الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA
FACULTE DES SCIENCES ET DE LA TECHNOLOGIE

Département du Génie Électrique et de l'Informatique



Projet de Fin d'Etude

Présenté pour l'obtention du Diplôme
De Master en Génie Électrique

Spécialité : Systèmes Intelligents et Communicants

Thème :

Vérification du Locuteur

Proposé et dirigé par

M. Billel ESSAID

Préparées et présentées par,

Melle. Kamilia ALOUANI

Mme. Soumeya BENTERKIA

Année Universitaire 2013-2014

« Le projet est le brouillon de l'avenir.

Parfois, il faut à l'avenir des centaines de brouillons »

« Jules Renard »

DEDICACE

Je dédie ce modeste travail :

A ma chère Mère

A mon cher papa

A mon cher Mari Aboubakre

A mes frères, Hamza, Youcef

A ma perle sœur Assia

A mon cher neveu Chiheb Eddine

A Toute ma famille

A tous mes amies…

SOUMEYA

DEDICACE

Je dédie ce modeste travail

A ma chère et tendre mère

A mon cher père

A Mon cher frère

Chakib

A mes amies

Kaouther, Nouha, Amina,Rima,Samira

A tous mes ami(e)s, tous ceux que j'aime, tous ceux qui m'aiment et tous

KAMILIA

REMERCIEMENT

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

Toute nos reconnaissances et remerciement vont à Mr le professeur ESSAID.B qui a accepté de nous encadrer, merci infiniment d'avoir nous orienté, conseillé et encouragé.

Nos sincères remerciements vont également à tous les Membres du jury, qui font l'honneur de juger notre travail.

Nos remerciements à un groupe d'amies qui nous a aidés pour effectuer les tests de notre application, merci infiniment les copines (les imposteurs).

Grand Merci à tous…

ملخص

التحقق من المتكلم هي عملية بيوميترية تهدف الى تأكيد أو نفي هوية الفرد استنادا إلى دراسة الخصائص التي ينفرد بها صوته، وتسمى هذه الدراسة تحليل الصوت.

الهدف من هذه الدراسة هو اقتراح تطبيق التحقق من المتحدث، يتم تنفيذ تحليل الصوت على كلام المتحدث المراد التحقق منه و على كلام المتحدث الذي بقاعدة بيانات و ذلك بتطبيق طريقة MFCC و من ثم الانتقال الى المقارنة بين الخصائص الناتجة عن عملية MFCC ، تتم هذه العملية من خلال حساب المسافة الاقليدية بين نتائج ال MFCC لكل من المتحدثين الاثنين.

الكلمات المفتاحية: التحقق من المتكلم، البيوميترية ، تحليل الصوتية ، MFCC ، النوافذ ، FFT ، DCT

Résumé

La vérification automatique du locuteur est un mécanisme de contrôle biométrique qui vise à confirmer ou infirmer l'identité d'un individu d'après une étude des caractéristiques spécifiques de sa voix, cette étude est appelée analyse acoustique.

L'objectif de cette thèse est de proposer une application de vérification du locuteur, une analyse acoustique est effectuée sur le signal parole du locuteur à confirmer et le locuteur de référence, en basant sur la méthode MFCC, ensuite passer à une tâche de comparaison entre les deux vecteurs MFCC 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 MFCC.

Mots clés : vérification du locuteur, biométrie, analyse acoustique, MFCC,FFT, DCT

Abstract

Speaker verification is a biometric task which aims to confirm or deny the identity of a person basing on a study of the specific characteristics of his voice, this study called acoustic analysis.

The object of this thesis is to propose an application of speaker verification, an acoustic analysis is performed on the speaker's voice signal to be confirmed and the reference speaker's voice, basing on MFCC method, then move to the comparison between the two MFCCs vectors in order to take the final decision accept or reject, this task is performed by the method of calculating the Euclidean distance between the two MFCCs vectors.

Keywords: speaker verification, biometric, acoustic analysis, MFCC, DCT,FFT,windowing

Sommaire

Sommaire

LISTE DES FIGURES ..... **I**

LISTE DES TABLEAUX ..... **II**

LISTE DES ACRONYMES ..... **III**

INTRODUCTION GENERALE **1**

CHAPITRE I :

Reconnaissance Automatique

Du Locuteur

1. Introduction **3**

2. La Biometrie **3**

3. Le système biométrique **4**

3.1. Identification 4
3.2. Authentification 4

4. Caractéristiques communes des systèmes biométriques **5**

4.1. Unicité 5
4.2. Caractère public d'une donnée biométrique 5
4.3. Mesure d'un système biométrique 5

5. Application de la biométrie **6**

6. Principales technologies biométriques **7**

6.1. Empreintes digitales 7
6.2. Visage 7
6.3. Iris 8
6.4. Rétine 8
6.5. Voix 9

7. Comparaison entre les différentes technologies biométriques **10**

8. Le Système de la reconnaissance automatique du locuteur RAL **11**

8.1. Le signal parole 12
8.2. Application d'un système RAL 13
8.3. Identification automatique du locuteur 14
8.4. Vérification automatique du locuteur 15
8.5. Performance d'un systéme VAL 17

9. Conclusion **18**

Sommaire

CHAPITRE II :

Extraction des paramètres du signal de parole

1. Introduction 19

2. Extraction de paramètres (Analyse acoustique) 19

2.1. Prétraitement 20
2.2. Analyse à court terme 20
2.3. Méthodes d'analyse acoustique. 20
2.3.1. LPC (Linear Prediction Coding) 20
2.3.2. PLP (Perceptral Linear Predection) 21
2.3.3. MFCC (Mel Frequency Cepstrum Coefficients). 21
2.3.4. Etude comparative entre les différentes méthodes 22

3. Analyse acoustique utilisant MFCC 22

3.1. Acquisition du signal 23
3.2. Préaccentuation 24
3.3. Segmentation 24
3.4. Fenêtrage 25
3.5. Analyse cepstral 26
3.5.1. Transformer de Fourier discréte DFT 26
3.5.1. Echelle Mel 27
3.5.1. Transformer en cosinus discréte 28

4. Méthodes de classifications 29

4.1. Méthode indépendante du texte (Mélange de Gaussiennes GMM) 29
4.2. Méthode dépendante du texte (Modèle de Markov Cachés HMM) 30

5. Méthodes de vérification 30

5.1. Quantification vectorielle VQ 30
5.2. Distance euclidienne 31

6. Conclusion 32

CHAPITRE III : Conception

1. Introduction 33

2. Environnement de travail 33

2.1. Environnement matériel 33

Sommaire

2.2. Outil de developpement (Matlab R2012b) 33

3. Description de systéme de vèrification du locuteur 34

3.1. Signal de parole (Locuteur) 34
3.2. Phase 1 (Analyse acoustique) 35
- Etape 1 37
- Etape 2 38
- Etape 3 40
- Etape 4 40
- Etape 5 41
- Etape 6 42

✓ Normalisation……43

3.3. Phase 2 Comparaison(distance euclidienne) 44
- Décision 45
- Tests 45
- Résultats de tests 46

4. Présentation de l'application 50

4.1. Présentation de l'interface 50

5. Conclusion 54

Conclusion Générale 55

Bibliographie 56

Liste des

Figures

LISTE DES FIGURES

Figure I.1. *Empreinte digitale acquise par un capteur optique* 7
Figure I.2. *Reconnaissance du visage* 7
Figure I.3. *Image d'Iris* 8
Figure I.4. *Détails d'une rétine* 9
Figure I.5. *Mot de passe vocal* 9
Figure I.6 *Comparaison des avantages et inconvénients applicatifs de différentes technologies biométriques* 10
Figure I.7. *Signal parole du mot « bonjour »* 13
Figure I.8. *schéma de principe de l'identification du locuteur* 14
Figure I.9. *Exemple d'indentification du locuteur* 15
Figure I.10. *Schéma de principe de vérification du locuteur* 16
Figure I.11. *Exemple e vérification du locuteur* 16
Figure II.1.*Prétraitement acoustique du signal* 19
Figure II.2. *Calcul des paramètres LPC* 20
Figure II.3. *Calcule des paramètres PLP* 21
Figure II.4. *Diagramme de bloc MFCC* 23
Figure II.5.*Visualisation d'un signal parole* 24
Figure II.6. *Fenêtre de hamming* 26
Figure II.8.*L'éhelle Mel* 27
Figure II.9. *La répartitiondes filtres triangulaires sur les échelles Fréquentielle et Mel* 27
Figure III.1. *Architecture général de notre application* 34
Figure III.2. *Représentation temporelle du signal parole de mot « vérification »* 35
Figure III.3. *Organigramme de L'analyse acoustique* 36
Figure III.4. *Signal parole « vérification » après préaccentuation* 38
Figure III.5. *Tracé des segments du signal préaccentué* 38
Figure III.6. *Schéma de Découpage en trames* 39
Figure III.7. *Signal segmenté après fenêtrage par Hamming* 40
Figure III.8. *Représentation fréquentielle du signal* 41
Figure III.9. *Banc de filtre Mel* 42
Figure III.10. *Coefficients MFCC* 43
Figure III.11. *Interface graphique* 50
Figure III.12. *Entrer le fichier audio ''.wav''* 51
Figure III.13. *Menu : Donner votre mot de passe* 52
Figure III.14. *Entrer le fichier ''.wav'' à tester* 52
Figure III.15. *Accès autorisé* 53
Figure III.16. *Enregistrement de mot de passe oralement* 53
Figure III.17. *Accès refusé* 53

Liste des

Tableaux

LISTE DES TABLEAUX

Tableau III.1. Paramètres utilisé en MFCC 37

Tableau III.2. Calcule des FR du test1 47

Tableau III.3. Calcule des FAdu test1 47

Tableau III.4. Calcule des FR du test2 48

Tableau I.5. Calcule des FA du test2 48

Tableau III.6.Calcule des FR du test3 49

Tableau III.7. Calcule des FA du test3 49

Liste des

Acronymes

LISTE DES ACRONYMES

DFT: Transformer de Fourier discrete

DTW: Dynamic time warping

FFT: Dynamic time warping

GMM: Gaussian Mixture Model

HMM: Hidden Markov Model

HTK: Hidden Markov Model Toolkit

IAL : Identification automatique du locuteur

LFCC: Linear Frequency Cepstral Coefficients

LPC: Linear Prediction Coding

DCT: Discret CosinTransform

LPCC: Linear Prediction Cepstral Coefficients

MFCC: Mel Frequency Cepstral Coefficients

mp3: Mpeg Audio Layer 3

mp4: Mpeg Audio Layer 4

mp6: Mpeg Audio Layer 6

MSG: Modulation SpectroGram

PIN: Personal Identification Number

PLP: Perceptual Linear Prediction

RAL: Reconnaissance Automatique de la Parole

RAP: Reconnaissance Automatique de la Parole

RIF: Response Impulsionelle Finie

DE: Distance Euclidienne

VAL: Vérification Automatique du Locuteur

Introduction

Générale

Introduction Générale

Durant la dernière décennie, on a vu l'apparition de plus en plus de domaines et d'applications nécessitant l'authentification des personnes : contrôle d'accès a des bâtiments sécurisés, transmission de données personnelles, recherche et surveillance d'individus, contrôle aux frontières, justice, domotique ... Ces applications utilisent généralement des technologies d'authentifications basées sur la biométrie. On parle dans ce cas-là d'une authentification biométrique. L'authentification biométrique est la reconnaissance automatique d'une personne en utilisant des traits distinctifs, i.e., des caractéristiques physiques (biologiques) ou traits comportementaux personnels automatiquement mesurables, robustes et distinctifs qui peuvent être utilisées pour identifier ou vérifier l'identité prétendue d'un individu. Certaines technologies biométriques constituent une solution efficace relativement simple et pas chère qui assure de bonnes performances en reconnaissance.

Elles offrent beaucoup plus d'avantages que les méthodes classiques d'authentification telles que l'utilisation de mots de passe ou de clés et badges d'accès, qui sont très vulnérables au vol. Les technologies biométriques fournissent en effet encore plus de sureté et de confort d'utilisation.

On trouve deux grandes taches en reconnaissance : l'identification et la vérification. Dans la première, le système biométrique pose et essaye de répondre a la question suivante : qui est la personne X ? Le système requiert une information biométrique et la compare avec chaque information stockée dans la base de données des utilisateurs autorisés ; c'est une comparaison un parmi plusieurs. Une des applications d'identification est la recherche de criminels et de terroristes en utilisant des données de surveillance. Dans la vérification, le système biométrique demande à l'utilisateur son identité et essaye de répondre à la question : est-ce la personne X ? Dans cette tâche, l'utilisateur annonce son identité par l'intermédiaire, par exemple, d'un numéro d'identification ou d'un nom d'utilisateur, puis le système sollicite une information biométrique provenant de cette personne pour la comparer avec la donnée préalablement enregistrée qui correspond à l'identité prétendue. C'est une comparaison vrai/faux. Le système trouvera finalement ou non une correspondance entre les deux. La vérification est communément employée dans des applications de contrôle d'accès et de paiement par authentification.

Au cours de ce travail, nous nous intéressons à la vérification automatique du locuteur en mode dépendent de texte, la vérification automatique du locuteur VAL consiste à comparer un signal de parole de la base de donnée (client) avec le signal parole à tester s'il est client ou imposteur, quand une personne veut accéder à un moyen quelconque, le système VAL lui a demandé de prononcer le mot de passe, ce dernier peut être un simple mot ou une phrase, dès que la personne termine le système commence à analyser le signal parole du message prononcé pour qu'il le compare avec le signal parole de la base de donnée, cette analyse est effectué par l'extraction des paramètres acoustique des deux signaux, ensuite le système effectuera une comparaison entre les paramètres acoustiques résultant de chaque signal, cette comparaison permettra au système VAL de décider si la personne est la bonne personne (client) ou la personne est un imposteur.

Nous concluons que le système VAL qu'on va réaliser est basé essentiellement sur deux phases primordiale indispensables qui sont, l'extraction des paramètres et la comparaison, il existe plusieurs méthodes pour effectuer ces deux phases.

Dans le cadre de ce projet nous avons adopté aux calculs des coefficients MFCC pour la phase d'extraction de caractéristiques et au calcul de la distance euclidienne pour la deuxième phase, cette dernière sert à calculer la distance euclidienne DE entre le vecteur des coefficients MFCC du signal de référence et le vecteur des coefficients MFCC du signal test, le résultat est une valeur à comparer avec un seuil défini pour vérifier le locuteur.

L'objectif de cette thèse est de proposer une application de vérification du locuteur et améliorer les performances de certains problèmes. Le choix de la méthode de distance euclidienne DE comme une méthode de comparaison est pour proposer une autre méthode différente que les méthodes de classification tel que GMM et HMM, nous avons rencontré plusieurs thèses travaillons avec ces méthodes par contre la méthode de DE est une méthode proposé mais nous n'avons rencontrer aucune thèse travaillant avec, pour cela nous l'avons adopté pour voir son efficacité en VAL et sa performance en terme de taux des faux rejet et fausse acceptation du système VAL.

CHAPITRE I :

Reconnaissance Automatique

Du Locuteur

1. 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 [1].

Ce chapitre 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. Avant d'aborder cette étape, on va décrire dans ce chapitre les principes généraux et les problèmes de la RAP, ainsi que les différentes étapes constituant un tel système. Nous allons parler sur le système de Reconnaissance du locuteur et ses deux modes principaux, l'identification et la vérification du locuteur

Si on dit reconnaissance de la parole on dit la biométrie vocale, on va présenter alors, le principe de la biométrie avec ses technologies, empreinte digitale, forme de la main ou des doigts de la main, forme du visage, rétine de l'œil, iris de l'œil, reconnaissance de l'écriture, rythme de frappe au clavier, forme des veines de la main et la reconnaissance de la voix.

2. La Biométrie

Durant la dernière décennie, on a vu l'apparition de plus en plus de domaines et d'applications nécessitant l'authentification des personnes : contrôle d'accès à des bâtiments sécurisé, transmission de données personnelles, recherche et surveillance d'individus, contrôle aux frontières, justice, domotique... Ces applications utilisent généralement des technologies d'authentification basées sur la biométrie. On parle dans ce cas là d'une authentification biométrique. L'authentification biométrique est la reconnaissance automatique d'une personne en utilisant des traits distinctifs, autrement dit, des caractéristiques physiques (biologiques) ou traits comportementaux personnels automatiquement mesurables, robustes et distinctifs qui peuvent être utilisés pour identifier ou vérifier l'identité prétendue d'un individu [7].

Dans cette partie du chapitre 1, on va parler sur la biométrie en générale, sa définition, ses caractéristiques, ses technologies et ses applications…

3. Le système biométrique

Un système de contrôle biométrique est un système automatique de mesure basé sur la reconnaissance de caractéristiques propres à un individu : physique, comportement…

Le mot « biométrie » utilisé dans le domaine de la sécurité est une traduction de l'anglais « biometrics » qui correspond en fait au mot français anthropométrie.

Le mot français biométrie définit l'étude mathématique des variations biologiques à l'intérieur d'un groupe déterminé [2].

La biométrie est basée sur l'analyse de données liées à l'individu et peut être classée en trois grandes catégories :

- Analyse basée sur l'analyse morphologique. (empreinte digitale, forme de la main, trait du visage, réseau veineux de la rétine, iris de l'œil, voix, etc.)
- Analyse basée sur l'analyse comportementale. (dynamique du tracé de signature, frappe sur clavier).

L'identité d'un individu est l'ensemble des données de fait et de droit qui permettent d'individualiser quelqu'un :

- La vérification de l'identité conduit à l'identification,
- La preuve de l'identité conduit à l'authentification.

3.1. *Identification*

La vérification de l'identité se fait à partir d'une pièce d'identité (document officiel) : ni l'iris de l'œil, ni l'empreinte, ni la voix ne peut donner l'identité. Les personnes faisant l'objet d'une identification ont volontairement déposé leur identité.

La vérification de l'identité demande une base de référence et le but est de vérifier que l'identité de l'individu qui se présente existe bien dans la base de référence.

3.2. *Authentification*

L'authentification est réalisée en deux temps :

- Vérification de l'identité : la personne déclare son identité en se présentant au contrôle d'accès.
- Preuve de l'identité : Les éléments biométriques (empreinte, voix, visage, iris…) de la personne sont comparés avec le gabarit de cette personne, afin de vérifier si son identité est bien la bonne.

4. Caractéristiques communes des systèmes biométriques

4.1. *Unicité*:

Pour identifier ou authentifier une personne au sein d'une population donnée, il est nécessaire que la donnée biométrique utilisée soit unique à cette personne. L'empreinte digitale, la rétine et l'iris sont réputés pour présenter des caractéristiques uniques au sein de très large populations. En particulier, ces technologies permettent de distinguer les vrais jumeaux, et l'empreinte digitale est reconnue juridiquement comme identifiant d'un individu [6].

4.2. *Caractère public d'une donnée biométrique*

Un code personnel (PIN) est secret et doit le rester pour qu'un système de contrôle d'accès fonctionne. Une caractéristique biométrique n'est pas secrète. Elle peut être plus ou moins facilement capturée et imitée. Un système de contrôle d'accès biométrique doit donc prendre en compte cette menace et éliminer les artefacts construits pour le tromper [6].

4.3. *Mesure d'un système biométrique*

Un système biométrique n'utilise pas toute l'information contenue dans l'image ou le signal capté. Il en extrait certaines caractéristiques, ce qui réduit la quantité d'information, donc la capacité du système à reconnaitre l'unicité d'une donnée. Puis il effectue un calcul et obtient un résultat à partir des données recueillies.

Sa robustesse dépend du nombre de critères retenus et de la méthode de modélisation (ou de calcul) utilisée [6].

5. Application de la biométrie

Parmi les applications de la biométrie, nous citons :

- ❖ Contrôle d'accès physiques aux locaux
  - Salle informatique
  - Site sensible (service de recherche, site nucléaire)
- ❖ Contrôle d'accès logiques aux systèmes d'informations
  - Lancement du système d'exploitation
  - Accès au réseau informatique
  - Commerce électronique, paiement en ligne
  - Transaction (financière pour les banques, données entre entreprises)
  - Signature de document (lot de fabrication de médicaments)
  - Tous les logiciels utilisant un mot de passe
- ❖ Equipements de communication
  - Terminaux d'accès à Internet
  - Téléphones portables
- ❖ Machines & Equipements divers
  - Coffre-fort avec serrure électronique
  - Distributeur automatique de billets
  - Casier sensible (club de tir, police)
  - Cantine d'entreprise, cantine scolaire (pour éviter l'utilisation d'un badge par une personne extérieure et améliorer la gestion)
  - Casier de piscine (plus d'objet à porter sur soi)
  - Contrôle des adhérents dans un club, carte de fidélité
  - Contrôle des temps de présence
- ❖ Etat / Administration
  - Fichier judiciaire
  - Le Fichier national automatisé des empreintes digitales
  - Le fichier national automatisé des empreintes génétiques
  - Titres d'identité (carte d'identité, passeport, permis, titre de séjour)
  - Services sociaux (sécurisation des règlements)
  - Services municipaux
  - Système de vote électronique [6].

6. Principales technologie biométriques

6.1. *Empreintes digitales*

Empreinte digitale est la caractéristique d'un doigt. On le croit que chaque empreinte digitale est unique. Chaque personne a ses propres empreintes digitales avec l'unicité permanente. Ainsi les empreintes digitales sont utilisées depuis longtemps pour l'identification et l'investigation juridique [3].



Figure I.1. *Empreinte digitale acquise par un capteur optique*

C'est la technologie la plus ancienne, très fiable et la plus répandue, il existe différents types de captures, plus ou moins chers, plus ou moins robustes aux impostures (faux doigts) [4].

6.2. *Visage*

La reconnaissance de la forme du visage est la technologie la plus commune et populaire. Elle est la plus acceptable parce que on peut l'utilise à distance sans la collaboration avec objet.

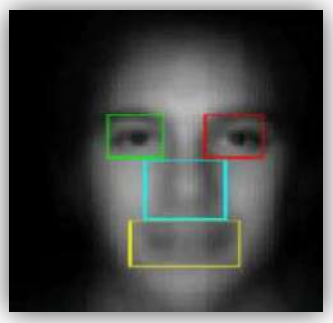


Figure I.2. *Reconnaissance du visage*

Utiliser une caméra permet d'acquérir la forme du visage d'un individu et puis retirer certaines caractéristiques. Les caractéristiques essentielles pour la reconnaissance du visage sont: les yeux, la bouche, le tour du visage, le bout du nez, etc. Le logiciel doit être capable d'identifier un individu malgré différents artifices physiques (moustache, barbe, lunettes, etc.)[3].

6.3. *Iris*

La reconnaissance de l'iris est considérée comme la technologie biométrique la plus exacte. L'iris est la partie colorée de l'œil qui entoure la pupille noire. Son inspection attentive révèle de nombreuses structures détaillées. Ces structures se développent dès le plus jeune âge et sont considérées comme uniques, même chez les vrais jumeaux. Les structures de l'iris resteront constantes jusqu'à la mort.

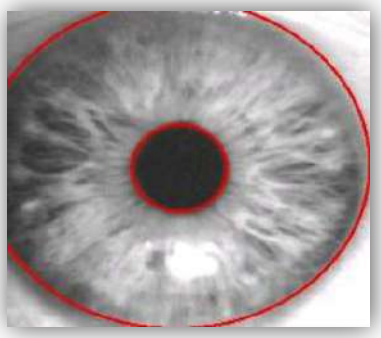


Figure I.3. *Image d'Iris*

Pour reconnaître un individu, on extrait des paramètres de l'image de l'œil. Puis, on compare ces paramètres avec tous les paramètres précédemment extraits et sauvegardés. La comparaison qui produira le meilleur score permettra d'identifier le nom de l'individu [3].

6.4. *Rétine*

La reconnaissance de la rétine est une méthode assez ancienne, les motifs formés par les veines sous la surface de la rétine sont uniques et stables dans le temps. Ils ne peuvent être affectés que par certaines maladies. Pour ces raison, la reconnaissance de la rétine est actuellement considérée comme une des méthodes biométriques les plus sures [3].

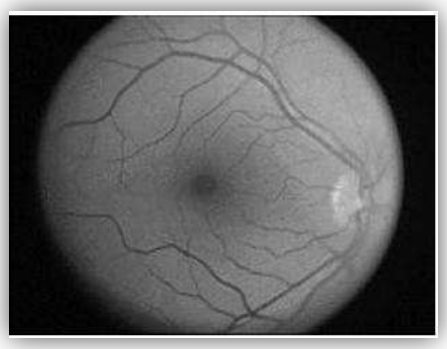


Figure I.4. *Détails d'une rétine*

6.5. *Voix*

L'identification de la voix est considérée par les utilisateurs comme une des formes les plus normales de la technologie biométrique, car elle n'exige aucun contact physique avec le lecteur du système.

Les systèmes d'identification de la voix se concentrent sur les seules caractéristiques de voix qui sont uniques à la configuration de la parole d'un individu. Ces configurations de la parole sont constituées par une combinaison des facteurs comportementaux et physiologiques.



Figure I.5. *Mot de passe vocal*

Lors d'une tentative d'accès, on compare la voix d'un utilisateur à la signature de la voix stocké dans la base des données. L'utilisateur doit prononcer une phrase (donnée ou pas) et un logiciel en extrait les paramètres.

Ensuite, le logiciel effectue la classification dans le cas d'une identification. Enfin, la fraude est possible par enregistrement. Par contre, les imitateurs essayent habituellement de reproduire les caractéristiques vocales les plus évidentes à l'oreille humaine et ne recréent pas

les caractéristiques moins accessibles qu'un système automatisé d'identification de voix analyse [3].

Comme il existe d'autres technologies de biométrie telle que, la signature dynamique, dynamique de frappe au clavier, Forme des veines de la main [5].

Dans le cadre de ce projet, ce qui nous intéresse c'est la reconnaissance de voix.

7. Comparaison entre les différentes technologies biométriques

Plutôt que de comparer les performances des diverses technologies (empreintes, visage, main…), il faut surtout tenir compte de l'environnement de leur usage, (facilité de : saisie, d'analyse, de stockage, de vérification). Chaque technologie possédant des avantages et des inconvénients, acceptables ou inacceptables suivant les applications.

Ces solutions ne sont pas concurrentes, elles n'offrent ni les mêmes niveaux de sécurité ni les mêmes facilités d'emploi.

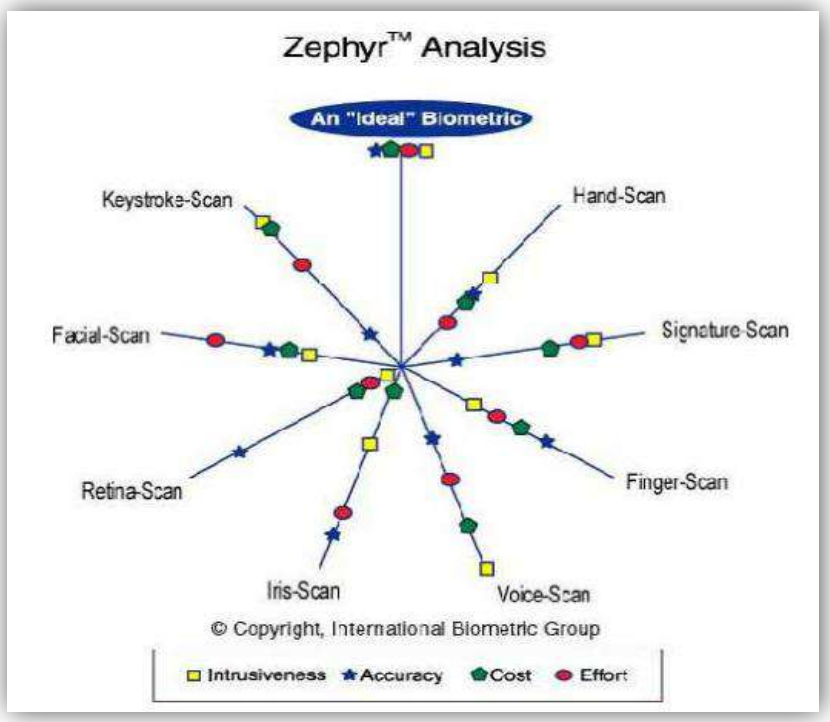


Figure I.6. *Comparaison des avantages et inconvénients applicatifs de différentes technologies biométriques d'après www.biometricgroup.com.*

La figure 1.6 appelée analyse de Zephyr présente une évaluation des avantages et inconvénients applicatifs de différents systèmes d'authentification biométrique basés sur

différents attributs biométriques courants. Sur cette figure, les quatre critères d'évaluation des systèmes sont :

- La précision (accuracy) : mesure objective des performances obtenues par le système,
- Le coût (cost) : mesure objective du coût de mise en œuvre et de fonctionnement du système,
- L'ergonomie (effort) : mesure subjective du nombre et de la difficulté des démarches à suivre par l'utilisateur lors de son utilisation,
- Le caractère intrusif (intrusiveness) : mesure subjective, dépendant fortement de la culture de l'utilisateur et permettant d'évaluer sa perception du système [6].

Selon les critères d'évaluation choisis, aucune caractéristique biométrique n'est globalement supérieure à une autre.

La biométrie vocale est considérée par la plupart des utilisateurs comme l'une des technologies biométriques la plus naturelle: elle n'est pas intrusive, n'exige aucun contact physique avec le système. De plus, elle correspond à une manière usuelle pour chaque individu de reconnaitre l'un de ses proches. Un autre avantage de la biométrie vocale est qu'elle est souvent la technologie la plus adaptée pour de nombreuses applications telles que la sécurisation de transaction bancaire téléphonique, etc. [6].

8. Le système de la reconnaissance automatique du locuteur RAL

La reconnaissance du locuteur est une des branches de l'authentification biométrique, qui se réfère à la reconnaissance automatique de l'identité des personnes en utilisant certaines de leurs caractéristiques intrinsèques. Outre la voix, il y a beaucoup d'autres modèles physique et comportementaux pour l'authentification biométrique, par exemple : les yeux, le visage, les empreintes, la signature, …etc.

Pratiquement, la sélection d'un modèle biométrique prometteur devrait prendre en compte au moins les considérations suivantes : la robustesse, la précision, l'accessibilité et l'acceptabilité. Par rapport à ces critères de sélection, parmi toutes les technologies d'authentification biométriques, la reconnaissance du locuteur est probablement la plus naturelle et économique pour les systèmes de communication homme-machine et cela en raison de la collecte de données parole est beaucoup plus pratique que les autres motifs, et plus important encore, la parole et le mode dominant d'échange d'information pour les

êtres humains et tend à être le mode dominant pour l'échange d'information pour les système de communication homme-machine.

La reconnaissance du locuteur peut être divisée en deux tâches: vérification et identification qu'on va les présenter par la suite.

Les systèmes de reconnaissance du locuteur RAL utilisent la variabilité interlocuteurs de production de la parole pour séparer les individus entre eux. Mais il y a aussi d'autres facteurs de variation qui s'ajoutent et qui altèrent la parole et nuisent à la robustesse des systèmes de reconnaissance du locuteur, comme les variabilités intra-locuteur.

Pour améliorer les performances de reconnaissance, la voix peut être utilisée en complément d'une autre modalité ; on parle dans ce cas là d'une authentification multimodale.

8.1. Le signal parole

- **La parole** est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps.
- **La parole humaine** est un flux continu constitué d'une suite de mots, eux-mêmes étant constitués d'un enchainement de phonèmes et de bruits articulatoires.
- **Un phonème** c'est la plus petite unité discrète ou distinctive d'un mot, auquel elle peut faire perdre son sens par substitution. Par exemple, «chou » et « pou » se distinguent par leurs phonèmes initiaux.

  Dans le signal de parole, on différencie des phonèmes voisés et des phonèmes non voisés
- La langue arabe contient 34 phonèmes, la langue anglais plus de 40 phonèmes et la langue française a 37 phonèmes
- **Phonèmes voisés** sont quasi périodiques, au contraire des **phonèmes non voisés**. En réalité, les phonèmes non voisés n'utilisent pas les cordes vocales, comme quand on siffle et qu'on utilise la bouche comme une cavité résonante, c'est le cas de certains phonèmes, 's' et 'f' entre autres, c'est aussi le cas quand on chuchote, exemple montre les zones voisés et les zones non voisés du signal parole « **bonjour** » sur la figure ci-dessus :

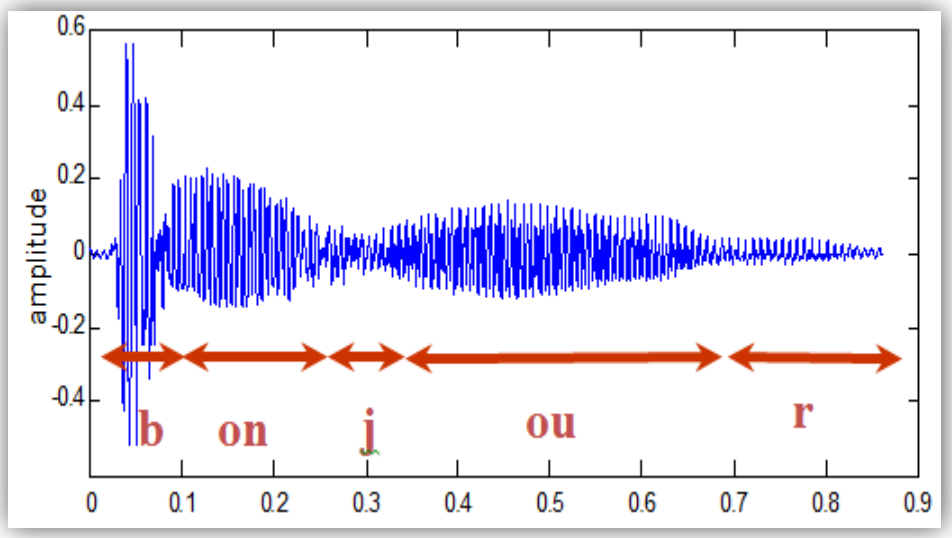


Figure I.7. *Signal parole du mot « bonjour »*

- **Pitch** est la fréquence fondamentale F0, un phonème voisé est quasi périodique ou pitch de période P (seconde) et de fréquence $F0 = \frac{1}{P}(Hz)$.

8.2. Applications d'un système RAL

Les avantages que l'on attend de la reconnaissance du locuteur sont multiples. Le système RAL 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. 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.

Un système RAL peut être divisé en deux grandes classes : l'identification et la vérification.

8.3. Identification automatique du locuteur

L'identification automatique du locuteur (IAL) a été une des premières utilisations de la RAL (Atal, 1976). En IAL, la liste des locuteurs à identifier est connue du système. Le système doit pouvoir décider, à partir d'un échantillon de voix, à quelle identité connue du système correspond l'échantillon, figure (Figure I.8) :

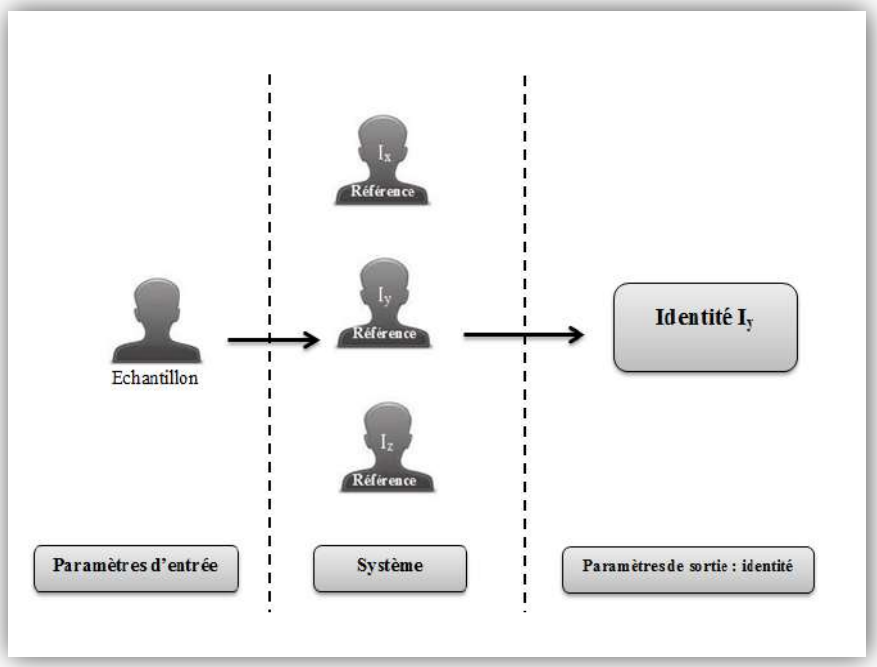


Figure I.8. *Schéma de principe de l'identification du locuteur*

L'identification automatique du locuteur se découpe en deux étapes : une étape d'apprentissage et une étape de test. À partir d'un ensemble d'enregistrements de voix de chaque locuteur, le système apprend un modèle pour chaque locuteur lors de l'étape d'apprentissage. Lors de l'étape de test, le système confrontera l'échantillon de voix qu'il recevra aux différents modèles qu'il aura déjà appris afin de déterminer si l'identité du locuteur est déjà connue.

En fonction de l'application, deux types de décisions sont possibles. Prenons le cas d'un centre d'appel téléphonique qui enregistre les conversations de ses employés. Un système d'IAL peut être utilisé pour classer chacun des enregistrements en fonction de l'employé concerné. Dans ce cas, la liste des locuteurs possibles est connue du système (tous les

employés), et aucun autre locuteur non employé ne peut être concerné. Dans ce cas, l'application présuppose que l'ensemble des locuteurs possible est fermé et connu du système. Le système de IAL choisira alors, parmi la liste, le modèle de locuteur le plus ressemblant à l'enregistrement de test.

En revanche, dans une application utilisant un ensemble ouvert de locuteurs possibles, le système d'IAL ne connait pas tous les locuteurs possibles. Dans ce cas, en plus de déterminer le locuteur le plus vraisemblable, le système a la possibilité de rejeter l'échantillon de test en ne renvoyant aucune identité connue pour cet échantillon [7].

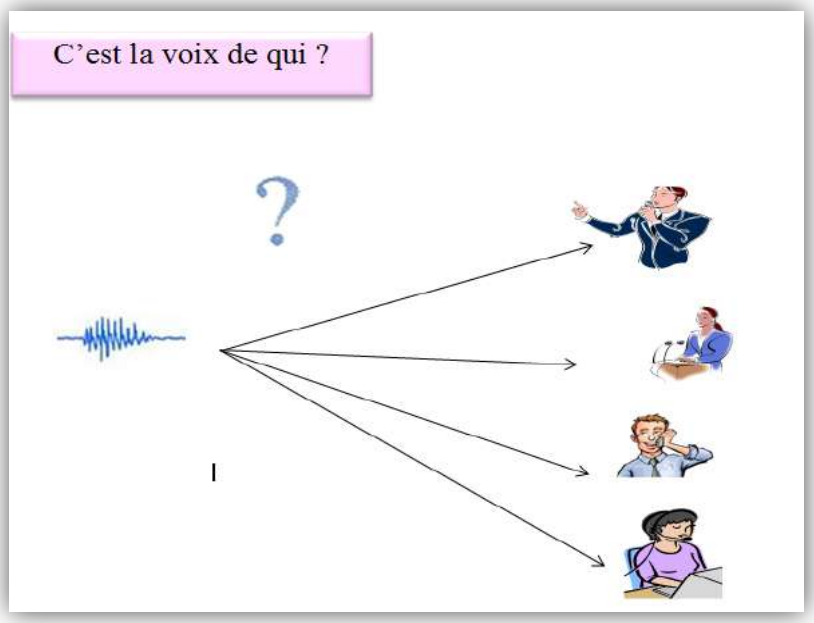


Figure I.9. *Exemple d'indentification du locuteur*

8.4. Vérification automatique du locuteur

La vérification automatique du locuteur (VAL) permet de décider si l'identité revendiquée par un locuteur est compatible avec sa voix. Il s'agit donc de trancher entre deux hypothèses : soit le locuteur est bien le locuteur autorisé (on l'appelle aussi locuteur client), c'est-à-dire que son identité correspond à celle qu'il revendique, soit le locuteur est un imposteur qui cherche à se faire passer pour la personne qu'il n'est pas. À partir d'un échantillon de voix de référence, et d'un échantillon de voix de test, le système va donc devoir dire si oui ou non les deux locuteurs correspondent [7].

La figure ci-dessus montre le fonctionnement d'un système VAL :

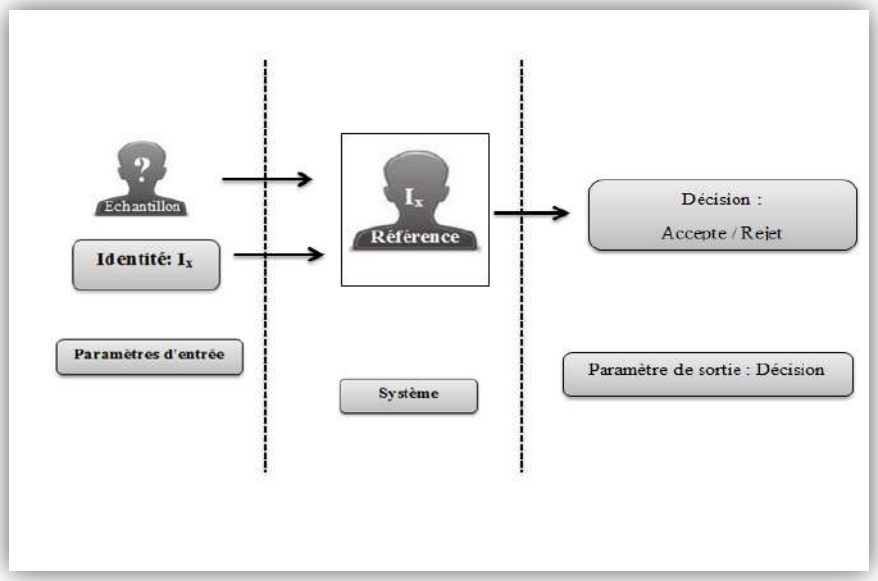


Figure I.10. *Schéma de principe de vérification du locuteur*

Les systèmes de VAL sont très dépendants des différences entre les échantillons de voix de référence et les échantillons de tests. Accepter un locuteur qui devrait être rejeté peut avoir de lourdes conséquences, en particulier dans les applications où un haut niveau de sécurité est demandé. (Contrôle aux frontières, système bancaire, identification judiciaire, etc.) [7].

La vérification du locuteur consiste à déterminer si l'identité proclamée d'un message vocal correspond à la véritable identité du locuteur. Et c'est l'objectif principal de cette mémoire.

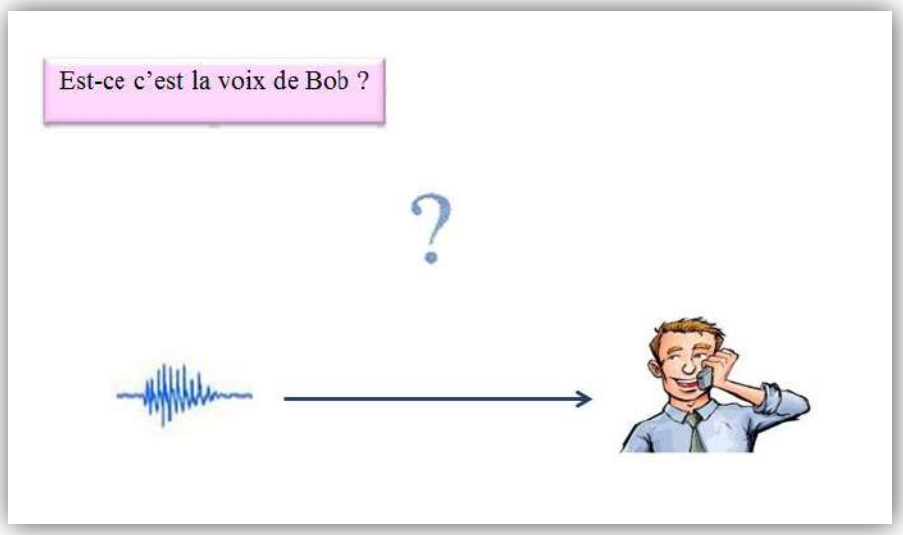


Figure I.11. *Exemple de vérification du locuteur*

Un système de vérification du locuteur peut être dépendant ou indépendant de texte :

- **Dépendante de texte :** être dépendante du texte il faut prononcer un texte précis pour être reconnu, un système est dit dépendant de texte s'il connait exactement le texte parlé.
- **Indépendante de texte :** peu importe ce qui est dit, Un système est dit indépendant du texte si la reconnaissance du locuteur ne lui impose pas de prononcer une séquence préétablie (prompteur ou mot de passe).

8.5. Performance d'un système de VAL

Un système de VAL peut être confronté à deux types de tests :

- **Tests clients :** lors desquels l'échantillon biométrique présenté au système correspond à l'identité clamée
- **Test imposteur :** lors desquels l'échantillon biométrique présenté au système provient d'un individu inconnu du système.

Le système doit répondre à chaque tentative d'authentification auquel il fait face par une décision binaire. Il peut donc engendrer deux types d'erreurs :

- **Faux rejet (FR) :** cette erreur sera commise quand le système rejette un client.
- **Fausse acceptation (FA)** : erreur commise lorsqu'un imposteur est malencontreusement accepté en tant qu'utilisateur légitime lors d'un test imposteur) [7].

Les mesures de performances d'un système de vérification se basent principalement sur le taux des fausses acceptations et le taux de faux rejets calculés comme le montrent les formules ci-dessous.

$$FA = \frac{nombre\ de\ tests\ \text{à}\ fausse\ acceptation}{nombre\ de\ test}$$

$$FR = \frac{nombre\ de\ tests\ \text{à}\ faux\ rejet}{nombre\ de\ test}$$

9. Conclusion

Dans ce chapitre nous avons introduit le concept de la biométrie en générale, ses technologies et ses applications. Nous avons aussi présenté le système de reconnaissances de locuteurs avec ses deux phases identification et vérification du locuteur, basant sur les systèmes de vérification du locuteur qui est le but de notre travail.

CHAPITRE II :

Extraction des paramètres du signal de parole

1. Introduction

Le signal parole est un signal acoustique qui contient des informations, la reconnaissance automatique du locuteur (RAL) ne considère que l'information acoustique que le signal contient. Dans un environnement bruité il est moins précis.

Le traitement de la parole peut être réalisé à trois niveaux différents. Un niveau de traitement de signal, un deuxième niveau de traitement de phonème, les phonèmes de la parole sont acquis et traitées, « Le phonème est l'unité de base du parole », un troisième niveau de traitement est connu comme le traitement au niveau de mot, ce modèle se concentre sur l'entité linguistique de la parole.

Dans ce chapitre, on décrira les méthodes d'extraction des caractéristiques acoustiques, comme LPC, PLP et MFCC [9].

2. Extraction de paramètres (Analyse acoustique)

Le signal vocal transporte plusieurs informations comme le message linguistique, l'identité du locuteur, ainsi que ses émotions, la langue adoptée, etc. Un système RAP consiste à récupérer seulement le message linguistique indépendamment des autres informations. Dans un tel système, l'analyse acoustique consiste à extraire du signal vocal un ensemble de paramètres pertinents dans le but de réduire la redondance du signal vocal pour une tâche de reconnaissance de la parole. Le nombre de ces paramètres doit rester raisonnable, afin d'éviter la nécessité d'un grand espace mémoire ce qui accroît le coût de calcul dans le module de décodage. Ces paramètres doivent être discriminants en rendant les sons de base facilement séparables. Ils doivent être robustes au bruit [1].

Le calcul des paramètres acoustiques est réalisé par une chaîne de prétraitement selon les étapes suivantes (figure II.1) :

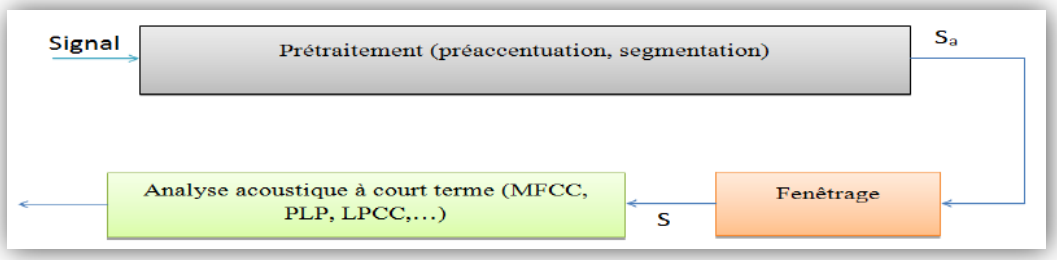


Figure II.1. *Prétraitement acoustique du signal vocal*

2.1.Prétraitement

Le prétraitement est une étape importante du processus du calcul des vecteurs acoustiques. L'objectif du prétraitement est de modifier le signal parole, X(n), pour qu'il soit plus convenable à l'étape d'extraction des paramètres. Le fichier de données du phonème prononcé par l'utilisateur est extrait du fichier audio et est échantillonné à 16 kHz. Une fois que le signal vocal est numérisé en temps et en amplitude, il peut être stocké et traité par un ordinateur.

Les étapes de prétraitements (élimination du silence, préaccentuation, segmentation, fenêtrage), vont être détaillées par la suite.

2.2. Analyse à court terme

Chaque trame fenêtrée du signal est ensuite convertie en un vecteur acoustique constitué d'un ensemble réduit de paramètres.

Différentes méthodes existent pour la transformation d'une trame fenêtrée de signal en un vecteur acoustique.

2.3.Méthode d'analyse acoustique

2.3.1. *LPC (Linear Prediction Coding)*

Il est souhaitable de compresser le signal pour une transmission efficace, un signal numérique est compressé avec la transmission pour une utilisation efficace des canaux ou media sans fil [9].

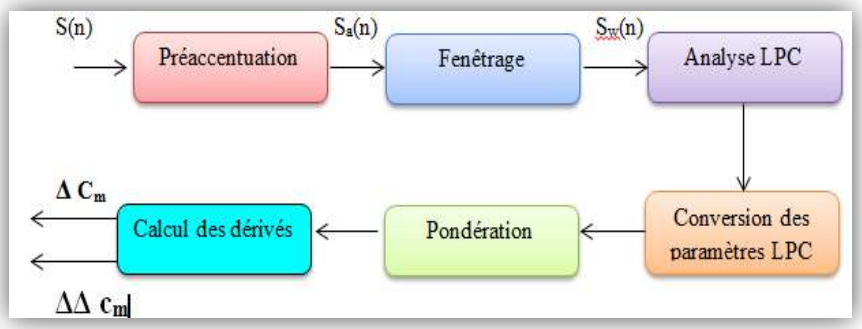


Figure II.2. *Calcul des paramètres LPC*

LPC calcule le spectre de puissance d'un signal. LPC est l'un des plus puissantes techniques d'analyse de signal parole et elle a gagné une popularité en tant que technique d'estimation des formants, les étapes de calcul LPC sont montrées sur la figure II.2.

2.3.2. PLP (Perceptual Linear Prediction)

Cette méthode de PLP (Prediction lineaire perceptuelle) est développée par Hermansky. PLP sert à supprimer l'information non pertinente du signal parole et donc améliore le taux de reconnaissance de la parole. Elle est identique à LPC, sauf que ses caractéristiques spectrales ont été transformées pour correspondre aux caractéristiques humaines [9]. La figure II.3 montre les étapes de calcul PLP

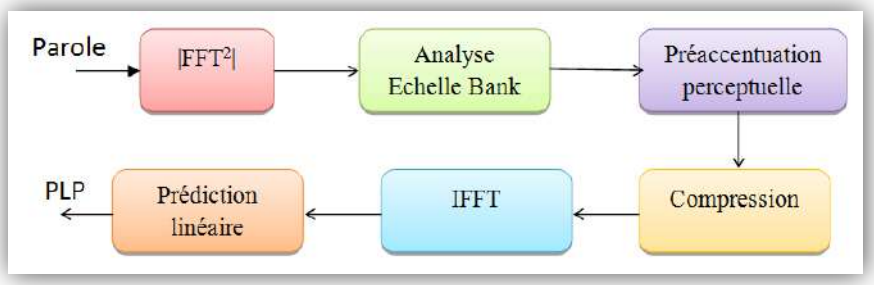


Figure II.3. *Calcule des paramètres PLP*

2.3.3. MFCC (Mel Frequency Cepstrum Coefficients)

La méthode la plus répandue et dominante utilisée pour extraire les caractéristiques spectrales est le calcul des coefficients cepstraux Mel-fréquences (MFCC).

Le principe de calcul des MFCC est issu des recherches psycho-acoustiques sur la tonie (La sensation de hauteur tonale) et la perception des différentes bandes de fréquences par l'oreille humaine. La FFT passe dans un banc de filtres à l'Echelle de Mel. Cette échelle non linéaire tient principalement compte du fait que la perception des intervalles change suivant la zone du spectre à laquelle les hauteurs qui les composent appartiennent. Le principal intérêt de ces coefficients est d'extraire des informations pertinentes en nombre limité en s'appuyant à la fois sur la production (théorie Cepstrale) et à la fois sur la perception de la parole (échelle Mel) [15].

Le calcul se déroule de la manière suivante :

- La FFT est calculée sur les trames.
- Cette dernière est filtrée par un banc de filtres triangulaires répartis le long de l'échelle de Mel.
- Le logarithme module de l'énergie de sortie du banc de filtres est calculé.
- Une Transformée en Cosinus est appliquée.
- Seuls les premiers coefficients sont conservés.

Nous citons, qu'il existe d'autres méthodes d'extraction comme RC, LFCC, et LPCC.

2.3.4. Etude comparative entre les différentes méthodes

Des études comparatives entre différents types de paramètres acoustiques ont été effectuées, pour déterminer ceux qui représentent mieux le signal vocal. Une étude comparative classique a été effectuée entre plusieurs représentations du signal vocal: cepstres en sortie d'un banc de filtres en échelle Mel (MFCC) ou en échelle linéaire (LFCC), coefficients de prédiction linéaire (LPC) ou de réflexion (RC), cepstre calculé à partir des coefficients autorégressifs (LPCC). Dans cette étude, les MFCC donnent les meilleurs résultats, ce qui montre plus généralement l'intérêt d'un prétraitement par banc de filtres, d'une échelle fréquentielle non linéaire, et de la représentation cepstrale.

En se basant sur les résultats de recherche décrits précédemment, nous avons retenu dans notre travail les coefficients MFCC. Par la suite on va détailler le MFCC, ses étapes et ses phases nécessaires, les méthodes de calculs des coefficients MFCC avec les résultats que nous avons obtenus [1].

3. Analyse acoustique utilisant MFCC

L'extraction des paramètres acoustiques d'un signal est une tâche importante pour produire une reconnaissance performante.

MFCC est une technique basé sur la perception auditive humaine qui ne peut pas percevoir des fréquences supérieures à 1Khz, en d'autres termes, MFCC est basé sur la variation connue de la largeur de bande passante perceptible par l'oreille humaine [11].

Cette technique consiste à exploiter les propriétés du système auditif humain par la transformation de l'échelle linéaire des fréquences en échelle de Mel [1].

L'analyse MFCC comporte plusieurs étapes qui sont représentées dans la figure II.4.

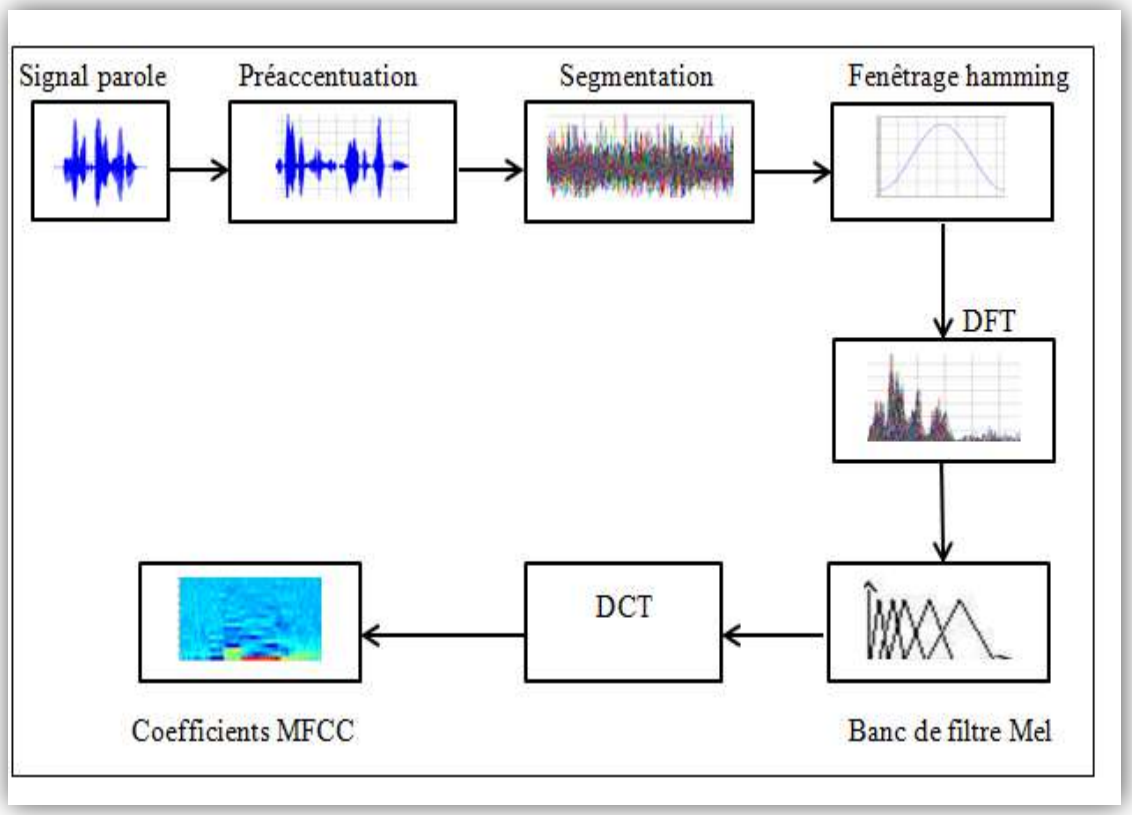


Figure II.4. *Diagramme de bloc MFCC*

3.1. Acquisition du signal

Une fois le signal de la parole est produit, il sera stocké dans l'ordinateur en passant par le microphone, ce dernier joue un rôle très important dans la qualité de l'enregistrement qui affectera bien entendu la justesse des résultats finaux. Mais avant cela, le signal de la voix doit être échantillonné c'est -à-dire transformer de la forme continue à une forme discrète, alors le signal sera décrit par une suite numérique de la forme :

$$\boldsymbol{X[n] = X(nT)} \qquad \text{(II.1)}$$

Où T est la période d'échantillonnage et on déduit la fréquence F = 1 / T.

Enfin, le signal de la parole est stocké dans la machine sous format (.wav).

Un exemple d'un signal parole est présenté sur la figure II.5 :

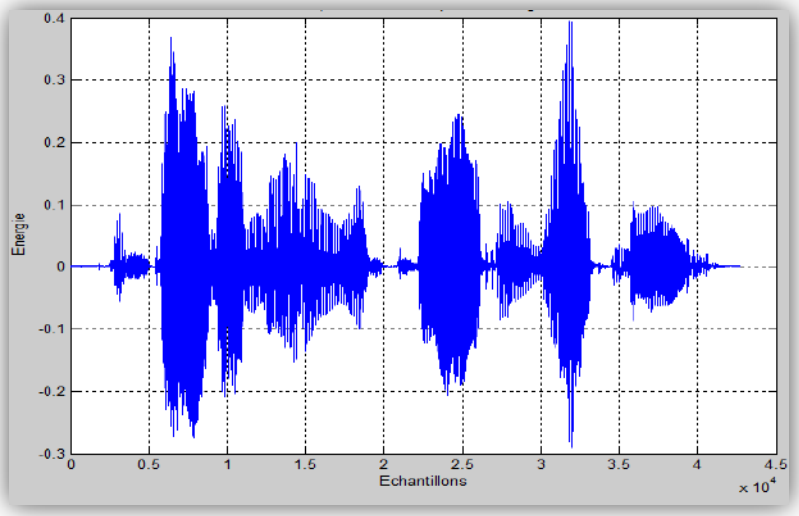


Figure II.5. *Visualisation d'un signal parole*

3.2. Préaccentuation :

Si on observe le signal de la voix produit par un humain, on trouve que les sons qui sortent par la cavité nasale ont une haute fréquence mais aussi une faible amplitude ce qui crée une difficulté pour distinguer clairement ces sons. Pour remédier à ce problème, on appliquera un filtre sur notre signal et donc sur toutes les trames et on aura une amplification des composantes de haute fréquence. La transformé en Z de ce filtre est donné par :

$$\mathbf{H(z) = 1 - \alpha\, Z^{-1}} \qquad \text{(II.2)}$$

Avec $0.9 \leq \alpha \leq 10$

Cette étape est utilisée pour l'amélioration du signal. Son but est de soulever les hautes fréquences des composantes spectrales dans le signal de parole [15].

3.3. Segmentation

Cette étape est nécessaire car le signal parole est non stationnaire, en effet, les études ont pu montrer que le signal peut être considéré comme quasi-stationnaire dans des intervalles de temps inférieur à 30 ms.

Cette étape sert à diviser le signal de parole en trames de longueur courtes afin d'obtenir une voix quasi stationnaires. Le signal de la voix ne peut pas être considéré comme un signal stable à long terme que ses propriétés varient considérablement dans le temps [15].

3.4.Fenêtrage

La segmentation du signal en trames produit des discontinuités aux frontières des trames. Si on passe directement au domaine fréquentiel moyennant la transformée de Fourier et on observe le spectre on trouve certainement des informations provenant du signal lui-même mais aussi d'autres provenant du découpage de signal en trame, ces discontinuités se manifestent par des lobes secondaires. Ces effets sont réduits en multipliant les échantillons $\{S_a(n)\}\ n = 0 \dots N-1$ de la trame par une fenêtre de pondération $\{w(n)\}\ n = 0 \dots N-1$ [1].

$$\boldsymbol{S(n) = w(n).S_a(n)} \qquad \textbf{(II.3)}$$

Une fenêtre utile doit être capable de préserver l'information spectrale et avoir moins de fuites spectrales.

Une fenêtre rectangulaire (avec fuite spectrale) est définie comme suit [10] :

$$\boldsymbol{W(n) = \begin{cases} 1, & n = 0, 1, \dots N-1 \\ 0, & ailleur \end{cases}} \qquad \textbf{(II.4)}$$

Une fenêtre de hamming (avec moins de fuite spectrale) est définie comme suit :

$$\boldsymbol{W(n) = \begin{cases} 0.54 - 0.46\, cos\left(\frac{2\pi n}{N-1}\right), & n = 0, 1, \dots N-1 \\ 0, & ailleur \end{cases}} \qquad \textbf{(II.5)}$$

La figure II.6 montre la fenêtre de hamming :

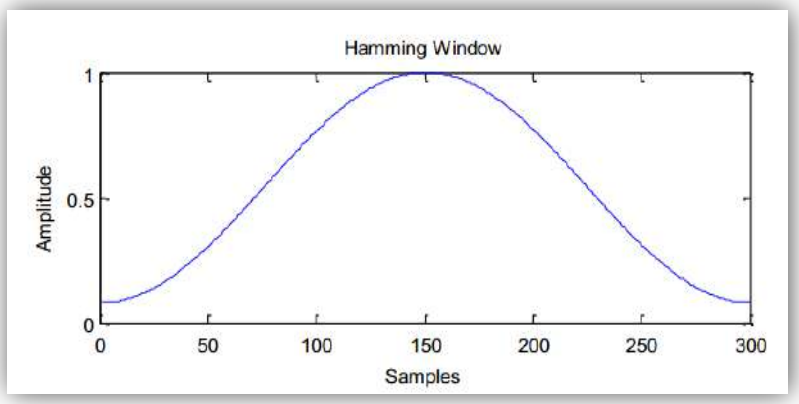


Figure II.6. *Fenêtre de hamming*

Chaque trame est de longueur N échantillons, tel que les trames adjacentes sont chevauché de M échantillons, on applique une fenêtre à chaque trame pour réduire la discontinuité à la fin de chacune d'elles. On a utilisé la fenêtre de Hamming décrite par la formule (II.5).

3.5.Analyse cepstral

3.5.1. Transformer de Fourier Discrète DFT

La transformée de Fourier discrète (DFT) est calculé par transformée de Fourier rapide (FFT). Cet algorithme est largement utilisé pour évaluer le spectre de fréquences de la parole et convertit chaque trame de N échantillons à partir du domaine temporel au domaine fréquentiel. La FFT est calculée sur l'ensemble des N échantillons.

Pour chacune des N bandes de fréquences discrètes, un nombre complexe X [k] représente l'amplitude et la phase de la composante de fréquence dans le signal original. La transformée de Fourier discrète (DFT) d'un signal x (n) est donnée par:

$$\boldsymbol{X[K] = \sum_{n=0}^{N-1} x[n]\, e^{-j2\frac{\pi}{N}Kn}} \qquad \textbf{(II.6)}$$

Où $k \in \mathbb{Z}$ est l'indice de spectre composant discret, $X[K] \in \mathbb{R}$ est la trame d'entrée, Ainsi, seuls les coefficients donnés par $K = 0,1, \dots, N-1$ sont nécessaires pour représenter le spectre d'énergie de la DFT, en outre si $X[K]$ à la longueur de $\mathrm{N} = 2^{\mathrm{q}}, \mathrm{q} \in \mathbb{N}$

Le spectre d'énergie peut être représenté comme suit :

$$\mathbf{|X[K]|^2 = \left|\sum_{n=0}^{N-1} x[n]\, e^{-j2\frac{\pi}{N}Kn}\right|^2} \qquad \textbf{(II.7)}$$

Par conséquent, si la longueur de X[K] ne correspond pas 2^q,remplissage à zéro sera appliqué au signal de sorte que la FFT peut être utilisé enfin.

Lorsque nous effectuons FFT sur une trame, nous supposons que le signal à l'intérieur d'une trame est périodique est continu. Si ce n'est pas le cas, nous pouvons toujours effectuer FFT mais les discontinuités au premier et dernier point de la trame est susceptible d'introduire des effets indésirables dans la réponse en fréquence. Pour faire face à ce problème, nous avons deux stratégies [17]:

- Multiplier chaque trame par une fenêtre de Hamming pour augmenter sa continuité aux premier et dernier points.
- Prendre une trame d'une dimension variable de telle sorte qu'elle comporte toujours un nombre entier multiple du nombre des périodes fondamentales du signal de parole.

La deuxième stratégie rencontre des difficultés dans la pratique depuis l'identification de la période fondamentale n'est pas un problème trivial. De plus, les sons non vocaux n'ont pas une période fondamentale à tous. Par conséquent, on adopte généralement la première stratégie précisément de multiplier la trame par une fenêtre de Hamming avant d'effectuer la FFT.

3.5.2. Echelle Mel

Échelle Mel est appliquée en raison de placer plus l'accent sur les composantes de basse fréquence. C'est parce que, l'information portée par les composantes de basse fréquence du signal de parole est plus important que les composantes de fréquence élevée. Elle se distingue par le fait qu'il est linéaire pour les fréquences inférieures à 1KHz alors qu'il est logarithmique pour les autres fréquences. La formule approximative qui fait correspondre chaque fréquence en Hz une fréquence Mel est la suivante :

$$f_{mel} = 2595 \times \log(1 + \frac{F(Hz)}{700}) \qquad \textbf{(II.8)}$$

Et la transformation inverse f_{mel}^{-1} est donnée par :

$$f_{mel}^{-1} = 700 * [e^{\frac{f_{mel}}{2595}} - 1] \qquad \textbf{(II.9)}$$

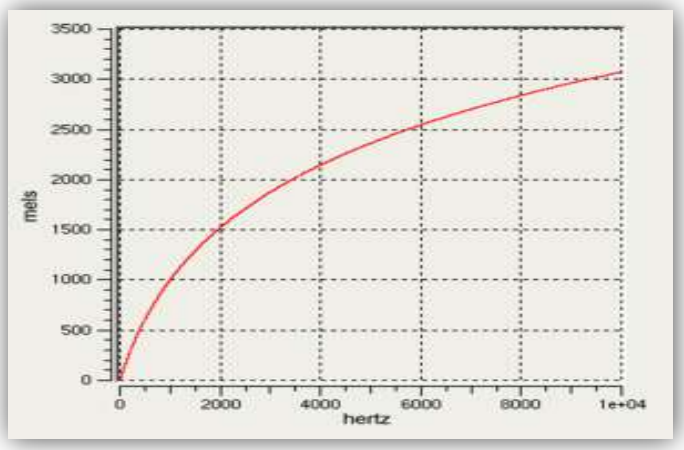


Figure II.7. *L'échelle Mel*

Maintenant que l'échelle Mel est connue, nous essayons de voir son utilité : des études psychophysiques ont montré que la perception humaine de la parole ne suit pas une échelle linéaire dans le domaine fréquentiel mais une échelle logarithmique, c'est pour cette raison que l'échelle Mel a été introduite afin de reproduire la sélectivité de l'oreille [16].

Donc, on applique un banc de K filtres triangulaires sur chaque trame, ces filtres sont positionnés uniformément sur l'échelle Mel contrairement à l'échelle Hz. Cette répartition est illustrée dans la figure II.11:

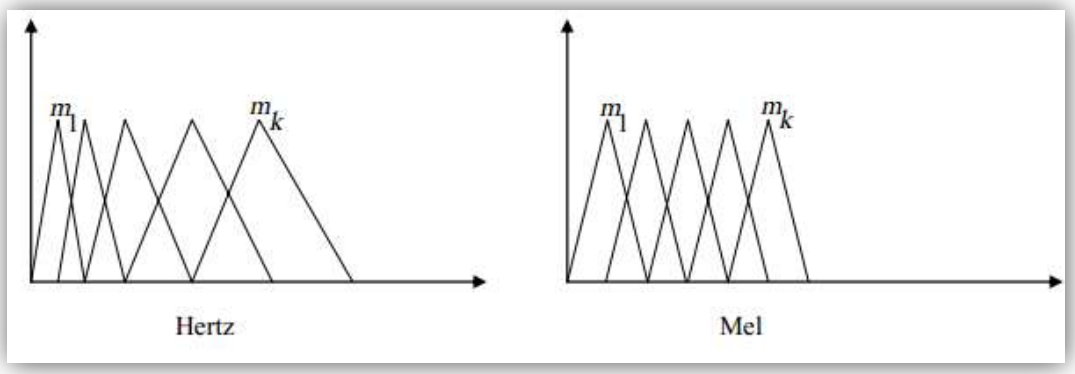


Figure II.8. *La répartition des filtres triangulaires sur les échelles Fréquentielle et Mel.*

3.5.3. Transformée en cosinus discrète

Une fois que nous avons les énergies de banc de filtres, nous prenons le logarithme d'eux. Ceci est également motivé par l'oreille humaine, nous n'entendons pas l'intensité sonore sur une échelle linéaire.

Une transformée en cosinus discrète (DCT) exprime une séquence finie de points de données en termes de somme de cosinus fonctions oscillant à des fréquences différentes. C'est le processus de convertir le logarithme de spectre Mel dans le domaine temporel en utilisant DCT. Le résultat de la conversion est appelée Mel Frequency Cepstrum Coefficient. L'ensemble est appelé coefficient de vecteurs acoustiques. Par conséquent, chaque énoncé d'entrée est transformé en une séquence de vecteur acoustique [12].

4. Méthodes de classification

Le but des modèles de classification est de classer les objets d'intérêt c'est-à-dire les séquences de vecteurs acoustiques extraits à partir du signal de parole, en une ou plusieurs catégories ou classes. Chaque classe est ici rattachée à un locuteur.

4.1.Méthode indépendante du texte (Mélange de Gaussiennes GMM)

Les systèmes de reconnaissance du locuteur ont pour but de reconnaître le locuteur, c'est-à-dire la personne qui parle. Cette reconnaissance peut être dépendante du texte (il faut prononcer un texte précis pour être reconnu) ou indépendante du texte (peu importe ce qui est dit).

Depuis quelques années la façon la plus courante de représenter un locuteur en reconnaissance indépendante du texte est d'utiliser des mélanges de gaussiennes (GMM pour Gaussian Mixture Model).

C'est approche qui a été proposé par Conrad SANDERSON, elle consiste à transformer le signal de départ en plusieurs vecteurs de coefficients DCT, puis modéliser leur distribution selon une combinaison linéaire de plusieurs gaussiennes qui vont représenter un modèle d'une personne.

Cette technique est venue pour améliorer les performances des HMM, elle a prouvé une efficacité surprenante surtout en matière de précision et de temps d'exécution [16].

Ce modèle GMM est créé à partir de données audio de nombreux locuteurs représentatifs d'une population générique.

4.2. Méthodes dépendantes du texte (Modèle de Markov Cachés HMM)

Un système dépendant du texte est caractérisé par le fait que la structure partielle ou globale de la parole est connue. Il existe quatre sortes de systèmes dépendant du texte: les systèmes pour lesquels le client doit prononcer un mot de passe fixé, les systèmes ou il peut choisir son mot de passe, ceux où il doit prononcer des mots fixés mais dont il choisit l'ordre et ceux qui utilisent un prompteur. Cependant dans tous ces cas, l'information a priori connue par le système légitimiste une classification différente de celles vues précédemment.

Les modèles de Markov cachés (HMM) sont utilisés depuis plusieurs années pour la détection et la reconnaissance du locuteur, reposant sur certains coefficients de la transformée en cosinus discrète (DCT) comme source d'observations [16].

Un autre phénomène variable dans la parole provient des variations de prononciation de chaque locuteur, et des différences entre locuteurs. Il est aisément compréhensible que deux personnes, même si elles prononcent le même énoncé, n'ont pas exactement le même résultat acoustique. Cela s'accentue encore entre hommes et femmes.

5. Méthodes de vérification

5.1.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 [13].

5.2.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 : $P = (p1, p2 \dots pn)$ et $Q = (q1, q2 \dots qn)$, est donnée par

$$\boldsymbol{d = \sqrt{\sum_{i=1}^{n}(pi - qi)2} = \sqrt{(p1 - q1)2 + (p2 - q2)2 + \cdots + (pn - qn)2}} \qquad \textbf{(II.10)}$$

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 (MFCC). Si la distance est supérieure au seuil, le locuteur est dit escroc sinon le système va reconnaitre le locuteur selon sa base de données [14].

Notons que la distance euclidienne est la méthode utilisée dans notre conception, L'algorithme de la vérification est comme suit :

- **a)** premièrement, le signal original est transformé en un vecteur acoustique suivant les étapes d'analyse acoustique (MFCC dans notre cas) soit E1 le vecteur normalisé des caractéristiques.
- **b)** pour un signal inconnu, calculer le MFCC 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 [14].

6. Conclusion

Dans notre système de vérification du locuteur, on a adopté pour l'extraction des paramètres acoustiques l'analyse cepstrale et le filtrage. Ces deux techniques combinées ensemble ont prouvé une efficacité accrue dans ce domaine car et comme on l'a vu précédemment les coefficients cepstraux fournissent que l'information nécessaire à la reconnaissance du locuteur ainsi que le filtrage Mel formatent ces coefficients en inspirant de la perception de l'oreille humaine, ensuite on a passé à la comparaison qui est une phase nécessaire pour la vérification de l'identité du locuteur, basant sur la méthode de distance euclidienne, par la suite on va présenter les procédures des calculs et la conception de notre système.

Chapitre III :

Conception

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 de vérification du locuteur.

2. Environnement de travail

Dans cette section, nous présenterons les environnements matériels et logiciel de notre travail.

2.1.Environnement matériel

Afin de mener à bien ce projet, il a été mis à notre disposition un ensemble de matériels dont les caractéristiques sont les suivantes :

Un ordinateur HP avec les caractéristiques suivantes :

Processeur : Intel® Core™ i5 CPU M 460 @ 2.53GHz 2.53 GHz

RAM : 4.00 Go (3.80 Go utilisable)

Disque dure : 1 To

OS : Microsoft Windows 7

2.2. Outil de développement (Matlab R2012b)

On a eu recours lors de l'élaboration de notre système à MATLAB R2012b.

Matlab et son environnement interactif est un langage de haut niveau qui permet l'exécution de tache nécessitant une grande puissance de calcul et dont la mise en œuvre sera bien plus simple et rapide qu'avec des langages de programmation traditionnels tels que le C, C++. L'utilisation du logiciel Matlab nous permettra d'être au plus près avec le signal et d'acquérir rapidement, avec une autonomie au niveau de la manipulation et du traitement des signaux de parole.

3. Description de système de vérification du locuteur

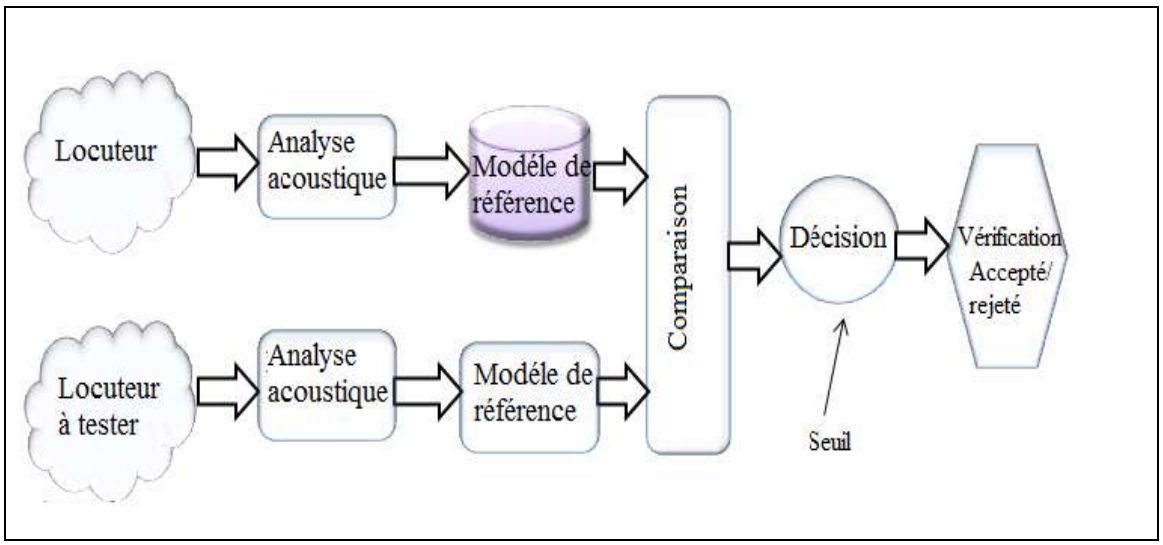


Figure III.1. *Architecture général de notre application*

Notre système se décompose en deux phases. (Figure III.1). L'analyse acoustique et la comparaison

3.1. *Signal de parole (Locuteur)* :

Lecture du fichier audio "client " enregistré sous Matlab avec la commande wavrecord, ses données sont stockées dans un fichier « .wav » qui correspond à un locuteur. Le signal parole (voix) a été échantillonné avec une fréquence d'échantillonnage de 16 kHz

On a effectué cette étape avec la commande suivante :

```
[voix, Fe] = wavread  ('client.wav');
```

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

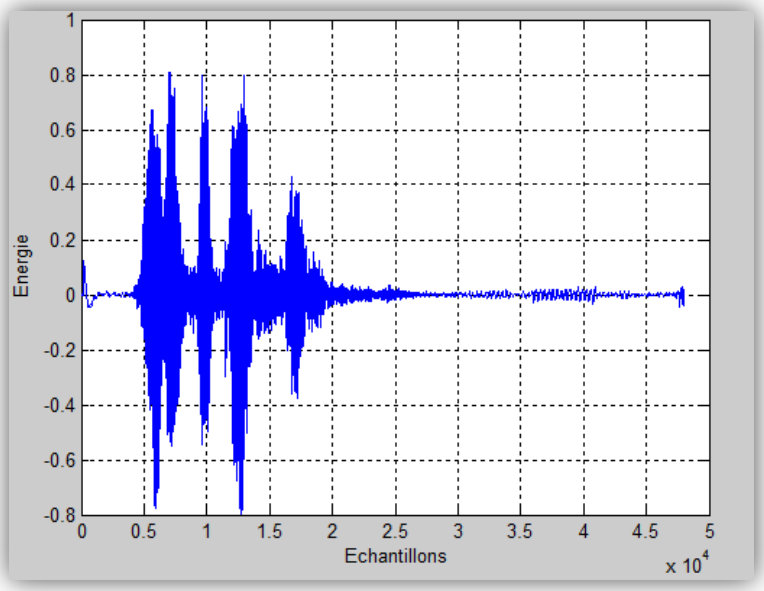


Figure III.2. *Représentation d'échantillons du signal parole de mot « vérification »*

3.2. Phase 1 (Analyse acoustique)

On procède cette phase afin de réduire la redondance du signal de parole, en extraire qu'un ensemble de paramètres parmi toutes les données disponibles, choisissant les paramètres de MFCC expliqué dans le chapitre II, dans le but de diminuer la quantité de calcul et de stockage lors du traitement et de vérification ,elle est construite par 6 étapes. La figure (III.2) représente l'organigramme de toutes les étapes d'analyse acoustique.

Les paramètres utilisés sont montrée sur le tableau (Tableau III.1)

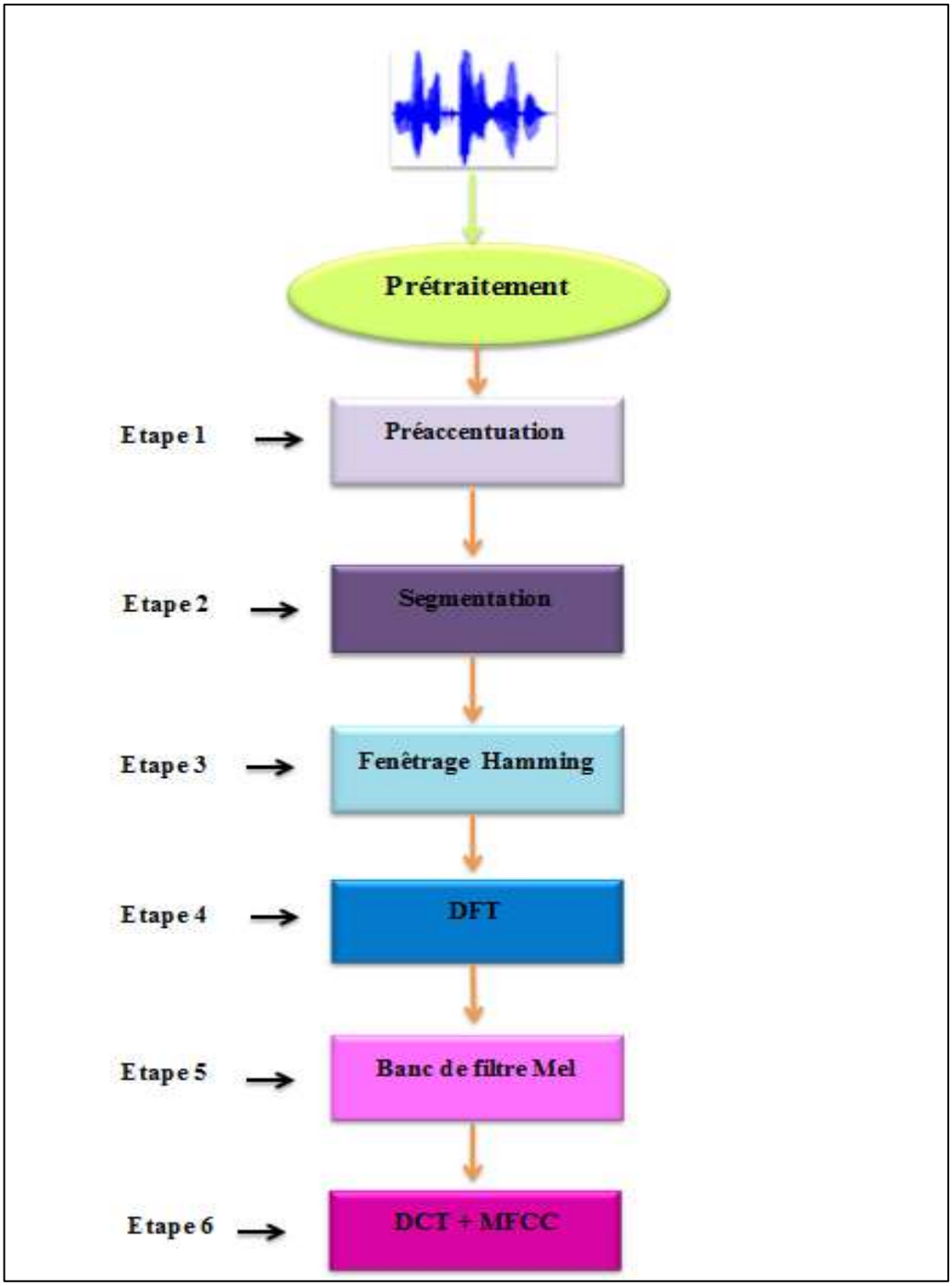


Figure III.3. *Organigramme de L'analyse acoustique*

Tableau III.1. Paramètres utilisé en MFCC

Paramètres	Valeurs
duree_enreg	**3s**
long= fe * duree_enreg	**48000 échantillons**
temps_station	**15ms**
taille_trame = (long * temp_station / duree_enreg)	**240 échantillons**
taille_chevauch= taille_trame / 2	**120 échantillons**
n_trame= (long-1) / taille_chevauch	**399 trames**
n_filtre	**40**

- **Etape1 :**

Avant la préaccentuation du signal, on effectue cette étape pour préparer le signal et le rend convenable pour les futurs traitements, le résultat de cette étape (élimination du silence, amplification) est le signal « voix1 ».

Ensuite, le signal échantillonné « voix1 » est préaccentué afin de relever les hautes fréquences qui sont moins énergétiques que les basses fréquences. Cette étape consiste à appliquer un filtre numérique à réponse impulsionnelle finie, montré dans le code ci-dessus

```
alpha=0.95;
preacc = filter( [1 ,-alpha], 1, voix1 );
```

Le tracé du signal parole préaccentué sur la figure III.4 :

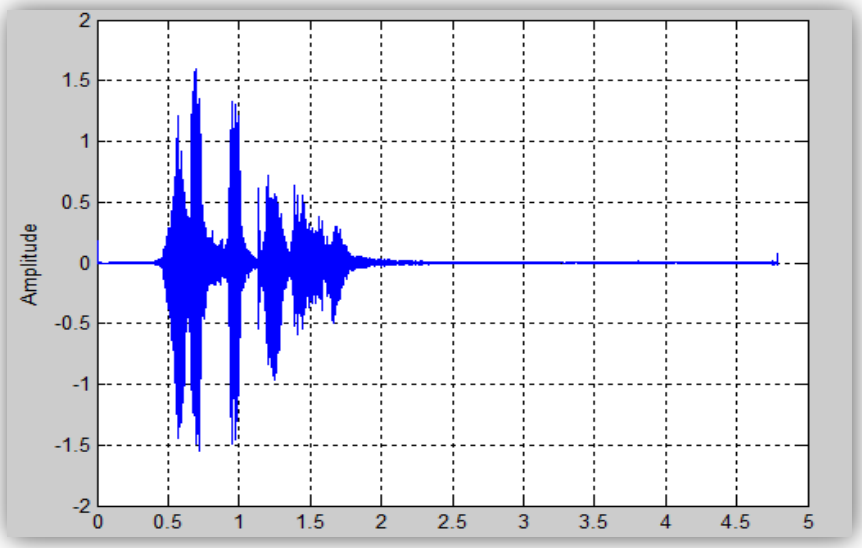


Figure III.4. *Signal parole « vérification » après préaccentuation*

- **Etape 2 :** Le signal de parole préaccentué est ensuite **segmenté**.

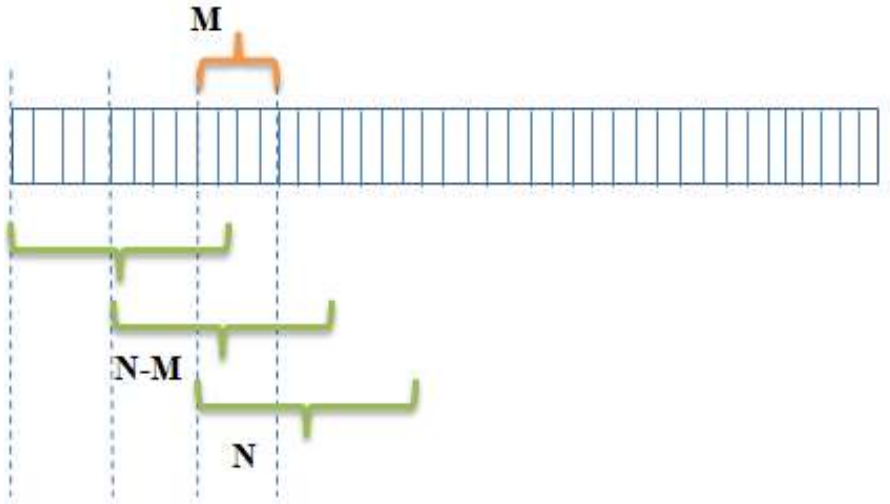


Figure III.5. *Schéma de Découpage en trames*

Son code est donné comme suit :

```
a1(1:taille_trame,1:n_fenetre)=0;
a1(:,1)=preacc(1:taille_trame);
for j=2:n_fenetre
a1(:,j)=preacc((taille_trame-taille_chevauch)*(j-1)+1:(taille_trame-taille_chevauch)*…
(j-1)+taille_trame);
end;
```

Le tracé des segments du signal parole préaccentué est comme montre la figure suivante:

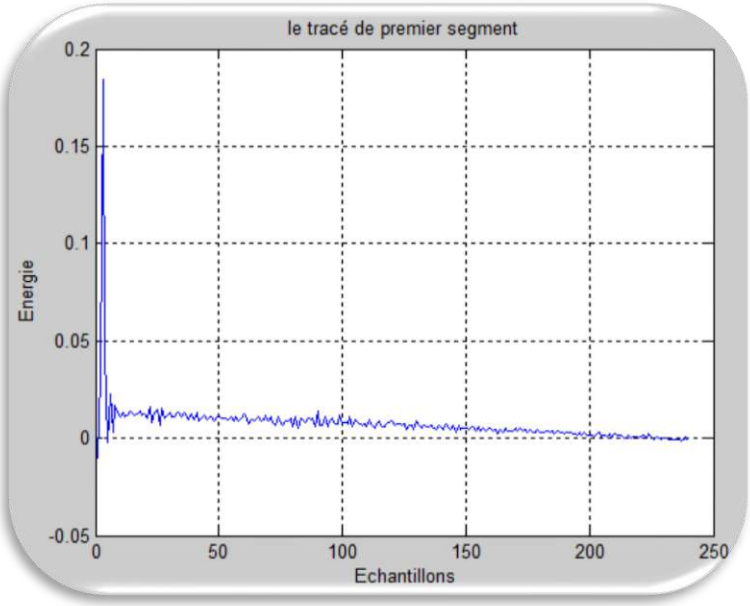


Figure III.6. *Tracé de premier segment du signal préaccentué*

On découpe le signale en trames de taille_trame= 240 échantillons. Ces trames adjacentes sont séparées par taille_chevauch=120.

La première trame est constituée de 240 échantillons. La seconde trame commencera à partir de 120ème échantillons, et qu'il se chevauche par taille _trame – taille_chevauch (240-120=120 échantillons). Le processus se poursuivra jusqu'à ce que tous les trames du signal « preacc » est pris en compte.

- **Etape3 :**

On applique une fenêtre à chaque trame pour réduire la discontinuité à la fin de chacune d'elles. On a utilisé la fenêtre de Hamming (figure II.6), Son code est définit comme suit :

```
h1= hamming (taille_trame);
b1(:,j)= a1(:,j).* h1;
```

Chaque trame du signal segmenté « a1 » sera multipliée par une fenêtre de hamming « h1 » afin de garder la continuité du première et dernière de la trame.

Le résultat de fenêtrage du signal « a1» est le signal de sortie « b1» montré sur la figure III.7.

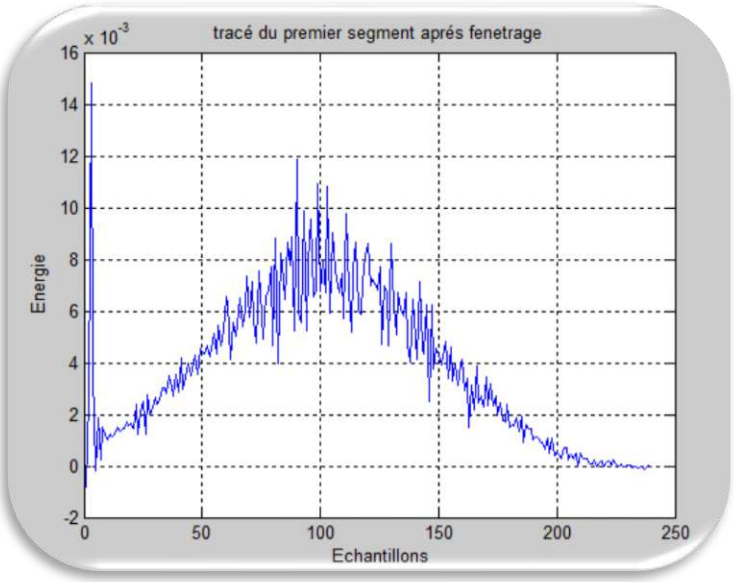


Figure III.7. *premier segment après fenêtrage par Hamming*

Comme on le voit sur la figure, la partie centrale du signal est ameliorée tandit que les ehchantillons du debut et de fin sont attenués suite à l'utilisation de fenêtrage de Hamming

- **Etape 4 :**

C'est la première étape de traitement, elle est basée sur la transformée de Fourier rapide principalement. La FFT convertit chaque trame de 240 échantillons du domaine temporel vers le domaine fréquentiel. La FFT est un algorithme rapide pour mettre en œuvre la transformée de Fourier discrète (DFT), cela se fait par code Matlab suivant:

```
c1(:,j)=fft(b1(:,j));
```

Ensuite, les nombres complexes dans la matrice FFT sont représentés comme réel par le calcule de son modules (spectre de puissance) avec la commande *abs*.

La figure III.8 montre la représentation fréquentielle du signal

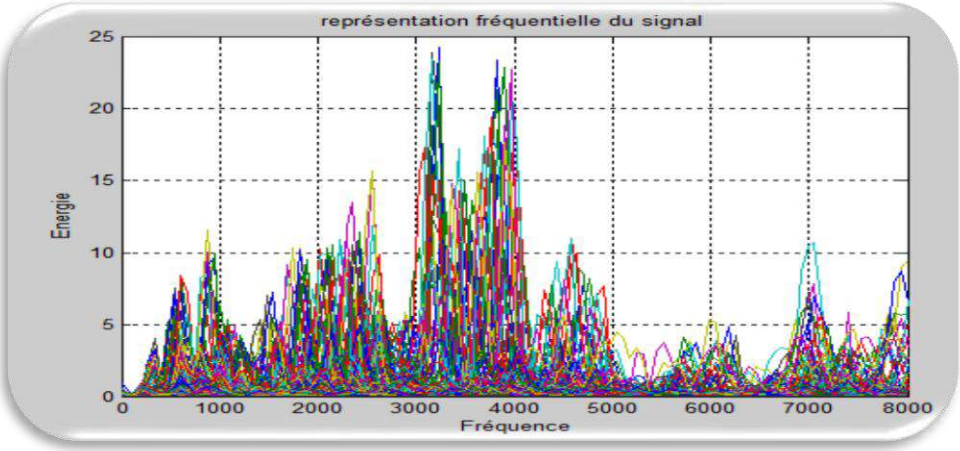


Figure III.8. *Représentation fréquentielle du signal*

- **Etape 5 :**

Cette étape consiste à créer un banc de filtre Mel, on applique un banc de 40 filtres triangulaires sur chaque trame.

Par ce code Matlab nous avons créé le banc de filtre Mel :

```
for c = 1:n_filtre

    increment = 1/(ind_centre(c)-ind_debut(c));
    for i = ind_debut(c)+1:ind_centre(c)
        mel(c,i) = (i -ind_debut(c))*increment;
    end
     decrement = 1/(ind_stop(c) ind_centre(c));
    for i = ind_centre(c):ind_stop(c)
  mel(c,i) = 1- ((i -ind_centre(c))*decrement);
    end
end
```

Voici une représentation du banc de filtres réalisé sur la figure III.9 :

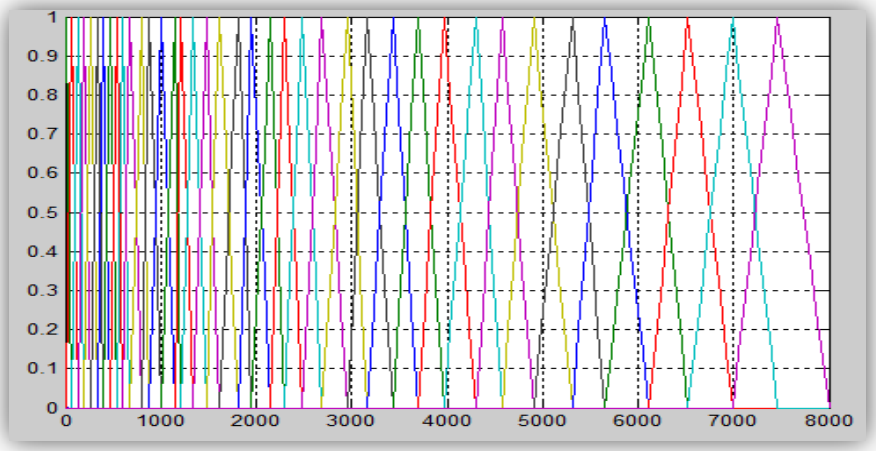


Figure III.9. *Banc de filtre Mel*

- **Etape 6 :**

Dans cette dernière étape, nous utilisons la transformée en cosinus discrète (DCT) pour convertir le log de spectre de l'échelle Mel vers le domaine temporel. Le résultat de la conversion est MFCC. Son code est comme suit :

```
for j=1:n_trame
n=floor(taille_trame/2);

mfcc1(:,j)=dct(log(mel*abs(spec_puiss(1:n,j).^2)))
;
 end ;
```

Suit à ce code on aura les coefficients MFCC comme le montre la figure III.10 :

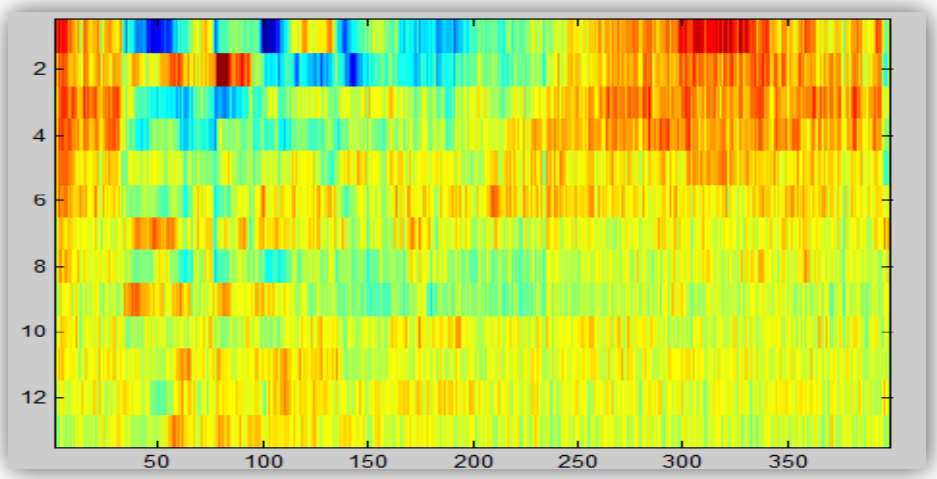


Figure III.10. *Coefficients MFCC*

La figure III.10 montre les coefficients MFCC du signal parole de référence du mot prononcé « vérification ».

Ces coefficients doivent être normalisés pour effectuer la prochaine étape qui est la comparaison entre les coefficients MFCC du signal référence et les coefficients MFCC du signal à tester.

- ✓ **Normalisation**

A la fin du calcul des MFCC, nous avons effectué une « normalisation » des vecteurs d'observations. Cette étape consiste à centrer et à univarier les vecteurs. La nécessité de ce traitement peut être discutée. Imaginons qu'un coefficient soit biaisé et diverge anormalement. Il entrainera alors tous les autres dans son biais. Ceci peut poser un problème pour le suivi car alors tous les coefficients seraient mauvais. Le système ne pourrait plus se caler sur les coefficients non biaisés.

On a testé notre application avec des coefficients mfcc non normalisés et nous avons remarqués la différence si on n'effectue pas la normalisation sur les coefficients, donc la normalisation améliore la fiabilité de notre système de vérification. La rapidité de la comparaison est bien plus grande quand les coefficients utilisés sont normalisés.

La normalisation est basée sur le calcul des centroid.

Pour cela, nous avons créé une fonction de postnormalisation comme suit :

```
function cb=postnorm(n_fenetre,var)
```

Et pour trouver les valeurs de centroid de chaque colonne :

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

Et pour trouver les valeurs de centroid de chaque ligne :

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

avec :

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

Tel que :

```
var(:,j)=var(:,j)-m;
m=mean(var(:,j));
```

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

L'appel est comme suit :

```
[Signal]=postnorm(n_fenetre,mfcc_BD);
```

3.3. Phase 2 : Comparaison (distance euclidienne)

On compare les paramètres du modèle à tester « TEST » d'un individu quelconque aux paramètres du modèle de référence « BD » 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(BD,TEST,n,k,c)
for j=1:n       % finding euclidian distance
  x(j)=sqrt(sum((BD(1,1)-TEST((n*(k-1))+j,1)).^2))
  y(j)=sqrt(sum((BD(1,2:c)-TEST((n*(k-1))+j,2:c)).^2))
end
```

C : nombre de colonne, r : nombre de ligne, k=1,…,n : nombre de locuteur=1,

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

- **Décision :**

A partir de seuil choisit, cette phase fournit une décision qui constituera la réponse finale du système de vérification du locuteur.

On a confronté deux types de tests :

- Les tests clients lors desquels la parole présentée au système correspond à l'identité Proclamée.
- Les tests imposteurs lors desquels la parole présentée au système provient d'un Individu inconnu du système.

- **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 cela, nous avons effectué 20 tests dans une salle totalement vide pour éviter le maximum de bruit : 10 tests pour le client et 10 imposteurs prononçant le mot « Vérification »

Pour chaque test, on calcule les probabilités de fausse acceptation (FA) et de faux rejet (FR), on effectue les seuils de vérification jusqu'au l'obtention de minimum de (FA) et (FR), et le taux pour l'évaluation de notre système.

- **Résultat de test :**

A partir les valeurs de comparaison on définit une valeur de seuil, si nous avons des valeurs de x et y inférieurs au seuil (accepté) sinon (rejeté) 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. Le seuil à fixer se situe entre le score imposteur le plus élevé et le score client le plus faible.

On va présenter ci-dessous quelques résultats de test parmi les 20 tests effectués. Le client prononce le même code 10 fois, et chaque imposteur des 10 imposteurs prononce le même code.

❖ **Test 1** ⟶ **Seuil1 1 : x=0.04, y=5.5**

Tableau III.2. Calcule des FR du test1

Client	Rejeté	Accepté
Code 1	x=0.0125 y=5.9063	
Code 2		X=0.0314 Y=5.3973
Code 3		X=0.0178 Y=5.0307
Code 4	X=0.0420 Y=5.5767	
Code 5	X=0.0379 Y=6.0881	
Code 6		X=0.0070 Y=5.0901
Code 7	X=0.0425 Y=5.9217	
Code 8		X=0.0141 Y=5.0043
Code 9	X=0.0454 y=6.4031	
Code 10		X=0.0403 Y=5.0872

P(FR) =5/10 =50%

Tableau III.3. Calcule des FA du test1

	Rejeté	Accepté
Imposteur1	x=0.0463 y=5.0356	
Imposteur2	x=0.0469 y=4.9104	
Imposteur3	x=0.0503 y=6.1027	
Imposteur4	x=0.0514 y=5.9145	
Imposteur5	x=0.0345 y=5.7475	
Imposteur6	x=0.0404 y=6.0717	
Imposteur7	x=0.0642 y=6.0940	
Imposteur8	x=0.0697 y=6.8782	
Imposteur9	x=0.0668 y=6.5785	
Imposteur10	x=0.0969 y=5.8320	

P(FA)=0%

❖ **Test 2** ⟶ **Seuil1 2 : x=0.04, y=6.088**

Tableau III.4. *Calcul des FR du* test2

Client	Rejeté	Accepté
Code 1		x=0.0125 y=5.9063
Code 2		X=0.0314 Y=5.3973
Code 3		X=0.0178 Y=5.0307
Code 4	X=0.0420 Y=5.5767	
Code 5		X=0.0379 Y=6.0881
Code 6		X=0.0070 Y=5.0901
Code 7	X=0.0425 Y=5.9217	
Code 8		X=0.0141 Y=5.0043
Code 9	X=0.0454 y=6.4031	
Code 10	X=0.0403 Y=5.0872	

P(FR) =4/10 =40%

Tableau III.5. *Calcul des FA du test2*

	Rejeté	Accepté
Imposteur1	x=0.0463 y=5.0356	
Imposteur2	x=0.0469 y=4.9104	
Imposteur3	x=0.0503 y=6.1027	
Imposteur4	x=0.0514 y=5.9145	
Imposteur5		x=0.0345 y=5.7475
Imposteur6	x=0.0404 y=6.0717	
Imposteur7	x=0.0642 y=6.0940	
Imposteur8	x=0.0697 y=6.8782	
Imposteur9	x=0.0668 y=6.5785	
Imposteur10	x=0.0969 y=5.8320	

P(FA)=1/10= 10%

❖ **Test 3 : Seuil1 2 → x=0.0430, y=6.4050**

Tableau III.6. Calcule des FR du test 3

Client	Rejeté	Accepté
Code 1		x=0.0125 y=5.9063
Code 2		X=0.0314 Y=5.3973
Code 3		X=0.0178 Y=5.0307
Code 4		X=0.0420 Y=5.5767
Code 5		X=0.0379 Y=6.0881
Code 6		X=0.0070 Y=5.0901
Code 7		X=0.0425 Y=5.9217
Code 8		X=0.0141 Y=5.0043
Code 9	X=0.0454 y=6.4031	
Code 10		X=0.0403 Y=5.0872

P(FR) =1/10 =10%

Tableau III.7. Calcule des FA du test 3

	Rejeté	Accepté
Imposteur1	x=0.0463 y=5.0356	
Imposteur2	x=0.0469 y=4.9104	
Imposteur3	x=0.0503 y=6.1027	
Imposteur4	x=0.0514 y=5.9145	
Imposteur5	x=0.0345 y=5.7475	
Imposteur6		x=0.0404 y=6.0717
Imposteur7	x=0.0642 y=6.0940	
Imposteur8	x=0.0697 y=6.8782	
Imposteur9	x=0.0668 y=6.5785	
Imposteur10	x=0.0969 y=5.8320	

P(FA)=1/10= 10%

Dans ce cas aussi, La meilleure vérification dans ce test est celle où on a trouvé P(FA)=P(FR)=10% donc, notre système était capable de reconnaître 9 sur 10 clients et refusé 9 sur 10 imposteurs.

4. Présentation de l'application

On présente dans cette section les différents aspects du système de vérification du locuteur.

4.1. *Présentation de l'interface :*

Dans cette interface on illustre le fonctionnement du notre système de vérification du locuteur.

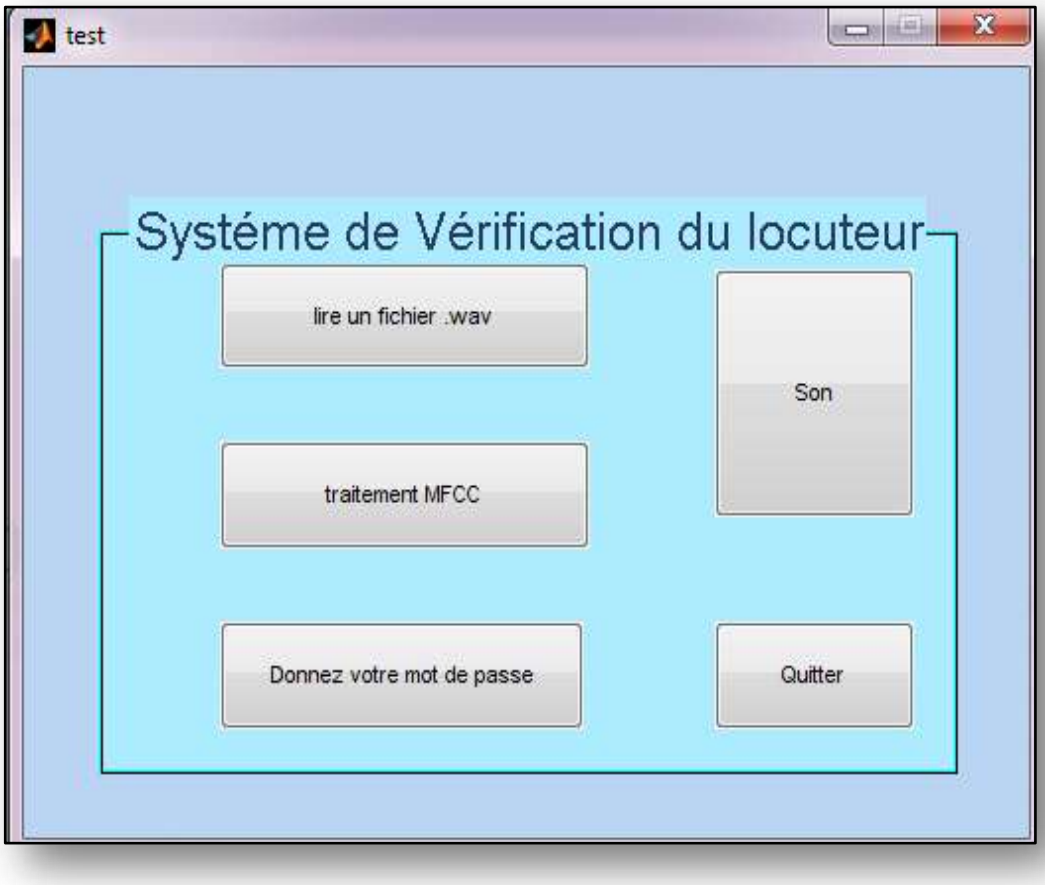


Figure III.11. *Interface graphique*

- Le Panel « Système de vérification du locuteur » regroupe cinq boutons :

1. *Lire un fichier audio :*

On clique sur le bouton ‘’ lire un fichier.wav’’, qui ouvre une fenétre permettant de parcourir le disque et choisir la base de donnée (client.wav).

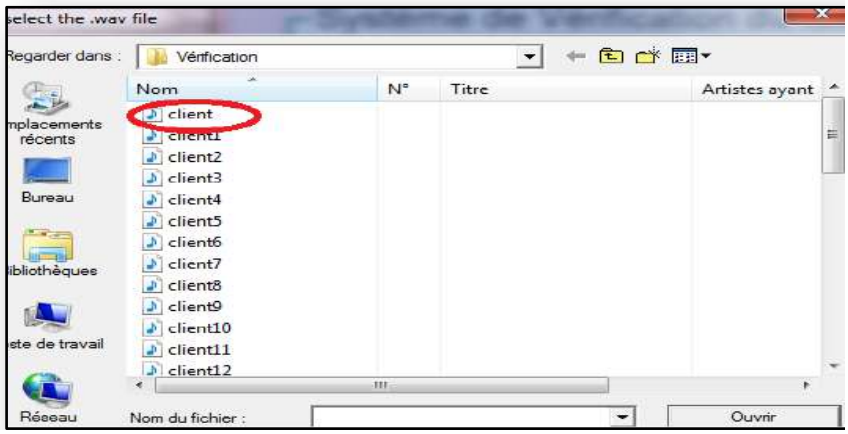


Figure III.12. *Entrer le fichier audio ‘’.wav’’*

2. *Son :*

Pour écouter la voix du locuteur sélectionné (client.wav) on clique sur le bouton ‘’Son’’.

3. *Traitement mfcc :*

Lorsqu’on clique sur le bouton ‘’traitement mfcc’’, une figure apparaitre (figure III.10) montre les coefficients MFCC du signal de la base de donnés.

4. *Donner votre mot de passe :*

On clique sur le bouton ‘’Donner votre mot de passe ‘’, une boite de message Menu apparait, voir (figure III.13) a deux boutons :



Figure III.13. *Menu : Donner votre mot de passe*

4.1. lire un fichier :

nous cliquons sur le bouton ‘’ lire un fichier ’’, une fenetre s’ouvre permettant de parcourir le disque et choisir le fichier de test déjà enregistré. Exemple : « client1 ».

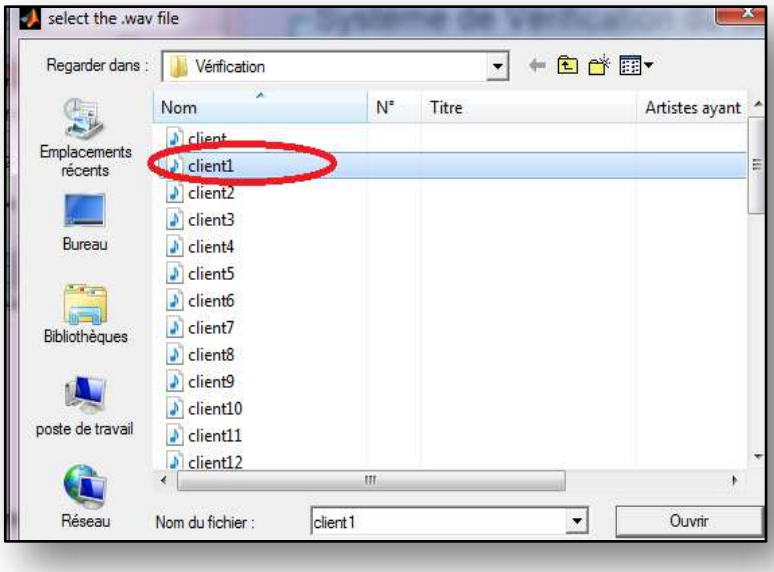


Figure III.14. *Entrer le fichier ‘’.wav’’ à tester*

Après la sélection du fichier audio, la boite de message de vérification apparait affichant le message « autorisé » (figure III.5).



Figure III.15. *Vous êtes autorisé*

4.2. Enregistrer un mot de passe :

Nous cliquons sur le bouton ‘’ Enregistrer un mot de passe ’’



Figure III.16. *Enregistrement de mot de passe oralement*

En enregistrant une voix, s’il reconnait la voix le message « autorisé » apparait sinon « refusé ».



Figure III.17. *Accès refusé*

5. Quitter :

C’est un bouton pour quitter l’interface.

Conclusion

Enfin, dans ce chapitre, on a illustré une architecture de notre application de vérification du locuteur ainsi que tous les modules qui le composent avec les différentes phases (acquisitions des signaux, traitement MFCC, comparaison et les différents tests).

A travers ces tests, on peut confirmer l’efficacité de notre système de vérification du locuteur. En effet, On a réussi de fixer le seuil car on a enregistrée une meilleure performance pour la distance euclidienne entre les centroids avec P(FA)=P(FR)=10%.

Conclusion

Générale

Conclusion Générale

Le but de notre projet était de créer un système de vérification du locuteur, et l'appliquer à un discours d'un locuteur inconnu. En étudiant les caractéristiques extraites de la parole inconnue et les comparer aux caractéristiques extraites stockées pour chaque haut-parleur différent afin d'identifier le locuteur inconnu.

Ce travail s'inscrit dans le domaine de reconnaissance automatique du locuteur, par le biais de leurs voix. 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 la biométrie, ses applications et ses technologies, on a présenté une comparaison entre ses différentes technologie, en premier lieu, on a introduit la biométrie vocale qui est notre point de recherche, on a parlé sur le système RAL, ses applications et ses tâches nécessaires qui sont l'indentification et la vérification, en basant sur la tache de VAL qui est notre objectif dans cette thèse. Tout ça était présenté dans le premier chapitre.

Ensuite, on a présenté les phases nécessaires d'un système de VAL, qui sont l'analyse acoustique et la vérification, on a illustré les différentes méthodes PLP, LPC, MFCC... avec une étude comparative entre eux, où on a choisi MFCCs car ils offrent des meilleurs résultats, pour cela on a l'adopté pour l'extraction de paramètres des signaux de parole dans notre projet. Ainsi que dans la deuxième phase (vérification), on a définit des méthodes de classifications, ensuite les méthodes de vérification, où on a choisit pour la comparaison la méthode de distance euclidienne, car elle offre une meilleure performance.

On a décrit l'architecture de notre application de VAL puis on a introduit les étapes essentielles de calcul MFCC en présentant les codes sources Matlab et les résultats de chaque étape. Ensuite on a fait la normalisation des coefficients MFCCs, 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, ou le but est toujours de proposer une application plus performante avec moins de probabilité de fausse acceptation et surtout moins de faux rejet.

Références

Bibliographiques

Références Bibliographique

[1] Hacine Gharbi Abdenour « Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole », THÈSE Université Ferhat Abbas-Sétif Algérie, Faculté de Technologie Département d'Électronique, 2012

[2] « Techniques de contrôle d'accès par biométrie », Dossier Technique, Comission Technique de Sécurité Physique, 2003

[3] DANG Hoang Vu « Biométrie pour l'Identification », Rapport Final du TIPE, Institut de la Francophonie pour l'Informatique Vietnam, 2005

[4] Bernadette Dorizzi « Techniques et Usages Biométriques », Cours INT, Bernadette.Dorizzi@int-evry.fr, 2004

[5] Max Chassé « La biométrie au Québec » , Document d'analyse, Commission d'accès à l'information Québec, 2002

[6] Fiche d'information sur la biométrie, http://90plan.ovh.net/ ~addexa/ Net-Documentations/ Contr% F4le%20d'acc%E8s/Identification%20 Biom%E9trique/Info%20-biometrie.pdf

[7] HARRAG Abdelghani « Extraction des données d'une base: Application à l'extraction des traits du locuteur » THÈSE Doctorat Sciences, Université Ferhat ABBAS – Sétif faculté de technologie Département Electronique, 2011

[8] Vincent JOUSSE « identification nommée du locuteur : exploitation conjointe du signal sonore et de sa transcription », THÈSE Doctorat, ACADÉMIE DE NANTES UNIVERSITÉ DU MAINE, École Doctorale Sciences et Technologies de l'Information et Mathématiques Laboratoire d'Informatique, 2011

[9] Namrata Dave « Feature Extraction Methods LPC,PLP and MFCC In Speech Recognition », INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY, 1G H Patel College of Engineering, Gujarat Technology University, INDIA, 2013

Références Bibliographique

[10] Rishiraj Mukherjee, « Speaker Recognition Using Shifted MFCC », Graduate School Theses and Dissertations, University of South Florida Scholar Commons, 2012

[11] 12-2 MFCC : http://mirlab.org/jang/books/audiosignalprocessing/speech-Feature Mfcc.asp?title=12-2%20MFCC

[12] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[13] Kashyap Patel, R.K. Prasad, « Speech Recognition and Verification Using MFCC & VQ » International Journal of Emerging Science and Engineering, 2013

[14] Poornima.S, « Performance Study of Fusion in Multimodal Biometric Verification using Ear and Iris Features », International Conference on Research Trends in Computer Technologies, 2013

[15] Julien RACHEDI, « Reconnaissance et classification de phonèmes», Mémoire pour le Master Sciences et Technologie de l'UPMC Spécialité SAR Parcours ATIAM, Laboratoire d'accueil : IRCAM, Paris, 2005

[16] ROUIDI Houssam, « Conception et réalisation d'une Plateforme de fusion biométrique en score à base des machines à vecteurs de support (SVM) », THÈSE Pour l'obtention du diplôme d'ingénieur d'état en informatique Ecole nationale supérieure d'informatique (ESI- EX INI), Alger, 2008

[17] David Sierra Rodriguez, « Text-Independent Speaker Identification », THÈSE Master Science, AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY KRAKOW Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, 2008