = '00D' + 'ا' + 'ل' + 'LS'**

Cet exemple indique que la consonne lam 'ل' (représentée en Unicode par **'644'**), précédée de la consonne alif 'ا' (représentée en Unicode par **'627'**), et du caractère **'00D'** (début de phrase), et suivie d'une lettre solaire **'LS'**, n'aura aucune transcription phonétique, donc la lettre lam 'ل' est muette.

Règle6: **[Pho] = CG2 + CG1 + C**

Cette règle indique qu'un caractère **C**, précédé par des caractères **CG1, CG2**, sera transcrit par le phonème **Pho**.

Exemple: **[a:] = 'ل' + 'ا' + 'ـَ'**

Cet exemple indique que la voyelle courte fatha 'ـَ' (représentée en Unicode par **'64E'**), précédée de la consonne alif 'ا' (représentée en Unicode par **'627'**), et de la consonne lam **'ل'** (représentée en Unicode par **'644'**), sera transcrite par le phonème de la voyelle longue **[a:].**

Règle7: **[Pho1] + [Pho2] = CG + C + CD1 + CD2**

Avec :

CD1: Le premier contexte droit du caractère testé.
CD2: Le deuxième contexte droit du caractère testé.

Cette règle indique qu'un caractère **C**, précédé par le caractère **CG**, et suivi par des caractères **CD1, CD2**, sera transcrit par les phonèmes **Pho1, Pho2**.

Exemple: **[?] + [a] = '00D' + 'ا' + 'ل' + 'LS'**

Cet exemple indique que la consonne alif 'ا' (représentée en Unicode par **'627'**), précédée du caractère **'00D'** (début de phrase), et suivie de la consonne lam 'ل' (représentée en Unicode

par **'644'**), et d'une lettre solaire **'LS'**, sera transcrite par la succession des deux phonèmes **[?]** et **[a]**.

Règle8: **[Pho1] + [Pho2] = C + CD1 + CD2**

Cette règle indique qu'un caractère **C**, suivi par des caractères **CD1, CD2**, sera transcrit par les phonèmes **Pho1, Pho2**.

Exemple: **[u] + [w] = 'ـُ' + 'و' + 'VC'**

Cet exemple indique que la voyelle courte damma **'ـُ'** (représentée en Unicode par **'64F'**), suivie de la consonne waw **'و'** (représentée en Unicode par **'648'**), et d'une voyelle courte **'VC'**, sera transcrite par la succession des deux phonèmes **[u]** et **[w]**.

Afin de bien comprendre les règles de transcription graphème-phonème précitées. Nous présentons dans ce qui suit un exemple d'application de ces règles sur le mot 'الطَّعَام' (nourriture). Nous commençons par transcrire le mot 'الطَّعَام' (nourriture) lettre par lettre:

- La consonne **alif** 'ا' est transcrite par les phonèmes **[?] + [a]** selon la **règle7**.
- La consonne **lam** 'ل' est transcrite par la non-présence de phonème selon la **règle5**.
- La consonne **ta** 'ط' est transcrite par les phonèmes **[t`] + [t`]** selon la structure de la **règle4**.
- La gémination **chadda** 'ـّ' est transcrite par la non-présence de phonème selon la **règle1**.
- La voyelle **fatha** 'ـَ' est transcrite par le phonème **[a]** selon la **règle1**.
- La consonne ayn 'ع' est transcrite par le phonème **[H]** selon la **règle1**.
- La voyelle **fatha** 'ـَ' est transcrite par le phonème **[a:]** selon la **règle4**.
- La consonne **alif** 'ا' est transcrite par la non-présence de phonème selon la **règle3**.
- La consonne **mim** 'م' est transcrite par le phonème **[m]** selon la **règle1**.

Après avoir appliqué les règles de transcription graphème-phonème au mot 'الطَّعَام' (nourriture), nous obtenons la transcription phonétique suivante: **?at`t`aHa:m**

II.5 Conclusion:

La phase de la transcription phonétique du texte Arabe constitue une étape obligatoire et essentielle dans un système de synthèse de la parole à partir du texte. La phonétisation automatique du texte Arabe ne peut se faire sans une étude approfondie à divers plans de la langue Arabe.

Dans ce chapitre nous avons abordé l'ensemble des traitements linguistiques destiné à la synthèse de la parole à partir du texte de la langue Arabe. Ces traitements linguistiques sont effectués dans le but de générer la transcription phonétique la plus exacte possible du texte Arabe de sorte que le signal de la parole synthétisée soit intelligible.

Dans ce chapitre nous avons aussi présenté les différentes étapes que nous avons suivies pour aboutir à un système de phonétisation automatique du texte pour la langue Arabe, en détaillant l'ensemble des règles de transcription graphème-phonème.

La stratégie suivie pour transcrire le texte Arabe a été très bénéfique à divers plans soit sur la précision du résultat phonétique, soit sur le temps d'exécution, et soit sur la simplicité de la mise en œuvre. Nous avons constaté lors de l'élaboration de notre base de règles de transcription graphème-phonème, qu'il n'est pas nécessaire de considérer les contextes gauches et droits de manière systématique dans le processus de transcription.

Chapitre III: La génération des allophones

III.1 Introduction:

Le naturel et l'intelligibilité du signal de la parole synthétisé sont deux paramètres très importants dans un synthétiseur de la parole. Ces deux paramètres de la validation de la qualité de la synthèse de la parole ne dépendent pas que de la méthode utilisée pour la synthèse de la parole [Tab11], [Ras10], mais aussi dans la base de données acoustique utilisée lors de la production de la parole synthétique. D'où la nécessité d'intégrer les allophones dans la base de données acoustique afin de les utiliser pour améliorer la qualité de la parole synthétique produite.

La langue Arabe est une langue phonétiquement variante (incluant différentes réalisations acoustiques concernant la prononciation des phonèmes), d'où la nécessité de prendre en charge ces caractéristiques phonétiques afin de produire correctement de la parole à partir du texte Arabe.

Dans ce qui suit nous allons détailler la démarche que nous avons suivie pour la génération des différentes réalisations sonores (allophones) du parlé Arabe. Nous allons également décrire les différentes règles établies de la transformation des phonèmes en allophones.

III.2 Définition des allophones:

Les règles phonologiques appliquées sur les phonèmes génèrent les différentes réalisations acoustiques (allophones). Ces règles phonologiques ont pour but d'éclaircir l'effet des sons (phonèmes) les uns sur les autres afin de déterminer le son à produire à la fin de l'analyse phonologique.

Les allophones sont des unités acoustiques (sons de la parole). Ces derniers sont les différentes réalisations acoustiques d'un phonème (variante phonétique d'un phonème). Les allophones sont définis comme une multitude de formes articulatoires des phonèmes, caractérisés par différentes fréquences fondamentales, ainsi que par différents niveaux d'énergie (Figure28).

Les allophones d'un même phonème ne changent pas le sens d'un mot (pas d'information sémantique), pour exemple, la séquence phonétique [bath] ne diffère pas de sens à la séquence phonétique [bat]; avec [th] est un /t/ aspiré et [t] est un /t/ non aspiré).

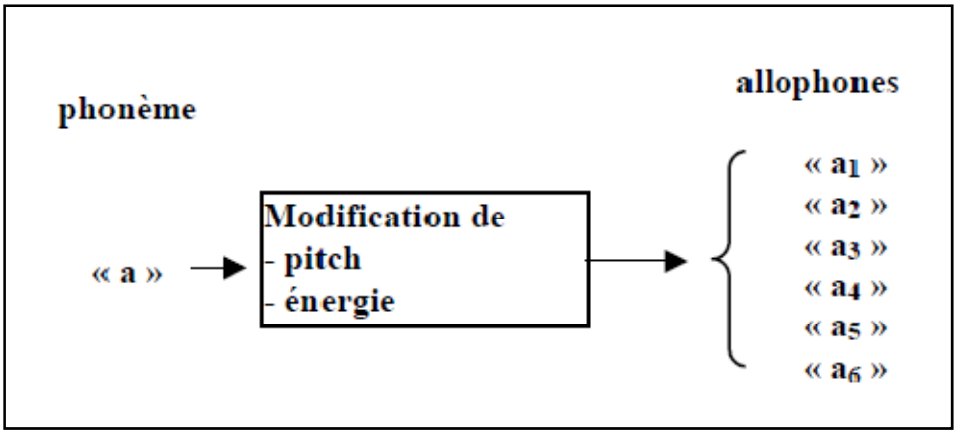


Figure28. Les différents allophones du phonème « a ».

III.3 Travaux réalisés:

Les travaux de recherche qui portent sur la génération des allophones dédié spécifiquement à la langue Arabe, en utilisant des règles de conversion phonème-allophone sont peu nombreux. Ces travaux réalisés se basent essentiellement sur l'ordre phonologique des phonèmes dans le mot afin de produire les allophones. Dans ce qui suit nous allons évoquer les différents travaux qui traitent l'aspect de la génération des allophones de la langue Arabe.

- Le travail fait par Yousif A. El-imam [Eli04] traite essentiellement les importantes variations phonétiques des sons Arabes qui comprennent la pharyngalisation des voyelles et des diphtongues, la nasalisation des voyelles et des diphtongues, et d'autres coarticulations anticipatoires. Ces traitements sont les suivants:

  - La nasalisation des voyelles qui se trouvent avant les consonnes nasales (mim 'م', noun 'ن').
  - Le traitement de l'emphase qui concerne les voyelles et les diphtongues (/aj/ lorsque la voyelle fatha ' ـَ ' est suivie par la consonne ya 'ي', et /aw/ lorsque la voyelle fatha ' ـَ ' est suivie par la consonne waw 'و'), et ainsi que les consonnes (lam 'ل' et ra 'ر'), lorsque elles se trouvent avant ou après les consonnes emphatiques.
  - Le remplacement des phonèmes de contre partie emphatique (ta 'ت', dal 'د', thal 'ذ', sin 'س') par les phonèmes emphatiques (ta 'ط', dad 'ض', zha 'ظ', sad 'ص') lorsque ces phonèmes de contre partie emphatique se trouvent près d'une syllabe qui contient un phonème emphatique ou un son pharyngalisé.

- Par contre le travail réalisé par Moustapha Elshafei [Els02] ignore la nasalisation, le traitement des diphtongues ainsi que le traitement des phonèmes de contrepartie emphatique. En contrepartie il met en valeur l'accentuation des voyelles (courtes, longues). Ce travail se résume dans les points suivants:

  - Les voyelles seront accentuées lorsqu'elles sont placées avant ou après les consonnes emphatiques (ta '**ط**', dad '**ض**', zha '**ظ**', sad '**ص**') et les consonnes pharyngales (kha '**خ**', ghayn '**غ**', qaf '**ق**').

  - L'accentuation des voyelles lorsqu'elles sont placées après les consonnes (lam '**ل**' et ra '**ر**'), si ces dernières sont pharyngalisées.

  - Les voyelles seront amincies lorsqu'elles sont placées avant ou après le reste des consonnes (pas de consonnes emphatiques ou pharyngales).

- L'étude faite par Mansour M. Al-ghamdi [Alg04], diffère des travaux cités par le fait que son étude traite en plus les cas suivants: le fusionnement '**الإدغام**', la conversion **'الإقلاب'** et la dissimulation **'الإخفاء'**. Ce travail se résume dans les points suivants:

  - L'accentuation des voyelles se fait lorsqu'elles sont placées après les consonnes emphatiques et les consonnes pharyngales.
  - L'accentuation **'التفخيم'** ou l'amincissement **'الترقيق'** des consonnes (lam '**ل**' et ra '**ر**') selon des règles phonologiques spécifiques.
  - La transformation du son de la consonne thal '**ذ**' par le son de la consonne zha '**ظ**' lorsque cette dernière se trouve après le phonème de la consonne thal '**ذ**', ainsi que la transformation du son de la consonne ta '**ت**' par les sons des consonnes ta '**ط**' et dal '**د**' lorsque ces dernières se trouvent respectivement après le phonème de la consonne ta '**ت**'. Ces règles phonologiques traitent l'aspect du fusionnement '**الإدغام**'.
  - La transformation du son de la consonne nun '**ن**' par le son de la consonne mim '**م**' lorsque les phonèmes des consonnes mim '**م**' et ba '**ب**' se trouvent respectivement après le phonème de la consonne nun '**ن**'. Ce type de règle traite la conversion **'الإقلاب'**.
  - La génération des allophones de tonalité nasale '**الغنة**' lorsque quelques phonèmes des consonnes (qaf '**ق**', kaf '**ك**', jim '**ج**', shin '**ش**', sad '**ص**', dad '**ض**', sin '**س**', zin '**ز**', ta '**ط**', dal '**د**', ta '**ت**', zha '**ظ**', thal '**ذ**', tha '**ث**', fa '**ف**') sont précédées par le phonème de la consonne nasale nun '**ن**'. Cette dernière prend le son de la consonne qui la suit mais avec une tonalité nasale '**الغنة**'. Ce traitement est la dissimulation **'الإخفاء'**.

III.4 Classification des consonnes Arabes:

Les consonnes de la langue Arabe sont classées en trois catégories:

- Consonnes toujours fortes '**المفخمة**'.
- Consonnes toujours faibles '**المرققة**'.
- Consonnes fortes, faibles (selon leurs contextes dans le mot).

III.4.1 Consonnes toujours fortes:

Les consonnes toujours fortes devraient être magnifiées une fois prononcées et sont de l'ordre de 7 consonnes [Har10]. Ces dernières sont partagées en deux catégories: la première catégorie ce sont les consonnes emphatiques: sad '**ص**', dad '**ض**', ta '**ط**', zha '**ظ**'. La deuxième catégorie sont les consonnes pharyngales: kha '**خ**', ghayn '**غ**', qaf '**ق**'. Ces trois consonnes sont moins magnifiées une fois prononcées que les consonnes emphatiques.

III.4.2 Consonnes toujours faibles:

Les consonnes toujours faibles devraient être amincies une fois prononcées et sont de l'ordre de 19 consonnes [Els02], qui sont: alif '**ا**', ba '**ب**', ta '**ت**', tha '**ث**', jim '**ج**', hha '**ح**', dal '**د**', thal '**ذ**', zin '**ز**', sin '**س**', shin '**ش**', ayn '**ع**', fa '**ف**', kaf '**ك**', mim '**م**', nun '**ن**', ha '**ه**', waw '**و**', ya '**ي**'.

III.4.3 Consonnes fortes, faibles:

Les consonnes fortes, faibles devraient être magnifiées ou amincies une fois prononcées (elles changent de forme acoustique selon leurs contextes dans le mot suivant des règles spécifiques) et sont de l'ordre de 2 consonnes [Els02], qui sont: lam '**ل**', ra '**ر**'.

III.5 Transcription phonème-allophone:

La démarche que nous avons suivie pour générer les allophones (voir Figure29), est fondée sur l'utilisation d'une base de règles de transcription phonème-allophone qui prend en charge toutes les spécificités phonologiques de la langue Arabe, afin de générer correctement les différentes réalisations acoustiques (les allophones).

Notre base de règles de transcription phonème-allophone que nous avons élaborée contient 30 règles qui traitent l'ensemble des 34 phonèmes pour enfin obtenir 48 sons. Ces règles traitent les cas suivants:

- Les phonèmes emphatiques.
- Les phonèmes pharyngaux.

- Les phonèmes nasaux (mim ‘م’, nun ‘ن’).
- Les phonèmes (lam ‘ل’ et ra ‘ر’).

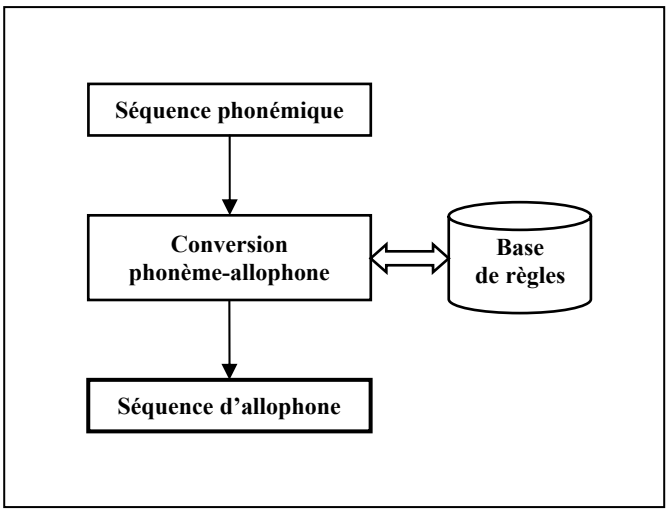


Figure29. Schéma représentatif de la génération des allophones.

Les règles établies s’appuient sur la position ou l’emplacement du phonème dans le mot à prononcer, donc l’influence des phonèmes voisins sur le phonème traité génère les allophones. De ce fait la forme des règles établies obéit à la forme suivante: chaque phonème testé est remplacé par un allophone selon son contexte gauche, ou son contexte droit, ou les deux à la fois.

[All] = CG + Pho , **[All] = Pho + CD** , **[All] = CG + Pho + CD**

Avec:

All: Allophone.
CG: Contexte gauche du phonème testé.
Pho: Phonème testé.
CD: Contexte droit du phonème testé.

Ces trois formes de règles indiquent qu’un phonème **Pho**, précédé d’un phonème **CG**, ou suivi d’un phonème **CD**, ou précédé et suivi par les deux contextes en même temps (**CG**, **CD**) aura comme réalisation acoustique l’allophone **All**.

La représentation des allophones générés est inspirée de la notation SAMPA puisque les séquences phonémiques sont déjà représentées par cette dernière. Les allophones qui sont toujours prononcés d’une façon amincie et magnifiée prennent la notation SAMPA pour leur représentation. Tandis que les autres allophones sont représentés en utilisant une représentation interne (voir Table9).

Lettres	Son aminci	Son magnifié	Lettres	Son aminci	Son magnifié	Son nasalisé
Alif ‘ا’	?		Zha ‘ظ’		D`	
Ba ‘ب’	b		Ayn ‘ع’	H		
Ta ‘ت’	t		Ghayn ‘غ’		G	
Tha ‘ث’	T		Fa ‘ف’	f		
Jim ‘ج’	Z		Qaf ‘ق’		q	
Hha ‘ح’	X		Kaf ‘ك’	k		
Kha ‘خ’		x	Lam ‘ل’	l	L	
Dal ‘د’	d		Mim ‘م’			m
Thal ‘ذ’	D		Nun ‘ن’			n
Ra ‘ر’	R	r	Ha ‘ه’	h		
Zin ‘ز’	z		Waw ‘و’	w		
Sin ‘س’	s		Ya ‘ي’	j		
Shin ‘ش’	S		Fatha ‘ـَ’	a	A	à
Sad ‘ص’		s`	Damma ‘ـُ’	u	U	ù
Dad ‘ض’		d`	Kasra ‘ـِ’	i	I	ì
Ta ‘ط’		t`				

Table9. Correspondance des allophones de la langue Arabe.

III.5.1 Les phonèmes emphatiques:

Lorsque les phonèmes des 6 voyelles (3 courtes, 3 longues) ou les phonèmes (lam ‘ل’ et ra ‘ر’) sont au voisinage des phonèmes emphatiques, ils seront transformés en sons magnifiés lors de la prononciation.

Exemple: **[A] = ‘s`’ + ‘a’**

Cet exemple indique que lorsque le phonème **‘a’** (son de la voyelle courte fatha ‘ـَ’), est précédé du phonème emphatique **‘s`’** (son de la consonne emphatique sad ‘ص’), aura comme réalisation acoustique l’allophone **[A]**.

Lorsque les phonèmes de contrepartie emphatique (ta ‘ت’, dal ‘د’, thal ‘ذ’, sin ‘س’) sont au voisinage des phonèmes emphatiques, ils seront transformés en phonèmes emphatiques qui leur correspondent. Les phonèmes de contrepartie emphatique (ta ‘ت’, dal ‘د’, thal ‘ذ’,

sin ‘س’) correspondent respectivement aux phonèmes emphatiques (ta ‘ط’, dad ‘ض’, zha ‘ظ’, sad ‘ص’).

Exemple: **[d`] = ‘s`’ + ‘d’**

Cet exemple indique que lorsque le phonème **‘d’** (son de la consonne dal ‘د’), est précédé du phonème emphatique **‘s`’** (son de la consonne emphatique sad ‘ص’), aura pour réalisation acoustique l’allophone **[d`]**.

III.5.2 Les phonèmes pharyngaux:

Lorsque les phonèmes des 6 voyelles (3 courtes, 3 longues) ou les phonèmes (lam ‘ل’ et ra ‘ر’) sont au voisinage des phonèmes pharyngaux, ils seront transformés en sons magnifiés lors de la prononciation.

Exemple: **[A] = ‘a’ + ‘q’**

Cet exemple indique que lorsque le phonème **‘a’** (son de la voyelle courte fatha ‘ـَ ’), est suivi du phonème pharyngal **‘q’** (son de la consonne pharyngale qaf ‘ق’), aura comme réalisation acoustique l’allophone **[A]**.

III.5.3 Les phonèmes nasaux:

Lorsque les phonèmes des 6 voyelles (3 courtes, 3 longues) sont au voisinage des phonèmes nasaux, ils seront transformés en sons d’une prononciation nasale.

Exemple: **[à] = ‘a’ + ‘m’**

Cet exemple indique que lorsque le phonème **‘a’** (son de la voyelle courte fatha ‘ـَ ’), est suivi du phonème nasal **‘m’** (son de la consonne nasale mim ‘م’), aura comme réalisation acoustique l’allophone **[à]**.

III.5.4 Les phonèmes (lam ‘ل’ et ra ‘ر’):

Le traitement des phonèmes (lam ‘ل’ et ra ‘ر’) est très spécial, à cause de leur variabilité phonétique dans le mot prononçable. Ces deux phonèmes peuvent prendre deux différentes réalisations acoustiques (magnifiées, amincies) selon leurs contextes dans le mot. Le traitement du phonème lam ‘ل’ est uniquement dans le mot allahu ‘اللَّهُ’ ainsi que dans le mot allahuma ‘اللَّهُمَ’. Alors que pour le phonème ra ‘ر’ est traité dans tous les mots sans exception.

Exemple1: **[R] = ‘r’ + ‘i’**

Cet exemple indique que lorsque le phonème **‘r’** (son de la consonne ra ‘ر’), est suivi du

phonème **‘i’** (son de la voyelle courte kasra ‘ ـِ ’), aura pour réalisation acoustique l’allophone **[R]**.

Exemple2: **[r] = ‘i’ + ‘r’ + ‘q’**

Cet exemple indique que lorsque le phonème **‘r’** (son de la consonne ra ‘ر’), est précédé du phonème **‘i’** (son de la voyelle courte kasra ‘ ـِ ’), et suivi du phonème pharyngal **‘q’** (son de la consonne pharyngale qaf ‘ق’), aura comme réalisation acoustique l’allophone **[r]**.

Afin de détailler toutes les règles utilisées pour générer les différentes réalisations acoustiques (allophones), nous avons décrit ces règles de transcription phonème-allophone sous la forme d'un code qui est donnée dans la “ Figure30”. Ces règles doivent être appliquées dans l’ordre qui est donné dans le code pour assurer l'efficacité de la transcription phonème-allophone.

```
DEBUT

Pour tout i variant de 1 jusqu’à  N
(N: La taille de la séquence phonémique)
{
// Traitement des voyelles
  Si (le phonème testé == son d’une voyelle)
     {
    Si (le cotexte gauche ou droit == phonème emphatique ou pharyngal)
       {
         Allophone(i) = correspondance du son magnifié
        }
     Sinon Si   (le cotexte gauche ou droit == phonème nasal)
          {
           Allophone(i) = correspondance du son nasalisé
           }
           Sinon
           {
           Allophone(i) = correspondance du son aminci
           }
        }
// Traitement des phonèmes de contrepartie emphatique
    Sinon Si (le phonème testé == son de contrepartie emphatique)
     {
    Si (le cotexte gauche ou droit== phonème emphatique)
       {
         Allophone(i) = correspondance du son emphatique
         }

    Sinon
          {
           Allophone(i)= correspondance du son de contrepartie emphatique
           }
      }
```

```
// Traitement du phonème ra 'ر'
   Sinon Si (le phonème testé == son de ra 'ر')
     {
     Si (le cotexte gauche == phonème kasra 'ـِ')  et
        (le cotexte droit == phonème emphatique ou  pharyngal)
         {
         Allophone(i)= correspondance du son magnifié
         }
      Sinon Si   (le cotexte gauche == phonème kasra 'ـِ')  et
                 (le contexte droit == phonème d'une consonne)  ou
                 (le cotexte droit == phonème kasra 'ـِ')
         {
         Allophone(i) = correspondance du son aminci
         }
       Sinon
         {
          Allophone(i) = correspondance du son magnifié
          }
      }
// Traitement du phonème lam 'ل'
   Sinon Si (le phonème testé == son de lam 'ل')
     {
      Si (le lam majestueux 'لام الجلالة')
        {
        Si  (le cotexte gauche == phonème fatha 'ـَ')  ou
            (le cotexte gauche== phonème damma 'ـُ')
          {
            Allophone(i)= correspondance du son magnifié
            }
         Sinon
           {
            Allophone(i)= correspondance du son aminci
            }
          }
       Sinon Si (le cotexte gauche ou droit == phonème emphatique ou pharyngal)
           {
             Allophone(i)= correspondance du son magnifié
            }
         Sinon
           {
            Allophone(i)= correspondance du son aminci
            }
      }
// Le reste des phonèmes
   Sinon
  {
   Allophone(i)= correspondance du son aminci
   }
}                                                    FIN
```

Figure30. Code de la génération des allophones.

III.6 Conclusion:

Cette phase de traitement linguistique consiste à transformer les phonèmes issus de la transcription graphème-phonème, en allophones. Ces derniers sont des sons (phones), qui sont les variantes de prononciation d'un phonème.

Au terme de ce chapitre nous avons réalisé un système qui génère correctement les différentes réalisations acoustiques (allophones) de la langue Arabe à partir d'une séquence phonémique.

La démarche que nous avons suivie afin d'obtenir les allophones se base essentiellement sur des règles de transcription phonème-allophone que nous avons élaborées. Dans ce chapitre nous avons détaillé toutes les règles établies, ainsi que le code de la génération des allophones afin de simplifier le processus de la transcription phonème-allophone.

La génération des différentes réalisations acoustiques (allophones) améliore la qualité du signal de parole synthétique issu des synthétiseurs de la parole dédiés à la langue Arabe, cette amélioration sera perçue lors de la phase d'écoute du signal de la parole synthétisée.

Chapitre IV: Les traitements acoustiques du synthétiseur de la parole

IV.1 Introduction:

La partie qui concerne la génération du signal de la parole synthétique (traitements acoustiques) est une phase très importante dans un système de la synthèse de la parole à partir du texte (Text-To-Speech) puisque la qualité de la synthèse de la parole dépend totalement de la qualité des traitements de la parole effectuée à cet étage.

Le traitement de la parole (destiné à la synthèse de la parole) est un domaine qui s'appuie sur les outils de traitement du signal dans le but de modifier les unités acoustiques de la parole de façon à avoir à la fin de ces traitements acoustiques, un signal de la parole artificielle de bonne qualité.

Dans ce chapitre nous allons aborder la phase des traitements de la parole du synthétiseur dans le but de produire de la parole artificielle à partir des séquences phonétiques issues des traitements linguistiques. Cette phase des traitements acoustiques consiste en la sélection des unités acoustiques préenregistrées à concaténer, stockées dans une base de données acoustique en utilisant les séquences phonétiques. Ensuite ces unités acoustiques subissent des traitements spécifiques au point de concaténation selon la nature des sons à concaténer (voisés, non voisés) afin de générer un signal de parole synthétique le plus naturel et intelligible possible.

Les travaux de recherche qui traitent de la synthèse de la parole par concaténation à partir du texte Arabe sont très nombreux pour exemples: les travaux réalisés par Mansour Al-ghamdi [Alg02], Moustafa Elshafei [Els02], Tahar Saidane [Saï05], Zouhir Zemirli [Zem06], Rashad [Ras10*], Hazem El-bakry [Elb11], Mazin Hamad [Ham11]. Ces derniers convergent vers un système de synthèse de la parole en concaténant les unités acoustiques issues de la segmentation du signal de parole naturelle, en intégrant les paramètres prosodiques (fréquence fondamentale, durée, intensité,…) issus des traitements linguistiques.

Pour notre part nous avons opté pour la synthèse de la parole par concaténation de diphones, cette dernière est une méthode simple à mettre en œuvre et donne de bons résultats soit sur l'intelligibilité ou sur le naturel de la voix synthétisée. Le diphone est une portion de parole allant de la partie stable d'un phonème à la partie stable du phonème qui le suit, portant essentiellement la transition entre deux phonèmes, information difficilement modélisée et pourtant essentielle sur le plan de l'intelligibilité.

IV.2 Structure générale des traitements acoustiques du synthétiseur:

Les traitements acoustiques du synthétiseur de la parole sont illustrés par la “Figure31”. Ces traitements acoustiques peuvent se résumer par les étapes suivantes:

- Sélection et chargement des diphones à concaténer.

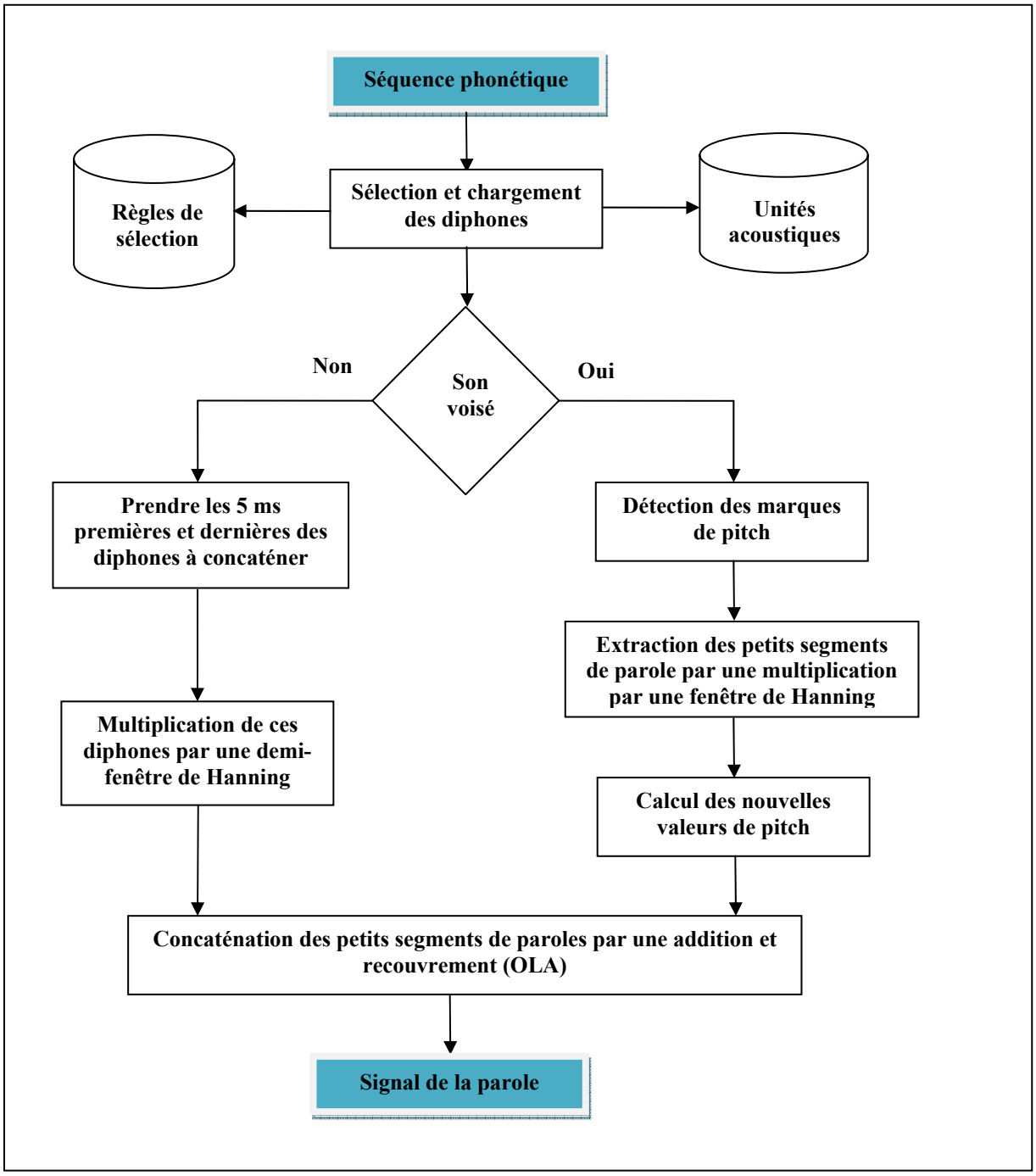


Figure31. Schéma représentatif des traitements acoustiques du synthétiseur de la parole.

- Traitement des sons voisés.
- Traitement des sons non voisés.
- Concaténation des segments de parole et synthèse de la parole.

IV.3 Base de données acoustique:

La constitution de notre base de donnés acoustique de type diphone est une tache difficile puisque la qualité de la synthèse de la parole est liée entièrement à la qualité des unités acoustiques réalisées. Cette tache comporte plusieurs étapes:

- Mise en œuvre d'un corpus de logatomes dédiés à la langue Arabe pour l'enregistrement de l'ensemble des unités acoustiques utilisées dans la synthèse de la parole.
- L'enregistrement de ce corpus de logatomes.
- Constitution de notre base de données acoustique par la segmentation des enregistrements sonores.

IV.3.1 Corpus de logatomes dédiés à la langue Arabe:

Le corpus de logatomes Arabes (suite de sons ou de syllabes sans signification) a été réalisé de façon à englober toutes les combinaisons possibles afin d'extraire toutes les unités acoustiques utilisées lors de la production de la voix synthétique.

La structure des logatomes a été inspirée du travail réalisé par Mansour Al-ghamdi [Alg02] à la différence d'utilisation de la consonne sin '**س**' (son non voisé) à la place de la consonne zin '**ز**' (son voisé) dans la constitution du logatome ce choix a été bénéfique puisque il nous a permit de simplifier la segmentation en localisant la cible à segmenter entre ces deux extrémités non voisées, même cette structure pourra être utile dans la segmentation automatique de la parole. Les structures des logatomes sont de la forme suivante:

- '**سَبَسْ**' /sabas/ et '**سَابَاسْ**' /sa:ba:s/ respectivement pour la voyelle courte fatha /a/ et la voyelle longue fatha /a:/.
- '**سُبُسْ**' /subus/ et '**سُوبُوسْ**' /su:bu:s/ respectivement pour la voyelle courte damma /u/ et la voyelle longue damma /u:/.
- '**سِبِسْ**' /sibis/ et '**سِيبِيسْ**' /si:bi:s/ respectivement pour la voyelle courte kasra /i/ et la voyelle longue kasra /i:/.

Ces structures sont spécifiques à la consonne ba '**ب**', donc il faut répéter le même processus pour les 27 consonnes restantes de la langue Arabe afin de construire tout le corpus de logatomes (voir Table10) et représenter tous les diphones de la langue Arabe.

Consonnes	Voyelle fatha		Voyelle damma		Voyelle kasra	
	courte	longue	courte	longue	courte	longue
Alif ‘ا’	سَأسْ	سَآسْ	سُؤُسْ	سُوؤُوسْ	سِئِسْ	سِيئِيسْ
Ba ‘ب’	سَبَسْ	سَابَاسْ	سُبُسْ	سُوبُوسْ	سِبِسْ	سِيبِيسْ
Ta ‘ت’	سَتَسْ	سَاتَاسْ	سُتُسْ	سُوتُوسْ	سِتِسْ	سِيتِيسْ
Tha ‘ث’	سَثَسْ	سَاثَاسْ	سُثُسْ	سُوثُوسْ	سِثِسْ	سِيثِيسْ
Jim ‘ج’	سَجَسْ	سَاجَاسْ	سُجُسْ	سُوجُوسْ	سِجِسْ	سِيجِيسْ
Hha ‘ح’	سَحَسْ	سَاحَاسْ	سُحُسْ	سُوحُوسْ	سِحِسْ	سِيحِيسْ
Kha ‘خ’	سَخَسْ	سَاخَاسْ	سُخُسْ	سُوخُوسْ	سِخِسْ	سِيخِيسْ
Dal ‘د’	سَدَسْ	سَادَاسْ	سُدُسْ	سُودُوسْ	سِدِسْ	سِيدِيسْ

Table10. Un échantillon de notre corpus de logatomes pour la langue Arabe.

IV.3.2 Enregistrement du corpus de logatomes:

L’étape d’enregistrement de notre corpus de logatomes Arabes a été effectuée malheureusement dans une chambre pas totalement sourde avec un locuteur ordinaire. Les enregistrements sonores ont été réalisés avec une fréquence d’échantillonnage de 16 kHz, et par l’utilisation d’un microphone dynamique avec une directivité cardioïde et une réponse en fréquence 80 Hz-15 kHz.

Le locuteur a été amené à respecter quelques conseils afin d’avoir des enregistrements sonores de bonne qualité, pour exemple la distance entre la bouche et le microphone de 2-5 cm lors de l’enregistrement des logatomes et de les prononcer avec le même rythme, la même intensité et avec une bonne élocution et d’une façon neutre.

Ces instructions que nous avons imposé ne sont pas toujours respectées par le locuteur ordinaire, ce qui engendre une dégradation de la qualité du signal de la parole synthétisée par l’utilisation de ces enregistrements sonores segmentés sous forme d’unités acoustiques de type diphone.

IV.3.2 Segmentation des enregistrements sonores:

La segmentation des enregistrements sonores a été faite de façon manuelle dans le but d’avoir plus de précision dans les unités acoustiques résultantes de la segmentation.

La segmentation manuelle (Figure32) a été caractérisée par trois (3) paramètres essentiels qui sont les suivants:

- La représentation temporelle du signal de parole.
- La représentation fréquentielle du signal de parole.
- L'écoute des enregistrements sonores.

Ces paramètres (essentiellement le paramètre d'écoute) ont été utilisés dans l'identification de la portion de la parole à segmenter.

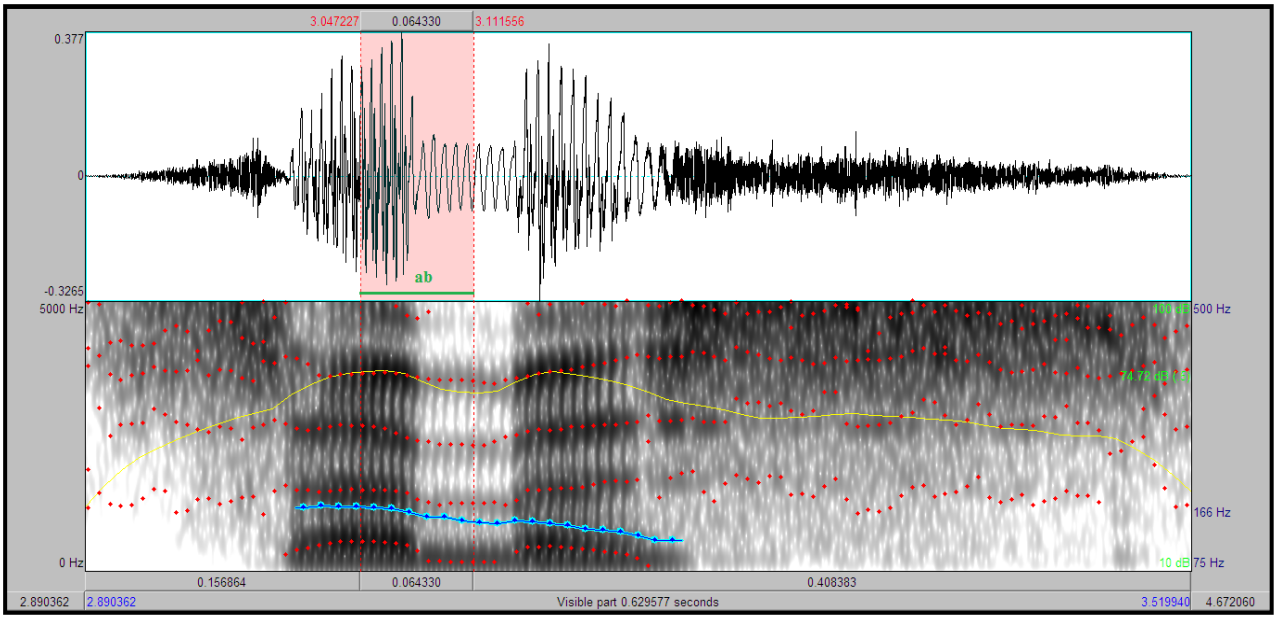


Figure32. Extraction du diphone /ab/ par la segmentation manuelle du son de logatome 'سَبَسْ' /sabas/.

Lors de la segmentation manuelle des enregistrements sonores nous avons essayé de respecter les durées citées ci-dessous, sauf si la consonne ou la voyelle enregistrée n'atteint pas la durée que nous avons imposé.

- Les durées des consonnes que nous avons imposées lors de la segmentation varient entre 25 ms à 30 ms dans un diphone et de 50 ms à 65 ms dans un phonème.
- Les durées des voyelles courtes que nous avons imposées lors de la segmentation varient entre 30 ms à 40 ms dans un diphone et de 35 ms à 40 ms dans un phonème.
- Les durées des voyelles longues que nous avons imposées lors de la segmentation varient entre 70 ms à 75 ms dans un diphone et de 70 ms à 80 ms dans un phonème.

Après avoir segmenté tous les enregistrements sonores, nous avons abouti à un ensemble de 336 diphones et 33 phonèmes, donc un total de 369 unités acoustiques.

IV.4 Synthèse de la parole:

Cette phase englobe tous les traitements de la parole du synthétiseur qui permettent de convertir les séquences phonétiques en un son de parole artificielle. Notre synthétiseur de la parole est basé sur la concaténation des diphones. Ces derniers subissent des modifications dans le domaine temporel au point de concaténation dans le but de réduire les discontinuités perceptibles au niveau de la région de concaténation afin d'améliorer la qualité de la parole synthétique générée.

Les modifications des unités acoustiques lors de la production de la parole artificielle ont été divisées en deux catégories selon la nature des sons à concaténer voisés ou non voisés (voir Table11).

Voisé (V) /Non Voisé (NV)	Graphème	Appellation	Phonème	Voisé (V) /Non Voisé (NV)	Graphème	Appellation	Phonème
NV	ء	Hamza	?	V	ع	Ayn	H
NV	ا أ إ ؤ ئ	Alif	?	V	غ	Ghayn	G
V	ب	Ba	b	NV	ف	Fa	f
NV	ت ة	Ta	t	NV	ق	Qaf	q
NV	ث	Tha	T	NV	ك	Kaf	k
V	ج	Jim	Z	V	ل	Lam	l
NV	ح	Hha	X	V	م	Mim	m
NV	خ	Kha	x	V	ن	Nun	n
V	د	Dal	d	NV	ه	Ha	h
V	ذ	Thal	D	V	و	Waw	w
V	ر	Ra	r	V	ي	Ya	j
V	ز	Zin	z	V	ـَ	Fatha	a
NV	س	Sin	s	V	ـُ	Damma	u
NV	ش	Shin	S	V	ـِ	Kasra	i
NV	ص	Sad	s`	V	آ ـَا ـَى	Longue fatha	a:
V	ض	Dad	d`	V	ـُو	Longue damma	u:
NV	ط	Ta	t`	V	ـِي	Longue kasra	i:
V	ظ	Zha	D`				

Table11. Classification des sons de la langue Arabe (voisé, non voisé).

IV.4.1 Sélection et chargement des unités acoustiques:

Avant toute modification des unités acoustiques à concaténer d'un synthétiseur de la parole, il faut toujours la précéder par l'étape de sélection et de chargement de ces dernières. Cette étape est primordiale et totalement basée sur des règles de sélection des unités acoustiques qui sont les diphones, puisque nous n'avons pas réalisé la syllabation des séquences phonétiques.

Les règles de sélection sont réalisées d'une manière à sélectionner les unités acoustiques de type diphone. Donc les règles de sélection varient selon le type d'unité acoustique utilisé lors de la synthèse de la parole.

La structure des règles élaborées se base totalement sur le contexte gauche et droit du son testé. Où chaque son testé obéit à la même structure de règle. Le chargement des sons appropriés se fait de manière automatique directement après la sélection des unités acoustiques. La structure des règles établies est de la forme suivante:

[D1] + [D2] = CG + S + CD

Avec:

D1: Le premier diphone.
D2: Le deuxième diphone.
CG: Contexte gauche du son testé.
S: Son testé.
CD: Contexte droit du son testé.

Cette forme de règle indique qu'un son **S**, précédé d'un son **CG**, et suivi d'un son **CD**, aura comme diphones **D1** et **D2**.

Afin d'éclaircir les règles de sélection établies nous donnons dans ce qui suit quelques exemples d'application de ces règles de sélection des unités acoustiques (diphones).

Exemple1: **[ib] + [ba] = 'i' + 'b' + 'a'**

Cet exemple indique que lorsque le son **'b'** (son de la consonne ba 'ب'), est précédé du son **'i'** (son de la voyelle courte kasra ' ـِ '), et suivi du son **'a'** (son de la voyelle courte fatha ' ـَ '), aura comme diphones **[ib] et [ba]**.

Exemple2: **[ut] + [ta:] = 'u' + 't' + 'a:'**

Cet exemple indique que lorsque le son **'t'** (son de la consonne ta 'ت'), est précédé du son **'u'** (son de la voyelle courte damma ' ـُ '), et suivi du son **'a:'** (son de la voyelle longue fatha ' ـَ '), aura comme diphones **[ut] et [ta:]**.

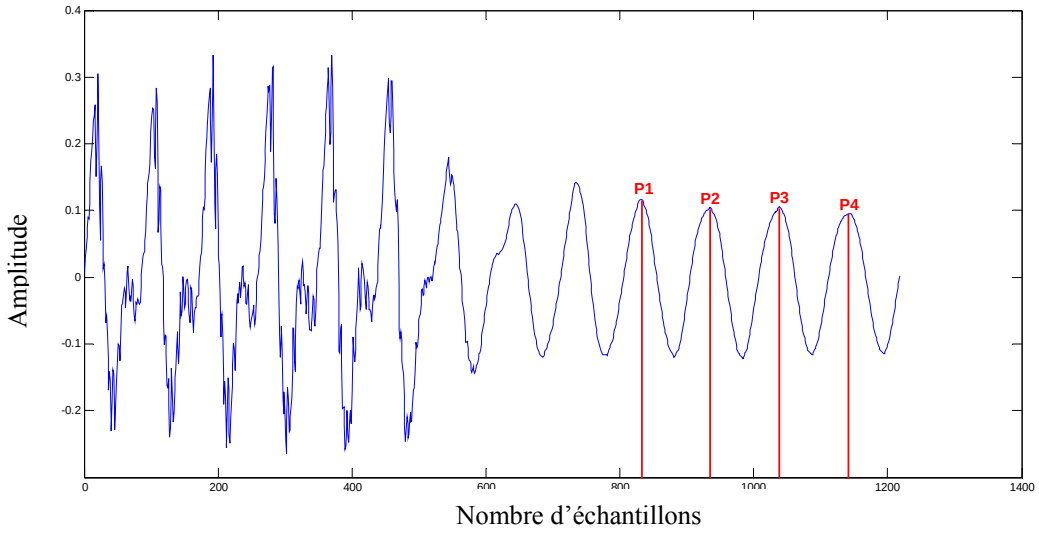
Après avoir sélectionné et chargé les bonnes unités acoustiques. Ces dernières subissent des traitements spécifiques selon le son à concaténer (voisé, non voisé).

IV.4.2 Sons voisés:

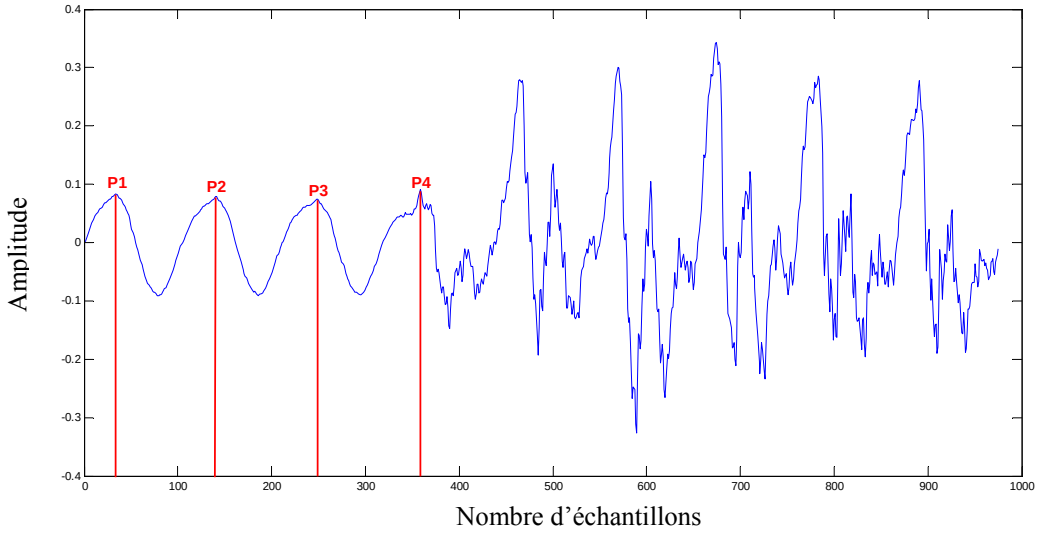
Les sons voisés à concaténer subissent uniquement des modifications au niveau de la fréquence fondamentale sans modifier la durée. Cette étape comporte deux phases de traitement (détection des marques de pitch et modification de la fréquence fondamentale).

La première phase consiste à détecter les marques de pitch des deux segments de parole à concaténer (voir Figure33) donc détecter que les quatre (4) dernières marques de pitch (trois (3) périodes) du premier segment de la parole voisée et les quatre (4) premières marques de pitch (trois (3) périodes) du deuxième segment de la parole voisée. La détection se fait en appliquant la démarche suivante:

- Appliquer la détection des marques de pitch que sur les dernières 25 ms du premier segment de parole et les premières 25 ms du deuxième segment de parole.
- Prendre un seuil de détection égal à 0.6 donc 60% de la plus grande valeur d'amplitude du segment de parole.
- Diviser la durée de 25 ms du segment de parole en deux segments de 12.5 ms chacun afin de définir deux seuils. Pour chaque segment de parole on extrait la plus grande valeur, cette démarche évite d'ignorer une vraie marque de pitch et donne à notre détecteur des marques de pitch une fiabilité.
- Détecter les pics pour les deux segments de parole de 12.5 ms en comparant la valeur d'échantillon testé avec la valeur qui le précède et qui le suit.
- Détecter les marques de pitch en comparant les valeurs des pics détectés du premier et du deuxième segment de parole avec leurs seuils appropriés.
- Corriger le résultat de la détection en comparant ces valeurs obtenues avec leurs valeurs qui les précèdent et qui les suivent. Si la fréquence entre les deux marques de pitch consécutives est supérieure à 640 Hz, il faut éliminer la plus petite valeur de la marque de pitch.
- Réappliquer le processus de la correction des marques de pitch jusqu'à éliminer toutes les fausses marques de pitch.



(a)



(b)

Figure33. Détection des marques de pitch des deux segments de parole à concaténer, (a) le premier segment de parole [ib], (b) le deuxième segment de parole [ba].

La deuxième phase consiste en la modification de la fréquence fondamentale (voir Figure34) en utilisant les marques de pitch déjà obtenues par la première phase. Cette phase utilise le principe de la méthode TD-PSOLA mais avec la différence de modifier uniquement les valeurs des trois (3) périodes dernières du premier segment et les valeurs des trois (3) périodes premières du deuxième segment. Le processus de la modification est le suivant:

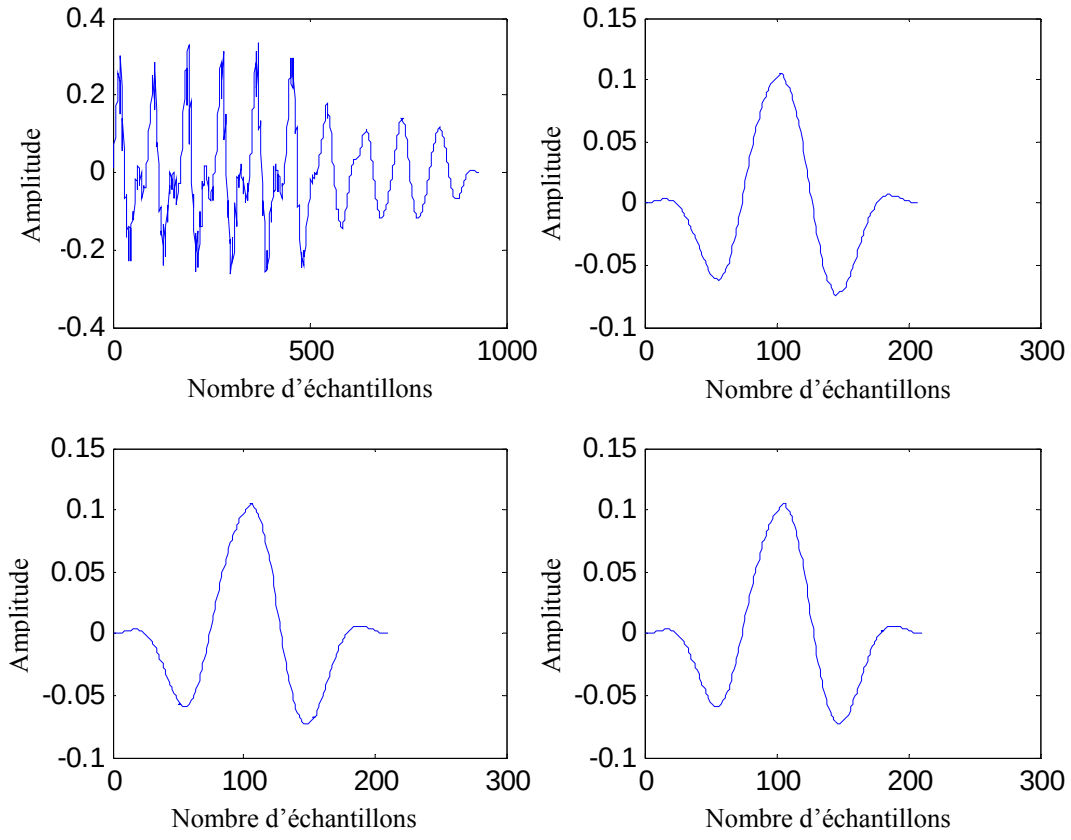
- Extraire les petits segments de parole de largeur deux fois la période du signal (2*T) en multipliant chaque trois (3) marques de pitch par une fenêtre de Hanning centrée sur la marque de pitch ciblée.

- Calcul des trois (3) périodes en utilisant les quatre (4) marques de pitch pour chaque segment de parole.

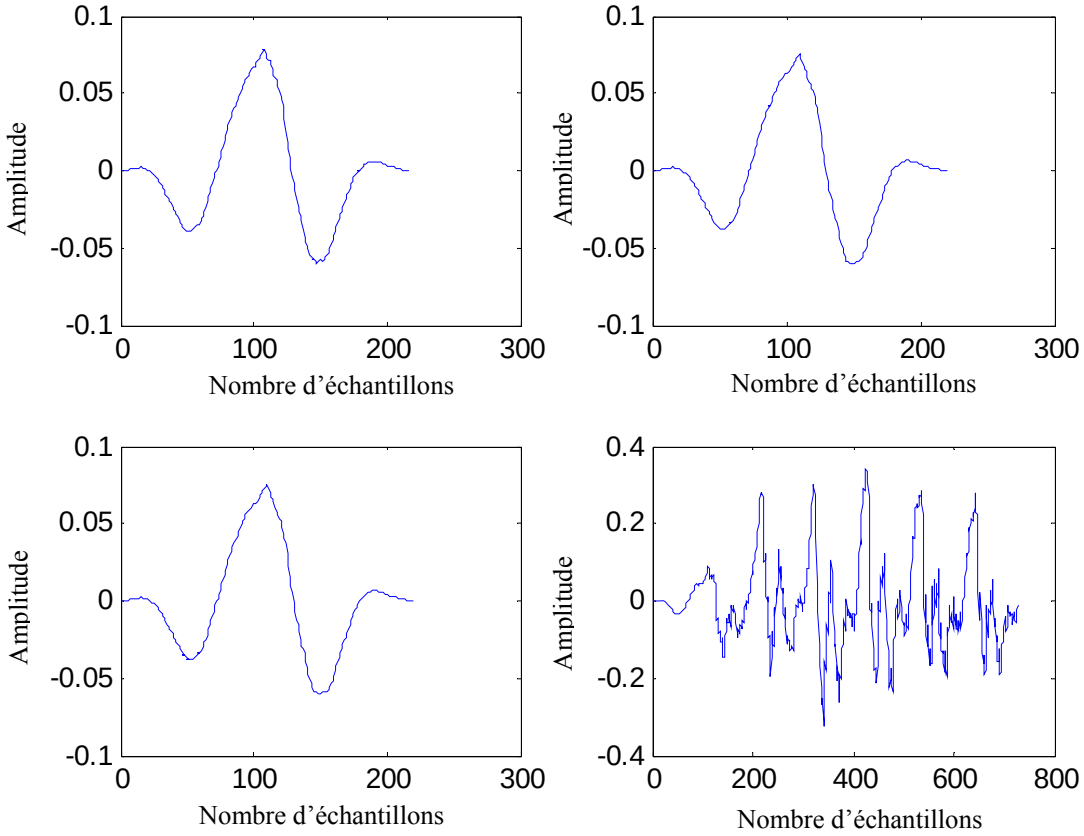
- Calcul de la période entre les deux (2) segments de parole à concaténer en calculant la valeur moyenne entre la dernière valeur de la période du premier segment et la première valeur de la période du deuxième segment.

- Calcul des nouvelles valeurs des périodes pour les deux (2) segments de parole en calculant la valeur moyenne entre la période cible et les deux (2) périodes qui la précèdent et qui la suivent. Après chaque calcul des nouvelles valeurs des périodes on fait une mise à jour aux valeurs de ces dernières afin de les réutiliser dans le calcul.

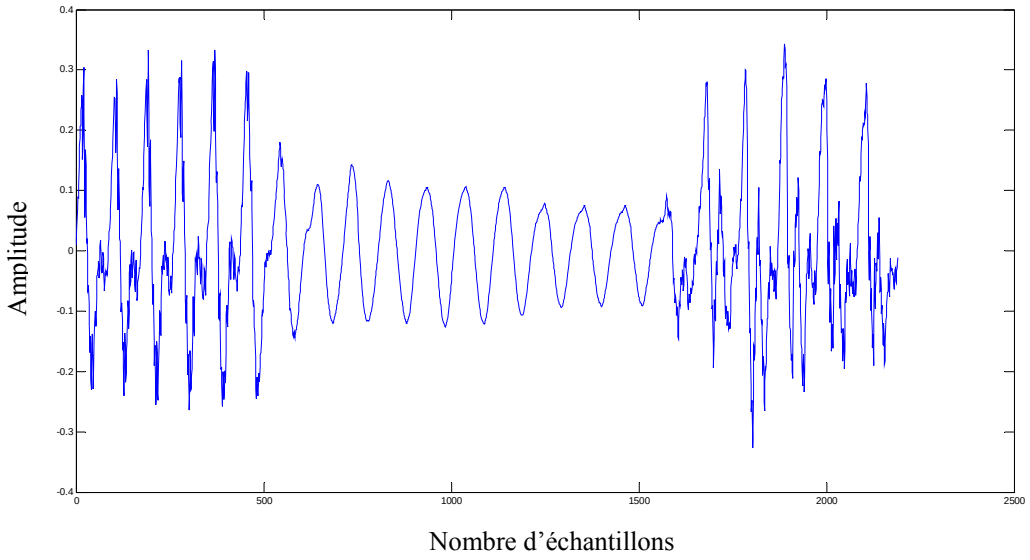
- La concaténation des petits segments de parole générés en introduisant les nouvelles valeurs des périodes calculées dans le but de modifier la fréquence fondamentale. Cette tâche est réalisée par le processus d'addition et recouvrement 'OLA'.



(a)



(b)



(c)

Figure34. Modification de la fréquence fondamentale, (a) extraction des petits segments de parole du son [ib], (b) extraction des petits segments de parole du son [ba], (c) concaténation des petits segments de parole par 'OLA' pour obtenir le son [iba].

IV.4.3 Sons non voisés:

En ce qui concerne les sons non voisés à concaténer. Nous les avons partagés en deux groupes de sons. Le premier groupe englobe les sons qui contiennent un silence dans leur réalisation acoustique qui sont de l'ordre de six (6) consonnes: alif 'ا', ta 'ت', tha 'ث', ta 'ط', qaf 'ق', kaf 'ك'. Les sons de ces dernières ont été concaténés directement (voir Figure35) en joignent les deux segments de parole sans aucun traitement spécifique, puisque les deux segments de parole contiennent à la fois un silence.

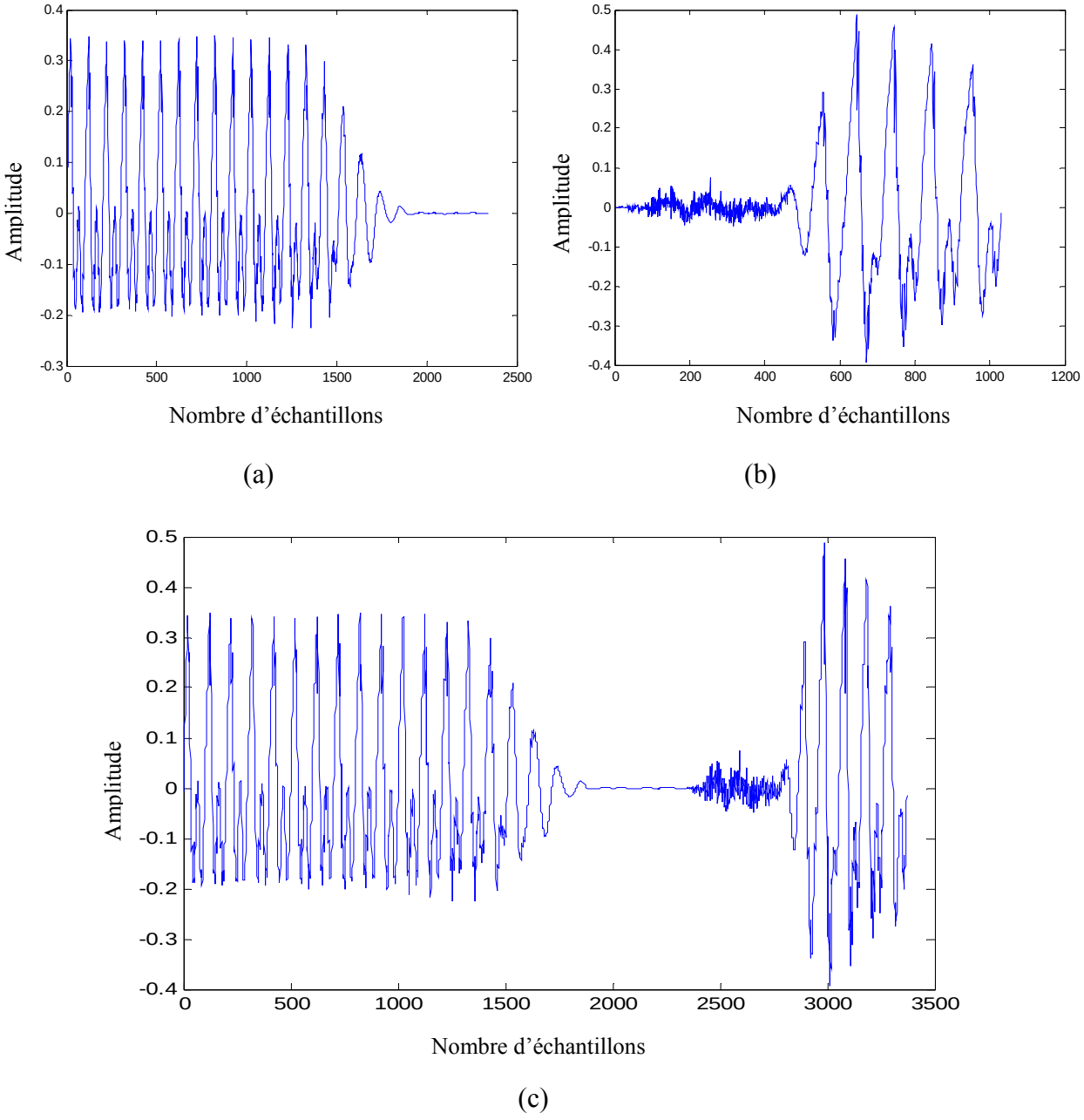
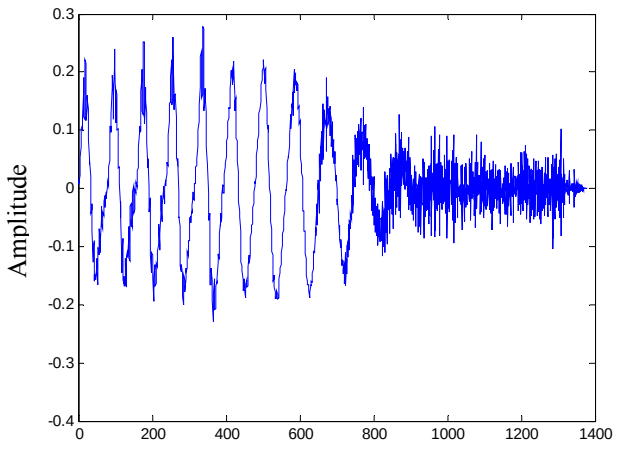


Figure35. Concaténation des sons non voisés, (a) le premier segment de parole [i:t], (b) le deuxième segment de parole [ti], (c) concaténation directe des deux segments de parole pour obtenir le son [i:ti].

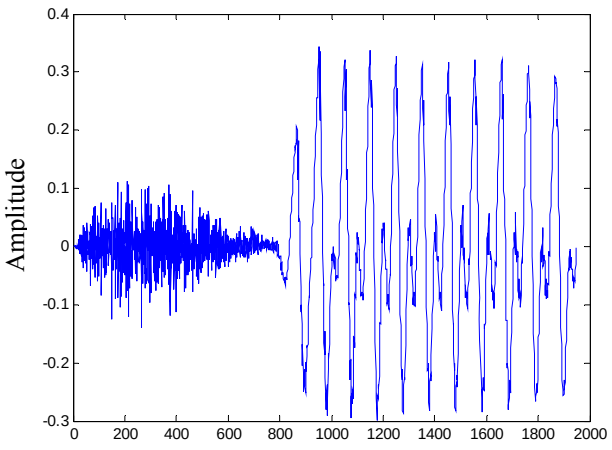
Le deuxième groupe contient le reste des sons non voisés qui ne rentrent pas dans le cadre du premier groupe de sons. Ces sons non voisés sont de l'ordre de sept (7) consonnes: hha '**ح**', kha '**خ**', sin '**س**', shin '**ش**', sad '**ص**', fa '**ف**', ha '**ه**'.

Afin de générer de la parole synthétique à partir de ces sons non voisés, nous avons procédé à un lissage temporel des extrémités aux points de concaténation (voir Figure36), du fait qu'ils ne présentent pas de structure périodique. La démarche suivie est simple à mettre en œuvre et elle est comme suit:

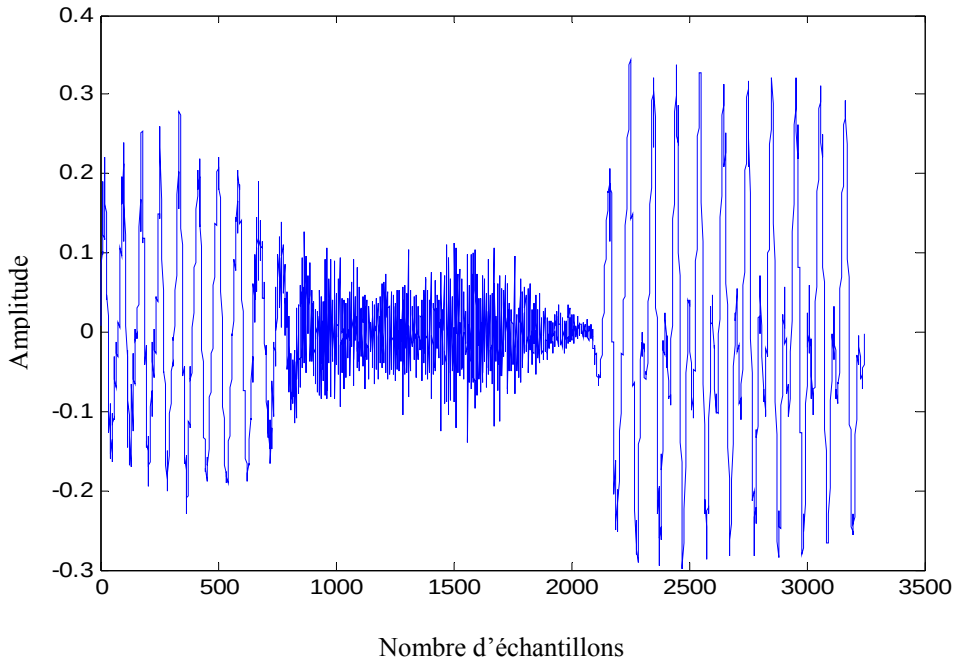
- Prendre une fenêtre de Hanning de largeur 10 ms. Cette dernière est devisée en deux (2) demi-fenêtres de largeur 5 ms. La première demi-fenêtre comporte les valeurs de la fenêtre de Hanning qui varient de la première valeur jusqu'à la valeur centrale de la fenêtre de Hanning. En outre La deuxième demi-fenêtre comporte les valeurs de la fenêtre de Hanning qui varient de la valeur centrale jusqu'à la dernière valeur de la fenêtre de Hanning.
- Multiplier les 5 ms dernières du premier segment de parole à concaténer par la deuxième demi-fenêtre.
- Multiplier les 5 ms premières du deuxième segment de parole à concaténer par la première demi-fenêtre.
- La concaténation des deux (2) segments de parole est réalisée par une addition et recouvrement 'OLA'.



(a)



(b)



(c)

Figure36. Concaténation des sons non voisés, (a) le premier segment de parole multiplié par une demi-fenêtre de Hanning [is], (b) le deuxième segment de parole multiplié par une demi-fenêtre de Hanning [si], (c) concaténation des deux segments de parole par 'OLA' pour obtenir le son [isi].

IV.5 Algorithme de la génération de la parole artificielle:

La démarche de la génération de la parole artificielle est basée sur des règles. De ce fait et vu la difficulté de comprendre le processus de la production de la parole synthétique, nous allons donner et détailler l'algorithme de la génération de la parole artificielle.

L'algorithme de la génération de la parole artificielle (voir Figure37) utilise les séquences phonétiques issues des traitements linguistiques afin de les faire correspondre aux sons à charger et à les traiter selon leur voisement ou non, pour que le signal de la parole généré soit d'une bonne qualité.

DEBUT

Pour tout i variant de 1 jusqu'à N
(N: La taille de la séquence phonétique)
{
// Traitement des pauses
Si le symbole testé est un espace en entre mots « _ »
Insérer un silence de 800 échantillons
Sinon Si le symbole testé est un espace entre consonne « - »
Insérer un silence de 400 échantillons

// Traitement des sons voisés (consonnes)
Sinon Si le son testé est un son voisé
Tester son contexte gauche et droit et charger les unités acoustiques appropriées
Détecter les quatre (4) dernières marques du pitch du premier segment de parole
Détecter les quatre (4) premières marques du pitch du deuxième segment de parole
Modifier les valeurs des trois (3) périodes des deux segments de parole (trois dernières périodes du premier segment et les trois premières périodes du deuxième segment)
Concaténer les segments de parole par une addition et recouvrement 'OLA'

// Traitement des sons non voisés (consonnes)
Sinon Si le son testé est un son non voisé
Si le son testé appartient au son des consonnes: alif 'ا', ta 'ت', tha 'ث', ta 'ط',qaf 'ق', kaf 'ك'
Tester son contexte gauche et droit et charger les unités acoustiques appropriées
Concaténer directement les deux segments de parole
Sinon
Tester son contexte gauche et droit et charger les unités acoustiques appropriées
Multiplier les 5 ms dernières du premier segment par une demi-fenêtre de Hanning
Multiplier les 5 ms premières du deuxième segment par une demi-fenêtre de Hanning
Concaténer les deux segments de parole par une addition et recouvrement 'OLA'

// Traitement des sons voisés à la fin du mot (voyelles)
Sinon Si le son testé est un son voisé à la fin du mot
Tester son contexte gauche et charger l'unité acoustique appropriée
Détecter les quatre (4) dernières marques du pitch du premier segment de parole
Détecter les quatre (4) premières marques du pitch du deuxième segment de parole
Modifier les valeurs des trois (3) périodes des deux segments de parole (trois dernières périodes du premier segment et les trois premières périodes du deuxième segment)
Concaténer les segments de parole par une addition et recouvrement 'OLA'
}

FIN

Figure37. Algorithme de la génération de la parole artificielle.

IV.6 Conclusion:

Ce chapitre a été consacré entièrement à la partie des traitements acoustiques du synthétiseur de la parole à partir du texte Arabe. Les traitements acoustiques effectués à ce stade ont pour objectif de générer un signal de la parole le plus proche possible de la voix humaine.

La création de la parole synthétique se fait à l'aide des séquences phonétiques générées au niveau des traitements linguistiques. Ces dernières sont utilisées pour sélectionner les unités acoustiques de type diphone à concaténer, qui se trouvent dans une base de données acoustique que nous avons créée à partir de l'enregistrement et la segmentation manuelle d'un corpus de logatomes pour la langue Arabe.

Dans ce chapitre nous avons détaillé les différents traitements de la parole du synthétiseur de la langue Arabe afin d'aboutir à la production de la parole artificielle à partir de n'importe quelle séquence phonétique. Ces traitements peuvent se résumer par la sélection des bonnes unités acoustiques à concaténer, suivis par les différents traitements réalisés au niveau de ces dernières selon le type de son (voisé et non voisé), et enfin la concaténation de ces segments de parole par addition et recouvrement 'OLA'.

.

Chapitre V: Evaluation:

V.1 Introduction:

La phase d'évaluation de n'importe quel système de la synthèse de la parole à partir du texte est une étape très importante, même primordiale afin de juger la fiabilité et la qualité de la synthèse de la parole à partir du texte.

Les critères de base pour mesurer les performances d'un système de la synthèse de la parole à partir du texte (Text-To-Speech) peuvent être énumérés comme la similitude avec la voix humaine (le naturel) et la capacité à être compris (l'intelligibilité). Le synthétiseur vocal doit maximiser ces deux critères d'évaluation.

L'évaluation d'un système de la synthèse de la parole à partir du texte peut se faire par diverses méthodes, en particulier par l'évaluation de la compréhension. Cette dernière est un moyen valable pour évaluer l'intelligibilité, les auditeurs mettront l'accent sur la reconnaissance des mots individuels, plutôt que de se concentrer sur le sens des phrases. En ce qui concerne l'évaluation du naturel de la voix synthétisée, elle se fait par le biais de la note d'opinion moyenne (MOS). Cette dernière est la mesure subjective la plus populaire et largement la plus utilisée pour l'évaluation du naturel des systèmes de TTS.

Pour notre part cette évaluation a été partagée en trois étapes à cause de la liaison qui se trouve entre les différents étages du synthétiseur de la parole à partir du texte Arabe (la partie linguistique et la partie acoustique), donc un bon résultat sur le plan acoustique est lié entièrement à un bon résultat sur le plan linguistique et inversement.

Les trois étapes de l'évaluation de notre synthétiseur de la parole à partir du texte Arabe sont comme suit:

- Evaluation de la transcription phonétique du texte Arabe.
- Evaluation de la génération des allophones.
- Evaluation de la qualité de signal de la parole synthétisée.

Dans ce qui suit nous allons détailler la démarche que nous avons suivie afin d'évaluer notre synthétiseur de la parole à partir du texte Arabe que soit sur le plan linguistique ou que soit sur le plan acoustique ainsi que les résultats trouvés pour les différentes étapes d'évaluation.

V.2 Evaluation de la transcription phonétique du texte Arabe:

Afin de tester notre système de phonétisation automatique nous avons pris un corpus de phrases Arabes réalisé par Mansour M. Al-ghamdi [Alg04*], ce corpus contient une liste de 367 phrases. Ces dernières ont été réalisées en combinant une liste de 229 mots riches et divers phonétiquement extraits d'un corpus de mots [Alg97] déjà réalisé.

La liste de 367 phrases est composée de 1835 mots, et de 12940 graphèmes. Le nombre de mots par phrase est de l'ordre de 2 à 9 mots par phrase.

Pour la vérification de la transcription phonétique générée automatiquement par notre système [Ime12], [Ime13], nous avons procédé à une transcription des 367 phrases manuellement afin de valider les résultats obtenus.

Les résultats obtenus (Table12) sont très encourageants puisque nous avons pu transcrire les 367 phrases correctement avec un taux de réussite égal à 100%. D'après Mansour M. Al-ghamdi ce corpus de phrases Arabe se caractérise par la richesse, l'équilibre et la diversité phonétique. Si un système de synthèse de la parole à partir du texte Arabe prononce correctement ce corpus de phrase Arabe, donc il pourra prononcer correctement n'importe quel autre texte Arabe, de ce fait si la synthèse est bonne donc la transcription phonétique est aussi bonne, à cet effet on peut dire que notre système de phonétisation automatique pourra transcrire correctement n'importe quel autre texte Arabe.

N°	Phrases Arabes	Transcription manuelle selon la notation **SAMPA**	Transcription automatique selon la notation **SAMPA**	Taux de réussite
1	مِنْ بَخْسِ نِعْمَةِ اللهِ دَفْنُهَا	min_bax-si_niH-mati_ lla:hi_daf-nuha:	min_bax-si_niH-mati_ lla:hi_daf-nuha:	100%
2	أبْجَلَنِي هَذَا الطَّعَامُ	?ab-Zalani:_ha:Da_ t`t`aHa:m	?ab-Zalani:_ha:Da_ t`t`aHa:m	100%
3	الأعْشَى الشَّاعِرُ مِنْ أدْهَى الشُّعَرَاءِ	?al-?aH-Sa_SSa:Hiru_ min_ ?ad-ha_ SSuHara:?	?al-?aH-Sa_SSa:Hiru_ min_?ad-ha_SSuHara:?	100%
4	مِنْ أضْأَلِ الطَّعَامِ رَغِيفُ القَمْحِ	min_?ad`-?ali_t`t`aHa:mi_ raGi:fu_l-qam-X	min_?ad`-?ali_t`t`aHa:mi_ raGi:fu_ l-qam-X	100%
5	رَأى النَّائِمُ أضْغَاثَ أحْلاَمٍ	ra?a_nna:?imu_?ad`-Ga:Ta_?aX-la:m	ra?a_nna:?imu_?ad`-Ga:Ta_?aX-la:m	100%
6	أضْهَى الرَّجُلُ إبِلَهُ	?ad`-ha_rraZulu_?ibilah	?ad`-ha_rraZulu_?ibilah	100%
7	كَانَ حَاتِمُ الطَّائِيُّ يُكْرِمُ الأضْيَافَ	ka:na_Xa:timu_t`t`a:?ijju_ juk-rimu_ l-?ad`-ja:f	ka:na_Xa:timu_t`t`a:?ijju_ juk-rimu_ l-?ad`-ja:f	100%
8	الأفْغَانُ مُقَاتِلُونَ أفْذَاذٌ	?al-?af-Ga:nu_ muqa:tilu:na_?af-Da:D	?al-?af-Ga:nu_ muqa:tilu:na_?afDa:D	100%
9	إنْطَلَقَتْ أفْوَاجُ الْحُجَّاجِ فِي جَوٍّ أغْبَرَ	?in-t`alaqat_?af-wa:Zu_ l-XuZZa:Zi_ fi:_Zawwin_ ?aG-bar	?in-t`alaqat_?af-wa:Zu_ l-XuZZa:Zi_ fi:_Zawwin_ ?aG-bar	100%
10	اللَّبَنُ مِنْ أغْنَى الأغْذِيَةِ	?a-llabanu_min_?aG-na_ l-?aG-Dijah	?a-llabanu_min_?aG-na_ l-?aG-Dijah	100%

11	أَحَطْتُ بِأَضْدَادِ هَذِهِ الْكَلِمَةِ	?aXat`-tu_bi?ad`-da:di_ha:Dihi_l-kalimah	?aXat`-tu_bi?ad`-da:di_ha:Dihi_l-kalimah	100%
12	يَعِيشُ كَثِيرٌ مِنَ الأَشْخَاصِ فِي أَحْزَانٍ	jaHi:Su_kaTi:run_mina_l-?aS-xa:s`i_ fi:_?aX-za:n	jaHi:Su_kaTi:run_mina_ l-?aS-xa:s`i_fi:_?aX-za:n	100%
13	حِينَمَا تَثُورُ الأَشْجَانُ تَتَغَيَّرُ الأَمْزِجَةُ	Xi:nama:_taTu:ru_l-?aS-Za:nu_tataGajjaru_l-?am-ziZah	Xi:nama:_taTu:ru_l-?aS-Za:nu_tataGajjaru_l-?am-ziZah	100%
14	أَحْضَرْتُ مِنْ الشَّارِقَةِ أَشْيَاءَ لِلْبَيْعِ	?aX-d`ar-tu_min_ SSa:riqati_?aS-ja:?a_ lil-baj-H	?aX-d`ar-tu_min_ SSa:riqati_?aS-ja:?a_ lil-baj-H	100%
15	أَكَلْتُ بَيْضَ الدَّجَاجِ	?akal-tu_baj-d`a_ddaZa:Z	?akal-tu_baj-d`a_ddaZa:Z	100%
16	فِي هَذِهِ الأَكْوَابِ أَمْشَاجٌ مِنَ الشَّرَابِ	fi:_ha:Dihi_l-?ak-wa:bi_ ?am-Sa:Zun_ mina_ SSara:b	fi:_ha:Dihi_l-?ak-wa:bi_ ?am-Sa:Zun_mina_ SSara:b	100%
17	أَمَّا فِعْلُ الْخَيْرِ فَحَسَنٌ	?amma:_fiH-lu_l-xaj-ri_ faXasan	?amma:_fiH-lu_l-xaj-ri_ faXasan	100%
18	فِي الْمُسْتَشْفَى أَمْصَالٌ ضِدَّ الأَمْرَاضِ	fi_l-mus-taS-fa:_?am-s`a:lun_d`idda_l-?am-ra:d`	fi_l-mus-taS-fa:_?am-s`a:lun_ d`idda_l-?am-ra:d`	100%
19	نَسْمَعُ الأَنْغَامَ عِنْدَ عَزْفِ الْمَعَازِف	nas-maHu_l-?an-Ga:ma_ Hin-da_ Haz-fi_ l-maHa:zif	nas-maHu_l-?an-Ga:ma_ Hin-da_Haz-fi_ l-maHa:zif	100%
20	أَنْذَرَ الرَّجُلُ صَدِيقَهُ بِوَقْفِ مُسَاعَدَتِهِ	?an-Dara_rraZulu_ s`adi:qahu_biwaq-fi_ musa:Hadatih	?an-Dara_rraZulu_ s`adi:qahu_biwaq-fi_ musa:Hadatih	100%

Table12. Résultats de la transcription phonétique pour une liste des 20 premières phrases de notre corpus de 367 phrases Arabe.

V.3 Evaluation de la génération des allophones:

Pour tester notre système de génération des allophones [Ime15], nous avons pris une liste d'une vingtaine (20) de phrases d'un corpus de 367 phrases Arabes [Alg04*]. Ces dernières ont été transcrites automatiquement par notre système de transcription phonétique avec succès.

Les résultats de la génération des différentes réalisations acoustiques (allophones) sont donnés par la Table13.

N°	Phrases Arabes	Génération des allophones
1	مِنْ بَحْس نِعْمَةِ اللهِ دَفْنُهَا	mìn_bAx-s`I_nìH-màti_ lla:hi_daf-nùha:
2	أبْجَلَنِي هَذَا الطَّعَامُ	?ab-Zalànì:_ha:Da_t`t`AHà:m
3	الأعْشَى الشَّاعِرُ مِنْ أدْهَى الشُّعَرَاءِ	?al-?aH-Sa_SSa:HIrU_ mìn_?ad-ha_SSuHArA:?
4	مِنْ أضْأل الطَّعَام رَغِيفُ القَمْح	mìn_?Ad`-?ali_t`t`AHà:mì_ rAGI:fu_L-qAm-X
5	رَأى النَّائِمُ أضْغَاثَ أحْلاَمٍ	rA?a_nnà:?ìmù_?Ad`-GA:Ta_ ?aX-là:m
6	أضْهَى الرَّجُلُ إبِلَهُ	?Ad`-ha_rrAZulu_?ibilah
7	كَانَ حَاتِمُ الطَّائِيُّ يُكْرِمُ الأضْيَافَ	kà:nà_Xa:tìmù_t`t`A:?ijju_ juk-Rìmù_l-?Ad`-ja:f
8	الأفْغَانُ مُقَاتِلُونَ أفْذاذٌ	?al-?af-GA:nù_ mUqA:tilù:nà_ ?af-Da:D
9	إنْطَلَقَتْ أفْوَاجُ الْحُجَّاجِ فِي جَوٍّ أغْبَرَ	?ìn-t`AlAqAt_?af-wa:Zu_ l-XuZZa:Zi_fi:_Zawwìn_ ?AG-bAr
10	اللَّبَنُ مِنْ أغْنَى الأغْذِيَةِ	?a-llabànù_mìn_?AG-nà_ l-?AG-D`Ijah
11	أحَطْتُ بِأضْدَادِ هَذِهِ الْكَلِمَةِ	?aXAt`-t`U_bi?Ad`-d`A:di_ ha:Dihi_l-kalìmàh
12	يَعِيشُ كَثِيرٌ مِنَ الأشْخَاصِ فِي أحْزَانٍ	jaHi:Su_kaTI:rUn_mìnà_ l-?aS-xA:s`I_ fi:_?aX-zà:n
13	حِينَمَا تَثُورُ الأشْجَانُ تَتَغَيَّرُ الأمْزِجَةُ	Xì:nàmà:_taTU:rU_ l-?aS-Zà:nù_ tatAGAjjArU_ l-?àm-ziZah
14	أحْضَرْتُ مِنْ الشَّارِقَةِ أشْيَاءَ لِلْبَيْعِ	?aX-d`Ar-tu_mìn_ SSa:RIqAti_?aS-ja:?a_lil-baj-H
15	أكَلْتُ بَيْضَ الدَّجَاجِ	?akal-tu_baj-d`A_ddaZa:Z
16	فِي هَذِهِ الأكْوَابِ أمْشَاجٌ مِنَ الشَّرَابِ	fi:_ha:Dihi_l-?ak-wa:bi_ ?àm-Sa:Zùn_ mìnà_SSArA:b
17	أمَّا فِعْلُ الْخَيْرِ فَحَسَنٌ	?àmmà:_fiH-lu_L-xAj-Ri_ faXasàn
18	فِي الْمُسْتَشْفَى أمْصَالٌ ضِدَّ الأمْرَاضِ	fi_l-mùs-taS-fa:_?àm-s`A:lùn_ d`Idda_ l-?àm-rA:d`
19	نَسْمَعُ الأنْغَامَ عِنْدَ عَزْفِ الْمَعَازِف	nàs-màHu_l-?àn-GA:mà_ Hìn-da_Haz-fi_l-màHa:zif

20	أنْذَرَ الرَّجُلُ صَدِيقَهُ بوَقْفِ مُسَاعَدَتِهِ	?àn-DArA_rrAZulu_ s`AdI:qAhu_biwAq-fi_ mùsa:Hadatih

Table13. Résultats de la génération des allophones pour une liste de 20 phrases Arabes.

V.4 Evaluation de la qualité de signal de la parole synthétisée:

Notre synthétiseur de la parole à partir du texte Arabe a été testé sur une base de dix (10) phrases en langue Arabe, extraite d'un corpus de 367 phrases Arabes [Alg04*]. Ce dernier se caractérise par la richesse, l'équilibre et la diversité phonétique. Ce corpus de phrases Arabes a été conçu spécialement pour évaluer les synthétiseurs de la parole à partir du texte.

Ces dix (10) phrases Arabes ont été transcrites automatiquement à travers notre système de transcription phonétique [Ime14], afin de générer les séquences phonétiques qui seront utilisées pour la production du signal de parole synthétique.

Afin de tester notre synthétiseur de la parole à partir du texte Arabe [Ime15*], nous avons procédé à une évaluation de l'intelligibilité de la synthèse de la parole en faisant écouter quatre (4) auditeurs (deux (2) hommes et deux (2) femmes) ces dix (10) phrases Arabes, en leur demandant de répéter les phrases Arabes qu'ils ont écoutées les unes après les autres. Les taux de réussite moyens pour les dix (10) phrases Arabes synthétisées par notre synthétiseur de la parole, sont de l'ordre de 86% pour les hommes et de 90% pour les femmes avec un taux total de réussite des phrases Arabes synthétisées de 88% pour l'ensemble des auditeurs.

Nous avons aussi évalué la qualité de la voix synthétisée par une note d'opinion moyenne (MOS) pour l'ensemble des dix (10) phrases Arabes synthétisées. La note peut varier entre un (1) (très mauvaise) et cinq (5) (excellente, comparable à la version d'origine) et elle est comme suit:

- '1' Très mauvaise.
- '2' Mauvaise.
- '3' Juste.
- '4' Bonne.
- '5' Excellente.

Les résultats obtenus sont encourageants mais il reste à régler quelques problèmes de non reconnaissance par les auditeurs de quelques sons de consonnes synthétisées, dus à une mauvaise élocution lors des enregistrements, ainsi qu'à la durée des sons insuffisante à les identifier. Ces problèmes peuvent être résolus à l'avenir afin d'améliorer la qualité de la voix synthétisée et d'atteindre un taux total de réussite supérieur à 95%. Les résultats détaillés de ces évaluations sont illustrés dans la Table14 et dans la Table15.

N°	Phrases Arabes synthétisées	Taux de réussite %	
		Pour les hommes	Pour les femmes
1	مِنْ بَخْسِ نِعْمَةِ اللهِ دَفْنُهَا	100	100
2	أَبْجَلَنِي هَذَا الطَّعَامُ	100	100
3	الأَعْشَى الشَّاعِرُ مِنْ أَدْهَى الشُّعَرَاءِ	69	71
4	مِنْ أَضْأَلِ الطَّعَامِ رَغِيفُ القَمْحِ	82	86
5	رَأَى النَّائِمُ أَضْغَاثَ أَحْلاَمٍ	70	72
6	أَضْهَى الرَّجُلُ إِبِلَهُ	61	78
7	كَانَ حَاتِمُ الطَّائِيُّ يُكْرِمُ الأَضْيَافَ	93	94
8	الأَفْغَانُ مُقَاتِلُونَ أَفْذَاذٌ	92	95
9	إِنْطَلَقَتْ أَفْوَاجُ الْحُجَّاجِ فِي جَوٍّ أَغْبَرَ	95	99
10	اللَّبَنُ مِنْ أَغْنَى الأَغْذِيَةِ	100	100

Table14. Les Résultats des taux de réussite pour les dix (10) phrases Arabes synthétisées.

Auditeurs	MOS	MOS	MOS
1er homme	3	3	3
2ème homme	3		
1ère femme	3	3	
2ème femme	3		

Table15. Note d'opinion moyenne (MOS) pour l'ensemble des dix (10) phrases Arabes synthétisées.

V.5 Conclusion:

Dans ce chapitre nous avons traité la phase d'évaluation de notre système de la synthèse de la parole à partir du texte Arabe. Les résultats de la transcription phonétique du texte Arabe on été satisfaisants mais notre système de phonétisation automatique pour la langue Arabe sera aussi testé sur d'autres corpus de phrases, afin de valider la fiabilité et la robustesse de notre système de transcription phonétique.

La démarche que nous avons suivie pour évaluer la qualité de la parole synthétisée a donné des résultats acceptables du fait que nous avons testé notre synthétiseur de la parole à partir du texte sur un corpus constitué de quelques mots difficiles à prononcer et peu utilisables dans la vie quotidienne ce qui a limité la qualité associée par l'évaluation.

Il reste à résoudre quelques problèmes afin d'améliorer la qualité du signal de la parole synthétisée et tester notre synthétiseur de la parole à partir du texte sur un corpus plus large de phrases Arabe afin de confirmer et valider les résultats obtenus.

CONCLUSION GENERALE

Le thème de notre thèse de Doctorat traite à la fois l'ensemble des traitements linguistiques, ainsi que des traitements acoustiques destinés à la synthèse de la parole de la langue Arabe. Ce thème de recherche a été abordé afin de faire avancer les travaux de recherche dans le domaine de la synthèse de la parole à partir du texte dédiée à la langue Arabe.

Les traitements linguistiques traités dans cette thèse ont comme but de convertir un texte en une séquence phonétique et utiliser ainsi ces symboles phonétiques pour produire de la parole synthétique. Cette conversion a été faite à travers l'utilisation à la fois d'une base de règles de transcription phonétique et d'un lexique des exceptions pour les mots qui ne sont pas pris en charge par ces règles de transcription. La démarche que nous avons suivie au cours des traitements linguistiques afin d'aboutir à la transcription phonétique du texte Arabe a donné de bons résultats puisque nous avons pu transcrire correctement n'importe quel texte Arabe.

Dans le but d'améliorer la qualité de la parole synthétisée nous avons opté pour la génération des différentes réalisations acoustiques (les allophones) en utilisant une méthode simple à mettre en œuvre qui est basé sur des règles de transcription phonème-allophone.

La phase des traitements acoustiques abordée dans le quatrième chapitre est liée entièrement à la phase des traitements linguistiques puisque elle utilise les séquences phonétiques générées par cette dernière pour produire de la parole synthétique. Dans cette phase nous avons choisi la technique de la synthèse de la parole la plus utilisée et la plus simple à implémenter, qui est la synthèse de la parole par concaténation d'unité acoustique, pour notre part, nous avons choisi les unités acoustiques de type diphone.

La stratégie adoptée pour produire de la parole synthétique est basée uniquement sur la modification des segments de la parole dans le domaine temporal, en modifiant la fréquence fondamentale des deux sons voisés à concaténer et en appliquant un lissage temporel des extrémités des deux sons non voisés à concaténer. Cette méthode de la synthèse de la parole à donné des résultats acceptables vu la qualité des enregistrements sonores réalisés ainsi que les traitements de la parole effectués à ce stade.

La phase de test de notre synthétiseur de la parole à partir du texte Arabe a été divisée en deux grandes parties. La première partie est consacrée à l'étape d'évaluation des traitements linguistiques. Tandis que la deuxième partie est dédiée à l'étape d'évaluation des traitements acoustiques.

Les résultats trouvés sont très encourageant, que ce soit sur le plan linguistique où nous avons généré des séquences phonétiques correctement à partir des séquences de graphèmes de la langue Arabe, ou que ce soit sur le plan acoustique où nous avons produit de la parole synthétique d'une qualité acceptable à partir d'une base de données acoustique de type diphone.

Au terme de ce travail de recherche axé sur la synthèse de la parole à partir du texte Arabe nous pouvons affirmer que les objectifs ont été atteints, puisque nous avons abouti à synthétiser de la parole à partir de n'importe quel texte Arabe et cela en utilisant notre système de synthèse. Cependant, il est évident, que des améliorations peuvent y être apportées et qui pourront être intégrées dans nos travaux de recherche future, tel que:

- Intégrer la prosodie dans notre système de synthèse.
- Intégrer l'étage de la voyéllation automatique du texte Arabe.
- Intégrer de nouveau mots dans le lexique des exceptions.
- Traitement des chiffres, dates, heures,…
- Augmenter le dictionnaire acoustique (triphones, mots enregistrés, …).
- Appliquer des traitements spécifiques sur les segments de la parole dans le domaine fréquentiel.
- Enregistrement de notre base de données acoustique par un locuteur professionnel et dans des conditions idéales.

BIBLIOGRAPHIE

[Alg97] M. Alghamdi, M. S. Basalamah, M. Alsini, S.A. Hussein, "Database of Arabic Sounds: Words," Proceedings of the 15th National Computer Conference, 1997. (In Arabic).

[Alg02] M. Al-ghamdi, M. Elshafei, H. Al-muhtasib, "Arabic Text-To-Speech: Speech Units," Proceeding of the 4th Workshop on Computer and Information Sciences, pp. 199-212, 2002.

[Alg04] M. Al-ghamdi, H. Al-muhtaseb, M. Elshafei, "Phonetic Rules in Arabic Script," King Saud University Journal: Computer Sciences and Information, vol. 16, pp. 1–25, 2004.

[Alg04*] M. Alghamdi, A. H. Alhamid, M. M. Aldasuqi, "Database of Arabic Sounds: Sentences," Technical Report, King Abdulaziz City of Science and Technology, 2003. (In Arabic).

[Als09] G. Al-Said and M. Abdallah, "An Arabic Text-To-Speech System Based on Artificial Neural Networks," Journal of Computer Science, vol. 5 (3), pp. 207–213, 2009.

[Bal03] S. Baloul, "Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé," Thèse de Doctorat, Université du Maine, Mai, 2003.

[Bou01] P. Boula de Mareüil, P. Célérier, T. Cesses, S. Fabre, C. Jobin, P.Y Le Meur, D. Obadia, B. Soulage, J. Toen, "Elan Text-To-Speech: un système multilingue de synthèse de la parole à partir du texte," Traitement Automatique des Langues, vol. 42, pp. 1–30, 2001.

[Bou05] F. Boukadida, N. Ellouze "Modélisation Statistique de la Durée des Voyelles en Parole Arabe," 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, March, 2005.

[Bra06] D. Braga, L. Coelho, F.G Vianna Resende, "A Rule-Based Grapheme-to-Phone Converter for TTS Systems in European Portuguese," International Telecommunications Symposium , pp. 328–333, 2006.

[Cha10] A. Chabchoub, A. Cherif, "Implementation of the Arabic speech synthesis with TD-PSOLA modifier," International journal of signal system control and engineering application, vol. 3 (4), pp. 77–80, 2010.

[Che01] A. Cherif, L. Bouafif, T. Dabbabi, "Pitch detection and formant analysis of Arabic speech processing," Elsevier: Applied Acoustics, vol. 62, pp. 1129–1140, 2001.

[Cro00] O. Crouzet, "Segmentation de la parole en mots et régularités phonotactiques: Effets phonologiques, probabilistes ou lexicaux ?," Thèse de Doctorat, Université Paris 5 - Rene Descartes, Décembre, 2000.

[Div08] M. Divay, E. Bruckert, "Text-to-speech formant synthesis for french," in Human Factors and Voice Interactive Systems, G. Bonneau et al., Ed. Springer US, vol. 498, pp. 381-416, 2008.

[Dut00] T. Dutoit, " Introduction au Traitement Automatique de la Parole," Faculté Polytechnique de Mons, Première édition, 2000.

[Dut00*] T. Dutoit, "Je parle, donc je suis ?," Un bilan des développements récents en traitement automatique de la parole, Faculté Polytechnique de Mons, 2000.

[Eke02] B. Eker, "Turkish text to speech system," Master's thesis, Bilkent University, January, 2002.

[Elb11] H. M. El-bakry, M. Z. Rashad, I. R. Isma'il, "Diphone based concatenative speech synthesis systems for Arabic language," in 10th WSEAS international conference on circuits, systems, electronics, control & signal processing, and the 7th WSEAS international conference on applied and theoretical mechanics, pp. 81–86, 2011.

[Eli89] Y. A. El-imam, "Unrestricted Vocabulary Arabic Speech Synthesis System," IEEE Transactions on Acoustic Speech and Signal Processing, vol. 37, pp. 1829-1845, 1989.

[Eli00] Y.A. El-Imam, Z.M Don, "Text-to-Speech Conversion of Standard Malay," International journal of speech technology, vol. 3, pp. 129–146, 2000.

[Eli04] Y.A El-imam, "Phonetization of Arabic: rules and algorithms," Computer Speech and Language, vol. 18, pp. 339–373, 2004.

[Els91] M.A. Elshafei, "Toward an Arabic text-to-speech system," The Arabian Journal of Science and Engineering, vol. 16, pp. 565–583, 1991.

[Els02] M. Elshafei, H. Al-muhtaseb, M. Al-ghamdi, “Techniques for high quality Arabic speech synthesis,” Elsevier: Information Sciences, vol. 140, pp. 255−267, 2002.

[Gha90] S. Ghazali, H. Habaili, M. Zrigui, “Correspondance graphème-phonème pour la synthèse de la parole arabe à partir du texte,” IRSIT Congrès dialogue homme machine, 1990.

[Ham11] M. Hamad, M. Hussain, “Arabic text-to-speech synthesizer,” in IEEE student conference on research and development, pp. 409–414, 2011.

[Har10] A. Harrag, “QSDAS: New quranic speech database for Arabic speaker recognition,” Arabian Journal of Science and Engineering. vol. 35, pp. 7–19, 2010.

[Ime12] F. Imedjdouben, A. Houacine, “Outil de transcription phonétique à partir du texte Arabe,” 11th African Conference on Research in Computer Science and Applied Mathematics, pp. 475–482, 2012

[Ime13] F. Imedjdouben, A. Houacine, “Automatic Phonetization of Arabic Text,” 4th International Conference on Computer Science and its Applications, Saida, Algeria, pp. 85–94, May, 2013. Edited by Springer Verlag: Modeling Approaches and Algorithms for Advanced Computer Applications, Studies in Computational Intelligence, vol. 488, pp. 85–94, May 2013.

[Ime14] F. Imedjdouben and A. Houacine, “Development of an automatic phonetization system for Arabic text-to-speech synthesis,” Springer: International Journal of Speech Technology, vol. 17, issue 4, pp. 417-426, 2014.

[Ime15] F. Imedjdouben, A. Houacine, “Generation of allophones for speech synthesis dedicated to the Arabic language,” First International Conference on New Technologies of Information and Communication, pp. 99-102, IEEE DOI:10.1109/NTIC.2015.7368754, 2015.

[Ime15*] F. Imedjdouben, A. Houacine, “Implementation of an Arabic TTS System Based on Concatenative Synthesis,” 3rd International Conference on Signal, Image, Vision and their Applications, pp. 273-276, 2015.

[Kha11] O. O. Khalifa, M.Z. Obaid, A.W. Naji, J.I. Daoud, “A Rule-Based Arabic Text-To-Speech System Based On Hybrid Synthesis Technique,” Australian Journal of Basic and Applied Sciences, vol. 5(6), pp. 342-354, 2011.

[Kro92] B.J. Kroger, “Minimal Rules for Articulatory Speech Synthesis,” in Signal Processing VI: Theories and Applications, J.P.H. van Santen et al., Ed. Elsevier

Science, pp. 331–334, 1992.

[Lem99] S. Lemmetty, “Review of Speech Synthesis Technology,” Master's thesis, Helsinki University of Technology, March, 1999.

[Lev93] S.E. Levinson, J.P. Olive, and J.S. Tschirgi, “Speech Synthesis in Telecommunications, Synthesis of speech from unrestricted text is now commercially viable for telecommunications applications,” IEEE Communications Magazine, pp. 46–53, 1993.

[Mna05] Z. Mnasri, F. Boukadida, N. Ellouze, “Analyse/Synthèse de parole par modélisation sinusoïdale et recouvrement addition,” 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2005.

[Nes06] I. Nesterenko, “Analyse formelle et implémentation phonétique de l’intonation du parler russe spontané en vue d’une application à la synthèse vocale,” Thèse de Doctorat, Université AIX-Marseille I − Université de Provence, Septembre, 2006.

[Pos04] M. Postel, “Introduction au logiciel Matlab,” Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, 2004.

[Qua07] V.M Quang, “Exploitation de la prosodie pour la segmentation et l’analyse automatique de signaux de parole,” Thèse de Doctorat, Institut national polytechnique de Grenoble et de l’Institut polytechnique de Hanoi, Septembre, 2007.

[Ras10] M. Z. Rashad, H. M. El-Bakry, I. R. Isma'il, N. Mastorakis, “An Overview of Text-To-Speech Synthesis Techniques,” 4th International Conference on Communications and Information Technology, pp. 84–89, 2010.

[Ras10*] M. Z. Rashad, H. M. El-Bakry, I. R. Isma'il, “Diphone speech synthesis system for Arabic using MARY TTS,” International Journal of Computer Science & Information Technology (IJCSIT), vol. 2, pp. 18–26, 2010.

[Ric10] G. Richard, “Effets sonores Réverbération,” Master MVA, 2010.

[Rou06] S. Rouibia, “Prise en compte de critères acoustiques pour la synthèse de la parole,” Thèse de Doctorat, L'école nationale supérieure des télécommunications de Bretagne en habilitation conjointe avec l'Université de Rennes 1, Septembre, 2006.

[Saï04] T. Saïdane, M. Zrigui, M. Ben Ahmed, “La transcription orthographique-phonétique de la langue Arabe,” In: RÉCITAL, 2004.

[Saï05*] T. Saïdane, M. Zrigui, M. Ben Ahmed, “Un système de synthèse de la parole arabe par concaténation de polyphèmes: Les résultats de l’utilisation d’un lissage linéaire,” 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2005.

[Saï05] T. Saïdane, M. Zrigui, M. Ben Ahmed, “Arabic speech synthesis using a concatenation of polyphones: the results,” In Lecture notes in computer science: Advances in artificial intelligence, B. Kégl and G. Lapalme, Ed. Springer Berlin Heidelberg, vol. 3501, pp. 406–411, 2005.

[Saï06] T. Saïdane, “Contribution à la synthèse automatique de la parole arabe,” Thèse de Doctorat, Université de la Manouba, Ecole Nationale des Sciences de l’Informatique, Septembre, 2006.

[She12] Y. Shen, J. Jia, L. Cai, “Detection on PSOLA-modified Voices by Seeking out Duplicated Fragments,” International Conference on Systems and Informatics, 2012.

[Ste09] T. Stefan-Adrian, M. Doru-Petru, “Rule-based Automatic Phonetic Transcription for the Romanian Language,” Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, pp. 682–686, 2009.

[Sud15] B. Sudhakar, R. Bensraj, “Development of Concatenative Syllable-Based Text to Speech Synthesis System for Tamil,” in Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, L.P. Suresh et al., Ed. Springer India, vol. 324, pp. 585–592, 2015.

[Tab11] Y. Tabet, M. Boughazi, “Speech synthesis techniques. a survey,” 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA), pp. 67–70, 2011.

[Tho07] S. Thomas, “Natural sounding text-to-speech synthesis based on syllable-like units,” Master's thesis, Indian institute of technology madras, May, 2007.

[Tok02] K. Tokuda, H. Zen, A.W. Black, “An HMM-based speech synthesis system applied to English,” in Proceedings of the IEEE Workshop on Speech Synthesis, pp. 227–230, 2002.

[Vee10] E. Veera Raghavendra, P. Vijayaditya, K. Prahallad, "Speech synthesis using artificial neural networks," National Conference on Communications (NCC), pp. 1–5, 2010.

[Wel97] Wells, J.C., "SAMPA computer readable phonetic alphabet," In Gibbon, D., Moore, R. and Winski, R. (eds.). Handbook of Standards and Resources for Spoken Language Systems. Berlin and New York: Mouton de Gruyter. Part IV, section B.

[Zek10] M. Zeki, O.O. Khalifa, A.W. Naji, "Development of an Arabic Text-To-Speech System," In: IEEE International Conference on Computer and Communication Engineering, pp. 1–5, 2010.

[Zem96] Z. Zemirli, N. Vigouroux, M. Sellami, "SYNTHAR+ : un système de pré-traitement de textes arabes en vue de leur synthèse orale sous le système Multivox," 3th African Conference on Research in Computer Science and Applied Mathematics, pp. 719–729, 1996.

[Zem06] Z. Zemirli, "ARAB_TTS: An Arabic Text To Speech Synthesis," In: IEEE International Conference on Computer Systems and Applications, pp. 976–979, 2006.