الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR.YAHIA FARES DE MEDEA

FACULTE DE LA TECHNOLOGIE

Département du Génie Électrique

Mémoire de Fin d'études

Présenté pour l'obtention du Diplôme
de Master en Télécommunications

Option : systèmes intelligents et communicants

THEME

Étude et implémentation de l`algorithme MRCG pour la reconnaissance des signaux de la parole

Préparé et présenté par :
Melle. CHEMMAM MESSAOUDA
Mme. BEN KHELFA KHEIRA

Proposé et dirigé par :
M. ESSAID Billel

ANNEE UNIVERSITAIRE: 2016/2017

Dedicace

J'ai l'honneur de dédier ce modeste travail

A Mes très chers parents:

vos prières et vos bénédictions m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que vous méritez pour tous les sacrifices que vous n'avez pas cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Que Dieu, le tout puissant, vous préserve et vous accorde santé, longue vie et bonheur.

A Cher Mon mari yahia et ma petit fille tassnim meriem

A Mes sœurs Ahlem et Nour eliman et Mes frères Sofiane, Abdenour et Katkout Abd el alim

Atous la famile benkhelfa et belaissa

A tout mes amis

A ma très chère partenaire Assia vous étiez vraiment une formidable copine.

KHEIRA

remerciement

En premier lieu on remercie le dieu Allah qui nous a donné la volonté et le courage pour la réalisation de ce modeste travail.

Nos remerciements s'adressent particulièrement au monsieur ESSAID Billel, promoteur de ce travail. Nous avons eu le privilège de travailler avec vous et d'apprécier vos conseils. Votre compétence et votre sens du devoir nous ont énormément marqués. Veuillez trouver ici l'expression de notre respectueuse considération.

Nos sincères remerciements vont également aux membres de jury qui nous ont fait l'honneur de juger ce travail.

Que tous ceux qui nous ont aidés de près ou de loin, trouvent ici l'expression de notre gratitude.

ملخص

الهدف من هذه الدراسة هو اقتراح تطبيق التحقق من المتحدث، يتم تنفيذ تحليل الصوت على كلام المتحدث المراد التحقق منه وعلى كلام المتحدث الذي بقاعدة البيانات وذلك بتطبيق طريقة MRCG ومن ثم الانتقال الى المقارنة بين الخصائص الناتجة عن عملية MRCG ، تتم هذه العملية من خلال حساب المسافة الإقليدية بين نتائج MRCG لكل من المتحدثين الاثنين .

الكلمات المفتاحية التعرف على الكلام ، إشارة الكلام ، تحليل الصوتية MRCG، قوقعي.

Résumé

L'objectif de cette mémoire est de proposer une application de reconnaissance de parole, une analyse acoustique est effectuée sur le signal parole du locuteur à confirmer et le locuteur de référence, en basant sur la méthode MRCG, ensuite passer à une tâche de comparaison entre les deux vecteurs MRCG pour prendre la décision d'acceptation ou de rejet, cette comparaison est effectuer par la méthode de calcul de la distance euclidienne entre les deux vecteur MRCG.

Mots clés : reconnaissance de parole, signal de parole, analyse acoustique, MRCG , cochléagramme.

Abstract

The objective of this thesis is to propose a speech recognition application, an acoustic analysis is performed on the speaker speech signal to be confirmed and the reference speaker, based on the MRCG method, then proceed to a task of comparison between The two MFCG vectors to take the decision of acceptance or rejection, this comparison is carried out by the method of calculating the euclidean distance between the two MRCG vectors.

Key words: speech recognition, speech signal, acoustic analysis, MRCG, cochleagrame.

Sommaire

LISTE DES FIGURES……I

LISTE DES TABLEAUX ……II

LISTE DES ACRONYMES.....III

Introduction générale 1

1. Introduction.....2

2. Signal de la parole.....2

2.1 .Les caractéristiques d'un signal de la parole.....2

2.1.1. La fréquence fondamentale 3

2.1.2. L'énergie.....3

2.1.3. Le spectre.....3

3.1. La cochlée 4

3.2. Membrane basilaire (BM).....4

3.2.1. IHC et OHC.....5

4. Extraction des paramètres acoustiques des signaux de la parole.....5

4.1. Les paramètres MFCC (Mel Frequency cepstral Coefficients) 6

4.2 Les paramètres LPC (Linear Predictive Coefficients) 6

4.3. Les paramètres PLP (Perceptual Linear Prediction) 7

5. Reconnaissance de la parole 7

5.1. Historique.....7

5.3. Caractérisation de la RAP 9

5.3.1. Mode et style de prononciation 9

5.3.2. Mode d'utilisateurs 9

5.4. Application de la reconnaissance de la parole 10

5.5. Difficultés de la reconnaissance de la parole 11

5.5.1. La Redondance 12

5.5.2. Variabilité 12

5.5.3. Continuité Et Coarticulation 13

5.5.4. Conditions D'enregistrement 13

5.6. Méthodes de reconnaissances 13

5.6. 1.Méthodes de classification 13

5.6.2Methode de vérification 15

6. Conclusion 17

1. Introduction 18

2 .MRCG 18

2.1 .Principe de MRCG 18

2.2. Construction de MRCG 19

3. Deltat et Double Deltat 22

4. Conclusion 23

1. Introduction 24

2. Environnement de travail 24

2.1. Outil de développement (Matlab R2008a) 24

3. Description de système de vérification du locuteur 25

3.1. Construction de la base de données 25

3.2. Analyse acoustique(MRCG) 27

3.3. La comparaison 27

3 .4. Dicision 29

3 .4. 1. Tests 29

Conclusion 36

Conclusion générale 37

Bibioghraphie 38

LISTE DES FIGURES

Figure I.1 Chaine de Reconnaissance Vocal .....**Erreur ! Signet non défini.**

FigureII.1. Principe de MRCG .....19**Erreur ! Signet non défini.**

FigureII.2.la réponse impulsionnelle du filtre10 .....**Erreur ! Signet non défini.**

Figure II.3. Le signal filtré par le 10ème filtre Gammatone.....**Erreur ! Signet non défini.**

Figure II.4. Paramètres MRCG.....**Erreur ! Signet non défini.**

FigureIII.1. Architecture général de notre application .....**Erreur ! Signet non défini.**

Figure III.2. Le code MATLAB pour enregistrer le son. .....**Erreur ! Signet non défini.**

Figure III.3. Représentation du signal parole de mot «Bienvenue » .... **Erreur ! Signet non défini.**

Figure III.4. *entrer numéro personne existe dans l'application* .**Erreur ! Signet non défini.**

Figure III.5. *entrer numéro l'echantillon de chaque personne* 35

Figure III.6. *Choisir le fichie .wa* …… 35

Figure III.7.le son est correct 36

Figure III.8.le son est incorrect 36

LISTE DES TABLEAUX

Tableau III.1.performance de système……30

Tableau III.2.performance de système……31

Tableau III.3.performance de système ……32

LISTE DES ACRONYMES

ARPA: Advanced Research Projects Agency.

BM: Basilar Membrane

DTW: Dynamic Time Warping.

ERB: Equivalent Rectangular Bandwidth.

GTF: Gammatone Filter.

HMM: Hidden Markov Models.

IHC: Inner Hair Cells.

LPC: Linear Predictive Coefficients.

MFCC: Mel Frequency Cepstral Coefficients.

MRCG : Multi-Resolution Cochleagram

OHC: Outer Hair Cells.

PLP: Perceptual Linear Predictive.

VQ: vector Quantification

RAP : Reconnaissance Automatique de Parole

SVM: Supports Vectors Machine.

TFCT : La Transformée de Fourier à Court Terme

Introduction générale

Un système de RAP est un système qui a la capacité de détecter à partir du signal vocal la parole et de l'analyser dans le but de transcrire ce signal en une chaîne de mots ou phonèmes représentant ce que la personne a prononcé. ces propriétés offrent une grande variété d'applications comme l'aide aux handicapés, la messagerie vocale, les services vocaux dans les téléphones portables (''name dialing'', numérotation automatique, reconnaissance de commandes vocales), la production de documents écrits par dictée, etc.

Ces méthodes et modèles travaillent à partir d'une information extraite du signal de parole considérée comme pertinente. cette extraction est effectuée par une analyse acoustique qui conduit à rassembler cette information sous le terme de vecteur de paramètres acoustiques dont la dimension et la nature sont déterminants pour accéder à de bonnes performances des systèmes de RAP. les différents types de paramètres acoustiques couramment cités dans la littérature sont les coefficients : LPC, LSF, MSG, LPCC, LFCC, PLP, MFCC, MRCG etc.

L'extraction des coefficients MRCG est une nouvelle technique d'extraction des paramètres acoustiques (caractéristiques), qui montre sa capacité dans les systèmes RAP.

Le mémoire est constitué de trois chapitres :

Le premier chapitre est une introduction générale au domaine du signal de la parole, ses caractéristiques, son système de production et quelques rappelles essentiels.

Dans une seconde partie, nous présenterons le modèle proposé dans ce mémoire pour l'extraction de caractéristiques : le Codage MRCG

Enfin, le troisième chapitre détaille les différentes étapes utilisées dans notre application. nous démarrons par l'acquisition du signal de parole des différents locuteurs, la modélisation de ce dernier avec les modèles MRCG et aussi la réalisation d'une application de vérification du locuteur.

1. Introduction

La parole est la faculté de communiquer la penser par un système de sons articulés; c'est le moyen de communication privilégié entre les humain qui sont les seul être vivants à utiliser un tel système structuré.

L'information d'un message parlé réside dans les fluctuations de la pression de l'air, engendrées puis émises par l'appareil phonatoire, ces fluctuations constituent le signal vocal, elles sont détectées par l'oreille, laquelle procède à une certaine analyse.

Les résultats sont transmis au cerveau, qui interprète un message vocal, est une suite d'image auditives, éléments minimaux dénués de sens, mais dont l'association permet de distinguer les éléments constitutifs des niveaux supérieurs (syllabes, mots, phrase,…)

Le contenu d'un message vocal au sens strict est simplement son intelligibilité ;

Au sens large il faut bien sûr prendre en compte toutes les intonations, ce qui en a accroit fortement la richesse par rapport au message écrit. [1]

Un système de reconnaissance automatique de la parole (RAP) est un système qui a la capacité de détecter la parole et de l'analyser dans le but de générer une chaîne de mots ou phonèmes représentant ce que la personne a prononcé.

Dans ce chapitre, nous nous présentons les caractéristiques du signal de la parole, ainsi que les méthodes d'extraction des paramètres utilisés dans les systèmes de reconnaissance automatique de parole (RAP).

2. Signal de la parole

La parole est un signal réel, continu, d'énergie finie et non stationnaire. sa structure est complexe et variable avec le temps.

2.1 .Les caractéristiques d'un signal de la parole

Le signal vocal est généralement caractérisé par trois paramètres: sa fréquence fondamentale, son énergie et son spectre.

2.1.1. La fréquence fondamentale

Elle représente la fréquence du cycle d'ouverture/fermeture des cordes vocales. cette fréquence caractérise seulement les sons voisés, elle peut varier :

- De 80Hz à 200Hz pour une voix masculine,
- De 150Hz à 450Hz pour une voix féminine,
- De 200Hz à 600Hz pour une voix d'enfant. [2]

2.1.2. L'énergie

Elle est représentée par l'intensité du son qui est liée à la pression de l'air en amont du larynx. l'amplitude du signal de la parole varie au cours du temps selon le type de son, et son énergie dans une trame est donnée par :

$$E = \sum_{n=0}^{N-1} s^2(n) \qquad \textbf{(I.1)}$$

Avec N : la taille de la trame.

2.1.3. Le spectre

L'enveloppe spectrale ou spectre représente l'intensité de la voix selon la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme.

La quasi stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal vocal sur des fenêtres de durée généralement comprise entre *20ms* et *30ms* appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse.

La transformée de fourier à court terme (*TFCT*) d'un signal échantillonné est par définition la transformée du signal pondéré

$$\hat{s}(k) = \hat{s}\left(f = \frac{k}{N}\right) = \sum_{n=0}^{N-1} s(n) \,.\, w(n) . exp^{-\frac{j2\pi nk}{N}} \quad 0 \leq k \leq N-1 \qquad \textbf{(I.2)}$$

Où : *N* : Le nombre de points prélevés

s(k) : Spectre complexe.

s(n): Segment analysé.

w(n) : Fenêtre de temps.

Le spectre de puissance (appelé aussi densité spectrale de puissance de la transformé de fourier) est donné par:[3]

$$(k)|^2 , 0 \leq k \leq N/2 \qquad (I.3)$$

3. **Le système auditif humain**

Anatomiquement, l'oreille humaine se compose de trois sections: l'extérieur, le milieu et l'intérieur de l'oreille, l'oreille externe recueille du son et la propage vers l'oreille moyenne.

Dans l'oreille moyenne, trois petits os fournissent une adaptation d'impédance et transforment les sonne à l'oreille interne. l'oreille interne a une structure de récepteur auditif appelée la cochlée, qui transforme le son d'entrée acoustique en signaux adaptés par les neurones.

3.1. La cochlée

La cochlée (est un tube enroulé sur lui-même, qui ressemble à une coquille d'escargot. l'étrier sur sa base se connecte à l'oreille moyenne. si une cochlée était déballée, soit environ 35 mm de long. la structure du tube de cochlée est divisée en trois chambres (la canal vestibulaire, canal de cochlée et canal tympanique) par la membrane de reissner et la membrane basilaire.

3.2. Membrane basilaire (BM)

La membrane basilaire est une structure de support rigide pour les cellules sensorielles limaçon. la membrane basilaire, mesurée de la base cochèle à l'apex, est d'environ 35 mm de longueur. sa largeur et sa rigidité changent sur sa longueur. de base à apex, la largeur de la membrane basilaire augmente d'environ 6, et la dureté varie d'un facteur de 100, avec la partie

la plus rigide à la base. sur la membrane basilaire il existe trois rangées de cellules capillaires externes (OHC) et une rangée de cellules capillaires intérieures (IHC). il y a environ 12000 OHC et environ 3500 IHC sur une membrane basilaire humaine.

3.2.1. IHC et OHC

Les cellules capillaires intérieures (IHC) et les cellules ciliées externes (OHC) sont des cellules différentes de l'anatomie et des points de vue fonctionnels. les IHC sont servis par environ 95% du nerf fibers dans une cochlée, ce qui prouve que les IHC sont des récepteurs auditifs.

OHC ont des structures motrices et fournissent des commentaires positifs pour améliorer les vibrations à Certains endroits sur la membrane basilaire.

L'IHC a des cheveux au sommet et des nerfs nerveux attachés à sa base. fonctionnellement, un IHC est quelque peu similaire à un transistor. La flexion des cheveux cellulaires contrôle un courant ionique à l'intérieur de l'IHC semblable à un transistor " gate " Le courant ionique induit des tirs dans les neurones afférents.

le processus d'innervation est également appelé processus de transduction ihc, dans lequel le les mouvements mécaniques sont transduits dans les essais de neurones. si le son d'entrée est un signal périodique, les fibres nerveuses afférentes montrent une fringure neuronale à taux élevé / faible périodique activités [4]

4. Extraction des paramètres acoustiques des signaux de la parole

Le signal de parole est 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.

La phase d'extraction de caractéristiques doit être faite avec soin, car elle contribue directement aux performances du système global. les codeurs les plus couramment utilisés sont le codage linéaire prédictif (Linear Predictive Coefficients « LPC »), les coefficients cepstral (Mel Frequency Cepstre Coefficients « MFCC ») ou bien le codage linéaire prédictif

perceptuel (Perceptual Linear Predictive PLP) .le codage MFCC et le codage PLP ont la propriété d'intégrer des connaissances du modèle auditif humain

4.1. Les paramètres MFCC (Mel Frequency cepstral Coefficients)

Cette technique consiste à calculer les coefficients cepstraux sur une échelle en mel qui se rapproche de la perception fréquentielle de l'oreille

Pour calculer MFCC, un signal d'entrée est divisé en 20 ms avec un décalage de trame de 10 ms. nous appliquons un fenêtrage dans chaque trame et dériver le spectre de puissance en utilisant la transformée de fourier à courte durée. puis nous convertissons le spectre de puissance en échelle de mel. finalement, la compression logarithmique et la transformée cosinus discrète (DCT) appliqué pour calculer le MFCC.

4.2 Les paramètres LPC (Linear Predictive Coefficients)

L'analyse par prédiction linéaire LPC est forte utilisée dans le codage de parole dans le but de réduire la redondance du signal vocal, ou pour extraire des paramètres pertinents pour la reconnaissance de parole. l'estimation des coefficients de la fonction de transfert du conduit vocal est faite en supposant connaître le signal d'excitation. pour les sons non voisés, le signal d'excitation est un bruit blanc de moyenne nulle et de variance unité. pour les sons voisés, cette excitation est une suite d'impulsions d'amplitude unité. la fonction de transfert du conduit vocal dans le domaine Z est donnée par :

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1-A(z)} \qquad (I.4)$$

$$A(z) = \sum_{k=1}^{p} a_k{}^{z-k} \qquad (I.5)$$

Où $A(z)$ est le prédicteur linéaire, (a_k) sont les coefficients de prédiction, $S(z)$ est le signal de parole produit en sortie, $U(z)$ est le signal d'excitation et G est un gain.

Le signal de parole *S(n)* à la sortie du modèle est donc représenté par la somme d'une combinaison linéaire des échantillons précédents et de la fonction d'excitation, tel que

$$S(n)= \sum_{k=0}^{n} a_k s(n-k) + Gu(n) \qquad (I.6)$$

Le modèle de prédiction exploite le fait que les échantillons successifs du signal de parole sont corrélés ; d'où l'intérêt de ce modèle dans le codage de la parole dans le sens où il permet de représenter la parole juste par ses paramètres pertinents, sans redondance. Signalons également que les coefficients sont choisis de façon à minimiser l'erreur quadratique de prédiction sur chaque segment de la fenêtre d'analyse.

4.3. Les paramètres PLP (Perceptual Linear Prediction)

La prédiction linéaire perceptuelle exploite les connaissances du système auditif humain pour paramétrer la parole en introduisant des mécanismes psycho acoustiques de l'oreille humaine.

l'analyse perceptive repose sur l'analyse par prédiction linéaire (LPC). dans cette analyse, les coefficients sont calculés suite à la résolution d'équations obtenus par la transformée de fourier inverse du module au carré de la transformée de fourier du signal.

l'analyse perceptive introduit au niveau fréquentiel des bandes critiques. cette intégration se fait avec un banc dont les fréquences centrales sont espacées linéairement selon l'échelle Bark. cette intégration se justifie par le fait que l'oreille se comporte comme un banc de filtres. [5]

5. Reconnaissance de la parole

5.1. Historique

Les études scientifiques sur la reconnaissance de la parole ont commencé au début des années 1950, mais ont donné leurs premiers résultats tangibles dans les années 70 avec le projet ARPA (Advanced Research Projects Agency [Newell et al, 1973 ; Klatt, 1977]. leur principal but a été, et est toujours, de concevoir un système de reconnaissance de la parole robuste et efficace, pour obtenir un outil s'intégrant *naturellement* dans la communication

entre l'homme et la machine et facilitant son interaction. pour le commun des hommes, les robots et des machines qui communiquent en langage naturel avec les acteurs des films de science-fiction, est l'image immédiate qui s'impose. cependant, sans tomber dans cet extrême, de nombreuses applications d'un tel outil de reconnaissance automatique de la parole (RAP) peuvent dès maintenant être envisagées.

5.2. Introduction

La parole est l'un des moyens les plus directs d'échange de l'information, utilisés par l'homme. cet avantage a donné naissance à plusieurs travaux de recherche dont l'objectif est la conception des systèmes permettant de reconnaître la séquence des mots parlés.

Un système de reconnaissance automatique de la parole (RAP) est un système qui a la capacité de détecter la parole et de l'analyser dans le but de générer une chaîne de mots ou phonèmes représentant ce que la personne a prononcé. cette analyse se fonde sur l'extraction des paramètres descriptifs de la parole. cependant le signal parole ne contient pas seulement des informations sur le texte parlé mais aussi des informations sur le locuteur, la langue, les émotions dont leur extraction n'est pas l'objectif de la RAP. cette thème s'intéresse à une étape primordiale de la RAP, en l'occurrence la sélection des paramètres descriptifs pertinents pour la tâche de reconnaissance des mots parlés.

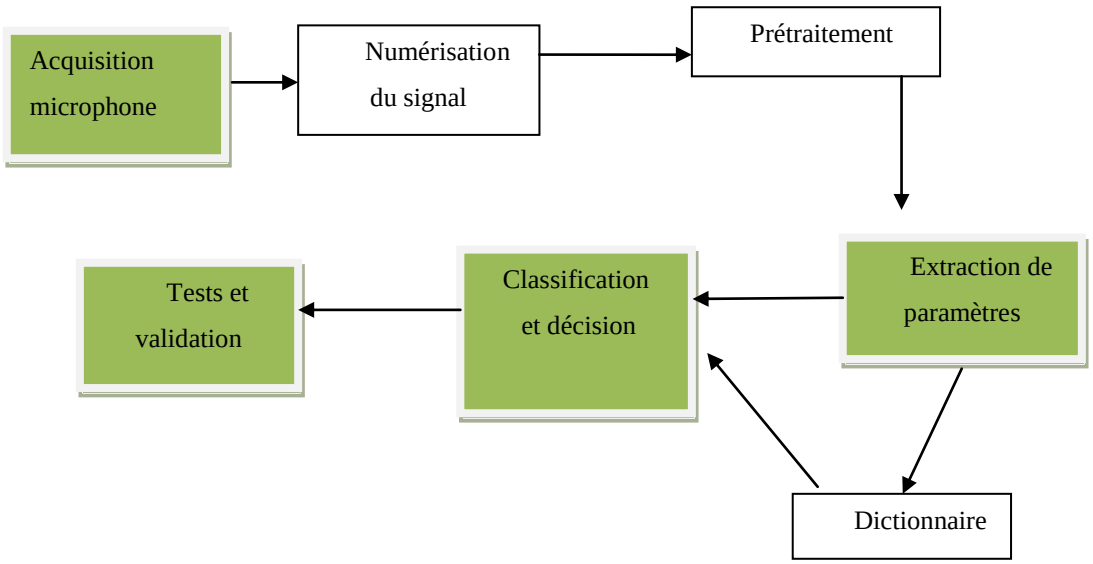


Figure I.1.*Chaine de Reconnaissance Vocal*

5.3. Caractérisation de la RAP

La reconnaissance automatique de la parole (RAP) est un processus qui convertit le signal acoustique de parole en un ensemble de mots. les systèmes de RAP peuvent être caractérisés par plusieurs paramètres.

5.3.1. Mode et style de prononciation

Il existe trois modes de prononciation distincts :

- mots isolés : chaque mot est prononcé isolément ; une pause de durée importante sépare les mots ;
- *mots connectés :* le système reconnaît des séquences de quelques mots sans pause volontaire entre eux (exemple : reconnaissance de chiffres connectés ou de nombres quelconques, etc.) ;
- *parole continue :* les mots sont prononcés naturellement sous forme de phrases aux énoncés plus ou moins longs.

5.3.2. Mode d'utilisateurs

Le système de reconnaissance de la parole peut se distinguer suivant deux types d’utilisation différents :

- *dépendant du locuteur*: le système est adapté à la voix d'un locuteur particulier.
- *indépendant du locuteur* : le système n'est pas adapté à la voix d'un locuteur particulier et peut être utilisé par un locuteur quelconque.

5.3.3. Environnement

Les changements de conditions environnementales (bruit, changement de microphone, acoustique de la salle, etc.) peuvent influer sur le taux de reconnaissance du système. la plupart des systèmes vont fonctionner correctement dans un environnement aux caractéristiques acoustiques et sonores proches de l'environnement dans lequel s’est fait l'entraînement mais les performances vont se dégrader notablement si les conditions environnementales sont très différentes.

De nombreuses équipes de recherche étudient des moyens de rendre les systèmes de reconnaissance insensibles, c'est-à-dire robustes, aux changements des conditions environnementales, soit en réalisant des moteurs de reconnaissance capables de s'adapter eux mêmes aux nouvelles conditions avec, par exemple, des techniques d'apprentissage adaptative, des techniques de modèles évolutionnaires , soit en faisant précéder les moteurs de reconnaissance par des modules de traitement du signal capables d'améliorer le signal de parole à l'entrée du système de reconnaissance.

Ces dernières techniques étudient des modules de réhaussement de la parole dans le bruit, ou bien des modules d'annulation d'écho des salles, ou encore du filtrage adaptatif pour éliminer les bruits.

5.4. Application de la reconnaissance de la parole

les avantages que l'on attend de la reconnaissance de la parole sont multiples. elle libère complètement l'usage de la vue et des mains (contrairement à l'écran et au clavier), et laisse l'utilisateur libre de ses mouvements. la vitesse de transmission des informations est supérieure, dans la rap à celle que permet l'usage du clavier. enfin tout le monde ou presque sait parler, alors que peu de gens sont à l'abri des fautes de frappe et d'orthographe.

Ces avantages sont à l'origine d'une grande variété d'applications comme :

- L'aide aux handicapés.
- La messagerie.
- L'avionique.
- La commande de machines ou de robots.
- Le contrôle de qualité et la saisie des données.
- L'accès à distance : téléphone, internet.
- La dictée vocale.

Toutes ces applications bénéficient de l'évolution technologique qui se traduit par l'apparition de composants intégrés spécialisés (en traitement du signal pour la programmation dynamique) et du développement des techniques et des méthodes algorithmiques de plus en plus performantes.

Enfin l'insertion d'un système RAP dans son environnement réel d'utilisation dépend de son contexte d'application et de ses conditions d'utilisation ce qui peut vite rendre le dispositif très complexe. un système RAP peut être décrit selon quatre grands axes graduant cette complexité :

- La dépendance du locuteur (système optimisé pour un locuteur bien particulier) ou l'indépendance du locuteur (système pouvant reconnaître n'importe quel locuteur).
- Le mode d'élocution : mots isolés, mots connectés, mots-clés, parole continue lue ou parole continue spontanée.
- La complexité du langage autorisé : taille du vocabulaire et difficulté de la grammaire.
- La robustesse aux conditions d'enregistrement : systèmes nécessitant de la parole de bonne qualité ou fonctionnement en milieu bruité.

Tout système correspond à un compromis entre ces axes, choisi en fonction du but à atteindre. les systèmes réalisés de la RAP sont conçus pour des applications spécifiques. cela conduit à une restriction de l'univers du dialogue homme-machine. en effet, la conception de systèmes capables de comprendre la langue orale dans son intégralité est pour l'instant beaucoup trop complexe.

5.5. Difficultés de la reconnaissance de la parole

Le signal de parole est l'un des signaux les plus complexes à caractériser et analyser car c'est un sujet à une grande variabilité. cette complexité est liée à la production du signal de parole, ainsi qu'à l'aspect technologique.

Le signal de parole varie non seulement avec les sons prononcés, mais également avec le locuteur, l'âge, les émotions, la santé, l'environnement.

De plus, la mesure du signal de parole est fortement influencée par la fonction de transfert du système de reconnaissance (les appareils d'acquisition et de transmission), ainsi que par le milieu ambiant.

Ainsi, l'obstacle majeur pour améliorer les performances d'un système RAP provient de la grande complexité du signal vocal due à la combinaison de plusieurs facteurs, principalement

la redondance du signal acoustique, la grande variabilité intra et interlocuteurs, les effets de la coarticulation en parole continue, ainsi que les conditions d'enregistrement.

5.5.1. La Redondance

Le signal vocal présente un caractère redondant. il contient plusieurs types d'information : les sons, la syntaxe et la sémantique de la phrase, l'identité du locuteur et son état émotionnel.

Bien que cette redondance assure une certaine résistance du message au bruit, elle rend l'extraction des informations pertinentes pour la RAP plus délicate de part la multi modalité des sources d'information.

5.5.2. Variabilité

Le signal vocal de deux prononciations à contenu phonétique égal est différent pour un même locuteur (variabilité intralocuteur) ou pour des locuteurs différents (variabilité interlocuteur).

En effet, lorsque la même personne prononce deux fois le même énoncé, on constate des variations sensibles sur le signal vocal causées par :

- L'état physique, par exemple, la fatigue ou le rhume.
- Les conditions psychologiques, comme le stress.
- Les émotions du locuteur.
- Le rythme lié à la durée des phonèmes (façon dont s'exprime le locuteur) et l'amplitude (voix normale, voix chuchotée, voix criée).

Cependant la variabilité interlocuteur est a priori la plus importante. elle s'explique par :

- Les différences physiologiques entre locuteurs.
- Les habitudes acquises en fonction du milieu social et géographique comme les accents régionaux.

Cette variabilité rend très difficile la définition d'invariants et complique la tâche de reconnaissance. ainsi, il faut pouvoir séparer ce qui caractérise les phonèmes, de l'aspect particulier à chaque locuteur.

5.5.3. Continuité Et Coarticulation

La production d'un son est fortement influencée par le son qui le précède et qui le suit en raison de l'anticipation du geste articulatoire. la localisation correcte d'un segment de parole isolé de son contexte est parfois impossible. évidement la reconnaissance des mots isolés bien séparés par un silence est plus facile que la reconnaissance des mots connectés. en effet, dans ce dernier cas, non seulement la frontière entre mots n'est plus connue mais, de plus, les mots deviennent fortement articulés.

5.5.4. Conditions D'enregistrement

L'enregistrement du signal de parole dans de mauvaises conditions rend difficile l'extraction des informations pertinentes indispensables pour la reconnaissance des mots contenus dans ce signal. en effet, les perturbations apportées par le microphone (selon le type, la distance, l'orientation) et l'environnement (bruit, réverbération) compliquent beaucoup le problème de la reconnaissance.

Pour illustrer l'ensemble de ces difficultés, un système de RAP doit, en définitive, être capable de décider "qu'un [a] prononcé par un adulte masculin est plus proche d'un [a] prononcé par un enfant, dans un mot différent, dans un environnement différent et avec un autre microphone, que d'un [o] prononcé dans la même phrase par le même adulte masculin"[6]

5.6. Méthodes de reconnaissances

On a deux méthodes de reconnaissance :

5.6. 1.Méthodes de classification

Il existe plusieurs méthodes de classification parmi cette méthodes on a :

5.6.1. 1. Les modèles de markov cachés HMM (Hidden Markov Models)

Les modèles de markov cachés (Hidden Markov Models ou HMM) ont connu une grande importance depuis leur introduction en traitement de la parole. la plupart des systèmes de RAP utilisent les HMM pour modéliser les mots ou les unités élémentaires de la parole telle

que les phonèmes, les syllabes. de nombreux travaux de recherches ont montré l'efficacité de ces modèles en reconnaissance de la parole continue ou isolée, indépendamment du locuteur pour de petits et grands vocabulaires.

5.6.1.2. Machine à vecteurs de Supports SVM (supports vectors machine)

Les SVM sont des classificateurs binaires et statiques. c'est à dire qu'ils permettent de créer une surface de décision entre deux classes définies dans un même espace. pour cela, ils construisent une frontière de décision par projection des caractéristiques provenant d'un espace d'origine dans un espace de caractéristiques de dimension supérieure (voir infini) dans le but de rendre les classes linéairement séparables.

L'hyperplan choisi est celui qui maximise la marge de séparabilité entre les deux ensembles de données. la sélection de l'hyperplan dans un espace de caractéristiques nécessite d'évaluer un produit scalaire dans cet espace. ce qui peut être très coûteux en temps et en complexité si l'espace est de très grande dimension. heureusement, ce calcul n'est pas obligatoire grâce à une opération mathématique appelé kernel. le kernel calcule le produit scalaire de deux points dans l'espace de dimension supérieur sans avoir à les projeter.

de plus, cette technique n'est pas appliquée seulement au svm, mais aux groupes des machines à apprentissage appelés kernel machines. toutes utilisent des kernels pour appliquer les données d'un espace d'entrée en un espace de caractéristiques de dimension supérieur. ils sont capables de créer une séparation non linéaire d'un espace d'entrée. il existe de nombreuses kernel machines, les plus connues relevant vector machines, bayes point machine, gaussian processes, least squares SVM, méthodes de boosting et kernel fisher discriminant.

Les kernel machines sont apparues comme des solutions viables et compétitives dans de nombreux domaines. cependant, il existe de nombreux problèmes à leur utilisation pour la reconnaissance de la parole.

5.6.1.3. Déformation temporelle dynamique DTW (dynamic time warping)

Cette méthode est la solution la plus simple à mettre en œuvre en reconnaissance de la parole. elle consiste à effectuer une comparaison dynamique entre un vecteur de référence et

un vecteur de test. on cherche à apparier l'ensemble des points du premier vecteur avec le second. on associe un poids à chaque appariement possible. l'utilisation de la programmation dynamique permet de trouver l'appariement minimisant ce coût. les formes références sont obtenues par extraction des caractéristiques. la forme test peut avoir subit des transformations (rotations, translation, changement de taille). on utilise pour cela une méthode de mesure de similarité. la distance finale entre une forme de référence et une forme de test est calculée comme la somme des distances partielles entre vecteurs de paramètres le long du chemin optimal. des poids liés à la suppression et à l'insertion d'informations, sont imposés pour calculer le meilleur chemin

Cette

5.6.1.4. Les réseaux de neurones

Les réseaux de neurones disposent de propriétés qui sont intéressantes du point de vue de méthode a l'avantage d'être simple et de nécessiter que peu de données d'apprentissage. par contre, elle est limitée par la taille du vocabulaire à reconnaître et à la reconnaissance mono locuteur.la reconnaissance de la parole.

Ils peuvent apprendre des fonctions fortement non linéaires qui permettent de reconnaître une classe de forme et, simultanément, de rejeter les autres classes. ils constituent des approximateurs de fonctions très puissants.

De plus, ils ne nécessitent aucune donnée statistique à l'initialisation contrairement aux modèles de markov cachés.

Enfin, ils présentent des structures parallèles régulières facilement implantables sous forme logicielle et matérielle [7].

5.6.2Methode de vérification

5.6.2.1 Quantification vectorielle VQ

Un système de reconnaissance du locuteur doit pouvoir estimer la probabilité de distributions des vecteurs de caractéristiques calculées. il est impossible de stocker chaque vecteur qui génère le mode d'apprentissage, étant donné que ces distributions sont définies sur

un espace de grande dimension. il est souvent plus facile de commencer par quantifier chaque vecteur de caractéristique avec un processus appelé quantification vectorielle (VQ), c'est un processus qui prend un grand nombre de vecteurs acoustiques et produit un petit ensemble de vecteurs qui représente les centres de gravité de la distribution. La technique de VQ consiste à extraire un petit nombre de vecteur de caractéristiques représentatives comme un moyen efficace de caractériser un locuteur spécifique, il n'est pas possible de stocker chaque vecteur généré

5.6.2.2 .Distance euclidienne

Dans la phase de reconnaissance du locuteur, une voix d'un locuteur inconnu est représenté par une séquence de vecteurs caractéristiques {x1, x2…x1}, et ensuite elle est comparée avec le dictionnaire de base de données. afin de vérifier ou bien identifier le locuteur, cela peut être fait en mesurant la distance de distorsion de deux vecteurs basé sur la minimisation de la distance euclidienne [13].

La distance euclidienne est la distance ordinaire entre les deux points que l'on peut mesurer avec une règle. La distance euclidienne entre deux points : P=(p1,p2…pn)et Q=(q1,q2…qn) , est donnée par

$$\mathbf{d}=\sqrt{\sum_{i=1}^{n}(pi-qi)^2}=\sqrt{(p1-q1)+(p2-q2)+\cdots(pn-qn)^2} \qquad \textbf{(I.7)}$$

La comparaison est effectuée entre le vecteur test et le vecteur de référence. le vecteur de caractéristiques généré du signal parole test est comparé avec le vecteur de caractéristiques du signal parole de base de données. la similarité entre les deux signaux est calculée par la distance euclidienne entre les vecteurs acoustique (MRCG). si la distance est supérieure au seuil, le locuteur est dit escroc sinon le système va reconnaitre le locuteur selon sa base de données.

Notons que la distance euclidienne est la méthode utilisée dans notre conception (chapitre 3),

L'algorithme de la vérification est décrit comme suit :

a) premièrement, le signal original est transformé en un vecteur acoustique suivant les étapes d'analyse acoustique (MRCG dans notre cas) soit E1 le vecteur normalisé des caractéristiques.
b) pour un signal inconnu, calculer le MRCG pour obtenir un vecteur de caractéristiques soit E2.
c) calculer la distance euclidienne D entre le vecteur de caractéristiques E1 et le vecteur de caractéristiques E2.
d) si la distance D est supérieure à une certaine valeur de seuil, alors le vecteur de signal inconnu E2 est loin de vecteur du signal de base de données E1. Dans ce cas, la personne est rejetée. Sinon,la personne est considérée comme authentifié [8] .

6. Conclusion

Dans ce chapitre nous avons introduit le concept de signal de parole en générale, et la représentation acoustique du signal de parole. nous avons aussi présenté le système de reconnaissance automatique de la parole avec ses méthodes, difficultés et leurs applications.

1. Introduction

Dans ce chapitre, une nouvelle méthode d'extraction de paramètres appelé : cochléagramme multi-résolution (MRCG), principalement conçu pour obtenir les meilleures performances de séparation que les autres paramètres : PLP, LPC, MFCC Outre les fonctionnalités existantes est étudiée. cette méthode encode les distributions de puissance multi-résolution dans la représentation temps-fréquence d'un signal.

Une combinaison de quatre cochléagrammes À différentes résolutions est effectuée pour construire les coefficients MRCG. Le cochléagramme à haute résolution capte les informations locales, tandis que les trois autres cochléagrammes à basse résolution captent les caractéristiques spectro-temporel à différentes échelles.

2 .MRCG

Dans cette section, nous présentons la fonction MRCG. cette technique a montré son avantage sur de nombreuses caractéristiques acoustiques dans des problèmes de reconnaissance de la parole.

L'idée clé de la MRCG est d'intégrer l'information locale et l'information globale de la parole par une extraction multi-échelle.

1a, MRCG est une concaténation de 4 cochléagrammes avec différentes tailles de fenêtres et différentes longueur du trames.[9]

2.1 .Principe de MRCG

Voici une description des principales étapes de MRCG :

1. Compte tenu du signal d'entrée, générer des réponses du banc de filtres gammatone.
2. Calculer les 64 premiers cochléagramme CG1 avec une longueur de trame de 20 ms et un décalage de trame de 10 ms. Un Log est appliqué sur chaque unité T-F.
3. De la même manière, calculer le deuxième cochléagramme CG2 avec une longueur de trame de 200 ms et un décalage de trame de 10 ms. Un Log est appliqué sur chaque unité T-F.

4. Calculer CG3 et CG4 en faisant la moyenne de CG1 en utilisant des fenêtres carrées de 11x11 et 23x23.
5. Concaténer CG1-4 et ajouter des traits delta et double delta.

La figure suivante, montre le principe de l'algorithme d'extraction des paramètres de signal de la parole comme il est décrit précédemment.

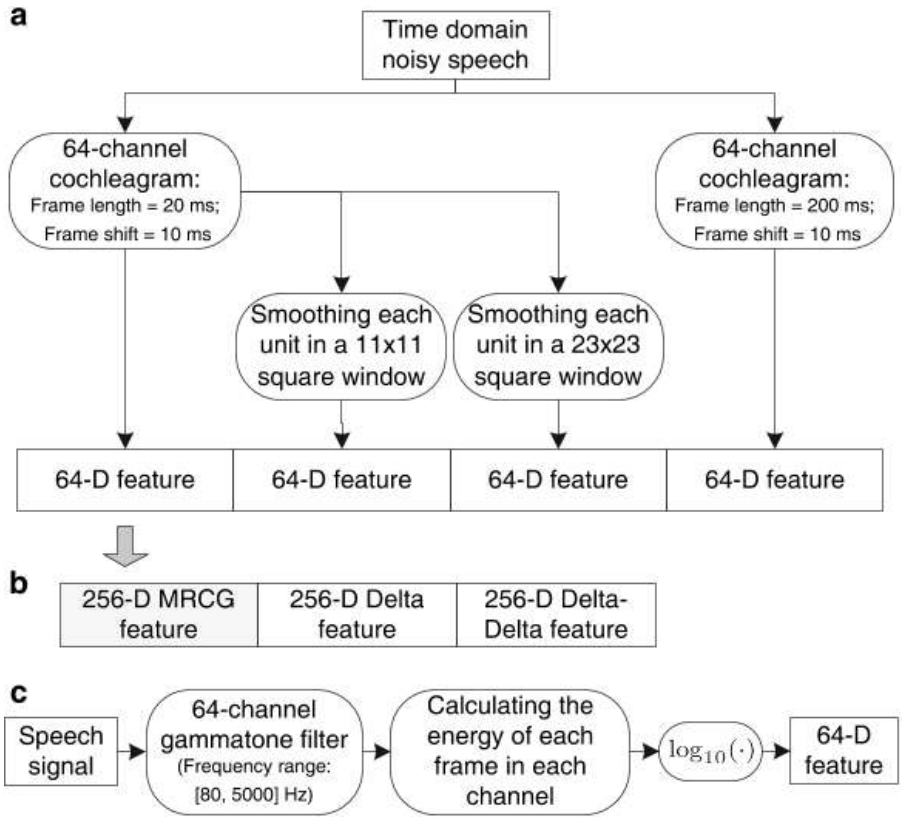


Figure II.1. *Principe de MRCG*

2.2. Construction de MRCG

Pour calculer le cochléagramme, nous passons d'abord un signal d'entrée dans un banc de filtres gammatone, où la réponse impulsionnel d'un filtre gammatone est donnée par :

$$G_{fc}(t) = t^{N-1} exp\,[-2\pi t b(f_c)] cos(2\pi f_c t) u(t) \qquad (II.1)$$

Où N est l'ordre du filtre; *b(fc)* est la bande passante du filtre et qui détermine en grande partie la durée de la réponse impulsionnelle; et *u(t)* est une fonction échelon. pour imiter les

filtres auditifs humains, les fréquences centrales Sont uniformément espacées sur la largeur de bande rectangulaire équivalente (ERB).

La relation entre *b(fc)* et *fc* est donnée par l'équation II.2.

$$b(f_c) = 1.019*ERB(f_c) = 1.019*24.7*(4.37*f_c/1000+1) \quad (II.2)$$

Nous avons fixé la bande fréquentielle de travail à [50-8000].

Après avoir obtenu des signaux de réponse du banc de filtres gammatone, nous faisons une segmentation de chaque signal de la réponse du banc de filtres en trames de 20 ms avec un décalage de trame de 10 ms. nous dérivons ensuite, le cochléagramme en calculant la puissance de chaque trame à chaque canal.

Chaque unité T-F dans le cochléagramme ne contient que des informations locales, ce qui peut ne pas être suffisant pour estimer le masque. pour compenser cela, la fonctionnalité mrcg fournit des information énergétique de chaque unité T-F.

D'une façon similaire, nous calculons CG2 avec une longueur de trame de 200ms, et un chevauchement de 10ms.

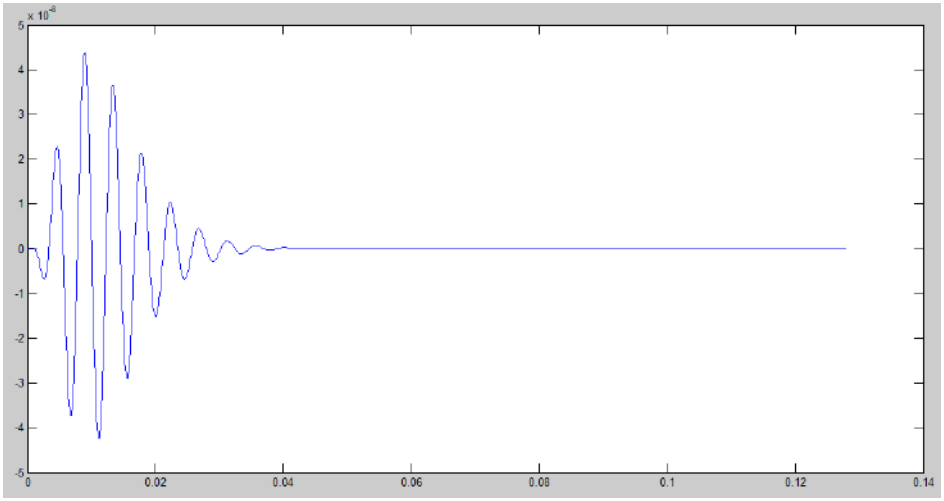
Le CG3 est dérivé en moyennant CG1 sur une fenêtre carrée de 11 canaux de fréquences sur 11 trames de temporels centrée sur une unité T-F donnée.

CG4 et calculé d'une façon similaire à CG3 sauf que la fenêtre utilisée est de taille 23x23.

Faisons une concaténation de CG1 à CG4 pour obtenir les coefficients MRCG, c'est une matrice de dimension 64x4 pour chaque trame temporelle.

Après avoir calculé les coefficients MRCG à 256 dimensions, nous calculons ensuite ses dérivatives deltas et ses doubles deltas, puis combiner tous les trois dans une caractéristique de 768-dimensions [10].

La figure suivante montre la réponse impulsionnelle du $10^{ème}$ filtre du banc de filtre gammatone.



FigureII.2.*la réponse impulsionnelle du filtre10*

Le banc de filtres gammatone est couramment utilisé pour simuler le mouvement de la membrane basilaire dans la cochlée en fonction du temps, dans lequel la sortie de chaque filtre modélise la réponse en fréquence de la membrane basilaire à un seul endroit. le banc de filtres est normalement défini d'une manière telle que les fréquences centrales de filtre sont répartis en fréquence en proportion de leur largeur de bande, dite échelle ERB (Glasberg et Moore, 1990). L'échelle ERB est approximativement logarithmique, sur laquelle les fréquences centrales des filtres sont espacées linéairement. [11]

La figure suivante montre la sortie du 10ème filtre du banc gammatone.

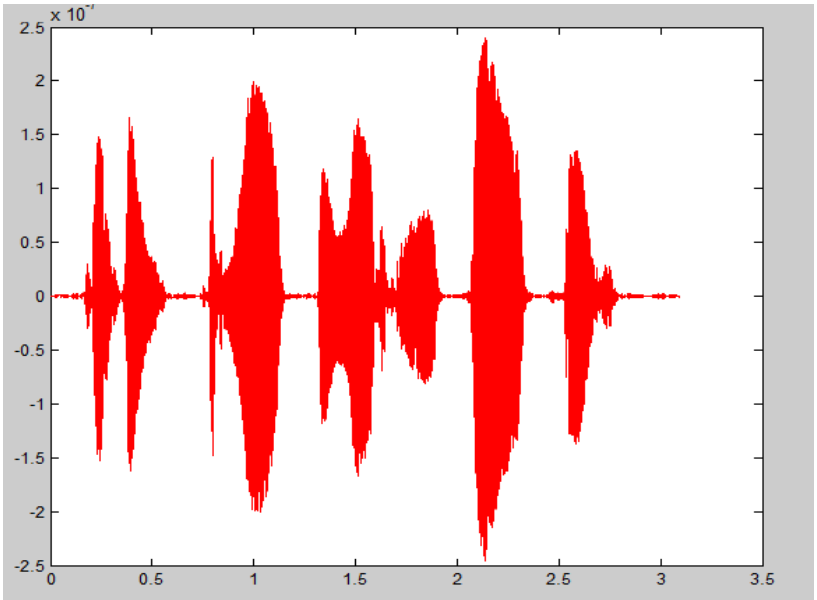


Figure II.3. *Le signal filtré par le 10ème filtre Gammatone*

Le cochléagramme représente le comportement de la membrane basilaire dans la cochlée. la cochlée est la partie interne de l'oreille.

Le cochléagramme est une représentation temps-fréquence du son d'entrée passant par la banque GTF et les modèles IHC.

L'analyse de cochléagramme est un outil utile pour la perception auditive. selon certains chercheurs, le cerveau humain peut effectuer une opération d'autocorrélation sur les signaux neuronaux et d'analyser les informations d'histogramme pour obtenir les périodes de hauteur fondamentales du son d'entrée. [12]

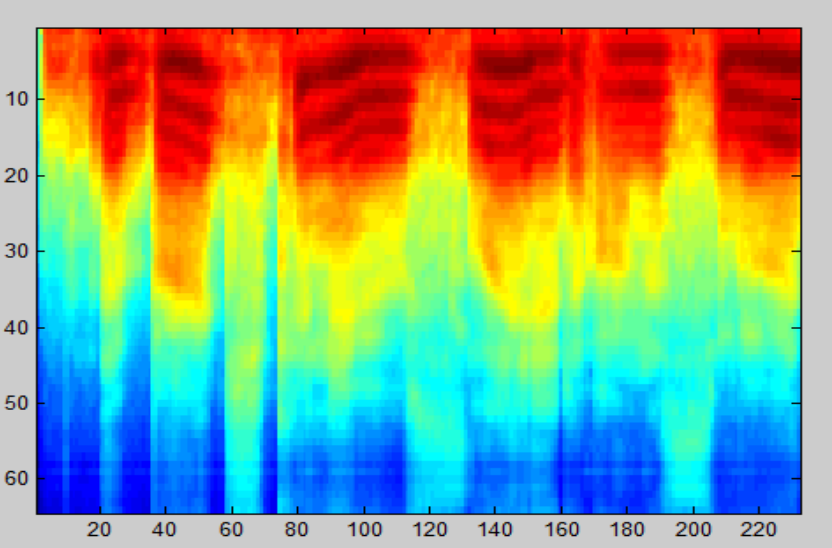


Figure II.4. *Paramètres MRCG*

3. Deltat et Double Deltat

dans le traitement de la parole, les fonctionnalités delta () et double-delta () sont largement utilisés pour extraire la dynamique temporelle. ajouter ces fonctionnalités est une technique de post-traitement dans l'extraction des paramètres de signal de la parole.

Ces caractéristiques permettent d'améliorer la précision de la reconnaissance en ajoutant une caractérisation du temps. pour une séquence cepstrale de courte durée C[n], le delta-cepstral des caractéristiques sont généralement définies comme :

$$D[n] = C[n + m] - C[n - m] \qquad \textbf{(II .3)}$$

Où n est l'indice des trames d'analyse et en pratique m est environ 2 ou 3. de même, les caractéristiques cepstrales à double delta sont définis en termes d'une opération delta ultérieure sur le delta-cepstral. [13]

4. Conclusion

Dans ce chapitre nous avons expliqué la nouvelle technique d'extraction des paramètres statiques et dynamiques de la parole, les coefficients MRCG. c'est une technique qui joue sur le principe de multi-résolution.

1. Introduction

pour bien exploiter les notions théoriques des deux parties de ce travail, nous allons décrire dans ce chapitre en détails le fonctionnement de notre système. ce chapitre met l'accent sur la conception d'une application ayant pour objectif de réaliser un système de vérification du locuteur, avec une utilisation d'un algorithme MRCG. il couvrira tous les aspects de la partie pratique de notre travail de recherche en englobant les divers schémas. ces derniers, à travers des algorithmes et des méthodes appliqués, illustrent dans la mesure du possible tout ce qui a été présenté dans la partie théorique.

2. Environnement de travail

2.1. Outil de développement (Matlab R2008a)

est un langage de programmation de quatrième génération émulé par un environnement de développement du même nom ; il est utilisé à des fins de calcul numérique. développé par la société The MathWorks, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et fortran. les utilisateurs de MATLAB (environ un million en 20041) sont de milieux très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel que pour la recherche. Matlab peut s'utiliser seul ou bien avec des toolbox (« boîte à outils ») [14].

3. Description de système de vérification du locuteur

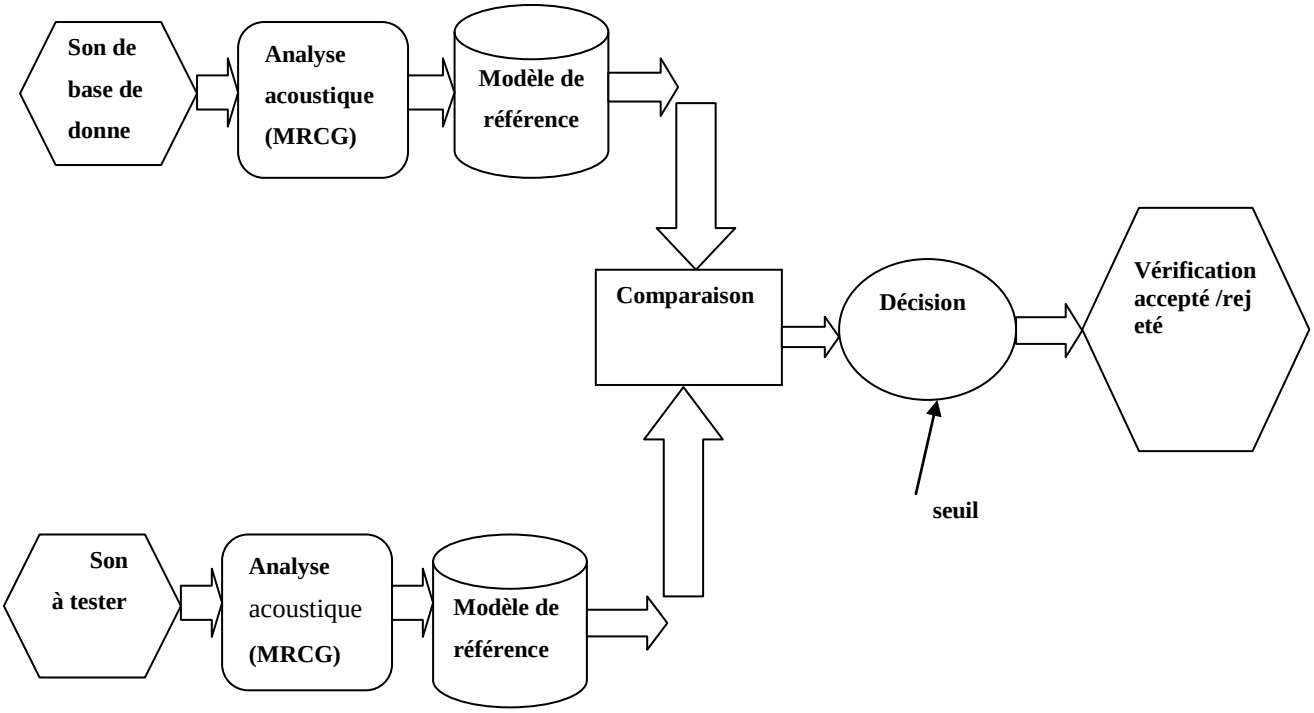


Figure III.1. *Architecture général de notre application*

3.1. Construction de la base de données

Le mot ***Bienvenue*** a été enregistré par 2 personnes. chacun a répété le mot 3 fois.

Dans cette partie nous proposons le code MATLAB pour enregistrer le son à partir de microphone et aussi pour lire un son. la figure ci-dessous montre le code associé pour l'enregistrement

```
%------------Enregister le son :-------------%
%------le son est automatiquement enregistré sous format .wav
fs=16000.
Temps=1.5;
%------------Fonction d'enregistrement-----------%
Son_enregistre= wavrecord(Temps*fs,fs,'int16')
Wavplay(Son_enregistre,fs);%-----Fonction pour écouter le son
```

Figure III.2. *Le code MATLAB pour enregistrer le son.*

D'abord, la fonction d'enregistrement exige au programmeur la définition de certains paramètres avant d'enregistrer le son à savoir : le nombre d'échantillons, le nombre de bit par échantillon et la fréquence d'échantillonnage. Le format de la commande est comme suit

```
« son_enregitré = wavrecord (le nombre d'échantillon, fréquence
d'échantillonnage, nombre de bit par échantillon) »
```

Ensuite, nous choisissons aussi bien le temps d'enregistrement « Temps = 1.5 secondes», que la fréquence d'échantillonnage « fs = 16000 HZ ». nous savons que le nombre d'échantillon est égal à : la durée d'enregistrement multipliée par la fréquence d'échantillonnage. nous obtenons à la fin la taille de vecteur « son_enregitré = 24000 x1 ».

Pour lire un son à partir d'un disque dur, nous utilisons la commande «wavread» qui associe un vecteur de valeurs au son lu. la figure qui suit montre le code Matlab associé pour la lecture.

```
[son_enregitré, fs, nbit] = wavread ('A8.wav') ;

 Wavplay (A8, fs) ; -------------fonction pour écouter le son
```

Le tracé du signal parole client du mot «Bienvenue» est montré par la figure III.2 :

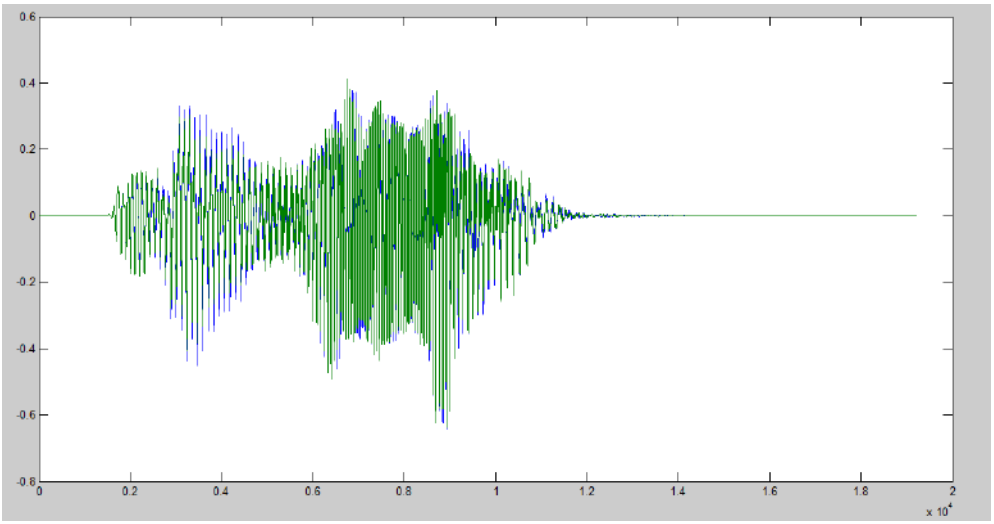


Figure III.3. *Représentation du signal parole de mot «Bienvenue »*

3.2. Analyse acoustique(MRCG)

Comme on a déjà expliqué le MRCG dans le chapitre précédant, on applique cette méthode pour les sons de la base de donné.

```
for i=1:nbr_par
    for j=1:nbr_ech_par
 [namefile,pathname]=uigetfile('*.wav','select the .wav file ');
         [V,fs]=wavread(namefile);
         mrcg = MRCG(V, fs);
    end
end
```

Sachant que nbr_par: nombre de personne=2, et nbr_ech_par :nombre d'échantillons de chaque personne =3.

Donc pour sélectionner et lire les sons on utilise la fonction suivant :

```
    [namefile, pathname]=uigetfile ('*.wav','select the .wav file ');
    [V, fs]=wavread(namefile);
```

3.3. La comparaison

Phase1

A la fin du calcul des MRCG, nous avons calculé les centroid des coefficients MRCG

et pour trouver les valeurs de centroid de chaque colonne :

```
for j=1:c1
     cbc1(1,j)=sqrt(sum(var(1:r1,j).^2))/r1;
end;
```

Et pour trouver les valeurs de centroid de chaque ligne :

```
for j=1:r1
        cbr1(1,j)=sqrt(sum(var(j,1:c1).^2))/c1;
end;
```

Avec :

```
[r1,c1]=size(var)
```

Tel que :

```
cb (1,1:r1)=cbr1 (1, :);
cb (1, r1+1:r1+c1)=cbc1 (1, :);
```

On mettre les valeurs des centroids dans un vecteur « «cb » contient les valeurs de colonnes de '1' jusque' r1' après les valeurs centre de ligne de' r1+1' jusque 'r1+c1'

Puis on fait l'appel à la fonction « centroid » dans le programme principal, et dans notre cas on le fait deux fois, un appel pour le MRCG de signal de référence « BD » de chaque échantillon et un deuxième fois pour le signal à tester « TEST ».

L'appel est comme suit :

```
cd ((nbr_ech_par*(i-1))+j,:)=centre(mrcg);
cb(1,:)=centr(mrcg); % ------- signal tester----
```

On stoke le vecteur cd dans un base de donne cdb « `cdb=cd;` »

```
save ('BD','cdb')
load BD
```

Phase 2 : Comparaison (distance euclidienne)

On compare les paramètres du modèle à tester «cb » d'un individu quelconque aux paramètres du modèle de référence « cdb » issue de centroid extrait du signal client utilisant la méthode de distance euclidienne. ce type de méthode réalise une bonne performance dans le texte dépendant pour la vérification du locuteur. Sa fonction peut être donnée comme suit :

```
function op1=cmpr(cb,cdb,nbr_ech_par,k,r,c)

op1(1:nbr_ech_par)=0;

for j=1:nbr_ech_par % finding euclidian distance

  x(j)=sqrt(sum((cb(1,1:r)-cd((nbr_ech_par*(k-1))+j,1:r)).^2));

    y(j)=sqrt(sum((cb(1,r+1:r+c)-cd((nbr_ech_par*(k-))+j,r+1:r+c)).^2));

end;
```

Avec k: nombre de personne dans la base de donnée, r: nombre de ligne, c: nombre de colonnes,

La première équation x(j) sert à calculer la distance entre les points de ligne du chaque vecteur, et la 2ème y(j) entre les composantes de la suite du vecteur, on a calculé deux valeurs pour nous aider à effectuer le seuil de décision.

3 .4. Dicision

3 .4. 1. Tests **:**

Trouver un seuil performant n'est pas facile en réalité, pour un système de vérification du locuteur dont la vérification est basé sur la comparaison, la définition du seuil n'a pas une loi claire ou bien un algorithme à suivre, donc ce n'était pas facile pour nous de trouver le seuil qui convient.

Et donc nous avons adopté à faire plusieurs test et définir un seuil manuellement. pour la construction de la basse de donnés est faite pour 3 échantillons de parole pour chaque

personne, pour le teste nous avons d'abord choisi des signaux de la base de donne ensuite de signaux qui n'existe pas de la base formé.

3 .5.vérification

A partir les valeurs de comparaison on définit une valeur de seuil, si nous avons des valeurs de x et y inférieurs au seuil (Correct) sinon (Incorrect) jusqu'à obtention de seuil performant. Pour cela nous avons fixé au début un seuil provisoire et le modifier à chaque fois suivant les résultats obtenues.

On va présenter ci-dessous quelques résultats de test :

- **Test 1 Seuil1 1 :** x(i)=0.001 && y(j)<0.5

Signal de test	**Résultat**
Vrai signal 1(A1)	Correct
Vrai signal 2(A2)	Correct
Vrai signal 3(A3)	Correct
Vrai signal 4(B1)	Correct
Vrai signal 5(B2)	Correct
Vrai signal 6(B3)	Correct
Faux signal 1(A4)	Incorrect
Faux signal 2(A5)	Incorrect
Faux signal 3(B4)	Incorrect
Faux signal 4(B5)	Incorrect
Faux signal 5(C1)	Incorrect
Faux signal 6(C2)	Incorrect
Faux signal 7(C3)	Incorrect
Faux signal 8(C4)	Incorrect
Faux signal 9 (C5)	Incorrect

Tableau III.1.performance de système

	Nombre totale	Nbr d'acceptation	Nombre de rejet	Performance
Son acceptée	10	6	4	60%
Son rejeté	5	0	5	100%

❖

❖ **Test 2 Seuil 2 :** x(i)=0.9 && y(j)<0.9

Signal de test	Résultat
Vrai signal 1(A1)	Correct
Vrai signal 2(A2)	Correct
Vrai signal 3(A3)	Correct
Vrai signal 4(B1)	Correct
Vrai signal 5(B2)	Correct
Vrai signal 6(B3)	Correct
Faux signal 1(A4)	Incorrect
Faux signal 2(A5)	Incorrect
Faux signal 3(B4)	Correct
Faux signal 4(B5)	Correct
Faux signal 5(C1)	Correct
Faux signal 6(C2)	Correct
Faux signal 7(C3)	Incorrect
Faux signal 8(C4)	Incorrect
Faux signal 9 (C5)	Incorrect

Tableau III.2.performance de système

	Nombre totale	Nbr d'acceptation	Nombre de rejet	Performance
Son acceptée	10	8	2	80%
Son rejeté	5	2	3	60%

❖ **Test 3 Seuil1 3 :** x(i)=0.8 && y(j)<0.7

Signal de test	Résultat
Vrai signal 1(A1)	Correct
Vrai signal 2(A2)	Correct
Vrai signal 3(A3)	Correct
Vrai signal 4(B1)	Correct
Vrai signal 5(B2)	Correct
Vrai signal 6(B3)	Correct
Faux signal 1(A4)	Incorrect
Faux signal 2(A5)	Incorrect
Faux signal 3(B4)	Correct
Faux signal 4(B5)	Incorrect
Faux signal 5(C1)	Incorrect
Faux signal 6(C2)	Correct
Faux signal 7(C3)	Incorrect
Faux signal 8(C4)	Incorrect
Faux signal 9 (C5)	Incorrect

Tableau III.2.performance de système

	Nombre totale	**Nbr d'acceptation**	**Nombre de rejet**	**Performance**
Son acceptée	10	7	3	70%
Son rejeté	5	1	4	80%

Dans ce cas aussi, La meilleure vérification dans ce test est celle où on a trouvé La performance de système est 70% pour les sons juste, et de 80% pour les sons faux.

Voici le code principal de notre application

```
clear
clc
nbr_par=2;
nbr_ech_par=3
for i=1:nbr_par
    for j=1:nbr_ech_par
[namefile,pathname]=uigetfile('*.wav','selectionner le
fichier.wav');
          [V,fs]=wavread(namefile);
          mrcg = MRCG(V, fs);
          [r,c]=size(mrcg);
          cd((nbr_ech_par*(i-1))+j,:)=centr(mrcg);
    end
end
 cdb=cd;
 save ('BD','cdb')
 h = msgbox('Entrer le son teste','Ok','help');
 waitfor(h)
[namefile,pathname]=uigetfile('*.wav','selectionner le
fichier.wav');
 [test,fs]=wavread(namefile);
          mrcg = MRCG(test, fs);
          [r,c]=size(mrcg);
         cb(1,:)=centr(mrcg);
  load BD.mat;
nmatch(1:nbr_par)=0;
for k=1:nbr_par
   op=cmpr(cb,cdb,nbr_ech_par,k,r,c);
   nmatch(k)=nnz(op);
end;

%display(nmatch);
disply(nmatch)
```

Si en exécute cette code on obtient :

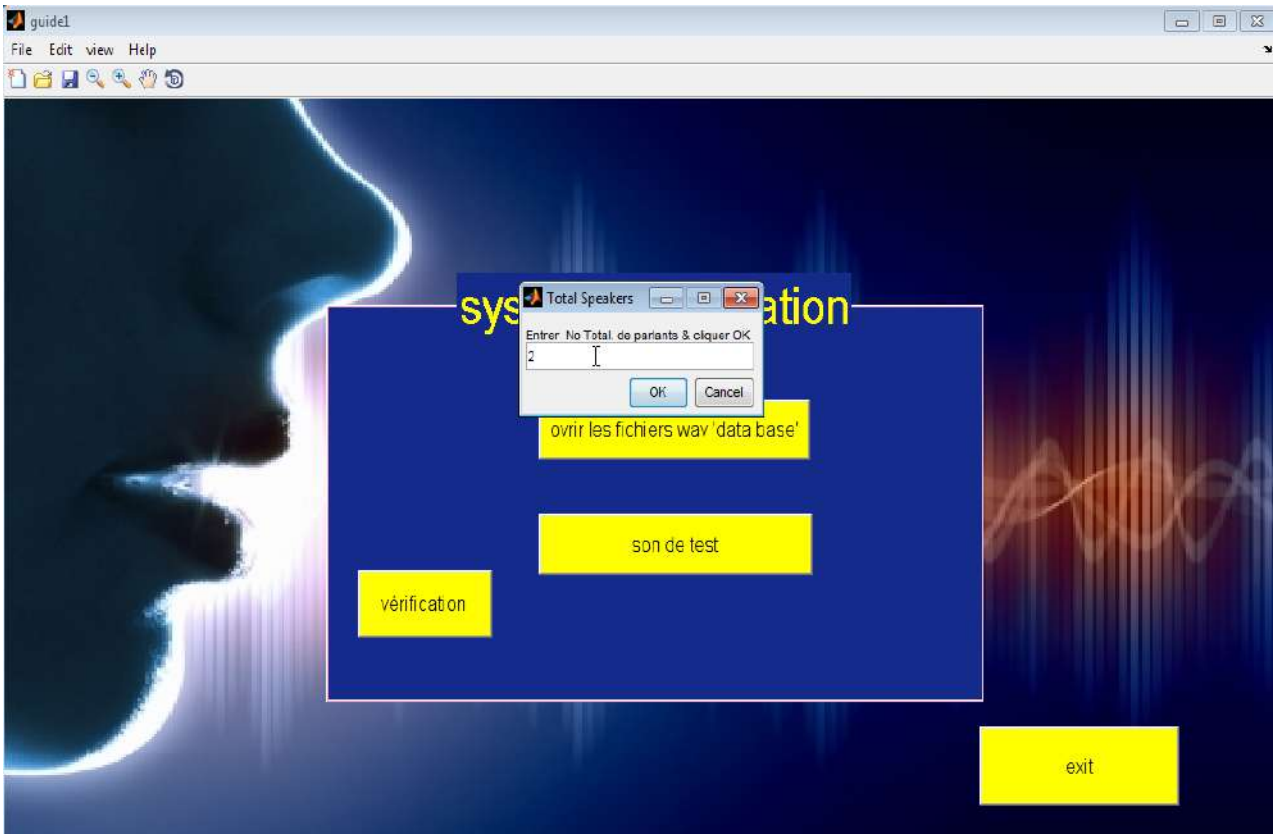


Figure III.4. *entrer numéro personne existe dans l'application*

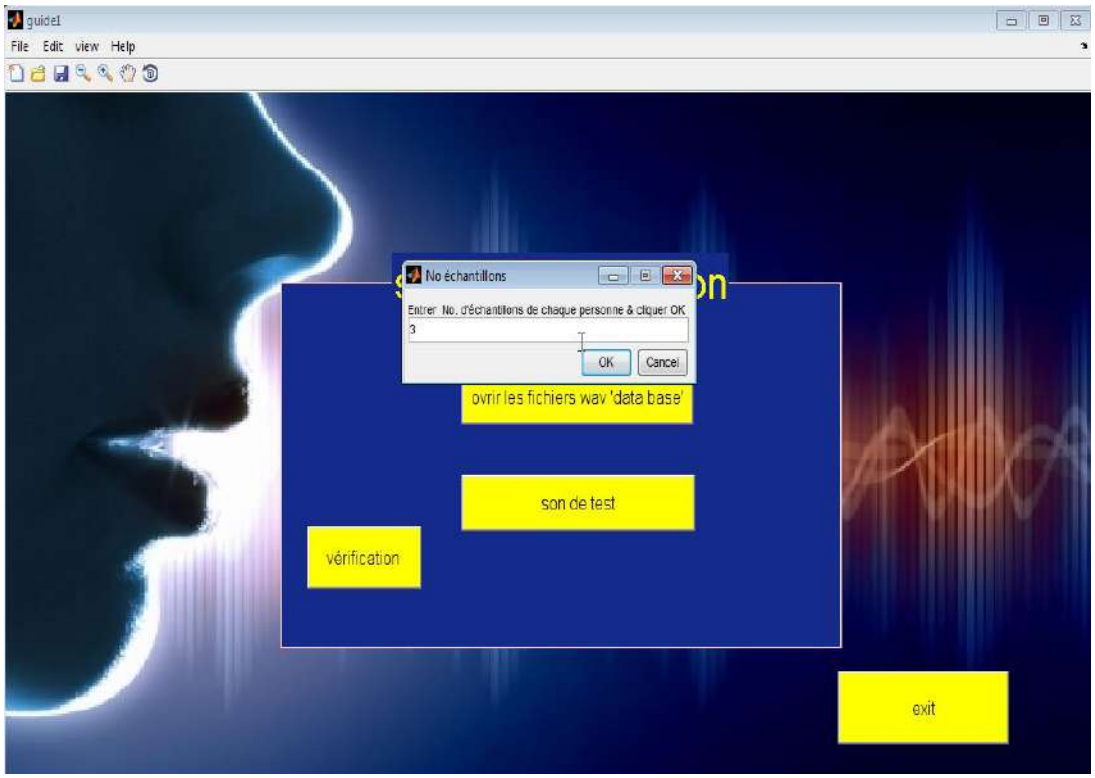


Figure III.5. *entrer numéro l'echantillon de chaque personne*

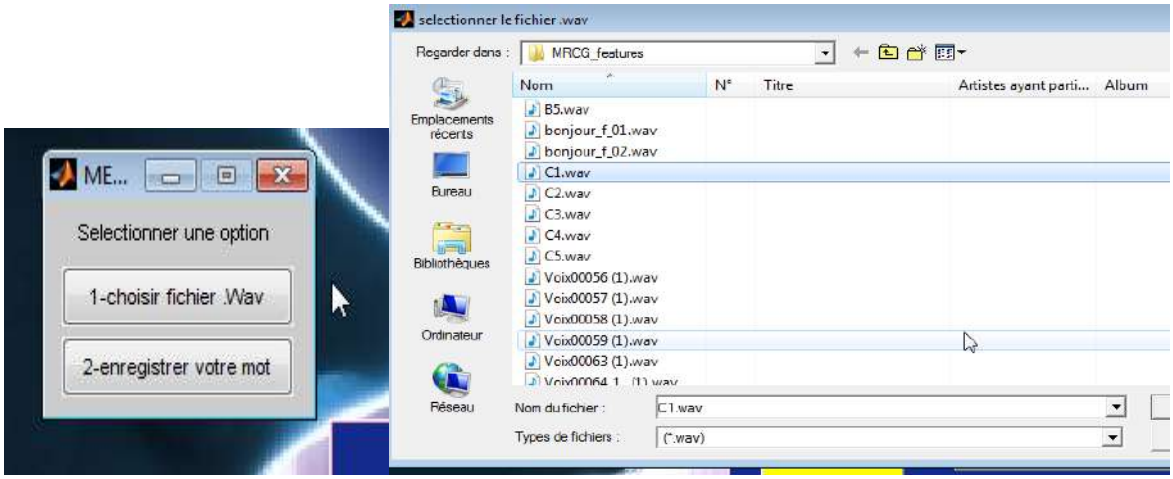


Figure III.6. *Choisir le fichier .wav*

S'il reconnait Le son de teste afficher



Figure III.7.*le son est correct*

Sinon



Figure III.8.*le son est incorrect*

Conclusion

Dans ce chapitre nous avons décrit notre système en expliquant comment on a créé notre base donnée, ensuite nous avons cité l'ensemble des processus appliqués pour la conception de notre application. Nous avons commencé par un organigramme qui décrit notre système puis, nous avons expliqué les étapes en commençant par l'enregistrement du son jusqu'à l'arrivé à la décision.

Conclusion générale

Le travail présenté dans ce mémoire consiste à étudier un nouveau modèle d'extraction de caractéristiques du signal de la parole et leur utilisation dans les systèmes de reconnaissance de la parole.

Ce travail s'inscrit dans le domaine de la RAP. Cette reconnaissance consiste à vérifier l'identité d'une personne à partir de sa voix conjointement. Utilisées principalement pour des raisons de sécurité et de confidentialité, ces systèmes de vérification sont souvent développés dans des applications de télésurveillance, télé service et d'accès à des endroits sécurisés.

Tout d'abord, on a donné une généralité sur MRCG et on a introduit les étapes essentielles de calcul MRCG en présentant les codes sources Matlab et les résultats de chaque étape. Ensuite on a fait la normalisation des coefficients MRCG, cette dernière est une tâche nécessaire avant de passer à la comparaison.

Enfin, la reconnaissance automatique du locuteur reste un domaine vaste et très important, où le but est toujours de proposer une application plus performante avec moins de probabilité de fausse acceptation et surtout moins de faux rejet

Bibioghraphie :

[1]rené boite, mura kunt, traitement de la parole, lausanne, presses polytechnique romandes, 1987

[2]rené boite, mura kunt, traitement de la parole, lausanne, presses polytechnique romandes, 1987

[3]yassamine azziza, modélisation AR et ARMA de la parole pour une vérification robuste du locuteur dans un milieu bruité en mode dépendant du texte, théme de magister, universite ferhat abbas –setif1- ufas (algerie),2013

[4] xiaofeng yang, auditory domain speech enhancement, master (engineering), queen's university kingston, ontario, canada may 2008

[5] ould ramoul amina , yahia helali nouha ,contribution a la conception d'un correcteur automatique des regles de recitation du coran « tajwid » (application a l'idgham),theme master,université yahia fares medea ,2014

[6]hacine gharbi abdenour , sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole ,thèse de doctorat, université ferhat abbas-sétif,2012

[7] ould ramoul amina , yahia helali nouha contribution a la conception d'un correcteur automatique des regles de recitation du coran « tajwid » (application a l'idgham),theme master,université yahia fares medea ,2014

[8] kami lia alouani , soumeya benterkia,vérification de locuteur, theme master , université yahia fares medea ,2014

[9] xiao-lei zhang, deliang wang,boosted deep neural networks and multi-resolution cochleagram features for voice activity detection, the ohio state university, columbus, oh, usa,2014

[10] jitong chen, yuxuan wang, and deliang wang a feature study for classification-based speech separation at low signal-to-noise ratios , fellow, ieee

[11]http://staffwww.dcs.shef.ac.uk/people/n.ma/resources/ratemap/cochleagram representation of sound

[12] xiaofeng yang, auditory domain speech enhancement, master (engineering), queen's university kingston, ontario, canada may 2008

[13] Kshitiz Kumar1, Chanwoo Kim2 and Richard M. Stern, Language Technologies Institute, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

[14] « https://fr.wikipedia.org/wiki/MATLAB consulté le 05/2017