

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA  
RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU  
FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE  
Département d'Electronique



# Mémoire de fin d'études

Présenté en vue de l'obtention du diplôme  
D'Ingénieur d'Etat en Electronique  
Option : Communication

## Thème

*Reconnaissance Automatique  
du Locuteur par la méthode du  
Taux de Passage par Zéro*

Proposé et dirigé par :  
*Mr. S. HADDAB*

Etudié par :  
*Melle .CHABANE Mounia*  
*Melle .BENSAFIA Kahina*

Promotion 2007/ 2008

# *Remerciements*

*Nous tenons à remercier :*

- *Le bon Dieu qui nous a donné la bonne santé et la persévérance durant notre cursus.*
- *Nos parents et tous ceux qui nous ont aidé pour mener à terme ce travail.*
- *Notre promoteur Mr Haddab.S.*
- *Les membres de jury.*
- *Toutes les personnes qui ont participé aux enregistrements.*
- *Et tous les lecteurs.*

# DEDICACE

*Il est bien trop modeste comme travail  
mais c'est le fruit de tant d'effort  
tout au long de ce semestre,  
je le dédie :*

*Tout d'abord à ma très chère mère YAMINA qui m'a beaucoup  
aidé financièrement et encouragée pour que je puisse  
mener mes études "Que Dieu la Protège".*

*À mes frères: Sadak et sa fiancée Nouara, Abdessallem, Kader,  
Farid et Meziane.*

*À mes sœurs adorées : Sabiha, Zina et son mari Ali , Akila et  
Rezika.*

*À mes trois anges : Le beau goss Abdelhak, la ménionne Mérième  
et le petit naïf Nassim.*

*À mes amis : Mouloud et Karim.*

*À mes amies : Ghenima et sa sœur Hayete, ynda, Chahira, Souad,  
Houra, Yasmina Ouiza, Nedjma, Malika , Lamia, Kahina<sup>2</sup>,  
Fadhila, Lamia, Farida, , ... ainsi leurs familles.*

*À mes cousines et cousins.*

*À tata Ourida, tata Malika et ma tante adorée Zohra.*

*Je les remercie tous pour leurs soutiens.*

*À mon binôme et fidèle amie Monika et toute sa famille.*

*À la mémoire de mon père MEZIANE :*

*"Que Dieu Tout Puissant t'accueille en son Vaste Paradis".*

*À toute la promo 2007/2008.*

*Et à tous ceux et celles que j'ai oublié.*

**Kahina BENSALIA.**

# Dédicaces

*Il est bien trop modeste comme travail mais c'est le fruit de tant d'effort tout au long de ce semestre, je le dédie :*

*Tout d'abord à mon très chér père RACHID et chère mère YAMINA qui ont tang sacrifié pour moi ,et qui m'ont beaucoup aidé financièrement et encouragée pour que je puisse mener mes études.*

*Merci et que DIEU vous garde.*

*A mes frères : Tahar et son épouse Souad ,Boualem, ,Nadir et Hakim  
A mes sœurs adorées : Safia et son marie Azzedine ,Siham, Djahida qui m'a vraiment suivie et la petite Lydia que j'admire.*

*A mes deux anges Amoune et Nanita.*

*Sans oublier ma grande mère DJOUZA que Dieu la Protège et NANA FATIMA.*

*A mes amis : Mouloud et Karim .*

*A mes amies : Chahira, Dalila K, Dalila B, Nedjma, Malika , Lamia B..  
je les remercie tous pour leurs soutiens.*

*A Mr Melouk Ramdane qui depuis que j'ai été son élève m'a dirigé vers la bonne voie et m'orienté vers les études.*

*A mon binôme et fidèle amie Kahina et sa famille.*

*Bien sur sans t'oublier à Kamel et merci pour tes encouragements.*

*A toute la prômo 2007/2008*

*Et à toutes les personnes que j'ai oubliées.*

*Mounia*

**CHABANE**

# SOMMAIRE

<i>Introduction générale</i> .....	1
------------------------------------	---

## **Chapitre I :**      *Systeme de production de la parole*

<b>Introduction</b> .....	2
<b>I-1-Description et fonctionnement de l'appareil vocal</b> .....	2
I-1-1-Description de l'appareil vocal.....	2
I-1-2-fonctionnement de l'appareil vocal.....	5
<b>I-2- Taxonomie des sons (théorie de classification)</b> .....	10
I-2-1-la phonologie(phonétique fonctionnelle).....	10
I-2-2-la phonétique(phonétique articulatoire).....	10
<b>I-3-Le signal vocal</b> .....	11
I-3-1-Description et caractéristiques.....	11
I-3-2-Variabilité du signal vocal.....	12
Variabilité intralocuteur.....	12
Variabilité interlocuteur.....	12
Variabilité due à l'environnement.....	13
<b>I-4-la fréquence fondamentale</b> .....	13
<b>I-5-Acquisition et mise en forme du signal de parole</b> .....	15
<b>Conclusion</b> .....	18

## **Chapitre II :**      *L'analyse du signal de parole*

<b>Introduction</b> .....	19
<b>II-1-Les différentes méthodes d'analyse de la parole</b> .....	19
<b>II-1-1-La méthode directe</b> .....	19
II-1-1-1- La glottographie photoélectrique.....	19
II-1-1-2- La glottographie électrique.....	20
II-1-1-3- La glottographie ultrasonore.....	20
<b>II-1-2-La méthode indirecte</b> .....	20
<b>II-1-2-1-L'analyse temporelle</b> .....	20
II-1-2-1-1- Le taux de passage par zéro.....	21
II-1-2-1-2- L'énergie du signal.....	22
II-1-2-1-3- La fonction d'autocorrelation.....	22
<b>II-1-2-2-L'analyse fréquentielle</b> .....	25
II-1-2-2-1-L'analyse spectrale par le spectrogramme.....	25
II-1-2-2-2- L'analyse spectrale par le cepstre.....	29
II-1-2-2-3- L'analyse spectrale par la prédiction linéaire.....	31
II-1-2-2-4- L'analyse spectrale par la TFCT.....	33
II-1-2-2-5- L'analyse par les ondelettes.....	35
<b>Conclusion</b> .....	39

## **Chapitre III :**      *La Reconnaissance Automatique du Locuteur (RAL)*

<b>Introduction</b> .....	40
<b>III-1- Les différentes tâches de la reconnaissance du locuteur</b> .....	40
III-1-1- Les tâches relevant de l'identification.....	40
III-1-2- Les tâches relevant de la vérification.....	41

<b>III-2- Le mode d'élocution</b> .....	41
III-2-1-Mode indépendant du texte .....	41
III-2-2-Mode dépendant du texte.....	41
<b>III-3- Les différents systèmes de la reconnaissance du locuteur</b> .....	42
III-3-1- Système dépendant du locuteur(monolocuteur).....	42
III-3-2- Multilocuteur.....	42
III-3-3- Système indépendant du locuteur.....	42
<b>III-4-La structure du système de reconnaissance du locuteur</b> .....	42
III-4-1- Paramétrisation.....	43
III-4-2- Apprentissage et modélisation.....	43
III-4-3- Mesure de similarité.....	44
III-4-4- Décision.....	45
<b>III-5- Les limites et les difficultés de la RAL</b> .....	45
<b>III-6- Evaluation des performances de la RAL</b> .....	45
<b>III-7- Les domaines d'application de la RAL</b> .....	46
<b>III-8- Principe de la méthode du taux de passage par zéro pour la RAL</b> .....	47
<b>Conclusion</b> .....	48

## Chapitre IV :     *L'application du TPZ à la RAL*

<b>Introduction</b> .....	49
<b>IV-1-Les ressources</b> .....	49
IV-1-1-Matériel utilisé.....	49
IV-1-2-Langage de programmation .....	49
<b>IV-2- Conditions d'enregistrement</b> .....	50
<b>IV-3- Les étapes d'enregistrement</b> .....	50
<b>IV-4- Fonctionnement du système de reconnaissance et application</b> .....	51
IV-4-1- La phase d'apprentissage.....	51
IV-4-2- La phase de reconnaissance.....	54
<b>IV-5-Interprétation des résultats</b> .....	57
<b>Conclusion</b> .....	62
<b>Conclusion générale</b> .....	63

**Annexe**

**Abréviation et Glossaire**

**Bibliographie**

# *Introduction générale*

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

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.

S'il n'est pas en principe de parole sans cerveau humain pour la produire, l'entendre, et la comprendre, les techniques modernes de traitement de la parole tendent cependant à produire des systèmes automatiques qui remplacent plusieurs fonctions telles que le codage, l'analyse, la synthèse et la reconnaissance.

Notre étude est consacrée à la reconnaissance automatique du locuteur (RAL) par la méthode de taux de passage par zéro (TPZ) dans un système où le nombre du locuteur est limité en mode dépendant du texte. Et pour cela on a réparti notre travail comme suit :

Le chapitre I nous a introduit dans l'univers du son vocalique, sa constitution, sa fabrication, et ses caractéristiques temporelles et fréquentielles ainsi son acquisition.

Après avoir eu une idée sur les éléments constitutionnels de la parole, le chapitre II nous invite à une analyse du signal vocal, et à ce propos on a cité quelques systèmes et méthodes d'analyse.

Après cette analyse du signal de parole, on passe pour le traiter en reconnaissance, où on a essayé de donner un aperçu sur les différentes techniques, les difficultés, et les divers systèmes et leurs critères mis en évidence pour la reconnaissance automatique du locuteur, cela sera le but du chapitre III.

Au chapitre IV on a développé et on a appliqué le programme concernant la méthode du taux de passage par zéro pour la reconnaissance du locuteur.

Enfin en terminant notre travail par une conclusion générale.



Le processus de production de la parole est complexe et entraîne l'activation de nombreux muscles et articulateurs dont les commandes sont parfaitement synchronisées et contrôlées par le système nerveux. Dans ce chapitre, nous proposons d'étudier ce système afin de mieux comprendre le rôle des articulateurs les plus importants intervenant dans le processus de production de la parole. Nous allons donner un aperçu des éléments constituant le système vocal, ainsi qu'une brève description du système de phonation. Cette dernière va nous permettre de caractériser et d'identifier les grandes classes des sons élémentaires et expliquer les variations de ces sons. Puis nous terminons par l'explication du principe d'acquisition du signal de parole.

## **I – Système de production de la parole :**

### **I-1 Description et fonctionnement de l'appareil vocal :**

#### **I-1-1 Description de l'appareil vocal:**

L'appareil vocal est susceptible de produire une grande variété de sons parlés. Cette production est assurée grâce à un ensemble d'organes qui se compose des poumons, du larynx et du conduit vocal dans lequel on retrouve le pharynx, la cavité buccale et les cavités nasales.

La figure I-1 représente une coupe de l'appareil phonatoire humain.

#### **➤ La source : Les poumons :**

Les poumons constituent la source principale d'énergie dans la production des sons. C'est un générateur d'air qui alimente le larynx à travers la trachée. Cette dernière ne joue pas de rôle actif, mais assure la relation entre les poumons et le larynx.

#### **➤ Le larynx :**

Le larynx est situé dans la région moyenne du cou. Il se compose d'une série de muscles mobiles qui entourent une cavité située dans la partie supérieure de la trachée.

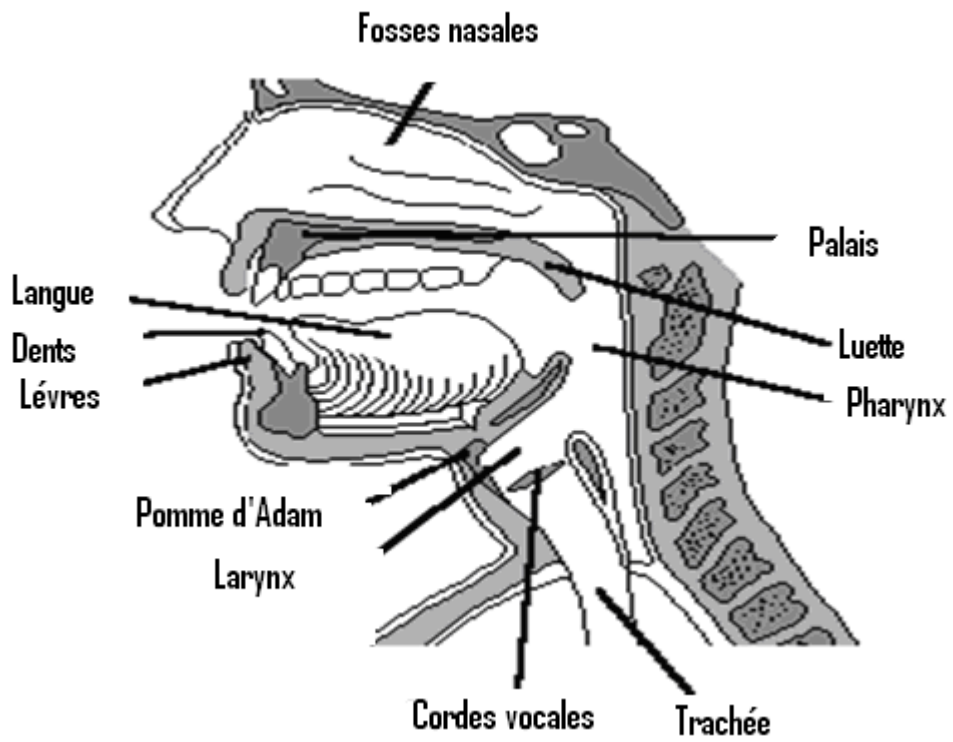


Figure I-1 : Coupe de l'appareil phonatoire humain [d'après Mella 93]

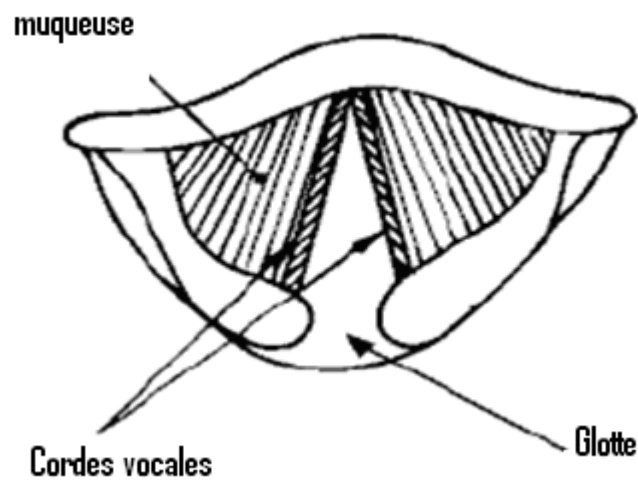


Figure I-2 : Vue de haut du larynx

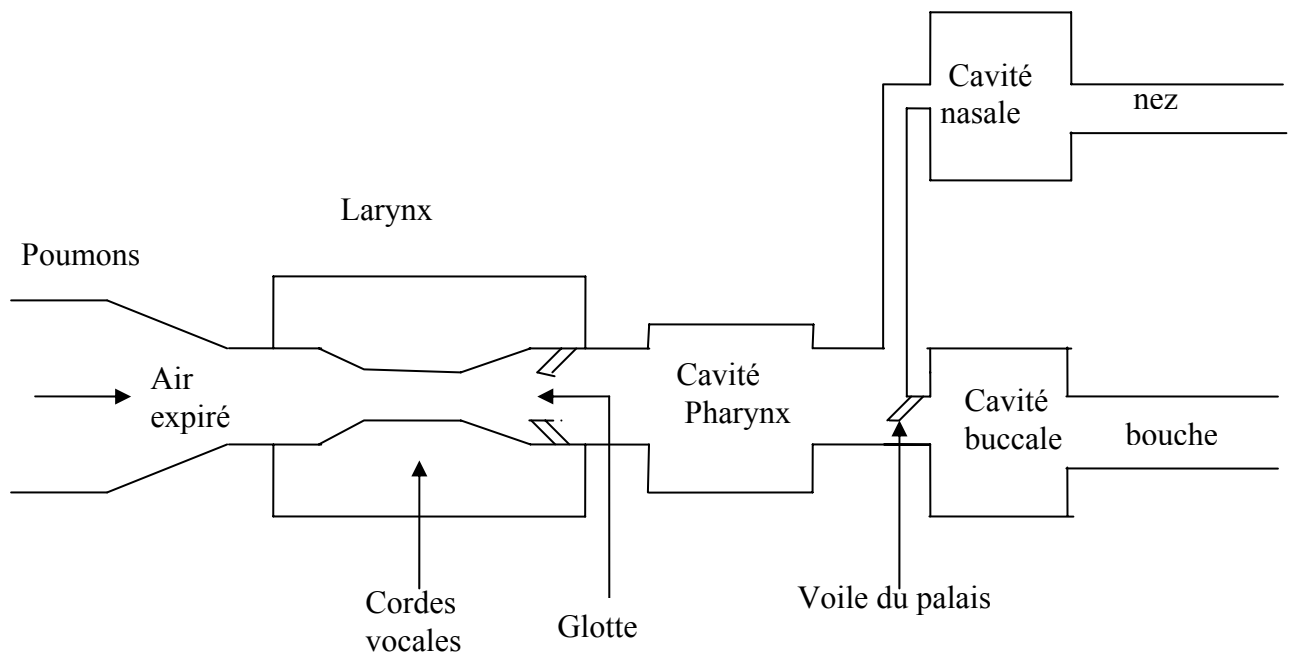
Les cordes vocales sont, en fait, des muscles et des ligaments, placés en travers du larynx, et qui peuvent fermer complètement le larynx puis s'écarter progressivement. Elles déterminent alors une ouverture triangulaire appelée « *glotte* » (figure I-2). C'est le degré de cette ouverture qui détermine la nature du flux d'air qui va exciter le conduit vocal.

Le larynx, essentiellement les cordes vocales, jouent le rôle d'un vibreur.

### ➤ Le conduit vocal :

Le conduit vocal comporte des résonateurs en série et en dérivation, formés essentiellement de la « *cavité pharyngo-buccale* » et de la « *cavité nasale* » respectivement. Ce sont les muscles articulaires qui commandent les mouvements des mâchoires, de la langue, du voile de palais, et des lèvres qui régissent l'évolution du conduit vocal, donc ses caractéristiques acoustiques. Ces dernières permettront de distinguer les différents sons, selon deux fonctions: générateur de bruit et résonateur.

La figure ci-dessous représente un modèle simplifié de l'appareil phonatoire.



**Figure I-3 :** Modèle simplifié de l'appareil phonatoire.

### I-1-2 Fonctionnement de l'appareil vocal :

#### ➤ Production de l'onde glottique :

L'air pulmonaire rencontre un premier obstacle : les cordes vocales. Pendant la phonation, l'action des différents muscles tend à rapprocher les cordes vocales ce qui provoque la fermeture de la glotte. Sous l'effet de la pression amont, les cordes vocales se mettent à vibrer laissant passer l'air par impulsion. C'est ainsi que se forme l'onde glottique dont la période d'oscillation, ou fréquence fondamentale, appelée « Pitch », est déterminée par la masse et la tension des cordes vocales, ainsi que la pression subglottique. Ceci est à la base de la formation des sons « *voisés* » ou sonores (Figure I-4-a).

Dans le cas où les cordes vocales sont relâchées, la glotte s'ouvre et l'air passe librement à travers cette dernière permettant alors la production des sons « *non voisés* » ou *sourds* (figure I-4-b).

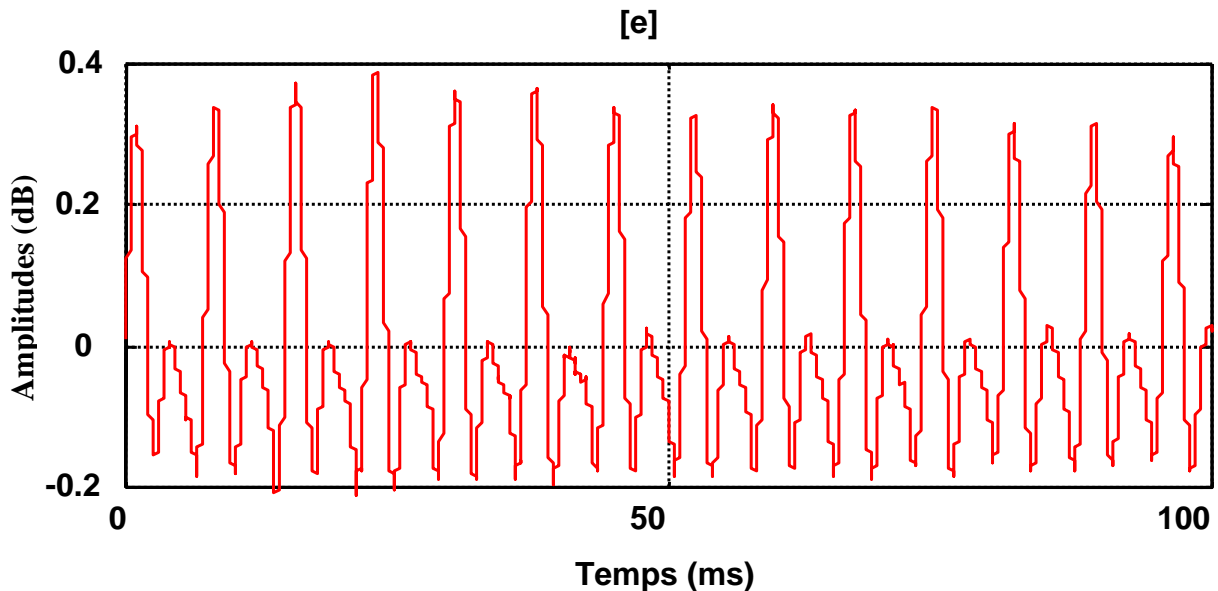


Figure I-4-a : Son voisé.

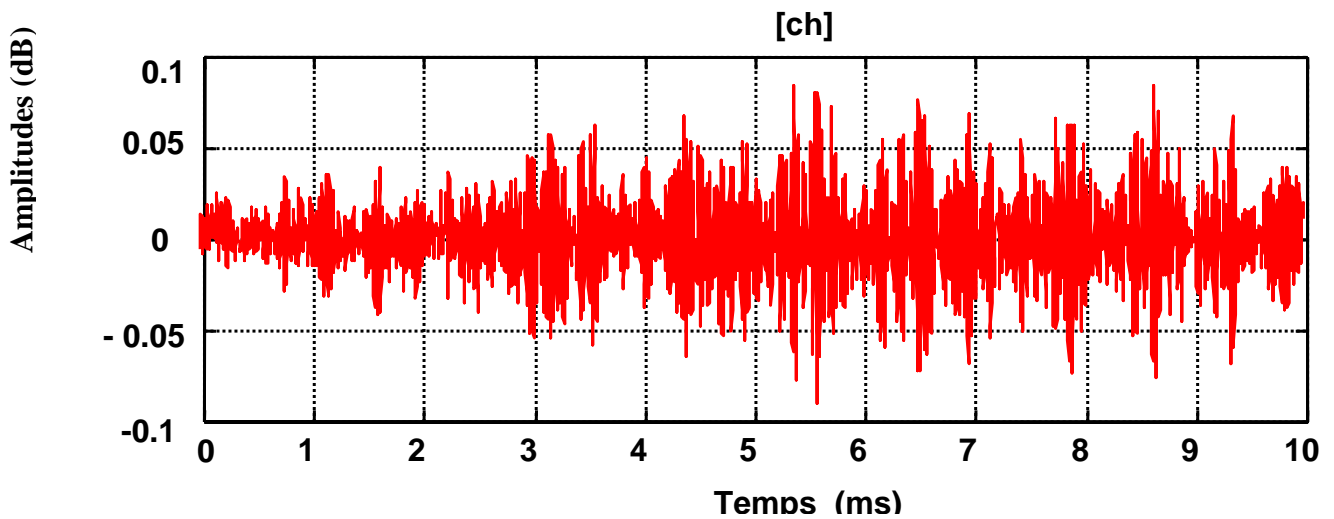


Figure I-4-b : Son non voisé.

➤ **Fonction résonatrice du conduit vocal :**

L'onde glottique subit des modifications lors de la traversée du conduit vocal. Les positions de la mâchoire et de la langue déterminent des cavités, jouant le rôle de caisse de résonance, qui vont renforcer certaines régions du spectre acoustique. Les maxima de la courbe de réponse en fréquence du conduit vocal sont appelés « *Formants* » (voir figure I-4). Les trois premiers formants sont essentiels pour caractériser le spectre vocal. Alors que les formants d'ordre supérieur ont une influence plus limitée.

La fréquence du premier formant varie de 200Hz à 900Hz.

La fréquence du deuxième formant varie de 500Hz à 2500Hz.

La fréquence du troisième formant varie de 1500Hz à 3500Hz.

Chaque voyelle a ses propres formants. Donc, pour n'importe quel spectrogramme, on peut distinguer les voyelles par leurs formants (Figures I-5-(a et b)).

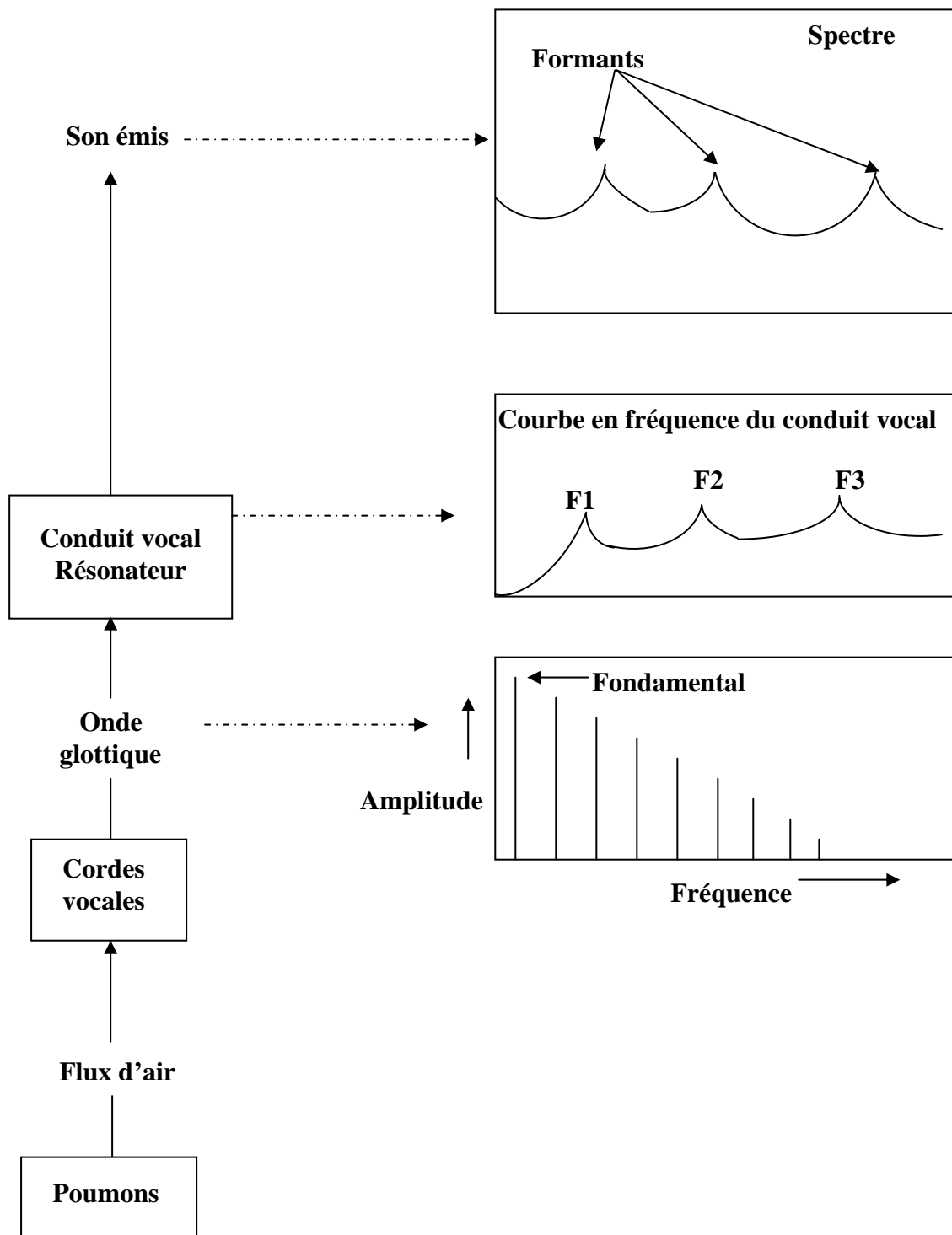


Figure I-4 : Production d'une voyelle d'après [Sundberg-80]

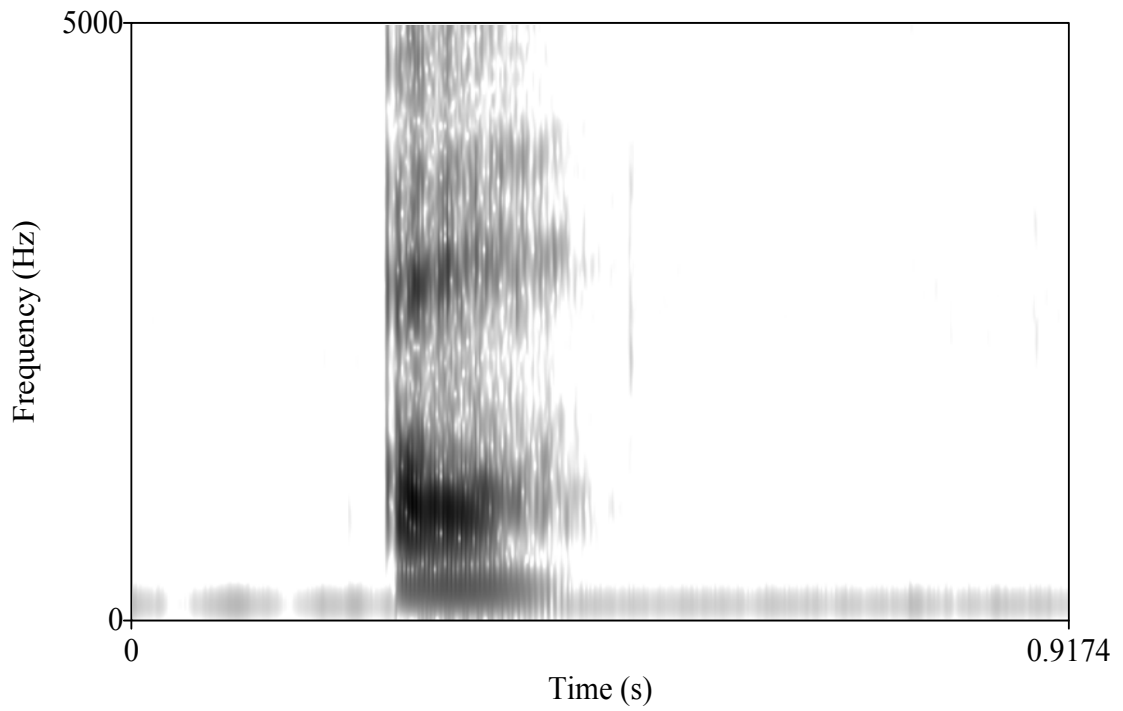


Figure I-5-a : Spectrogramme du son /a/.

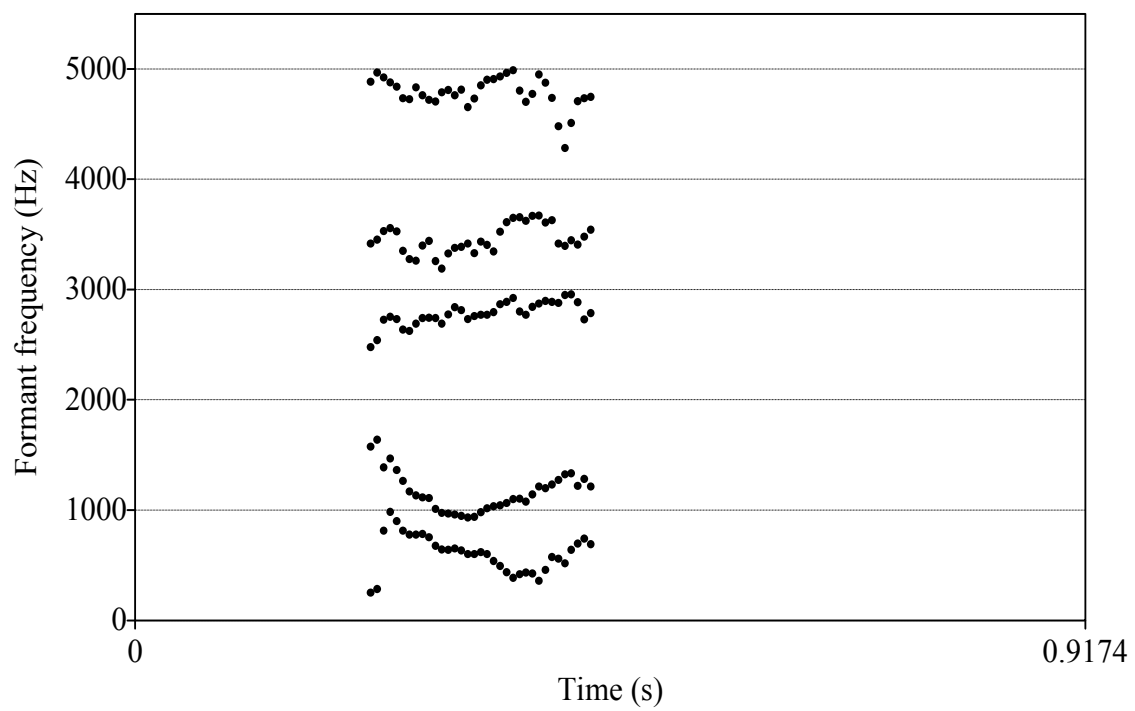


Figure I-5-b : L'évolution des formants du son /a/.

➤ **Fonction génératrice de bruit du conduit vocal :**

L'onde glottique peut rencontrer :

- Un obstacle partiel tel un rétrécissement du conduit vocal et un relâchement des cordes vocales, entraînant l'apparition d'un bruit de friction qui caractérise les sons fricatifs (/s/, /f/, /ch/, /z/, /v/, /j/).
- Un obstacle total produisant une augmentation de la pression derrière l'obstacle suivi d'un relâchement brusque, qui engendre la formation des sons « occlusifs » ou « plosifs » (/p/, /t/, /k/, /b/, /d/, /g/).

Le schéma ci-dessous décrit la fonction des parties essentielles de l'appareil phonatoire (fig I-6-a), ainsi que le modèle qui correspond à chaque partie (figI-6-b).

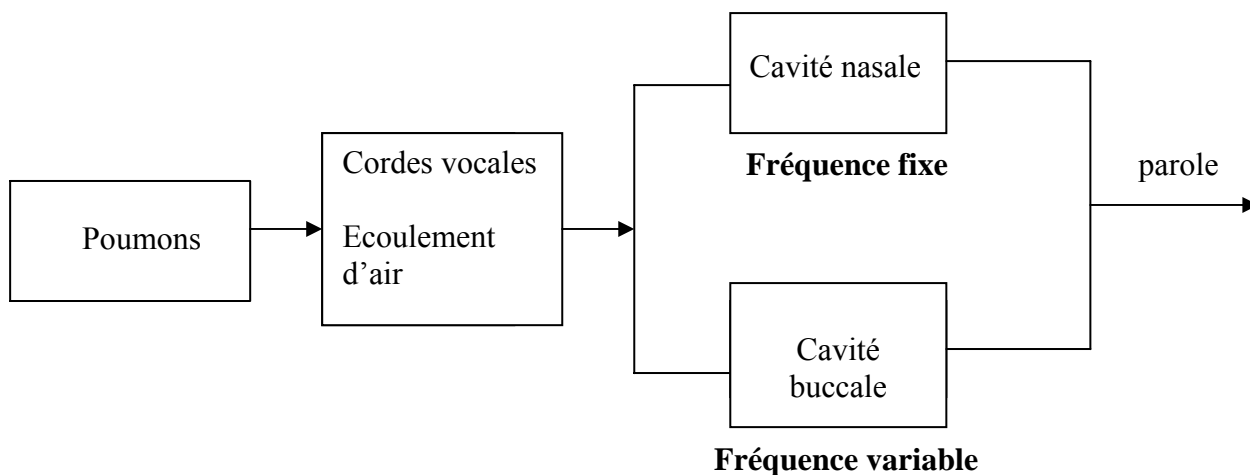


Figure I-6-a : les parties de l'appareil phonatoire

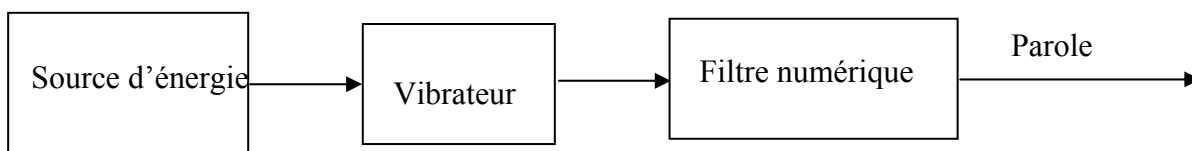


Figure I-6-b : Modèle de production de la parole.

## I-2 Taxonomie des sons (la théorie de classification) :

La taxonomie des sons est définie de deux manières, grâce à la phonétique et à la phonologie.

### ❖ La phonologie (phonétique fonctionnelle) :

Caractérise et compare les principes qui définissent l'organisation des sons des différentes langues. Elle essaie donc plus particulièrement d'expliquer les différences qui peuvent exister entre la transcription phonétique d'un mot du langage et la transcription phonétique exacte du mot qui est effectivement prononcé.

La phonologie introduit la notion d'unité abstraite du discours : « les phonèmes » et leurs dérivés (diphonème, triphonème, syllabes....).

**N.B :** Les phonèmes sont les éléments sonores les plus brefs qui permettent de distinguer différents mots. Il existe 37 phonèmes qui permettent de décrire le français parlé.

Exemples [p] [b] :

- pas / bas
- paie / baie.
- pot / beau

### ❖ La phonétique (phonétique articulatoire) :

La phonétique peut être définie comme l'étude de l'articulation de phonème. Elle peut être considérée descriptive, associant chaque son de la langue à un symbole et à une classe.

Il est intéressant de grouper les sons de parole en classe phonétique, en fonction de leur mode articulatoire (caractér

ise la forme et la façon dont la constriction est produite par le conduit vocal). On distingue généralement quatre classe principales : les voyelles, les semi-voyelles, les liquides et les consonnes.

#### ➤ Les voyelles :

Elles se caractérisent de tous les autres sons par le degré d'ouverture du conduit vocal. Si ce dernier est suffisamment ouvert pour que l'air poussé par les poumons le traverse sans obstacle, il y a production d'une voyelle [i, ɨ, e, ə, ε, Φ, a, œ, y, u, o, a]. Les voyelles se caractérisent, principalement, par le voisement qui crée des formants. Ces formants sont des

zones fréquentielles de forte énergie dont les fréquences peuvent s'élever jusqu'à 5 KHz, mais ce sont principalement les formants en basse fréquence qui caractérisent les voyelles.

➤ **Les consonnes :**

Elles sont caractérisées par une grande constriction du conduit vocal qui peut être un rétrécissement, une fermeture totale, cette caractérisation peut apparaître au niveau des lèvres, du palais ou des dents. Il existe deux sortes de consonnes :

▪ **Les consonnes occlusives :**

Lorsqu'il y a occlusion complète à un endroit du conduit vocal puis ouverture brutale, la consonne est dite occlusive (/p/, /t/, /k/, /b/, /d/, /g/).

▪ **Les consonnes fricatives :**

Le rétrécissement du conduit vocal et le relâchement des cordes vocales entraînent l'apparition d'un bruit de friction (/s/, /f/, /ch/, /z/, /v/, /j/).

▪ **Les consonnes nasales :**

Lorsqu'il y a occlusion du conduit et intervention de la cavité nasale, la consonne est nasale (/m/, /n/, /gn/).

➤ **Les semi-voyelles:** Elles combinent certaines caractéristiques des voyelles et des consonnes (/j/, /w/).

➤ **Les liquides :** Sont très similaires aux voyelles, mais leur durée et leur énergie sont plus faibles. Elles sont sonores (/l/, /r/).

### **I-3-Le signal vocal :**

Nous allons citer quelques caractéristiques du signal de parole, afin de faire ressortir les problèmes posés lors de son traitement.

#### **I-3-1-Description et caractéristiques:**

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.

Le système auditif humain est sensible dans une gamme allant de 800Hz à 8000Hz, les limites extrêmes sont respectivement 20Hz et 20KHz.

Le signal de parole est continu, non stationnaire, et d'énergie finie. Il présente une grande variabilité de sons selon le locuteur et les conditions d'environnement. Sa structure est complexe et variable dans le temps :

- Tantôt périodique (plus exactement pseudo-périodique) pour les sons voisés.
- Tantôt aléatoire pour les sons fricatifs.
- Tantôt impulsionnelle dans les phases explosives des sons occlusifs.

Selon la forme et l'évolution des organes constituant l'appareil vocal, différents sons peuvent être produit :

La vibration des cordes vocales engendre les sons voisés.

Le relâchement des cordes vocales engendre les sons non voisés.

Dans la première catégorie de sons, la forme du signal glottique est sensiblement triangulaire, son spectre est riche en harmonique (figure I-4-a), caractérisée principalement par le pitch.

Dans la deuxième catégorie de sons, le signal est apériodique (figure I-4-b), son spectre est relativement uniforme sur une large bande de fréquence (bruit blanc).

### **I-3-2-Variabilité du signal vocal :**

On distingue trois types de variabilité :

#### **➤ Variabilité intra-locuteur :**

Un même locuteur ne peut produire deux fois le même signal acoustique pour une même phrase. Cette variabilité s'explique par les variations de l'état physique et moral du locuteur. Une maladie des voies respiratoires peut ainsi dégrader la qualité du signal de parole. Le rythme d'élocution est également difficile à maîtriser précisément.

#### **➤ Variabilité interlocuteur :**

Cette variabilité s'explique par les différences physiologiques de l'appareil de production selon les locuteurs. Les variations des caractéristiques géométriques du conduit vocal, du volume des cavités buccale et nasale influencent sur la fréquence des formants. Les

variations de la pression subglottique et la longueur des cordes vocales entre locuteurs expliquent la différence de la fréquence fondamentale des locuteurs.

De nombreuses études ont montrés l'influence du sexe du locuteur sur la fréquence fondamentale (en moyenne : 100 à 150 Hz pour un homme, 200 à 350 Hz pour une femme) et sur les formants.

➤ **Variabilité due à l'environnement :**

La variabilité due à l'environnement peut provoquer une dégradation du signal de parole sans que le locuteur ait modifié son mode d'élocution.

Cette variation est considérée comme un bruit.

**I-4- Fréquence fondamentale :**

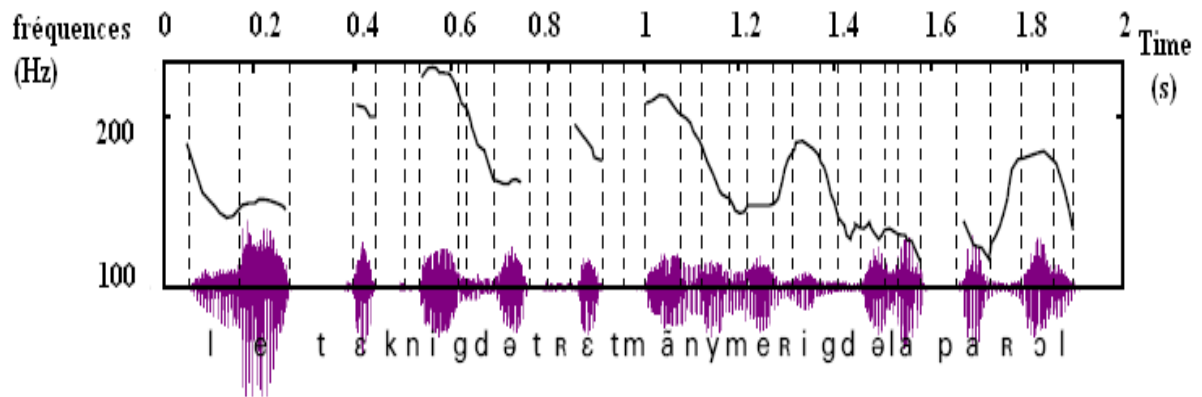
La fréquence fondamentale, appelée aussi pitch, caractérise la fréquence d'ouverture et de fermeture des cordes vocales. Typiquement, la période fondamentale des différents sons voisés varie entre 2ms et 20ms.

La figure I-7 donne l'évolution temporelle de la fréquence fondamentale de la phrase « *les techniques de traitement numérique de la parole* ». On constate qu'à l'intérieur des zones voisées la fréquence fondamentale évolue lentement dans le temps. Elle dépend essentiellement de l'âge et du sexe du locuteur, elle s'étend approximativement de 100 à 150Hz pour un homme, 200 à 350 Hz pour une femme et 200 à 600Hz pour les enfants.

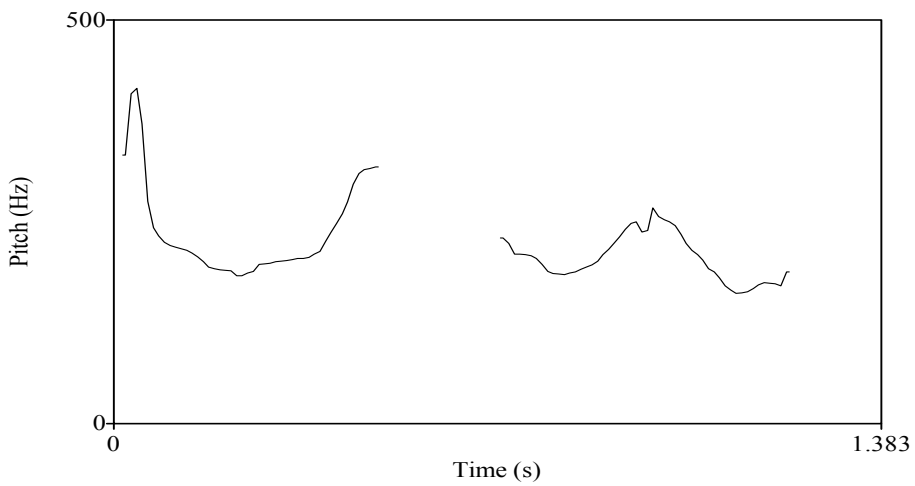
La fréquence fondamentale peut présenter des variations considérables chez un même locuteur :

- Selon le type de phrase prononcée.
- Selon l'état émotif et l'attitude du locuteur.

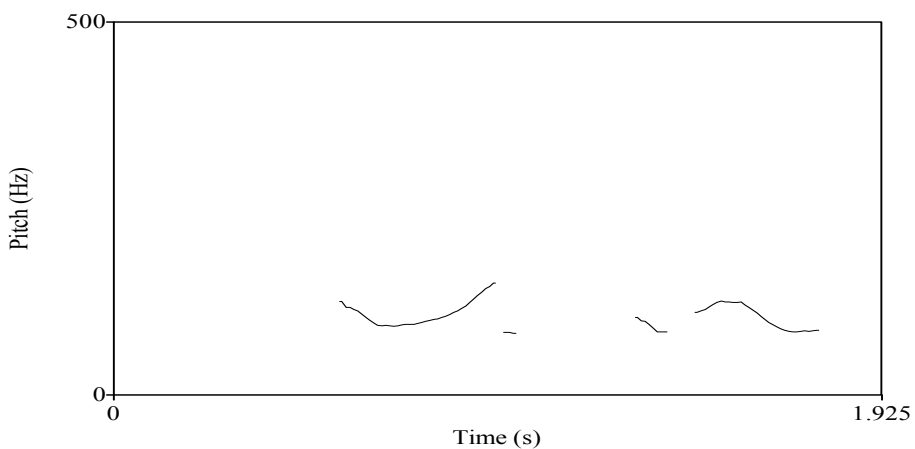
Les figures I-8-a et I-8-b représentent l'évolution temporelle du pitch pour l'homme et pour la femme respectivement dans la phrase « *pouvoir c'est vouloir* ». En utilisant le logiciel du traitement de son « *praat* ».



**Figure I-7 :** Evolution de la fréquence fondamentale dans la phrase « *les techniques de traitement numérique de la parole* ». Les sons non voisés sont associés à une fréquence nulle.



**Figure I-8-a :** Evolution du pitch de la femme dans la phrase « *pouvoir c'est vouloir* »  
Le pitch évolue de 160.994Hz à 408.3263Hz



**Figure I-8-b :** Evolution du pitch de l'homme dans la phrase « *pouvoir c'est vouloir* »  
le pitch évolue de 82.211Hz à 149.409Hz.

#### I-4 Acquisition du signal vocal :

L'acquisition de données consiste à recueillir les signaux de leurs sources et à les numériser pour les stocker, les analyser et les présenter sur un PC. Différentes composantes sont à prendre en compte lors de la construction d'un système d'acquisition du signal de parole :

➤ **Transducteurs et capteurs (microphone) :**

Une fois que le son a été émis par le locuteur, il est capté par un microphone qui transforme les vibrations sonores en courant électrique (l'énergie mécanique en énergie électrique).

➤ **Le filtre d'accentuation :** pour renforcer les aigus dans les hautes fréquences.

➤ **L'amplificateur :**

Il est placé juste à la sortie du capteur, son rôle est d'amplifier le signal issu du microphone qui est de faible énergie, il assure la protection contre surcharge, court circuit et échauffements, de manière à ce qu'il soit traitable par les autres étages. Le signal ne dépasse pas généralement les 4mv.

➤ **Le filtre d'anti-repliement (filtre de garde):**

Le signal électrique est ensuite appliqué à l'entrée d'un filtre passe bas dit filtre d'anti-repliement dont l'utilité sera justifiée par le théorème d'échantillonnage. Il est destiné à éliminer toutes les fréquences qui sont supérieures à la demi-fréquence d'échantillonnage.

On passe maintenant à la numérisation du signal vocal qui est une opération indispensable afin de transmettre le signal vocal sur les réseaux de transmission de données. On utilise un convertisseur analogique numérique (CAN) qui effectue les opérations suivantes :

**Première opération : Echantillonnage.**

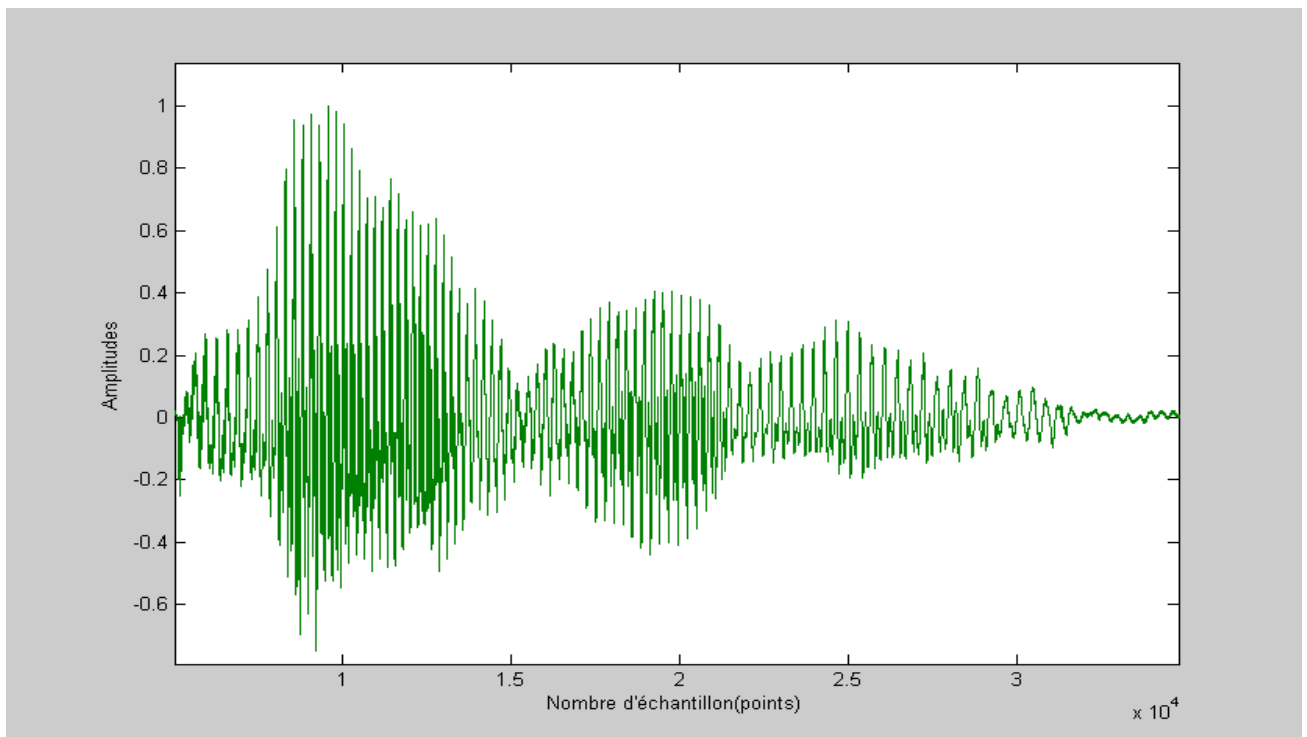
L'échantillonnage consiste à passer d'un signal à temps continu en une suite discrète de valeurs (valeurs mesurées à intervalles réguliers). On appelle « période d'échantillonnage » la

durée entre la prise de deux échantillons successifs, et elle est notée «  $T_e$  ». Son inverse est appelée « *fréquence d'échantillonnage* », elle est notée :  $F_e = 1/T_e$ .

Dans le cas où on souhaite reconstituer le signal analogique, il faut poser une contrainte dans le choix de ' $T_e$ ', cette contrainte découle du « *théorème d'échantillonnage* » ou de « Shannon ». La fréquence  $F_e = \frac{1}{T_e}$  doit valoir au moins le double de la fréquence maximale ' $F_{max}$ ' du signal c'est à dire :  $F_e \geq 2 \cdot F_{max}$ .

Les fréquences inférieures à la demi fréquence d'échantillonnage ne seront pas affectées par l'effet de repliement, donc si on souhaite représenter correctement des fréquences jusqu'à 20kHz, il est nécessaire d'échantillonner le signal à une fréquence supérieure ou égale à 40kHz. Un filtre appelé « *anti-repliement* » est disposé en amont de la conversion afin d'éliminer toutes les composantes de fréquences supérieures à la demi fréquence d'échantillonnage  $F_c < \frac{1}{2} \cdot F_e$ .

La figure I-10 représente un signal parole « *bienvenue* » échantillonné à  $F_e = 44.1$  KHz.



**Figure I-10 :** Exemple d'échantillonnage d'un signal parole « *bienvenue* »  $F_e = 44100$  Hz.

### Deuxième opération : Quantification.

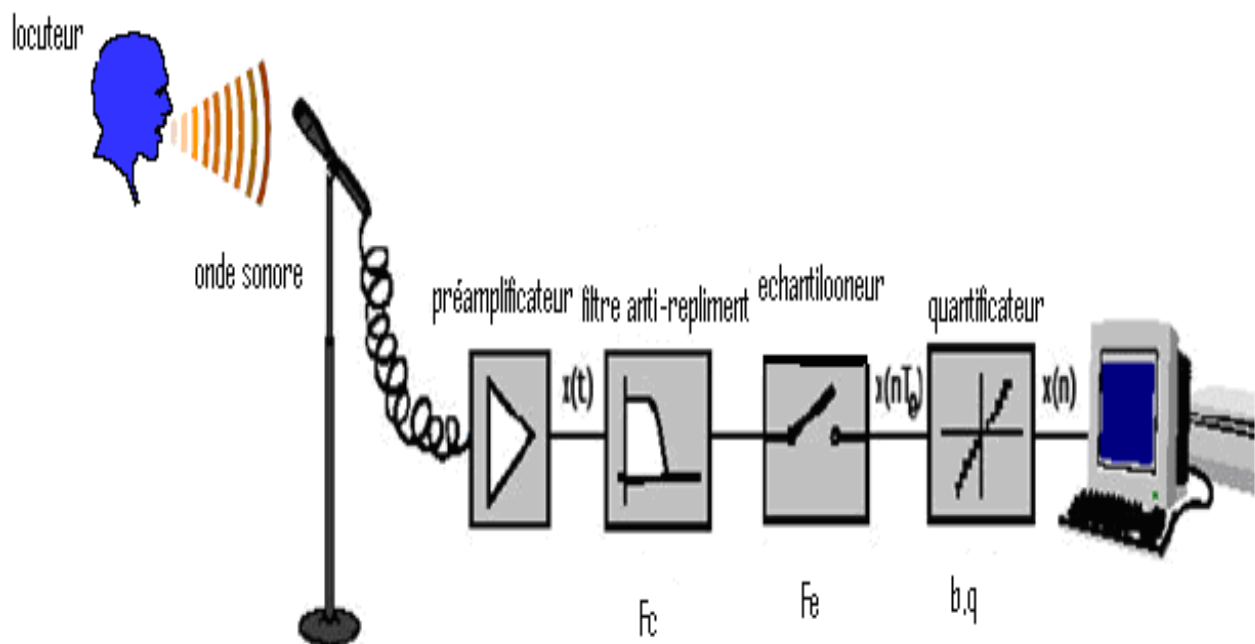
Elle permet de mesurer l'amplitude de l'onde sonore à chaque pas de l'échantillonnage. L'amplitude mesurée du signal est codée sur 'b' bits, donc en  $2^b$  classes. L'approximation de la valeur analogique par une valeur numérique discrète provoque une distorsion du son, dite bruit de quantification. Pour limiter ce bruit on choisit un nombre de classe élevé.

L'erreur relative de quantification vaut en moyenne  $(1/2^b)$  tel que b est le nombre de bits utilisé pour le codage des échantillons.

La quantification permet de mesurer l'amplitude de l'onde sonore à chaque pas de l'échantillonnage.

Dans le cas d'une quantification linéaire (c'est-à-dire le pas de quantification 'q' est constant sur toute la plage de conversion) on exprime cette erreur sous la forme d'un rapport signal à bruit (SNR : signal to noise ration).

Les étapes d'acquisition du signal vocal sont représentées sur la figure I-11.



**Figure I-11** : Acquisition d'un signal acoustique. La fréquence de coupure filtre de garde, la fréquence d'échantillonnage, le nombre de bits et le pas de quantification sont respectivement notés  $F_c$ ,  $F_e$ ,  $b$  et  $q$ .

**Conclusion :**

Ce chapitre qui porte un aperçu sur le système de production de la parole, a permis de donner une description simple du système phonatoire, et le rôle qui joue ce dernier dans la production du signal de parole. Par ailleurs, nous avons également abordé certaines caractéristiques de la parole, afin de comprendre les différents aspects du signal vocal tel que l'analyse et la reconnaissance, qui constituent le but des chapitres qui suivent.

L'analyse de la parole est une étape indispensable à toute application de synthèse, de codage, et de reconnaissance. Elle permet de fournir une description du signal acoustique et l'extraction des paramètres significatifs, tel que la fréquence fondamentale, les formants et l'énergie. Il existe alors, des méthodes qu'on peut classer en deux catégories; directe et indirecte. Dans ce chapitre on va se baser sur la deuxième catégorie dont le but est de traiter le signal de parole.

## **II- L'analyse du signal de parole :**

### **II-1-Les différentes méthodes d'analyse du signal de parole :**

Il existe deux principes différents d'analyse, directe et indirecte, selon que l'on s'intéresse aux grandeurs anatomiques, ou bien aux signaux électriques correspondants respectivement.

#### **II-1-1- La méthode directe :**

Cette méthode consiste à étudier le comportement des organes de l'appareil vocal tel que la glotte, le larynx et les cordes vocales. On utilise alors quelques techniques assurant cette analyse.

L'objectif est d'obtenir une image de la fréquence fondamentale. Les techniques utilisées sont les suivants :

##### **II-1-1-1- La glottographie photoélectrique :**

Elle utilise un principe optique qui consiste à mesurer les variations continues de la surface de la glotte par l'intermédiaire d'un photo-transistor inséré par le nez dans le pharynx, et qui mesure l'intensité lumineuse, à travers la peau du cou, au niveau de la glotte. Le signal électrique délivré par le photo-transistor représente la variation de l'air de la glotte. C'est la glottographie photoélectrique.

**II-1-1-2- La glottographie électrique :**

Elle s'effectue au moyen de deux électrodes fixées sur le cou sans modifier les conditions normales de phonation, elle permet la mesure des variations de l'impédance du tissu musculaire du larynx.

**II-1-1-3- La glottographie ultrasonore :**

A travers la réflexion des ondes ultrasonores sur des cristaux piezo électriques placés autour du cou, on peut mesurer le temps de fermeture et d'ouverture des cordes vocales.

**II-1-2- La méthode indirecte :**

C'est cette méthode qui nous intéresse car elle traite le signal électrique délivré par un microphone. On peut traiter ou analyser le signal temporellement ou fréquentiellement selon les paramètres à extraire. Nous allons expliquer le principe des différentes méthodes utilisées en précisant, à chaque fois, les paramètres caractéristiques entrant en jeu dans le domaine temporel et dans le domaine fréquentiel.

**II-1-2-1- L'analyse temporelle :**

Elle fournit des paramètres à partir du traitement effectué sur la représentation temporelle du signal.

De nombreuses mesures sur cette représentation ont été élaborées, parmi elles nous nous intéressons à celles qui suivent :

- Le taux de passage par zéro (TPZ).
- L'énergie du signal.
- La fonction d'autocorrelation.

**Remarque :**

Avant tout calcul, il est nécessaire de mettre en forme le signal de parole avec le fenêtrage, qui consiste à segmenter le signal en trames, Chaque trame est constituée d'un nombre N fixe d'échantillons de parole. En général, N est fixé de telle manière que chaque trame corresponde à environ 20 ms de parole (durée pendant laquelle

la parole peut être considérée comme stationnaire). La fenêtre la plus utilisée est celle de Hamming. Pour plus de détails voir l'annexe.

### II-1-2-1-1- Le taux de passage par zéro (T P Z) :

Pour un signal échantillonné, il y a passage par zéro lorsque deux échantillons successifs sont de signes opposés. Le taux de passage par zéro à court terme peut être estimé par la formule suivante :

$$TPZ(x) = \frac{1}{N} \sum_{n=1}^{N-1} a_n \quad \text{Avec } a_n = 1 \quad \text{si } x_{n+1} \cdot x_n < 0 \quad \dots\dots\dots \text{II-1[9]}$$

$$a_n = 0 \quad \text{sinon}$$

$n$  : Nombre d'échantillon du signal de parole.

$N$  : Nombre d'échantillon par fenêtre.

$x_n$  : Échantillon du signal.

$x_{n+1}$  : Échantillon suivant du signal.

Le TPZ est un indice très simple à calculer. Il représente le nombre de fois que le signal, dans sa représentation amplitude/temps, passe par la valeur centrale (zéro). Du fait de leur nature aléatoire, les sons non voisés (voir chapitre I) possèdent un taux de passage par zéro supérieur à celui des sons voisés.

Cette propriété est exploitée pour arriver à une classification des signaux en sons voisés et non voisés.

Nous avons pu observer que le TPZ présente une moyenne de 12 pour les sons voisés (a, o, y, u) et de 35 pour les sons non voisés (p, t, k) [7].

Un exemple de variation du taux de passage par zéro d'une phrase de notre corpus est donné en figure II-1.

### II-1-2-1-2-Analyse par l'énergie :

L'énergie à court terme sert à détecter les silences, elle est élevée s'il y a présence d'un signal (son), et petite s'il n'y a pas de son c'est à dire présence d'un silence.

Les sons voisés présentent une valeur d'énergie plus grande à celle des sons non voisés.

L'énergie  $E$  en décibel (dB) est définie par :

$$E = 10 \log \sum_{n=0}^{N-1} |x(n)|^2 \dots\dots\dots\text{II-2}[7]$$

$x(n)$  : Est un signal de 'n'échantillons.

$N$  : Nombre d'échantillon par fenêtre.

La figure II-2 présente la variation de l'énergie  $E$  de la phrase « *il s'arrêtait tout l'été ici* »

Il existe toutefois un paramètre plus efficace pour déterminer le taux de voisement d'un signal. C'est ce qu'on va voir dans le paragraphe suivant.

### II -1-2-1-3- L'autocorrelation :

L'autocorrelation est la convolution du signal avec lui-même, elle mesure la similitude du signal  $x(t)$  avec une version décalée de  $x(t)$ .

Elle est défini par la formule suivante :

$$R(k) = \sum_{l=0}^N x(l) \cdot x(l+k) \dots\dots\dots\text{II-3}[8]$$

Où  $x(t)$  : est le signal à l'instant  $t$  et  $x(t+k)$  est le signal à l'instant  $t+k$ .

$t$  et  $k$  sont des instants du signal  $x(t)$ .

$N$  : représente le nombre de point de chaque fenêtre.

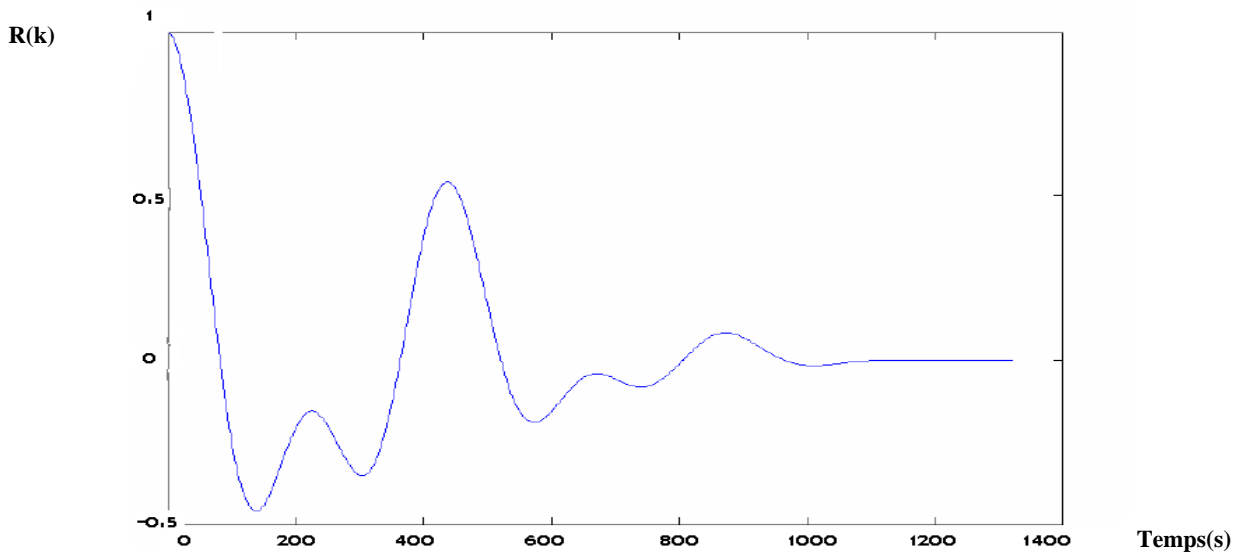
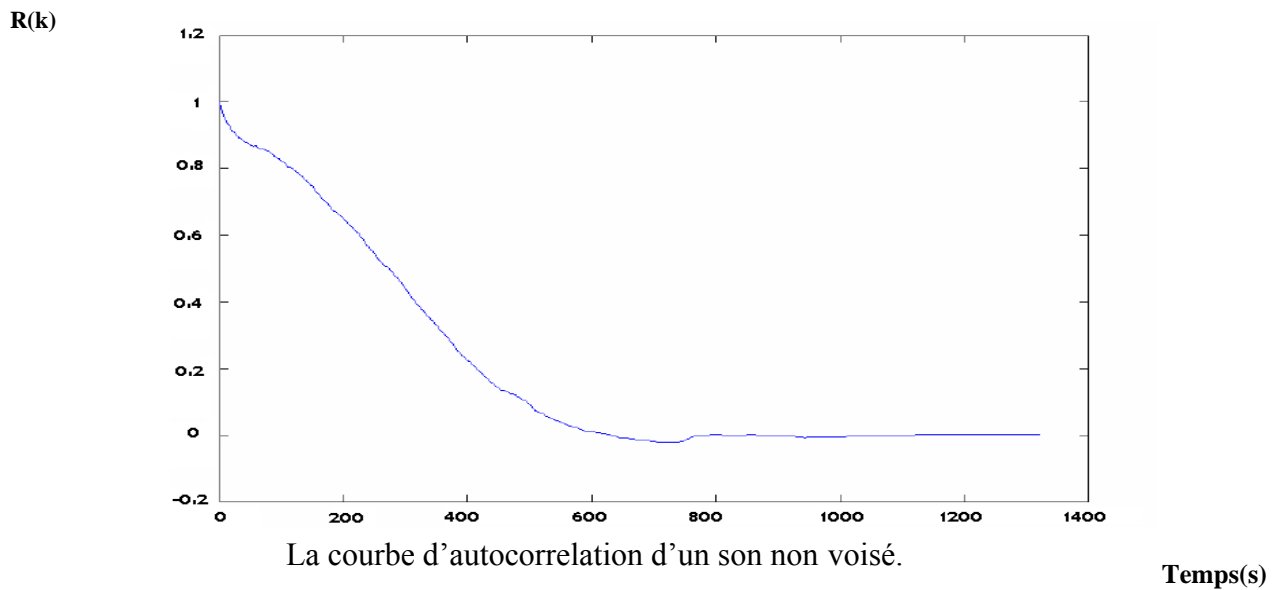
L'autocorrelation est une méthode utilisée pour distinguer un son voisé d'un son non voisé, elle est couramment utilisée pour déterminer la fréquence fondamentale.



La figure II-3 représente les courbes d'autocorrelation du son voisé et non voisé.

Le résultat d'une autocorrelation dans le cas d'un son voisé (possédant une fréquence fondamentale  $F_0$ ) est une suite de lobes espacés de  $n_0$  échantillons ( $n_0$  détermine la distance entre deux lobes successifs elle est équivalente à la période  $T_0$ ). La fréquence fondamentale peut être déterminée en prenant l'inverse de la distance entre les deux premiers lobes de la deuxième moitié de la courbe ( $F_0 = \frac{1}{n_0} = \frac{1}{T_0}$ ).

Si le son n'est pas voisé, la courbe d'autocorrelation décroît sans pics jusqu'à zéro.



**Figure II-3** : Autocorrelation d'un son voisé et non voisé.

### **II -1-2-2- Analyse fréquentielle :**

On passe maintenant à la méthode la plus courante pour le traitement du signal de parole qui est l'analyse fréquentielle, pour extraire les propriétés spectrales d'un signal vocal qui présentent un intérêt majeur pour la perception auditive.

L'analyse spectrale a pour but de caractériser la répartition d'énergie ou de puissance d'un signal en fonction de la fréquence. Parmi les outils permettant l'analyse spectrale du signal vocal, nous citons :

- L'analyse spectrale par spectrogramme.
- L'analyse spectrale par le cepstre.
- L'analyse spectrale par la prédiction linéaire (LPC : Linear Predictive Coding).
- L'analyse spectrale par la TFCT (Transformée de Fourier à Court Terme).
- L'analyse spectrale par les ondelettes.

#### **II-1-2-2-1- Le Spectrogramme :**

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un « spectrogramme » à l'aide d'un « sonographe ».

Le spectrogramme est une représentation tridimensionnelle :

- Le temps en abscisses.
- La fréquence en ordonnées.
- L'énergie est représentée par un niveau de gris pour un temps et une fréquence donnée.

On peut générer des spectrogrammes à larges bandes (fenêtre de Hamming par exemple de 4ms et déplacement de 2ms) ou des spectrogrammes à bandes étroites (fenêtre de 30ms) selon la durée des fenêtres de pondération.

La figure II-4 représente les deux types de spectrogrammes. Les spectrogrammes à bandes larges sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10ms) ; ils mettent en évidence l'enveloppe spectrale du signal, ils permettent par conséquent de visualiser les formants. Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres.

Les spectrogrammes à bande étroite 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.

Un exemple d'un signal sonore est son spectrogramme est donné à la (figure II-5). La phrase « *le sage ne dit pas se qu'il sait* » prononcée par un homme. La fenêtre utilisée est celle de Hamming, avec le logiciel de traitement du son "praat".

Fréquences (Hz)

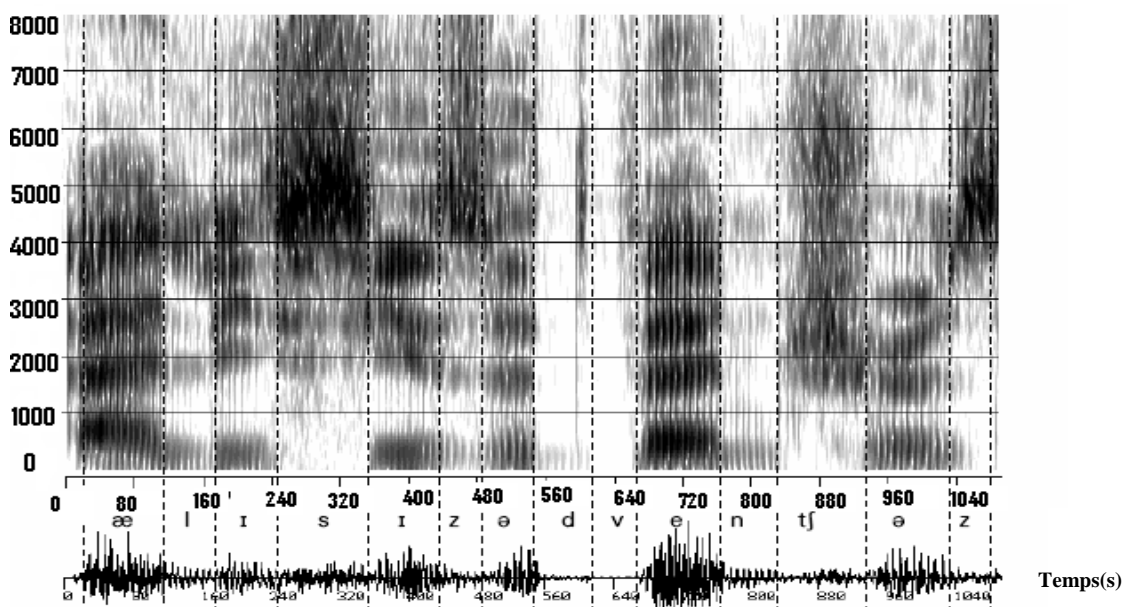
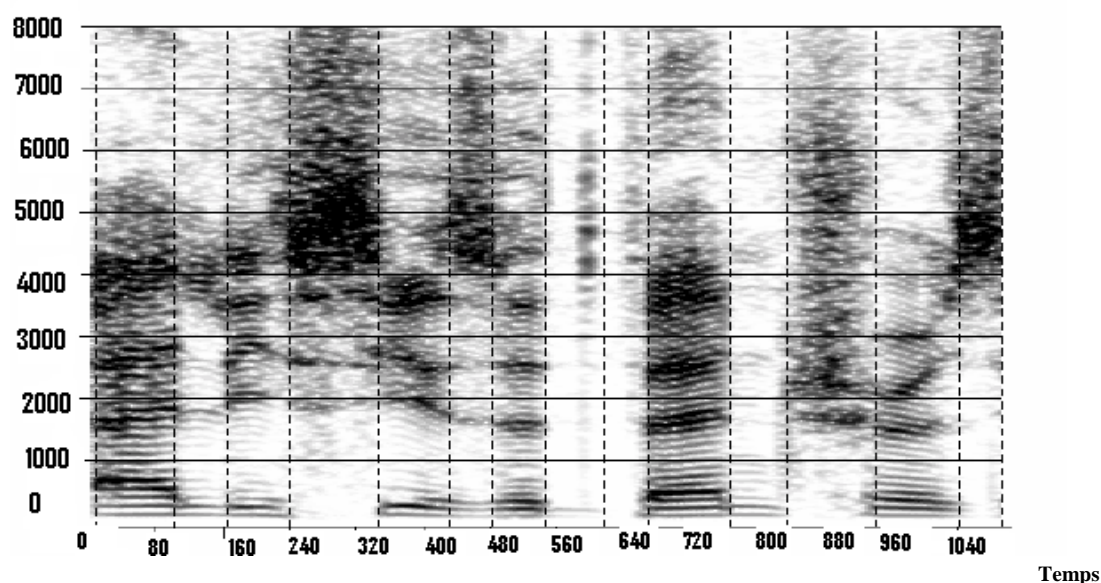
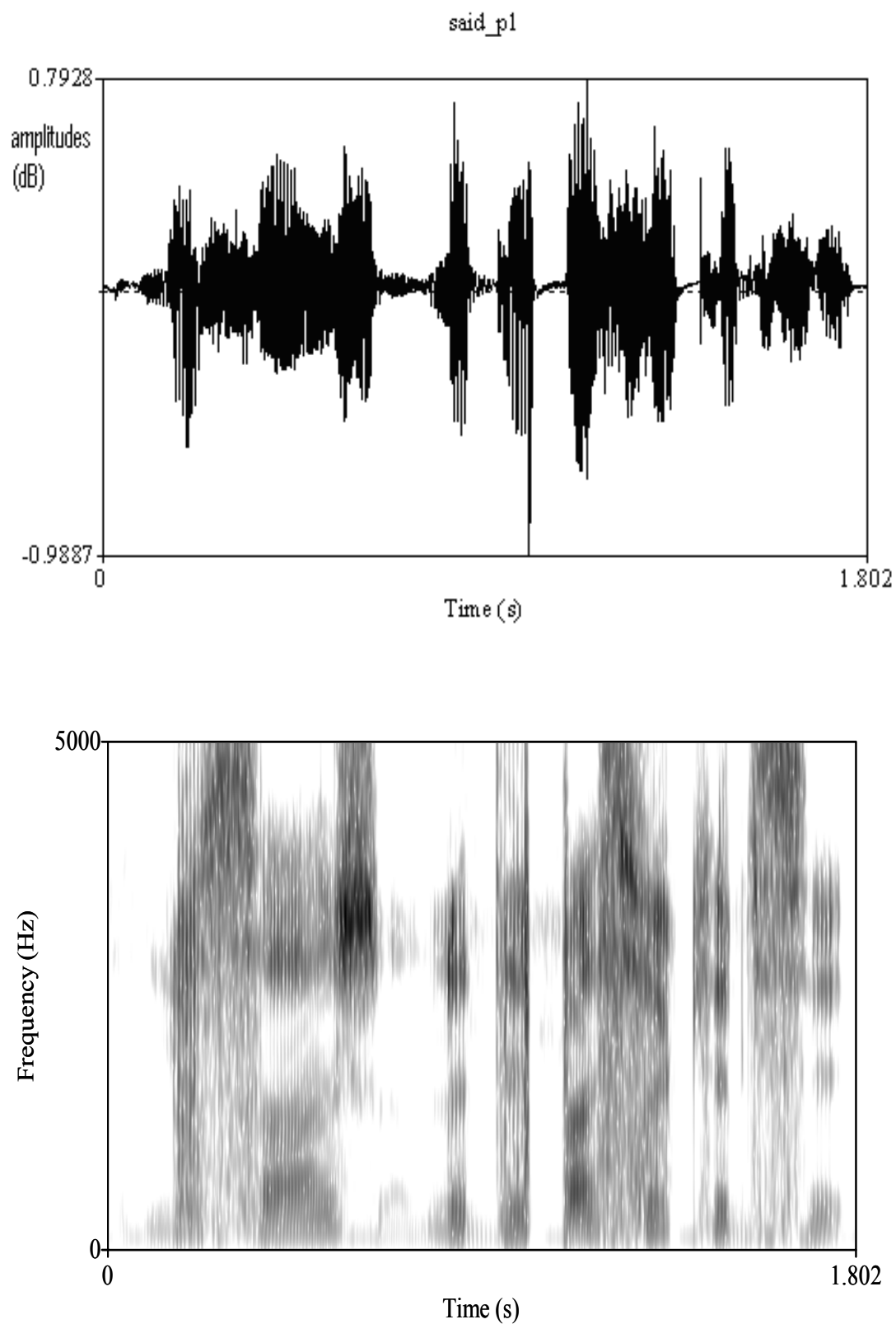
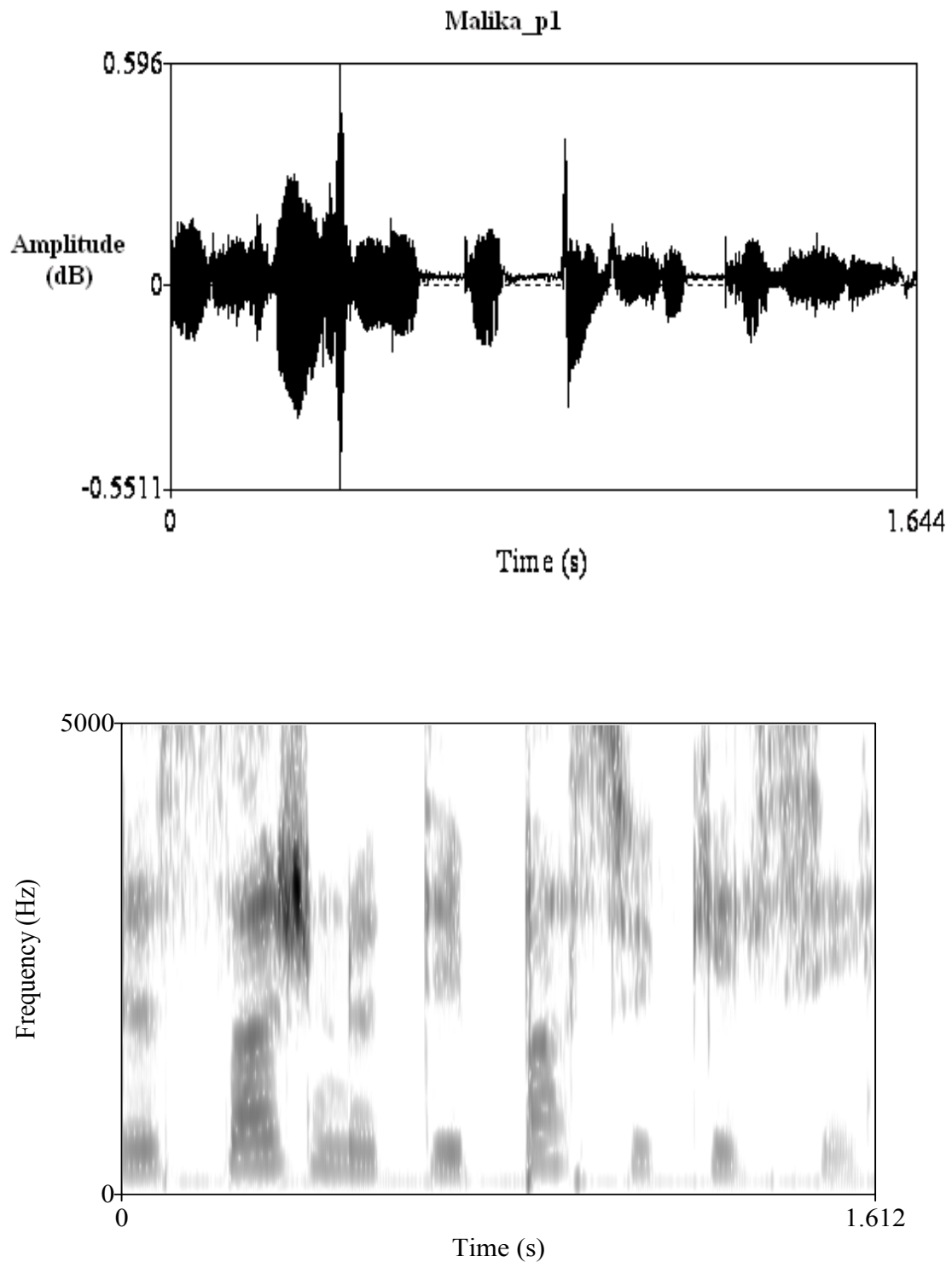


Figure II-4 : Spectrogramme à large bande (en bas), et à bande étroite (en haut) de la phrase « *Alice's adventures* ».



**FigureII-5-a** : Le spectrogramme de la phrase « *Le sage ne dit pas ce qu'il sait* » prononcée par un homme.



**FigureII-5-b** : Le spectrogramme de la phrase « *Le sage ne dit pas ce qu'il sait* » prononcée par une femme.

Les spectrogrammes des figures II-5-a et b montre la différence qui existe entre la femme et l’homme. D’où la possibilité de faire la reconnaissance du locuteur avec les spectrogrammes.

**II-1-2-2-2-Analyse spectrale par le cepstre :**

Le spectre donné par FFT (Transformée de Fourier Rapide, Fast Fourier Transform en anglais), contient des renseignements sur la source et le conduit vocal, mais leur intermodulation rend difficile la mesure de  $F_0$  (fréquence fondamentale) et celle des formants qui caractérisent respectivement la source et le conduit. Le cepstre est une méthode visant à séparer la contribution de la source et celle de conduit par déconvolution dont le principe est de situer dans un espace où l’opérateur de convolution (\*) correspond à un opérateur d’addition (+).

On fait l’hypothèse que le signal vocal  $s(t)$  est le résultat de la convolution du signal de la source  $e(t)$  par le filtre correspondant au conduit vocal  $h(t)$  (figure II-6).

$$S(t) = e(t) * h(t) \dots\dots\dots \text{II-4}[11]$$

- En passant aux transformées de Fourier :

$$S(f) = E(f).H(f) \dots\dots\dots \text{II-5}[11]$$

- On prend le logarithme du module (on ne s’intéresse pas à l’information de phase), afin de transformer le produit en somme.

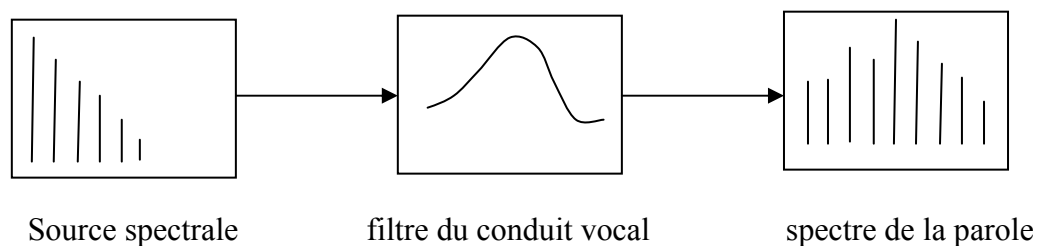
$$\text{Log}|S(f)| = \text{Log}|E(f)| + \text{Log}|H(f)| \dots\dots\dots \text{II-6}[11]$$

- Par transformation inverse ( $TF^{-1}$ ) (nous permet de revenir dans le domaine temporel) on obtient le cepstre :

$$TF^{-1}[\text{Log}|S(f)|] = TF^{-1}[\text{Log}|E(f)|] + TF^{-1}[\text{Log}|H(f)|] \dots\dots \text{II-7}[11]$$

$$C(t) = e'(t) + h'(t) \dots\dots\dots \text{II-8}[11]$$

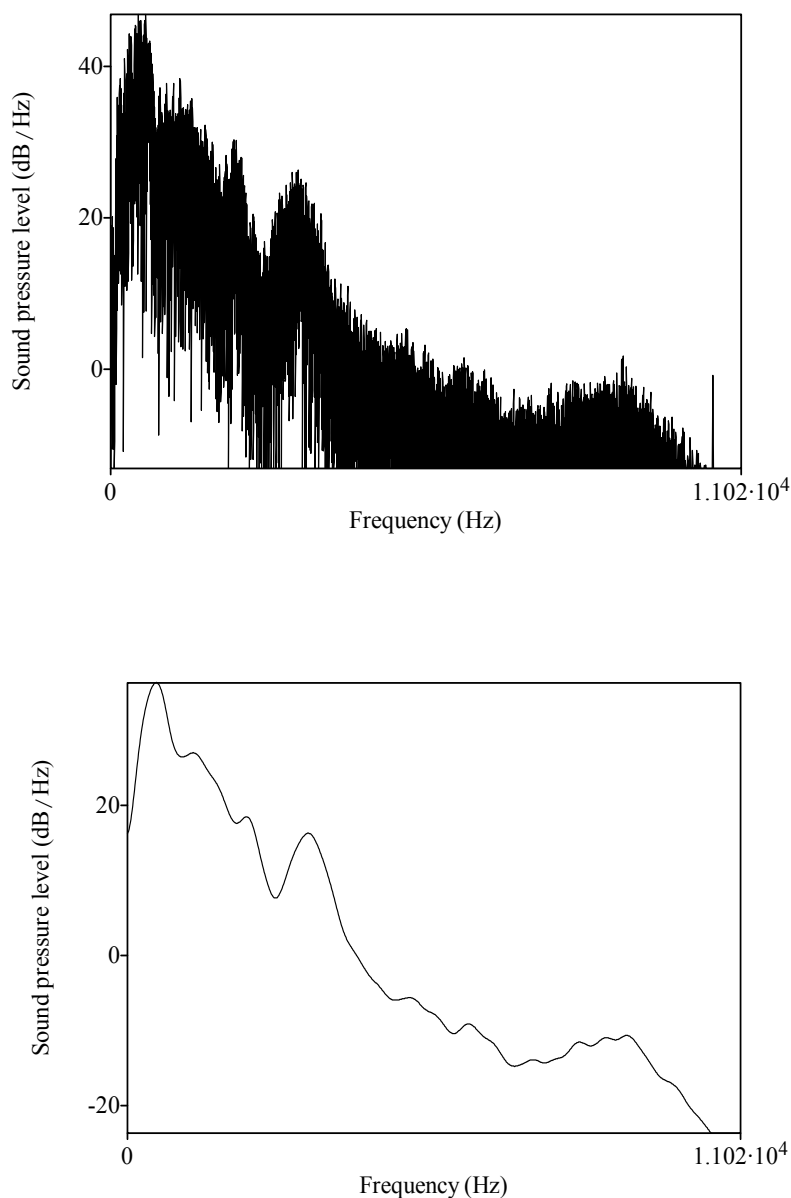
$$S(t) = e(t) * h(t) \longrightarrow C(t) = e'(t) + h'(t).$$



**Figure II-6 :** Modèle spectral source-filtre.

En posant l'hypothèse que  $e'(t)$  est une séquence d'impulsion séparés de  $n_0$  échantillons (ou  $n_0$  correspond à  $F_0$ ) et que  $h'(t)$  décroît rapidement. On peut séparer les contributions de la source et du filtre en ne conservant que les  $n < n_0$  premiers coefficients cepstraux. Le spectre du cepstre pour  $n < n_0$  permettra d'obtenir un spectre lissé, débarrassé des lobes dus à la contribution de la source, et il ne contient que des informations sur le conduit, ses extrêmes correspondent aux résonances du conduit vocal (Formants). Les pics périodiques visibles au-delà de  $n_0$  correspondent aux impulsions de la source.

La figure II-7 représente un lissage cepstral du spectre de fréquence d'un signal sonore.



**Figure II -7:** Représentation spectrale d'un signal et son lissage.

### **II-1-2-2-3-L'analyse par prédiction linéaire (LPC) :**

Le « Linear Predictive Coding » repose sur la connaissance du modèle de production de la parole (la source et le conduit). Dans ce modèle nous considérons que sur les intervalles d'environ 20ms à 30ms, le son est le résultat du filtrage d'un signal appelé excitation qui est :

- Soit un train de Dirac qui est produit par les cordes vocales, le son est alors dit voisé et est caractérisé par le « pitch ».
- Soit un bruit blanc si les cordes vocales ne vibrent pas, le son est alors dit non voisé.

Le filtre correspond au conduit vocal nous le considérons comme linéaire et autorégressif, d'où le terme « linéaire prédictive » qui signifie que le résultat à un instant donné peut être prédit par la connaissance des résultats aux instants précédents.

Le but du système LPC est de déterminer le formant du signal de discours. La solution est une équation de différence, qui exprime chaque échantillon du signal comme une combinaison linéaire d'échantillons précédents. Une telle équation est appelée une prédiction linéaire ; d'où le nom de Codage Prédicatif Linéaire.

$$S(n) = e(n) - [a(1)s(n-1) + a(2)s(n-2) + \dots + a(p)s(n-p)] \quad \dots\dots\dots \text{II-9 [3]}$$

Où: S (n) représente le signal à l'instant n, e (n) est l'excitation.

Les coefficients : a (1), a (2), ..., a (p) sont des coefficients LPC et p l'ordre du filtre.

### **Modèle de LPC manqué.**

#### **II-1-2-2-4 Analyse fréquentielle par la transformée de Fourier à court terme**

##### **(TFCT) :**

La transformée de Fourier (TF) tout court, a un intérêt majeur dans le traitement des signaux mais à plusieurs limitations.

En effet l'analyse de Fourier permet de connaître les différentes fréquences contenues dans un signal c'est-à-dire son spectre de fréquence, mais ne permet pas de savoir à quel instant ces fréquences ont été émises, autrement dit ; la TF donne une information globale et non locale. Cette perte de localité est un inconvénient pour analyser des signaux non stationnaires ou quasi-stationnaire notamment signal de parole.

La TF est définie par :

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) \exp(-j\omega t) dt \quad \dots\dots\dots \text{II-10 [3]}$$

$\omega = 2\pi f$  : Etant la pulsation

La TF discrète est :

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j \frac{2\pi}{N} k.n\right) \quad \dots\dots\dots \text{II-11[3]}$$

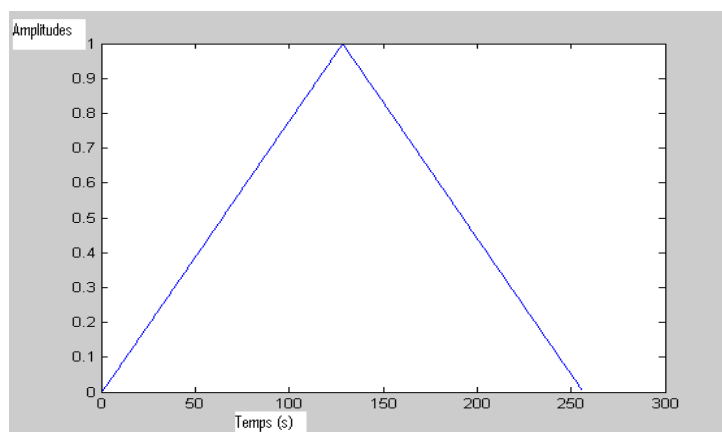
$N$  : Le nombre de point de la suite temporel  $x(n)$  .

$$f = \frac{k}{N} \text{ avec } k=0, 1, 2, \dots, N-1 \text{ et } n=0, 1, 2, \dots, N-1.$$

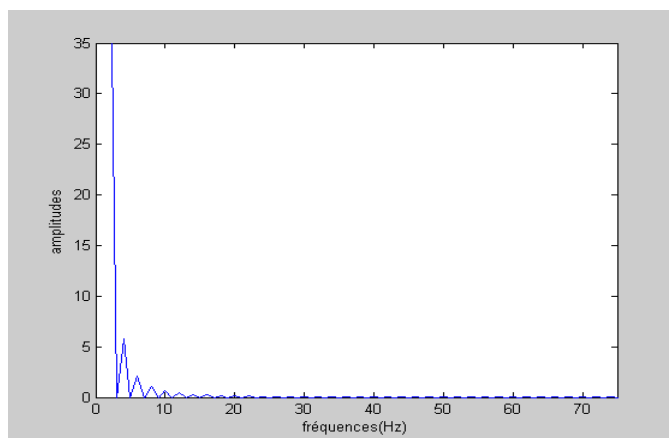
Afin de réduire le temps de calcul de la transformée de Fourier discrète (TFD), COOLEY et TUQEY en 1965 misent au point la transformée de Fourier rapide (FFT), elle permet d'obtenir des spectres en temps réel.

Pour concrétiser ça, on prend un signal réel 'un triangle', (figure II-8) et on calcul sa transformée de Fourier.

On remarque que sur la répartition fréquentielle du signal, l'allure de  $X(\nu)$  ne permet pas de connaître l'allure de  $x(t)$ , c'est-à-dire à quel instant apparaît le triangle.



La fonction triangle.



La TF du signal triangle.

**Figure II-8:** La transformée de Fourier d'un signal.

Pour remédier à cette déficience, on a mis en œuvre la Transformée de Fourier à Court Terme (TFCT) appelée aussi la Transformée de Fourier à Fenêtre Glissante (TFFG).

La TFCT consiste alors, à appliquer au signal une fenêtre d'observation glissante (en général de Hamming) qui sera translatée dans le domaine temporel de sorte que les hypothèses de stationnarité soient localement satisfaites et on applique la TF non pas au signal global  $s(t)$ , mais à chacune des tranches ainsi découpées. On obtient ainsi le spectre « local ». La totalité des spectres « locaux » indique alors comment le spectre varie au court du temps.

Sur l'axe horizontal on porte le numéro de la fenêtre (c'est-à-dire le temps) et sur l'axe vertical la fréquence. En effet on suppose que le signal est stationnaire sur toute la longueur de chaque fenêtre.

Ces fenêtres sont utilisées afin d'améliorer sensiblement le spectre, alors on la fait glisser devant le graphe, on obtient une famille de coefficients à deux paramètres  $f$  et  $\tau$ .

Pour un signal  $s(t)$ , la TFCT est définie par :

$$\text{TFCT}(\tau, f) = \int_{-\infty}^{+\infty} s(t) \exp(-2j\pi ft) W(t-\tau) dt \quad \dots\dots\dots \text{II-12 [3]}$$

$W(t-\tau)$  : représente la fenêtre d'analyse en la déplaçant sur le domaine de la fréquence définit du signal  $s(t)$ .

$\tau$  : est le paramètre de translation de la fenêtre d'analyse.

La TFCT est la transformée d'un produit donc c'est la convolution des transformées.

$$\text{TFCT} = F(\omega) * W(\omega) \quad \dots\dots\dots \text{II-13}$$

$\omega = 2\pi f$  : La pulsation.

La fréquence  $f$  est localisée autour de l'abscisse  $\tau$  du signal temporel, alors cette formule nous donne des indications sur tous ce qui se passe autour de  $\tau$  pour la fréquence  $f$ .

Malheureusement cette représentation (TFCT) présente les inconvénients suivants :

- La taille de la fenêtre étant fixe, cette représentation ne paraît pas adaptée à la représentation des fonctions ou des signaux ayant des composantes de taille très différentes de la taille de la fenêtre.
- La représentation de l'analyse impose un compromis insurmontable entre la résolution temporelle et la résolution fréquentielle.
- La détermination de l'intervalle temporel dans lequel on peut considérer le signal comme stationnaire est un inconvénient pratique.

La taille de la fenêtre d'analyse est assujettie au principe d'incertitude d'Heisenberg qui stipule que la résolution conjointe temps-fréquence est minorée :

$$\Delta t \Delta f > 1/4\pi$$

La figure II-9 représente un exemple d'un signal vocal de la phrase « *pouvoir c'est vouloir* » et application de la transformée de Fourier sur une fenêtre de 16000 points.

Afin de remédier aux problèmes de la TFCT, une nouvelle transformée appelée « *Transformée en Ondelette* » a été proposée. Elle offre une bonne localisation dans les deux domaines (c'est à dire en temps et en fréquence), et utilise des fenêtres mieux adaptées au contenu fréquentiel du signal à traiter.

#### II-1-2-2-4 L'analyse par les ondelettes :

##### ➤ Transformée en Ondelette :

La transformée en Ondelette est un opérateur dans lequel on remplace la sinusoïde de la transformée de Fourier par une famille de translations et dilatations d'une même fonction dite Ondelette.

L'analyse d'un signal par l'Ondelette est réalisée par une fonction analysante spécifique  $\Psi$  appelée '*Ondelette mère*'. Le nom d'Ondelette vient du fait que c'est une fonction oscillante (comme une onde). Cette dernière est positionnée dans le domaine temporel afin de sélectionner la partie du signal à traiter. Elle peut être contractée ou dilatée afin de se concentrer sur une gamme de fréquence et pour couvrir tous les domaines de définition du signal à traiter.

Le but principal recherché à travers l'analyse en ondelettes est de faire varier la fenêtre d'analyse en fonction des variations du signal à étudier. La fonction ondelette est un autre outil de traitement du signal, c'est une fonction localisée en temps et en fréquence.

Une fonction ondelette générée, par dilatation (ou contraction) et translation (selon l'axe temporel), génère une famille de fonctions ondelettes qui est appelée la fonction ondelette d'analyse ou l'ondelette mère notée «  $\Psi$  » (figure II-10). Cet ensemble contient toutes les versions translatées (par  $b$ ) et dilatées (par  $a$ ) de la fonction d'analyse.

L'ondelette mère peut engendrer une famille d'ondelette qui s'écrit sous forme l'équation suivante : avec  $a > 0$  et  $b$  réel.

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \dots\dots\dots\text{II-14 [3]}$$

Où «  $a$  » traduit changement d'échelle et «  $b$  » la translation dans le temps.

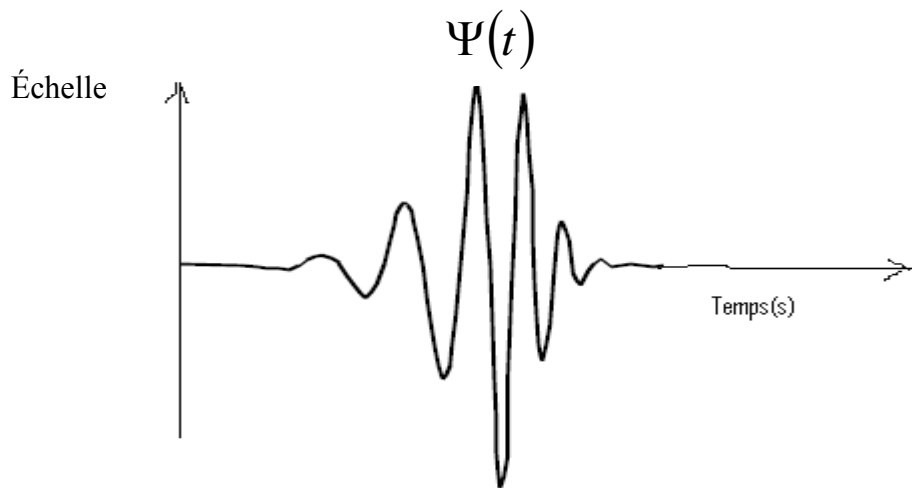
La fonction à deux variables  $S(a,b)$  est une représentation temps - échelle appelée « *transformation en ondelettes* ».

$$S(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \cdot \Psi^*\left(\frac{t-b}{a}\right) dt \dots\dots\dots\text{II-15 [3]}$$

Où :  $\Psi^*$  représente le complexe conjugué de  $\Psi$ .

La transformée en ondelette est un outil qui nous offre la possibilité d'avoir une précision en temps lorsqu'on analyse des hautes fréquences et une précision en fréquence quant on a une information basse fréquence.

Il existe plusieurs ondelettes offrant une grande diversité d'analyse. Il existe également diverses transformées en ondelette (continue ou discrètes).

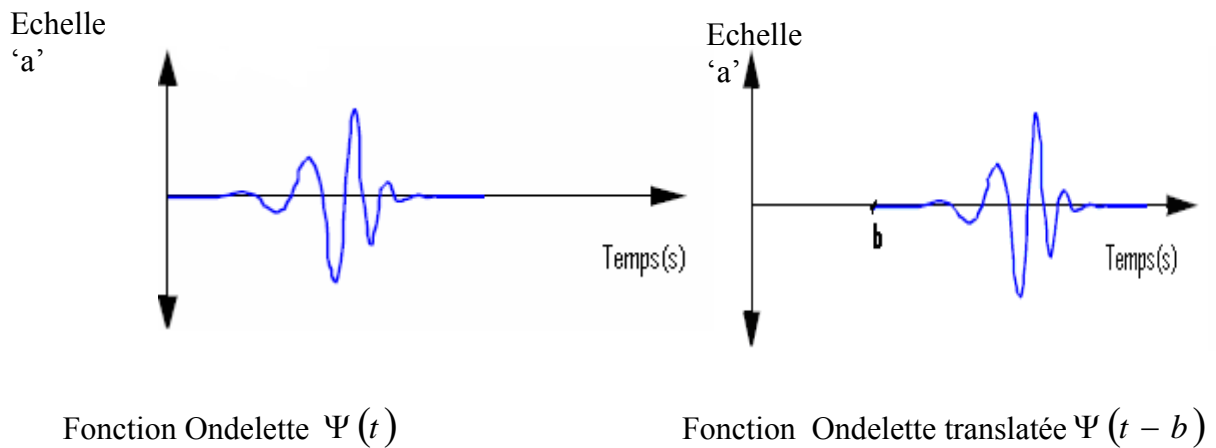


**Figure II-10 :** Ondelette mère.

➤ **La translation de l'Ondelette :**

Le terme translation est employée dans le même sens qu'elle a été employé dans la TFCT; il est lié à l'endroit de la fenêtre, car la fenêtre est décalée sur le signal. Ce terme, évidemment, correspond à l'information de temps dans le domaine de transformation.

Exemple d'une Ondelette translatée est donnée en figureII-11-a.



**figureII-11-a :** Translation d'une Ondelette.

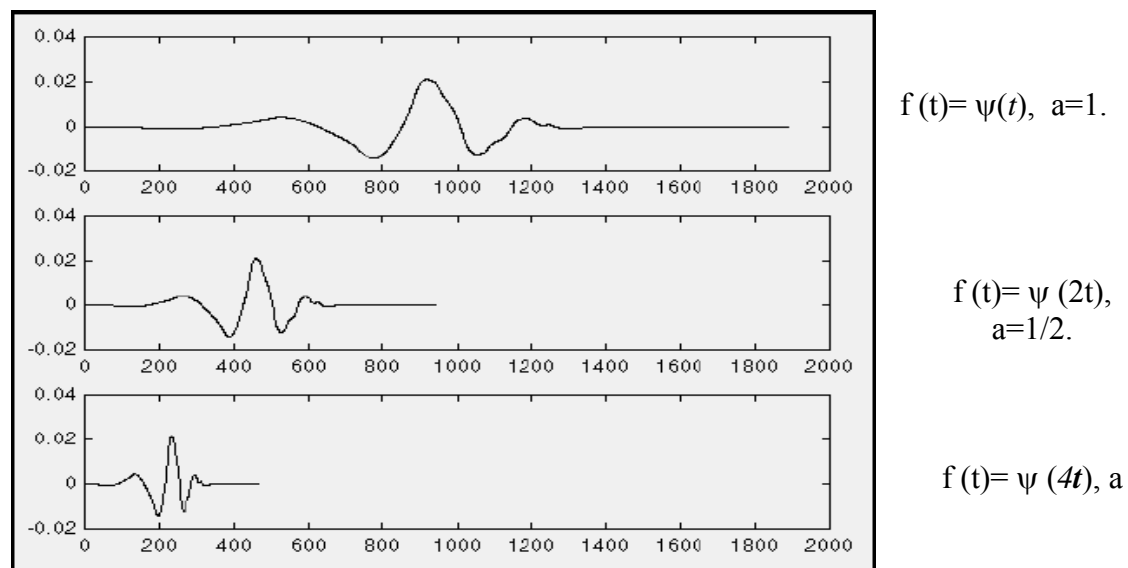
➤ **La dilatation de l'ondelette :**

Nous n'avons pas un paramètre de fréquence, comme nous l'avons eu avant pour la TFCT. Au lieu de cela, nous avons le paramètre d'échelle 'a' (scale) qui est défini comme 1/fréquence. Le terme fréquence est réservé à la TFCT.

L'échelle, comme une opération mathématique, *dilate* ou *comprime* un signal.

Grandes échelles correspondent aux signaux dilatés et les petites échelles correspondent aux signaux comprimés.

Exemple d'une dilatation d'une Ondelette est donné en figure II-11-b.



**Figure II-11-b :** Dilatation de l'Ondelette

**Conclusion :**

Nous avons brièvement parcouru dans ce chapitre les méthodes utilisées pour extraire les paramètres acoustiques qui seront fournis au système de reconnaissance. Nous n'avons fait que survoler ces techniques de calcul des coefficients représentatifs du signal vocal. Ces derniers seront fournis au système de reconnaissance du locuteur, c'est le but du chapitre suivant.

Les reconnaissances ont pour mission de décoder l'information portée par le signal vocal à partir des données fournies par l'analyse. On distingue fondamentalement deux types de reconnaissances en fonction de l'information qu'on cherche à extraire du signal vocal :

➤ La reconnaissance automatique de la parole (RAP) :

On s'intéresse au message linguistique, son objectif est de reconnaître ce qui est dit.

➤ La reconnaissance automatique du locuteur (RAL) :

Son objectif est de reconnaître l'identité d'une personne à l'aide de sa voix, autrement dit, reconnaître la personne qui parle. La variabilité de la parole entre locuteur (variabilité interlocuteur) est l'essence même de la RAL. Sans cette variabilité, il serait impossible de reconnaître une voix parmi plusieurs voix possibles.

La reconnaissance automatique du locuteur, contrairement à la reconnaissance automatique de la parole s'intéresse tout particulièrement aux informations extra-linguistiques véhiculées par un signal vocal. Pourtant, la RAL a très souvent bénéficié des avancées de la RAP. Ainsi de nombreuses techniques ont été appliquées en RAP avant d'être adaptées au domaine de la RAL.

Dans ce chapitre on s'intéresse à la reconnaissance automatique du locuteur dont on va décrire les divers systèmes, les phases et les tâches de la reconnaissance. Puis les limites et les performances d'un système de reconnaissance du locuteur, ainsi que ses domaines d'application.

### **III-1- Les différentes tâches de la reconnaissance du locuteur :**

Les tâches courantes en reconnaissance automatique du locuteur peuvent être classées en deux grandes catégories : l'identification et la vérification

#### **III-1-1- Les tâches relevant de l'identification du locuteur :**

Ces tâches consistent à rechercher, parmi un ensemble de locuteur connu, le locuteur possédant la référence la plus proche d'un message vocal donné. Deux modes sont proposés en l'identification automatique du locuteur :

- L'identification en ensemble fermé : pour lequel le locuteur est identifié parmi un nombre connu de locuteurs.
- L'identification en ensemble ouvert : pour lequel le locuteur à identifier n'appartient pas forcément à cet ensemble (un locuteur inconnu a pu prononcer le message).

### **III-1-2- Les tâches relevant de la vérification :**

Ces tâches consistent, après que le locuteur a présenté son identité, à vérifier l'adéquation de son message vocal avec la référence acoustique du locuteur qu'il prétend être. C'est une décision en tout ou rien (acceptation ou rejet).

### **III-2- Le mode d'élocution :**

Comme dans la reconnaissance automatique de la parole dont le mode d'élocution caractérise la façon dont on peut parler au système (mots isolés, mots connectés, paroles continues), la reconnaissance automatique du locuteur a aussi plusieurs modes d'élocution.

**III-2-1- Mode indépendant du texte :** la reconnaissance du locuteur est indépendante du contenu de la phrase prononcée.

**III-2-2- Mode dépendant du texte :** la phrase à prononcer pour être reconnue est fixée dès la conception du système c'est-à-dire le locuteur prononce un mot ou une phrase clef.

Les niveaux de dépendance au texte sont classés suivant les applications :

- **Système à texte libre :** le locuteur est libre de prononcer ce qu'il veut. Dans ce mode, les phrases d'apprentissages et de test sont différentes.

- **Système à texte suggéré (imposé par le système) :** un texte différent à chaque session et pour chaque personne, est imposé au locuteur et déterminé par la machine. Les phrases d'apprentissage et de test peuvent être différentes.

- **Système dépendant du vocabulaire :** Le locuteur prononce une séquence de mots issus d'un vocabulaire limité. Dans ce mode, l'apprentissage et le test sont réalisés sur des textes constitués à partir du même vocabulaire.

- **Système personnalisé dépendant du texte (système à mot de passe individuel fixe) :** Chaque locuteur a son propre mot de passe. Dans ce mode, l'apprentissage et le test sont réalisés sur le même texte.

### III-3- Les divers systèmes de reconnaissance :

Un système de reconnaissance peut être utilisé sous plusieurs modes :

#### III-3-1- Système dépendant du locuteur (monolocuteur) :

Dans ce cas, le système de reconnaissance est configuré par un locuteur spécifique.

Ce mode de fonctionnement est principalement utilisé pour de petits vocabulaires car il est demandé à l'utilisateur de prononcer tous les mots du vocabulaire une ou plusieurs fois, afin de les apprendre et de s'adapter à sa prononciation. Ce mode bien que très contraignant (difficile) a les meilleures performances. Il faut cependant noter qu'il est souvent demandé de prononcer les mots plusieurs fois, car il est impossible de prononcer un mot deux fois exactement de la même façon, il existe toujours des variations dans le débit d'élocution, l'accentuation ou l'intonation du mot prononcé. Il existe deux méthodes permettant une reconnaissance monolocuteur:

##### a. La méthode analytique :

Qui est basée sur la reconnaissance de syllabes, et essai de recomposer les mots en fonction de l'enchaînement de celle-ci.

##### b. La méthode globale : Permettant de reconnaître le mot dans son ensemble.

**III-3-2- Système multilocuteur :** le système de reconnaissance est élaboré pour un groupe restreint de personnes.

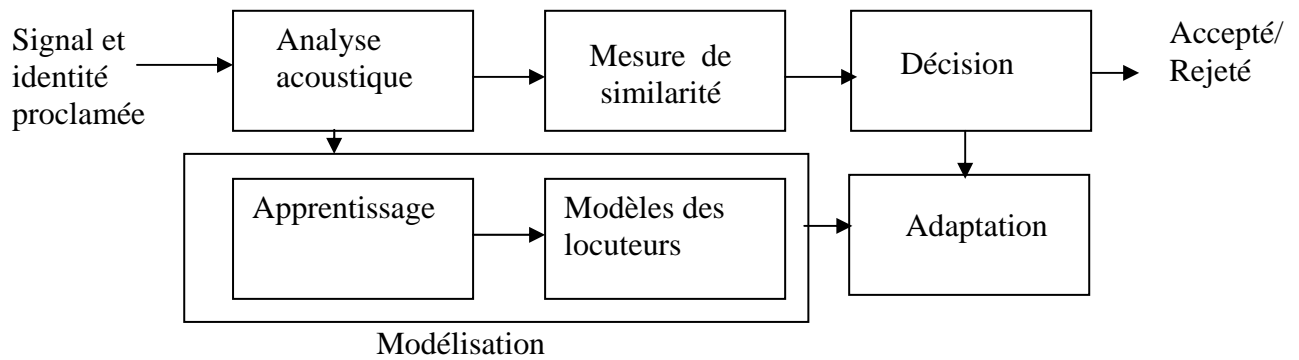
**III-3-3- Système indépendant du locuteur :** dans ce cas le système de reconnaissance est capable de reconnaître n'importe qui sans apprentissage préalable, tout locuteur peut utiliser le système de reconnaissance.

### III-4- La structure du système de reconnaissance automatique du locuteur (RAL) :

La RAL peut être interprétée comme une tâche particulière de reconnaissance de formes (reconnaissance de texte, reconnaissance de la langue, reconnaissance des émotions et reconnaissance du locuteur).

C'est une succession de modules dont l'étape finale est de reconnaître une forme particulière, c'est-à-dire le signal de parole que l'on met à l'entrée de cette chaîne.

Un système automatique de reconnaissance vocale, se divise généralement en quatre modules qui sont présentés par la figure III-2.



**FigureIII-1** : schéma typique d'un système de reconnaissance du locuteur.

#### III-4-1- Paramétrisation :

Un premier module de traitement du signal réalise l'analyse acoustique du signal de parole sur des fenêtres temporelles de courte durée, de 10 à 30 ms. A l'issue de cette étape, le signal est représenté par des vecteurs acoustiques de paramètres qui permettent de discriminer (distinguer) différents locuteurs. Idéalement, les paramètres (ou traits acoustiques) doivent être fréquents, facilement mesurables, peu sensibles à la variabilité intra-locuteur, non affectés par le bruit ambiant ou les variations dues au canal de transmission.

En pratique, il est très difficile de réunir tous ces attributs en même temps. Il ressort également que les seuls types de paramètres vraiment pertinents et utilisables efficacement, sont les paramètres de l'analyse spectrale et éventuellement les paramètres prosodiques qui réunissent l'énergie, la durée et la fréquence fondamentale.

#### III-4-2- Apprentissage et modélisation des locuteurs :

Dans le cadre de la reconnaissance vocale (du locuteur), nous modélisons les prononciations des différents locuteurs pour le même message vocal. En étudiant la parole des locuteurs sur des prononciations de même vocabulaire, nous pouvons distinguer des variabilités caractéristiques du signal de parole nous permettant de séparer les locuteurs les uns des autres (variabilités inter-locuteurs) et d'autre intrinsèques au locuteur (variabilité intra-locuteur).

De manière à modéliser les caractéristiques qui dépendent du locuteur, nous utilisons des algorithmes (techniques) capable de capturer les point qui différent entre les représentations spectraux et temporelles, issus des différents locuteurs, constituant ainsi des modèles ou des références du locuteur.

Il nous faut des mesures qui permettront de donner une valeur de similitude ou de distance entre le modèle des locuteurs et un motif inconnu dont on cherche à déterminer l'origine. Nous proposons ici un aperçu des approches qui fournissent les meilleurs résultats en reconnaissance vocale.

- a. L'approche vectorielle :** Le locuteur est représenté par un ensemble de vecteur de paramètres dans l'espace acoustique. Ses principales techniques sont la reconnaissance à base de DTW (reconnaissance par déformation temporelle dynamique, en anglais Dynamique Time Warping) et par quantification vectorielle (QV).
- b. L'approche statistique :** Consiste à représenter chaque locuteur par une densité de probabilité dans l'espace des paramètres acoustiques .Elle couvre les techniques de modélisation par les modèles de Markov caché par les mélanges de gaussiennes (MMG).
- c. L'approche connexionniste :** Consiste principalement à modéliser les locuteurs par des réseaux de neurones.
- d. L'approche relative :** il s'agit de modéliser un locuteur relativement par rapport à d'autres locuteurs de référence dont les modèles sont bien appris.

**III-4-3- Mesure de similarité :** Un modèle de calcul de ressemblance est utilisé durant la reconnaissance, il a en charge de mesurer la similarité entre les paramètres acoustique de signal prononcé et une référence correspondant à une personne donnée.

La stratégie mise en jeu dans cette partie de système de reconnaissance vocale dépend essentiellement du type d'application : identification ou vérification.

Dans une application de vérification du locuteur, la procédure consiste à fixer un seuil sur la mesure de similarité : en dessous, le locuteur est rejeté, au dessus, le locuteur est accepté.

Dans une application de l'identification du locuteur, la stratégie est assez simple puisqu'il s'agit d'évaluer la similarité des caractéristiques mesurées avec toutes les références correspondantes à chacun des locuteurs autorisés. Le locuteur identifié est celui pour lequel la similarité est la plus grande.

#### **III-4-4-Module de décision**

Le module de décision, fournit la réponse du système. On peut également introduire un module d'adaptation pour augmenter les performances du système de reconnaissance.

#### **III-5 - Les limites de la RAL (difficultés) :**

La reconnaissance automatique du locuteur rencontre plusieurs obstacles qui la rendent très limitée en pratique, nous citons ici ceux qui sont les plus évident :

- L'environnement au moment de l'acquisition du signal notamment la présence ou non de bruits.
- La qualité des équipements utilisés tels que le canal de transmission de l'information, le microphone et la carte son.
- Toute altération de l'appareil phonatoire peut modifier la qualité de l'émission c'est-à-dire la voix évolue avec l'âge, avec l'état physiologique ou pathologique du locuteur (Maladie : rhume ; émotion : stress, joie).
- Les bruits de bouche, les hésitations, les mots hors vocabulaire sont autant d'interférences supplémentaires sur le signal de parole.
- La nature très variable du signal de parole est un facteur délicat (sensible) à gérer.
- Facteurs socioculturels ; le locuteur peut changer l'accent, l'intonation qui diffèrent selon le sexe, l'origine sociale, régionale ou nationale, ces facteurs influencent la forme et le contenu du message.
- Facteurs subjectifs telles que la prononciation, la vitesse d'élocution.
- La quantité de données de reconnaissance (durée des énoncés)

Malgré ces difficultés, on peut dire d'une manière générale que l'amélioration des techniques utilisées permet d'envisager des conditions de fonctionnement de plus en plus proche des situations rencontrées en pratique.

#### **III-6- Evaluation des performances en reconnaissance du locuteur :**

Le domaine de la reconnaissance automatique de la parole fait preuve d'un grand dynamisme, au niveau de la recherche du développement et de l'évolution des performances. Les chercheurs actuels sont concentrés sur les limitations pratiques de la reconnaissance

automatique du locuteur et sur la présentation, la diffusion et l'interprétation des résultats des systèmes.

Et pour garantir un niveau de performance acceptable pour les applications de la reconnaissance du locuteur, plusieurs caractéristiques sont généralement nécessaires :

- Environnement calme et microphone de bonne qualité.
- Les conditions d'enregistrement et du traitement du signal audio sont connues et contrôlées.
- Des données de paroles, enregistrées dans des mêmes conditions que le signal de test sont disponibles pour référencer un locuteur dans le système.
- L'usage des techniques sophistiquées pour modifier ou déguiser la voix, doit être interdit aux locuteurs.
- La mesure de ressemblance est étalonnée au cours d'expériences réalisées dans les conditions contrôlées citées précédemment.
- Les locuteurs souhaitent être acceptés par le système, ils coopèrent avec celui-ci.
- L'usage d'un système de synthèse de la parole n'est pas autorisé.

### **III-7- Domaines d'applications de la reconnaissance du locuteur :**

Les applications essentielles des systèmes de la reconnaissance du locuteur incluent :

#### **➤ Applications sur site :**

La personne doit être physiquement présentée en un lieu précis.

- Serrures vocales (pour des locaux, un compte informatique ....)
- Interactivité matérielle (retrait d'argent à un guichet automatique,...).

#### **➤ Applications liées aux télécommunications :**

La vérification s'opère à distance :

- Accès à des services pour des abonnés, ou des données confidentielles.
- Transaction à distance

#### **➤ Applications commerciales :**

- Associer un même mot de passe pour une petite population de locuteur (membre d'une famille, d'une société).
- Protection de matériel contre le vol.

➤ **Applications judiciaires :**

- Recherche de suspects, d'éléments de preuve, de preuve.
- Les juges, les avocats, les enquêteurs de police ou de la gendarmerie souhaitent utiliser des procédés de reconnaissance vocale pour mener une enquête ou confirmer un verdict de culpabilité ou d'innocence.

➤ **Applications stratégiques :**

- Ecoutes téléphoniques.
  - Protection de la démocratie.
  - Intrusion dans la vie privée.

**III-8- Principe de la méthode du 'taux de passage par zéro' pour la RAL :**

Ce principe repose sur le calcul du taux de passage par zéro de chaque signal des différents locuteurs des références et de test, ensuite comparer les résultats du test avec ceux des références afin de reconnaître le locuteur désiré. En basant sur l'équation suivante on va faire les programmes nécessaires pour cette tâche.

$$TPZ(x) = \frac{1}{N} \sum_{n=1}^{N-1} a_n \quad \text{avec} \quad a_n = 1 \quad \text{si} \quad x_{n+1} \cdot x_n < 0 \dots\dots\dots \text{III-1}$$

$$a_n = 0 \quad \text{sinon}$$

$n$  : Est le nombre d'échantillon du signal  $x(t)$ .

$N$  : Est la longueur de la fenêtre.

$x_n$  : Est l'amplitude du signal à l'instant  $n$  et  $x_{n+1}$  : est l'amplitude du signal à l'instant  $(n+1)$ .

$a_n$  : Est un coefficient.

**Conclusion :**

Ce chapitre qui porte un aperçu sur la reconnaissance automatique du locuteur, a dégagé les différents systèmes de RAL ainsi ses phases. Cette étude nous a montré les limites et les différents problèmes rencontrés dans la pratique. Le chapitre suivant, nous permettra de compléter et de mieux comprendre l'aspect pratique de la méthode du TPZ pour la reconnaissance du locuteur.

Après avoir expliqué dans les chapitres précédents, toutes les notions théoriques nécessaires à notre travail, nous allons maintenant franchir l'aspect pratique par une implémentation de la méthode du taux de passage par zéro (TPZ), pour la reconnaissance automatique du locuteur (RAL) en mode dépendant du texte (les locuteurs prononcent le même message), afin d'évaluer sa validité par une analyse des résultats obtenus.

#### **IV-1- Ressources :**

##### **IV-1-1-Matériel utilisé:**

Pour la mise en œuvre de notre travail, nous avons utilisé :

- **Un micro-ordinateur portable.**
- **Un microphone analogique (casque) :**

Il suffit de raccorder le microphone à l'entrée « Mic » de la carte son située à l'arrière de l'ordinateur à l'aide des fiches numériques, qui permettent la conservation fidèle du son numérique lors de son transfert d'un appareil à l'autre.

- **Des logiciels de traitement du son et d'enregistrement d'ondes sonores :**

Ils devraient offrir des fonctions de traitement audio et d'édition. Il existe plusieurs logiciels du traitement du son, nous citons : Gold Wave ; Praat ; Cool Edit ; Sound Forge, ils supportent les différents formats (mono et stéréo) de son, ils permettent le traitement et l'analyse de ce dernier, l'édition des formes d'ondes, et un choix varié de fréquence d'échantillonnage.

- **Des locuteurs :**

On possède de huit ( 8) locuteurs, chacun prononce la même phrase « *pouvoir c'est vouloir* » à la même fréquence d'échantillonnage qui est égale à 44.1 KHz.

Les échantillons de référence sont :

*Mounia, Kahina, Samira, Ammina, Naiima, Mohmed, Djamel, Azzize.*

Les échantillons de test sont :

*Mounia, Kahina, Samira, Ammina, Naiima, Mohmed, Djamel, Azzize, Ammine, Chafia, Djahid, Moulod.*

##### **IV-1-2-Le langage de programmation :**

Le langage utilisé est Matlab (Matrix Laboratory) qui est un logiciel interactif, développé par Math Works, destiné notamment au traitement numérique des données. Il est

particulièrement efficace lorsque celles-ci sont présentées sous la forme de vecteurs ou de matrices. Ce logiciel possédant une puissance et une flexibilité concernant la gestion des données manipulées par le programme, en plus de ça, il est simple à utiliser, précis, robuste et rapide.

#### **IV-2-Les conditions d'enregistrement :**

Avant d'expliquer les étapes d'enregistrement et les différentes phases de reconnaissance, on donne les conditions dont lesquelles sont effectuées les enregistrements :

- L'identification du locuteur se fait en ensemble fermé (nombre du locuteur limité).
- La nature du message vocal manipulé par ce système de reconnaissance est en mode dépendant du texte c'est à dire les personnes prononcent la même phrase.
- Les enregistrements des signaux se font dans une salle calme, afin d'éviter le maximum possible le bruit.
- Les conditions d'enregistrement des signaux de références seront les mêmes que celles des signaux de test.

On passe maintenant à la procédure suivie pour l'enregistrement, l'édition et le traitement du son.

#### **IV-3- Les étapes d'enregistrement sur support informatique des sons des locuteurs :**

On utilisant le logiciel du traitement de son « Praat » on passera par les étapes suivantes :

##### **Etape 1 :**

Pour débiter un nouvel enregistrement, il faut cliquer sur 'New' et choisir les formats de qualité et de durée.

Les formats : mono ou stéréo ; tel que les sons enregistrés en stéréo ont un volume deux fois plus élevé qu'en mono.

Qualité : 44100Hz.c'est la fréquence d'échantillonnage.

##### **Etape 2 :** commencer l'enregistrement.

A l'aide de la fenêtre de contrôle (se trouvant dans Tool), commencez l'enregistrement en cliquant sur le cercle rouge. Le locuteur prononce la phrase, et dès qu'il fini on clique sur 'stop'.

##### **Etape (3) :** Traitement et l'édition de son.

Editer le son à l'aide de la fenêtre 'Edit'. Choisir les traitements désirés à l'aide de la fenêtre 'Effects'.

**Etape (4) : Sauvegarder le son.**

On choisit le format de sauvegarde sous la fenêtre 'File'. De préférence enregistrer le fichier en format 'Wave' (format sous Windows).

La même procédure est effectuée pour l'enregistrement de chaque locuteur de références et de tests.

#### **IV-4- Fonctionnement du système de reconnaissance du locuteur :**

La reconnaissance du locuteur se déroule en deux phases :

1. La phase d'apprentissage.
2. La phase de reconnaissance.

##### **IV-4-1-La phase d'apprentissage :**

La phase d'apprentissage est une procédure de création et de mémorisation sur un disque ou sur la mémoire centrale d'un ordinateur, d'un vocabulaire de référence sous forme de paramètres extrait du signal de parole, utile à la comparaison du mot inconnu (test) avec le mot de référence pour l'identification de locuteur lors de phase de reconnaissance.

Les références 'xi' des différents locuteurs prononcent la phrase de l'application prévue « *pouvoir c'est vouloir* ».

Une fois que les sons sont enregistrés sous le format '*nom du fichier.wav*', on va effectuer l'opération qui permet de charger les signaux sous « *Matlab* ».

##### **➤ Chargement des signaux sous Matlab :**

On va copier tous les signaux enregistrés vers le « work » de '*Matlab*', ensuite pour analyser un signal sonore, on commence par importer le fichier (*son .wav*) à l'aide de la commande *wavread : x=wavread ('son.wav')*.

Ces sons sont tous prétraités puis conservés dans un dictionnaire de référence en tant que image acoustique.

Ensuite, analyser les différents signaux et appliquer à chaque signal l'opération de filtrage (débruitage) dans ce cas on utilise les '*ondelettes*'. Suivie d'une opération de fenêtrage en

découpant chaque signal en ' $m$ ' trames de 30ms (recouvrement de 10ms). La fenêtre utilisée est celle de Hamming tel que  $N=1000$  points.

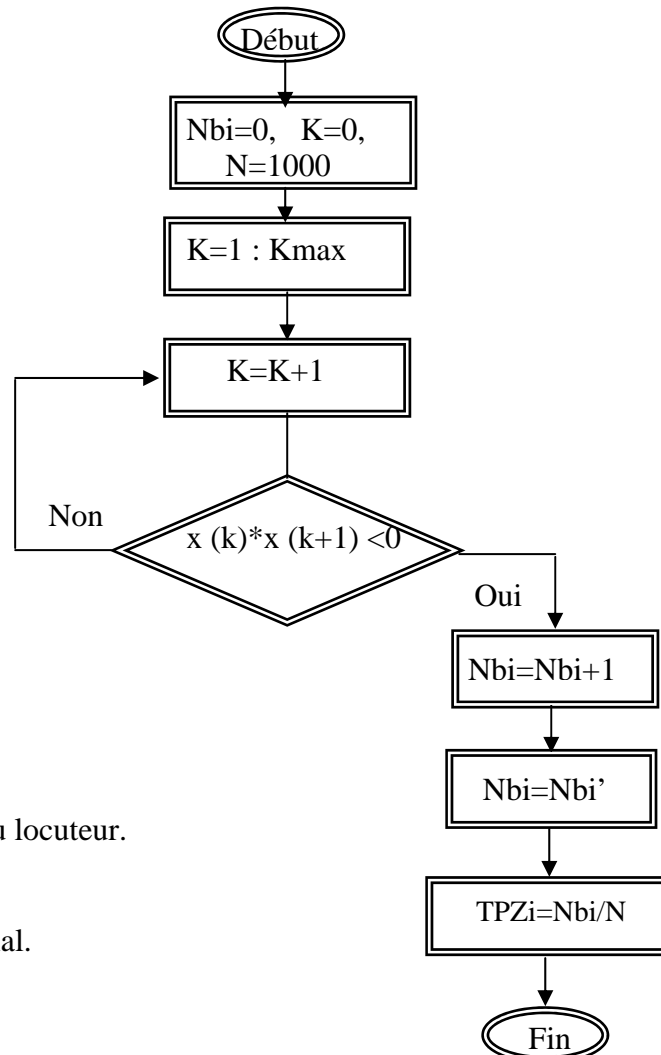
Après on va extraire des caractéristiques du locuteur, dans notre cas c'est le nombre de passage par zéro de chaque son du locuteur 'Nbi' ou bien le TPZ qui est égale à  $Nbi/N$ .

Enfin on va stocker les données dans un dictionnaire de référence.

$x_i = \text{Nom\_référence} = [\text{Mounia}, \text{Kahina}, \text{Samira}, \text{Ammina}, \text{Naiima}, \text{Mohmed}, \text{Djamel}, \text{Azzize}]$   
 $= [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]$ .

➤ **L'organigramme du programme qui calcul le taux de passage par zéro :**

Le programme qui calcul le taux de passage par zéro de chaque signal est donné sur l'organigramme suivant.



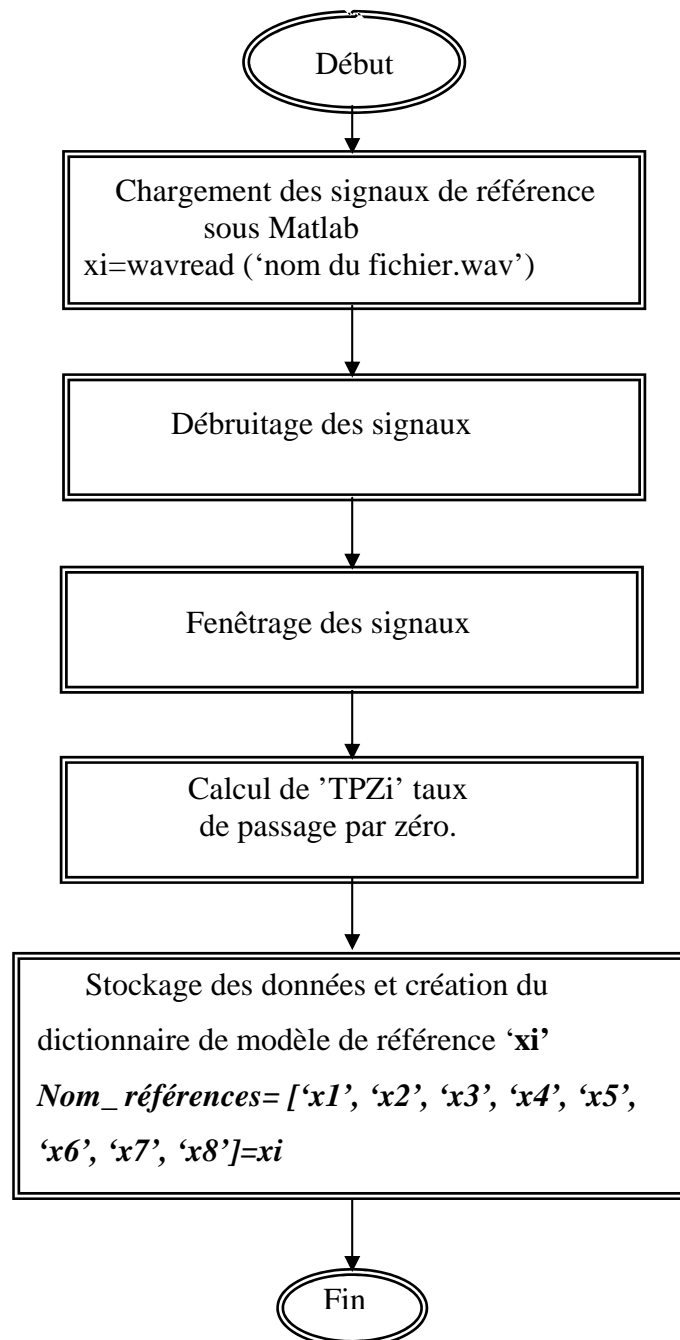
$i=1 : 10$  représente l'indice du locuteur.

$K$  : nombre de points du signal.

**Figure IV-1 :** L'organigramme du programme qui calcul le taux de passage par zéro.

➤ **L'organigramme du programme de la phase d'apprentissage :**

Le schéma suivant illustre les étapes suivies lors de la phase d'apprentissage.



**Figure IV-2 :** L'organigramme du programme de la phase d'apprentissage.

**IV-4-2-La phase de reconnaissance (phase de test) :**

On passe maintenant à la phase de test où le signal à reconnaître 'y' subit le même prétraitement que la phase précédente.

Le signal test sera soit parmi l'ensemble des signaux de référence soit un étrangère.

Les enregistrements de la nouvelle acquisition sont :

[*Mounia, Kahina, Samira, Ammina, Naiima, Mohmed, Djamel, Azzize, Ammine, Chafia, Djahid, Moulod*].

**➤ Mesure de ressemblance :**

La mesure de ressemblance (similarité) s'effectue par la comparaison entre le locuteur de la nouvelle acquisition (le signal test) et les locuteurs du dictionnaire de références.

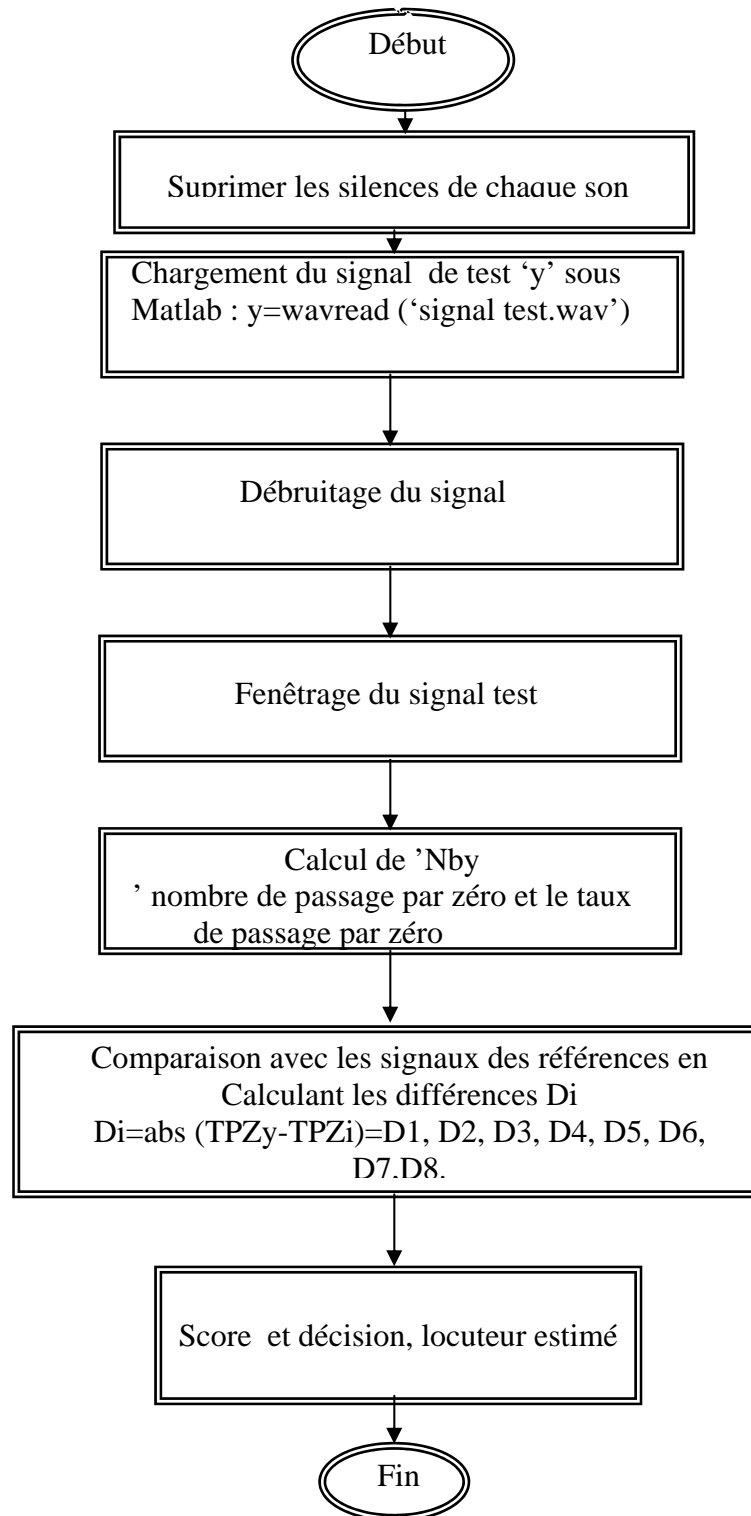
Cette comparaison se fait par le calcul de la différence 'Di ' entre le taux de passage par zéro 'TPZy' du locuteur test et les taux de passage par zéro des diverses références 'TPZ i'.

Le programme va comparer le locuteur qui a parlé (le signal test) avec ceux qui sont en mémoire depuis la phase d'apprentissage.

Une fois que les différences 'Di' sont obtenues, il suffit de rechercher la plus petite valeur parmi celles-ci, c'est-à-dire min (Di), et de donner le résultat du locuteur correspondant.

➤ **L'organigramme du programme de la phase de test :**

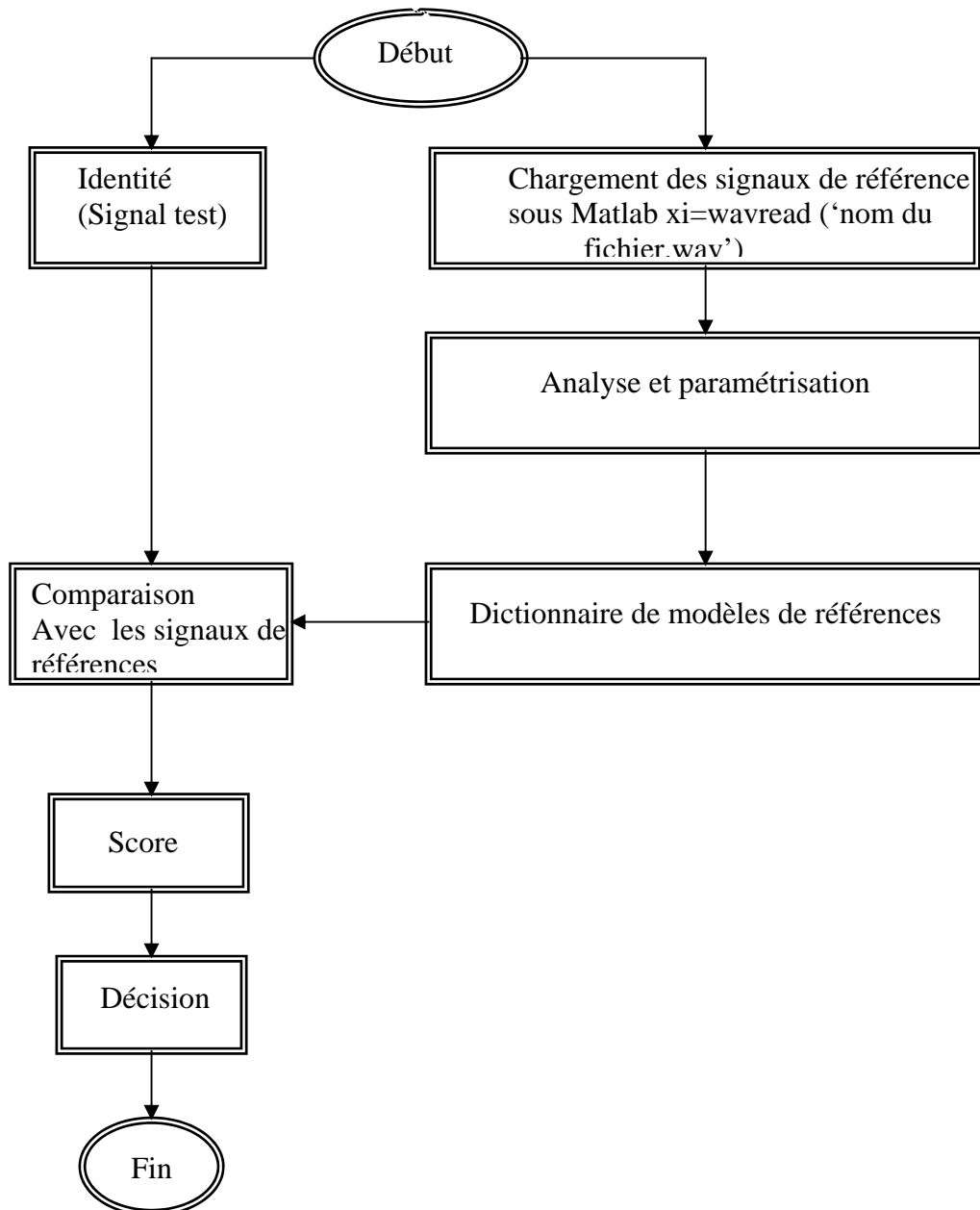
Le schéma de la figure IV-3 illustre l'organigramme du programme de la phase de test.



**Figure IV-3 :** L'organigramme du programme de la phase de test.

**IV-4-3-L'organigramme général:**

Le schéma suivant illustre l'organigramme général du programme de la reconnaissance du locuteur.



**Figure IV-4 :** L'organigramme général du programme de la reconnaissance du locuteur.

**IV-5- Interprétation des résultats :**

Après avoir donné les différents programmes concernant le système de reconnaissance automatique du locuteur ainsi leurs applications, on a obtenu les résultats suivants :

- **Les résultats de la phase d'apprentissage :**

On a disposé de huit (8) locuteurs associés à des signaux  $x_i$  telle que  $i=1\dots 8$ , où chacun prononce le même message « *pouvoir c'est vouloir* ». Les signaux de référence  $x_1, x_2\dots x_8$ , après leurs enregistrement et leurs chargement, le « workspace » qui se trouve dans la fenêtre « view

» de Matlab nous a permet d'obtenir le nombre d'échantillon ' $k_i$ ' de chaque signal, ainsi en divisant ce dernier sur la longueur de la fenêtre de Hamming  $N=1000$ , après fenêtrage on obtiendra le nombre de trames ' $m_i$ ' de chaque signal, ensuite on a effectué l'opération de débruitage par les ondelettes sur les différents signaux et cela à l'aide d'une simple instruction « *dwt* ». Tous les signaux sont alors prés à les analysés afin d'extraire les caractéristiques dont on a besoin tels que le nombre de passage par zéro ' $N_{bi}$ ' et le taux de passage par zéro ' $TPZ_i = N_{bi}/N$ '.

Le système stocke ces paramètres ainsi l'identité correspondante dans une base de références.

Le tableau n°1 récapitule les résultats de la phase d'apprentissage.

Les signaux de référence ' $x_i$ '	Le nombre d'échantillons du locuteur ' $k_i$ '	Le nombre de trames après le fenêtrage ' $m_i$ '	Le nombre de passage par zéro de chaque signal ' $Nb_i$ '	Le taux de passage par zéro $TPZ_i=Nb_i/N$ , $N=1000$
$x_1='Mounia'$	$k_1=84672$	$m_1=84$	$Nb_1=266$	$TPZ_1= 0.266$
$x_2='Kahina'$	$k_2=85113$	$m_2=85$	$Nb_2=250$	$TPZ_2= 0.250$
$x_3='Samira'$	$k_3=77616$	$m_3=77$	$Nb_3=275$	$TPZ_3= 0.275$
$x_4='Ammina'$	$k_4=103194$	$m_4=103$	$Nb_4=246$	$TPZ_4= 0.246$
$x_5='Naiima'$	$k_5=64827$	$m_5=64$	$Nb_5=221$	$TPZ_5=0.221$
$x_6='Mohmed'$	$k_6=88641$	$m_6=88$	$Nb_6=278$	$TPZ_6= 0.278$
$x_7='Djamel'$	$k_7=79380$	$m_7=79$	$Nb_7=231$	$TPZ_7=0.231$
$x_8='Azzize'$	$k_8=91987$	$m_8=91$	$Nb_8=197$	$TPZ_8=0.197$

**Tableau n°1 : Résultats de la phase d'apprentissage.**

- **Les résultats de la phase de test :**

Dans cette phase le locuteur se présente devant le système et produit un énoncé de test qui sera le même énoncé que celui de référence, puisque notre système de reconnaissance est en mode dépendant du texte.

Les signaux de la nouvelle acquisition (test) ont subi la même procédure que celle effectuée pour les signaux de références c'est-à-dire on a précisé la longueur de chaque signal ' $k_i$ ', le nombre de trames ' $m_i$ ', le nombre de passage par zéro du signal test ' $Nb_{y_i}$ ' et le taux de passage par zéro ' $TPZ_{y_i}$ '. Les signaux de test peuvent être soit des signaux appartenants aux signaux de références, soit des étrangères, on prend les huit (8) signaux de référence et quatre (4) signaux étrangères ( $x_{test}=x_1\dots\dots\dots x_{12}$ ).

Le tableau n°2 récapitule les résultats de la phase de test.

Les signaux de test 'y'	Le nombre d'échantillons du locuteur 'ki'	Le nombre de trames après le fenêtrage 'mi'	Le nombre de passage par zéro de chaque locuteur 'Nby <sub>i</sub> '	Le taux de passage par zéro TPZ <sub>y</sub> =Nby/N N=1000
$x_1='Mounia'$	77175	77	Nby <sub>1</sub> = 260	TPZ <sub>y1</sub> = 0.260
$x_2='Kahina'$	90405	90	Nby <sub>2</sub> = 256	TPZ <sub>y2</sub> = 0.256
$x_3='Samira'$	68796	68	Nby <sub>3</sub> = 276	TPZ <sub>y3</sub> = 0.276
$x_4='Ammina'$	99666	99	Nby <sub>4</sub> = 241	TPZ <sub>y4</sub> = 0.241
$x_5='Naiima'$	63504	63	Nby <sub>5</sub> = 224	TPZ <sub>y5</sub> = 0.224
$x_6='Mohmed'$	95256	95	Nby <sub>6</sub> = 282	TPZ <sub>y6</sub> = 0.282
$x_7='Djamel'$	78498	78	Nby <sub>7</sub> = 230	TPZ <sub>y7</sub> = 0.230
$x_8='Azzize'$	92610	92	Nby <sub>8</sub> =171	TPZ <sub>y8</sub> =0.171
$x_9='Ammine'$	88641	88	Nby <sub>9</sub> =240	TPZ <sub>y9</sub> =0.240
$x_{10}='Chafia'$	104076	104	Nby <sub>10</sub> =287	TPZ <sub>y10</sub> =0.287
$x_{11}='Djahid'$	96138	96	Nby <sub>11</sub> =247	TPZ <sub>y11</sub> =0.247
$x_{12}='Moulod'$	72324	72	Nby <sub>12</sub> =301	TPZ <sub>y12</sub> =0.301

Tableau n°2 : Résultats de la phase de test.

- **Les résultats de la comparaison et la mesure de ressemblance :**

Dans le cas de la reconnaissance en ensemble fermé, on a constaté que quelque soit le signal test pris, nous a donné des bons résultats ; le système a pu identifier le locuteur estimé.

**Exemple 1 :**

On a supposé que le signal test est le signal  $y_{test} = 'Mounia' = x_1$ , puis on a appliqué le programme de la reconnaissance, et pour vérifier l'identité de cette personne, on a comparé le TPZ<sub>y</sub> du signal test avec le TPZ<sub>i</sub> de chaque signal de références, cette

comparaison s'effectue par le calcul de la valeur absolue de la soustraction entre le TPZ du signal test et ceux des références.  $D_i = \text{abs}(\text{TPZ}_y - \text{TPZ}_i)$ . On a obtenu les différences

$D_i = [D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8]$  sous forme d'un vecteur.

Ensuite pour ressortir le locuteur correspondant, on a choisi parmi ces différences celle qui possède la plus petite valeur de  $D_i$  qui est donnée par  $m$  ( $m = \min(D_i)$ ).

### Prise de décision :

L'application du programme nous a donnée « 'indice' =1 » qui correspond à la position du locuteur sur l'ensemble de références ' $x_i$ ' = [ $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ ] = [Mounia, Kahina, Samira, Ammina, Naiima, Mohmed, Djamel, Azzize]. Le signal correspondant c'est ' $x_1$ ' et le locuteur estimé correspond au nom de l'indice de références.

Nom=Nom ref (indice, :)

Nom=Nom ref (1, :)

Nom=MOUNIA

Le tableau n°3 donne les résultats de la comparaison et la prise de décision.

Calcul des diverses soustractions $D_i = \text{abs}(\text{TPZ}_y - \text{TPZ}_i)$	Prise de décision
$D_1 = \text{abs}(\text{TPZ}_y - \text{TPZ}_1) = 0.260 - 0.266$	$D_i = \begin{pmatrix} 0.0060. \\ 0.01 \\ 0.015 \\ 0.014 \\ 0.039 \\ 0.018 \\ 0.029 \\ 0.063 \\ 0.025 \end{pmatrix}$
$D_2 = \text{abs}(\text{TPZ}_y - \text{TPZ}_2) = 0.260 - 0.250$	
$D_3 = \text{abs}(\text{TPZ}_y - \text{TPZ}_3) = 0.260 - 0.275$	
$D_4 = \text{abs}(\text{TPZ}_y - \text{TPZ}_4) = 0.260 - 0.246$	
$D_5 = \text{abs}(\text{TPZ}_y - \text{TPZ}_5) = 0.260 - 0.221$	
$D_6 = \text{abs}(\text{TPZ}_y - \text{TPZ}_6) = 0.260 - 0.278$	
$D_7 = \text{abs}(\text{TPZ}_y - \text{TPZ}_7) = 0.260 - 0.231$	
$D_8 = \text{abs}(\text{TPZ}_y - \text{TPZ}_8) = 0.260 - 0.197$	

**Tableau n°3** : résultats de la comparaison et de la décision.

Dans le cas de la reconnaissance en ensemble ouvert c'est-à-dire le signal test est un signal inconnu qui n'appartient pas aux signaux de références.

**Exemple 2** : le signal de test est  $y_{test} = \text{'Chafia'}$

L'application du programme de reconnaissance nous a donné directement le nom du test.

Nom=nom-test.

Nom=CHAFIA.

Le tableau n°4 donne les résultats de la comparaison et la prise de décision.

Calcul des diverses soustractions entre $TPZ_i$ et $ZPZ_y$ , ' $D_i$ '= $\text{abs}(TPZ_y - TPZ_i)$	Prise de décision
$D_1 = \text{abs}(TPZ_y - TPZ_1) = 0.287 - 0.266$	$D_i = \left( \begin{array}{c} 0.021 \\ 0.037 \\ 0.012 \\ 0.041 \\ 0.066 \\ 0.009 \\ 0.056 \\ 0.090 \end{array} \right)$ [m, indice]= Min (D) m=0.009 indice=6 nom=nom-test <b>Nom-test='Chafia'</b>
$D_2 = \text{abs}(TPZ_y - TPZ_2) = 0.287 - 0.250$	
$D_3 = \text{abs}(TPZ_y - TPZ_3) = 0.287 - 0.275$	
$D_4 = \text{abs}(TPZ_y - TPZ_4) = 0.287 - 0.246$	
$D_5 = \text{abs}(TPZ_y - TPZ_5) = 0.287 - 0.221$	
$D_6 = \text{abs}(TPZ_y - TPZ_6) = 0.287 - 0.278$	
$D_7 = \text{abs}(TPZ_y - TPZ_7) = 0.287 - 0.231$	
$D_8 = \text{abs}(TPZ_y - TPZ_8) = 0.287 - 0.197$	

**Tableau n°4** : résultats de la comparaison et de la décision.

**Conclusion :**

Les résultats obtenus montrent l'efficacité de la méthode du taux de passage par zéro pour la reconnaissance automatique du locuteur en mode dépendant du texte, en respectant le rythme de la prononciation et en évitant le maximum possible le bruit.

# *Conclusion générale*

Cette thèse s'inscrit dans le domaine du traitement de la parole, nous nous sommes intéressés à la reconnaissance automatique du locuteur en mode dépendant du texte et pour cela nous avons donné un aperçu sur le système de production de la parole, ses caractéristiques, son acquisition et sur son analyse par les différentes méthodes temporelles et fréquentielles.

On a choisit la méthode du taux de passage par zéro (TPZ) pour la reconnaissance automatique du locuteur dont on a expliqué ses divers systèmes et ses différentes tâches, ainsi ses applications et ses limites.

L'application de cette méthode sur le signal de parole, nous a permis d'obtenir des bons résultats c'est-à-dire reconnaître le locuteur qui parle dans le mode dépendant du texte (la phrase est fixée), mais dans des conditions d'enregistrements bien respectées et contrôlées.

Dans le cas de la reconnaissance automatique du locuteur en mode indépendant du texte (la phrase change), l'application de cette méthode ne donne pas des bons résultats, car il est impossible de prononcer un mot deux fois exactement de la même façon. Dans ce cas l'utilisation d'autres méthodes est nécessaire, parmi ces dernières il existe les modèles de Markov caché, les réseaux de neurones et la quantification vectorielle.

A ce jour, il n'y a pas de procédé scientifique qui permette de caractériser de manière unique la voix d'une personne ou d'identifier avec une certitude absolue un individu à partir de sa voix.

## I- Etude de quelques fenêtres

### I-1- Définition de la fenêtre de pondération :

Une fenêtre est une fonction régulière, lentement variable et bien localisée (ce qui signifie qu'elle est nulle en dehors d'une certaine zone)

Le rôle principal de la fenêtre est de limiter la durée d'un signal infini. La position de la fenêtre est choisie de manière à conserver les échantillons importants du signal à traiter.

Quant au choix de la forme (type), il dépend principalement de :

- La largeur du lobe principal.
- L'amplitude maximum des oscillations des lobes secondaires.
- La décroissance des lobes secondaires en dB/octave.

### I-2- Etude de quelques fenêtres :

**I-2-1- Fenêtre rectangulaire (à bande étroite) :** Elle est définie par :

$$W_R(n) = \begin{cases} 1 & 0 \leq n < N \\ 0 & n < 0 \text{ et } n \geq N \end{cases} \dots\dots\dots[9]$$

$n=0\dots\dots N-1$       $N$  : Le nombre d'échantillon par fenêtre.

Voilà la courbe représentant la fenêtre rectangulaire.

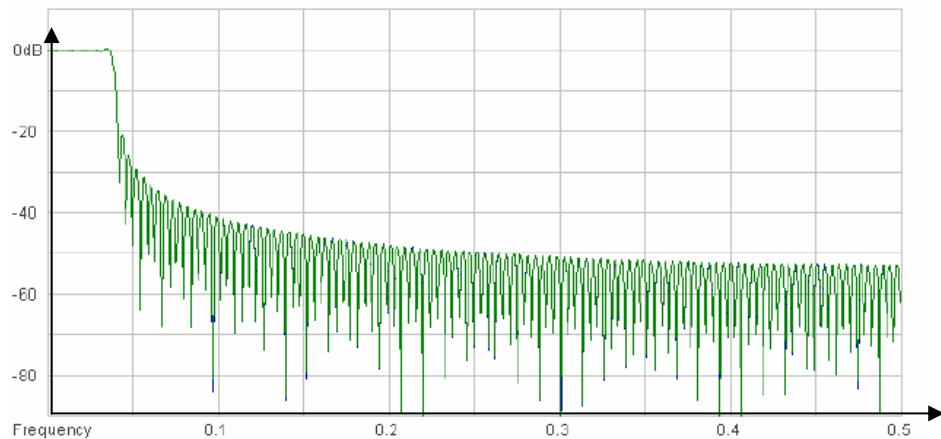


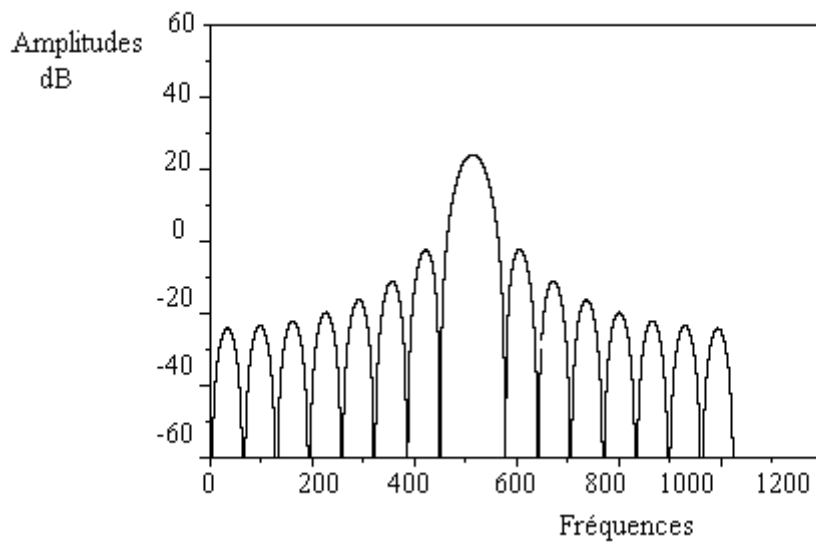
Figure 1 : Fenêtre Rectangulaire (256points)

**I-2-2- La fenêtre triangulaire (à bande étroite):** Elle est définie par l'équation suivante :

$$W_T(n) = \begin{cases} \frac{2n}{N} & 0 \leq n < \frac{N}{2} \\ \frac{2(N-n)}{N} & \frac{N}{2} \leq n < N-1 \\ 0 & \text{ailleurs} \end{cases} \dots\dots\dots[9]$$

$n=0\dots\dots N-1$ ,  $N$  : le nombre d'échantillon par fenêtre.

La courbe représentant

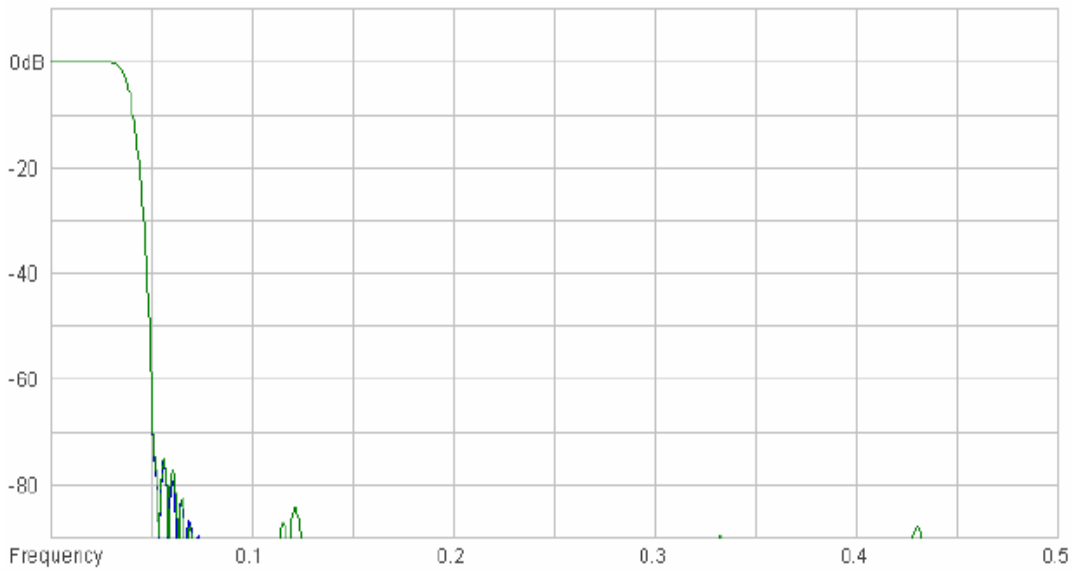


**Fig -2** la fenêtre triangulaire

**I-2-3- La fenêtre de Blackman :** Elle est définie par la formule suivante :

$$W(n) = \begin{cases} 0.42 + 0.46 \cos\left(2\pi \frac{n}{2N+1}\right) + 0.08 \cos\left(2\pi \frac{4n}{2N+1}\right) & 0 \leq n < N-1 \\ 0 & \text{Ailleurs} \end{cases} \dots\dots\dots[9]$$

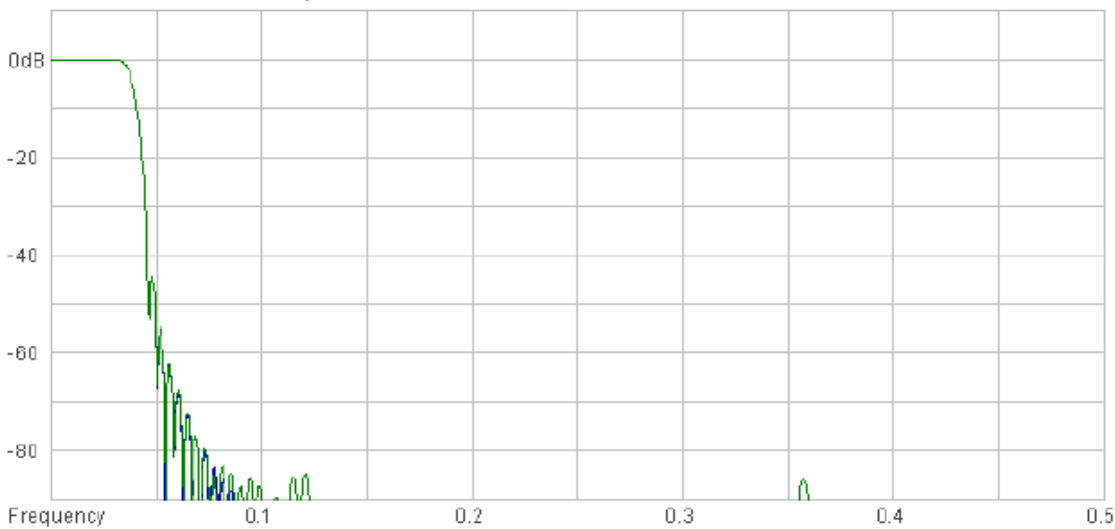
La courbe représentant la fenêtre de blackman est la suivante :



**Fig 3 : La fenêtre de Blackman**

**I-2-4- La fenêtre de Hanning (Von Hann) (à bande large) :** Elle est définie par la formule suivante

$$W_v(n) = \begin{cases} 0.5 - 0.5 \cos\left(\frac{2\pi n}{N}\right) & 0 \leq n < N - 1 \\ 0 & \text{ailleurs} \end{cases} \dots\dots\dots [9]$$



**I-2-5- La fenêtre Hamming (à bande étroite) :** Sa formule est donnée par :

$$W(n) = \begin{cases} 0.54 + 0.46 \cos \frac{2n\pi}{N} & 0 \leq n < N - 1 \\ 0 & \text{ailleurs} \end{cases} \dots\dots\dots[9]$$

$n=0\dots\dots\dots N-1$ ,  $N$  : le nombre d'échantillon par fenêtre.

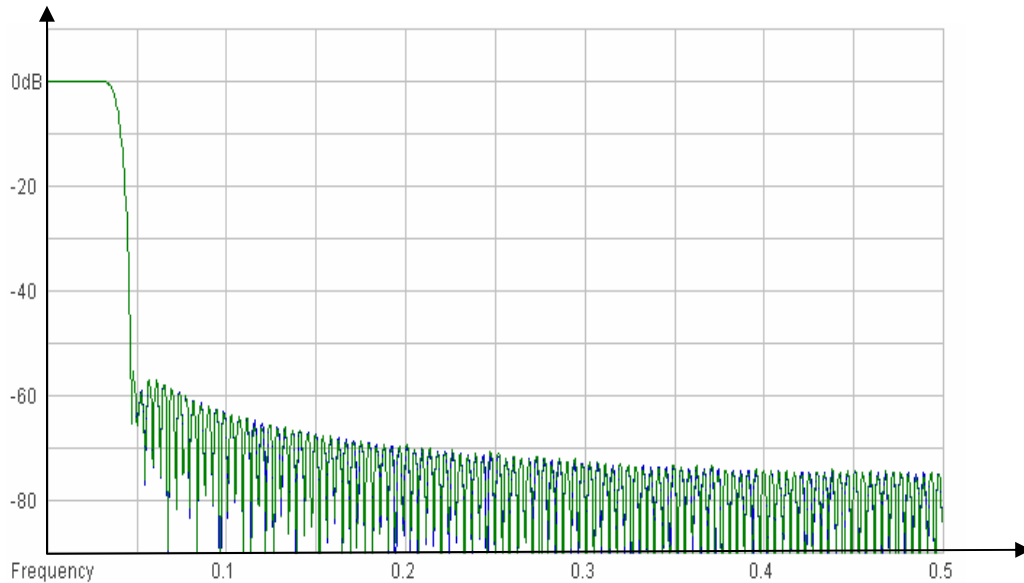


Figure 5 : Fenêtre Hamming (256points,  $F_c=10\text{kHz}$ ).

Le tableau suivant résume les paramètres associés aux différentes fenêtres.

Fenêtres	Largeur du lobe principale	Amplitudes des lobes secondaires	Décroissances des lobes secondaires en dB/octave
Rectangulaire	$2/N$	- 13 dB	-6 dB
Triangulaire	$4/N$	-27 dB	-12 dB
Hamming	$4/N$	-43 dB	-6 dB
Hanning	$4/N$	-32 dB	-18 dB
Blackman	$11/N$	-57dB	-74 dB

**I-3- Interprétation et choix de la fenêtre:**

L'étude comparative des différentes fenêtres d'analyse, montrent clairement sur quoi doit se porter notre choix: -La position.

-La forme.

-La taille (nombre d'échantillons).

Nous remarquons que la position est choisie de manière à conserver les échantillons importants du signal à traiter. Quant à la forme, son choix dépend de la largeur du pic central et de l'amplitudes des lobes secondaires. En ce qui concerne la taille des échantillons, il suffit de réduire les oscillations en limitant N (nombre d'échantillon de la fenêtre).

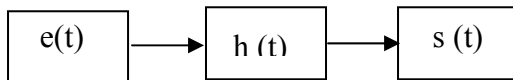
Malgré la simplicité de la fenêtre rectangulaire, elle n'est pas adaptée à l'analyse de la parole. C'est la fenêtre de Hamming que nous utilisons à cause de ses propriétés les plus favorables. D'autre part il convient de signaler que le choix de cette dernière nécessite un plus grand temps de calcul.

**II- convolution et corrélation :**

Par définition la convolution est une grandeur qui permet de déterminer l'allure de signal de sortie d'un système donné sans connaître son schéma interne.

L'objectif de la corrélation est de trouver une relation entre un signal aléatoire reçu dans un intervalle de temps et un autre signal aléatoire décalé dans le temps.

Tout système est caractérisé par sa réponse impulsionnelle  $h(t)$  tel que :



$$s(t) = h(t) * e(t) = \int_{-\infty}^{+\infty} e(\tau) \cdot h(t - \tau) d\tau = \int_{-\infty}^{+\infty} e(\tau) \cdot h[-(\tau - t)] d\tau \dots\dots\dots [10]$$

La fonction d'autocorrélation  $C_{xx}$  est définie par l'équation suivante :

$$C_{xx}(\tau) = 1/T \int_{-\infty}^{+\infty} x(\tau) \cdot x(\tau + t) d\tau \dots\dots\dots [10]$$

La fonction d'intercorrélation  $C_{xy}$  est définie par l'équation suivante :

$$C_{xy}(\tau) = 1/T \int_{-\infty}^{+\infty} x(\tau) \cdot y(\tau + t) d\tau \dots\dots\dots [10]$$

Les opérations de convolution et de corrélation sont liées par la relation suivante :

$$C_{xy}(\tau) = x(\tau) * y(-\tau) \dots\dots\dots [10]$$

## **Les abréviations :**

**CAN** : Convertisseur Analogique Numérique.

**TPZ** : Taux de Passage par Zéro.

**LPC** : Linear Predictive Coding (Coefficients de la Prédiction Linéaire).

**TF** : Transformée de Fourier.

**TFCT** : Transformée de Fourier à Court Terme.

**TFD** : Transformée de Fourier Discrète.

**FFT** : Fast Fourier Transforme (la Transformée de Fourier Rapide).

**TFFG** : Transformée de Fourier à Fenêtre Glissante.

**RAP** : la Reconnaissance Automatique de la Parole.

**RAL** : la Reconnaissance Automatique du Locuteur.

**QV** : la Quantification Vectorielle.

**DTW** : acronyme de Dynamique Time Warping (Déformation Dynamique Temporelle)

**HMM (MMC)** : Hidden Markov Models (Modèles de Markov Cachés)

**MMG** : Modèles par les Mélanges de Gaussiennes.

# GLOSSAIRE

**Le bruit** : Le bruit représente tout ce qui interfère avec la récupération d'un signal. Il est mesuré en énergie et est par définition aléatoire.

**Le bruit blanc** : Un bruit blanc est un bruit dont le spectre d'amplitude est constant sur toute la bande de fréquence.

**La carte son** : Est un élément intégré à l'ordinateur, permet de récupérer dans l'ordinateur le son enregistré dans l'appareil interne, le son sera ensuite enregistré dans un fichier sur le disque dur, puis travaillé, modifié avec des logiciels de traitement de son.

**Le spectre** : Intensité en décibel du signal en fonction de la fréquence à un instant donné.

**La prosodie** : Partie de la phonétique qui étudie l'intonation, l'accentuation, le rythme, les pauses, la durée des phonèmes.

**Onde sonore** : Conversion numérique d'un signal analogique externe ou numérique. Lorsqu'il y a une oscillation des particules d'un milieu de part et d'autre d'un point ou état d'équilibre.

**Réverbération** : Son qui résulte de l'ensemble de toutes les réflexions qui peuvent se produire dans un endroit.

**Signal analogique** : Un signal analogique est un signal dont les valeurs varient d'une façon continue dans le temps.

**Signal numérique** : Un signal numérique est un signal dont les valeurs varient d'une façon valeur discrète dans le temps.

**La compression** : rendre plus compact / réduire les crêtes de façon à limiter la distorsion.

# Bibliographie

[1] : P. Hall « **Techniques cepstrales améliorées pour l'extraction d'enveloppe spectrale et la détection du pitch** ».

Paris-1983. (Bibliothèque du département d'électronique)

[2] : Christian TAVERNIER« **Faites parler vos montages** ».

Paris-1984. (Bibliothèque du département d'électronique)

[3] : Mariani JOSEPH « **Analyse, synthèse et codage de la parole tome 1 et 2** ». Lavoisier-2002 (Bibliothèque du département d'électronique)

[4] : Jean-Paul HATON, Jean-Marie PIERREL, Guy PERENNOU, Jean CAELEN et Jean-Luc GAUVAIN. « **Reconnaissance automatique de la parole** ».

Paris-1991. (Bibliothèque du département d'électronique)

[5] : Henri MELONI.« **Fondement et perspectives en traitement automatique de la parole** ».

AUPELF-UREF-1996. (Bibliothèque du département d'électronique)

[6] : André QUINQUIS.« **Traitement du signal sous matlab** ».

Paris-2000. (Bibliothèque du département d'électronique)

[7] : Madame AMRANE – STIET Malika. « **Paramétrisation et segmentation du signal de parole** »

Thèse de magister en électronique. UMMTO-2002. (Bibliothèque centrale Hasnaoua).

**[8] : BOUSEKSOU Boualem « Reconnaissance automatique de la parole par la méthode globale ».**

Thèse de magister option électronique acoustique et physiologique de la

Parole. Institut de linguistique et de phonétique. Alger-1983. (Bibliothèque centrale Hasnaoua).

**[9] : TIGZIRI née IBRAHIM Noura « Reconnaissance automatique de la parole dans un contexte multilocuteur »**

Thèse de magister option électronique acoustique et physiologique de la parole.

Institut de linguistiques et de phonétique. Alger-1985 (Bibliothèque centrale Hasnaoua).

***Sites Internet :***

[10] : <http://www.enl.UCA.lgary.ca>

[11] : <http://tcts.fpms.ac.be/cours/1005-08/speech/>

[12] : <http://www.speech.cs.cmu.edu/comp.speech/>

**Les sites Internet pour télécharger les logiciels du traitement de son :**

- Goldwave (éditeur de sons *wav*) : <http://www.goldwave.com/>
- Praat: <http://www.syntrillium.com/Praat/>
- Sound Forge: <http://www.sonicfoundry.com/>

```

% LE PROGRAMME DU SYSTEME DE RECONNAISSANCE
.....
% INITIALISATION
.....
clc
clear all
close all
.....
% LE PROGRAMME DE LA PHASE D'APPRENTISSAGE ET CREATION DES REFERENCES.
.....
% LE TRAITEMENTS DES SIGNAUX DE REFERENCES
.....
x1=wavread('MOUNIA R');
figure(1);
plot(x1);
[t1,t2]=size(x1);
.....
% LE FENETRAGE DES SIGNAUX AVEC LA FENETRE DE HAMMING
.....
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:84
    for j=1:N
        y(j)=x1(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen;
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
.....
% LE DEBRUITAGE DES SIGNAUX AVEC LES ONDELETTES
.....
z1=dwt(z,'db1');
figure(5);
plot(z1);
.....
% CALCUL DU NOMBRE DE PASSAGE PAR ZERO 'Nbi' DU SIGNAL DEBRUITE
.....
Nb1=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb1=Nb1+1;
        Nb1=Nb1 '
    else k=k+1;
    end;
end;
.....
% LE CALCUL DU TAUX DE PASSAGE PAR ZERO (TPZ)
.....
TPZ1=mrdivide(Nb1,N)
.....

x2=wavread('KAHINA R');
figure(1);
plot(x2);
[t1,t2]=size(x2);
N=1000;
fen=hamming(N);

```

```

figure(2);
plot(fen);
z=[];
y=[];
for i=1:85
    for j=1:N
        y(j)=x2(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);

Nb2=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb2=Nb2+1;
        Nb2=Nb2';
    else k=k+1;
    end;
end;
TPZ2=mrdivide(Nb2,N)
.....
x3=wavread('SAMIRA R');
figure(1);
plot(x3);
[t1,t2]=size(x3);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:77
    for j=1:N
        y(j)=x3(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);
Nb3=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb3=Nb3+1;
        Nb3=Nb3';
    else k=k+1;
    end;
end;
TPZ3=mrdivide(Nb3,N)

```

```

.....
x4=wavread('AMINNA R');
figure(1);
plot(x4);
[t1,t2]=size(x4);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:103
    for j=1:N
        Y(j)=x4(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);
Nb4=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb4=Nb4+1;
        Nb4=Nb4';
    else k=k+1;
    end;
end;
TPZ4=mrdivide(Nb4,N)
.....
x5=wavread('NAIIMA R');
figure(1);
plot(x5);
[t1,t2]=size(x5);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:64
    for j=1:N
        Y(j)=x5(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);
Nb5=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;

```

```

        Nb5=Nb5+1;
        Nb5=Nb5 '
    else k=k+1;
    end;
end;
TPZ5=mrdivide(Nb5,N)
.....
x6=wavread('MOHMED R');
figure(1);
plot(x6);
[t1,t2]=size(x6);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:88
    for j=1:N
        Y(j)=x6(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);
Nb6=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb6=Nb6+1;
        Nb6=Nb6 '
    else k=k+1;
    end;
end;
TPZ6=mrdivide(Nb6,N)
.....
x7=wavread('DJAMEL R');
figure(1);
plot(x7);
[t1,t2]=size(x7);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:79
    for j=1:N
        Y(j)=x7(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);

```

```

plot(z1);
Nb7=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb7=Nb7+1;
        Nb7=Nb7'
    else k=k+1;
    end;
end;
TPZ7=mrdivide(Nb7,N)
.....
x8=wavread('AZZIZE R');
figure(1);
plot(x8);
[t1,t2]=size(x8);
N=1000;
fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
for i=1:91
    for j=1:N
        y(j)=x8(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
z1=dwt(z,'db1');
figure(5);
plot(z1);
Nb8=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nb8=Nb8+1;
        Nb8=Nb8'
    else k=k+1;
    end;
end;
TPZ8=mrdivide(Nb8,N)
.....
% STOCKAGE DES SIGNAUX DANS LE DICTIONNAIRE DE REFERENCES
.....
nom_ref=['MOUNIA';'KAHINA';'SAMIRA';'AMINNA';'NAIIMA';'MOHMED';'DJAMEL';'AZZIZE']
Nbi=[Nb1,Nb2,Nb3,Nb4,Nb5,Nb6,Nb7,Nb8]
.....
% LE PROGRAMME DE LA PHASE DE RECONNAISSANCE
.....
% TRAITEMENT DU SIGNAL DE LA NOUVELLE ACQUISITION(LE TEST)
.....
nom_test='MOUNIA T'
ytest=wavread('MOUNIA T');
Figure(1);
plot(ytest);
[t1,t2]=size(ytest);
.....
%FENETRAGE DU SIGNAL TEST
.....
N=1000;

```

```

fen=hamming(N);
figure(2);
plot(fen);
z=[];
y=[];
L=length(ytest)
for i=1:(mrdivide(L,N))
    for j=1:N
        y(j)=ytest(((i-1)*N)+j);
    end
    z=[z;y];
    h=y.*fen';
    zfen=[z;h];
end
figure(3);
plot(zfen(10,:));
figure(4);
plot(z(10,:));
.....
% DEBRUITAGE DU SIGNAL TEST
.....
z1=dwt(z, 'db1');
figure(5);
plot(z1);
.....
% CALCUL DU NOMBRE DE PASSAGE PAR ZERO 'Nby' DU TEST
.....
Nby=0;
k=1;
for k=1:498;
    k=k+1;
    if z1(k)*z1(k+1)<0;
        Nby=Nby+1;
        Nby=Nby';
    else k=k+1;
    end;
end;
.....
% CALCUL DU TAUX DE PASSAGE PAR ZERO 'TPZy'
.....
TPZy=mrdivide(Nby,N)
Nby=[Nby]
Nbi=[Nb1,Nb2,Nb3,Nb4,Nb5,Nb6,Nb7,Nb8]
TPZi=[TPZ1,TPZ2,TPZ3,TPZ4,TPZ5,TPZ6,TPZ7,TPZ8]
.....
%LA COMPARAISON ENTRE LE SIGNAL TEST ET LES DIFFERENTS SIGNAUX DEREFERENCES
.....
% CALCUL DES DIFFERENCES 'Di' ENTRE 'TPZy' ET 'TPZi'
.....
Di=[abs(TPZy-TPZ1);abs(TPZy-TPZ2);abs(TPZy-TPZ3);abs(TPZy-TPZ4);
abs(TPZy-TPZ5);abs(TPZy-TPZ6);abs(TPZy-TPZ7);abs(TPZy-TPZ8)]
.....
% RECHERCHE DE LA PLUS PETITE VALEUR DE 'Di' ET LE RESULTAT DU LOCUTEUR
CORRESPONDANT
.....
[m,indice]=min(Di)
if ([m,indice])>=[0.001,indice] &([m,indice])<[0.03,indice]
    nom=nom_ref(indice,:)
else
    nom=nom_test
end;

```