الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique

Mémoire de Projet de Fin d'Études

présenté par

KHATiRWalid

&

BENHAMRiMohamed

pour l'obtention du diplôme Master en Télécommunicationoption Télécommunication et Réseaux

Thème

Identification automatique d'un locuteur par la méthode MFCC

Proposé par : Mr Djebari et Mr Hemri.

Année Universitaire 2015-2016

Remerciements

Nous tenons à remercier en premier lieu Dieu le tout puissant qui nous a doté de toute la force nécessaire à l'aboutissement de ce mémoire.

Nos remerciements vont aussi à notre promoteur le Dr Djebari et le doctorant Hemri qui nous a fait bénéficier de ses nombreuses et riches connaissances durant toute la durée de ce travail.

Nous remercions les membres du jury; qui nous honorent par la lecture de notre mémoire et de leur présence, le jour de notre soutenance.

Sans Oublier tous ceux qui ont contribué, de près ou de loin à mener à terme ce travail. Notamment nos enseignants qui nous ont instruits Tout au long de notre long parcours.

Dédicaces

A mes très chers parents,

Nul mot ne pourra exprimer mes sentiments et ma gratitude envers vous,

A mon frère et àmes sœurs,

Je vous souhaite beaucoup de bonheur et de réussite.

A toute ma grande famille.

A mes chers amis.

Amon binôme Mohamed.

A tous ceux qui m'aiment, A tous ceux que j'aime,

Je dédie le fruit de mon projet de fin d'études.

Walid

Dédicaces

Nous dédions ce travail à :

Ma mère, source de tendresse et d'amours pour leurs soutiens tout le long de notre vie scolaire.

Mon père, qui nous a toujours soutenu et qui a fait tout possible pour nous aider.

Nos frères et nos sœurs, que nous aimons beaucoup.

Notre grande famille.

Nos cher ami (e) s, et enseignants.

Tout qu'on collaboré de prés ou de loin à l'élaboration de ce travail.

Mohamed

Introduction Générale

La reconnaissence automatique de locuteur est l'identification d'une personne à partir des caractéristiques de sa voix. Il y a une différence entre la reconnaissance du locuteur (reconnaissance qui parle) et la reconnaissance de la parole (en reconnaissant ce qui est dit). Ces deux termes sont souvent confondus, et "la reconnaissance vocale" peut être utilisé pour les deux. En outre, il y a une différence entre l'acte d'authentification (communément appelée vérification du locuteur ou de l'authentification du locuteur) et l'identification.

On va parler dans ce mémoire de la généralités de la parole après on va discuter le methode de reconnaissanse MFCC et enfin on voir le progamme qui nous montre ça.

Finalement, il y a une différence entre la reconnaissance du locuteur (reconnaissance qui parle) et diarization du locuteur (Reconnaître quand le même locuteur parle). Reconnaissant le locuteur peut simplifier la tâche de traduire la parole dans les systèmes qui ont été formés sur les voix des personnes spécifiques ou il peut être utilisé pour authentifier ou vérifier l'identité d'un locuteur dans le cadre d'un processus de sécurité [1].

Résumé

ملخص: في هذا المشروع، نقترح طريقة للتعرف التلقائي للمتكلم. النظام المستخدم يعتمد على برنامج مثبت على جهاز الكمبيوتر مجهز بميكروفون. تم تنفيذ المشروع باستخدام لغة البرمجة ماتلاب.

كلمات المفاتيح : التعرف التلقائي للمتكلم, لغة البرمجة ماتلاب.

Résumé : Dans ce projet, nous proposons une méthode de reconnaissance automatique de locuteur. Le système utilisé se base sur programme installé sur un ordinateur muni d'un microphone. Le projet est réalisé en utilisant le langage de programmation Matlab.

Mots clés : Reconnaissance automatique de locuteur; langage de programmation Matlab.

Abstract: In this project, we propose an automatic speaker recognition method. The used system is based on a program installed on a computer equipped with a microphone. The project is realized, using the Matlab programming language.

Keywords: automatic speaker recognition**;** Matlab programming language**.**

Table des matières

Table des matières

Liste des figures

Listes des acronymes et abréviations

LPC : Linear Prédictive Coding

LPCC : Linear Predictive Coefficients Cepstral

MFCC : Mel Frequency Cepstral Coefficients

FFT : Fast Fourier transform

TFD : Transformée de Fourier Discrète

DCT : Discret Cosinus Transform

FFT $^{-1}$ **:** Transformée Fourier Inverse

DCT $^{-1}$ **:** Transformée en Cosinus Inverse

RAP : Reconnaissance Automatique de la Parole

MATLAB : Matrix Laboratory

SFS : Speech Filing System

1.1.Introduction

Comme on le souligne souvent, il n'y a pas dans le corps humain un organe responsable de la parole: la parole se fait par la collaboration de différents organes (qui servent normalement à respirer, mastiquer etc.). Le son est fondamentalement une vibration de l'air. Cet air est fourni par les poumons; la vibration est fournie par les plis vocaux; enfin des mouvements compliqués de la langue (principalement) et des lèvres donnent au son une « couleur » particulière [2].

Dans ce chapitre, nous présentons la physiologie anatomique du mécanisme de production de la parole humaine.

1.2.Physiologie des organes de la phonation

Trois groupes d'organes assument les fonctions essentielles dans l'acte de parole (fig 1.1), ou phonation :

- ✓ l'appareil respiratoire, (diaphragme, poumons, trachées), soufflerie qui fournit l'énergie et la quantité d'air nécessaire ;
- ✓ le larynx, organe vibrant, où naît le son ;
- ✓ Le conduit vocal, formé des cavités résonantes supra-laryngées (pharynx, bouche, nez) où s'effectue l'articulation proprement dite par les changements de forme du tractus vocal.

  Ces changements resultent surtout des mouvements des lèvres, de la langue, du voile du palais (dont l'abaissement fait intervenir une cavité supplémentaire, les fosses nasales) et de la mâchoire inférieure [3].

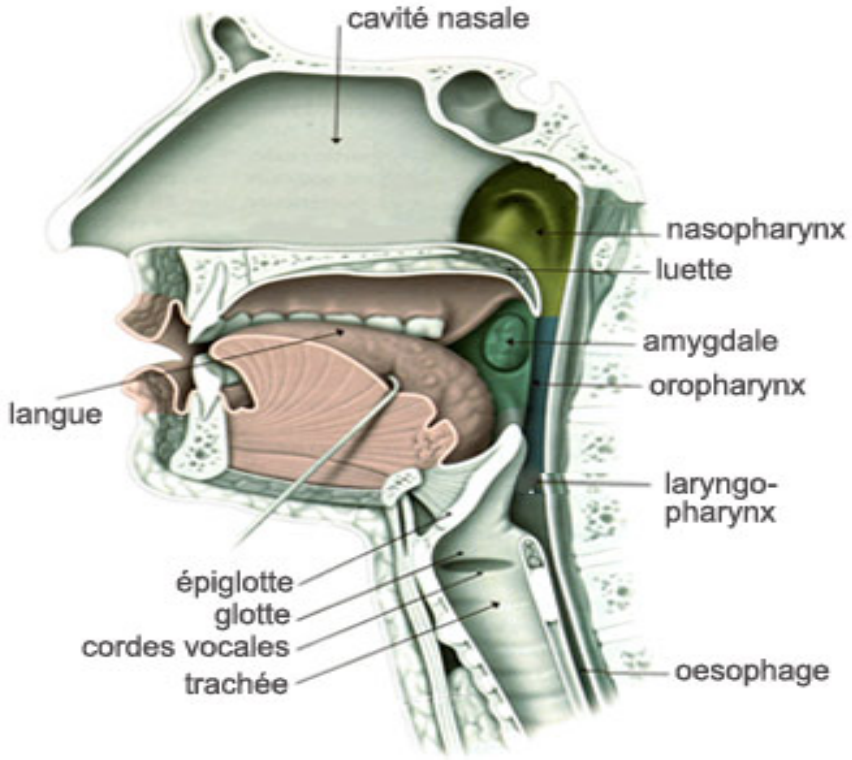

Figure 1.1 : Appareil phonatoire humain [4].

1.2.1. L'appareil phonatoire humain

Le langage parlé nécessite tout un ensemble d'organes permettant la production de sons. Cet ensemble est appelé l'appareil phonatoire (fig 1.2). Il est constitué successivement de :

- ✓ Partie sub-glottique ou appareil respiratoire (diaphragme poumons, trachée) qui fournit l'énergie nécessaire à la phonation en insufflant l'air vers la partie glottique ;
- ✓ Partie glottique ou larynx (ensemble de cartilages, ligaments ligaments et muscles) muscles) contenant contenant les cordes vocales (replis tendus horizontalement qui, sous l'effet des muscles, muscles, jouent un rôle de valve vis-à-vis de l'air des poumons libérant ainsi un flux d'air vers la partie supra-glottique).
- ✓ Partie supra-glottique ou conduit vocal, formé des cavités orales (pharyngienne et buccale), à géométrie variable, en fonction des éléments articulateurs (langue, mâchoire inférieure, lèvres) et des cavités nasales, à géométrie fixe, pouvant être couplées aux cavités orales par abaissement du voile du palais [3].

Chacun de ces éléments du corps humain est nécessaire à la production du langage parlé.

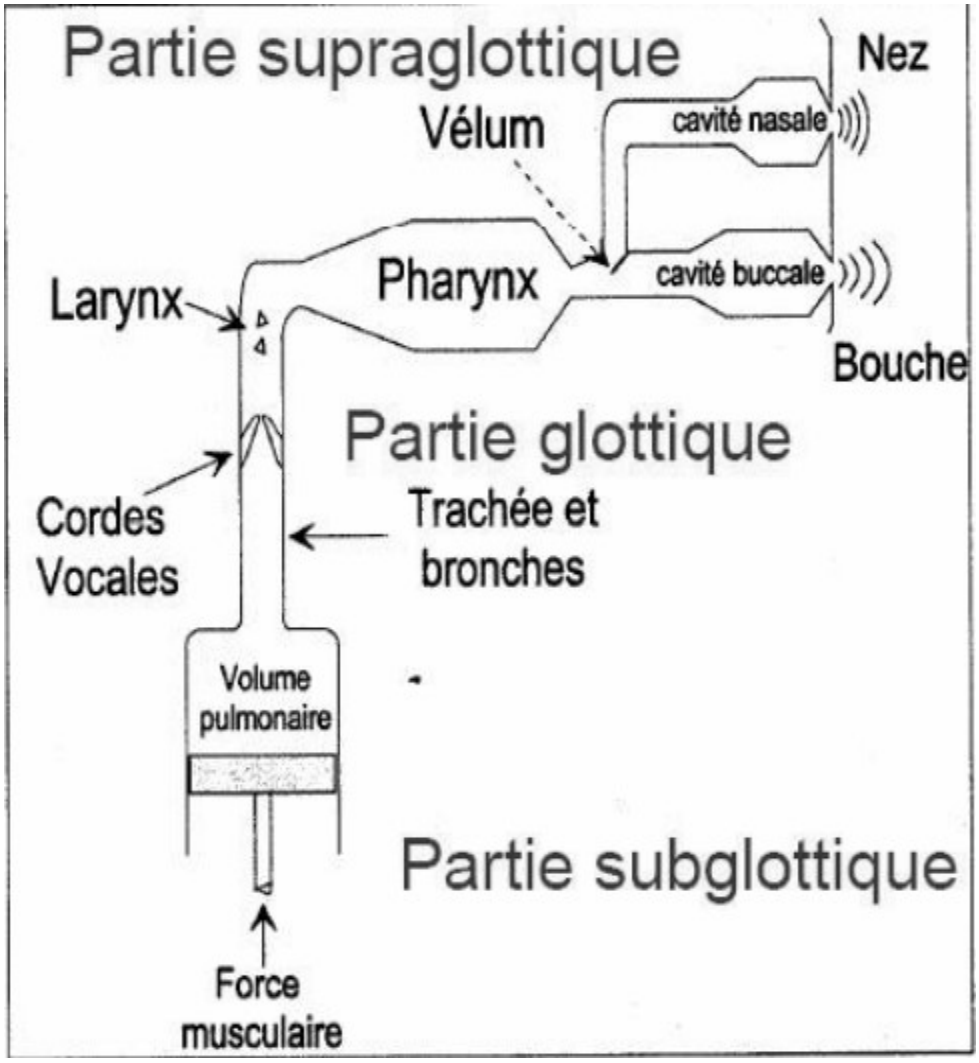

Figure 1.2 : L'appareil phonatoire humain schématisé [3].

1.2.2. Principe de production d'un son :

Au cours de la phonation, les sons de la parole sont produits soit par les vibrations des cordes vocales (sources de la phonation), soit par l'écoulement turbulent de l'air dans le conduit vocal ou à l'issue d'un relâchement d'une occlusion de ce conduit (source de bruit). La parole se distingue des autres sons par ses caractéristiques acoustiques ayant leurs origines dans l'interdépendance des paramètres de la source et du conduit vocal.

Pour commencer l'étude du langage, appréhender sa complexité et se rendre compte de sa spécificité, la phonétique est un terrain de choix. En effet, l'une des premières questions qui se posent est de savoir ce que sont les sons, et comment ils sont produits. La question n'est pas simple, car elle peut être envisagée de deux points de vue : le point de vue acoustique (propriétés physiques) et le point de vue articulatoire .

1.2.3. Le lieu d'articulation :

L'étude des sons du langage humain est envisagée, en phonétique articulatoire, sous l'angle de la production. Cette discipline comporte un volet sur la physiologie, consacré à la connaissance des organes de la phonation et un volet descriptif portant sur le rôle des différents organes dans la production des sons du langage.

Dans un processus de communication parlée, pour une langue donnée, l'articulation n'est en fait qu'un processus de production des sons par différents mouvements des articulateurs, lesquels changent la configuration du conduit vocal. Grâce à leur unité élémentaire appelée phonème, ces sons permettent alors la distinction des différents mots.

En fonction des possibilités offertes par l'appareil vocal, chaque langue a adopté un ensemble particulier de sons distinctifs ou phonèmes qui sont les éléments sonores les plus brefs permettant de distinguer les différents mots.

Il est possible d'opérer une classification des sons d'une langue à partir de critères articulatoires, c'est-à-dire à partir de la manière dont ils sont produits par les organes de la parole. Cette classification oppose principalement les voyelles et les consonnes. Il existe une troisième classe de sons, les semi-voyelles, appelées ainsi car elles peuvent être assimilées aux voyelles et aux consonnes.

Les critères permettant de classifier les sons d'une langue sont les suivants :

- le mode articulatoire a trait à la qualité du passage de l'air dans le canal buccal. La réalisation des voyelles implique un passage libre de l'air le long du canal buccal. Le degré d'ouverture de la cavité buccale permet de distinguer quatre types de voyelles : les voyelles ouvertes, mi-ouvertes, mi-fermées et fermées. Pour les consonnes, deux modes articulatoires sont à distinguer. Le passage de l'air est totalement bloqué ou obstrué lors de la production des consonnes occlusives. Le passage est rétréci suffisamment pour permettre l'émission d'un bruit continu lors de la réalisation des consonnes constrictives ou fricatives ;
- la résonance orale ou antirésonance nasale est fonction de la fermeture ou de l'ouverture de l'accès vers les fosses nasales. Lors de la production de voyelles ou de consonnes nasales, le voile du palais est abaissé et permet le passage de l'air à la fois par le canal buccal et par les fosses nasales, ce qui confère aux sons une coloration particulière. Les voyelles et les consonnes produites sont alors dites " *nasales* ". Lorsque le voile du palais est relevé et bien accolé à la paroi pharyngale, l'air ne passe que par la cavité buccale, donnant naissances aux sons vocaliques et consonantiques dits oraux ;

- le rôle des cordes vocales détermine le caractère sourd ou sonore des différentes articulations. Lorsque les cordes vocales vibrent, les sons seront dits voisés ou sonores par opposition aux sons non voisés ou sourds . La réalisation des voyelles implique la mise en vibration des cordes vocales. Pour les consonnes, l'absence ou la présence de ces vibrations détermine leur caractère sourd ou sonore ;
- le lieu d'articulation se situe nécessairement dans la partie supérieure du canal buccal dans une zone allant de la lèvre supérieure jusqu'à la paroi pharyngale. C'est le point duquel l'articulateur se rapprochera ou avec lequel il entrera en contact ;
- l'articulateur est constitué par la région inférieure du canal buccal. Il s'agit de la lèvre inférieure et des différentes parties de la langue. La réalisation de toute articulation implique un rapprochement plus ou moins grand ou un contact franc entre l'articulateur et le lieu d'articulation ;
- le rôle des lèvres détermine le caractère labialisé ou non labialisé d'une articulation. En effet, toute articulation peut être accompagnée ou non d'une projection des lèvres [5].

1.2.4. Classification des phonèmes

En générales la langue est composée de plus d'une trentaine de phonème repartis en plusieurs classes.

1.2.4.1. Les voyelles

Elles sont caractérisées acoustiquement par la présence des maxima spectraux ; c'est-à-dire des zones de fréquence où les harmoniques sont particulièrement intenses (formant). Elles sont générées par une vibration laryngienne des cordes vocales appelées fréquences fondamentales dont les valeurs varient entre 80 et 180 Hz pour les hommes et de 130 et 290 Hz pour les femmes. On distingue deux types de voyelles :

- voyelles nasales : dont le conduit nasal est couplé à la cavité buccale et l'émission se fait à la fois par les narines et par la bouche ; (exp : bl**anc,** b**on,** l**in**, br**un**) ;
- voyelles orales : sont des sons voisés ; chacune d'elles correspond à une configuration particulière du conduit vocal, sans intervention de la cavité nasale qui est alors isolée par fermeture du voile du palais (exp : pl**a**t, l**ai**t, b**o**l, bl**é**, roue**,** p**i**le).

1.2.4.2. Les semi-voyelles ou semi-consonnes

Ils **s**ont des phonèmes intermédiaires entre les voyelles et les consonnes. Quand on les prononce, on entend le timbre d'une voyelle auquel s'ajoute le frottement d'une consonne spirante. Leur fréquence d'emploi est liée à la vitesse du débit de la parole, plus celui-ci est rapide, plus il y aura de semi-voyelle (exp : hier, h**u**it, **ou**i).

1.2.4.3. Les consonnes

Ce sont des sons résultants d'une fermeture partielle (constriction) ou totale (occlusion) du conduit vocal lors du passage de l'air phonatoire ; elles peuvent être voisées ou non voisées ; nasales ou orales; les consonnes sont classées selon les trois principaux types suivants :

- fricatives (constrictives) : dans cette classe sont regroupés les sons produits par la friction de l'air dans le conduit vocal lorsque celui-ci est rétréci au niveau des lèvres, des dents ou de la langue. Cette friction produit un bruit de hautes fréquences et peut être voisée [v], [z] (exp : **v**erre, Asie) ou sourde [f], [s], [š] (exp : **f**er, a**ss**is, **ch**ou) ;
- plosives (occlusives) : un son plosive est produit par une occlusion momentanée du conduit vocal, en un point donné suivi par une ouverture brusque, et peut être sonore (exp : **b**asse, **d**oux, **g**oût) ou sourde (exp : **p**asse, **t**oux, **c**ou) ;
- nasales : dans ce cas le conduit vocal est fermé et l'air s'écoule par la cavité nasale. On distingue deux consonnes nasales, toutes les deux voisées :
  - [m] : dont le lieu d'articulation est labial (exp : **m**asse) ;
  - [n] : dont le lieu d'articulation est dentaire (exp : **n**ous, si**gn**al).

1.2.4.4. Les sonantes

Se caractérisent par une structure de formants et elles ne possèdent que peut ou pas de bruit, plusieurs sous-classes existent : les vibrantes telles que le [r], les liquides tels que le [l] **:**

- les vibrantes : il s'avère qu'il en existe une seule qui est le [r] (exp : **r**ue) qui est produite par une vibration de la langue et est caractérisée par une structure de formant interrompu par des intervalles de silences très courts, résultat du battement de la langue ;
- les liquides : il en existe une seule qui est produite par une obstruction partielle du conduit buccal et un écoulement latéral [l] (exp : **lent**). Au plan spectral, elle est caractérisée par une structure de formant similaire à celle des voyelles.

Pour mieux illustrer les différents modes d'articulation des phonèmes qui sont classifiés dans la (figure 1.3).

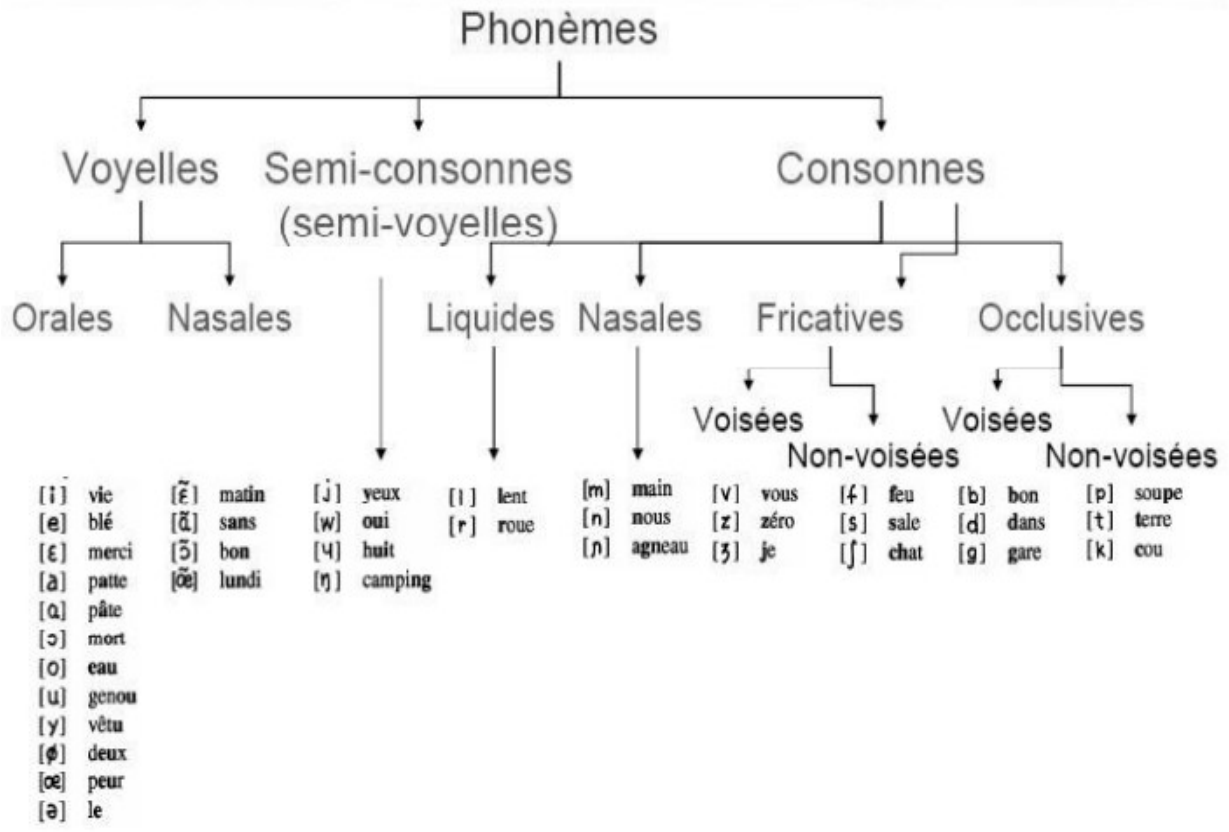

Fig 1.3 : Classification des phonèmes.

Le problème de la parole en traitement ou en reconnaissance dépend surtout du type de l'application ou son contexte, c'est pour cela que l'on a introduit la notion de variabilité de la parole.

1.3. Problèmes de variabilité de la parole

Il y a d'abord le problème de la variabilité intra et inter-locuteurs. Le système est-il dépendant du locuteur (optimisé pour un locuteur bien particulier) ou indépendant du locuteur (pouvant reconnaître n'importe quel utilisateur)?

Evidemment, les systèmes dépendants du locuteur sont plus faciles à développer et sont caractérisés par de meilleurs taux de reconnaissance que les systèmes indépendants du locuteur étant donné que la variabilité du signal de parole est plus limitée. Cette dépendance au locuteur est cependant acquise au prix d'un entraînement spécifique à chaque utilisateur. Ceci n'est cependant pas toujours possible. Par exemple, dans le cas d'applications téléphoniques, il est évident que les systèmes doivent pouvoir être utilisés par n'importe qui et doivent donc être indépendants du locuteur. Bien que la méthodologie de base reste la même, cette indépendance au locuteur est cependant obtenue par l'acquisition de nombreux locuteurs (couvrant si possible les différents dialectes) qui sont utilisés simultanément pour l'entraînement de modèles

susceptibles d'en extraire toutes les caractéristiques majeures. Une solution intermédiaire parfois utilisée est de développer des systèmes capable de s'adapter (de façon supervisée ou non supervisée) rapidement au nouveau locuteur.

Le système reconnaît-il des mots isolés ou de la parole continue? Evidemment, il est plus simple de reconnaître des mots isolés bien séparés par des périodes de silence que de reconnaître la séquence de mots constituant une phrase. En effet, dans ce dernier cas, non seulement la frontière entre mots n'est plus connue mais, de plus, les mots deviennent fortement articulés (c'est-à-dire que la prononciation de chaque mot est affectée par le mot qui précède ainsi que par celui qui suit - un exemple simple et bien connu étant les liaisons du français).

Dans le cas de la parole continue, le niveau de complexité varie également selon qu'il s'agisse de texte lu, de texte parlé ou, beaucoup plus difficile, de langage naturel avec ses hésitations, phrases grammaticalement incorrectes, faux départs, etc. Un autre problème, qui commence à être bien maîtrisé, concerne la reconnaissance de mots clés en parole libre. Dans ce dernier cas, le vocabulaire à reconnaître est relativement petit et bien défini mais le locuteur n'est pas contraint de parler en mots isolés. Par exemple, si un utilisateur est invité à répondre par « oui » ou « non », il peut répondre « oui, s'il vous plait ». Dans ce contexte, un problème qui reste particulièrement difficile est le rejet de phrases ne contenant aucun mots clés [6].

1.3.1. Variabilité intra-locuteur

La variabilité intra-locuteur identifie les différences dans le signal produit par une même personne. Cette variation peut résulter de l'état physique ou moral du locuteur. Une maladie des voies respiratoires peut ainsi dégrader la qualité du signal de parole de manière à ce que celui-ci devienne totalement incompréhensible, même pour un être humain. L'humeur ou l'émotion du locuteur peut également influencer son rythme d'élocution, son intonation ou sa phraséologie.

Il existe un autre type de variabilité intra-locuteur lié à la phrase de production de parole ou de préparation à la production de parole. Cette variation est due aux phénomènes de coarticulation.

1.3.2. Variabilité inter-locuteur

La variabilité inter-locuteur est un phénomène majeur en reconnaissance de la parole. Elle concerne les différences du signal vocal des locuteurs différents. Ces différences sont liées à l'âge, l'accent régional, le sexe, etc.

1.3.3. Variabilité due à l'environnement

La variabilité liée à l'environnement peut, parfois, être considérée comme une variabilité intra-locuteur mais les distorsions provoquées dans le signal de parole sont communes à toute personne soumise à des conditions particulières. La variabilité due à l'environnement peut également provoquer une dégradation du signal de parole sans que le locuteur ait modifié son mode d'élocution. Cette variation est considérée comme du bruit.

La variabilité environnementale due au locuteur peut tout d'abord être de nature physiologique.

Ainsi, un système mécanique provoquant une déformation du conduit vocal provoquera immanquablement une variation dans le signal de parole produit. Ces contraintes physiques sont généralement rencontrées dans les systèmes de transport où une posture particulière, ou une accélération lors du déplacement, pourront provoquer une déformation. Les moyens de transport peuvent également entraîner d'autres déformations du signal, d'origine psychologique. Le bruit ambiant peut ainsi provoquer une déformation du signal de parole en obligeant le locuteur à accentuer son effort vocal. Enfin, le stress et l'angoisse que certaines personnes finissent par éprouver lors de longs voyages peuvent également être mis au rang des contraintes environnementales susceptibles de modifier le mode d'élocution [7].

1.4. Extraction des paramètres du signal vocal

Depuis quelques années, le codage de la parole connaît un regain d'intérêt au sein de la communauté scientifique. En effet, les applications actuelles en parole donnent de bons résultats mais dans des environnements limités. Dans la chaîne de traitement, le codage occupe une place fondamentale. Il est utilisé pour l'analyse, la compression, la synthèse et il effectue aussi l'extraction des traits utilisés pour la reconnaissance des formes. Actuellement, on trouve deux grandes familles de codage qui sont les codages temporels (LPC, LPCC, etc.) et les codages paramétriques ou fréquentiels (Banc de Filtres, Cepstre, MFCC, etc…) [8].

1.4.1. Le codage LPC et LPCC

Le codage prédictif linéaire (LPC Linear ***P**rédictive **C**oding)* est une méthode de codage et de représentation de la parole**.** Elle repose principalement sur l'hypothèse que la parole peut être modélisée par un processus linéaire. Il s'agit donc de prédire le signal à un instant « **n** » à partir des ***p*** échantillons précédents donnée par l'équation (1.01). La parole n'étant cependant pas un processus parfaitement linéaire, la moyenne mobile que constitue la somme pondérée du signal sur ***p*** pas de temps introduit une erreur qu'il est nécessaire de corriger par l'introduction du terme ***e(n)*** [9].

Le codage par prédiction linéaire consiste donc à déterminer les coefficients a_k obtenus de la fonction de transfert (aussi appelés coefficients de réflexion) qui minimisent l'erreur ***e(n),*** ceci en fonction d'un ensemble de signaux constituant un corpus d'apprentissage.

$$s(n) = \sum_{k=1}^{p} a_k \ . \ s(n-k) + e(n) \qquad (1.01)$$

Les coefficients LPC sont basés sue le modèle de production de la parole, qui considère que l'appareil de production (cordes vocales et conduit vocal) est constitué d'une source (source pseudo-périodique ou source de bruit) et d'un filtre se comportant comme un résonateur. Les coefficients LPCC *(**L**inear **P**redictive **C**oefficients **C**epstral*) sont calculés à partir des coefficients LPC, ceci permet de minimiser l'accroissement du cout de calcul [10].

1.4.2. Les paramètres MFCC

Le principe de calcul des MFCC (Mel Frequency Cepstral Coefficients) est issu des recherches psycho-acoustique sur la perception des différentes bandes de fréquences par l'oreille humaine. Le principal intérêt de ces coefficients est extraire des informations pertinentes en nombre limité en s'appuyant à la fois sur la production (théorie Cepstrale) et sur la perception de la parole (échelle des Mels).

Plusieurs étapes sont nécessaires pour transformer un fichier audio en Cepstre MFCC. On détaillera dans le prochain chapitre la construction pas à pas de ces coefficients [11].

1.5. Méthode de reconnaissance de la parole

Les différentes méthodes appliquées pour la reconnaissance de la parole dépendent étroitement des types de problèmes à résoudre : taille de vocabulaire, mode mono-locuteur ou multi-locuteur, indépendant du locuteur, mots isolés, mots enchaînées ou parole continue. Il existe deux approches différentes pour la reconnaissance de parole[12].

1.5.1. Approche analytique

Cette méthode procède à une segmentation en unité de base (syllabe, demi syllabe, mot, etc.). Elle fait appel à plusieurs niveaux supérieurs (ou étapes de compréhension).

Cette méthode présente l'avantage de permettre la reconnaissance de parole continue pour grands vocabulaires, en contexte multi-locuteurs puisqu'on ne mémorise qu'un nombre restreint d'éléments, indépendant de la taille vocabulaire. Il se pose donc deux problèmes : segmenter et identifier une forme inconnue en ses éléments constituants. Les approches qui utilisent une méthode analytique, telles que la masse d'informations. En plus contenu sémantique des renseignements sur l'âge, le sexe, etc. Le problème majeur est donc un problème de réduction d'information. Que l'on essaye de résoudre par des techniques de traitement du signal : FFT, LPC, etc… [13].

1.5.2. Approche globale

Contrairement à l'approche analytique, le processus de reconnaissance globale, ne tient pas compte de la structure des propriétés phonétiques des mots, seule la forme acoustique est considérée. L'idée de base de cette méthode est de donner au système au moins une image acoustique de chaque mot qu'il devait identifier par la suite. Cette approche nécessite deux phases :

- l'apprentissage : où on constitue le dictionnaire à partir des images acoustiques des mots, et définit le modèle associé à chaque mot.
- les RN : les réseaux de neurones sont en fait, des classificateurs qui présentent l'avantage de pouvoir mémoriser et généraliser des exemples. L'acquisition des connaissances se fait par l'apprentissage et consiste simplement à minimiser une erreur quadratique globale dans l'espace des coefficients de pondérations de manière à rendre la sortie la plus similaire possible à l'entrée.

Un intérêt particulier est accordé à ces modèles que nous utilisons dans l'approche de reconnaissance dans les chapitres suivants.

1.6. Système de reconnaissance

On peut classer les différents systèmes de reconnaissance suivant leurs performances techniques, à savoir la reconnaissance de mots isolés par rapport à parole continue, systèmes mono locuteurs par rapport à systèmes multi locuteur, etc.

1.6.1. Reconnaissance des mots

Il y a plusieurs types de reconnaissance des mots.

1.6.1.1. Reconnaissance des mots isolés

La technique de la reconnaissance des mots isolés est basée sur l'approche globale ou les mots à reconnaître sont comparés à un dictionnaire de références acoustiques conçu lors de la phase d'apprentissage.

On rencontre ce type de reconnaissance aussi bien dans les systèmes de pilotage d'application les plus simples, correspondant à des situations connues à l'avance (jeu de commandes d'un système d'exploitation, réponses à des questions ou à des choix proposés par un serveur vocal, etc.) que dans les applications haut de gamme, telles que la dictée automatique[14].

1.6.1.2. Reconnaissance de parole continue

Cette technique qui apporte un confort d'utilisation indéniable est beaucoup plus complexe que la précédente en raison de phénomène de coarticulation ou de liaisons entre des mots contigus.

1.6.1.3. Reconnaissance de mots enchaînés

La méthode de la reconnaissance de la parole continue ne peut pas être appliquée aux mots enchaînés. En effet une simple analyse de l'enveloppe du signal ne peut donc pas détecter les frontières entre les mots. Pour palier à ce problème, une segmentation des mots s'impose[15].

1.6.2. Reconnaissance de locuteur

On distingue deux systèmes de la reconnaissance.

1.6.2.1. Système mono-locuteur

En raison de la variabilité importante du signal de parole entre locuteurs différents, de nombreux systèmes ne peuvent fonctionner qu'avec un seul locuteur. Les plus simples se contentent de stocker et de rapprocher les différentes prononciations d'un même mot. Ce qui suppose de la part de l'utilisateur un entrainement préalable du système. D'autres possèdent déjà une représentation standard des unités phonétiques, réalisées par le constructeur, complétées par une phase d'apprentissage durant laquelle on améliore le modèle en fonction des caractéristiques de la voix de l'utilisateur.

1.6.2.2. Système multi-locuteur

Le système multi-locuteur est plus précisément adapté aux applications visant un large public. Leur mise au point requiert cependant de la part du constructeur un travail important ; en effet, une liste de mots est présentées à un grand nombre de personnes, c'est la phase d'apprentissage dans le but est de créer un dictionnaire de référence.

1.7. L'usage de la reconnaissance de la parole

La reconnaissance de la parole présente pleine d'intérêts tels que :

- l'amélioration du confort de l'utilisateur : l'un des plus grands avantages que peut offrir cette application est de mettre à la disposition de l'utilisateur une interface avec un dispositif électronique ou mécanique plus simple que ce qui lui est proposé habituellement à savoir le clavier.
- l'augmentation de l'efficacité : la reconnaissance est utile dans des situations couramment appliquées (mains occupées, yeux occupées). Le but ici est d'améliorer la productivité en ouvrant un canal supplémentaire pour la transmission de l'information.
- offre de nouvelles possibilités : dans ce cadre, l'assistance aux personnes handicapées dans l'accomplissement de certaines tâches domestiques tel que l'accès au téléphone s'avère plus que motivante, car l'amélioration apportée à l'utilisateur est la plus évidente.

1.8. Conclusion

Dans ce chapitre nous avons décrit les principaux éléments caractérisant le signal de la parole. Nous avons aussi abordé des particularités de la langue afin d'extraire les caractéristiques acoustiques et articulatoires de ses phonèmes. Les différentes applications de la RAP ont également été présentées les types de reconnaissance qui seront dans les chapitres suivants un élément décision.

2.1.INTRODUCTION :

Nous avons vu dans le chapitre précédent que le signal de parole est un signal complexe et redondant. Il possède une grande variabilité. Pour que le système de reconnaissance de la parole fonctionne efficacement, les informations caractéristiques et invariantes doivent être extraites du signal de parole. Cette procédure consiste à associer au signal de parole une série de vecteurs de paramètres généralement acoustiques, spectraux ou cepstraux. Il existe un grand nombre de paramètres pour représenter le signal de la parole parmi lesquels on peut citer : les paramètres LPC (**L**inear **P**rédictive **C**oefficients), les LPCC (**L**inear **P**rédictive **C**oefficients **C**epstral) et NPC (**N**eural **P**rédictive **C**oding) ou bien encore les coefficients MFCC (**M**el-Frequency **C**epstral **C**oefficients). Que nous allons approfondir dans le cadre de ce travail.

2.2 PRETRAITEMENT DU SIGNAL DE PAROLE :

Avant d'aborder l'analyse acoustique, il est recommandé de faire subir au signal vocal un prétraitement pour lui donner une représentation moins redondante, tout en permettant une extraction assez précise des paramètres pertinents qui caractérisent le signal de la parole. Le prétraitement se présente en plusieurs étapes qui sont schématisées dans la figure 2.1.

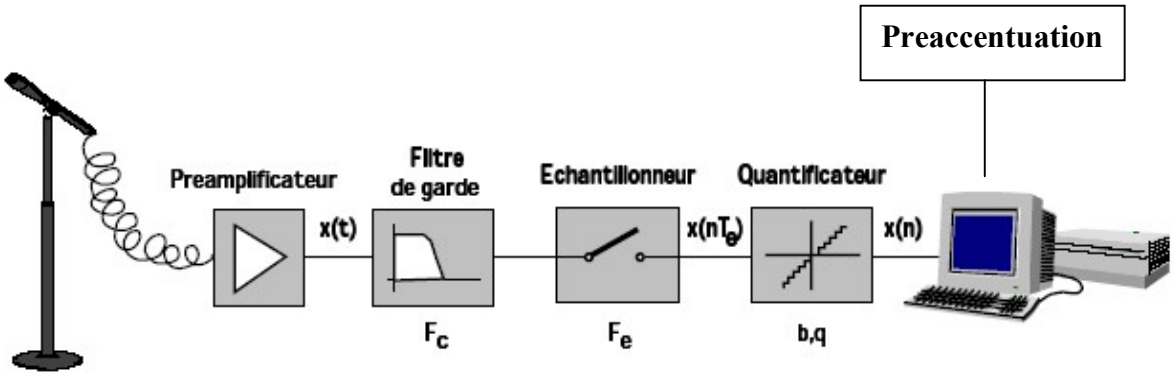

Figure 2.1 : Etapes de prétraitement du signal de parole [16]**.**

2.2.1 Filtre de garde

Afin de réduire le coût du traitement numérique d'une façon notable, on doit limiter le spectre en utilisant un filtre dont la fréquence de coupure f_c est choisie en fonction de la fréquence d'échantillonnage.

2.2.2 Echantillonnage :

Le signal de la parole étant analogique, il s'avère nécessaire de le numériser avant tout traitement. Cette opération consiste en l'échantillonnage du signal qui est présenté dans la figure 2.2.

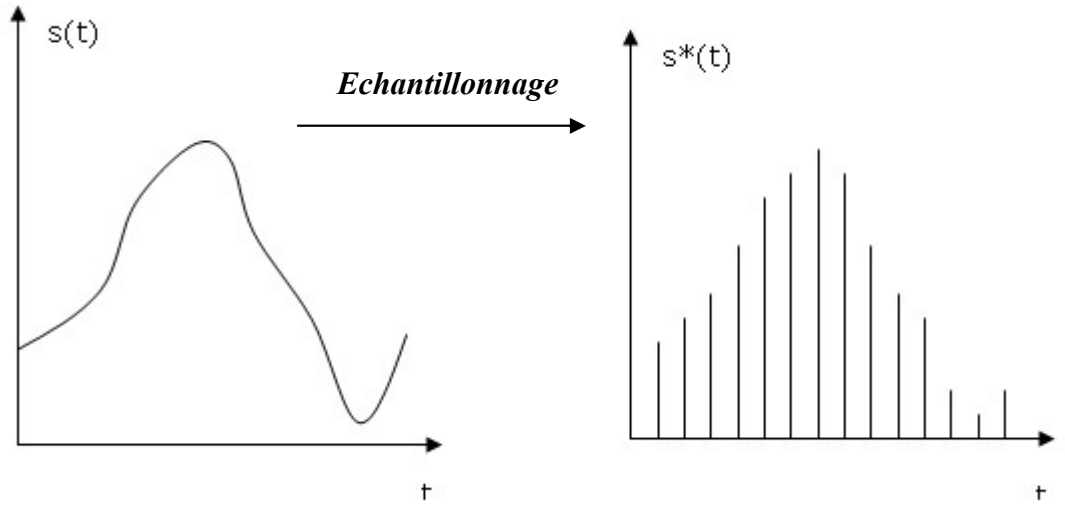

Figure 2.2 : Echantillonnage.

D'après Shannon, la perte d'information entre le signal analogique et le signal discret correspondant est nulle si et seulement si on a :

$$f_e \geq 2 \times f_{\max}$$

f_e : la fréquence d'échantillonnage.

$f_{\max}$: la fréquence maximale que contient le signal à traiter.

2.2.3 La quantification :

La quantification définit le nombre de bits sur lesquels on veut réaliser la numérisation. Elle permet de mesurer l'amplitude de l'onde sonore à chaque pas de l'échantillonnage. Le choix de la fréquence d'échantillonnage est aussi déterminant pour la définition de la bande passante représentée dans le signal numérisé qui est présenté dans la figure 2.3.

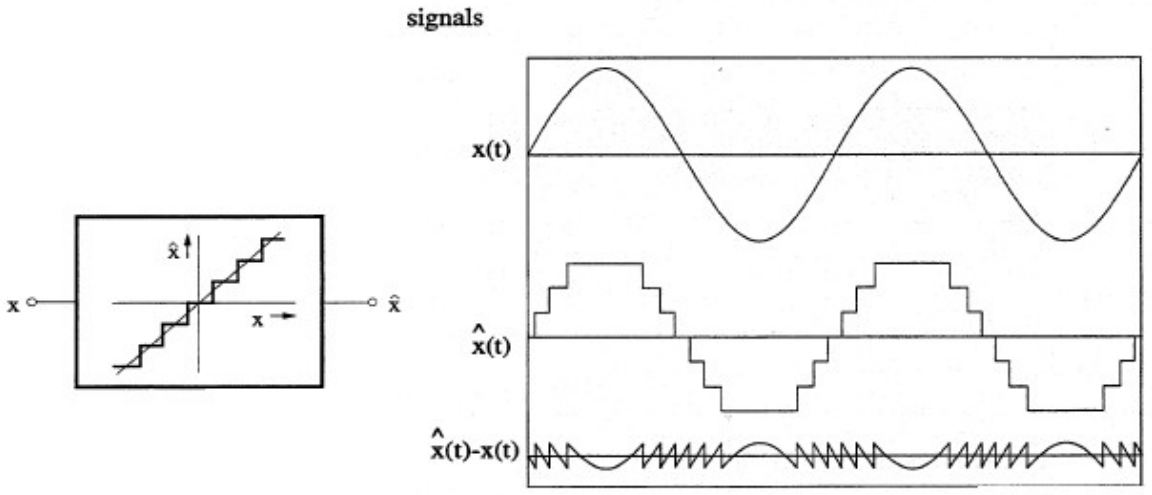

Figure 2.3 : La quantification.

2.2.4 Préaccentuation :

En général, le signal vocal se caractérise par une perte de 6 dB/octave, due à l'influence de la source d'excitation et au rayonnement des lèvres. Une perte de 6 dB/octave veut dire que les hautes fréquences ont une énergie plus faible que celle des basses fréquences. Pour palier à cet inconvénient la préaccentuation permet d'égaliser les sons aigus avec les sons graves **(**voir figure (2.4)).

L'opération consiste à faire passer le signal à travers un filtre de transmittance

$$H(z) = 1 - az^{-1} \qquad (2.00)$$

Le facteur de préaccentuation est pris entre 0.9 et 1 (souvent 0.95). comme conséquence, la préaccentuation introduit une légère distorsion spectrale [17].

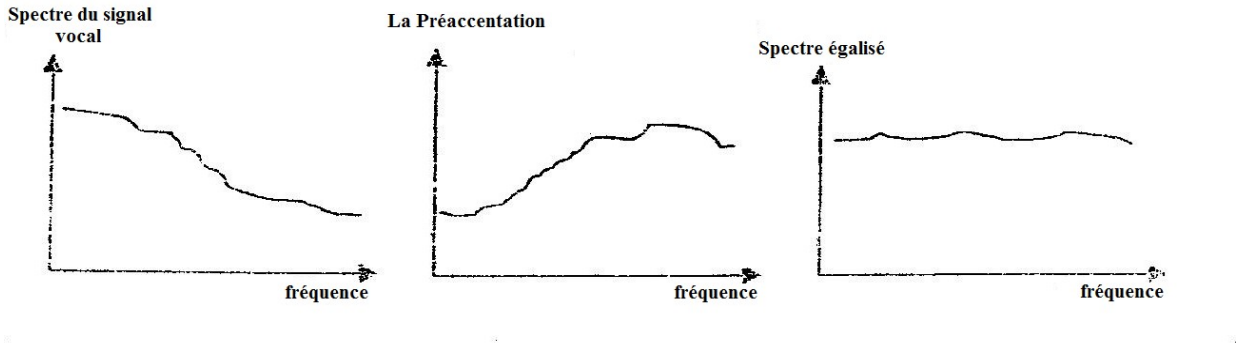

Figure 2.4 : la préaccentuation.

2.3 Extraction des paramètres à partir des coefficients MFCC (Mel Frequency Cepstral Coefficients)

La masse d'informations résultantes du signal vocal après sa numérisation véhicule une grande quantité d'éléments. Le problème majeur est la réduction de ce nombre. On essaye alors de résoudre cette tare par des techniques de traitement du signal pour sa réduction tout en retenant les paramètres pertinents.

Dans ce qui suit nous allons expliquer le choix ainsi que les techniques utilisées pour l'extraction des caractéristiques du signal vocal**.**

Le principe de base est que le signal vocal S(n) est produit par un signal excitant g(n) de la source glottique (qui représente l'élément commun à tous les phonèmes) traversant un système linéaire passif (qui représente l'élément discriminant de chaque phonème) dont la fonction de transfert est donnée par l'équation (2.01) de réponse impulsionnelle b(n) qui a la contribution du conduit.

On peut écrire :

$$S(n)=g(n) * b(n) \qquad (2.01)$$

Pour déconvoluer plus aisément S(n), il suffit de transposer le problème par homomorphisme. Pour cela, nous passons dans le domaine spectral par transformée de Fourier pour obtenir :

$$S(f)=G(f).B(f) \qquad (2.02)$$

S(f) : transformée de Fourier du signal S(n).

G(f) : transformée de Fourier du signal g(n).

B(f) : transformée de Fourier du signal b(n).

En prenant le logarithme, on obtient alors la somme des termes :

$$\log|S(f)| = \log|G(f)| + \log|B(f)| \qquad (2.03)$$

Cette opération nous permettra de séparer l'élément discriminant par rapport a celui qui n'est pas discriminent

En pratique, cette transposition par homomorphisme est réalisée par les étapes suivantes :

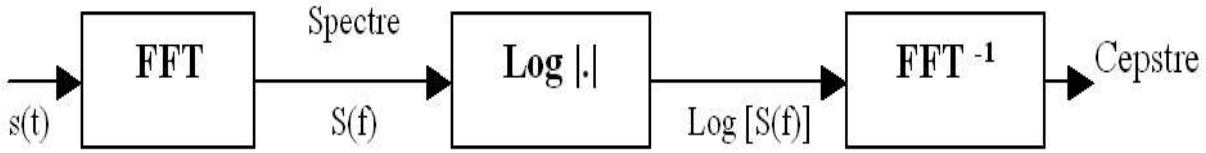

Figure 2.5 : Différentes étapes de l'analyse cepstrale.

$$S^{+}(n) = g^{+}(n) + b^{+}(n) \qquad (2.04)$$

Les $S^{+}(n)$ sont les coefficients cepstraux linéaires prenant leurs valeurs dans un domaine pseudo temporel réel appelée quéfrence [18].

La structure du signal de la parole permet de dire que :

- $g^{+}(n)$ se réduit théoriquement à une séquence d'impulsions espacées de $\boldsymbol{n_0 = f_e / f_0}$ échantillons ou $\boldsymbol{f_e}$ est la fréquence d'échantillonnage et $\boldsymbol{f_0}$ est la fréquence du fondamental.
- $b^{+}(n)$ décrois rapidement en 1/n et devient négligeable pour n>n_0.

Dans ces conditions, on peut admettre que la première valeur du cepstre reflète la contribution du conduit, alors que les pics périodiques au delà de n_0 reflètent les impulsions de la source.

Pour séparer ces deux contributions, on peut utiliser un simple filtre passe bas qui permettrait d'éliminer la composante de la période fondamentale.

Pour clarifier et visualiser les étapes pour aboutir aux résultats de l'extraction des caractéristiques on développera dans ce qui suit chaque bloc dans le schéma synoptique représenté dans la figure (2.6).

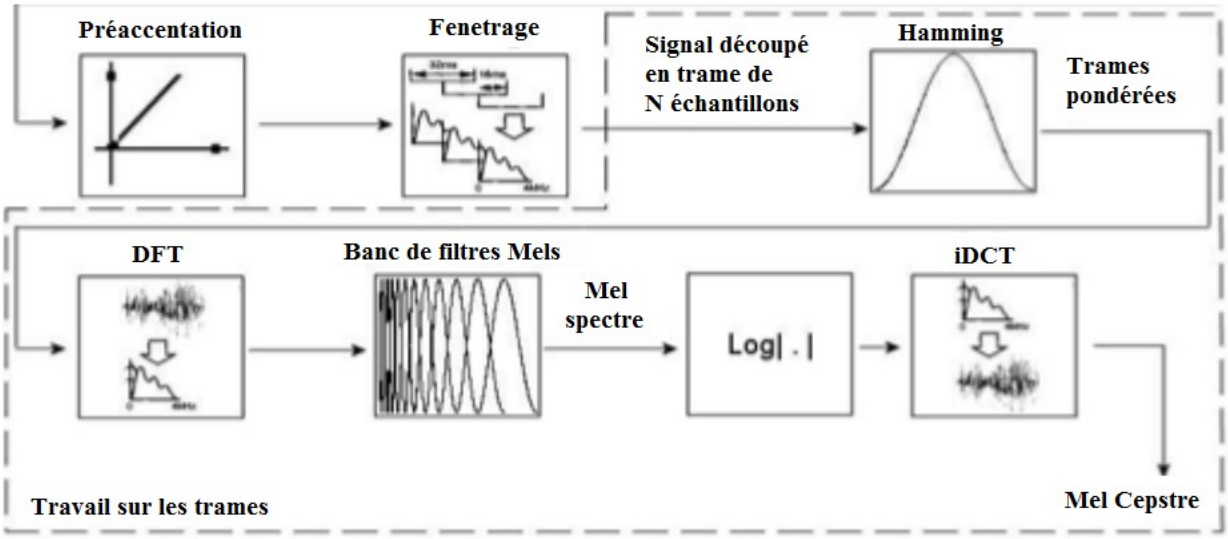

Figure 2.6 : Schéma synoptique des étapes d'extraction des paramètres MFCC [11].

2.3.1 Fenêtrage :

Le signal vocal est un signal non stationnaire; il présente une évolution lente dans le temps. Le but du fenêtrage est de découper le signal de parole en petites tranches (chacune de durée 20 ms environ) ou il peut être considéré localement comme quasi- stationnaire.

En outre, et pour l'évolution lente du signal vocal, le fenêtrage permet le traitement en temps réel et il facilite aussi l'analyse des signaux sur la machine. Les ressources d'une machine étant limitées, le signal ne peut pas être traité dans sa globalité.

Il existe plusieurs types de fenêtres d'analyse on représente quelques unes :

- **Fenêtre rectangulaire :**

Elle est définie par :

$$f(nT)=\begin{cases}1 \xrightarrow{Si} |nT| < T'. \\ 0 \xrightarrow{Si} ailleurs.\end{cases} \qquad (2.05)$$

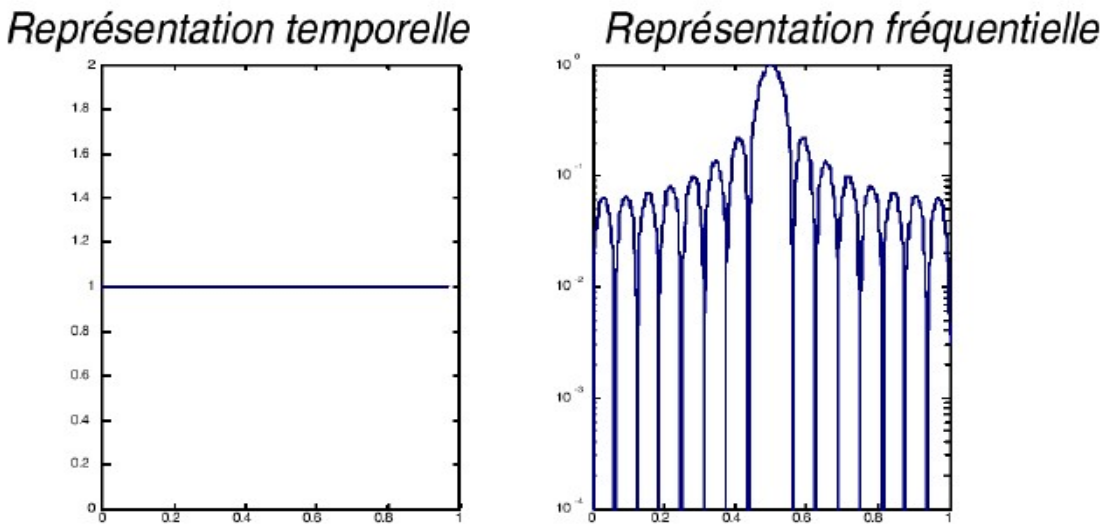

Figure 2.7 : Fenêtre rectangulaire et son spectre.

❖ **Fenêtre de Hanning :**

Elle est définie par :

$$f(nT)=\begin{cases} 0.5\left(1-\cos\left(\dfrac{n\pi T}{T'}\right)\right) \xrightarrow{Si} |nT| < T' \\ 0 \xrightarrow{Si} ailleurs. \end{cases} \qquad (2.06)$$

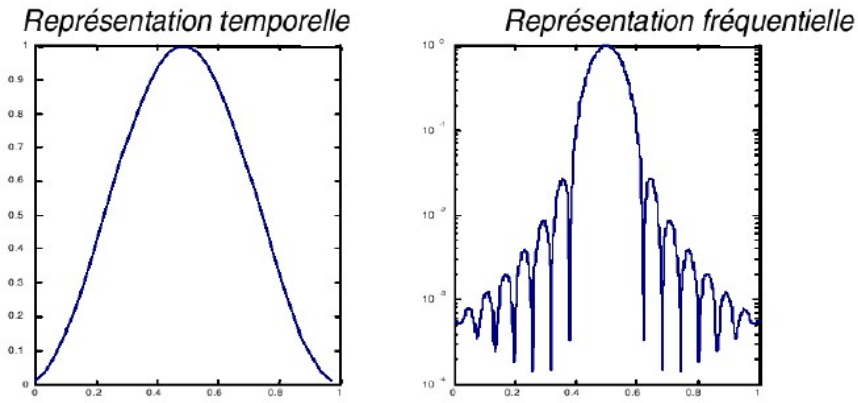

Figure 2.8 : Fenêtre de Hanning et son spectre.

❖ **Fenêtre de Hamming :**

Elle est définie par :

$$\mathbf{f(nT)}=\begin{cases}0.54-0.46.\cos\left(\dfrac{\pi nT}{T'}\right) \xrightarrow{Si} |nT| < T'. \\ 0 \xrightarrow{Si} ailleurs\end{cases} \qquad (2.07)$$

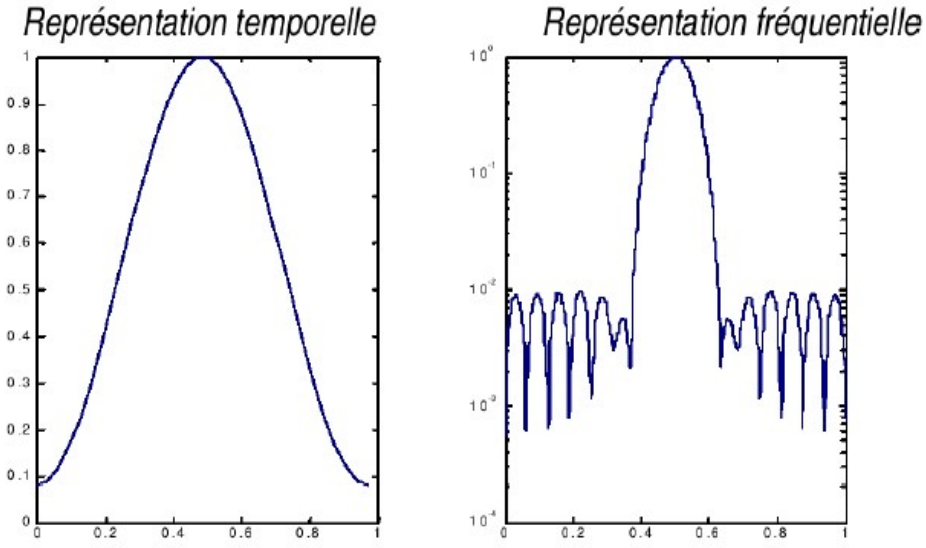

Figure 2.9: Fenêtre de Hamming et son spectre.

Parmi ces fenêtres, la fenêtre de hamming est la plus convenable à la parole, car elle entraîne un minimum de distorsion spectrale du signal de parole, par rapport aux autres fenêtres. (Atténuation du rapport du lobe principal au lobe secondaire est égale à - 41dB, c'est-à-dire que la concentration d'énergie dans le lobe principal est égale à 99.96%).

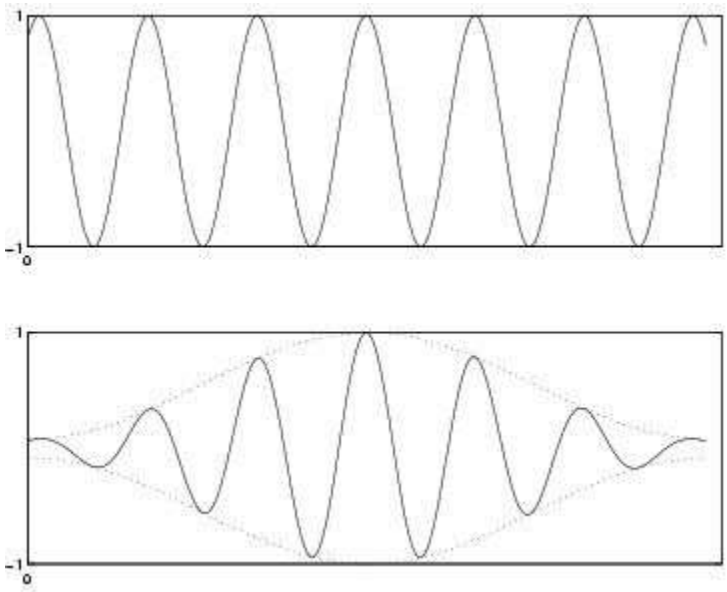

Figure 2.10 : Représentation d'un signal sinusoïdale non pondéré puis pondéré par la fenêtre de Hamming.

2.3.2 FFT (Fast Fourier Transform) :

La transformée de Fourier rapide (acronyme anglais : *FFT* ou *Fast Fourier Transform*) est un algorithme de calcul de la transformée de Fourier discrète (TFD). Ainsi, pour le temps de calcul de l'algorithme rapide peut être 100 fois plus petit que le calcul utilisant la formule de définition de la TFD.

Cet algorithme est couramment utilisé en traitement numérique du signal pour transformer des données du domaine temporel au domaine fréquentiel.

2.3.3 Banc de filtres Mels :

L'échelle spectrale dans le calcul du cepstre est linéaire en fréquence. Cependant, les études physiques et perceptives de l'oreille indiquent qu'elle est sensible à une échelle presque logarithmique de la fréquence [19].

Pour s'approcher donc du modèle de l'audition, on fait appel a une échelle pseudo logarithmique appelée échelle MEL linéaire pour des fréquences allant de 0 à 1 KHz, et logarithmique au delà.

2.3.4 Calcul des coefficients dans l'échelle MEL :

Les paramètres MFCC utilisent une échelle fréquentielle non linéaire.

La fréquence Mel-échelle est définie par :

$$B(f) = 2595 \log\left(1 + \frac{f}{700}\right) \qquad (2.08)$$

Où ***f*** est la fréquence en Hz, ***B(f)*** est la fréquence Mel-échelle de *f*.

Soit un signal discret $\{s[n]\}$ avec $0 < \delta(n) < \delta(N-1)$, N est le nombre d'échantillons d'une fenêtre

analysée, F_s est la fréquence d'échantillonnage, la transformée de Fourier discrète $S[k]$ est obtenue :

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j2\pi nk/N} \qquad avec \quad 0 \le k < N \qquad (2.09)$$

Le spectre du signal est multiplié avec des filtres triangulaires dont les bandes passantes sont équivalentes en domaine mel-fréquence. Les points frontières $B[m]$ des filtres en Mel-fréquence sont calculés ainsi :

$$B[m] = B(f_1) + m\frac{B(f_h) - B(f_1)}{M+1} \qquad 0 \le m \le M+1 \qquad (2.10)$$

Avec :

M : le nombre de filtres.

f_h : la fréquence la plus haute.

f_l : la fréquence la plus basse pour le traitement du signal.

Dans le domaine fréquentiel, les points f [m] discrets correspondants sont calculés par l'équation:

$$f[m]=\left[\frac{N}{F_S}\right]B^{-1}\left[B(f_1)+m\frac{B(f_h)-B(f_1)}{M+1}\right] \qquad (2.11)$$

Où B^{-1} est la transformée de Mel-fréquence en fréquence.

$$B^{-1}(b)=700*\left(10^{b/2595}-1\right) \qquad (2\text{-}12)$$

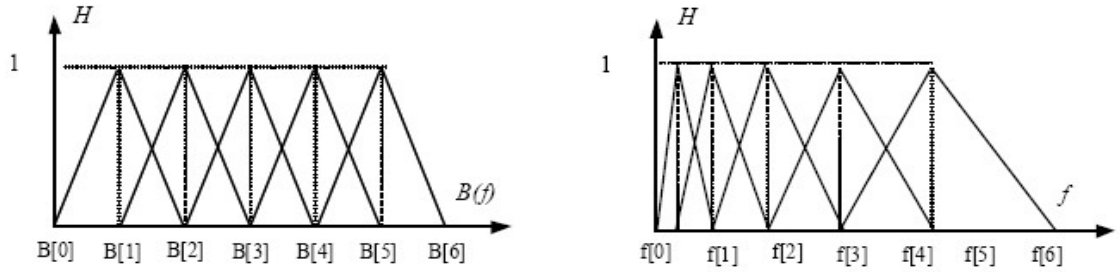

Figure 2.11 : Les filtres triangulaires passe-bande en Mel-fréquence **(B(*f*))** et en fréquence **(*f*).**

Le coefficient $H_m[k]$ de chaque filtre est déterminé par le système suivant :

$$H_m[k]=\begin{cases}0 & si \quad k\le f[m-1]\\ \dfrac{k-f[m-1]}{f[m]-f[m-1]} & si \quad f[m-1]\le k\le f[m]\\ \dfrac{f[m+1]-k}{f[m+1]-f[m]} & si \quad f[m]\le k\le f[m+1]\\ 0 & si \quad k\ge f[m+1]\end{cases} \qquad (2.13)$$

Pour un spectre lissé et stable, à la sortie des filtres un logarithme d'énergie (ou un logarithme de spectre d'amplitude) est calculé [16] :

$$E[m]=\log\left[\sum_{k=0}^{N-1}|S[k]|^2 H_m[k]\right] \qquad \text{avec} \quad 0\le m<M \qquad (2.14)$$

2.3.5 Les coefficients Cepstraux :

C'est l'étape finale, on transforme les données dans l'échelle des Mels, fréquentielle donc vers l'échelle des temps. Le résultat de cette étape sera les MFCC proprement dit. Il suffit d'effectuer l'inverse de la transformée de Fourier. Dans la pratique, on effectue une transformée en FFT $^{-1}$ (DCT $^{-1}$) indiquées dans les équations (2.15) et (2.16), ce qui revient au même puisque la transformée en Cosinus inverse donne la partie réelle de la transformée de Fourier ; or ici on a que des réels. Il faut noter que la transformée en sinus donnera la partie imaginaire de la transformée de Fourier.

$$C[n] = \sum_{m=0}^{M-1} E[m] \cos\left(\frac{\pi n\left(m+\frac{1}{2}\right)}{M}\right) \quad \text{Avec} \quad 0 \le n < M \qquad (2.15)$$

2.5 CONCLUSION :

L'extraction des caractéristiques du signal de parole est une étape fondamentale dans le traitement du signal vocal. Dans notre travail nous nous sommes intéressés à l'étude des paramètres MFCC.

Et on voit l'étude des paramètres MFCC se réalise par plusieurs étapes comme suit :

Etape 1 : Découper le signal en plusieurs fenêtres qui se recoupent entre elles. Nous appliquerons la MFCC à chaque fenêtre.

Etape 2 : Afin de diminuer la distortion spectrale nous appliquons une fenêtre de Hamming au signal. Par la suite nous multiplions cette fonction par le signal à transformer, nous minimisons ainsi la distortion spectrale créée par le recoupement.

Etape 3 : Appliquer ensuite la FFT à la fenêtre pour en ressortir l'amplitude, on obtient donc le spectre.

Etape 4 : On passe à l'échelle de Mel. En effet, après des études sur l'ouie humaine, il a été montré que l'homme se base sur une échelle fréquentielle spécifique, pour simuler l'oreille humaine, il faut passer par un banc de filtre, un filtre pour chaque fréquence que l'on cherche. ; Ces filtres ont une réponse de bande passante triangulaire.

Etape 5 : Pour finir, nous travaillons avec le Cepstre, nous convertissons le spectre logarithmique de Mel en temps au moyen de la DCT (Discret Cosinus Transforme), FFT (Fast Fourier Transforme).

3.1 Présentation de Langage de programmation utilisé :

Pour l'implémentation et le développement de l'application nous avons utilisé le langage Matlab.

Le langage Matlab :

MATLAB (abréviation de *Matrix LABoratory*) est un environnement complet, ouvert et extensible pour le calcul et la visualisation. Il dispose de plusieurs fonctions mathématiques, scientifiques et techniques. L'approche matricielle de MATLAB permet de traiter les données sans aucune limitation de taille et de réaliser des calculs numériques et symboliques de façon fiable et rapide.

Quelles sont les particularités de MATLAB ?

MATLAB permet le travail interactif soit en mode commande, soit en mode programmation, tout en ayant toujours la possibilité de faire des visualisations graphiques. MATLAB possède les particularitéssuivantes:

- la programmation facile,
- la continuité parmi les valeurs entières, réelles et complexes,
- la gamme étendue des nombres et leurs précisions,
- la bibliothèque mathématique très compréhensive,
- l'outil graphique qui inclut les fonctions d'interface graphique et les utilitaires,
- la possibilité de liaison avec les autres langages classiques de programmations (C ou Fortran).

Matlab permet d'écrire assez simplement une interface graphique pour faire une application

Interactive utilisable par des utilisateurs. Une interface graphique comprend des menus, des boutons, des "ascenseurs", des cases à cocher, des listes de choix, des zones de texte.

3.2 Architecture du système de reconnaissance :

L'architecture du système de reconnaissance comprend les modules suivants (fig 3.1).

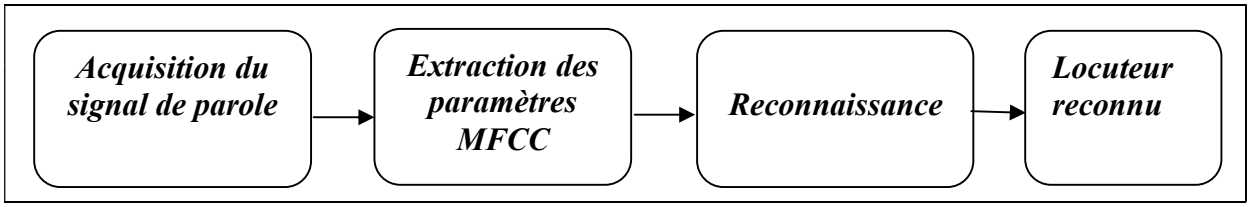

Figure 3.1 : Schéma global du système de reconnaissance.

3.3 Calcul des paramètres MFCC :

La Reconnaissance Automatique de la Parole (RAP) repose sur l'extraction des paramètres du signal acquis. La méthode d'analyse que nous avons choisie est l'analyse cepstrale (chapitre 2), en prenant les MFCC, le calcul de ces derniers est schématisé par un organigramme (fig 3.2).

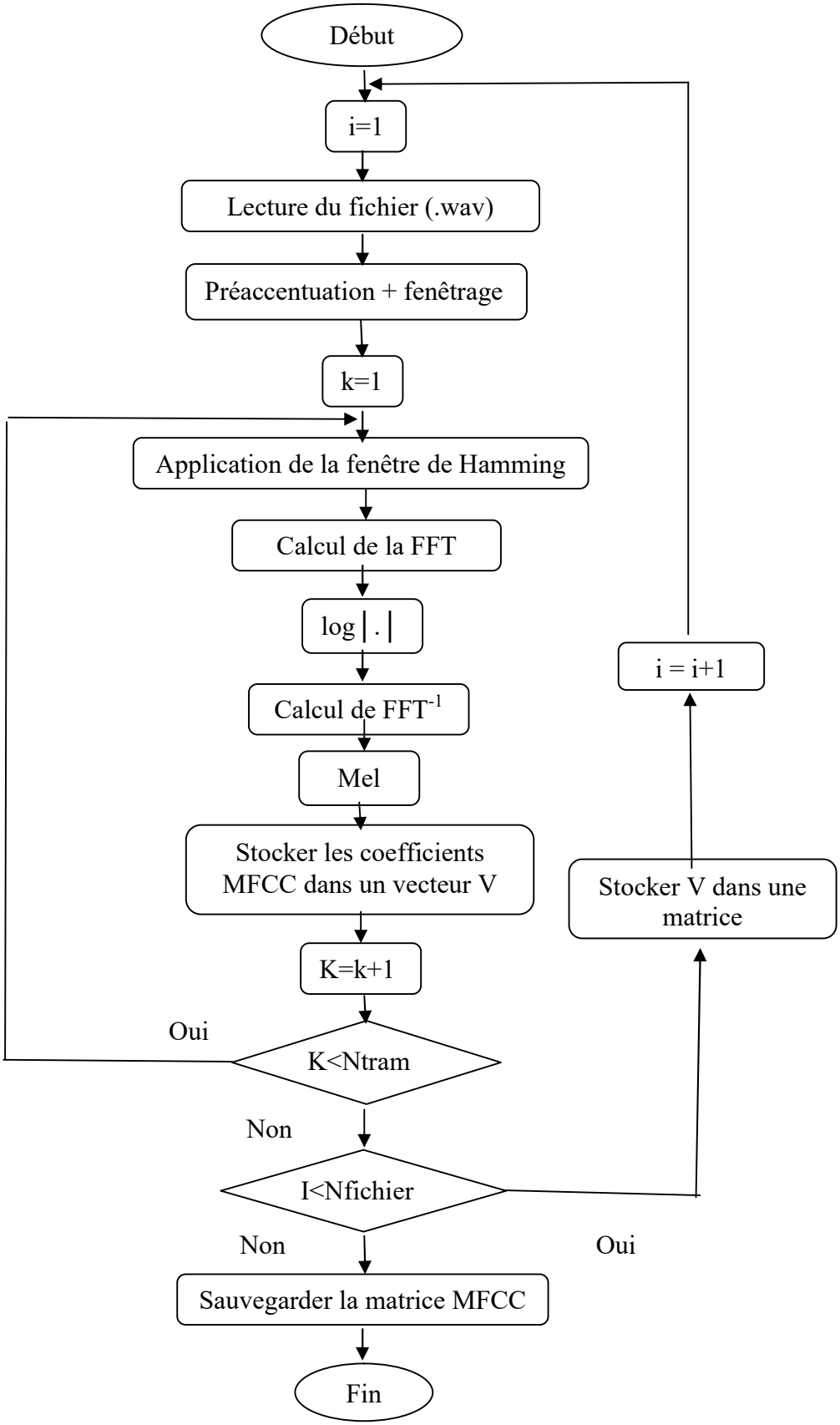

Figure 3.2 : Organigramme de l'extraction des paramètres MFCC.

Avec : **k** **:** numéro de la trame.

Ntram **:** nombre de la trame.

i **:** numéro de fichier à traiter.

Nfichier : nombre de fichiers.

V **:** vecteur du paramètre MFCC.

3.4 Interfaces de l'application :

Nous avons créé deux applications, la première application pour le taux de reconnaissance de l'ordinateur et la deuxième pour le teste avec d'autres fichiers vocaux (voir Fig 3.3 ,Fig 3.4).

Nous allons découvrir l'espace de travail de notre application :

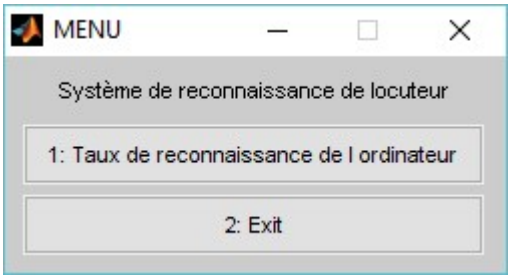

Figure 3.3: Interface de l'application 1.

Figure 3.4: Interface de l'application 2.

3.5 Plan de travail :

✓ 1ere application :

Avant de commencer la 1ère application il faut d'abord enregistrer deux fichiers vocaux**(.wav)** pour chaque personne, on met un dans le dossier d'entrainement **(train)** et l'autre dans le dossier de test **(test).**

Et on fait ça avec le logiciel **SFS** (Speech Filing Sytem).

En passant par les étapes suivantes : (fig 3.5,fig 3.6)

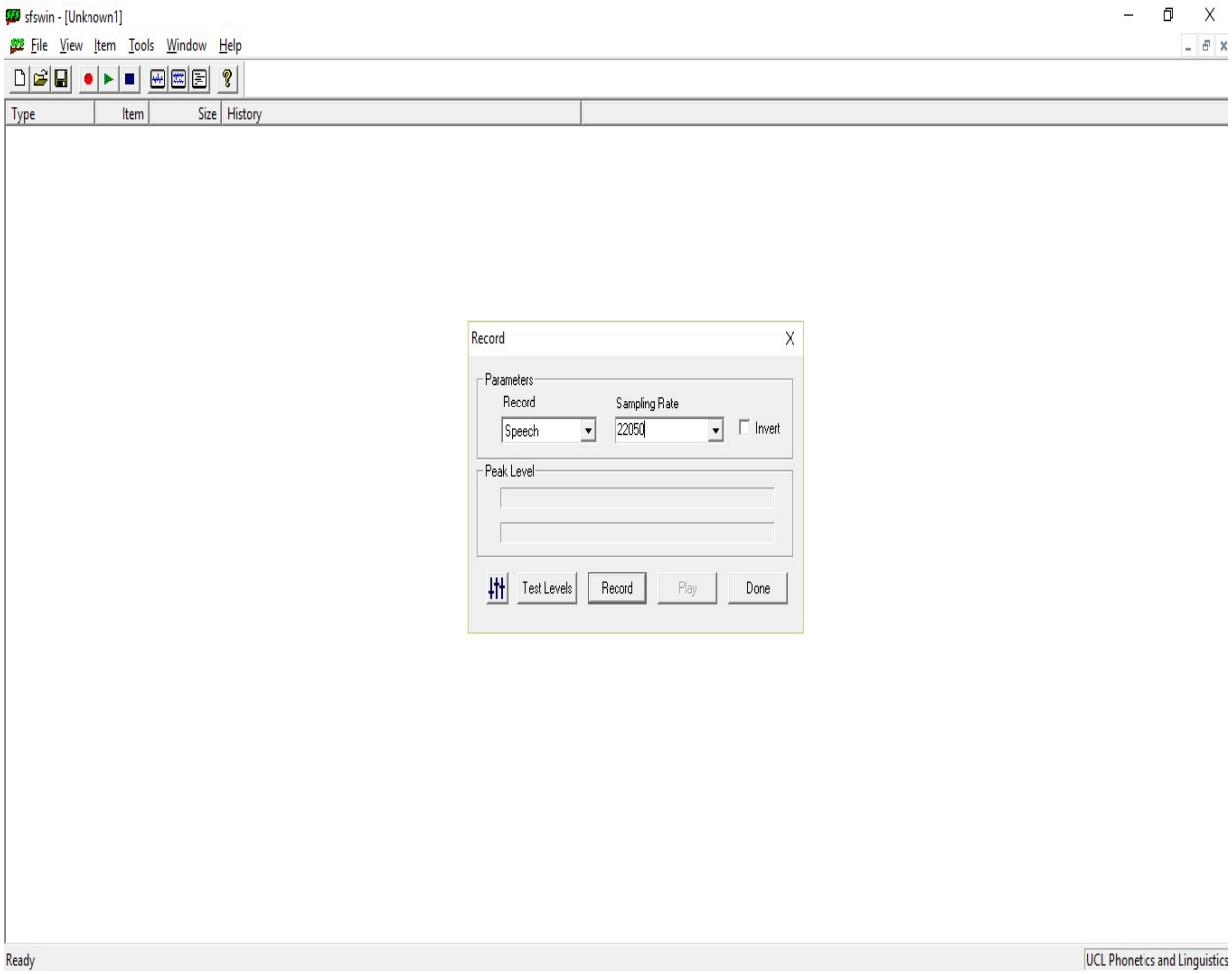

Figure 3.5: Fenêtre principale du logiciel SFS.

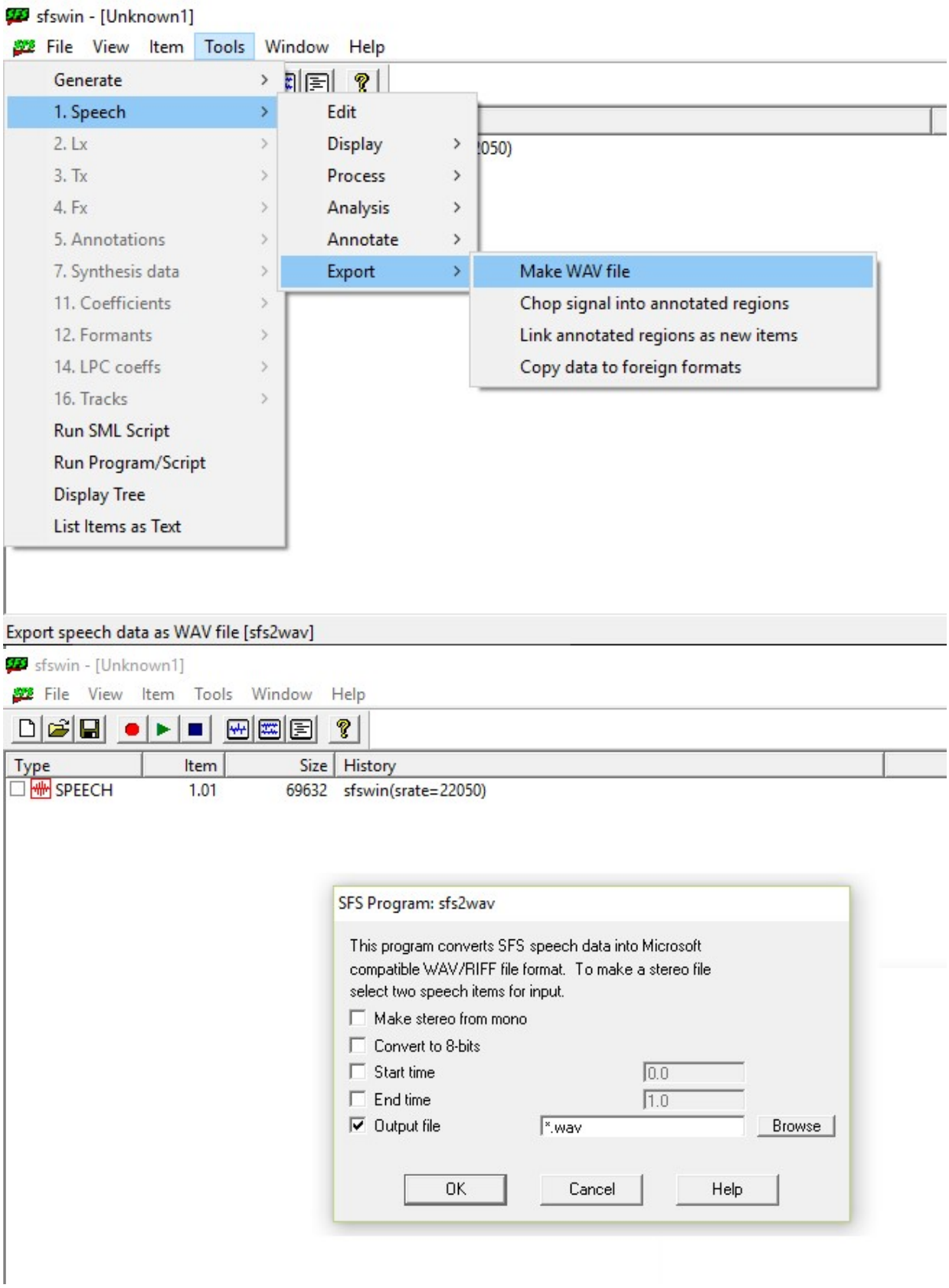

Figure 3.6 : Interface du logiciel SFSWAV.

Après la conversion les fichiers vocaux en format **.wav,** on clique sur le bouton 1 **(taux de reconnaissance de l'ordinateur)** et on voir ça dans les figures suivantes (fig 3.7,fig 3.8,fig 3.9,fig 3.10).

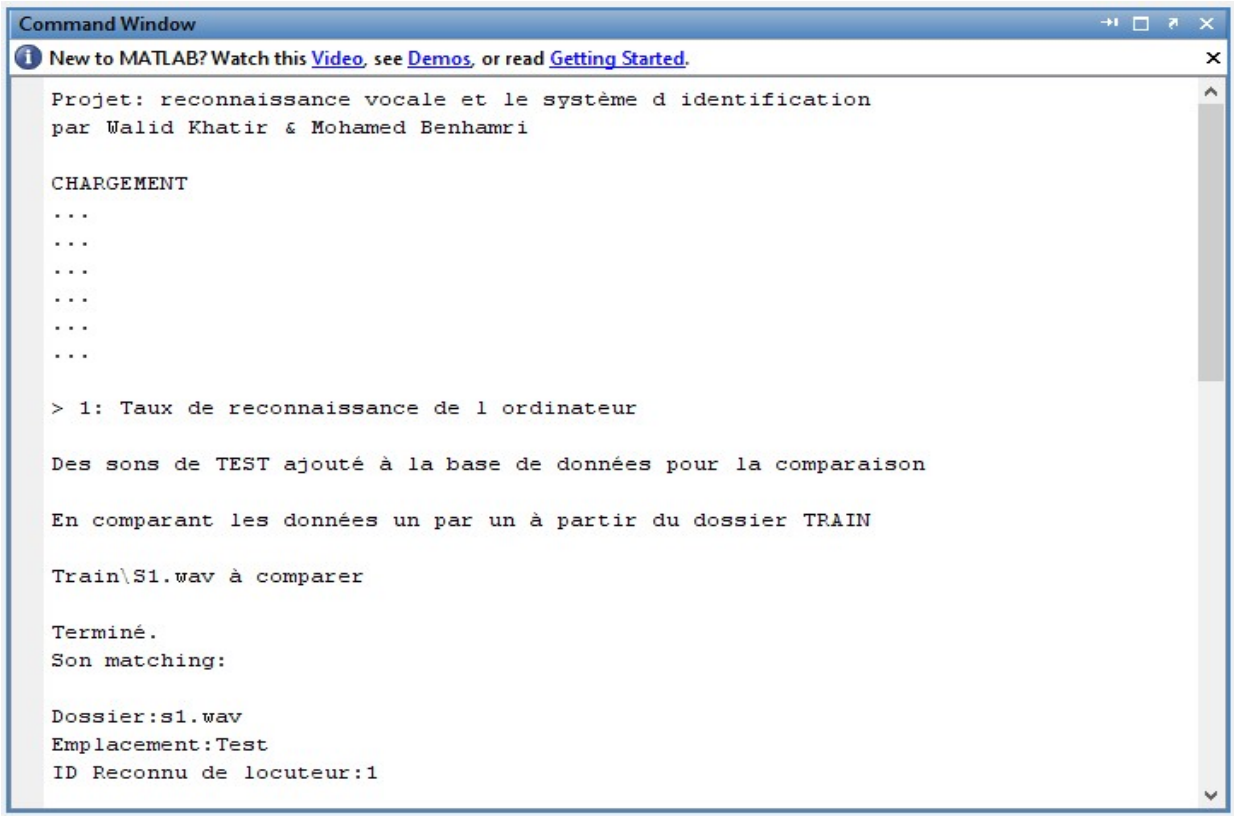

Figure 3.7 : Résultats du test.

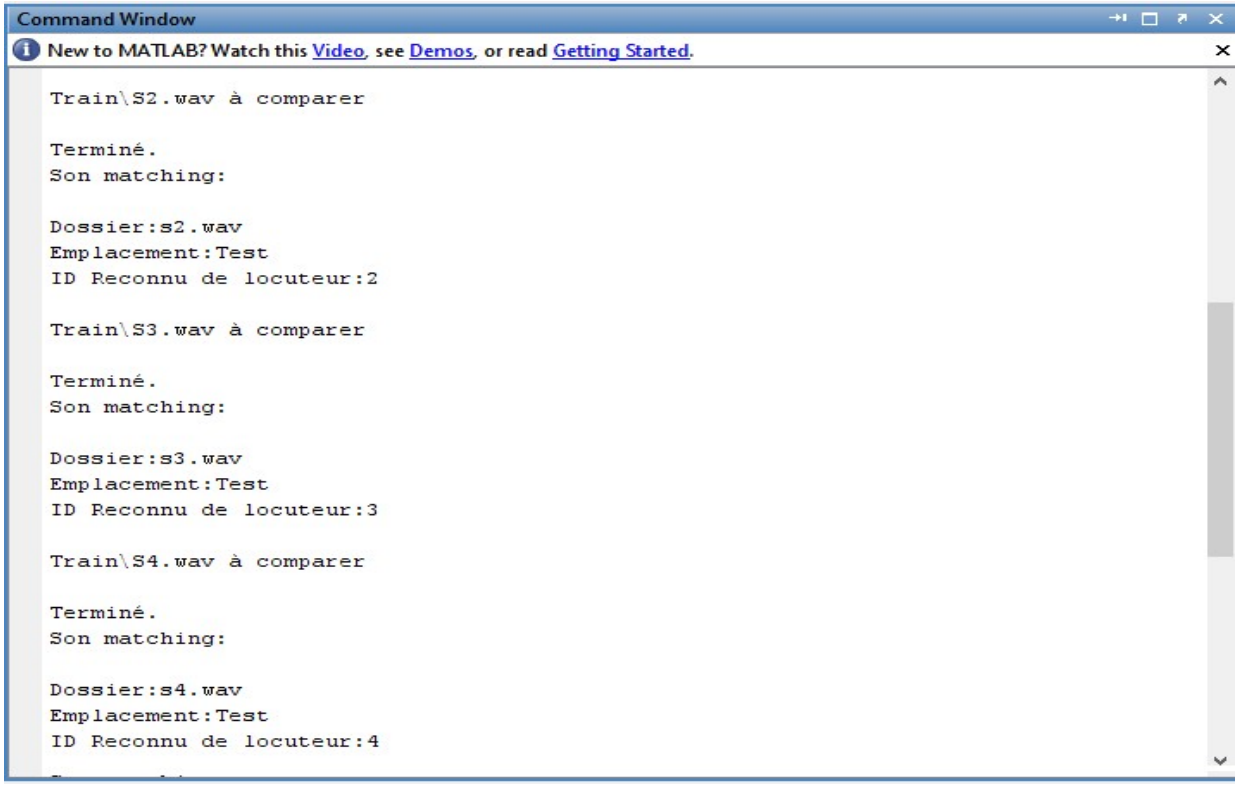

Figure 3.8 : Résultats du test.

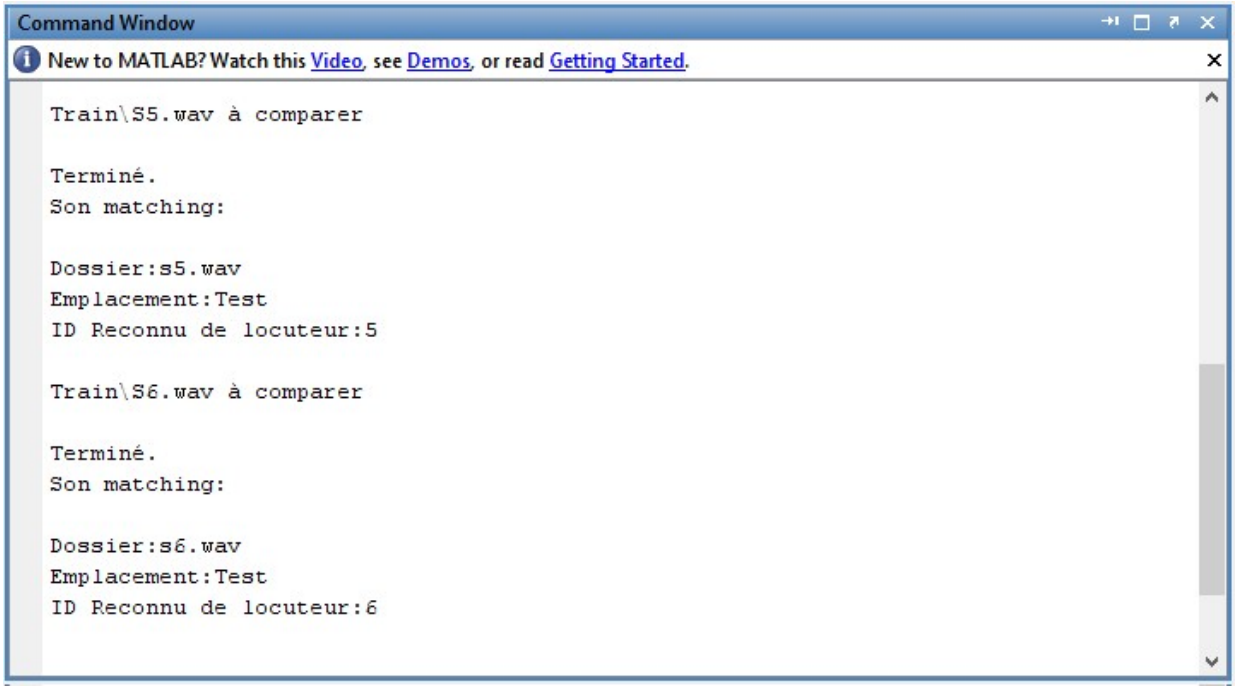

Figure 3.9 : Résultats du test.

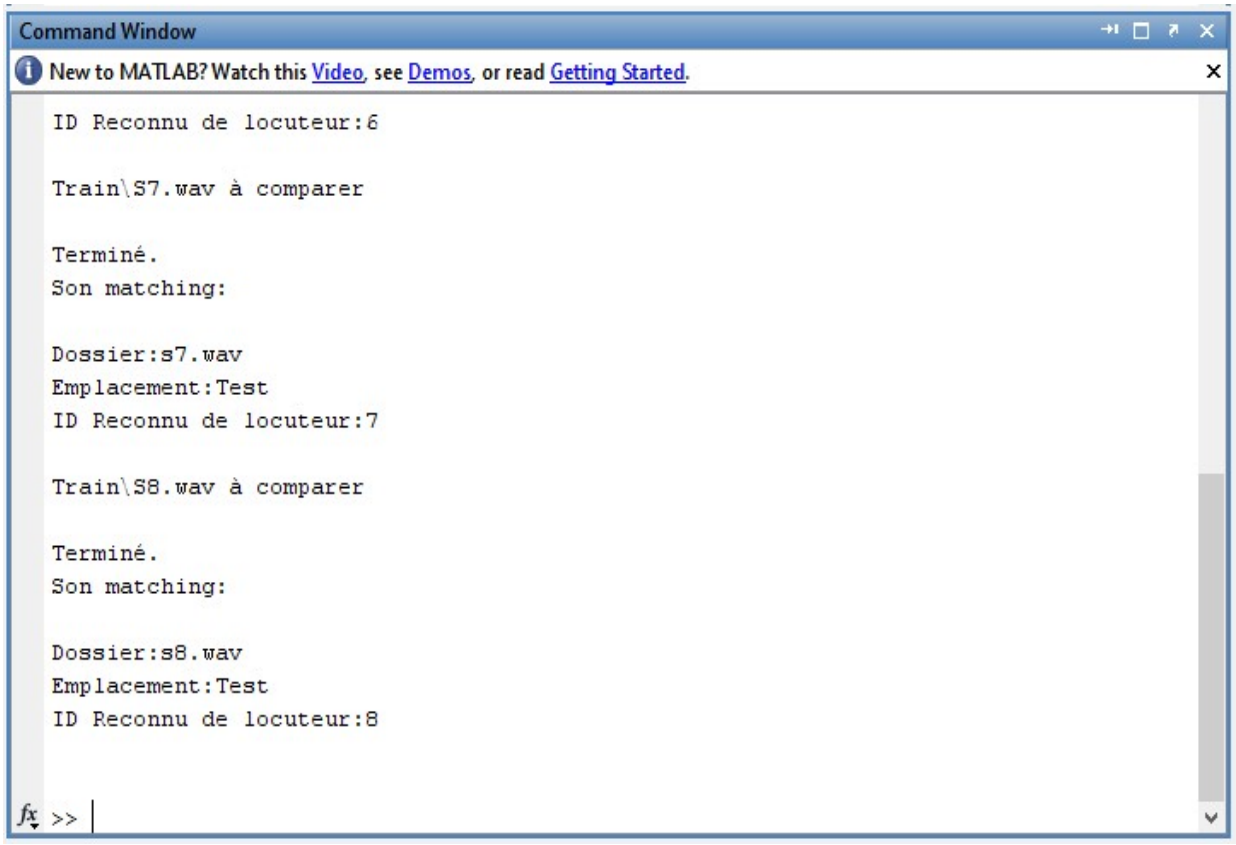

Figure 3.10 : Résultats du test.

✓ 2 ème application:

Description d'espace de travail :

Nous allons découvrir l'espace de travail de notre application :

1er Bouton pour ajouter un nouveau son à la base de données à partir du Microphone.
2ème Bouton pour la reconnaissance de locuteur à partir du microphone.
3ème Bouton pour afficher les informations de la base de données.
4ème Bouton pour supprimer la base de données.
5ème Bouton pour sortir.

Pour entamer cette application, il faut qu'il y ait des fichiers vocaux déjà enregistré dans la base de données, et pour faire cela on clique sur le bouton 1 et on passe par les étapes suivantes :

- On donne un numéro ID pour le locuteur.
- On parle dans le microphone.

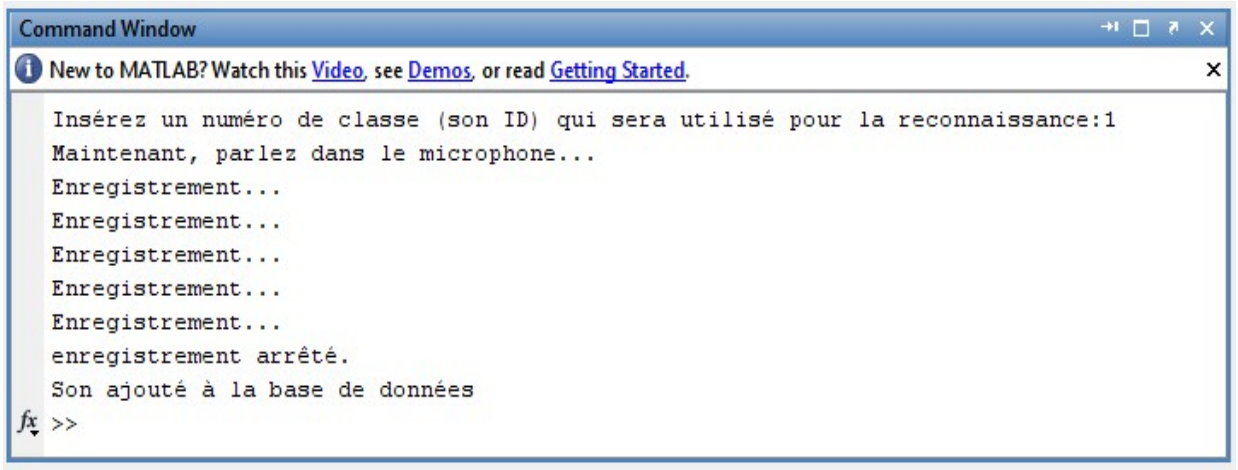

Figure 3.11 : Les étapes pour ajouter un son à la base de données.

Et pour faire la reconnaissance d'un locuteur il faut cliquer sur le bouton 2 et Parler dans le microphone.

Et on voit le résultat en bas de la fenêtre.

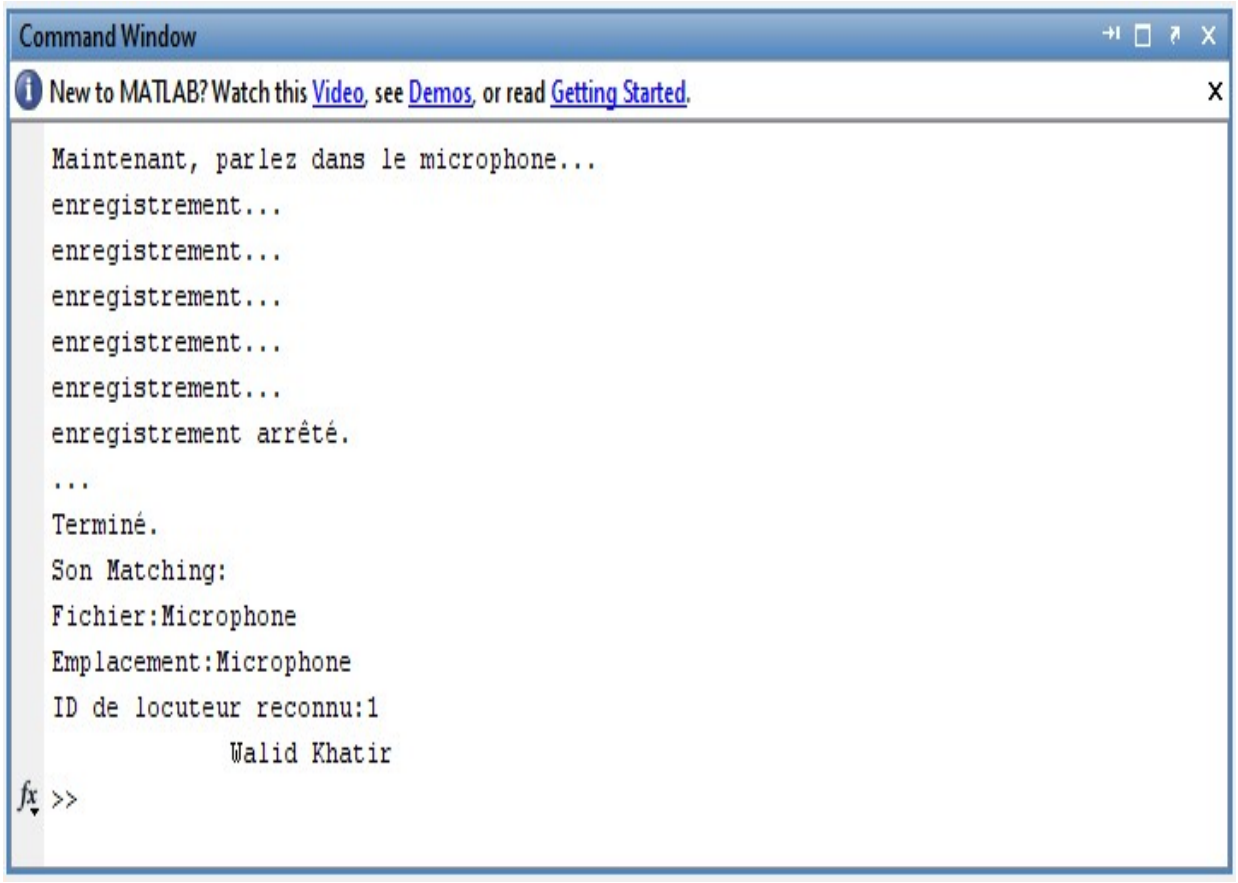

Figure 3.12 : Résultats du test de l'application 2.

3.6 Résultats expérimentaux :

L'application a été testée sur 10 personnes (voir les figures), le temps d'exécution pour chaque personne est 5 secondes. Le taux de réussite est de 90%, c'est à dire 9 sur 10 personnes ont été reconnues correctement. Les résultats trouvés sont enregistré automatiquement dans la base de données.

Figure 3.13 : Résultats du test de l'application 2 pour le locuteur 1.z ;ç

Figure 3.14 : Résultats du test de l'application 2 pour le locuteur 2.

3.7 Conclusion :

Le passage de l'algorithme a l'application est la phase la plus importante, d'après plusieurs tests nous avons réussi a trouvé une méthode de reconnaissance automatique de locuteur. Les résultats présentés montre bien la fiabilité de notre application.

Au cours de notre travail, nous avons réalisé une application dédiée à la reconnaissance automatique de locuteur, en utilisant un microphone relié à l'ordinateur.

En premier lieu, nous avons présenté quelques notions de base de la parole.

En deuxième lieu, nous avons introduit les différentes méthodes utilisant pour la reconnaissance automatique de locuteur (MFCC).

A la fin nous avons décrit l'espace de travail de notre application, ainsi que les différents résultats obtenus.

A travers ce travail, nous avons pu acquérir beaucoup de la connaissance dans le domaine de la parole. Premièrement, nous avons utilisé les outils appris au cours de notre formation, et enrichi nos connaissances dans le domaine de la parole. On a aussi appris à utiliser le langage de programmation Matlab.

Bibliographie

[1] https://en.wikipedia.org/wiki/Speaker_recognition.

[2] research.jyu.fi/phonfr/10.html

[3] http://www.claudegabriel.be/Cine%20acoustique%209.pdf

[4] tpelangage.e-monsite.com/pages/l-origine-physiologique/c-l-appariel-phonatoire.html

[5] portal.tpu.ru/shared/i/itn/student/tab5/phonetique%20articulatoire.pdf

[6] tcts.fpms.ac.be/cours/1005-07-08/speech/parole.pdf Cour **T. Dutoit**, Mons, le 20 octobre 2000 «Introduction au Traitement Automatique de la Parole».

[7] : J.P.ZERLING « Articulation et coarticulation dans les groupes occlusive-voyelle en français », thèse de doctorat de 3 éme cycle, Université de Nancy2, Nancy(France) **(1979).**

[8]: B.H.JUENG, L.R.RABINER, and J.G.WILPON « On the use of band pass liftering in speech recognition» .IEEE. « Translation on acoustic and speech signal processing, 35 (7): 947-954, **(1987).**

[9]: Cours codage 2 année master USTO.

[10] : M.SAHRAOUI, DJ.BOUMAZA « Traitement de signal de la parole par le filtre de Kalman, LPCC et LPC », thèse d'ingénieur, institut d'électronique, Université de Blida **(2004).**

[11] : Rachedi Julien Mémoire Master 2005 « Reconnaissance et classification de phonèmes ».

[12]: http://parole.loria.fr.

[13]: http://www.mass.u-bordeau.fr/2-mmc/ANN.

[14] : http://admi.net/evariste.

[15] : C.AZARA, H.AISSAOUI « Expériences de reconnaissance automatique de la parole à base de module de Markov cachés discrets », thèse d'ingénieurs, institut d'informatique USTHB, Alger **(1999).**

[16] :Morseli Khalida, Hannous Hassina «Application des MFCCs à la reconaissance des phonèmes arabes», Mémoire d'ingéniorat 2007 Université Saad Dahleb Blida.

[17] : JEAN. PAUL. HATON, « Reconnaissance automatique de la parole », GERVEL- FGH , **(1991).**

[18] : CALLIOPE « La parole et son traitement automatique », collection technique et scientifique des télécommunications, Edition Masson **(1989).**

[19] : L.MENSOR, A.SLIMANI « Reconnaissance des chiffres manuscrits par les réseaux de neurones artificiels », thèse d'ingénieur, institut d'électronique, USTHB, Alger **(2001)**.