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

DÉPARTEMENT D'ÉLECTRONIQUE

Projet de Fin d'Etudes
En vue de l'obtention du diplôme d'Ingénieur d'Etat en Electronique

Présenté par :

Mr TAIBI MOHAMED AMINE

Thème :

Gestion Automatique Vocale des Files d'Attente dans les Services Publics Algériens

Soutenu le : 22 Juin 2014

Devant le Jury :

H. BOUSBIA-SALAH	MCA	ENP	Président
M. GUERTI	Professeur	ENP	Rapporteur
L. HAMAMI	Professeur	ENP	Examinateur

Promotion : juin 2014

Remerciements

Tout d'abord je tiens à remercier Allah le tout Puissant qui m'a guidé durant mes études et qui m'a donné la volonté, la force et la patience afin de réaliser ce modeste travail.

Je remercie particulièrement ma promotrice Mme. GUERTI Mhania, Professeur à l'Ecole Nationale Polytechnique, sans qui, ce travail n'aurait pas pris cette forme ; je lui témoigne toute ma gratitude et reconnaissance pour ses encouragements, et le temps consacré avec une attention et une extrême patience, je la remercions pour tous ses conseils qui m'ont permis d'avoir une vision plus éclaire sur le travail.

Je viens humblement exprimer mes remerciements à Mr Kabache Mahrez enseignent à l'ISMAS, qui m'a aidé pour les enregistrements du corpus.

Je tiens à remercier les Membres du jury qui ont accepté d'évaluer ce travail :

- *Mr. BOUSBIA-SALAH Hichem, MCA à l'Ecole Nationale Polytechnique, d'avoir accepté de présider le jury de mon PFE.*
- *Mme. Hamami Latifa, Professeur à l'Ecole Nationale Polytechnique, d'avoir accepté de faire partie de mon jury.*
- *Je remercie également :*
- *Mr. AIT CHEIKH Mohamed, chef du département d'Electronique, pour les conseils qu'il m'a prodigués pendant toute notre période d'étude, malgré ses multiples responsabilités.*
- *Le personnel de la bibliothèque de l'ENP et surtout Mr Ami Salah, qui ont mis à ma disposition la documentation nécessaire.*
- *Enfin, un grand merci à tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce PFE.*

Dédicaces

Je dédie ce travail :

A tous les chahids qui ont contribué à la libération de notre cher pays.

Mes chers parents qui m'ont soutenu durant toutes les phases de mes études ; Je remercie mon cher père qui m'a toujours aidé et encouragé durant toute la période de mon PFE. Je remercie également ma chère et adorable mère, pour son amour, sa patience, ses sacrifices et son éternel soutien durant toutes mes années d'études.

Je leur dis : merci et qu'Allah vous garde pour moi.

A tous ceux qui ont contribué à ma formation et à mon éducation.

Je dédie ce travail aussi à toute ma famille :

Mon frère Abderrahmane.

Mes sœurs: Houda, Wafaa, Yasmine, Abir et surtout la petite Farah.

Mes tantes : fatma Zohra, Djamila, Hayat et Fatiha ainsi que leurs époux.

Mes cousins et mes cousines : Amine, Imane, Abderrahim, Aymen, Hichem, Abdelghani, et Ibrahim « Barhouma ».

Tous les amis qui ont partagé les bons moments que nous avons passés ensemble et particulièrement : Abdessalem, Akram, Alaeddine, Redha, Abdelmounaim, Fouad et Mohamed.

Je n'oublierais pas Hacen, Nassim, Halim, Loutfi, Abderraouf, Sami, Achraf, Youcef, Shamsou et Imed et mes copins de lycée Ibrahim, Rabah, Samir, Fouad.

A toutes les personnes que je connais et qui m'aiment.

TAIBI Mohamed Amine

ملخص

الهدف من عملنا هو إعداد نظام صوتي لإدارة قاعة الإنتظار في خدمات اتصالات الجزائر، باللغة العربية الفصحى والفرنسية. لتحقيق هذه الغاية، أنشأنا مدونة تضم جملتين ثابتتين و كلمات متغيرة(الأرقام). استخدمنا تقنية التسلسل الموجي بواسطة وصل الجملتين والكلمات المركبة، بعد التسجيلات التي تم إجراؤها في ظروف جيدة. قمنا بتقسيم التسجيل باستخدام برنامج PRAAT. لقد قمنا أيضا بالتحليلات الصوتية للتأكد من النوعية الجيدة للكلام المركب، ثم قمنا بتجميع الجملتين والكلمات المركبة باستخدام برنامج MATLAB، تقييم العمل تم باستخدام اختبارات لعشرة أشخاص. كانت النتائج المتحصل عليها جيدة.

كلمات المفاتيح : نظام صوتي، اللغة العربية الفصحى، اللغة الفرنسية، تركيب الكلام الإصطناعي، تحليل المدونة، تقنية التسلسل الموجي، اختبارات التقييم.

Résumé

L'objectif de notre travail est d'élaborer un **S**ystème Vocal de **G**estion **A**utomatique des **F**iles d'**A**ttente (**SGAFA)**, dans les services clientèles d'Algérie Telecom, en Arabe Standard et en Français. Pour arriver à cette fin, nous avons constitué un corpus comprenant deux phrases fixes et des mots variables (les numéros des tickets des clients et des guichets). Nous avons utilisé la synthèse par concaténation des phrases et mots combinés, après l'enregistrement qui a été faite dans des bonnes conditions à l'ISMAS. Nous avons fait une segmentation à l'aide du logiciel PRAAT. Nous faisons des analyses acoustiques pour confirmer la bonne qualité de la parole synthétique, puis nous concaténons les phrases fixes avec les mots variables par un programme sur MATLAB, notre travail a été évalué à l'aide des tests par 10 personnes.

La qualité obtenue est très bonne.

Mots clés : Système Vocal, Arabe Standard, Français, Synthèse de la Parole, Méthode de concaténation, Analyse de la parole, Tests d'évaluations.

Abstract

The objective of our work is to elaborate a **V**ocal **M**anagement **S**ystem **A**utomatic of the **W**aiting **H**all (**VMSAWH**), in customer services in Algeria Telecom in Standard Arabic and French languages; we established a corpus comprising two fixed phrases and variables words. We used synthesis by concatenating of compound words and phrases, after the recording was made in good conditions at ISMAS. We segmented using the software PRAAT. We make acoustic analyses to confirm the good quality of synthetic speech, then we concatenate the fixed sentences with the words variable by a program on MATLAB, our work was to evaluate using the tests for 10 people.

Key words : Voice System, Standard Arabic, French, Synthesis of the Word, Method concatenation, Speech analysis, Evaluation tests.

Table des matières

Chapitre 1 : GENERALITES SUR LA PAROLE

Chapitre 2 : NOTIONS SUR LA SYNTHESE DE LA PAROLE

Chapitre 3 : ANALYSE DU SIGNAL DU CORPUS

Chapitre 4 : GESTION AUTOMATIQUE DES FILES D'ATTENTE

Liste des Figures

page

Liste des Tableaux

Liste des Abréviations

CV : **C**ordes **V**ocales
TAP : **T**raitement **A**utomatique de la **P**arole
RAP : **R**econnaissance **A**utomatique de la **Pa**role
API : **A**lphabet **P**honétique **I**nternational
AS : **A**rabe **S**tandard
F_0 : **F**réquence **f**ondamentale
TTS : **T**ext-**T**o-**S**peech (Un Système de Synthèse à Partir du Texte)
TOP : **T**ranscription **O**rthographique-**P**honétique
LPC : **L**inear **P**rédictive **C**oding (Codage Prédictif Linéaire)
MFCC : **M**el **F**requency **C**epstral **C**oefficients
TF : **T**ransformée de **F**ourier
TFD : **T**ransformée de **F**ourier **D**iscrète
TFR : **T**ransformée de **F**ourier **R**apide (**FFT**)
AR : **A**uto **R**égressif
ARMA : **A**uto **R**égressif à **M**oyenne **A**justée
MA : **M**oyenne **A**justée
SPR : **S**ynthèse **P**ar **R**ègles

OEAO : **O**utils d'**E**nseignement **A**ssisté par **O**rdinateur

TALN : **T**raitement **A**utomatique du **L**angage **N**aturel
V-C-V : **Voyelle-C**onsonne-**V**oyelle
LAM : **L**aboratoire d'**A**coustique **M**usicale
LIMSI : **L**aboratoire d'**I**nformatique pour la **M**écanique et les **S**ciences de l'**I**ngénieur
ISMAS : **I**nstitut **S**upérieur des Métiers des Arts du Spectacle et de l'Audiovisuel

SGAFA : **S**ystème de **G**estion **A**utomatique de **F**iles d'**A**ttente

INTRODUCTION GENERALE

Le Traitement Automatique de la Parole (TAP) est un domaine de recherche immense et très vaste, et avec le développement du traitement du signal numérique et du traitement de la parole. Depuis le siècle dernier, cette évolution au traitement numérique du signal et à la télécommunication qui a conduit à des applications de Codage, de Reconnaissance Automatique de la parole et de la synthèse vocale.

La synthèse de parole est une technologie dont l'usage connaît un essor important, pour répondre notamment aux besoins des services des Télécommunications et pour reproduire un système qui peut remplacer la capacité humaine à l'aide de machines, tels que les services téléphoniques ou les services de messagerie électronique ou les services publics . Ces mises en service sont dues à une certaine confiance gagnée suite aux améliorations récentes de la qualité vocale qu'a connu cette technologie.

L'objectif de notre travail est de réaliser un système parlant qui fait l'appel aux personnes dans les services publics Algériens, nous utilisons la méthode de la synthèse de la parole par phrases et mots combinés enregistrés en AS et en Français.

Dans les services publics Algériens, nous avons assez problèmes avec l'organisation des services clientèles, c'est pour cela notre travail consiste de mettre un système qui fait la gestion automatique dans les files d'attente pour bien gérer et organiser nos services publics.

Les avantages de la gestion automatique c'est que ses annonces vocales des guichets faites pour ne pas subir la pression des personnes en attente et réduire le temps d'attente, gérer plus simplifier la gestion des clients qui vont être satisfait.

Pour notre travail, nous avons structuré notre PFE en quatre chapitres :

- ✓ dans le premier, nous allons voir d'une manière générale des notions sur la parole ainsi que la production de la voix humaine, l'appareil phonatoire et auditif de l'être humain. Après des notions de base sur les formants et sur les paramètres pertinents d'un signal acoustique ;

- ✓ le deuxième, nous donne un bref historique sur la synthèse de la parole, ses domaines d'application, Ensuite, nous expliquons les différentes techniques d'analyse du signal vocal ;

- ✓ dans le troisième, nous nous intéressons à analyser notre corpus, en étudiant les caractéristiques et les paramètres pertinents de ce signal vocal sur le spectrogramme (fréquence fondamentale (F0), formants et intensité). Nous présentons le logiciel d'analyse Praat ainsi que l'outil de programmation.

- ✓ Dans le dernier chapitre, nous introduisons les méthodes de la synthèse de la parole et les étapes de l'élaboration de notre corpus et son traitement, nous allons présenter aussi une simulation du **S**ystème de **G**estion **A**utomatique des **F**iles d'**A**ttente d'Algérie Telecom (**SGAFAAT**). Des tests d'évaluation subjectifs des résultats obtenus, ont été effectués par 10 personnes. Les résultats sont bons et de bonne qualité.

Nous terminons notre projet avec des conclusions générales et perspectives.

CHAPITRE 1

GENERALITES SUR LA PAROLE

1.1. Introduction

Ce chapitre est consacré à des notions fondamentales sur la parole, ainsi que, les caractéristiques des différents sons du langage. Des généralités sur le signal vont être également exposées.

1.2. Généralités sur la Parole

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

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

1.3. Production de la Voix Humaine

Le processus de production de parole est un mécanisme très complexe qui repose sur une interaction entre le système neurologique et physiologique. Il y a une grande quantité d'organes et de muscles qui entrent dans la production de sons des langues naturelles. Le fonctionnement de l'appareil phonatoire humain repose sur l'interaction entre trois grandes classes d'organes : les poumons, le larynx, et les cavités supra-glottiques.

1.3.1. Architecture de l'appareil phonatoire

Le modèle de la section suivante explicite le fonctionnement de l'appareil vocal, ainsi que le rôle joué par ses différents constituants lors du processus de production de la parole (Figure 1.1).

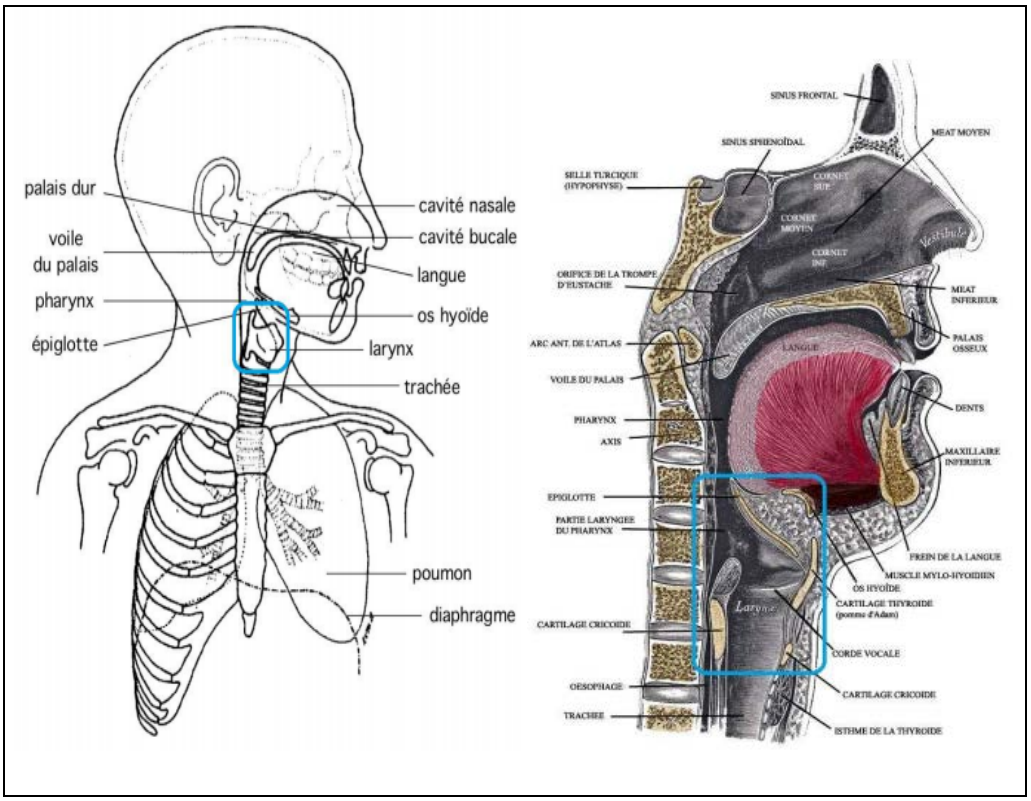


Figure 1.1 : Anatomie de l'appareil phonatoire humain, coupe sagittale [2]

L'organe le plus important dans la production de la voix c'est le larynx qui est l'organe clé car il contient les Cordes Vocales (**CV**), les deux rectangles bleus représentent la position du larynx sur l'appareil phonatoire.

Il y a aussi la fonction primordiale des poumons qui permet au corps de s'oxygéner. Cependant, les poumons fournissent aussi une source d'air qui est utilisée pour produire des sons.

1.3.2. Fonctionnement de l'appareil vocal

Le larynx est un ensemble de cartilages reliés par des muscles dont les CV, il est l'élément vibratoire du conduit vocal. Il faut rester souple pour qu'il puisse monter et descendre d'une façon très facile (Figure 1.2).

Une notion importante, qui fait le lien entre évaluation perceptive et fonctionnement physiologique est la notion de mécanisme laryngé.

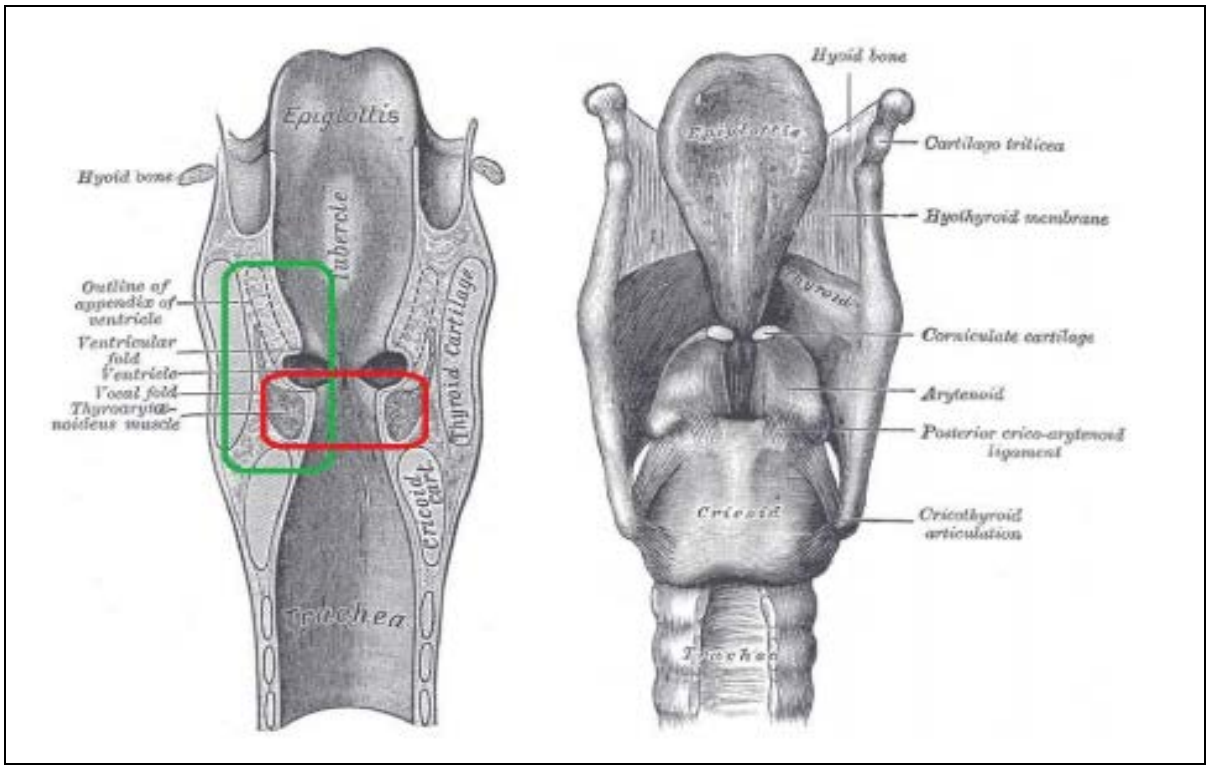


Figure 1.2 : Anatomie de larynx [2]

Les rectangles respectifs :

- Rouge montre la position des CV ;
- Vert // l'anatomie d'une seule CV et une bande ventriculaire.

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive. L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulée avant d'être appliquée au conduit vocal.

Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée. Les CV sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complétement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non-voisés (ou sourds).

Les sons voisés (ou sonores) résultent au contraire d'une vibration périodique des CV. Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales, les force à s'ouvrir, ce qui fait tomber la pression et permet aux CV de se refermer ; pour la plupart des sons, des impulsions périodiques de pression sont ainsi appliquées au conduit vocal, composé des cavités pharyngienne (pharyngo-buccale, nasale et labiale), Lorsque la luette est en position basse, la cavité nasale vient s'y ajouter en dérivation [3].

1.3.3. Production de l'onde glottique

L'air produit par excès de pression dans les poumons rencontre un premier obstacle qui est les cordes vocales (source d'excitation). Ces dernières accolées, sous l'effet de la pression sub-glottique se mettent à vibrer laissant passer l'air par impulsions. C'est ainsi que se forme l'onde glottique dont la fréquence d'oscillations notée F_0 (fréquence fondamentale ou pitch), est déterminée par la masse et la tension des CV ainsi que la pression sub-glottique. Quand elles vibrent, il y a émission de sons dits :

- sons voisés ou sonores
- sons non voisés ou sourds, qui sont assimilables à un bruit blanc [4].

1.3.4. L'épiglotte

C'est une structure cartilagineuse reliée au larynx qui coulisse vers le haut quand les voies aériennes sont ouvertes. Elle aide à obstruer l'entrée de la trachée au moment de la déglutition. Elle descend légèrement vers le bas, afin d'entrer en contact avec le larynx qui s'élève, formant ainsi un verrou au-dessus du larynx.

Il se peut que de temps à autre, lorsqu'on mange trop vite, des aliments liquides ou solides ingérés pénètrent dans le larynx avant que l'épiglotte n'ait pu se rabattre sur celui-ci. De tels cas peuvent s'avérer très dangereux du fait que les voies respiratoires peuvent se boucher et empêcher l'air de pénétrer dans les poumons.

1.3.5. La luette

La luette ou uvule est une saillie allongée mobile qui termine le voile du palais et qui contribue, lorsqu'elle se détache de la paroi pharyngale, à permettre à l'air provenant des poumons et du larynx de se diriger non seulement vers la bouche, mais également vers les

fosses nasales. Lorsque la luette s'appuie sur la paroi pharyngale, elle empêche l'air de pénétrer dans les fosses nasales et ne le laisse s'échapper que par la bouche (articulations orales) [5].

1.3.6. La langue

La langue est une masse musculaire divisée en trois parties :

- la pointe (apex) qui sert d'articulateur pour les articulations apicales, le dos pour les articulations pré médio ou post-dorsales, et la racine dans le cas des articulations radicales. Elle constitue l'articulateur principal des différents sons ;
- la langue permet le blocage d'air venant des poumons pour produire les consonnes occlusives, le resserrement de la cavité buccale inhérent à la production des consonnes constrictives, lorsqu'elle demeure suffisamment éloignée de la voûte du palais, elle permet la réalisation des différentes voyelles [5].

1.4. Notions sur les Formants

Parmi les différentes méthodes d'évaluation objectives des dysfonctionnements articulatoires, les paramètres prosodiques (mélodie, intensité et durée).

1.4.1. Timbre

Le timbre est la qualité du son qui permet de reconnaître la voix d'une personne, ou l'instrument de musique à l'origine d'un son.

Le caractère timbré de la voix serait lié au renforcement du formant du chanteur (région de 2 - 4 kHz) et aussi à la richesse spectrale dans les fréquences moyennes. Les représentations de la répartition d'énergie spectrale par bandes de fréquences confirment ces hypothèses. La voix détimbrée pourrait être assimilée à une atténuation du spectre dans la zone du formant du chanteur, la présence restreinte d'harmoniques aigus. Le son peut alors paraître terne, aggravé ou avec la présence de souffle (Figure 1.3).

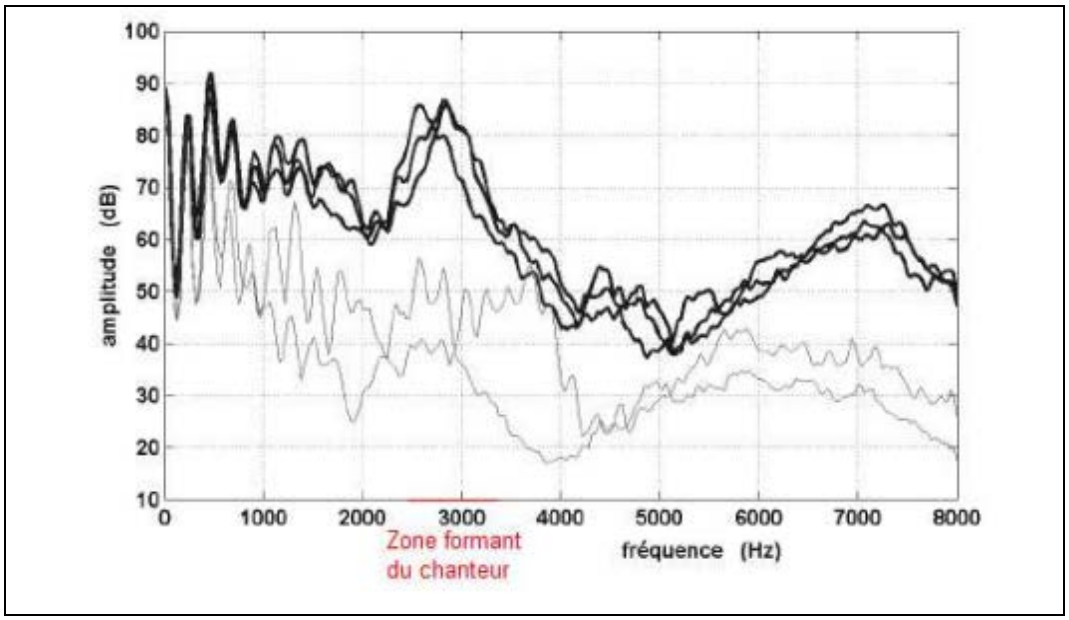


Figure 1.3 : Spectre d'amplitude d'une phrase [6]

1.4.2. Formants

C'est la fréquence de résonance du conduit vocal, la position de trois premiers formants (F_1, F_2, F_3) caractérise le timbre (Figure 1.4).

Pour les formants qui sont supérieurs indiquent l'état de locuteur.

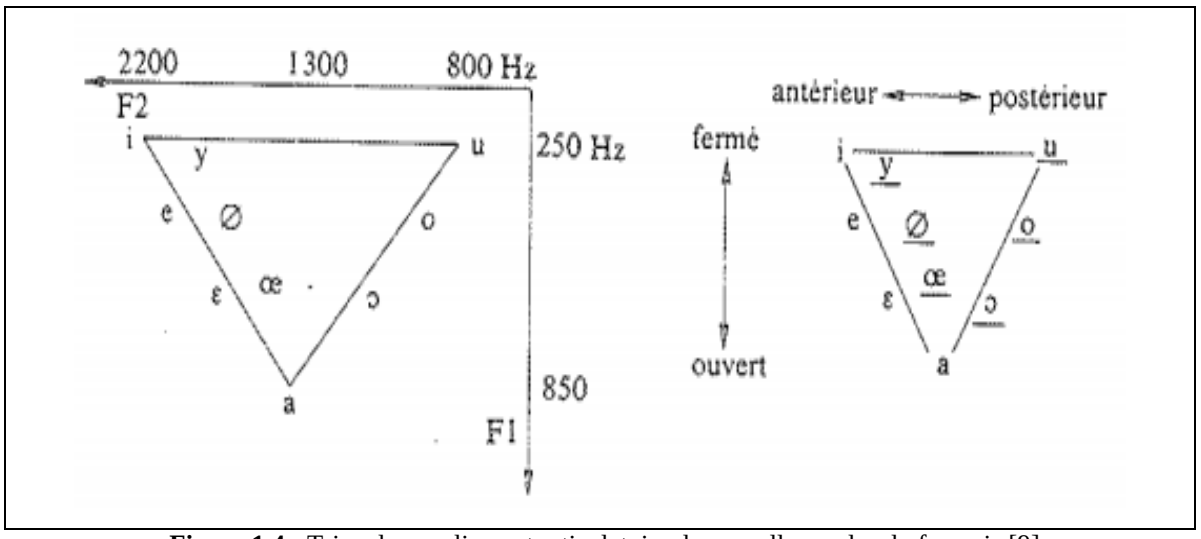


Figure 1.4 : Triangle vocalique et articulatoire des voyelles orales du français [9]

L'appareil phonatoire étant constitué de différentes cavités. Lors du passage de l'air à travers ces cavités il est amplifié et subit différentes transformations dues aux degrés d'ouverture et de fermeture au niveau de chaque cavité, à la position de la langue, des

lèvres, etc. Ces cavités possèdent des fréquences de résonance qui renforcent certaines régions du spectre de sources excitatrices.

Les maxima de la courbe de réponse en fréquences du conduit vocal sont appelés Formants. Chaque son a ses formants caractéristiques. Sur un spectrogramme, les formants sont représentés par des bandes noires (le degré de noirceur correspondant à l'énergie) (Figure 1.5).

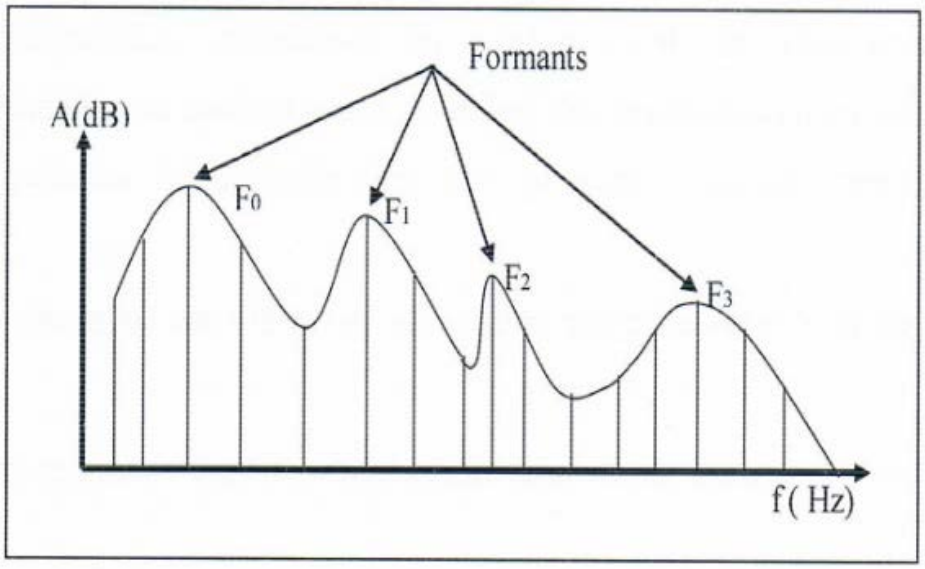


Figure 1.5 : Représentation des Formants d'un son voisé [8]

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

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

Les valeurs des formants sont très influencées par le lieu d'articulation. Ils donnent une image de la configuration articulatoire du conduit vocal, car elles correspondent aux

fréquences de résonance du conduit vocale, de même des expériences qui restent à vérifier ont montré que la position fréquentielle des trois premiers formants caractérise le timbre vocalique [8]:

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

1.4.3. Phonème

Le phonème est le plus petit élément acoustique, autrement dit, la plus petite unité distinctive de la chaîne parlée, c'est-à-dire la plus petite unité de son capable de produire un changement de sens.

Dans la langue française, il y a 37 phonèmes, 16 voyelles, 18 consonnes et 3 semi-voyelles.

1.4.4. voyelles

Les voyelles sont caractérisées par le degré d'ouverture, elles sont susceptibles de varier en énergie et correspondent à une disposition articulatoire ouverte.

Voilà le triangle vocalique de la langue française.

Ce triangle vocalique montre la position de la langue dans la cavité buccale.

1.4.5. consonnes

Les consonnes correspondent à l'émission de bruits qui ont leur origine dans des obstructions du chenal respiratoire.

Il y a deux types de consonnes :

- consonne occlusive : c'est un Son produit par l'air qui rencontre un obstacle total ;
- consonne constrictive : son produit par l'air qui rencontre un obstacle partiel :
  - non voisée (sourde) : pas de vibration des CV
  - voisée (sonore) : il y a vibration des CV.

1.4.6. Semi-Consonnes

Les semi-consonnes correspondent à des articulations qui ne peuvent pas être considérées comme relevant du vocalisme ou du consonantisme.

Dans la langue française, nous trouvons 3 semi-consonnes (semi-voyelles).

Tableau 1.1 : Symboles de l'Alphabet Phonétique International utilisés dans la transcription du Français

Les voyelles orales

symboles	*mot-clé*	*autres graphèmes*
[i]	**lit**	stylo, île, maïs, meeting
[y]	**lune**	sûr, j'ai eu, aigüe
[u]	**tout**	où, goûter, football, août
[ə]	**je**	
[e]	**télé**	parler, nez, pied, et
[ɛ]	**mère**	faire, secret
[ø]	**feu**	nœud, jeûne
[œ]	**fleur**	cœur, club
[o]	**vélo**	sauter, peau, nôtre
[ɔ]	**pomme**	album, alcool
[a]	**patte**	
[ɑ]	pâte	

Les voyelles nasales

[ɑ̃]	**gant**	camp, cent, empereur, paon, Caen
[õ]	**bon**	ombre
[ɛ̃]	**lapin**	chien, pain, daim, imparfait, syndicat, sympa
[œ̃]	brun	parfum

Les consonnes

[s]	**se**	ce, poisson, citron, garçon, science, dix, démocratie
[z]	**zéro**	maison, dixième, blizzard, -s *en liaison* (plus◡actif)
[ʃ]	**chat**	short, schéma, fascisme
[ʒ]	**jeune**	âgé, mangeons
[f]	**fou**	affaire, pharmacie
[v]	**vin**	wagon
[R]	**rare**	beurre
[l]	**lait**	elle
[p]	**papa**	appartement
[b]	**bébé**	abbaye
[m]	**mais**	flamme
[n]	**non**	anniversaire
[t]	**table**	patte, sept, -d *en liaison* (un grand homme)
[d]	**dos**	addition
[k]	**car**	accord, qualité, képi, orchestre, ticket, coq
[g]	**gâteau**	bague, aggraver, second

Les semi-voyelles

[w]	**oui**	toit, loin, poêle, jaguar, aquarelle
[ɥ]	**puis**	continuer, linguistique
[j]	**pied**	travail, payer, grenouille

1.5. Paramètres d'un Signal Vocal

Nous voyons ces paramètres de deux points de vue :

- perceptif :

  Variations de la mélodie, de l'intonation, de l'énergie et du rythme ;

- acoustique :

  Modifications de la fréquence fondamentale, de l'intensité et de la durée.

En ce qui concerne notre travail, nous nous intéressons

1.5.1. Fréquence Fondamentale

Dans le domaine de la parole, nous avons le problème du son non voisé, car nous avons besoin un signal voisé pour avoir la fréquence fondamentale c'est une composante fréquentielle principale qui est la fréquence de vibration de la CV, chez les hommes, la fréquence fondamentale est plus petite par rapport à celles des femmes et des enfants (Figure 1.6).

L'analyse de la fréquence fondamentale donne beaucoup d'informations, par exemple nous pouvons savoir est ce qu'un homme ou femme ou enfant qui parle car leurs fréquence n'est pas la même à cause des CV, sa longueur et sa masse, et pour qu'on puisse analyser et faire la synthèse de la voix, on utilise une méthode parmi les plus connues comme la FFT et l'autocorrélation, etc.

Dans la voix humaine, il existe trois bandes de fréquence, comprises entre :

- 70 Hz - 250 Hz pour l'Homme ;
- 150 Hz - 500 Hz pour la Femme ;
- 200 Hz - 600 Hz pour l'enfant.

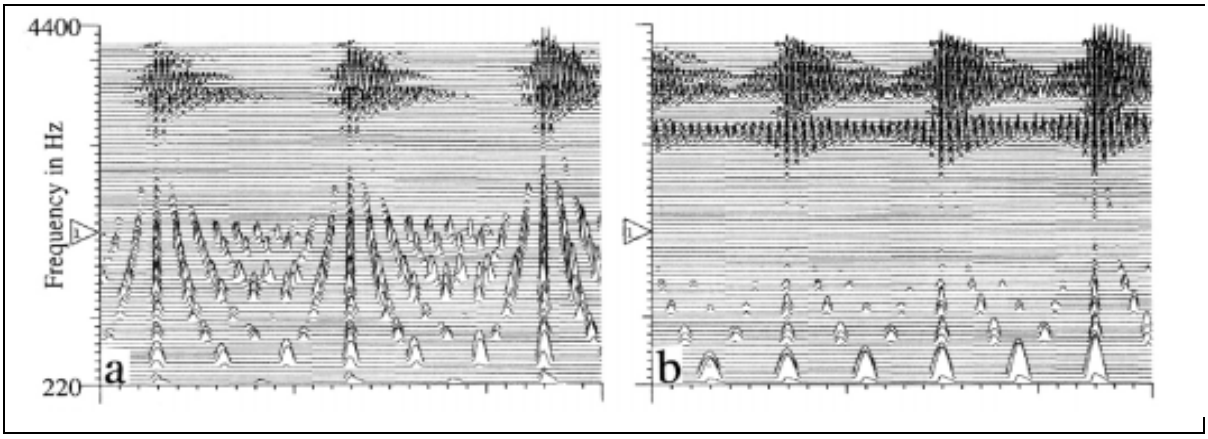


Figure 1.6 : Image auditive de voyelles isolées [a] et [i]

Cette figure est représentée sur le plan perceptif. La voyelle [a] a un F_0=100 Hz, et [i] a un F_0=125 Hz.

Le changement de la fréquence pendant la parole des personnes nous donnent l'intonation. Ceci relève du domaine de la prosodie.

1.5.2. Durée

La durée des sons dépend de la pression de l'air expiré, elle est très variable qui représente le temps pour prononcer un phonème.

Nous pouvons caractériser deux types de la durée : la durée observée, qui veut dire le temps pour faire activer les organes de phonation.

la durée perçue, est liée au mécanisme de la perception et elle est très utilisée dans le cas des occlusives puisqu'elles sont connues par une durée de réalisation discontinue.

Parmi les facteurs de variabilité de la durée d'un phonème, on peut citer :

- le type de la parole dont il est extrait : parole spontanée/lue, continue/mot isolé.
- la vitesse d'élocution.
- le mot, la phrase contenant le phonème : les durées des phonèmes diminuent si le nombre de syllabes augmente, la durée dépend aussi de la position du phonème dans le mot (début, fin de mot).
- les phonèmes adjacents

1.5.3. L'intensité(ou L'énergie)

Elle se caractérise sur la dimension faible – fort (en décibels).

Elle est souvent évaluée sur plusieurs trames de signal successives pour pouvoir mettre en évidence des variations. La formule de calcul de ce paramètre est :

- $E = \frac{1}{T} \sum_{n=1}^{T} X^2$ **(1.1)**

- énergie en décibel : $E_{dB} = 10 * \log_{10}(\frac{1}{T} \sum_{n=1}^{T} X^2)$

  (1.2)

1.6. Conclusion

Dans ce chapitre nous avons exposé certains concepts de base sur le traitement de la parole, et quelques caractéristiques du signal acoustique.

Les objectifs de ce chapitre sont de définir les notions que nous utiliserons dans notre travail. Cette partie théorique sur la parole sera complétée dans le chapitre suivant par une étude approfondie des systèmes de synthèse de la parole et ses variantes.

CHAPITRE 2

NOTIONS SUR LA SYNTHESE DE LA PAROLE

2.1. introduction

Ce chapitre nous présente l'état de l'art sur la synthèse vocale qui est le cadre technique de notre étude où les principales méthodes de synthèse de la parole seront exposées ainsi que les différents types d'unités de synthèse utilisables et les différents modules de traitement linguistiques et acoustiques présents dans un système de TTS.

Au premier lieu, nous allons voir l'évolution de la synthèse de la parole puis ces différentes applications et sa définition, ensuite nous expliquons les méthodes de la synthèse de la parole et nous donnons aussi les différentes techniques d'analyse du signal vocal.

2.2. Evolution de la Synthèse de la Parole

L'histoire de la synthèse de la parole est passée par trois grandes étapes technologiques, ces étapes existent aujourd'hui commercialement :

- la synthèse par règles ;
- la synthèse par concaténation de diphones ;
- la synthèse par sélection d'unités.

De point de vue du développement de l'industrie des machines de la synthèse vocale, Les premières machines parlantes voient le jour avec l'abbé Mical et Wolfgang von Kemplenau 18ème siècle, premières grandes simulations mécaniques des phénomènes de production de la parole humaine associant source vocale et résonateurs supraglottiques (Figure2.1) [9].

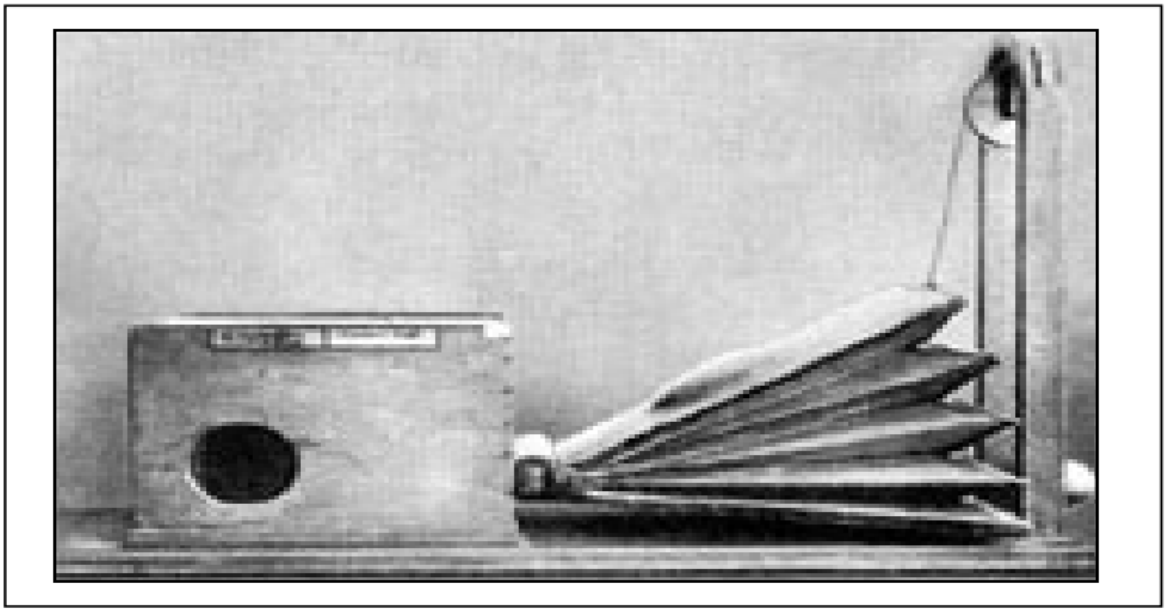


Figure2.1 : Machine à parler de Kempelen [9]

La parole est devenue très intéressante pour l'étudier par les chercheurs au 19^{éme} siècle de point de vue scientifique.

Parmi ces chercheurs, Joseph Faber avec son Euphonia (1830 - 1840) ; Charles Wheatstone avec le perfectionnement de la machine de Kempelen ; Alexander Graham Bell pour une version simplifiée de la reconstitution de Wheatstone et R.R. Riesz avec un appareil simulant les différentes sections du conduit vocal. Ce dernier permit une meilleure compréhension de la physiologie de l'appareil phonatoire humain, de la géométrie et du rôle de ses articulateurs [9].

Le 20ème siècle a vu des inventions plus développées à cause de l'apparition de l'électronique, J. C. Stewart a fait une machine qui est capable de reproduire des voyelles, après plus d'une quinzaine d'années, exactement en 1939, H. Dudley réalise un appareil mis au point par les laboratoires Bell à l'occasion de l'exposition universelle de New York, le VODER (Voice Operating Demonstrator), mais le premier synthétiseur de la parole s'est fait durant les années clinquantes, par le système pattern playback qui s'est créé par les laboratoires Haskins.

Le 20ème siècle vit le premier système de synthèse automatique à partir du texte en Français, qui a été conçu et réalisé au **L**aboratoire d'**A**coustique **M**usicale (**LAM**) de Paris VI et au **L**aboratoire d'**I**nformatique pour la **M**écanique et les **S**ciences de l'**I**ngénieur (**LIMSI**), après en 1974 l'apparition de l'ICOPHONE qui permet de modifier la F_0 des oscillateurs par programme ou manuellement, et les appareils de la synthèse de la parole sont toujours en développement Les machines sont toujours en développement, de nos jours, par exemple, synthétiseurs à formants, des synthétiseurs à prédiction linéaire, etc.

2.3. Quelques Applications de la Synthèse de la Parole

La synthèse de la parole est un domaine d'application très important pour l'informatique ou les télécommunications, elle est aussi un domaine de recherche très vivant en traitement automatique du langage naturel et en traitement du signal, pour cela, de nombreuses qui équipes consacrent leurs travaux à la synthèse de la parole à travers le monde.

2.3.1. Les Services De Télécommunications

Le marché des télécommunications au monde a récemment obligé les opérateurs de télécommunications de mettre leurs clients au confort. En particulier, ils cherchent

actuellement à fournir un maximum de services avec le moindre coût. Au niveau téléphonique, les synthétiseurs permettent précisément de rendre tout type d'information écrite disponible via le téléphone. Ils peuvent par exemple créer des serveurs vocaux pour faciliter la vie, diffusant les horaires des cinémas, des informations routières, l'état d'un compte en banque, des explications automatisées concernant la dernière facture de téléphone.

2.3.2. Aides aux personnes handicapées

Les personnes handicapées ont besoin vraiment de la parole. Le problème peut être soit d'origine mentale, soit d'origine motrice ou sensorielle.

La machine peut être d'un grand secours pour les personnes handicapées dans le second cas. Nous citons des exemples :

- lecture d'écrans ou de documents écrits pour non-voyants ;
- aides à la communication vocale pour les personnes muettes, laryngectomisées ou à infirmité motrice cérébrale ;
- journaux vocaux, etc.

2.3.3. Outils d'Enseignement Assisté par Ordinateur (OEAO)

Il existe beaucoup de systèmes qui sont des outils très utiles à l'apprentissage d'une nouvelle langue, en complément d'un cours avec un professeur, ou peut-être d'autre logiciel comme par exemple un système de dictées automatiques.

Les principales applications de la synthèse de la parole sont :

- applications grand public non téléphoniques :
  - ✓ domotique (alarmes, appareils domestiques parlants, etc.) ;
  - ✓ micro-informatique (jeux et CDROMs parlants, bureautique, etc.).
- applications industrielles :
  - ✓ serveurs d'alerte
  - ✓ fonction de vérification vocale dans les postes d'édition (correction des épreuves) ou de saisie d'informations écrites (bases de données).
- Télématique vocale :
  - ✓ serveurs vocaux d'informations

- ✓ serveurs de lecture vocale de FAX ou de messages électroniques (e-mails).
- les jeux vidéo

2.4. principe de la Synthèse de la Parole

Définition de la synthèse vocale

La synthèse de la parole est l'ensemble des dispositifs, matériels ou algorithmes, pour générer automatiquement de la parole artificielle. La synthèse de la parole consiste en la lecture par une voix synthétique d'un texte numérique.

Il existe plusieurs types de synthèse vocale, la version la plus complète étant la synthèse à partir d'un texte. Le but est de produire de la parole à partir d'un texte auparavant inconnu par le système.

Nous pouvons définir la synthèse de la parole par la génération automatique, par des dispositifs matériels et/ou des algorithmes, de parole artificielle. Il y a plusieurs types de synthèse vocale ; la plus complète est la synthèse à partir de texte (text – to - speech) où le but est de produire de la parole à partir d'un texte a priori inconnu.

Ce type qui s'appelle **TTS** (**T**ext – **T**o - **S**peech) permet de créer un signal de parole à partir d'un texte donné.

Quel que soit le type de synthèse (ce que nous étudierons par la suite), un système TTS comprend deux modules principaux :

- un module de traitement de texte composé de fonctions qui permettront de décomposer le texte en un ensemble d'unités phonétiques distinctes (phonèmes, diphones, positions contextuelles du mot dans la phrase…). Certains systèmes perfectionnés intègrent un traitement prosodique, c'est-à-dire, l'intonation dans la voix. On appelle cet ensemble les modules **TALN** (**T**raitement **A**utomatique du **L**angage **N**aturel)
- un synthétiseur vocal, prenant comme entrée la sortie du module précédent, et en déduisant le signal audio de parole.

Les synthétiseurs ont quant à eux la fonction inverse de celle des analyseurs et des reconnaisseurs de parole : ils produisent de la parole artificielle. On distingue fondamentalement deux types de synthétiseurs à partir d'une représentation :

- numérique, inverses des analyseurs, dont la mission est de produire de la parole à partir des caractéristiques numériques d'un signal vocal telles qu'obtenues par analyse ;
- symbolique¸ inverse des reconnaisseurs de parole et capables en principe de prononcer n'importe quelle phrase sans qu'il soit nécessaire de la faire prononcer par un locuteur humain au préalable. Dans cette seconde catégorie, on classe également les synthétiseurs en fonction de leur mode opératoire **:**

Les synthétiseurs à partir du texte reçoivent en entrée un texte orthographique et doivent en donner lecture.

Les synthétiseurs à partir de concepts, appelés à être insérés dans des systèmes de dialogue Homme-Machine, reçoivent le texte à prononcer et sa structure linguistique, telle que produite par le système de dialogue [10].

2.5. Architecture de la Synthèse de la Parole

Tout système **TTS** (**T**ext- **T**o -**S**peech) est généralement constitué de deux blocs de traitements principaux :

- un bloc de traitements linguistiques ;
- un bloc de traitements acoustiques.

Le premier bloc vise à analyser et à structurer le texte afin de déterminer un mode de prononciation cohérent, puis à transformer le texte analysé en une séquence de descripteurs symboliques décrivant les unités cible.

Le deuxième bloc consiste à générer un signal acoustique adapté à cette séquence symbolique.

L'architecture générale d'un système de synthèse de la parole à partir du texte. Les deux premières parties qui concernent les traitements de *haut niveau* permettent le passage de la représentation orthographique du texte en entrée à une représentation phonétique munie

d'une description prosodique. La dernière partie englobe les traitements de bas niveau du synthétiseur qui permettent la génération proprement dite du signal acoustique (Fig 2.2).

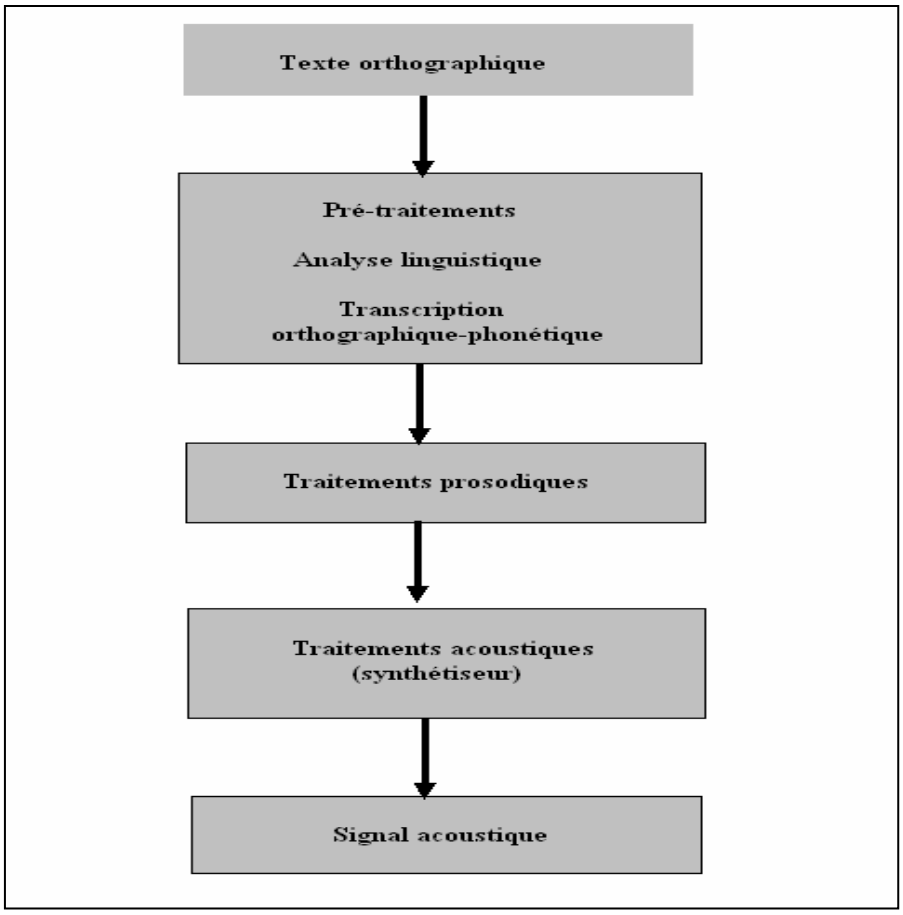


Figure 2.2 : Architecture générale d'un système de synthèse de la parole à partir du texte [11]

2.6. analyse et Modélisation du Signal de Parole

L'étude de l'évolution temporelle et fréquentielle d'un signal de parole permet de mettre en évidence les caractéristiques de ce signal. Cet objectif est atteint grâce aux méthodes modernes de traitement du signal qui permettent de calculer par exemple, la Transformée de Fourier (TF) d'un signal de parole pour déduire son spectre de puissance à court terme, et son spectrogramme qui représente l'évolution temporelle de ce spectre.

La quasi-stationnarité par morceaux du signal de parole est une hypothèse généralement admise qui permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation du signal stationnaire. Ces méthodes sont utilisées pour le traitement à court terme du signal de parole. Le traitement à long terme est quant à lui, assuré par le décalage temporel

sur le signal, de la fenêtre de traitement à court terme. Le signal de parole est ainsi progressivement analysé ou modélisé, sur des fenêtres du signal de durée généralement comprise entre 20 à 30 ms, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse ou du modèle.

2.7. Techniques de la Synthèse Vocale

Les techniques de synthèse dépendent de la stratégie adoptée. Dans le cas d'une synthèse basée sur la stratégie « system-models », la synthèse articulatoire figure comme étant la technique qui répond aux concepts de cette stratégie. Dans le cas de la synthèse fondée sur la stratégie « signal-models », deux familles de techniques sont utilisées : celles fondées sur un modèle source/filtre, et celles traitant le signal de parole directement dans le domaine temporel ou fréquentiel [13].

Pour les techniques de reconnaissance, d'analyse ou de synthèse de la parole, la fréquence d'échantillonnage peut varier de 08 jusqu'à 16 kHz. Le filtre de préaccentuation de transmittance H(z) est :

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

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

Le découpage du signal en trames produit des discontinuités aux frontières des trames, qui se manifestent par des lobes secondaires dans le spectre. Pour compenser ces effets de bord, nous multiplions en général préalablement chaque tranche d'analyse par une fenêtre de pondération de type fenêtre de Hamming notée W(n**)** [8].

$$W(n) = \begin{cases} 0.54+0.46.\ \mathrm{Cos}\ (\pi n/(n-1)) & \text{avec: } n \in [0,\ldots, n-1] \\ 0 & \text{Ailleurs} \end{cases} \qquad (2.2)$$

2.7.1. Synthèse articulatoire

La synthèse articulatoire est potentiellement considérée comme la technique la plus performante, car elle reflète théoriquement le processus physiologique. Cette technique est basée sur une modélisation géométrique du conduit vocal. Elle consiste à représenter le conduit vocal comme un tube de sections variables, avec des embranchements et des sections parallèles, puis à y simuler le trajet des ondes produites au niveau de la glotte. Les modèles d'écoulement d'air (mécanique des fluides), de sources et de propagation acoustique (phénomènes physiques), en association avec des modèles articulatoires (mécaniques), permettent de constituer un synthétiseur articulatoire complet, contrôlé par deux jeux de paramètres : les paramètres supra – laryngés qui commandent le modèle articulatoire, et un jeu de paramètres qui pilotent les cordes vocales (pression sub-glottique, longueur des cordes vocales et hauteur de la glotte au repos) [15].

La synthèse articulatoire est difficile à mettre en œuvre. Par ailleurs, comparée aux techniques alternatives, le volume de calcul est considérablement plus élevé. C'est pourquoi la synthèse articulatoire est très rarement utilisée dans les systèmes actuels. Mais, cette méthode a un grand potentiel, d'une part, pour sa haute qualité de synthèse, et d'autre part, pour l'approfondissement des connaissances acquises jusqu'à maintenant sur la production de la parole [16].

2.7.2. Méthodes paramétriques

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

Les avantages de cette approche sont la souplesse de l'analyse, l'introduction naturelle de l'information et les choix variés des espaces de représentations paramétriques. Dans le

cas de la modélisation du signal parole, nous n'avons accès qu'à une seule sortie du système alors que l'entrée n'est pas mesurée. Il en résulte un problème d'estimation non linéaire car nous ne disposons pas d'observation de l'onde glottique d'excitation. En conséquence, nous en sommes limités à faire quelques hypothèses relativement neutres sur l'entrée ; par exemple, bruit blanc à moyenne nulle et reporter tout l'effort de modélisation sur le système.

2.7.2.1. Codage Prédictif Linéaire (LPC)

Cette méthode connue de la production sous le sigle LPC (Linear Predictive Coding) se fonde sur les connaissances de la production de la parole et suppose que le modèle de production de la parole est linéaire selon le schéma (Figure.2.3).

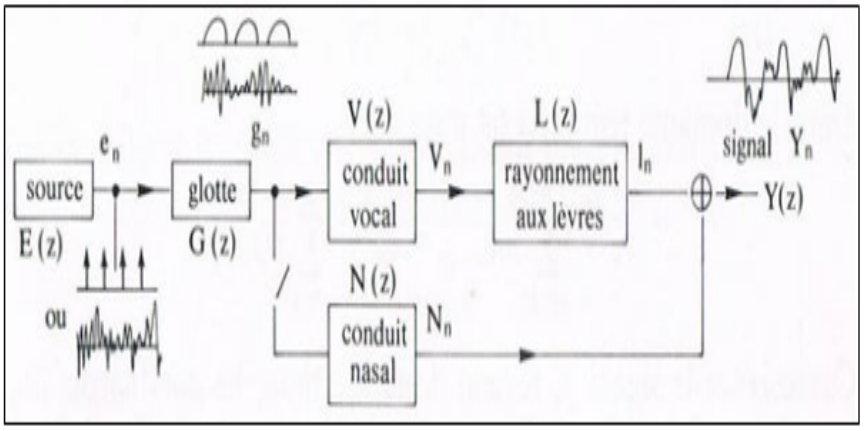


Figure 2.3 : Modèle général de production de la parole

Globalement, ce modèle peut se décomposer en deux parties : la source active, le conduit passif de manière plus détaillée, il peut se décrire de la manière suivante : l'onde est modélisée comme la sortie d'un filtre passe bas à deux pôles de fréquence de coupure d'environ 100 Hz (glotte), l'entrée « $\mathbf{e_n}$ » de ce filtre est un train d'impulsions de période « T_0 » pour les sons voisés ou un bruit blanc pour les sons non voisés (source).

Le modèle du conduit vocal est un filtre tout pôle (**AR** : **A**uto - **R**égressif) d'ordre 2M décomposable en une cascade de résonateurs à 2 pôles en série (tuyaux résonants). Le modèle du conduit nasal est un filtre pôle zéro ARMA (Auto Régressif à Moyenne

Ajustée) et le rayonnement aux lèvres peut se modéliser par un filtre tout zéro (**MA** : **M**oyenne **A**justée).

L'ensemble des conduits se comporte donc comme un système linéaire ARMA.

Modèle glottale :

$$G(z) = \frac{1}{(1-e^{-2\pi f_g T z^{-1}})^2} \quad (2.3)$$

Avec : $f_g = 100MH\ z$

Modèle du conduit vocal :

$$V(z) = \prod_{i=1}^{M} \left(\frac{1}{1-2e^{-2\pi B_i T}.\cos(2\pi F_i T)z^{-1}+e^{-4\pi B_i T} z^{-2}}\right) \quad (2.4)$$

F_i : Fréquence du formant Ne (i), B_i sa bande passante

Modèle du conduit nasal :

$$N(z) = \frac{1-2e^{-2\pi B'_N T}.\cos(2\pi F'_N T)z^{-1}+e^{-4\pi B_N T} z^{-2}}{1-2e^{-2\pi B_N T}.\cos(2\pi F_N T)z^{-1}+e^{-4\pi B_N T} z^{-2}} \quad (2.5)$$

Avec F_N et F'_N formant nasal ou anti formant nasal et respectivement, B_N et B'_N leurs bandes passantes.

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

$$H(z) = G(z).[1-\alpha).\mathrm{V(z)L(z)} + \alpha \mathrm{N(z)}] \quad (2.6)$$

Avec $0 \leq \alpha \leq 1$ pour un son nasal α=1 ; pour un son non nasal α=0.

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

$$H(z) = \frac{B(z)}{A(z)} \quad (2.7)$$

Dans le domaine temporel on aura :

$$y_n + \sum_{i=1}^{P} a_i y_{n-p} = e_n + \sum_{i=1}^{q} b_i e_{n-p} \quad (2.8)$$

Caractériser le signal y_n revient donc à estimer les coefficients $\{a_i ; b_i\}$.

Pour une source connue e_n (séquence d'impulsions ou bruit blanc). Souvent pour simplifier la résolution de ce problème, on suppose que $b_i = 0$, $i \geq 1$ ce qui rend le modèle AR [9].

2.7.2.2. Analyse cepstrale

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

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

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

Le but du cepstre est de séparer ces deux contributions par déconvolution. Il est fait l'hypothèse que un est soit une séquence d'impulsions (périodiques, de période T_0, pour les sons voisés), soit un bruit blanc pour les sons non voisés, conformément au modèle de production de la parole. Une transformation en Z permet de transformer la convolution en produit.

$$Y(z) = B(z).U(z) \tag{2.9}$$

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

$$\log|Y(z)| = \log|U(z)| + \log|B(z)| \tag{2.10}$$

Par transformation inverse, nous obtenons le cepstre. Dans la pratique, la transformation en

Z est remplacée par une **TFR** (**T**ransformé de **F**ourier **R**apide). L'expression du cepstre est donc :

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

Le cepstre qui ne fait appel à aucune information a priori sur le signal acoustique, est basé sur une connaissance du mécanisme de production de la parole. L'espace de représentation du cepstre ou espace quéferentiel est homogène par rapport au temps. Les premiers coefficients cepstraux contiennent l'information relative au conduit vocal. Cette contribution devient négligeable à partir d'un échantillon $\boldsymbol{n_0}$ qui correspond à la fréquence fondamentale $\boldsymbol{F_0}$ Les pics périodiques visibles au-delà de $\boldsymbol{n_0}$, reflètent les impulsions de la source.

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

La présence d'un pic important dans le cepstre renseigne d'une part, sur le caractère voisé ou non du son et d'autre part, constitue une bonne indication sur la fréquence fondamentale.

L'enveloppe spectrale du conduit vocal (structure formantique) est obtenue par une transformation supplémentaire (Figure 2.4).

Le spectre lissé débarrassé théoriquement de la contribution de la source ne contient que des informations sur le conduit vocal et en particulier sur ses extrema (Formants) [4].

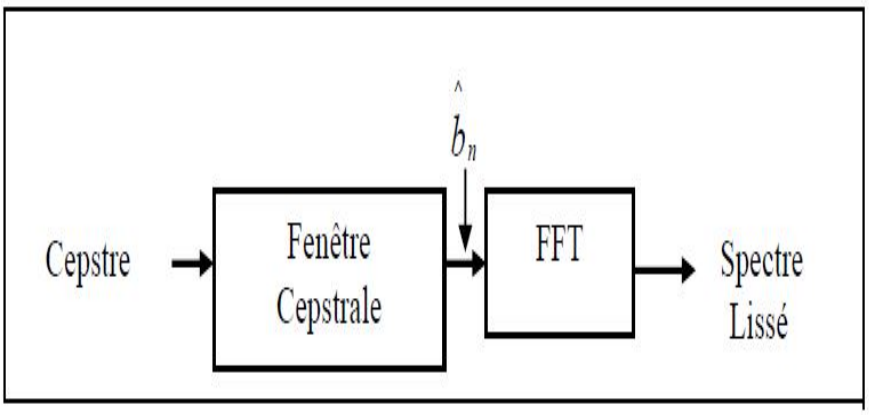


Figure 2.4 : Obtention de la structure formantique à partir du cepstre

2.7.3. Méthodes non paramétriques

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

L'analyse fréquentielle de la parole se ramène aux opérations de la **T**ransformée de **F**ourier **(TF)** et n'a d'intérêt que si elle s'applique à une période du signal vocal, donc sur une période assez courte.

2.7.3.1. Analyse par FFT

La FFT (Fast Fourier Transform ou transformée de Fourier rapide) est ici utilisée après échantillonnage du signal d'entrée basses fréquences. FFT est capable de capturer les signaux en temps réel avec une résolution spectrale très fine qui dépend du nombre de points N et de la fenêtre de pondération utilisée.

L'augmentation de la rapidité et de la résolution des convertisseurs analogique numérique permettra d'analyser des signaux à des fréquences de plus en plus élevées. En RAP, il est important de connaître l'évolution de ce spectre dans le temps.

Actuellement, les spectres sont obtenus numériquement par la **T**ransformée de **F**ourier **D**iscrète (**TFD)**, en particulier grâce à l'algorithme de la **T**ransformée de **F**ourier **R**apide (**TFR**) ou **F**ast **F**ourier **T**ransform (**FFT**) en Anglais. Cependant, le nombre de paramètres spectraux calculés sur une trame par FFT reste trop élevé pour un traitement automatique ultérieur. Pour une analyse très fine de la parole, la fenêtre de Hamming est déplacée à chaque fois de 128 points environ 10 ms (Figure 2.5) [12].

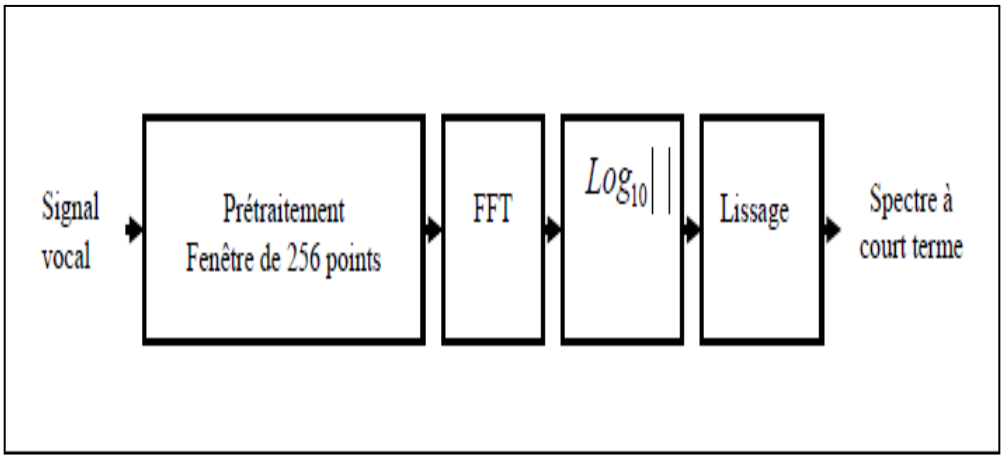


Figure 2.5 : Analyse numérique du signal parole par FFT

2.8. CONCLUSION

Dans ce chapitre, nous avons examiné les principales méthodes et techniques de la synthèse de la parole, ce qui nous a permis de passer en revue les différents paramètres et applications de la synthèse de la parole.

Le problème de synthèse s'est réduit à un problème de base de données et d'optimisation de la sélection d'unités. L'objectif est donc de réduire au maximum la modification du signal des unités de synthèse afin de préserver l'aspect naturel de la parole.

CHPITRE 3

ANALYSE DU SIGNAL DU CORPUS

3.1. Introduction

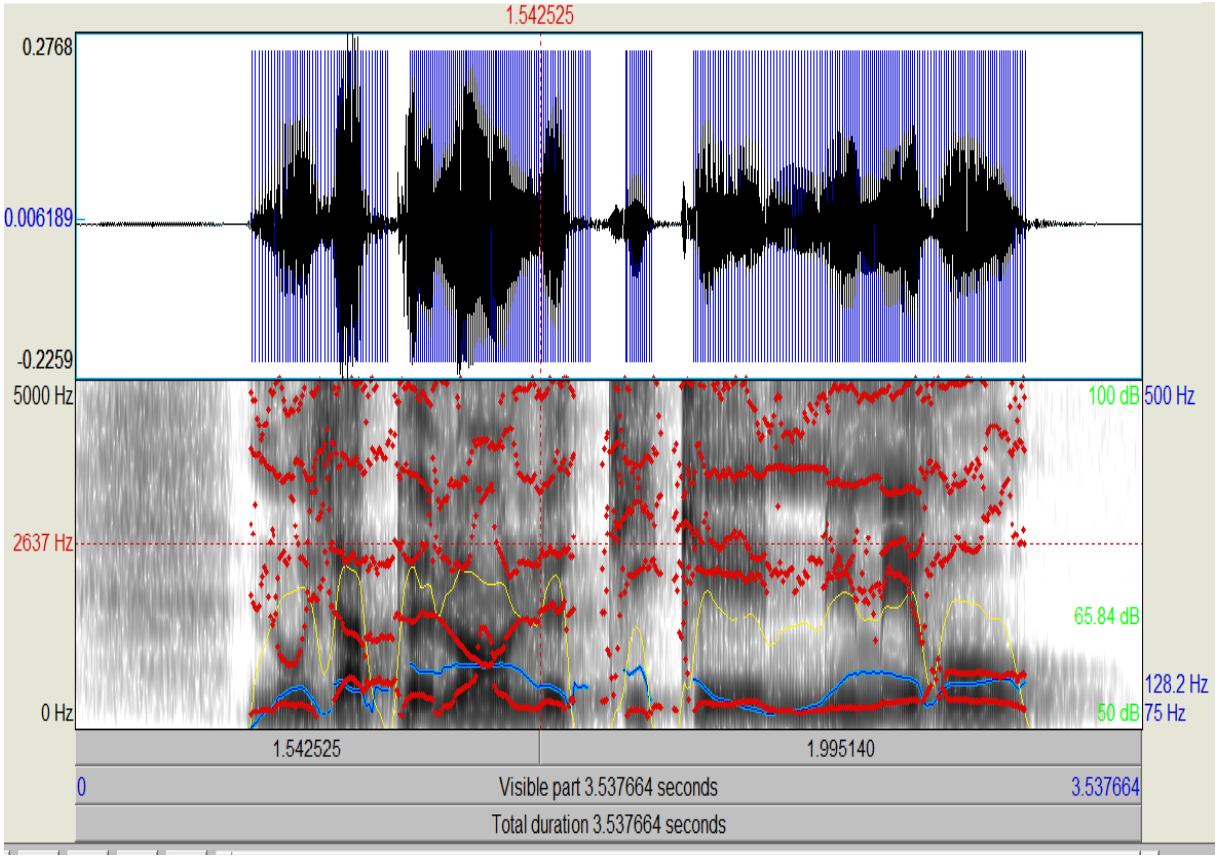
Le but de ce travail est de faire un système de synthèse de la parole qui est une combinaison de phrases fixes avec des mots variables, ce travail est utilisé dans les files d'attente d'Algerie Telecom pour la circulation normale des guichets, nous avons fait notre travail avec deux langues, la langue arabe et française, puisque ce travail contient une phrase et mot variable, donc nous utilisons la méthode de concaténation par phrases et mots combinés.

Ce chapitre nous allons nous intéresser au spectrogramme pour analyser notre signal, les outils Utilisés dans notre travail qui sont les deux logiciels Matlab et PRAAT et la méthode de synthèse utilisé **(SCUA).**

3.2. Spectrogramme

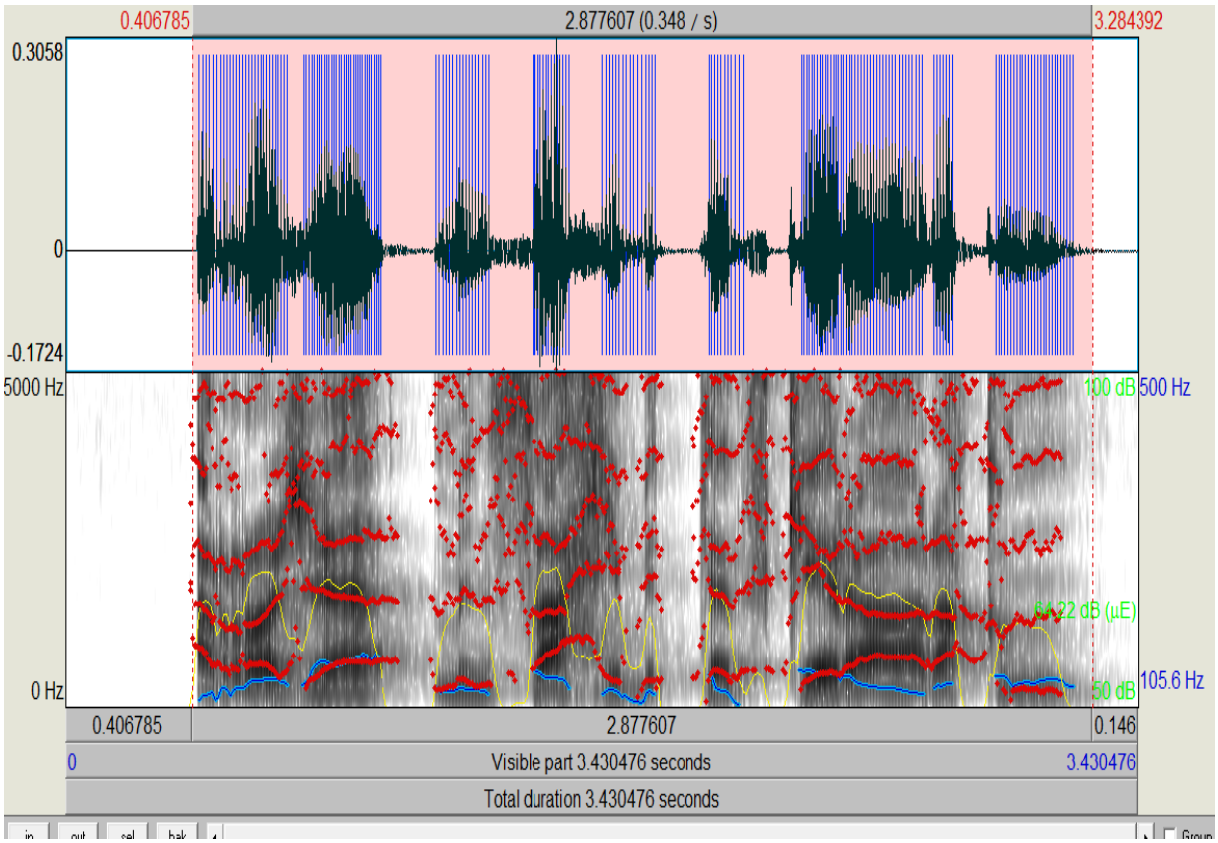
Le Spectrogramme et un outil de représentation 3D (temps, fréquence et énergie) graphique du spectre d'un rayonnement lumineux.

Nous verrons le spectrogramme de notre signal acoustique de travail qui est utilisé dans les guichets d'Algérie Telecom avec la langue Française (figuere3.1).



Figuere3.1 : Spectrogramme de la phrase fixe « Nous appelons le ticket N° »

Les services publics Algériens utilisent les deux langues, Arabe et Française, donc nous allons voir aussi le spectrogramme de notre signal en Arabe (Figuere3.2).



Figuere3.2 : Spectrogramme de la phrase fixe « الرجاء من صاحب التذكرة رقم »

D'après la durée de la fenêtre de pondération, nous pouvons distinguer deux types de spectrogrammes :

- larges bande**s** ou à bandes étroites selon la durée de la fenêtre de pondération. Les spectrogrammes à bandes larges sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10 ms); ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants. Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres.
- Les spectrogrammes à bandes étroites sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bandes horizontales. Le spectrogramme permet de mettre en évidence les différentes composantes fréquentielles du signal à tout instant.

Lecture de spectrogramme

La lecture de spectrogramme contient 4 étapes élémentaires :

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

Étape 2 : Savoir distinguer les consonnes et les voyelles :

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

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

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

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

- les voyelles antérieures, la distance entre F_1 - F_2 est supérieure à la distance entre F_2 - F_3 ;
- les voyelles postérieures, la distance entre F_1 - F_2est inférieure à la distance entre F_2 - F_3 ;
- les voyelles centrales, les formants sont plus (ou moins) équidistants [18].

3.3. Outils de Travail

Dans notre travail, nous avons deux types d'outils, l'outil d'analyse et l'outil de programmation.

3.3.1. Outil d'Analyse

Il existe beaucoup d'outil d'analyse que nous puissions utiliser comme PRAAT, CLAN, Speech Analysis, **S**peech **F**iling **S**ystem(**SFS**), WinPitch et Goldwave, tous ces logiciels nous permettent de visualiser le spectrogramme d'un signal de parole et sa forme d'ondes.

PRAAT

L'utilisation d'un outil d'analyse nous facilite de comprendre le signal acoustique, c'est pour cela nous utilisons PRAAT.

PRAAT est une application libre développée pour l'étude de sons vocaux par l'Institut de Phonétique d'Amsterdam. Il permet de faire des analyses assez poussées sur un signal vocal. Sa prise en main n'est pas immédiate, mais tout de même abordable. Au démarrage deux fenêtre s'ouvrent la fenêtre Objects, fenêtre principale qui gère les fichiers à analyser et la fenêtre Picture qui permet d'exporter des graphiques. Lorsque qu'un son a été chargé, la commande edit permet de faire afficher l'enveloppe temporelle du signal et un spectrogramme respectivement en haut et en bas de l'image (figure 3.3).

Dans notre travail, nous avons besoin que la fenêtre Objects.

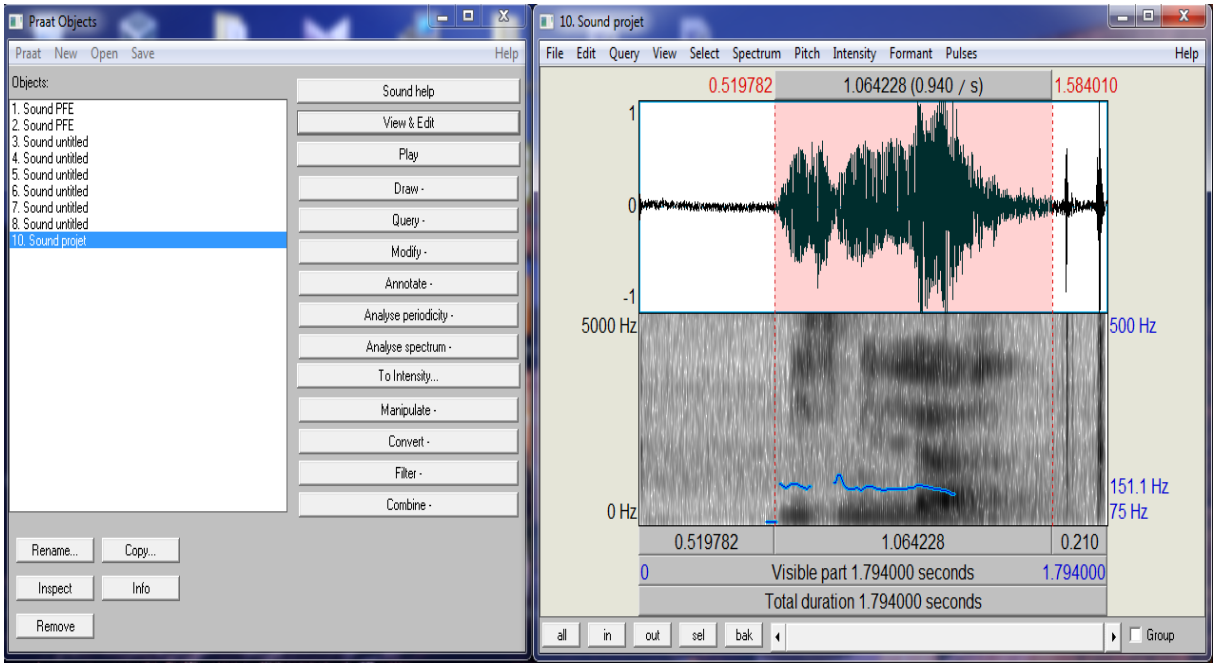


Figure 3.3 : Interface du logiciel PRAAT

Il est possible de faire des zooms pour visualiser la partie intéressante du signal. Le logiciel offre plusieurs options qui permettent d'afficher en surimpression du spectrogramme :

- la fréquence fondamentale : elle apparaît de couleur cyan. L'affichage peut se faire soit en Hz, soit en demi-tons par rapport à une valeur de référence par exemple à 440 Hz, pour l'obtenir, nous cliquons sur « Pitche » ensuite « Show pitch » . Les fréquences moyennes, maximum et minimum peuvent aussi être récupérées.
- Les formants : ils apparaissent en pointillés rouges, nous cochons sur « Formant » puis « Show formants » pour l'afficher.
- L'intensité : elle apparaît avec la couleur jaune, nous l'affichons avec l'outil « Intensity » après « Show Intensity ».

Le logiciel permet de faire afficher le spectre moyenné d'un signal. Il peut également être lissé grâce aux fonctions LPC et Cepstral smoothing. La valeur du centre de la gravité spectrale peut être obtenue dans la commande query de la fenêtre Objects.

En tant que logiciel de phonétique, il permet un affichage dans le plan F1-F2 des valeurs des formants.

Le logiciel PRAAT est assez complet et offre un nombre important de fonctionnalités pour une étude avancée de la voix (figure 3.4).

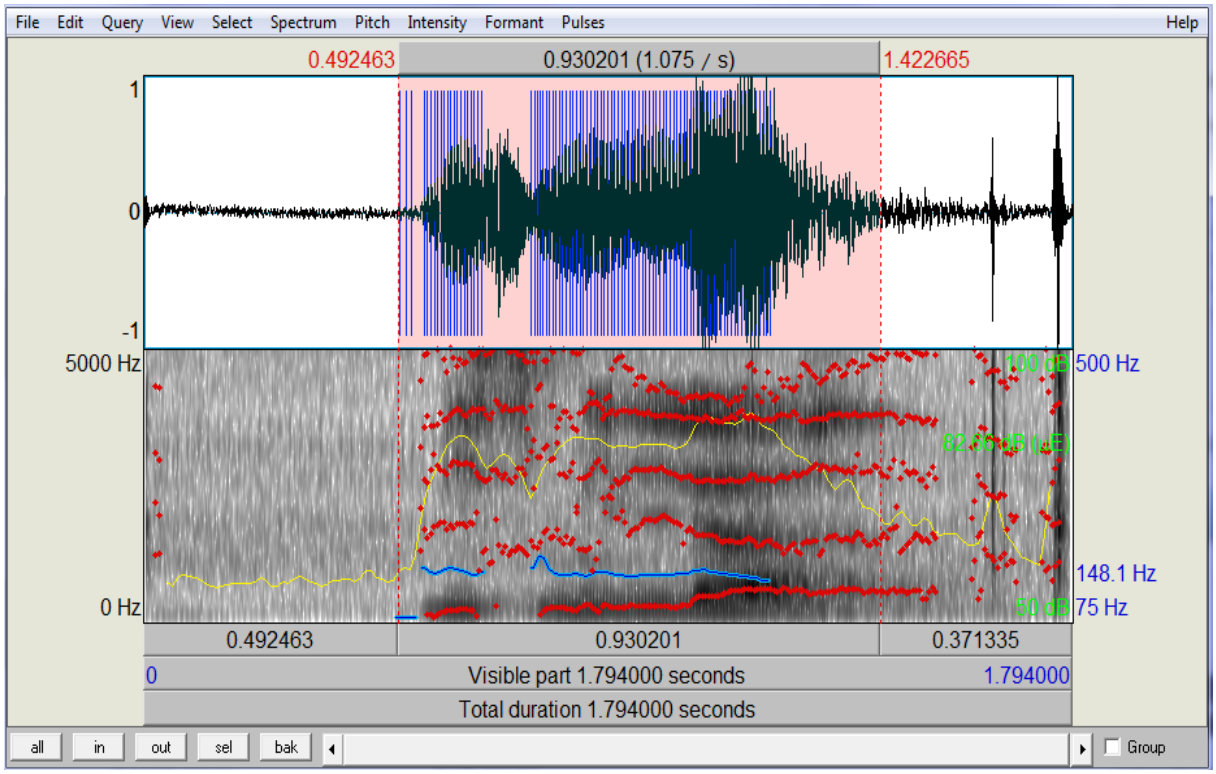


Figure 3.4 : Les propriétés du spectrogramme sur le logiciel PRAAT

3.3.2. Outil de Programmation

Pour lier entre une phrase fixe et les mots variables, nous utilisons **MATLAB** (MATrix LABoratory).

MATLAB est un langage de haut niveau et un environnement interactif pour le calcul numérique, la visualisation et la programmation (Figure 3.5).

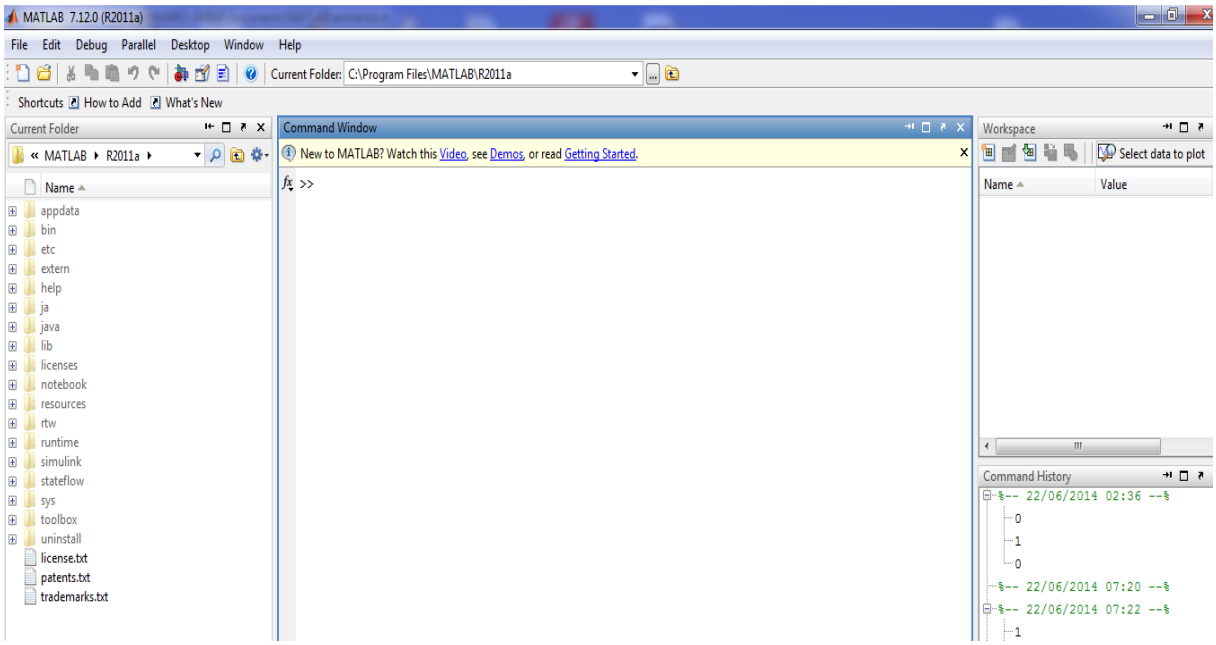


Figure 3.5 : Interface de MATLAB

3.4. les Méthodes de la Synthèse de la Parole

Nous enregistrons les signaux vocaux par un microphone. Les signaux qui ont été émis par le locuteur ce sont des signaux continus, d'énergie finie et non-stationnaire doivent être numérisé à l'aide d'un Convertisseur Analogique/Numérique.

Comme la voix humaine est constituée d'une multitude de sons, souvent répétitifs, le signal peut être compressé pour réduire le temps de traitement et l'encombrement en mémoire. Nous avons échantillé ces signaux pour que nous puissions les utiliser à la reconnaissance, l'analyse et la synthèse de la parole. Il existe beaucoup de méthodes appliquées à la synthèse de la parole.

3.4.1. Synthèse Par Règles (SPR)

La **S**ynthèse **P**ar **R**ègles est une méthode qui a eu beaucoup de succès dans le contexte de la synthèse de la parole à partir du texte. Des règles sont utilisées pour estimer les paramètres nécessaires. Cette approche est fondée sur un modèle de production du signal vocal, modèle commandé par un nombre restreint de paramètres. La synthèse se décompose alors en deux étapes : une transformation des informations phonético- prosodiques, à l'aide de règles contextuelles, en commandes permettant de spécifier l'évolution temporelle des paramètres du modèle de synthèse. Les paramètres ainsi déterminés sont utilisés pour synthétiser le signal acoustique.

Dans ce type de synthèse, les caractéristiques supra-glottiques sont modélisées à l'aide d'un filtre linéaire dont la Fonction de Transfert (FT) varie au cours du temps. Les paramètres utilisés pour le contrôle du filtre sont les paramètres formantiques, à savoir la fréquence centrale, la bande passante et l'amplitude des maxima significatifs de la FT du conduit vocal. Pour obtenir une parole intelligible, il suffit de spécifier les paramètres des 3 formants (voyelles) à 5 formants (consonnes) les plus importants, d'où la dénomination de synthèse par formants couramment employée pour ce type de synthèse. Une telle approche ne permet pas de restituer un signal de parole apparaissant naturel. La qualité médiocre obtenue résulte d'une part de la difficulté à modéliser suffisamment d'une maniére précise les trajectoires acoustiques et d'autre part, de la modélisation trop grossière du signal glottique.

Les synthétiseurs par règles ont principalement la faveur des phonéticiens et des phonologistes. Ils permettent une approche cognitive, générative du mécanisme de la

phonation. Ils sont basés sur l'idée que, si un phonéticien expérimenté est capable de «lire» un spectrogramme, il doit lui être possible de produire des règles permettant de créer un spectrogramme artificiel pour une suite de phonèmes donnés. Une fois le spectrogramme obtenu, il ne reste plus alors qu'à générer l'audiogramme correspondant (figure 3.6).

Dans un premier temps, on fait lire par un locuteur professionnel un grand nombre de mots, généralement de type **C**onsonne-**V**oyelle-**C**onsonne [**CVC**] et on les enregistre sous forme numérique. Les mots sont choisis de façon à constituer un corpus. On modélise alors ces données numériques à l'aide d'un modèle paramétrique de parole, qui a pour rôle de séparer les contributions respectives de la source glottique et du conduit vocal et de présenter cette dernière sous forme compacte, plus propice à l'établissement des règles.

On commence par inspecter globalement l'ensemble des données, de façon à établir la forme générale des règles à produire. On précise alors les valeurs numériques des paramètres intervenant dans ces règles (les fréquences des formants, ou les durées des transitions, par exemples) par un examen minutieux du corpus. Il est à remarquer que cette étape d'estimation est menée sur une seule voix : un moyennage interlocuteur aurait peu de signification dans ce contexte.

De même, les règles provenant de synthétiseurs déjà existants ne peuvent resservir que dans la mesure où elles modélisent des caractéristiques articulatoires générales plutôt que des particularités du locuteur ayant enregistré le corpus (sauf bien entendu si l'on cherche à produire des règles caractérisant précisément le passage d'une voix à une autre). La mise au point du synthétiseur s'achève par un long processus d'essais-erreurs, afin d'optimiser la qualité de la synthèse.

Les grands avantages de cette méthode, ce sont la grande souplesse d'utilisation, la facilité d'extension, et surtout la grande portabilité de ces systèmes facilitant leur intégration dans une large gamme de produits [12].

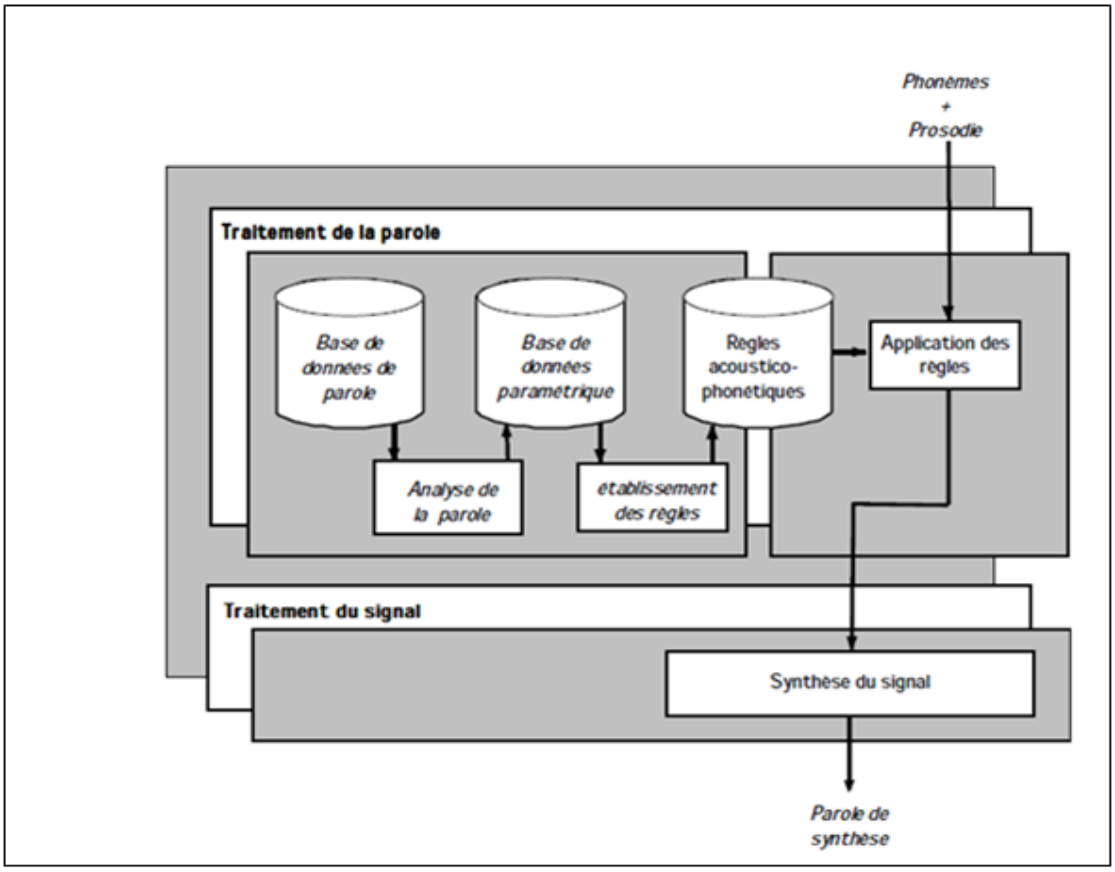


Figure 3.6 : Schéma de conception et fonctionnement typique d'un système de synthèse par règles [7]

3.4.2. Synthèse par Concaténation d'Unités Acoustiques (SCUA)

Avec les capacités de stockage de données devenues conséquentes, le choix se tourne d'avantage sur la synthèse par concaténation d'unité que sur celle par règles, limitée par son manque cruel de naturel. Elle ne consiste plus en une modélisation à proprement parlé du phénomène acoustique de la voix, mais d'une mise bout à bout de segments de paroles enregistrés. Ces segments ont une durée définie par les technologies, mais également par le cahier des charges. Ainsi, plus les segments seront long (mots, phrases, etc.), plus la qualité du résultat final sera intelligible et naturel, au prix d'un coût important en terme de stockage mais aussi de développement.

Historiquement, la première synthèse de ce genre est sans doute la synthèse par mots, que l'on peut toujours retrouver, par exemple dans les annonces ferroviaires. La simplicité du système n'a d'égal que son manque de naturel et de flexibilité, se contentant de lire des séquences d'échantillons sonores sans aucun traitement phonétique ou prosodique. On se tourne très vite vers d'autres solutions bien plus avantageuses (figure 3.7).

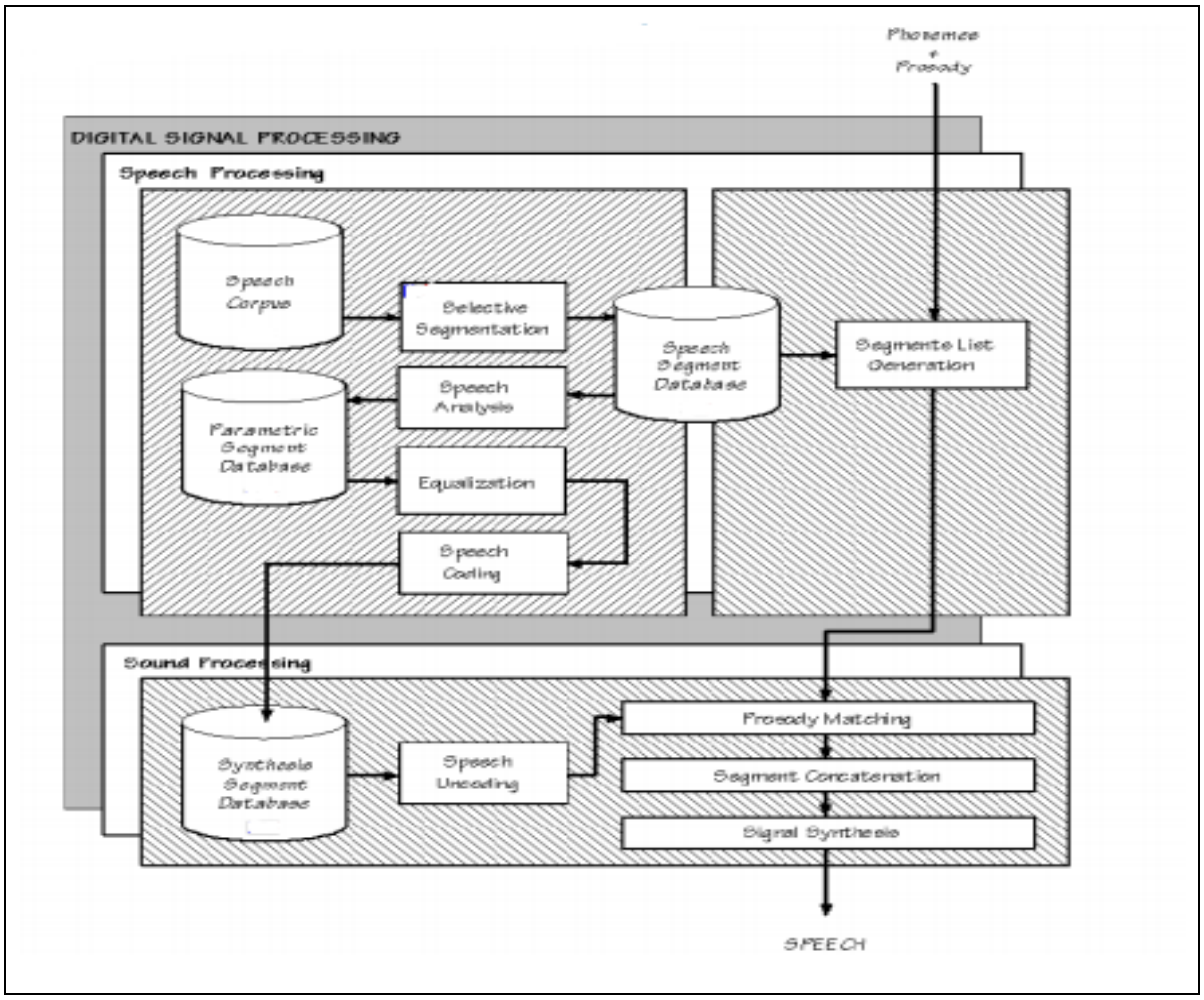


Figure3.7 : Schéma de conception et fonctionnement typique d'un système de synthèse par règles [7].

3.5. Conclusion

Dans ce chapitre nous avons exposé les logiciels utilisés pour l'exploitation de notre corpus et étudié aussi les méthodes de la synthèse de la parole. Par la suite, nous avons obtenu des spectrogrammes de notre travail en passant par plusieurs étapes d'analyse et de visualisation.

CHAPITRE 4

GESTION AUTOMATIQUE DES FILES D'ATTENTE

4.1. Introduction

Dans ce chapitre nous allons présenter une simulation d'un **S**ystème de Gestion Automatique des Files d'Algérie Telecom (**SGAFA**), avec son corpus, et finissons par des tests de perception subjective afin de pouvoir évaluer les résultats obtenus (signal vocal de sortie de l'interface) ce qu'il concerne l'intelligibilité et l'aspect naturel.

4.2. Système de Gestion Automatique des Files d'Attente

L'objectif de notre travail est d'élaborer un **S**ystème Vocal de **G**estion **A**utomatique des **F**iles d'**A**ttente (**SGAFA)**, dans les services clientèles d'Algérie Telecom. Cette gestion donne des annonces vocales qui seront faites pour le prochain client qui passera au guichet qui est vide. Notre système va se déclencher automatiquement par un lancement de signal vocal, en Arabe Standard et en Français qui annonce le numéro de ticket avec le numéro du guichet que doit choisir le client.

4.3. Elaboration du Corpus

Dans le but de notre travail, nous avons utilisé un corpus en parole continue des phrases en Arabe Standard et en Français, prononcées par deux personnes (locutrice et locuteur arabophone), cet enregistrement qui a deux copies, en left et right, nous choisissons l'enregistrement que, ce corpus contient des expressions utilisées pour des annonces vocales automatiques des files d'attente d'Algérie Telecom. Ce corpus contient 4 phrases fixes (2 en Arabe Standard et 2 en Français) et plus les 40 numéros qui sont les mots variables (20 en Arabe Standard et 20 en Français), nous avons donc en global 44 enregistrements.

Ce corpus est surnommé **GAFA** : **G**estion **A**utomatique des **F**iles d'**A**ttente.

Nous justifions le choix de ce type de corpus (parole continue au lieu de l'utilisation de logatomes) par le fait qu'il est préférable d'étudier les segments dans un continuum vocal pour pouvoir prendre en considération les effets de coarticulation existants entres les phonèmes.

4.3.1. Enregistrement de Corpus

L'enregistrement du corpus a été faite à l'**I**nstitut **S**upérieur des **M**étiers des **A**rts du **S**pectacle et de l'Audiovisuel (ISMAS) d'Alger, nous avons enregistré notre corpus selon les conditions d'enregistrement suivantes :

- la fréquence d'échantillonnages : Fe = 48 kHz et le codage : 24 bits ;
- le format d'enregistrement : multiple mono (stéréo) ;
- logiciel utilisé Pro Tools version 8 ;
- la chambre sourde ;
- le type de parole : phrases continu en arabe et en Français ;
- les signaux acoustiques sont enregistrés en format (WAV).

4.3.2. Equipement utilisés en enregistrement

Le matériel utilisé est :

- Microphone dynamique professionnel unidirectionnel Electrodynamique [Beyer dynamic M 69 TG] (figure 4.1) ;
- Station Pro Tools Version 8 qui enregistre les sons avec une bonne qualité (fig 4.2) ;
- une cabine Speaker : c'est une chambre isolée et contient des Microphones et des casques, séparé à la cabine technique avec un verre transparent et isolant (figure 4.3) ;
- une cabine technique : contient une table de mixage, une carte d'acquisition, un micro-ordinateur, des Haut-parleurs, des Microphones (figure. 4.3).
- Table de mixage (figure 4.4).

Figure 4.1 : Microphone Beyer dynamic M 69 TG

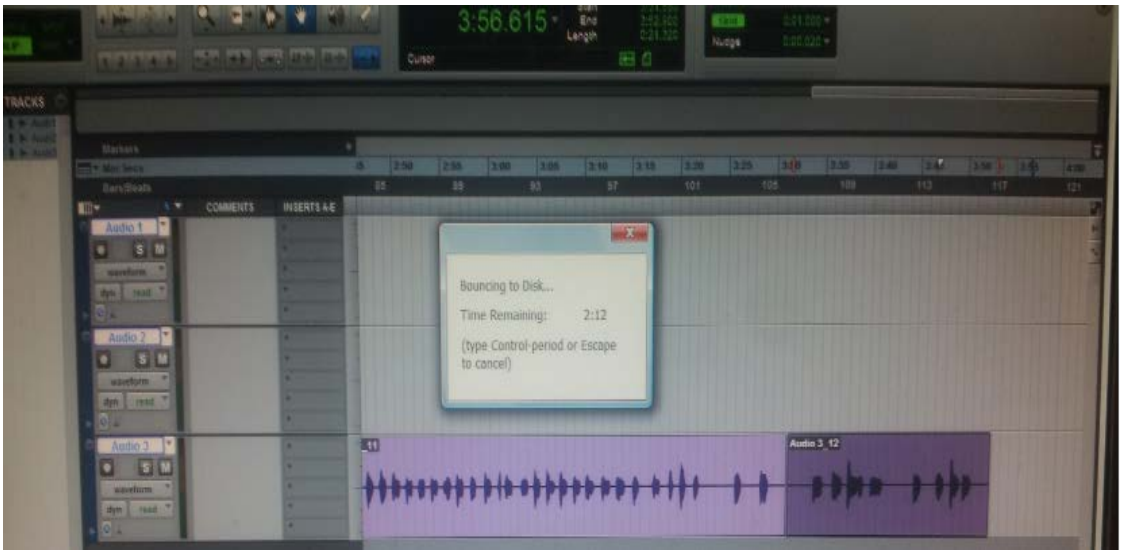


Figure 4.2 : Station Pro Tools version 8

Figure 4.3 : Cabine Speaker + cabine technique

Figure 4.4 : Table de mixage

4.4. ALGORITHME DE SIMULATION DU SGAFA

Nous avons élaboré un algorithme, qui est basé sur :

- l'enregistrement du corpus ;
- les fichiers sonores ;
- la base de données ;
- la sélection des segments choisis ;
- les annonces vocales (sortie orale).

L'enregistrement du corpus contient des fichiers sonores comprenant : deux phrases fixes et des mots variables (numéros). Nous avons nommé chaque fichier des segments sonores afin d'avoir une référence dans la base de données.

Les sélections des segments appropriés : les fichiers se divisent en deux types de catégories, des fichiers en Arabe Standard et les autres en Français.

Chaque catégorie est divisée en deux :

- deux phrases fixes ;
- mots variables.

Dans la Sélection des segments, nous allons choisir les fichiers sonores tel qu'à chaque fois lorsqu'un guichet va être vide nous définissons le mot variable, c'est-à-dire le numéro suivant, les deux phrases fixes sont répétitives (figure 4.5).

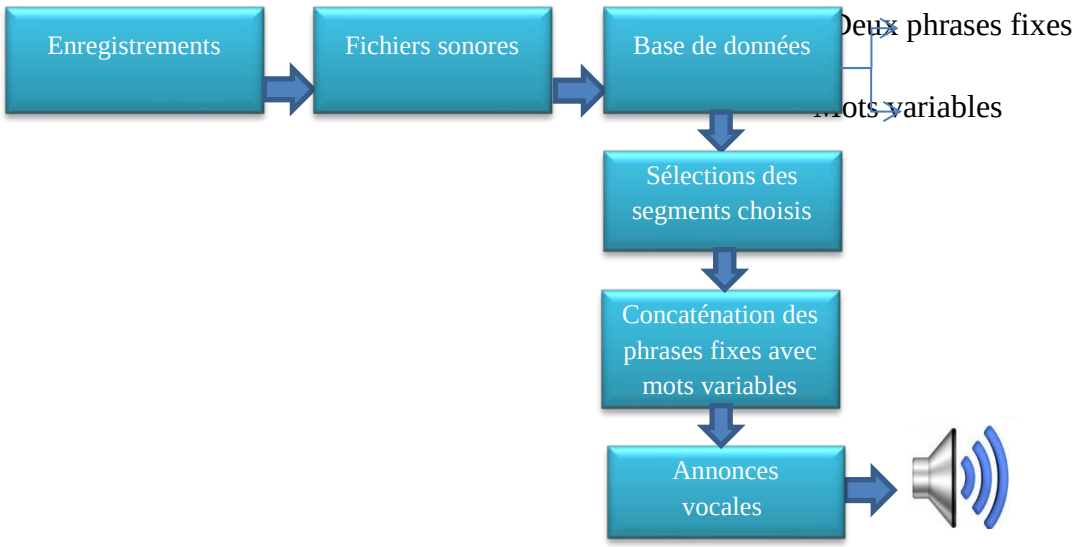


Figure 4.5 : Algorithme du SGAFA

4.5. Organigramme du SGAFA

Dans l'organigramme, nous allons mettre une incrémentation à chaque fois que le guichet sera vide. Nous appelons automatiquement le ticket suivant, si tous les guichets sont occupés, nous reviendrons au début et nous vérifions les guichets est ce qu'ils sont vides, c'est -à- dire nous faisons une boucle fermée (Figure 4.6).

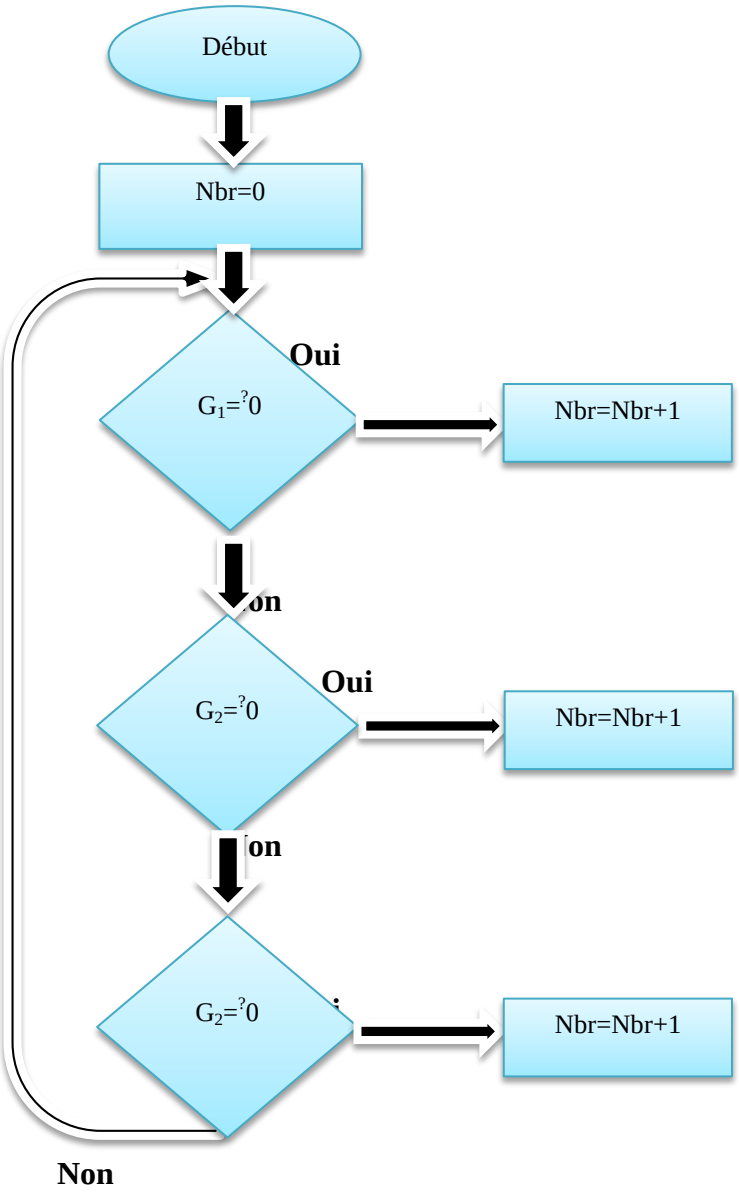


Figure 4.6 : Organigramme du SGAFA

4.6. Synthèse par Concaténation des Phrases et Mots Combines

La synthèse par concaténation d'unités est une méthode éprouvée en synthèse de la parole, et est actuellement la méthode permettant d'obtenir la meilleure qualité dans ce domaine. Celle-ci consiste à concaténer des segments provenant d'une base de données pré-enregistrée. Selon le système, ces segments peuvent correspondre à des phonèmes, des diphones, ou des unités plus longues pouvant aller jusqu'à des mots ou morceaux de phrases entiers comme dans notre travail. Le choix des segments est alors effectué grâce à des fonctions de coût.

Les discontinuités au niveau des jonctions entre ces segments sont ensuite lissées, et des transformations sont appliquées au signal pour lui donner la prosodie désirée.

Cette approche, en conservant un signal au plus proche du son original utilisé, permet une qualité d'élocution et de timbre naturelle très proche de la parole humaine (figure 4.7).

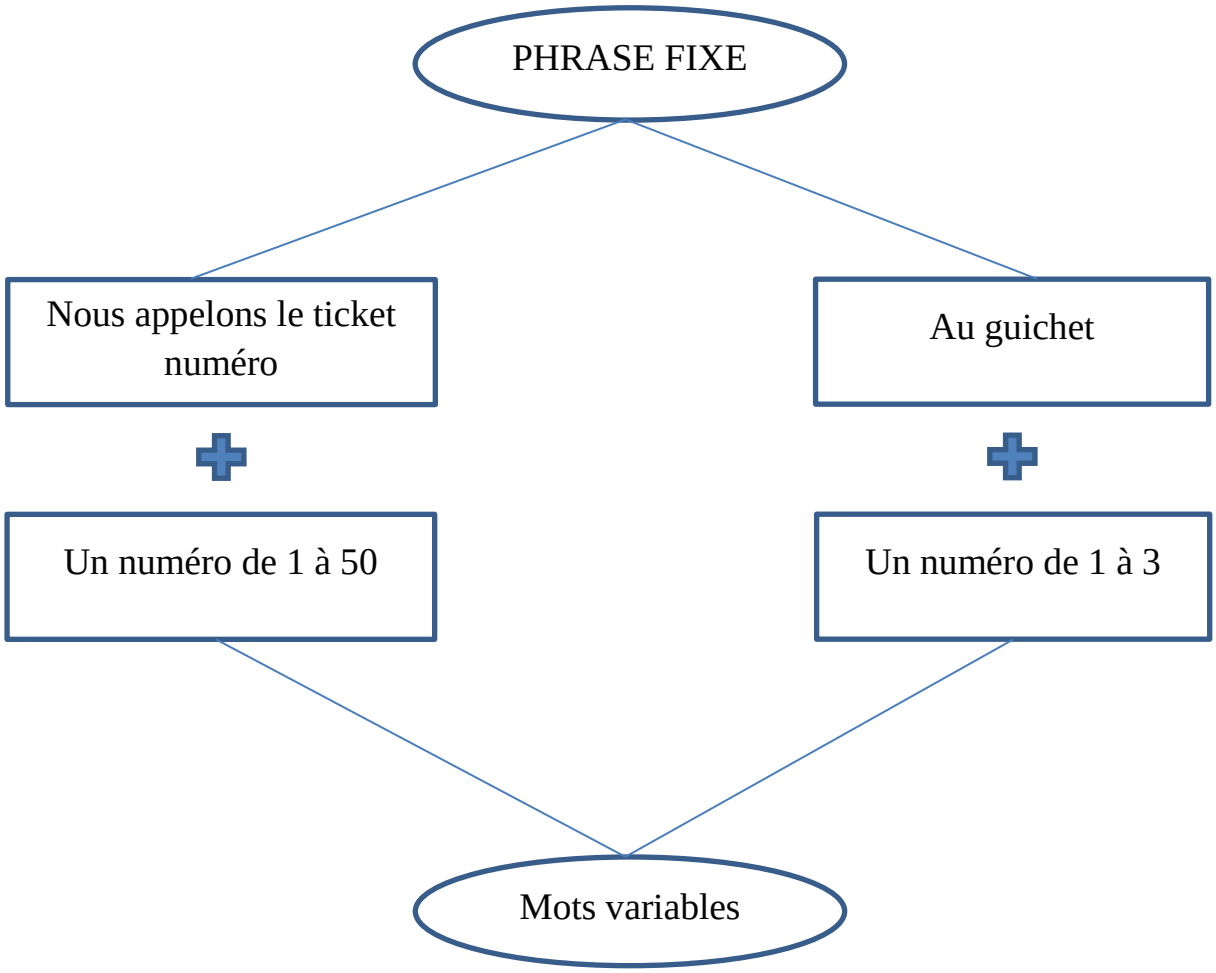


Figure 4.7 : Assemblage des parties fixes avec les parties variables

Cet organigramme fait la concaténation entre les phrases fixes qui sont deux et les mots variable qui sont des nombres de 1 jusqu'à 20.

Nous avons fait les deux phrases en Français et en Arabe Standard aussi avec aussi les numéros en Arabe (Tableau 4.1).

Nbr : c'est les nombres de tickets qui s'incrémentent toujours de 1 tel que $0 < Nbr < 20$

G_i : les guichets qui sont soit vides ou pleins tel que $1 < i < 3$ et avec :

- 0 implique que le guichet est vide ;
- 1 implique que le guichet est plein.

Tableau 4.1 : Les Phrases Fixes et les Mots Variables

Français	Les Phrases en Français
Phrase 1	Nous appelons le ticket numéro
Phrase 2	Au guichet numéro
Arabe	Les Phrases en Arabe
Phrase 1	الرَّجَاء مِنْ صَاحِب التَّذْكِرَة رَقَمْ
Phrase 2	التَّقَدَّمْ إِلَى الشُّبَاكْ رَقَمْ

Français	Les Numéros en Français
Mots Variables	Un, deux, trois, quatre, cinq,……., dix-huit, dix-neuf, vingt
Arabe	Les Numéros en Arabe
Mots Variables	واحد، إثنان، ثلاثة، أربعة، خمسة،.....، ثمانية عشر، تسعة عشر، عشرون

4.7. Tests d'Evaluation du SGAFA

Dix personnes ont participé au test d'évaluation subjective du SGAFA.

En fonction de leurs réponses, nous avons établi cinq niveaux de qualité : Mauvais, Médiocre, Passable, Bon et Excellent.

Les scores obtenus sont définis dans le Tableau 4.2. Une autre représentation des résultats sous forme d'histogrammes, est illustrée dans la figure 4.8.

Tableau 4.2 : Evaluation du SGAFA

Qualité / Résultats	Mauvais	Médiocre	Passable	Bon	Excellent
Nombres de Personnes	0	0	1	5	4
Pourcentage (%)	0	0	10	50	40

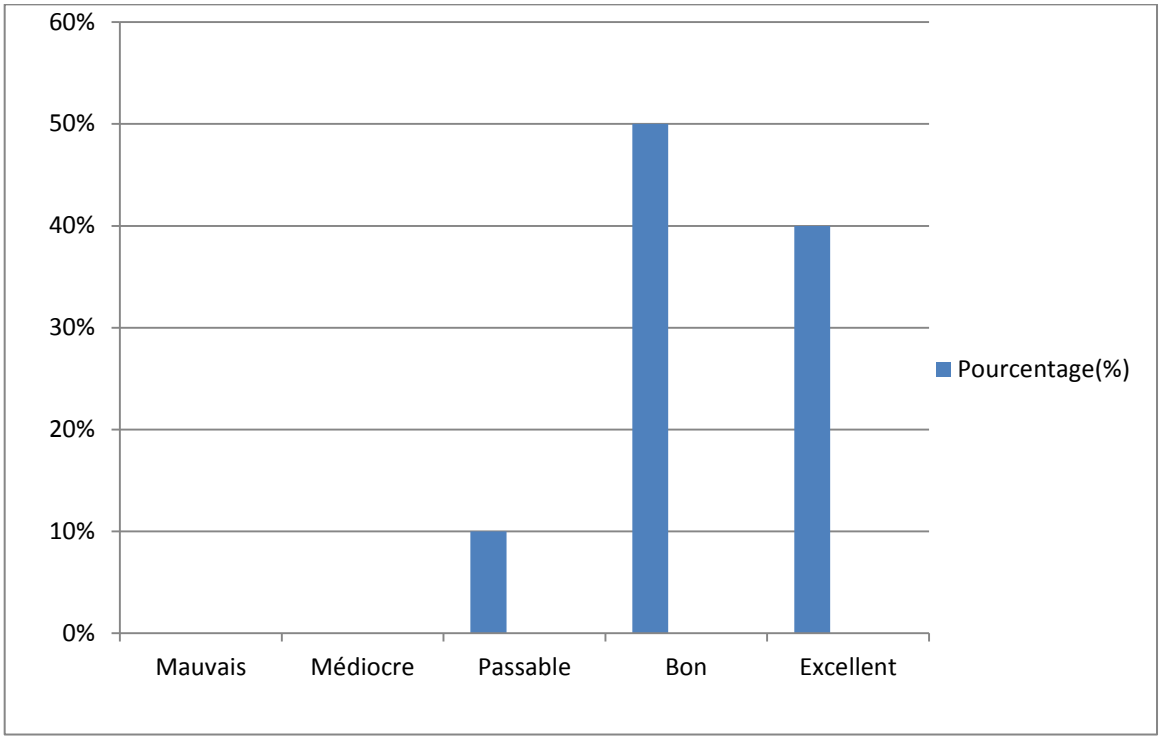


Figure 4.8 : Evaluation sur la parole synthétisée par 10 personnes

4.8. Interprétation

D'après la figure 4.8, les résultats du SGAFA que nous avons obtenus sont de bonnes qualités, c'est-à-dire qu'ils sont intelligibles et naturels. Ce qui montre que notre système est fiable et il peut aider les clients dans les services clientèles d'Algérie Telecom.

4.9. Avantages du SGAFA

Chaque système dans notre vie a des avantages pour aider les gens et faciliter la vie quotidienne, et notre système a des avantages que nous allons citer :

- éviter les conflits entre les clients ;
- ne pas subir la pression des personnes en attente ;
- faciliter la gestion des services clientèle et réduire le temps d'attente ;
- les clients vont être satisfaits ;
- avoir ses tâches planifiées et rappelées
- faciliter d'extension.

4.10.Conclusion

Nous avons exposé dans ce chapitre l'algorithme et l'organigramme de simulation de la synthèse de parole par concaténation de mots et phrases combinés. Nos résultats expérimentaux correspondant aux tests de perception subjective, subis par 20 personnes, ont montré le bon enregistrement du corpus GAFA ainsi que la bonne segmentation manuelle de logiciel PRAAT, ce qui implique qu'il y a eu compréhension des phrases prononcées avec une bonne qualité de la parole synthétique.

CONCLUSIONS GENERALE ET PERSPECTIVES

Les analyses acoustiques et les tests sur les signaux acoustiques que nous avons utilisés dans notre travail, nous ont permis de bien apprendre des notions sur le traitement de la parole et nous avons aussi compris le fonctionnement d'un système de synthèse vocale.

Le but de notre travail est de concevoir un système de synthèse de la parole pour la gestion automatique des files d'attente d'Algérie Telecom. Pour parvenir à cet objectif, nous avons élaboré un corpus composé de deux phrases fixes et des mots variables, prononcés par une locutrice pour la langue Française, et un locuteur pour la langue Arabe.

La synthèse par concaténation de la parole est simple et elle est capable de produire des annonces de haute qualité qui rapproche du naturel.

Notre système de gestion automatique, est d'une bonne qualité : intelligible et naturel. Cette qualité dépend des techniques et des méthodes de synthèse utilisées, mais également du soin apporté à la modélisation linguistique et prosodique.

Comme perspectives à ce travail, il serait très intéressant de faire une étude évaluative pour améliorer la qualité de la parole synthétique.

Un affichage pour les annonces vocales des guichets pour faciliter la gestion des services clientèle :

- Utilisation d'autres langues (Anglaise, Chawia, Kabyle, etc.) ;
- Mettre une interface graphique pour le bouton de guichets s'ils sont vides ;
- Utilisation d'une technique ou ajuster les paramètres prosodiques (la F_0, la durée, l'intensité) pour améliorer la qualité de la voix synthétique.

REFERENCES BIBLIOGRAFIQUES

[1] A. Chentir, Etude de la Microprosodie en vue de la Synthèse de la parole en Arabe Standard, Thèse de Doctorat, Ecole Nationale Polytechnique -Alger (Algérie), 01/10/2009.

[2] T. Hézard, Production de la voix : exploration, modèles et analyse/synthèse, Thèse de Doctorat, Université Pierre et Marie Curie-PARIS VI (France), 9/12/2013.

[3] T. Hueber, Synthèse de la parole à partir d'imagerie ultrasonore et optique de l'appareil vocal, PFE, Ecole Supérieure de Chimie Physique Electronique –Lyon (France).

[4] M. Aissiou, Application des Algorithmes Génétiques au Décodage Acoustico- Phonétique de la parole en Arabe Standard, Ecole Nationale Polytechnique-Alger (Algérie), 30/06/2008.

[5] Z. Benselama, pathologie du langage parlé Arabe : cas des sigmatismes occlusifs et constrictifs, Thèse de Doctorat : Ecole Nationale Polytechnique-Alger (Algérie), 15/12/2007

[6] M.Garnier , Approche de la qualité vocale dans le chant lyrique : perception, verbalisation et corrélats acoustiques. Mémoire de DEA., ATIAM Université de Paris VI, 2003.

[7] T. Dutoit, Introduction au traitement automatique de la parole notes de cours /DEC2, Faculté Polytechnique de Mons, LCTS Lab, France, 2000.

[8] S. Djeghiour, Application des Réseaux de Neurones à la synthèse de la Parole En Arabe Standard, Mémoire de Magister, Ecole Nationale Supérieure Des Sciences Humaines, Bouzaréah-Alger (Algérie), 2011.

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

[10] A. Ounnas, synthèse de la parole en Arabe Standard, Mémoire de Magister, Ecole Nationale Polytechnique Alger (Algérie), Décembre 2011.

[11] S. Baloul, Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé, Thèse de Doctorat, Spécialité : Informatique, Université du Maine, Le Mans, France, 2003.

[12] B. Mohamed El Amine & B. Moussaab, Synthèse de la Parole par Unités Variables en Vue d'un Guide Touristique en Algérie, PFE, Ecole Nationale Polytechnique-Alger (Algérie), 26/06/2013.

[13] A. Zaki, Modélisation de la prosodie pour la synthèse de la parole en AS à partir du texte. Thèse de Doctorat en Automatique, productique, signal et image. Université Bordeaux I, France, 2004.

[14] S. Rouibia, Prise en compte de critères acoustiques pour la synthèse de la parole. Thèse de Doctorat en Traitement du signal et Télécommunications, Ecole Nationale Supérieure des Télécommunications de Bretagne en habilitation conjointe avec l'université de Rennes 1, France, 2006.

[15] C. H. Shadle and R. I. Damper, Prospects for Articulatory Synthesis : A Position Paper. Proceedings of 4th ISCA Workshop on Speech Synthesis, Pitlochry, pp. 121-126, 2001.

[16] http://univ.ency-education.com/uploads/1/3/1/0/13102001/indpet2an-cours_matlab.pdf

[17] B.Soufiane, Annonces Vocales Automatiques des Stations d'Arrêt du Tramway d'Alger, PFE, Ecole Nationale Polytechnique-Alger (Algérie), Juin 2013.