

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



DÉPARTEMENT D'ÉLECTRONIQUE

Laboratoire de Signal et Communications

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

Présentée par :

Mr BOUALAM Mohamed El Amine

Thème :

**Synthèse de la Parole par méthode TD-PSOLA**

Soutenue le : 26 Juin 2013

Devant le Jury :

C. LARBES	Professeur	ENP	Président
M. GUERTI	Professeur	ENP	Rapporteur
M. MAAMRI	CC	ENP	Examineur

**Promotion: Juin 2013**



## ملخص :

الهدف من عملنا هذا هو إعداد نظام لتركيب الكلام بالاعتماد على تقنية TD-PSOLA ، بإجراء تغييرات عرضية بتغيير التردد الابتدائي و الزمن ، TD-PSOLA تعتمد على تقسيم الإشارة الصوتية إلى نوافذ موحدة الزمن مع دور التردد الابتدائي وهذا للحفاظ على الانسجام مع الأخذ بعين الاعتبار الشكل الزمني لدور التردد الابتدائي

**الكلمات المفتاحية:** تركيب الكلام ، TD-PSOLA ، التغييرات العرضية ، التردد الابتدائي.

## Résumé :

L'objectif de notre travail est d'élaborer un système de synthèse de la parole en se basant sur la méthode TD-PSOLA (Time Domain -Pitch Synchronous Overlap and Add), en effectuant des modifications prosodiques sur la fréquence fondamentale et le temps. TD-PSOLA est basée sur la décomposition du signal de la parole en fenêtres recouvrantes synchronisées sur les périodes du fondamental, afin de préserver la cohérence tout en tenant compte de la structure temporelle des périodes du fondamental

**Mots clés :** Synthèse de la Parole, TD-PSOLA, modifications prosodiques, Fréquence Fondamentale.

## Abstract :

The objective of our work is to develop a system of speech synthesis based on method TD-PSOLA (Time Domain - Pitch Synchronous Overlap and Add), by carrying out prosodic modifications over the fundamental frequency and time. TD-PSOLA is based on the decomposition of the speech signal in synchronized recovering windows over the periods of fundamental, in order to preserve coherence while taking account of the temporal structure of the periods to the fundamental.

**Key words:** Speech Synthesis, TD-PSOLA, prosodic modifications, Fundamental Frequency.

# DEDICACES

Je dédie ce modeste travail à celle qui m'a donné la vie, le symbole de tendresse, qui s'est sacrifiée pour mon bonheur et ma réussite, à ma mère.

A mon père, école de mon enfance, qui a été mon ombre durant toutes les années des études, et qui a veillé tout au long de ma vie à m'encourager, à me donner l'aide et à me protéger.

Que dieu les gardes et les protège.

A mes sœurs Fatiha, Ibtissem et Zakia et à toutes ma famille.

A mon binôme Moussaab.

A tous mes amies.

A tous ceux qui m'aiment.

A tous ceux que j'aime.

Je dédie ce travail

*Mohamed El Amine*

# Remerciements

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

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

*Je remercie les membres du jury, qui m'ont fait l'honneur de participer au jugement de ce mémoire.*

*J'exprime ma reconnaissance à Monsieur **LARBES chérif**, Professeur à l'Ecole Nationale Polytechnique, d'avoir accepté de présider le jury de mon mémoire.*

*Je remercie également Monsieur **MAMMERI Mohamed**, Maître de conférences à l'Ecole Nationale Polytechnique, d'avoir accepté de faire partie de mon jury.*

*Je tiens à remercier également l'ensemble des enseignants qui ont contribué à ma formation.*

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

---

# LISTE DES ABRÉVIATIONS

---

<b>TAP</b>	: Traitement <b>A</b> utomatique de la <b>P</b> arole
<b>RAP</b>	: <b>R</b> econnaissance <b>A</b> utomatique de la <b>P</b> arole
<b>API</b>	: <b>A</b> lphabet <b>P</b> honétique <b>I</b> nternational
<b>F<sub>0</sub></b>	: <b>F</b> réquence fondamentale
<b>F<sub>1</sub>, ..., F<sub>5</sub></b>	: Formants
<b>TTS</b>	: <b>T</b> ext- <b>T</b> o- <b>S</b> peech (Un <b>S</b> ystème de <b>S</b> ynthèse à <b>P</b> artir du <b>T</b> exte)
<b>OCR</b>	: <b>O</b> ptical <b>C</b> haracter <b>R</b> ecognition (un système de reconnaissance optique des caractères)
<b>LPC</b>	: <b>L</b> inear <b>P</b> redictive <b>C</b> oding (Codage <b>P</b> redictif <b>L</b> inéaire)
<b>TFD</b>	: <b>T</b> ransformée de <b>F</b> ourier <b>D</b> iscrete
<b>TFR</b>	: <b>T</b> ransformée de <b>F</b> ourier <b>R</b> apide ( <b>FFT</b> )
<b>AR</b>	: <b>A</b> uto <b>R</b> égressif
<b>ARMA</b>	: <b>A</b> uto <b>R</b> égressif à <b>M</b> oyenne <b>A</b> justée
<b>MA</b>	: <b>M</b> oyenne <b>A</b> justée
<b>SPR</b>	: <b>S</b> ynthèse <b>P</b> ar <b>R</b> ègles
<b>TFI</b>	: <b>T</b> ransformée de <b>F</b> ourier <b>I</b> nverse
<b>PSOLA</b>	: <b>P</b> itch <b>S</b> ynchronous <b>O</b> ver <b>L</b> ap and <b>A</b> dd
<b>CT</b>	: <b>C</b> ourt <b>T</b> erme
<b>OLA</b>	: <b>O</b> ver <b>L</b> ap and <b>A</b> dd
<b>SI</b>	: <b>S</b> ystème <b>I</b> nternational
<b>TD-PSOLA</b>	: <b>T</b> ime <b>D</b> omain- <b>P</b> itch <b>S</b> ynchronous <b>O</b> ver <b>L</b> ap and <b>A</b> dd

---

# Liste des Figures

page

Fig.1.1	Modèle simplifié de l'appareil phonatoire .....	3
Fig.1.2	les organes de la phonation.....	4
Fig.1.3	Représentation des Formants d'un son voisé .....	6
Fig.1.4	Classification des sons du langage .....	9
Fig.1.5	Relation acoustico-articulatoire des voyelles orales du Français.....	10
Fig.2.1	Modèle général de production de la parole .....	13
Fig.3.1	Fenêtrage du signal de parole.....	20
Fig.3.2	Exemple de signal à Court-Terme.....	21
Fig.3.3	Etape d'addition et recouvrement OLA.....	21
Fig.3.4	Signal synthétisé avec PSOLA.....	21
Fig.3.5	Algorithme de synthèse TD-PSOLA.....	22
Fig.3.6	Analyse du signal de parole.....	22

---

# Tables des Matières

<b>INTRODUCTION GENERALE.....</b>	<b>1</b>
<b>CHAPITRE 1 : NOTIONS GENERALES SUR LA PAROLE</b>	
<b>1.1 INTRODUCTION.....</b>	<b>2</b>
<b>1.2 QU'EST-CE-QUE LE TRAITEMENT AUTOMATIQUE DE LA PAROLE (TAP) ? .....</b>	<b>2</b>
<b>1.3 L'APPAREIL PHONATOIRE .....</b>	<b>2</b>
<b>1.3.1 Les voies aériennes inferieures.....</b>	<b>3</b>
<b>1.3.2 Le larynx .....</b>	<b>3</b>
<b>1.3.3 Le conduit vocal.....</b>	<b>3</b>
<b>1.4 LA PRODUCTION DE PAROLE.....</b>	<b>4</b>
<b>1.5 LES PARAMETRES PROSODIQUES ET ACOUSTIQUES D'UN SIGNAL VOCAL .....</b>	<b>4</b>
<b>1.5.1 La Fréquence Fondamentale.....</b>	<b>5</b>
<b>1.5.2 La durée .....</b>	<b>5</b>
<b>1.5.3 L'Intensité ou l'énergie.....</b>	<b>5</b>
<b>1.5.4 Les Formants .....</b>	<b>6</b>
<b>1.6 LA COMPLEXITE DE SIGNAL VOCAL.....</b>	<b>7</b>
<b>1.6.1 Continuité.....</b>	<b>7</b>
<b>1.6.2 Variabilités.....</b>	<b>7</b>
<i>1.6.2.1 Variabilité intra-locuteur .....</i>	<i>7</i>
<i>1.6.2.2 Variabilité interlocuteur.....</i>	<i>8</i>
<i>1.6.2.3 Variabilité contextuelle .....</i>	<i>8</i>
<b>1.6.3 Coarticulation.....</b>	<b>8</b>
<b>1.6.4 Redondance.....</b>	<b>8</b>
<b>1.7 CLASSIFICATION DES SONS .....</b>	<b>8</b>
<b>1.7.1 Les sons voisés .....</b>	<b>9</b>
<b>1.7.2 Les sons non voisés .....</b>	<b>9</b>
<b>1.7.3 Les voyelles .....</b>	<b>9</b>
<b>1.7.4 Les consonnes .....</b>	<b>10</b>
<b>1.7.5 Les semi-voyelles .....</b>	<b>10</b>



---

<b>1.8 CONCLUSION</b> .....	11
<b>CHAPITRE 2 : TECHNIQUES ET METHODES DE LA SYNTHESE DE LA PAROLE</b>	
<b>2.1 INTRODUCTION</b> .....	12
<b>2.2 DEFINITION DE LA SYNTHESE DE LA PAROLE</b> .....	12
<b>2.3 LE SYSTEME TEXT-TO-SPEECH (TTS)</b> .....	12
<b>2.4 TECHNIQUES D'ANALYSE DU SIGNAL VOCAL</b> .....	12
<b>2.4.1 Méthodes non paramétriques</b> .....	13
<b>2.4.2 Méthodes paramétriques</b> .....	13
2.4.2.1 <i>Codage prédictif Linéaire (LPC)</i> .....	13
2.4.2.2 <i>Analyse cepstrale</i> .....	14
<b>2.5 LES METHODES DE SYNTHESE DE LA PAROLE</b> .....	15
<b>2.5.1 Synthèse Par Règles (SPR)</b> .....	16
<b>2.5.2 Synthèse par concaténation d'unités acoustiques</b> .....	16
2.5.2.1 <i>Mise en œuvre</i> .....	16
2.5.2.2 <i>Synthèse fondée sur l'algorithme PSOLA</i> .....	17
<b>2.6 LES APPLICATIONS DE LA SYNTHESE DE PAROLE</b> .....	17
<b>2.7 CONCLUSION</b> .....	18
<b>CHAPITRE 3 : LA METHODE TD-PSOLA</b>	
<b>3.1 INTRODUCTION</b> .....	19
<b>3.2 PRINCIPE DE FONCTIONNEMENT DE LA TECHNIQUE PSOLA</b> .....	19
<b>3.3 LA TECHNIQUE TD-PSOLA</b> .....	20
<b>3.4 ALGORITHME DE SYNTHESE DE LA TECHNIQUE TD-PSOLA</b> .....	21
<b>3.4.1 Analyse du signal de parole</b> .....	22
<b>3.4.2 Prétraitement</b> .....	23
<b>3.4.3 Détection et détermination de Pitch</b> .....	23
<b>3.5 CONCLUSION</b> .....	23
<b>CONCLUSIONS GENERALES ET PERSPECTICES</b> .....	24
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	25



**Introduction  
générale**

# Introduction Générale

---

La communication par la voix est l'un des enjeux majeurs du dialogue Homme-Machine, puisque la voix véhicule à la fois un contenu linguistique explicite que l'on peut représenter sous forme écrite et un contenu non linguistique comme le type du locuteur, son attitude, ses gestes, etc. Cela rend le Traitement Automatique de la Parole (TAP) une composante fondamentale des sciences de l'ingénieur et un domaine de recherche actif, au croisement du traitement du signal numérique et du traitement symbolique du langage. Depuis les années 60, le TAP bénéficie d'efforts de recherche très importants, liés au développement des moyens et techniques de télécommunications et du traitement numérique de l'information. Ces efforts se sont concrétisés grâce à plusieurs applications du TAP, telles que le codage, la Reconnaissance Automatique et la synthèse de la parole.

Le but de notre travail qui s'inscrit dans le domaine de Traitement Automatique de la Parole, en particulier la synthèse de la parole est d'élaborer un système de synthèse de la parole et d'effectuer des modifications prosodiques de signal vocal en utilisant la méthode TD-PSOLA

L'algorithme PSOLA consiste à concaténer, à l'aide d'un lissage, des unités de parole pré-stockées en modifiant le pitch et la durée des segments. Cette technique est associée à la méthode de synthèse par concaténation.

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

- dans le premier chapitre nous allons décrire d'une manière générale des notions sur le traitement de la parole ainsi que sa production, l'appareil phonatoire humain, des spécifications du signal vocal et finissons par la classification des sons;
- le deuxième chapitre nous allons donner une brève définition de la synthèse de la parole, , En outre, nous étudions les différentes techniques d'analyse du signal vocal .Puis nous expliquons les méthodes de la synthèse de la parole ainsi que ses différentes applications.
- dans le dernier chapitre nous étudions la technique PSOLA qui permet de faire la synthèse d'un signal de parole et plus précisément TD-PSOLA ainsi son principe de fonctionnement

Nous terminons notre travail par des conclusions et perspectives.

# **Chapitre I**

**Notions sur la parole**

## 1.1 INTRODUCTION

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

## 1.2 QU'EST-CE-QUE LE TRAITEMENT AUTOMATIQUE DE LA PAROLE (TAP) ?

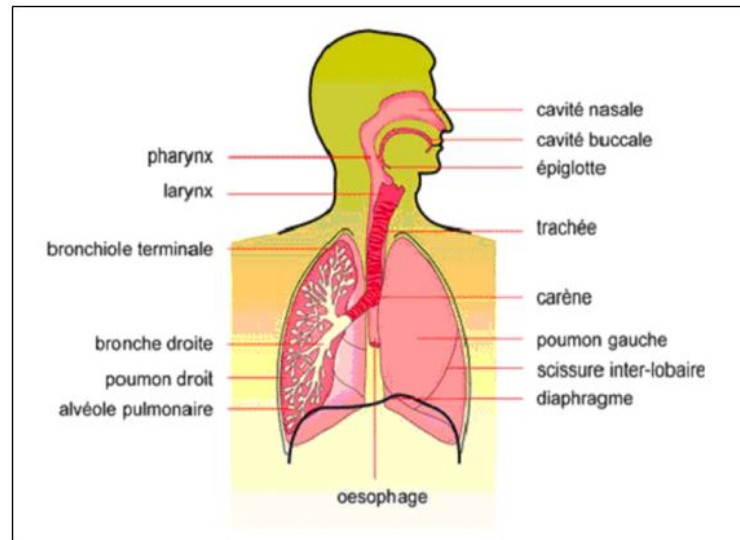
Le **Traitement Automatique de la Parole** est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage.

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

- **les analyseurs** de parole cherchent à mettre en évidence les caractéristiques du signal vocal tel qu'il est produit, ou parfois tel qu'il est perçu (on parle alors d'analyseur perceptuel), mais jamais tel qu'il est compris, ce rôle étant réservé aux reconnaisseurs.
- **les reconnaisseurs** ont pour mission de décoder l'information portée par le signal vocal à partir des données fournies par l'analyse.
- **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 deux types de synthétiseurs : les synthétiseurs de parole à partir d'une représentation numérique, et les synthétiseurs de parole à partir d'une représentation symbolique
- enfin, le rôle des **codeurs** est de permettre la transmission ou le stockage de parole avec un débit réduit, ce qui passe tout naturellement par une prise en compte judicieuse des propriétés de production et de perception de la parole [1].

## 1.3 L'APPAREIL PHONATOIRE

L'appareil phonatoire est l'ensemble des organes qui permettent de produire les sons constituant la voix. Elle est décomposé en trois parties correspondant à trois entités fonctionnelles différentes : les voies aériennes inférieures composées des poumons et de la trachée artère, le larynx, et le conduit vocal (Fig.1.1).



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

### 1.3.1 Les voies aériennes inférieures

Les voies aériennes inférieures correspondent à la partie de l'appareil phonatoire située dans le thorax et sont composées de deux poumons reliés à la trachée qui elle-même remonte jusqu'aux voies aériennes supérieures, les poumons jouent le rôle de réservoir de pression et permettent de générer l'écoulement d'air à l'origine de la production de sons et notamment des vibrations des cordes vocales.

### 1.3.2 Le larynx

Le larynx est l'organe qui fait la jonction entre la trachée et le pharynx. Il se situe dans la gorge et est donc le siège de la production des sons voisés qui implique la vibration des cordes vocales (Fig.1.2).

Lors de la production des sons voisés de la parole (comme les voyelles par exemple), c'est la vibration des cordes vocales qui constitue la source des ondes acoustiques.

### 1.3.3 Le conduit vocal

Le conduit vocal est la partie des voies aériennes supérieures située au-dessus du larynx, il est localisé dans la tête et constitué du pharynx et de deux cavités résonnantes séparées par le palais : la cavité orale (ou buccale) et la cavité nasale. Lorsque les cordes vocales vibrent, les ondes acoustiques générées se propagent dans ces cavités qui agissent comme un résonateur acoustique (Fig.1.2) [3].

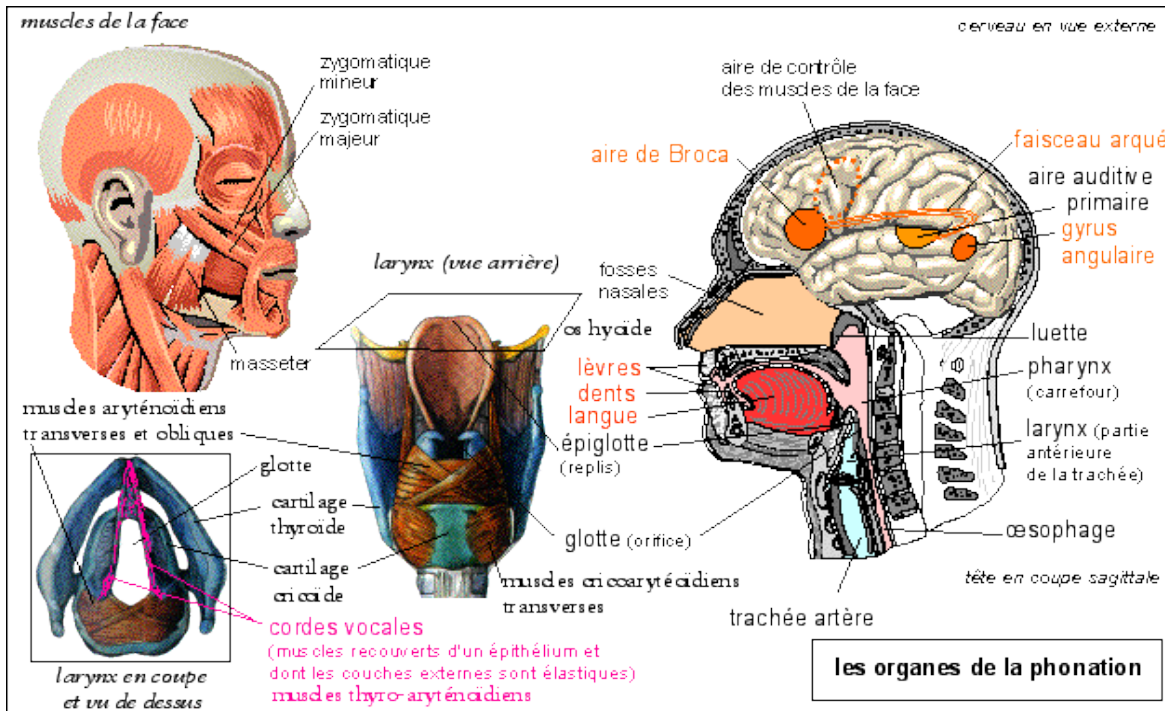


Figure 1.2 : les organes de la phonation [4]

## 1.4 LA PRODUCTION DE PAROLE

La production de la parole est l'opération la plus complexe de l'activité biologique humaine, et du monde vivant connu. Elle met en jeu un très grand nombre de muscles aux mouvements particulièrement précis, caractérisés par de très nombreuses unités motrices, et dont la synchronisation doit être parfaitement contrôlée pour créer l'objet sonore porteur de sens.

La production de la parole est un système dynamique, dont le comportement à un moment donné dépend de ses états antérieurs. Le système est donc dépendant d'une variable paramétrable fonction du temps qui dans ce cas est un geste articulatoire.

La phonologie articulatoire est basée sur la définition des phonèmes en termes de gestes, qui sont les unités d'action et les bases de contraste des items linguistiques, les atomes de la description phonologique. Les phonèmes sont donc définis par des groupes de gestes [5].

## 1.5 LES PARAMETRES PROSODIQUES ET ACOUSTIQUES D'UN SIGNAL VOCAL

La prosodie est une science de la linguistique qui étudie les éléments phoniques (l'accent, l'intonation, etc.) de n'importe quelle langue, et puisque la parole est un signal réel d'énergie

finie, continu, et non stationnaire ; les variations des paramètres prosodiques physiques (La fréquence fondamentale, la durée, et l'intensité) influencent de manière directe sur ces éléments phoniques.

Les caractéristiques prosodiques influencent directement sur l'intelligibilité de la parole synthétique.

### 1.5.1 La Fréquence Fondamentale

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

Elle permet de diviser l'ensemble des sons de parole en trois grandes macros classes [6]:

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

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

### 1.5.2 La durée

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

### 1.5.3 L'Intensité ou l'énergie

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

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



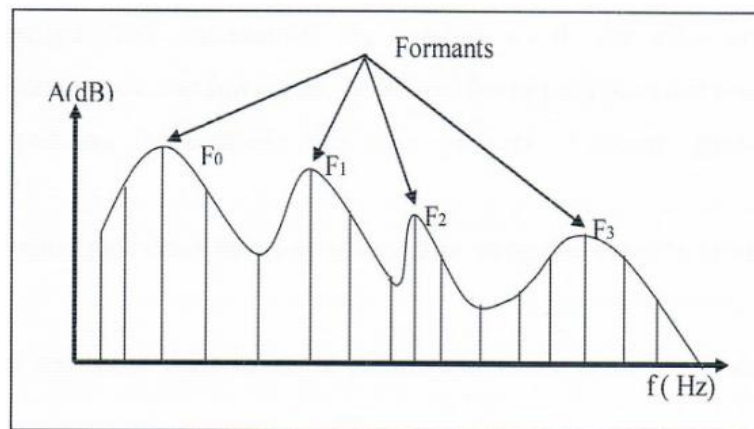
$$E_{db} = 10 * \log_{10} \left( \frac{1}{T} \sum_{n=1}^T x_n^2 \right) \quad (1.2)$$

### 1.5.4 Les Formants

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

L'appareil phonatoire étant constitué de différentes cavités Lors du passage de l'air à travers ces cavité il est amplifié et subit différentes transformations due 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 vocale sont appelés Formants. Chaque son à ses formants caractéristique. Sur un spectrogramme, les formants sont représentés par des bandes noires (le degré de noirceur correspondant à l'énergie) (Fig.1.3).

La fréquence fondamentale est responsable de la hauteur perçue d'un son. Les fréquences d'harmonique renforcées, responsables du timbre d'un son, sont elle aussi numérotées.  $F_1$  correspond à la 1 ère zone d'harmoniques renforcées,  $F_2$  à la 2 ème et ainsi de suite jusqu'à  $F_5$ .



**Figure 1.3** : Représentation des Formants d'un son voisé [7]

Généralement, nous pouvons aller jusqu'à cinq ou six formants pour produire une parole de très haute qualité. Les formants nous permettent de décrire aussi les cibles vocalique correspondant aux zones stables ainsi que les zones de transitions (passage entre deux son consécutifs) ce qui montre leur très grande 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 [7].

## **1.6 LA COMPLEXITE DE SIGNAL VOCAL**

La grande difficulté du TAP et en particulier celui de la **R**econnaissance **A**utomatique de la **P**arole (**RAP**) provient du caractère du processus de la communication parlée et des caractéristiques intrinsèques du signal vocal. La parole est un signal continu d'énergie finie, non stationnaire. Sa structure est complexe et variable dans le temps ; périodique ou plus exactement pseudo périodique pour les sons voisés, aléatoire pour les sons fricatifs et impulsionnel pour les sons occlusifs

### **1.6.1 Continuité**

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

### **1.6.2 Variabilités**

La parole présente une très grande variabilité qui résulte de plusieurs facteurs et ceci que ce soit pour un même ou plusieurs locuteurs (la fatigue, l'état émotionnel, l'âge, le sexe, l'origine géographique et le milieu social). Parmi ces facteurs, les perturbations apportées par le microphone (selon le type, la distance et l'orientation) et l'environnement (bruit et réverbération). De telles variations ne donnent pas naissance à de nouveaux phonèmes, puisqu'elles ne portent aucune information sémantique [8].

#### *1.6.2.1 Variabilité intra-locuteur*

La variabilité intra-locuteur concerne les différences de production du signal parole chez un même locuteur. Plusieurs critères peuvent être responsables de ces différences :

- la fatigue ;
- l'état émotionnel du sujet qui affecte le timbre et le rythme de la voix ;
- les maladies affectant les organes de la voix.

### 1.6.2.2 Variabilité interlocuteur

Des différences acoustiques apparaissent dans un mot prononcé par plusieurs locuteurs. En effet, des contrastes considérables peuvent se manifester suivant l'âge, le sexe, l'origine géographique et le milieu social.

### 1.6.2.3 Variabilité contextuelle

En effet les mouvements articulatoires peuvent être modifiés de façon à minimiser l'effort à produire pour les réaliser à partir d'une position articulatoire donnée, ou pour anticiper une position à venir. Ces effets sont connus sous le nom de *réduction*, *d'assimilation* et de *coarticulation*.

## 1.6.3 Coarticulation

Le signal de parole est constitué d'une succession d'unités différentes. Cependant, contrairement à ce qu'on pourrait croire, ces unités ne sont pas indépendantes les unes des autres mais s'influencent mutuellement : c'est le phénomène de coarticulation [9].

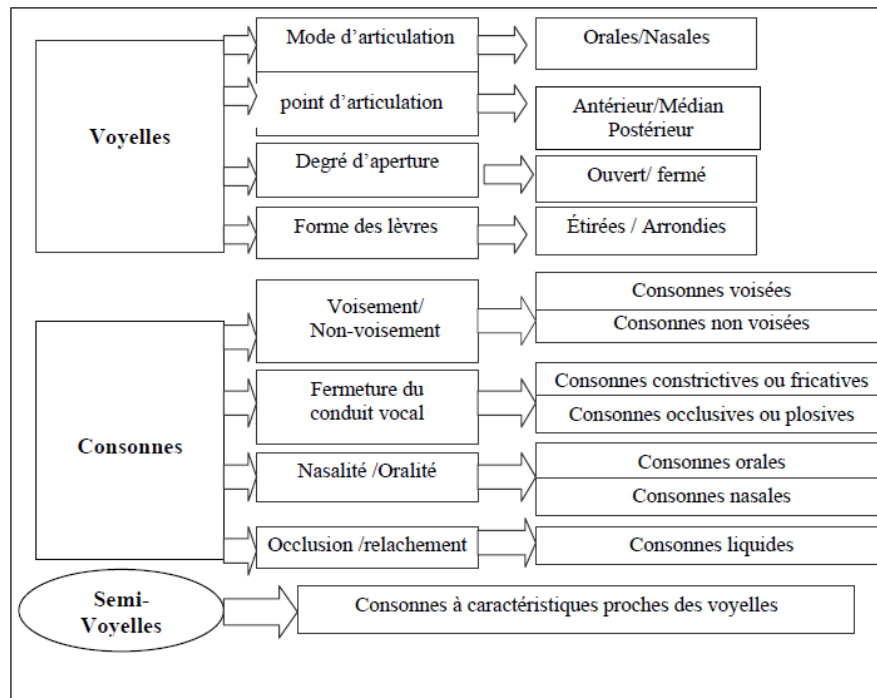
## 1.6.4 Redondance

Le signal de la parole est très redondant. Son traitement automatique nécessite, de réduire au maximum cette redondance afin de diminuer l'encombrement en mémoire et de limiter les durées du traitement, lequel doit se faire en temps réel. A l'inverse, le débit ne doit pas être trop faible pour conserver un bon rapport signal/bruit [8].

## 1.7 CLASSIFICATION DES SONS

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

Il est intéressant de grouper les sons de parole en classes phonétiques, en fonction de leur mode et lieu d'articulation. Le point d'articulation est l'endroit où vient se placer la langue pour obstruer le passage du canal d'air (Fig.1.4).



**Figure 1.4 :** Classification des sons du langage

### 1.7.1 Les sons voisés

Les vibrations des cordes vocales produisent les sons voisés (voyelles, semi-voyelles, consonnes nasales...etc.). Les cartilages sur lesquels s'accrochent les cordes vocales régularisent la tension des cordes, donc la fréquence des vibrations, au moyen des muscles du larynx s'appelle la fréquence fondamentale ou  $F_0$ .

### 1.7.2 Les sons non voisés

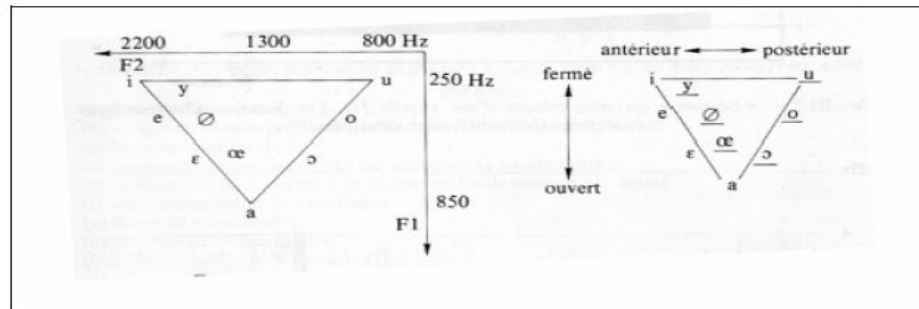
Le second mode d'excitation est obtenu par divers bruits produits par le passage de l'air en un point de resserrement du canal vocal ou par des bruits d'occlusion ou de plosion provoqués par la fermeture ou l'ouverture des lèvres, ou des chocs de la langue contre le palais. Dans cette catégorie de sons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

### 1.7.3 Les voyelles

Les voyelles diffèrent de tous les autres sons par le degré d'ouverture du conduit vocal. Quand ce dernier est suffisamment ouvert pour que l'air expiré par les poumons, le traverse sans obstacle, il y a production d'une voyelle. Le rôle de la cavité buccale se réduit alors à une modification du timbre vocalique. Si, au contraire, le passage se rétrécit par endroit, ou même s'il se ferme temporairement, le passage forcé de l'air donne naissance à un bruit : une consonne est

produite. Une voyelle se caractérise par un passage libre de l'air dans le conduit vocal et par les vibrations des cordes vocales.

Elles se différencient principalement les unes des autres par leur lieu d'articulation (position de la langue), leur degré d'ouverture (espace compris entre la pointe de la langue et le palais), et leur nasalisation. Nous distinguons ainsi, selon la localisation de la masse de la langue, les voyelles antérieures ou avant, les moyennes, et les voyelles postérieures (ou arrières), et, selon l'écartement entre l'organe et le lieu d'articulation, les voyelles fermées et ouvertes (Fig1.5).



**Figure 1.5** : Relation acoustico-articulatoire des voyelles orales du Français [10]

#### 1.7.4 Les consonnes

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

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

#### 1.7.5 Les semi-voyelles

Les semi-voyelles, quant à elles, combinent certaines caractéristiques des voyelles et des consonnes. Comme les voyelles, leur position centrale est assez ouverte, mais le relâchement soudain de cette position produit une friction qui est typique des consonnes. Enfin, elles sont assez difficiles à classer [8].

## 1.8 CONCLUSION

Dans ce chapitre nous avons exposé des notions de base sur le traitement de la parole, des spécifications du signal vocal.

Les objectifs de ce chapitre sont de définir les notions que nous utiliserons dans notre travail.

Cette partie théorique sera complétée dans le chapitre suivant par une étude sur systèmes de synthèse de la parole et ses variantes.

# **Chapitre II**

**Techniques et méthodes de la synthèse de la parole**

## 2.1 INTRODUCTION

Ce chapitre nous permet de présenter les principales techniques de la synthèse de la parole, en premier lieu, nous allons donner une brève définition de la synthèse de la parole, et le, En outre, nous étudions les différentes techniques d'analyse du signal vocal .Puis nous expliquons les méthodes de la synthèse de la parole ainsi que ses différentes applications.

## 2.2 DEFINITION DE LA SYNTHÈSE DE LA PAROLE

La synthèse de parole présente plusieurs avantages, elle est d'une part plus naturelle pour le grand public, elle est plus rapide et efficace qu'un message écrit court et le champ de vision reste libre pour effectuer une autre tâche de lecture.

Les deux principaux critères exigés par la synthèse de la voix sont l'intelligibilité et l'aspect naturel. Si de nos jours, le premier critère est atteint, le deuxième est encore au stade de développement. En effet, si les synthétiseurs reproduisent une voix tout à fait intelligible, les intonations et l'expressivité ne sont pas encore au point [9].

## 2.3 LE SYSTEME TEXT-TO-SPEECH (TTS)

Un Système de Synthèse à Partir du Texte (**TTS** : **Text-To-Speech**) est une machine capable de lire a priori n'importe quel texte à voix haute, que ce texte ait été directement introduit par un opérateur sur un clavier alphanumérique, qu'il ait été scanné et reconnu par un système de reconnaissance optique des caractères (**OCR** : **Optical Character Recognition**), ou qu'il ait été produit automatiquement par un système de Dialogue Homme-Machine. On définira donc plutôt la synthèse TTS comme la production automatique de phrases par calcul de leur transcription phonétique [1].

## 2.4 TECHNIQUES D'ANALYSE DU SIGNAL VOCAL

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



### 2.4.1 Méthodes non paramétriques

Le signal de parole peut être analysé dans le domaine temporel ou dans le domaine spectral par des méthodes non paramétriques, sans faire l'hypothèse d'un modèle pour rendre compte du signal observé. Les méthodes spectrales sont fondées sur la décomposition fréquentielle du signal sans connaissance a priori de sa structure fine. Une analyse spectrale du signal permet de mettre en évidence certaines caractéristiques de la production de la parole qui peuvent contribuer à l'identification phonétique. L'articulation des phonèmes a une influence directe sur la forme du conduit vocal et des cavités, et donc sur les résonances qui apparaissent dans l'enveloppe du spectre [8].

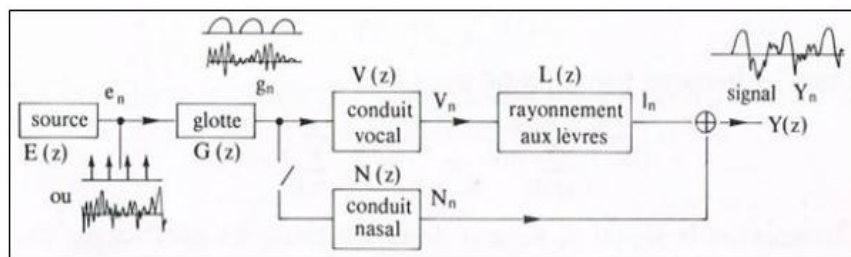
### 2.4.2 Méthodes paramétriques

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.

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

#### 2.4.2.1 Codage prédictif Linéaire (LPC)

Cette méthode connue de la production sous le sigle LPC (Linear Prédictive 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 (Fig.2.1).



**Figure 2.1** : Modèle général de production de la parole [10]

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 100Hz (glotte), l'entrée  $e_n$  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 vocale est un filtre tout pôle (AR : autorégressif) d'ordre  $2M$  décomposable en une cascade de résonateur à 2 pôle en série (tuyaux résonants). Le model du conduit nasal est un filtre pole zéro ARMA (auto régressif a moyen ajusté) et le rayonnement aux lèvres peut se modélisé par un filtre tout zéro (MA moyenne ajusté).

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

Modèle glottale :

$$G(z) = \frac{1}{(1 - e^{-2\pi f_g T} z^{-1})^2} \quad f_g = 100 \text{ Hz} \quad (2.3)$$

Modèle du conduit vocale :

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

$F_i$  : Fréquence du formant Ne (i),  $B_i$  sa band passante.

Modèle du conduit nasal :

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

Avec  $F_N$  et  $F'_N$  formant nasal et anti formant nasal et respectivement,  $B_N$  et  $B'_N$  leur bande passante [10].

#### 2.4.2.2 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.

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.

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

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

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

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

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

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 formantiques) est obtenue par une transformation supplémentaire

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) [8].

## 2.5 LES METHODES DE SYNTHESE DE LA PAROLE

On sait depuis longtemps que les transitions phonétiques contribuent plus à l'intelligibilité du signal vocal que les zones stables des phonèmes. On peut alors envisager de le faire de façons :

- explicite, sous la forme d'une série de règles décrivant formellement l'influence des phones les uns sur les autres ;
- implicite, en enregistrant des exemples de transitions entre phones dans une base de données de segments de parole, et en les utilisant tels quels comme unités de parole (en lieu et place des phones).

Cette alternative a donné lieu à deux grandes familles de synthétiseurs: la synthèse par règles et la synthèse par concaténation

### 2.5.1 Synthèse Par Règles (SPR)

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 [11]

### 2.5.2 Synthèse par concaténation d'unités acoustiques

Cette technique, qui repose sur l'utilisation de segments de signaux extraits de la parole naturelle, est la seule qui permet à ce jour de synthétiser des voix dont le timbre s'approche de celui d'un locuteur humain.

#### 2.5.2.1 Mise en œuvre

La synthèse proprement dite comprend trois étapes distinctes :

- **Sélection des unités acoustiques** : cette première étape consiste à choisir dans le répertoire d'unités acoustiques les unités qui seront effectivement utilisées pour synthétiser la succession de sons désirée. Cette étape est à peu près évidente quand les unités sont régulières (à l'instar des phonèmes et des diphtongues) : seule la présence de plusieurs versions pour le même segment est à prendre en considération. Cette étape est en revanche plus délicate pour les systèmes d'unités de taille variable. Pour une suite de sons donnée, plusieurs choix d'unités sont en général possibles. Il faut alors arbitrer entre les différentes décompositions avec des critères composites.

- **Ajustement des paramètres prosodiques** : les unités acoustiques pré-enregistrées possèdent une prosodie intrinsèque (les sons qui la composent ont une certaine durée et la fréquence fondamentale décrit un certain contour). Bien sûr, cette prosodie intrinsèque n'a que très peu de chances d'être conforme à la prosodie de synthèse, spécifiée par le module prosodique. Il va donc falloir utiliser une technique de traitement de signal pour ajuster aux valeurs cibles définies les paramètres prosodiques des unités de synthèse.

**-Concaténation des unités** : les unités acoustiques, quelles que soient les précautions prises lors de la sélection et de l'enregistrement des unités, ne possèdent pas exactement à leur frontière les mêmes caractéristiques acoustiques (en particulier énergétiques).

En l'absence de traitement, ces discontinuités vont engendrer des artefacts perceptibles et gênants. Il est donc important de lisser ces discontinuités en interpolant les trajectoires des différents paramètres caractéristiques de l'unité [11].

#### 2.5.2.2 Synthèse fondée sur l'algorithme PSOLA

L'algorithme **PSOLA (Pitch Synchronous Over Lap and Add)** consiste à concaténer, à l'aide d'un lissage, des unités de parole pré-stockées en modifiant le pitch et la durée des segments. Cette technique est associée à la méthode de synthèse par concaténation. L'algorithme PSOLA permet la synthèse d'une parole de haute qualité [7]

## 2.6 LES APPLICATIONS DE LA SYNTHÈSE DE PAROLE

Les applications actuelles de synthèse de la parole à partir du texte peuvent être regroupées en cinq grands domaines :

- aides pour personnes handicapées
  - lecture d'écrans ou de documents écrits pour non-voyants;
  - aides à la communication vocale pour personnes muets, laryngectomisés ou à infirmité motrice cérébrale;
  - journaux vocaux, etc.
- Outils d'Enseignement Assisté par Ordinateur (OEAO)
  - système de dictées automatiques;
  - système d'apprentissage des langues.
- applications industrielles
  - serveurs d'alerte, de surveillance de sites et de supervision de réseaux;
  - télémaintenance;
  - fonctions d'aide dans les postes de pilotage;
  - fonction de vérification vocale dans les postes d'édition (correction des épreuves) ou de saisie d'informations écrites (bases de données), etc.

- applications grand public non téléphoniques
  - domotique (alarmes, appareils domestiques parlants, etc.);
  - micro-informatique (jeux parlants, bureautique, etc.).
- télématique vocale
  - serveurs vocaux d'informations (la synthèse remplaçant la parole naturelle enregistrée pour des informations rapidement évolutives et disponibles sous forme textuelle);
  - serveurs de lecture vocale de FAX ou de messages électroniques (e-mails);
  - automatisation de services de renseignements (Annuaire, standards d'entreprises, etc.)[6].

## 2.7 CONCLUSION

Dans ce chapitre, nous avons abordé les principales méthodes et techniques de la synthèse de la parole.

Les deux principaux critères exigés par la synthèse de la voix sont l'intelligibilité et l'aspect naturel, d'où elle vise à améliorer le quotidien, mais n'oublions pas que si elle atteint le niveau de conversation d'un humain, elle engendrerait aussi sa substitution dans certains domaines augmentant ainsi l'emprise de la machine sur l'homme.

# **Chapitre III**

**La méthode TD-PSOLA**

### 3.1 INTRODUCTION

Ce troisième chapitre représente une étude de la technique qui permet de faire la synthèse d'un signal de parole, Soit la technique PSOLA. Nous nous intéressons principalement au principe de fonctionnement de cette technique, à l'algorithme de synthèse de TD-PSOLA et les méthodes de détection de pitch.

### 3.2 PRINCIPE DE FONCTIONNEMENT DE LA TECHNIQUE PSOLA

Depuis 20 ans, de nombreuses méthodes de modification du signal, reposant sur le principe de superposition/addition temporelle ont été proposées. Parmi les plus importantes, citons les méthodes TDHS (Time Domain Harmonic Scaling), SOLA (Synchronized Overlap-Add), WSOLA (Waveform Similarity Overlap-Add) [12].

La méthode PSOLA est une des variantes d'OLA, dans ces techniques, le fenêtrage ne se fait pas, à pas constant mais de manière synchrone de la fréquence fondamentale, ce qui exige un marquage précis de la fréquence fondamentale. Le taux de recouvrement est d'une période locale ( $\approx 50\%$ ) et chaque sommet d'une fenêtre (fenêtre de Hamming) coïncide avec un pic glottique dont la taille est le double de la période locale. Les pics sont alors déplacés suivant l'axe des temps de façon à épouser la forme du nouveau contour  $F(\tau)$  (contour lissé des fréquences) et leurs positions sont calculées par la formule (3.1)

$$t_{j+1} = t_j + \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} 1/F(\tau) \quad (3.1)$$

L'indice  $i$  renvoie aux instants avant modification alors que l'indice  $j$  renvoie aux instants après modification du contour. Dans les zones non voisées, les instants  $t_j$  sont régulièrement espacés d'une durée de l'ordre de 10ms.

La méthode PSOLA se distingue de ces méthodes par une synchronie à la période fondamentale tant à l'analyse qu'à la synthèse. Ceci permet un contrôle à la fois du déroulement de l'axe temporel et de la hauteur du signal.

Les différentes versions de PSOLA existantes fonctionnent selon le même principe. Le segment de signal de parole naturelle est subdivisé en un ensemble de signaux dits à Court- Terme (CT) en utilisant un fenêtrage synchronisé avec le pitch (trame voisée) et à intervalles fixes (trame non voisée).



Le pitch est augmenté ou diminué en agissant sur la distance entre les signaux à CT durant le processus de synthèse. La durée est gérée par suppression ou duplication des signaux à CT [12].

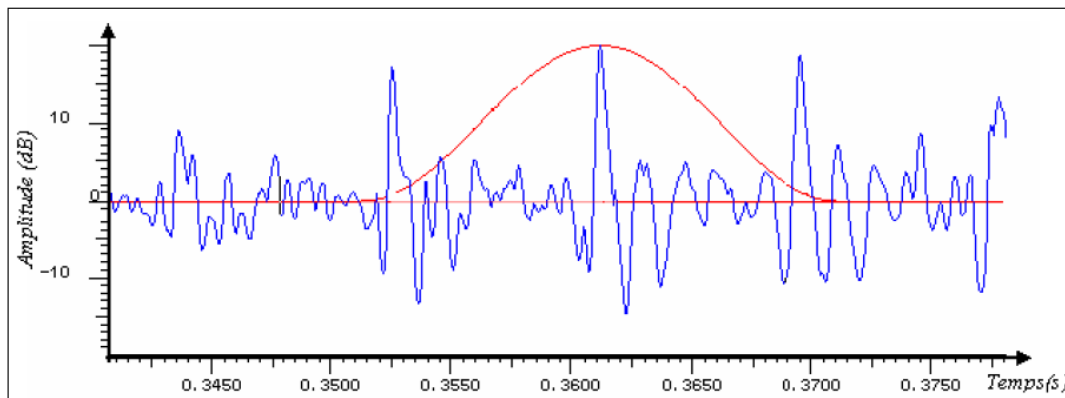
### 3.3 LA TECHNIQUE TD-PSOLA

La technique dite d'addition recouvrement des fenêtres temporelles synchrones avec le pitch **TD-PSOLA**; applique le principe de la réharmonisation spectrale directement sur le signal de parole. Appliquée à la synthèse par concaténation, elle conduit à une base de données « paramétriques » où les seuls paramètres stockés sont les marqueurs du pitch indiquant le milieu des fenêtres d'OLA dans les signaux de base de données des segments.

Ces marqueurs sont positionnés en synchronisme avec le pitch à l'aide d'un algorithme d'extraction de pitch, et régulièrement espacés sur les zones Non Voisées où aucune modification de pitch ne devra de toute façon être effectuée.

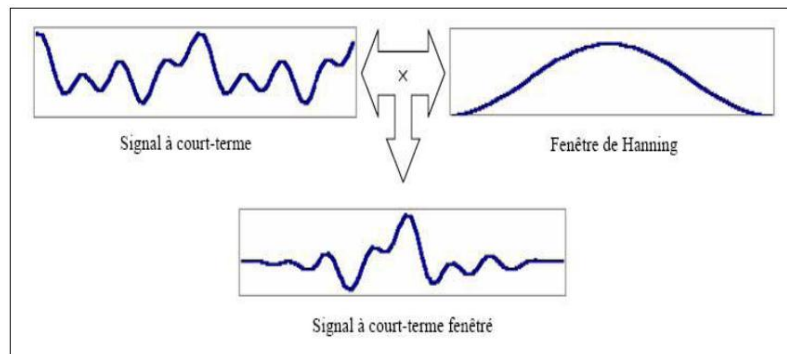
La fenêtre utilisée doit garantir l'atténuation des lobes secondaires, car elles seront candidates à une sommation ultérieure, et elles portent des informations sur l'identité des fenêtres voisines du signal.

On choisit souvent une fenêtre de Hamming ou une fenêtre Triangulaire, avec une longueur égale à deux fois la période du pitch du signal (Fig.3.1).



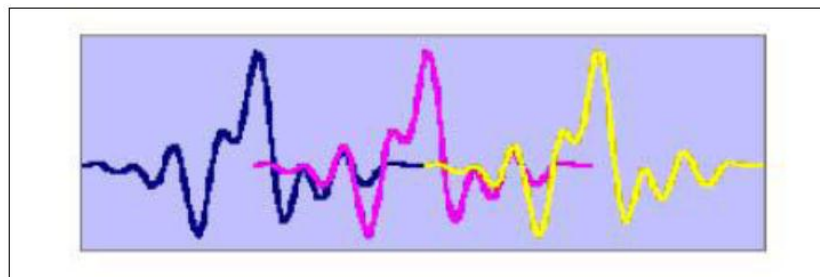
**Figure 3.1** : Fenêtrage du signal de parole

Une fenêtre plus large fait apparaître des harmoniques dans le spectre de signal synthétisé, une fenêtre plus courte n'approxime que très grossièrement l'enveloppe spectrale de signal original.



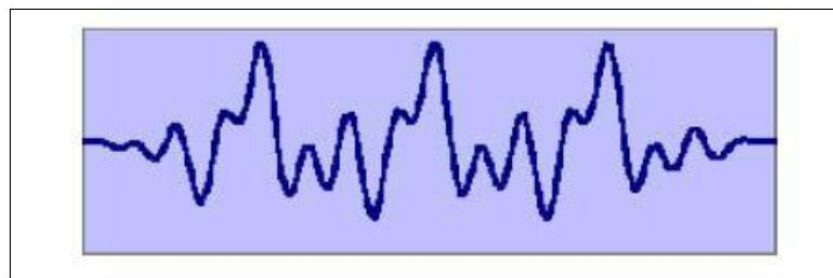
**Figure 3.2** : Exemple de signal à Court-Terme

Les signaux à CT sont recombinaison pour produire le signal de synthèse à l'aide d'une technique d'addition/recouvrement OLA (Fig.3.3).



**Figure 3.3** : Etape d'addition et recouvrement OLA [12]

Après la recombinaison des signaux à court terme par addition recouvrement OLA nous trouvons le signal synthétique (Fig.3.4).



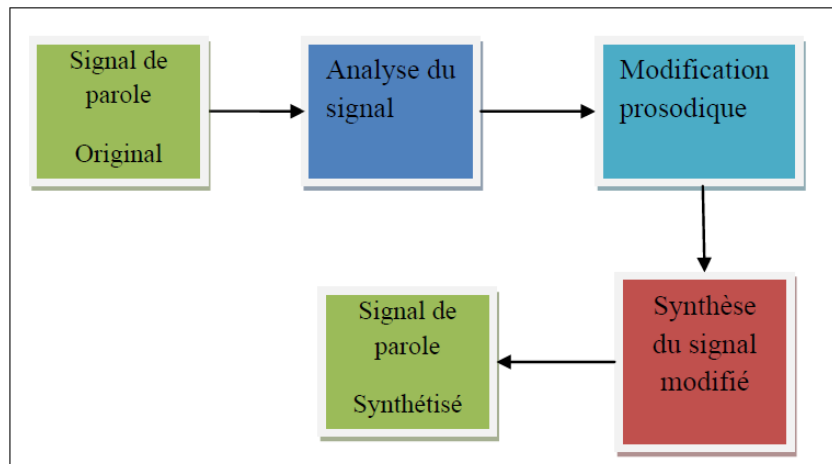
**Figure 3.4** : Signal synthétisé avec PSOLA [12]

### 3.4 ALGORITHME DE SYNTHÈSE DE LA TECHNIQUE TD-PSOLA

Après avoir présenté le principe de la technique de synthèse TD-PSOLA, nous pouvons décrire l'algorithme correspondant. Celui-ci requiert trois étapes principales:

- analyse du signal d'origine;

- modification prosodique apportée à ces signaux à CT;
- synthèse du signal modifié par recouvrement addition des signaux à CT.

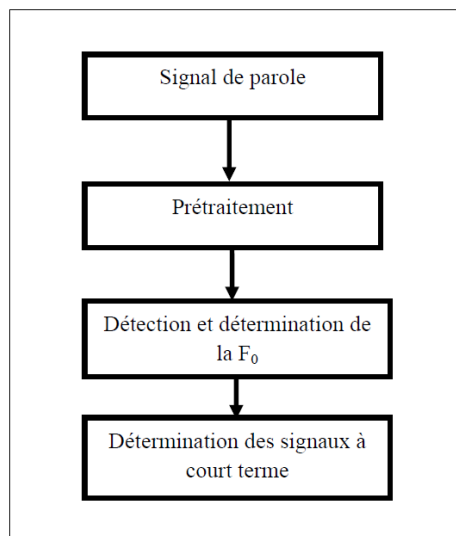


**Figure 3.5 :** Algorithme de synthèse TD-PSOLA

### 3.4.1 Analyse du signal de parole

Les opérations d'analyse à effectuer pour le marquage du signal sont les suivantes :

- prétraitement: séparation des composantes périodiques du signal (dites Voisées dans le cas de la voix) et des composantes aléatoires (dites Non Voisées);
- détection et détermination de la  $F_0$  avec une méthode appropriée;
- détermination des signaux à CT d'analyse.



**Figure 3.6 :** Analyse du signal de parole

### 3.4.2 Prétraitement

Cette phase est en général réservée à la préparation du signal issue d'un microphone. Elle consiste à choisir la durée de la trame d'analyse et du recouvrement afin de moins de compromettre :

- la condition de stationnarité souvent exigée par les algorithmes de traitement ;
- l'effet de bord lié aux fenêtres de pondération appliquées, et d'assurer ainsi la présence d'au moins une période du fondamental. La durée de la trame est généralement choisie entre 20 et 30ms avec un recouvrement de 30 à 50% ;
- la détection de silence garantissant la présence du signal utile ;
- la détection du voisement puisque la F0 n'est présent que sur une séquence Voisée ;
- une préaccentuation afin de rehausser l'énergie des hautes fréquences.

Dans cette phase de prétraitement, nous trouvons souvent d'autres techniques permettant d'améliorer la rapidité d'extraction. Il s'agit de toutes sortes de techniques de filtrage permettant d'atténuer ou même d'éliminer les formants d'ordres supérieur ou égal à 2, et de minimiser l'effet du bruit sur la détection. La décimation dans le rapport de 5 à 1 est souvent utilisée par des détecteurs à temps réels. Cependant la décimation ne peut être utilisée dans les systèmes où une grande précision est requise (telle que la reconnaissance, l'identification, la modification de la F0 etc.).

### 3.4.3 Détection et détermination de Pitch

La phase de traitement est réservée à l'extraction de la fréquence fondamentale et dépend donc de l'algorithme utilisé.

## 3.5 CONCLUSION

La technique PSOLA est une approche pour la manipulation de la parole, elle représente une étape importante dans le développement des techniques du traitement de la parole. Le développement des techniques de synthèse de la parole reflète une attention croissante à la nature physique de production de la parole.



**Conclusions générales  
et perspectives**

## CONCLUSIONS GENERALES ET PERSPECTIVES

L'objectif de notre travail était l'élaboration d'un système de synthèse de la parole et d'effectuer des modifications prosodiques de signal vocal en utilisant la technique TD-PSOLA.

Ce qui nous a permis d'avoir des notions sur le traitement de la parole. Il nous a permis également de comprendre le fonctionnement d'un système de synthèse de la parole. Il est logique de chercher à modifier la durée et la fréquence fondamentale du signal sans faire appel à un modèle de représentation. Parmi les techniques de modifications prosodiques étudiées jusqu'à présent, la plus efficace est l'algorithme PSOLA.

Nous nous sommes concentrés sur une variante dénommée la technique TD-PSOLA. C'est une technique de traitement du signal de parole dont l'objectif est de modifier les paramètres prosodiques de celui-ci. Si maintenant on la considère comme une technique de synthèse de parole; elle va avoir des résultats meilleurs car c'est une reproduction excellente de signal original.

La méthode PSOLA se distingue des autres méthodes par une synchronie à la période fondamentale tant à l'analyse qu'à la synthèse. Ceci permet un contrôle à la fois du déroulement de l'axe temporel et de la hauteur du signal

Comme suite à ce travail, il sera très intéressant de faire une étude comparative de toutes les variantes de la technique PSOLA telle que la FD-PSOLA et la LP-PSOLA, etc. et de tester leurs performances afin de montrer la bonne qualité de synthèse et de modifications prosodiques.

**Références  
Bibliographiques**

## REFERENCES BIBLIOGRAPHIQUES

- [1] T. Dutoit, Introduction au traitement automatique de la parole notes de cours /DEC2, Faculté Polytechnique de Mons, LCTS Lab, France, 2000.
- [2] [http : //www.infovisual.info](http://www.infovisual.info).
- [3] J. Cisonni, Modélisation et inversion d'un système complexe de production de signaux acoustiques Application à la voix et aux pathologies, Thèse de Doctorat, Institut Polytechnique De Grenoble, France, Novembre 2008.
- [4] <http://users.skynet.be/illusionsauditives/images/page802.gif>
- [5] L'étude instrumentale des gestes dans la production de la parole ; importance de l'aérophonométrie ; Manuscrit auteur, publié dans "Les Dysarthries, P. Auzou (Ed.) (2007) 115-117
- [6] O. Amine, Synthèse de la parole en arabe standard, Mémoire de Magister, ENP, Alger, Algérie, 2011.
- [7] G. Djeghiour, thèse magister : application des réseaux de neurones a la synthèse de la parole en arabe standard, Ecole nationale supérieure des sciences humaines; 2011.
- [8] A. Mohamed, Application des Algorithmes Génétiques au Décodage Acoustico-Phonétique de la parole en Arabe Standard, Thèse de Doctorat, ENP, Alger 2008
- [9] V A.Dubesset, La Langue française Parlée Complétée (LPC) : Production et Perception, Thèse de Doctorat, Institut National Polytechnique De Grenoble, France, 2005
- [10] Calliope, La parole et son traitement automatique, Collection Techniques et Scientifiques des Télécommunications. Préface de G. Fant, CNET/ENST, Ed. Masson, 1989.
- [11] E. Moulines, O. Cappé, synthèse de la parole à partir du texte Techniques de l'ingénieur, Vol. H1 960, p.6-7.
- [12] A. Chentir ; Etude de la Microprosodie en vue de la Synthèse de la parole en Arabe Standard ; thèse doctorat : Ecole nationale polytechnique d'Alger ; 01 Octobre 2009, Algérie