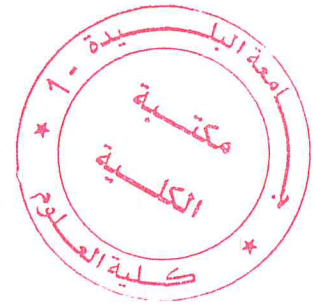


Université Saad Dahlab Blida1



Faculté des Sciences

Département d'informatique

Mémoire Présenté par :

DAHANE Hakima AMROUCHE Mahdia

En vue d'obtenir le diplôme de Master

Domaine : MI

Filière : informatique

Spécialité : Ingénierie des Logiciel

Option : Ingénierie des Logiciel

**Sujet : Plateforme Software d'un système de vérification de Locuteur
robuste au changement du canal de Transmission.**

Soutenu le : 23-06-2014

Encadreurs :

Devant le jury :

Mm Rezouge Président

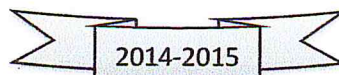
✓ Mr Bengherabi Messaoud

✓ Mm Fedila Meriem

Mm Gessoume Rapporteur

Promotrice : Mm Chikhi Imane.

Mm Zahra Examineur



Remerciement

On remercie *ALLAH* le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de Mr *BENGHERABI*, on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

Nos remerciements s'adressent à Mm *FEDILA* pour son aide pratique et son soutien moral et ses encouragements, pour ces précieuses aides pendant les moments difficiles.

On remercie aussi notre promotrice Mm *CHIKHI* pour leurs aides et leurs encouragements.

Nous tenons aussi à remercier mesdames et messieurs les membres du jury d'avoir accepté de juger notre travail et pour leur précieux temps accordé à l'étude de notre mémoire.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenus de près ou de loin.

Nous saisissons l'occasion pour remercier tout le corps professoral et administratif de l'USDB1

Dédicace

Je dédie ce modeste travail à :

A mes parents .Aucun hommage ne pourrait être à la hauteur de l'amour Dont ils ne cessent de me combler. Que dieu leur procure bonne santé et longue vie.

A celui que j'aime beaucoup et qui m'a soutenue tout au long de ce projet : mes soeurs *AMINA ,ASMA,HADIL,FARAH* et sans oublié *ALAA*.

A toute ma famille, et mes amis, A mon binôme *HAKIMA* et toute la famille *AMROUCHE*. Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

A. MAHDIA

Dédicace

Je dédie ce modeste travail à :

A mes parents .Aucun hommage ne pourrait être à la hauteur de l'amour Dont ils ne cessent de me combler. Que dieu leur procure bonne santé et longue vie.

A celui que j'aime beaucoup et qui m'a soutenue tout au long de ce projet : mes frère *FARES*, *YOUCEF* et sans oublié *SID ALI*.

A toute ma famille, et mes amis, A mon binôme *MAHDIA* et ma chère amie *SELMA* et toute la famille DAHANE. Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

D. HAKIMA

Liste des abréviations

RAL : Reconnaissance Automatique du Locuteur

TID : Calcule le Taux d'Identification

IAL : Identification Automatique du Locuteur

VAL : Vérification Automatique de Locuteur

ZCR : Zéro Crossing Rate

GMM : Modèles à Mélanges de Gaussiennes.

UBM : Universel Background Model (ou modèle de monde).

I-Vecteur : Identity Vecteur

DCF : Fonction de Coût de Décision

MFCC: Mel-Scale Frequency Cepstral Coefficients

FFT: Fast Fourier Transforme

ML: Maximum Likelihood

VAD: Voice Activity Detection.

EER : Equal Error Rate

FA: Fausse Acceptation

FR: Faux Rejet



ملخص:

في الوقت الحاضر، العديد من التطبيقات تتطلب مصادقة المستخدم يمكن الاعتماد عليها. لا يمكن أن يؤديها هذا التوثيق باستخدام بطاقة ذكية أو PIN (السحب من ATM) أو يمكن أن يتم ذلك باستخدام نظم التحقق من الهوية متطورة من خلال بصمة لتأمين الوصول إلى الأماكن الحساسة ومدخل مراقبة الوقت، والمسائل القانونية واستخدام الطب الشرعي DNA. ومع ذلك وبمجرد الوصول إلى التطبيق عن بعد (عبر شبكة الهاتف، على سبيل المثال)، يبقى الصوت السمة الحيوية الوحيدة المتاحة لمصادقة شخص. هذا العمل مصمم في إطار معرفة المتحدث على أساس أحدث التقنيات. I-vector

تقنية التعرف على المتحدث شهدت تقدما كبيرا في السنوات الأخيرة من خلال إدخال مفهوم I-vector لتمثيل المتحدث. يتميز هذا التمثيل الجديد بخاصية إمكانية تقليل كلام الخطاب إلى شعاع صغير الحجم مقارنة مع التمثيل السابق مع مزيد من التمييز.

في هذا المشروع، أدركنا منصة برمجية للتعرف على المتحدث تحتوي على منهجين الأكثر شعبية في النمذجة وهي

GMM-UBM (Gaussian Mixture Model / Universal Background Model) و I-vector.

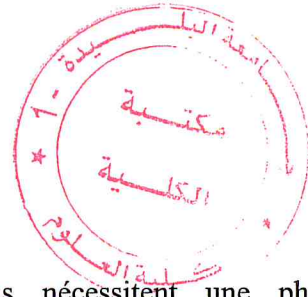
أجرينا دراسة على تأثير حالات عدم التطابق الناتجة من قنوات البث على أداء نظام التحقق من المستخدم بإحضار 39

متحدث جزائري. و تأكد النتائج تفوق و متانة عالية لتقنية I-vector بالمقارنة مع GMM-UBM

الكلمات المفتاحية:

القياسات الحيوية, GMM-UBM, I-vector, التعرف التلقائي على المتحدث (RAL), التحقق التلقائي من المتحدث

(VAL), التحديد التلقائي للمتحدث (IAL), ظروف عدم تطابق.



Résumé

De nos jours, un grand nombre d'applications nécessitent une phase d'authentification de l'utilisateur. Cette authentification peut être réalisée au moyen d'une carte à puce et d'un code confidentiel (retrait à un guichet automatique), au travers d'empreintes digitales (accès sécurisé à des locaux) ou d'empreintes génétiques (domaine juridique) ou encore grâce à la voix (serrure vocale). Dès lors qu'une application est accessible à distance (par le réseau téléphonique par exemple), la voix reste le seul élément disponible pour authentifier une personne. Cette thèse s'inscrit dans le cadre de la Reconnaissance Automatique du Locuteur RAL.

La reconnaissance du locuteur a connu ces dernières années une avancée significative avec un nouveau concept de représentation de l'énoncé de voix, désigné sous le terme de i-vector (Vecteur d'Identité). Ce type de représentation présente la particularité de se réduire numériquement à un vecteur de dimension faible, au regard des représentations précédentes, et pourtant très discriminant vis à vis du locuteur. Les travaux présentés dans ce mémoire s'inscrivent dans ce nouveau contexte.

Nous avons réalisé une plateforme logicielle de reconnaissance du locuteur englobant deux approches à savoir GMM-UBM (Gaussian Mixture Model/ Universal Background Model) et I-vector. Nous avons effectué une étude sur l'effet de variations des canaux de transmission sur les performances du système de vérification du locuteur en utilisant un *corpus* d'un *dialecte algérien* de 39 locuteurs. Les résultats trouvés ont montré que l'approche I-vecteur est plus robuste par rapport au GMM-UBM.

Mots-clés :

Biométrie, GMM-UBM, I-Vecteur, Reconnaissance Automatique du Locuteur, Identification automatique de locuteur (IAL), vérification automatique de locuteur (VAL), multisession.

Summary

Nowadays, many applications require a reliable user authentication. This authentication can be performed using a smart card or PIN code (withdrawal from an ATM) or can be done using sophisticated biometric systems through: a fingerprint to secure access to sensitive places and entrance time monitoring, legal matters and forensics using DNA. However, Once an application is accessed remotely (via the telephone network, for example), the voice remains the only available biometric trait to authenticate a person. The aim of this work is designing a framework of speaker recognition based on state of the art techniques.

Speaker recognition has witnessed a significant progress in recent year by the introduction of the concept of I-vector for speaker representation. This new representation is distinguished by its attractive property of reducing the speech utterance to a small size vector relative to previous representations with more distinctiveness.

In this project, we realized a software platform for speaker recognition containing two popular modeling approaches: namely GMM-UBM (Gaussian Mixture Model / Universal Background Model) and I-vector.

We conducted a study on the effect of the mismatch resulting from the variability of transmission channels on the performance of speaker verification system using a corpus of 39 Algerian speakers. The results confirm the superiority and the high robustness of the I-vector approach compared to the GMM-UBM.

Keywords:

Biometrics, GMM-UBM, I-Vector, Automatic Speaker Recognition (ASR), Automatic Speaker Verification (VAL), Automatic Speaker Identification (ALI), Mismatch conditions.

Sommaire

Introduction Générale.....	1
CHAPITRE I : Généralités sur la Biométrie	
I.Introduction.....	4
II.Définition de la Biométrie.....	4
III. Présentation des différentes techniques biométriques.....	5
IV. Structure d'un système biométrique	9
V. Evaluation des performances d'un système biométrique.....	10
V.1. En mode identification.....	10
V.2. En mode vérification	10
VI. Conclusion.....	12
CHAPITRE II : RECONNAISSANCE AUTOMATIQUE DE LOCUTEUR(RAL)	
I.Introduction.....	14
II. la Reconnaissance Automatique de locuteur (RAL).....	14
II.1. La Vérification Automatique de Locuteur (VAL).....	15
II.2. L'identification Automatique de Locuteur (IAL).....	15
III. Les étapes d'un Système de Reconnaissance Automatique du locuteur.....	16
III.1 Acquisition du signal de la parole	16
III.1.1. Source de variabilité du signal de la parole	17
III.1.2. Echantillonnage.....	17
III.2. Prétraitements.....	18
III.2.1. Découpage en trames.....	18
III.2.2. Préaccentuation	19
III.2.3. Elimination de silence	20
III.2.4. Fenêtrage	22
III.3. Extraction des paramètres acoustiques.....	22
III.3.1. Analyse Cepstrale	22



III.3.2. L'extraction des Coefficients MFCC	23
III.4. Modélisation	26
III.4.1. Quantification vectorielle	26
III.4.2 Modélisation par Mélange de Gaussiennes (GMM)	27
III.4.3. Modèle de Monde de Mélange de Gaussiennes UBM.....	27
III.4.4. Machine à Vecteurs de Support (SVM)	27
III.4.5. L'analyse Jointe de Facteur JFA (Joint Factor Analysis)	28
III.4.6. La Modélisation par Identity Vecteur (I-Vecteur).....	28
IV.Conclusion.....	28

CHAPITRE III : Modélisation de Locuteur

I. Introduction.....	30
II. Approche GMM (Gaussians Mixture Model)	30
III. Approche GMM-UBM (Modèle de Monde)	32
III.1. Estimation du modèle UBM.....	32
III.2 Calcule des scores.....	34
IV. Approche JFA L'analyse jointe de facteur (Joint Factor Analysis).....	35
V. Approche (I-Vecteur).....	36
V .1. Estimation de la matrice de variabilité totale T	37
V.2 L'extraction des I-Vecteur	38
V.3. Compensation de l'effet de canal.....	38
V.4 G-PLDA (Gaussian Probabilistic LDA).....	39
V.5. Calcul des scores.....	39
VI. Conclusion.....	39

Chapitre IV : Analyse et Conception

I. Introduction.....	41
II. Présentation de la démarche utilisée	41
II.1 Le Processus Unifie UP	41

III. Le langage UML.....	42
IV. Expression des besoins	42
IV.1 Identification des acteurs.....	42
IV. 2 Identification des cas d'utilisation	43
IV.3 Diagrammes des cas d'utilisation	43
IV.4 Diagrammes des classes	49
IV.4.1 Diagramme de classe relatif aux modèle GMM-UBM	50
IV.4.4 Diagramme de classe relatif aux modèle I-Vecteur.....	52
IV.5 Diagrammes de séquence	54
V. Conclusion.....	59
Chapitre V : Test et Implémentation	
I. Introduction	61
II. Description de la base de donnée utilisée	61
III. L'architecture du Logiciel.....	61
IV. Prétraitement et extraction des paramètres.....	61
IV.1 Prétraitement.....	61
IV.2 Extraction des paramètres MFCC	62
V. Protocole expérimental	62
V.1 Système de vérification de locuteur	62
V.2 Vérification du locuteur à base de GMMUBM	63
V.3 Vérification du locuteur à base de I-Vecteur	65
V.4 Comparaison entre les modélisations à base GMM-UBM et I-Vecteur.....	66
V.5 Nouvelle technique de concaténation de caractéristiques MFCC.....	67
VI. L'interface de l'application	68
VI.1 Environnement du travail	68
VI.2 Environnement matériel	68
VI.3 Présentation de l'interface de l'application	69

VI.3.1	La fenêtre principale de l'application	69
VI.3.2	Modélisation par GMM-UBM.....	70
V.3.3	Modélisation à base de I-Vecteur.....	73
VII.	Application en temps réel.....	77
VI.1	En mode de Vérification.....	77
VI.2	En mode d'identification	78
VIII.	Conclusion.....	79
	Conclusion Générale.....	80



Liste des Figure

- P. 5 Figure I.1** : Acquisition du visage.
- P. 5 Figure I.2** : Empreinte digitale.
- P. 6 Figure I.3** : Rétine.
- P. 7 Figure I.4** : Forme de la main.
- P.7 Figure I.5** : Signature.
- P.8 Figure I.6** : Frappes au clavier.
- P. 8 Figure I.7** : ADN.
- P. 9 Figure I.8** : Signal vocal.
- P. 10 Figure I.9** : Schéma général d'un système biométrique.
- P. 11 Figure I.10** Illustration du FRR et du FAR.
- P.11 Figure I.11** Relation entre TFR, TFA et EER.
- P.14 Figure II.1** : Structure d'un system de RAL.
- P. 15 Figure II.2** : Structure de système de vérification de locuteur.
- P. 16 Figure II.3** : La structure de système d'identification du locuteur.
- P. 16 Figure II.4** : Etapes d'un système de RAL.
- P. 17 Figure II.5** : L'appareil vocal.
- P. 18 Figure II.6** : Echantillonnage d'un signal.
- P. 19 Figure II.7** : Découpage en trames.
- P. 19 Figure II.8** : Le filtre de la préaccentuation.
- P. 21 Figure II.9** : Schéma de classification vocal / non vocal.
- P. 21 Figure II.10** : Du signal $s_1(n)$ avec silence au signal $x_1(n)$ sans silence en utilisant la fonction $VAD(n)$.
- P.22 Figure II.11** : Les étapes du fenêtrage.
- P.22 Figure II.12** : Le découpage en trames.
- P.23 Figure II.13** : les étapes de Calcul des coefficients MFCC.
- P.25 Figure II.14** : L'allure de la sélectivité de l'oreille humaine.
- P.25 Figure II.15** : Répartition des filtres sur les échelles Mels.
- P.26 Figure II.16** : Définition des classes de données en quantification vectorielle.
- P. 27 Figure II.17** : Approximation de la distribution d'un paramètre acoustique par une combinaison.
- P. 30 Figure.III.1** : Illustration de nuages acoustiques représentant l'identité d'un locuteur.
- P. 31 Figure.III.2** : Mélanges des gaussiennes en trois dimensions.

- P. 34 Figure III.3 :** Adaptation MAP d'un modèle GMM-UBM.
- P.34 Figure III.4 :** Architecture du système à base de la méthode GMM-UBM.
- P. 35 Figure III.5** Décomposition du supervecteur du locuteur.
- P. 37 Figure III.6 :** Architecture du système à base de la méthode I-vecteur.
- P. 41 Figure IV.1:** Cycle de vie d'UP.
- P. 43 Figure IV.2 :** diagramme de cas d'utilisation global.
- P. 44 Figure IV.3:** Diagramme de cas d'utilisation d'apprentissage à base de GMM-UBM.
- P.45 Figure IV.4:** Diagramme de cas d'utilisation de test à base de GMM-UBM.
- P.46 Figure IV.5 :** Diagramme de cas d'utilisation d'apprentissage à bas I-Vecteur.
- P.47 Figure IV.6 :** Diagramme de cas d'utilisation de création de modèle global à base I-Vecteur.
- P.47 Figure IV.7 :** Diagramme de cas d'utilisation de création des modèles train à base I-Vecteur.
- P. 48 Figure VI.8 :** Diagramme de cas d'utilisation de test à base I-Vecteur.
- P. 50 Figure VI.9:** Diagramme de classe relatif aux modèle GMM-UBM.
- P. 52 Figure VI.10 :** Diagramme de classe relatif aux modèle I-Vecteur.
- P.54 Figure IV.11:** Diagramme de séquence de Modèle de Monde (UBM) à base de GMM-UBM.
- P.55 Figure VI.12 :** Diagramme de séquence d'apprentissage des Modèles des Locuteur à base de GMM-UBM.
- P. 55 Figure VI.13 :** Diagramme de séquence de test d'identification à base de GMM-UBM.
- P.56 Figure IV.14 :** Diagramme de séquence de création de modèle globale (G-PLDA) à base I-Vecteur.
- P.57 Figure IV.15 :** Diagramme de séquence de création des Modèle de Client à bas I-Vecteur.
- P.58 Figure IV.16 :** Diagramme de séquence de Test en mode de vérification à base I-Vecteur.
- P.61 Figure V.1 :** Base de données d'apprentissage et base de données de test.
- P.62 Figure V.2 :** Les EER de système de vérification à base de GMM-UBM.
- P.64 Figure V.3 :** Les résultats de test (EER) selon l'espace de variabilité totale(T) de système de vérification à base d'I-Vecteur.
- P.65 Figure V.4 :** test d'influence de l'ordre LDA.
- P.66 Figure V.5** Les Courbes DET de GMM-UBM VS I-Vecteur.
- P.68 Figure V.6 :** la fenêtre principale de l'application.

- P.69 Figure V.7** : la fenêtre principale de modélisation par GMM-UBM.
- P. 70 Figure V.8** : Formulaire des Paramètres GMM-UBM.
- P.71 Figure V.9** : Charger Basse de donnée UBM.
- P.71 Figure V.10** : Charger Basse de donnée Train.
- P.72 Figure V.11** : Charger BD Test.
- P.72 Figure V.12** : Affichez les Scores.
- P.72 Figure V.13** : interface principale de modélisation par I-Vecteur.
- P.73 Figure V.14** : Formulaire des Paramètres I-Vecteur.
- P.74 Figure V.15** : chargement des BD UBM et Variabilité totale.
- P.74 Figure V.16** : Charger BD Train.
- P.75 Figure V.17** : calcule des scores.
- P.75 Figure V.18** : affichage les signaux et les résultats.
- P.76 Figure V.19** Acquisition d'une voix en temps réel.
- P.76 Figure V.20**: Thèse de Test.
- P.77 Figure V.21**: Thèse de Test en mode de Vérification.
- P.77 Figure V.22**: Thèse de Test en mode d'Identification.

Liste des Tableaux

P.42 Tableau IV.1 : Tableau descriptif des acteurs.

P.44 Table IV.2 : Table descriptive pour la Reconnaissance de locuteur à bas GMM-UBM et à bas I-Vecteur.

P.45 Table IV.3 : Table descriptive pour l'apprentissage à bas de GMM-UBM.

P.46 Table IV.4 : Table descriptive pour le Test à bas de GMM-UBM.

P.48 Table IV.5: Table descriptive pour la création des modèles des locuteurs train à base I-Vecteur.

P.49 Table IV.6 : Table descriptive pour le test à bas I-Vecteur.

P.51 Table IV.7 : Dictionnaire de données relatif aux GMM-UBM

P.51 Table IV.8 : Liste de méthodes relatives aux GMM-UBM

P.53 Table IV.9 : Dictionnaire de données de diagramme de classe relatif aux I-Vecteur.

P.54 Table IV.10 : Liste des méthodes relatives aux I-Vecteur.

P.63 Tableau V.1 : L'EER du système de vérification du locuteur dans les quatre cas de tests : Microphone vs à (Fixe, microphone et Mobile), Microphone vs à (Microphone, Fixe et Mob), Mobile vs à (Microphone, Fixe et Mobil) et 3Sessions vs à (Microphone, Fixe et Mobil).

P.65 Tableau V.2 : table comparative des valeurs de test du système de vérification données par les deux méthodes de modélisation GMM-UBM et I-Vecteur.

P.67 Tableau V.3 : table comparative des valeurs de test entre les deux méthodes (la moyenne de I-Vecteurs, Concaténation) selon les trois sessions du système de vérification à base de I-Vecteur.

Introduction Générale

La reconnaissance automatique du locuteur (RAL) est interprétée comme une tâche particulière de la biométrie. Ce domaine regroupe les problèmes relatifs à l'identification et la vérification du locuteur basées sur l'information contenue dans le signal parole.

Les systèmes RAL actuels reposent majoritairement sur des approches génératives; comme les modèles de mélange de gaussiennes (GMM), associée à une représentation acoustique de type Mel Frequency Cepstral Coefficient MFCC. Ce modèle probabiliste permet de modéliser finement la distribution multimodale des paramètres acoustique. L'approche standard GMM-UBM (Gaussian Mixture Model/ Universal Background Model) [1] utilise un modèle du monde UBM qui caractérise l'univers des locuteurs et rendre la décision optimale dans un *cadre bayésien*.

Les systèmes RAL basés sur le paradigme GMM-UBM disposent d'excellentes performances lorsqu'il s'agit de discriminer entre des voix de locuteurs acquises dans des conditions contrôlées. Cependant, le bruit, l'état du locuteur et le "mismatch" entre les données d'enregistrements provenant de téléphones cellulaires utilisées lors de la phase d'apprentissage et celles provenant de téléphones fixes utilisées lors de la phase de test peuvent dégrader de façon draconienne les performances.

Sachant que la plupart des enregistrements vocaux traités par les experts de la reconnaissance automatique du locuteur sont issus des communications téléphoniques, et comme les moyens et les appareils de communication sont très variées (fixe, internet, mobile...etc.) et ils sont en pleine évolution, le développement d'un système robuste et efficace de reconnaissance du locuteur est plus que nécessaire.

Une avancée significative a été enregistrée récemment avec l'introduction d'un nouveau concept de modélisation désigné sous le terme de Vecteur d'Identité ou Identity Vector (I-Vector) [2]. Ce type de représentation s'appuie sur le paradigme GMM-UBM et présente la particularité de se réduire à un vecteur de dimension faible, au regard des représentations précédentes, et pourtant très discriminant vis à vis du locuteur et robuste vis à vis les conditions d'enregistrement.

Notre principale contribution dans ce mémoire est la réalisation d'une plateforme logicielle de reconnaissance du locuteur englobant les techniques GMM/UBM (Gaussian

Mixture Model/ Universal Background Model) et I-Vector. La robustesse des deux approches aux variations des canaux de transmission a été évaluée en utilisant un *corpus* d'un *dialecte algérien* de 39 locuteurs. Cette base de données a été construite par le Centre du Développement des Technologies Avancées CDTA en coopération avec la société espagnole AGNITIO¹ qui est un leader mondial dans la reconnaissance de la voix et ses applications. La base de données nous a permis d'effectuer plusieurs tests d'évaluation sur les différents systèmes RAL. Les résultats de cette étude ont montré que l'approche I-vecteur est plus robuste par rapport au GMM-UBM.

Nous avons choisi d'articuler notre étude autour de cinq chapitres principaux :

Le **premier chapitre** présente des généralités sur la Biométrie. Il décrit le principe de fonctionnement des systèmes biométriques puis définit les outils utilisés pour évaluer leurs performances.

Le **deuxième chapitre** constitue une description schématique des systèmes de reconnaissance automatique du locuteur. De la paramétrisation acoustique du signal à la production d'une décision, les différentes étapes de traitement y sont présentées.

Le **troisième chapitre** présente les différentes méthodes de modélisation utilisées à savoir le GMM-UBM, JFA et I-Vecteur.

Le **quatrième chapitre** présente la Conception de notre système logiciel.

Le **cinquième chapitre** est consacré à la présentation de l'application développée ainsi que l'évaluation expérimentale de la vérification.

Le mémoire se termine par une conclusion qui présente un bilan du travail réalisé dans ce mémoire et expose les perspectives et les travaux futurs pour améliorer et compléter le travail présenté.

¹ [http : www.agnitio.es](http://www.agnitio.es)

CHAPITRE I : Généralités

sur la Biométrie

I. Introduction

La croissance internationale des communications, tant en volume qu'en diversité (déplacement physique, transaction financière, accès aux services...), implique le besoin de s'assurer de l'identité des individus. L'importance des enjeux, motive les fraudeurs à mettre en échec les systèmes de sécurité existants, Comme les systèmes biométriques qui sont reliés à l'extraction et l'analyse des caractéristiques des personnes par la dérivation des informations et /ou l'interprétation des axiomes humain [3].

Dans ce chapitre, nous allons définir la Biométrie puis nous allons décrire les principales techniques biométriques et les systèmes biométriques (structure et performance).

II. Définition de la Biométrie

Le mot biométrie désigne dans un sens très large l'étude quantitative des êtres vivants. La biométrie ou mesure (métrie) du vivant (bios) est, d'après l'encyclopédie Larousse², « l'étude statistique des dimensions et de la croissance des êtres vivants ». L'extension de la biométrie au domaine de la reconnaissance des personnes consiste à déterminer l'identité d'un individu grâce à des mesures quantitatives. Ces mesures peuvent avoir pour objet les caractéristiques morphologiques ou les caractéristiques comportementales de cette personne [4].

Pour que des caractéristiques collectées puissent être qualifiées de modalités biométriques, elles doivent être [5]:

- *universelles* : Exister chez tous les individus.
- *uniques* : Permettre de différencier un individu par rapport à un autre.
- *permanentes* : Autoriser l'évolution dans le temps.
- *enregistrables* : Collecter les caractéristiques d'un individu avec son accord.
- *mesurables* : Autoriser une comparaison future.

L'avantage de la biométrie est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être changées, perdues ou volées. Les méthodes biométrique peuvent aussi être utilisées en complément ou remplacement de mots de passe.

² <http://www.larousse.fr/encyclopedie/>

III. Présentation des différentes techniques biométriques

III.1. Visage

L'avantage de la reconnaissance faciale par rapport aux autres technologies biométriques ci est qu'elle est non intrusive (acceptée par le public), facile à utiliser [6], peut être utilisée à distance et elle ne nécessite qu'une caméra numérique ou webcam (capteur).

Les caractéristiques essentielles de la reconnaissance du visage varient d'un système à un autre : les yeux, la bouche, le tour du visage, le nez, etc. La difficulté de cette reconnaissance est l'instabilité de l'environnement : luminosité, arrière-plan, la distance de la caméra, etc.



Figure I.1 : Acquisition du visage.

III.2. Empreinte digitale

La reconnaissance d'empreintes digitales est la technique biométrique la plus ancienne. Cependant, elle est considérée comme une mesure biométrique assez mal acceptée par les utilisateurs à cause de l'association qui est souvent faite avec la criminologie.

L'empreinte digitale est la caractéristique d'un doigt. Chaque personne a ses propres empreintes digitales. L'empreinte digitale basée sur trois aspects, d'abord l'acquisition d'une empreinte digitale, l'extraction de ses caractéristiques puis leur comparaison. [7]



Figure I.2 : Empreinte digitale.

III.3. Rétine

La rétine est la paroi interne de l'œil sur laquelle se projettent les images que nous voyons. Cette paroi est tapissée par un réseau de vaisseaux sanguins, qui forment un motif unique pour chaque individu. Cette technique donne de très bonnes performances car le schéma sanguin est stable durant toute la vie mais les systèmes d'acquisition de la rétine sont couteux ; elle nécessite un système intrusif et peu agréable par le public à cause du processus de l'acquisition qui exige la stabilité devant le capteur pendant une durée de 10s ou plus : L'image est obtenue en projetant sur l'œil un rayon lumineux de faible intensité dans les fréquences visibles ou infrarouges. L'œil doit être situé très près de la tête de lecture et l'utilisateur doit fixer son regard sur un point déterminé pendant plusieurs secondes ce qui demande une grande coopération de sa part. Les personnes hésitent en général à approcher un organe aussi sensible que l'œil près de l'appareil de mesure ce qui explique pourquoi cette méthode est mal acceptée par le grand public [8].

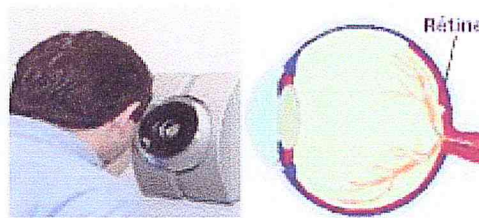


Figure I.3 : Rétine.

III.4. Forme de la main

Depuis 1993, les services américains de l'immigration se sont mis à l'heure de la biométrie à l'aéroport John-Fitzgerald-Kennedy de New York et à celui de Newark (New Jersey). Destiné à faciliter l'entrée sur le sol américain des étrangers effectuant de fréquents séjours, le système Inpass a consisté à enregistrer la forme de la main : largeur et épaisseur de la paume, taille et longueur des doigts, etc. [9]

La reconnaissance de la forme de la main consiste à déterminer les caractéristiques de la main d'un individu, la forme, la longueur, la large, la courbure des doigts, etc... Pour le fonctionnement du système, l'utilisateur place sa main sur un gabarit. Le tout est éclairé par une lumière infrarouge et l'image résultante est captée par une caméra digitale, La caractérisation est obtenue par mesure de la longueur et de la largeur de chaque doigts [10].

Cette modalité nécessite un capteur de grande taille ce qui empêche de l'utiliser dans les systèmes embarqués. De plus ce capteur est très cher par rapport à ceux des autres techniques. La reconnaissance basée sur la géométrie de la main n'est pas assez efficace dans le cas de blessure de la main, de deux personnes de la même famille ou dans le cas des jumeaux.

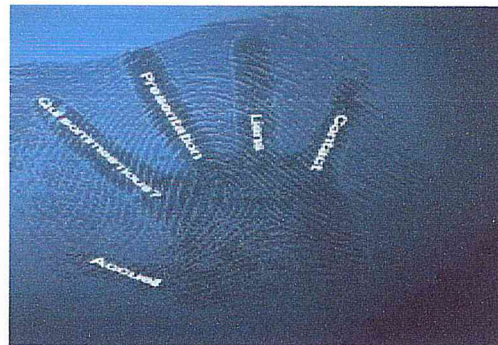


Figure I.4 : Forme de la main.

III.5. Signature

Elle est obtenue par une tablette graphique reliée à un ordinateur. La signature et surtout le mouvement réalisé par la main pour la fabriquer sont analysés en comparaison à plusieurs signatures de référence [10].

La signature est caractérisée par sa forme ou par sa dynamique. Il existe deux types de signatures hors-ligne et en ligne :

- La signature hors-ligne : qui est facile à falsifier. Elle est produite sur un papier.
- La signature en ligne : qui est produite sur une tablette graphique dédiée à capturer des caractéristiques dynamiques plus que l'image de la signature comme la vitesse et la pression exercée par le signataire, ce qui rend ce type de signatures difficile à falsifier.

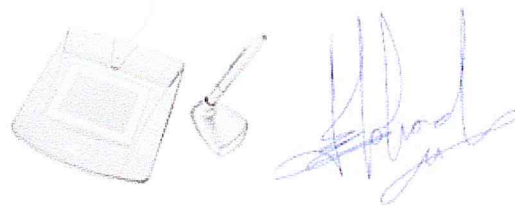


Figure I.5 : Signature.

III.6. Frappes au clavier

Cette modalité basée sur l'existence des caractéristiques uniques lorsqu'une personne tape sur un clavier. Par exemple, les travaux de Monroe et Rubin [11] ont montré qu'il est possible

de reconnaître une personne au rythme de sa frappe sur un clavier. L'avantage d'un système basé sur la dynamique de frappe au clavier est qu'il ne nécessite aucun équipement particulier de capture. En effet, il suffit d'installer un logiciel qui calcule le temps entre les frappes au clavier. Initialement, l'utilisateur tape plusieurs fois un mot de passe qui sert comme référence. Ce mot de passe est utilisé pour l'authentification conjointement avec les caractéristiques de frappe de l'utilisateur.

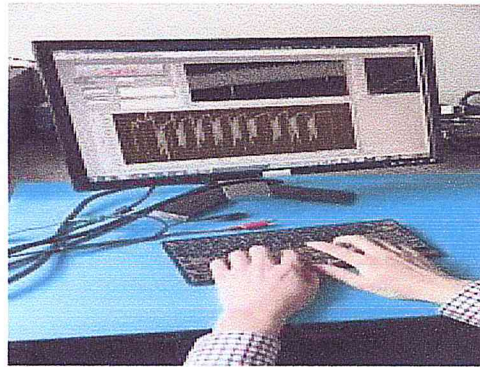


Figure I.6 : Frappes au clavier.

III.7. ADN

L'ADN (Acide Désoxyribo Nucléique) est une molécule contenant l'information génétique héréditaire. L'ADN est la modalité biométrique la plus sûre du monde car elle est stable et spécifique pour chaque individu. En 1995 et en Grande-Bretagne que la première banque d'ADN au monde est créée et commence à recueillir des échantillons d'ADN. Les États-Unis vont suivre en 1998 avec le *National DNA Index System* du FBI qui permet à tous les corps de police de comparer électroniquement des profils génétiques. Au tournant du millénaire, en 2000, la banque d'ADN de la Grande-Bretagne compte un million des profils génétiques. [12] Les systèmes basés sur l'analyse de l'ADN utilisent des techniques lourdes, coûteuses et qui nécessitent des délais de plusieurs semaines, ce qui interdit toutes les applications d'authentification en temps réel. C'est pour ça qu'ils sont utilisés généralement par les milieux judiciaires.

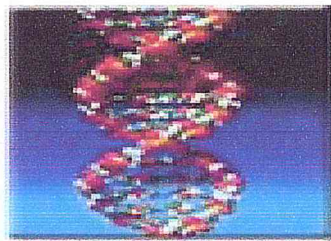


Figure I.7 : ADN.

III.8. Biométrie vocale :

Nos voix se sont un moyen de communiquer et encore un moyen fiable de nous reconnaître. Elles font partie intégrante de notre identité. C'est la raison pour laquelle les grandes entreprises se tournent aujourd'hui vers l'authentification vocale comme les banques. Contrairement à nos clés de voitures, et aux mots de passes ou codes PIN qu'on peut très souvent oublier, la voix est avec nous tout le temps. Cette sécurité et cette simplicité d'usage offerte par l'authentification biométrique vocale qui poussent les banques, les opérateurs de télécommunications et autres grandes organisations à choisir ce mode d'authentification.

La biométrie vocale, tout comme la reconnaissance et la synthèse vocale, s'est d'abord propagée dans les serveurs vocaux automatiques des centres d'appels. Mais aujourd'hui, elle est également utilisée dans des domaines aussi variés que l'authentification mobile et le paiement par cartes de crédit. Nous noterons toutefois que l'authentification biométrique par la voix ou les données vidéo peut être effectuée sans déranger les utilisateurs et que ces deux modalités offrent un niveau de sécurité relativement élevé [13], [14], [15].

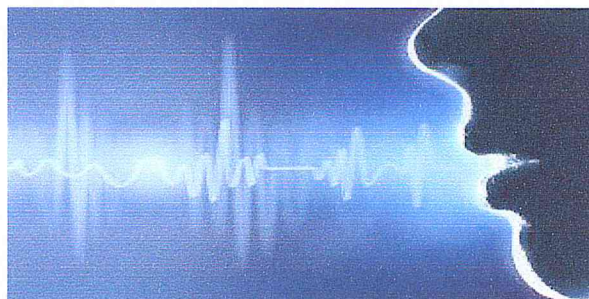


Figure I.8 : Signal vocal.

IV. Structure d'un système biométrique :

Le système biométrique se compose de deux grandes étapes illustrées par la figure I.9

- *L'apprentissage des personnes* : c'est la phase initiale qui ne s'effectuera qu'une seule fois. Elle comporte l'étape de création du modèle biométrique et l'étape de stockage du modèle avec l'identité déclarée en l'occurrence le nom, le prénom, etc.
- *Le test* : c'est la phase qui se déroulera à chaque fois qu'une personne se présentera devant le système [16]. cette phase comporte des étapes similaires aux étapes de l'apprentissage en plus de l'étape de comparaison et de décision.

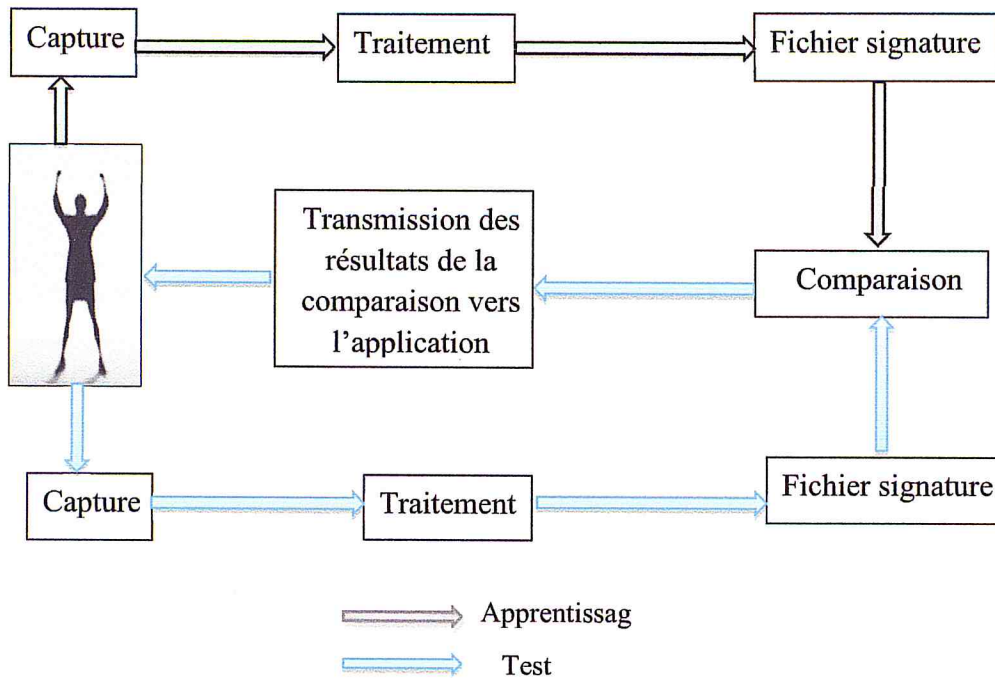


Figure I.9 : Schéma général d'un système biométrique.

V. Evaluation des performances d'un système biométrique

L'évaluation d'un système biométrique selon plusieurs critères : la vitesse d'exécution, la précision, la taille des paramètres stockés, etc. Dans ce qui suit, nous nous focalisons sur le deuxième aspect à savoir la précision, dans les deux modes identification et vérification.

V.1. En mode identification

Calcule le taux d'identification (*TID*) qui est le rapport entre le nombre de clients correctement identifiés par le système et le nombre total de clients [17].

$$TID = \frac{\text{Nombre de clients correctement identifiés}}{\text{Nombre totale des clients}} \quad I.1$$

V.2. En mode vérification

Il faut définir clairement trois critères principaux:

- **La fausse acceptation (FA) :** Elle correspond au cas où le système accepte un imposteur (une personne qui a proclamé une identité qui n'est pas la sienne).

$$FAR = \frac{\text{Nombre de FA}}{\text{Nombre d'imposteurs}} \quad I.2$$

- **Le faux rejet (FR) :** C'est quand le système rejette un client (une personne qui a proclamée sa vraie identité). Donc, les performances d'un système se mesurent par son taux de faux rejet (False Rejection Rate ou FRR) ou par son taux de fausse acceptation (False Acceptance Rate ou FAR) [18].

$$FRR = \frac{\text{Nombre de FR}}{\text{Nombre de clients}} \quad I.3$$

- Le troisième critère est connu sous le nom de **taux d'égale erreur** ("Equal Error Rate" ou EER). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où FRR = FAR, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations [18].

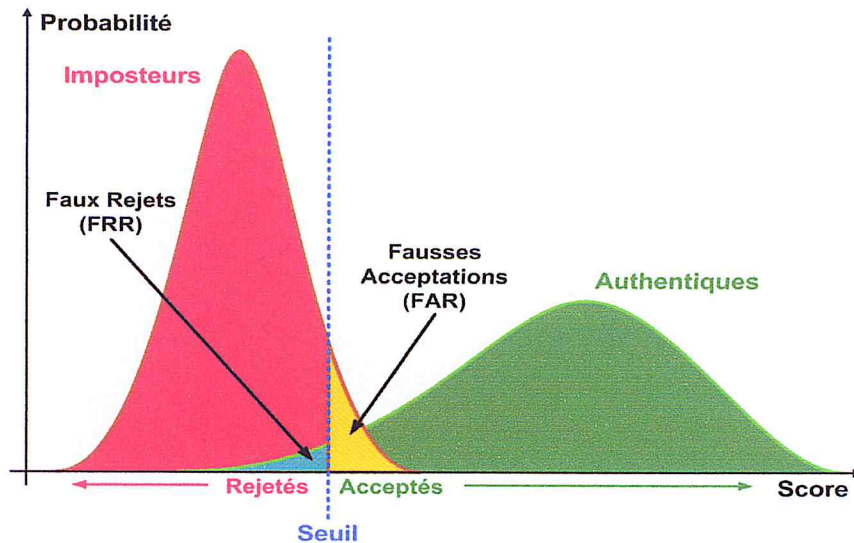


Figure I.10 Illustration du FRR et du FAR [31].

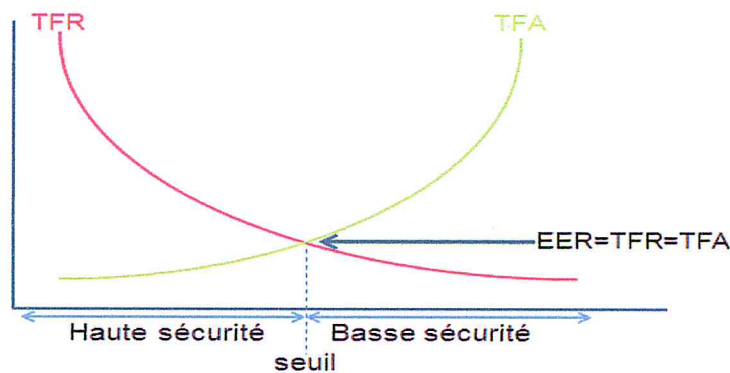


Figure I.11 Relation entre TFR, TFA et EER [31].

VI. Conclusion

Dans ce chapitre, nous avons passé en revue les principales technologies biométriques. Ensuite, nous avons présenté l'architecture de base d'un système biométrique. Dans le chapitre suivant, nous abordons la Reconnaissance Automatique de Locuteur notamment la vérification automatique de locuteur et l'identification automatique de locuteur.

CHAPITRE II :
RECONNAISSANCE
AUTOMATIQUE DE
LOCUTEUR(RAL)

I. Introduction

La reconnaissance automatique du locuteur RAL est le processus qui identifie automatiquement celui qui parle en se basant sur des informations individuelles incluses dans le signal de parole. La reconnaissance automatique du locuteur par leur voix est une tâche très complexe et nécessite la prise en compte de plusieurs paramètres. Le type d'application pour lequel cette reconnaissance est faite, impose l'utilisation des méthodes bien définies pour avoir des résultats significatifs. Il y'a plusieurs types de systèmes de reconnaissance automatique du locuteur. L'objectif principal de ce chapitre est de bien présenter les différents types des systèmes de reconnaissance automatique du locuteur.

II. la Reconnaissance Automatique de locuteur (RAL)

Les systèmes RAL ont pour mission de décoller l'information portée par le signal vocal. On classe les systèmes RAL comme suit (voir figure II.1) :

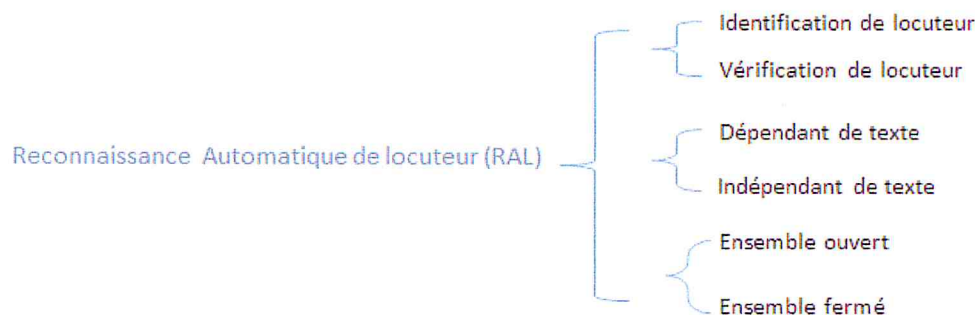


Figure II.1 : Structure d'un system de RAL.

- En reconnaissance automatique du locuteur, on fait la différence entre l'**identification** et la **vérification** du locuteur, selon que le problème est de vérifier que la voix analysée correspond bien à la personne qui est sensée la produire, ou qu'il s'agit de déterminer qui, parmi un nombre fini et destiné de locuteurs, a produit le signal analysé.
- On sépare reconnaissance du locuteur **dépendante du texte**, et reconnaissance **indépendante du texte**. Dans le premier cas, la phrase à prononcer pour être reconnu est fixée dès la conception du système ; et n'est pas précisée dans le deuxième. [19].
- En peut distinguer deux modes de décision et ça selon le type de l'application visée. L'identification en **ensemble fermé** suppose que le locuteur à identifier est forcément un des locuteurs connus du système ; la décision ou la réponse du système est ici l'identité

du locuteur dont la voix est la plus proche de celle testée. L'identification en **ensemble ouvert**, en revanche, ne fait aucun a priori concernant l'appartenance du locuteur de test à l'ensemble des locuteurs connus ; elle impose une étape supplémentaire dans le processus de décision afin d'accepter ou de rejeter le locuteur testé [20].

II.1. La Vérification Automatique de Locuteur (VAL)

Il s'agit de vérifier que la voix analysée correspond bien à la personne qui est censée de la produire. C'est le processus qui prend la décision d'accepter ou de rejeter l'identité d'un locuteur susceptible d'être la source d'un enregistrement vocal (figure II.2).

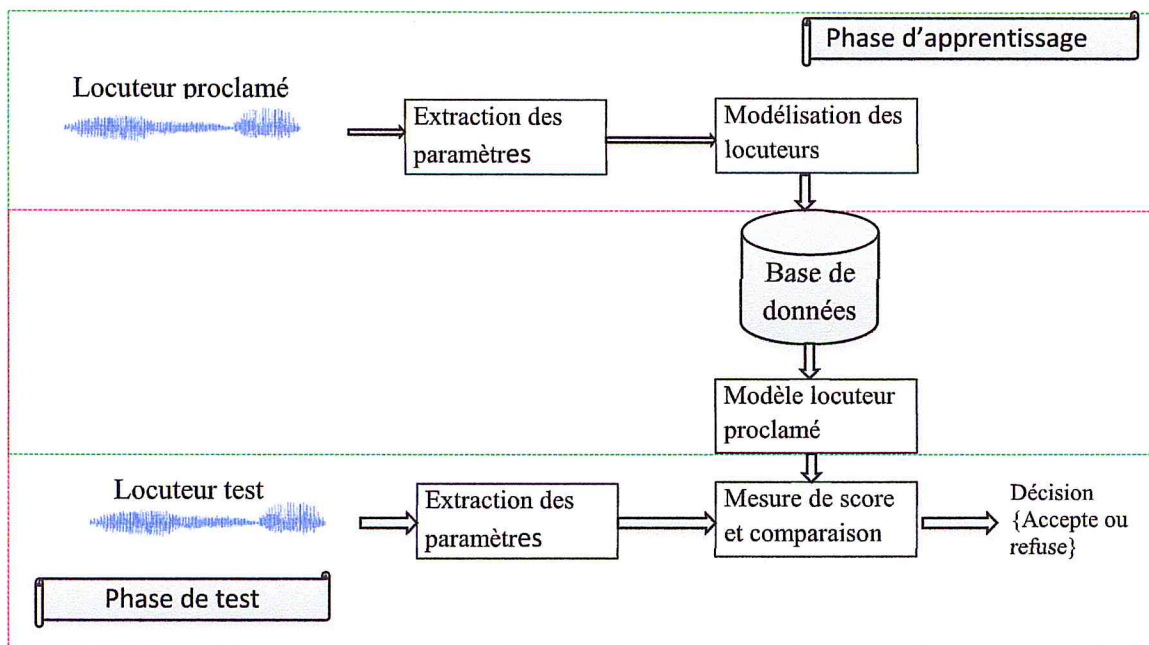


Figure II.2 : Structure de système de vérification de locuteur.

II.2. L'identification Automatique de Locuteur (IAL)

L'identification automatique du locuteur (IAL) fut historiquement l'une des premières applications de la RAL ([21], [22]), L'identification d'un individu est le processus qui consiste à décider quelle est, parmi une population connue, l'identité de l'utilisateur présent. Dans un premier temps, il est nécessaire pour tous les locuteurs de se faire connaître auprès du système, qui apprendra leurs caractéristiques à partir d'un ou plusieurs enregistrements de leur voix. Dans un second temps, lors de la phase de test, un individu se présente devant le système et compare sa voix de chacun des locuteurs connus puis la recherche de la voix la plus proche.

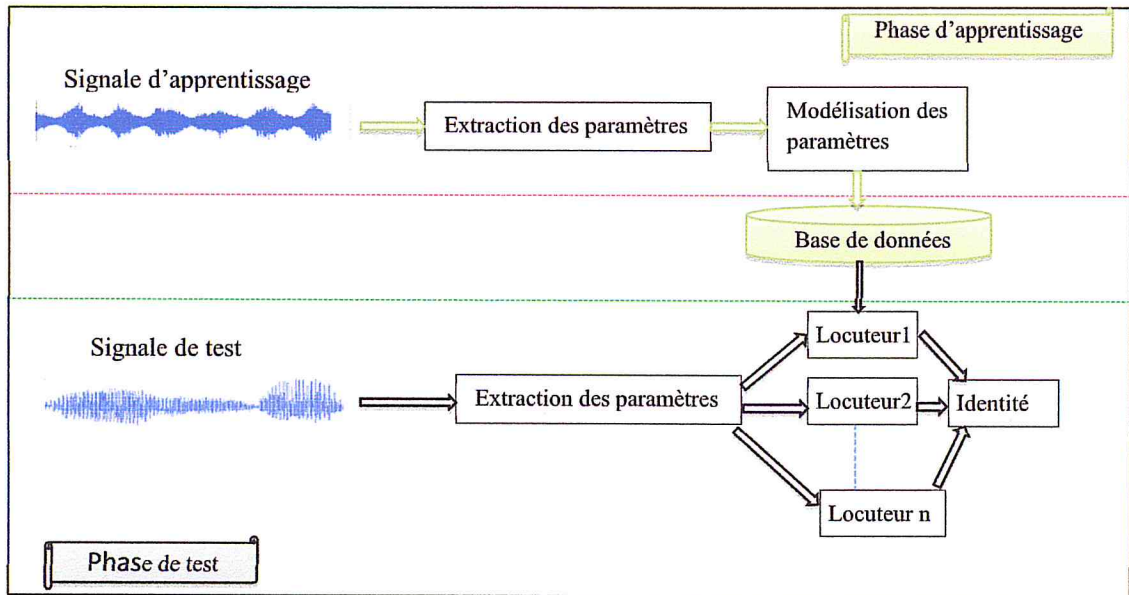


Figure II.3 : La structure de system d'identification du locuteur.

III. Les étapes d'un Système de Reconnaissance Automatique du locuteur

Les systèmes de reconnaissance du locuteur passent généralement par les étapes suivantes (Figure II.4):

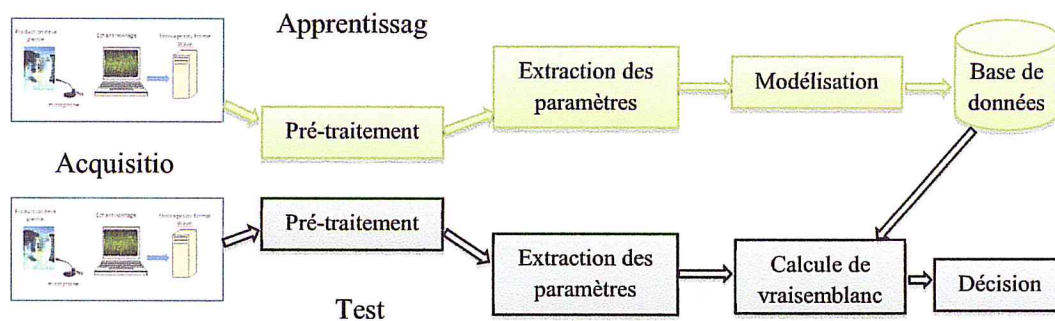


Figure II.4 : Etapes d'un système de RAL.

III.1 Acquisition du signal de la parole

Pour l'acquisition du signal, nous verrons différentes étapes par lesquelles passe un signal de parole depuis sa production dans le corps humain jusqu'à sa dernière transformation pour accéder à la phase des prétraitements.

III.1.1. Source de variabilité du signal de la parole

Le problème majeur rencontré dans la reconnaissance du locuteur provient de la variabilité de la parole ; aucun locuteur n'est capable de prononcer toujours une phrase de la même manière.

Différentes sources de dégradation du signal sont dues aux variations du locuteur, aux canaux de transmission utilisés et à l'environnement acoustiques [18].

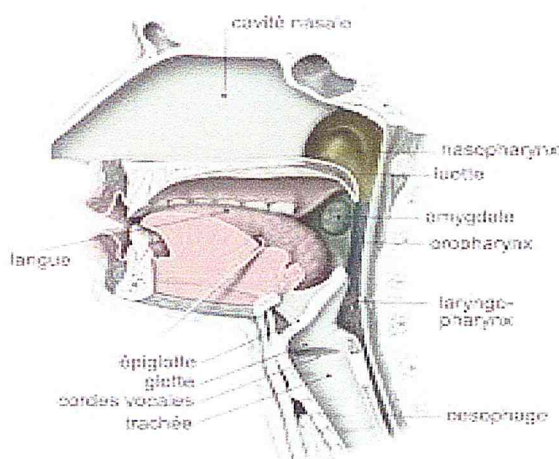


Figure II.5 : L'appareil vocal.

Les poumons se comportent comme un générateur d'air qui servira ensuite à alimenter le larynx. Le volume d'air est directement lié à l'énergie estimée sur de courts intervalles à partir du signal acoustique. Au niveau de larynx, on retrouve les cordes vocales qui en mode vibratoire génèrent des ondes glottiques de formes triangulaires, asymétriques et périodiques.

III.1.2. Echantillonnage

L'échantillonnage est le lien entre les signaux continu, dans notre cas le signal de la parole produit par le locuteur et les signaux à temps discret qui sont représentés dans la machine (figure II.6).

Donc L'échantillonnage est une opération qui décrit un signal par la donnée d'une suite numérique $X_n = X(nT_e)$ ou T_e est la période d'échantillonnage.

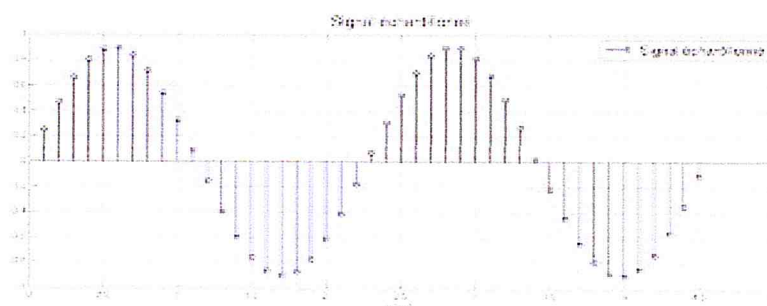


Figure II.6 : Echantillonnage d'un signal.

Le théorème de Shannon démontre qu'une fréquence d'échantillonnage supérieure à deux fois la fréquence la plus élevée contenue dans le signal éviterait un repliement du spectre du signal échantillonné et garantirait une représentation sans perte d'informations.

III.2. Prétraitements

En RAL le prétraitement effectué sur un signal vocal est un facteur important pour les performances du système. On va citer dans ce qui suit, quelques prétraitements standards utilisés par les systèmes de reconnaissance vocale actuels :

III.2.1. Découpage en trames

Après l'acquisition, le signal de parole qu'on aura n'est pas un signal stationnaire, le conduit vocal se déforme de façon continue et les paramètres du modèle sont donc variables dans le temps. Mais l'étude statistique qu'on veut faire exige la stationnarité alors on va le découper en trames de taille [20-30ms] chacune expérimentalement (le signal est quasi stationnaire dans cette plage) [23]. Ceci correspond à un segment de 256 échantillons du signal vocal (pour une fréquence d'échantillonnage dans [8,16] KHz) [2].

Entre chaque trame et la trame suivante, on applique un technique de chevauchement pour un taux de 25 ou 50% dans le but de préserver le maximum des paramètres, et pour éviter la discontinuité entre les trames ainsi que pour augmenter le nombre des échantillons statistiques à étudier (chaque trame est considérée comme un échantillon statistique).

$$\text{nbreTrame} = \frac{2 \times \text{NbEchantillons}}{\text{NbEchantillonParTrame}} - 1 \quad \text{II.1}$$

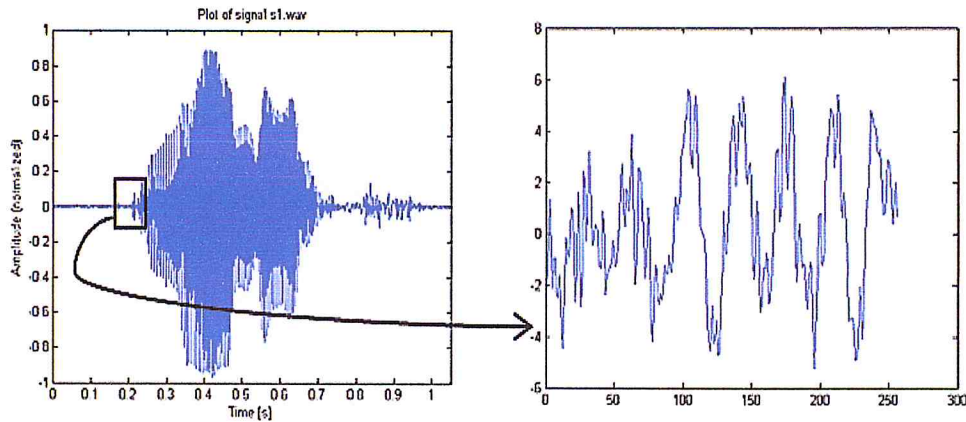


Figure II.7 : Découpage en trames.

III.2.2. Préaccentuation

Ce traitement est un filtre que nous appliquons sur toutes les trames et permet donc d'annuler l'effet nasal (l'atténuation des hautes fréquences) : c'est en effet dans les hautes fréquences que sont cachées une partie des paramètres qui serviront à la mesure de dissemblance. Le filtre le plus utilisé est le filtre à réponse impulsionnelle finie décrit ci-dessous:

$$H(z) = 1 - 0.95z^{-1} \quad \text{II.2}$$

La réponse de ce filtre peut être vue dans la Figure II.8.

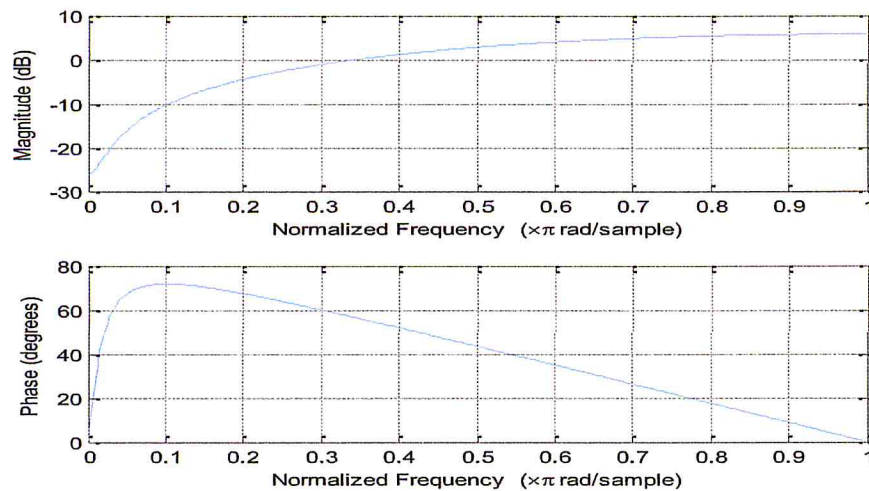


Figure II.8 : Le filtre de la préaccentuation.

III.2.3. Elimination de silence

Nous trouvons dans les signaux de parole des zones de silence, que signifie qu'il n'y a aucune information utiles et ce qui influe négativement sur les performances. Pour l'éliminer il existe plusieurs approches, on prend par exemple l'algorithme VAD (Voice Activity Detection).

➤ **Le principe de l'algorithme :**

Il s'agit de calculer le seuil de silence à partir des cinq premières trames qui sont considérées comme un silence généralement. On va comparer la valeur de chaque trame avec le seuil, La trame qui a une valeur inférieure au seuil est éliminée (on considérées comme un silence). Les mesures les plus utilisées pour trouver et éliminer le silence sont : l'énergie du signal, le rapport de passage par zéro.

a. L'énergie $E(x)$ d'une trame x :

Une trame de silence à une énergie quasi nulle, contrairement à une trame de parole. L'énergie d'une trame x pour un signal $s(n)$ est définie par :

$$E(x) = \sum_{n=0}^{l-1} s(n)^2 \quad \text{II.3}$$

Ou $s(n)$: est la valeur du signal dans l'échantillon.

l : la taille d'une trame en nombre d'échantillons.

b. Le taux de passage par zéro ou le ZCR (Zéro Crossing Rate) :

Le rapport de passage par zéro donne une mesure du nombre de fois où le signal change de signe ($ZCR(x)$ est le nombre de passage par zéro (zéro crossing Rate) pour la trame). Ce rapport est en général plus grand dans les régions non voisées.

Le ZCR d'une trame x d'un signal $s(n)$ est défini par :

$$ZCR(x) = \sum_{n=0}^{l-1} |sgn(s(n) - sgn(s(n-1)))| \quad \text{II.4}$$

Ou :

$$sgn(a) = \begin{cases} 1 & \text{si } a \geq 0 \\ 0 & \text{si } a < 0 \end{cases} \quad \text{II.5}$$

Cet algorithme utilise conjointement ces deux paramètres dans la fonction suivante :

$$w_s = E_s(x) * (1 - ZCR_s(x)) \quad \text{II.6}$$

Pour chaque trame, la valeur de $w_s(x)$ best comparé à un seuil pour décider de sa nature (silence ou parole).

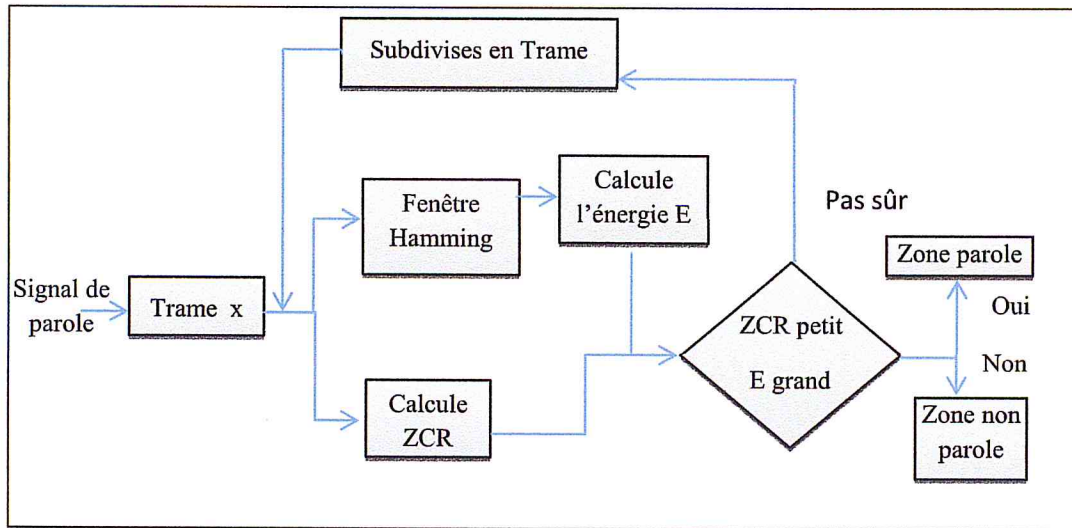


Figure II.9 : Schéma de classification vocal / non vocal [24].

➤ **Calcul de seuil**

On supposant que les 120 premières millisecondes (5 trames) de notre signal sont généralement dues silence. Le seuil de décision est donné par la formule :

$$Seuil = \mu_{w_s} + \alpha \delta_{w_s} \quad \text{II.7}$$

Avec : w_s c'est la fonction calculée dans (II.6)

$\alpha = \delta^{-08}$, μ_{w_s} est la moyenne des valeurs calculées par la fonction w_s pour les 5 premières trames et δ_{w_s} leurs écart type.

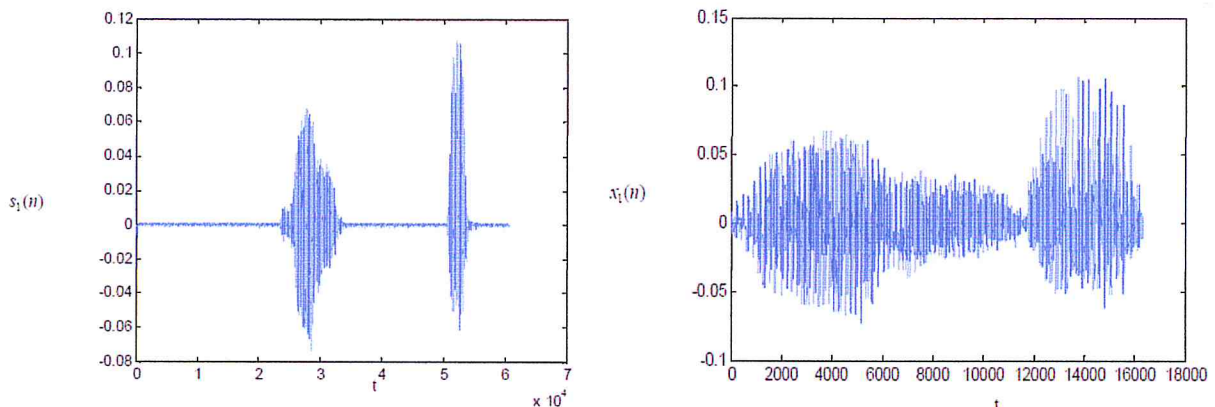


Figure II.10 : Du signal $s_1(n)$ avec silence au signal $x_1(n)$ sans silence en utilisant la fonction $VAD(n)$.

III.2.4. Fenêtrage

Le principe de Cette phase consiste à découper le signal $x_1(n)$ en trames et d'appliquer une fenêtre pour chacune d'elles. Voir la figure II.11

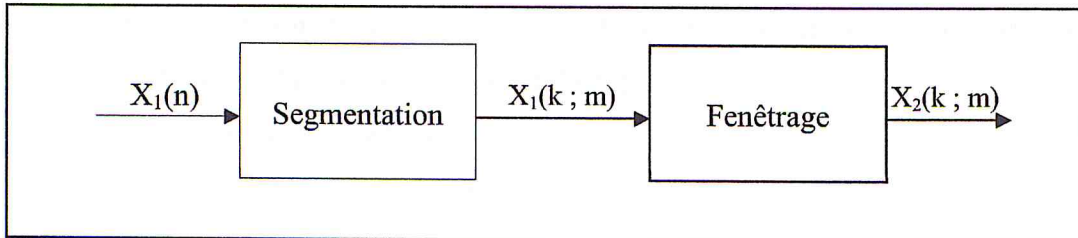


Figure II.11 : Les étapes du fenêtrage.

Chaque trame est de longueur de K échantillons, tel que les trames adjacentes sont séparées par P échantillons [20] voir la figure II.12.

Ensuite, on applique une fenêtre à chaque trame pour réduire la discontinuité à la fin de chacune d'elles. La fenêtre la plus utilisée est la fenêtre de *Hamming*, elle est définie comme suit :

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N - 1 \quad \text{II.8}$$

Avec N la longueur de trame

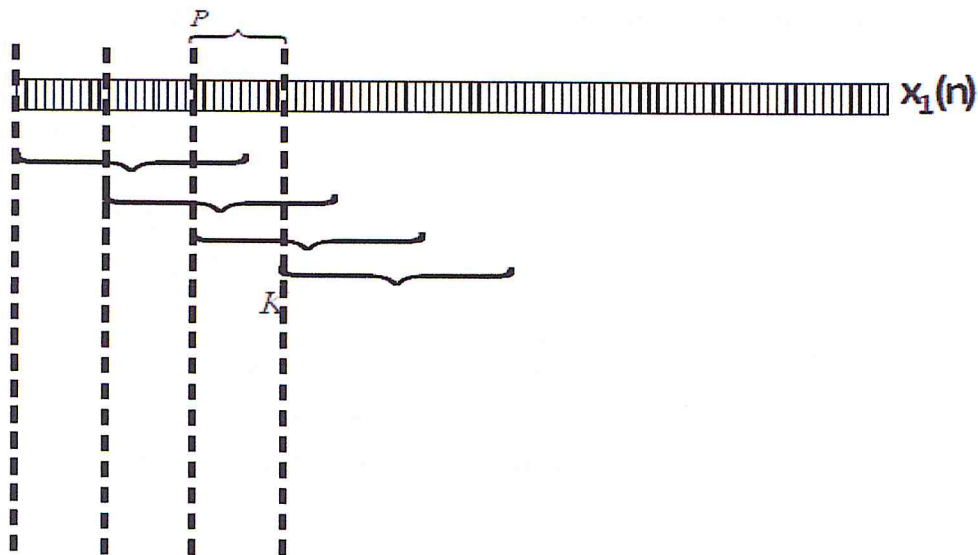


Figure II.12 : Le découpage en trames.

III.3. Extraction des paramètres acoustiques

III.3.1. Analyse Cepstrale :

Le cepstre est basé sur une connaissance du modèle de production de la parole, une modélisation du signal de parole consiste à définir ce signal comme le résultat de la

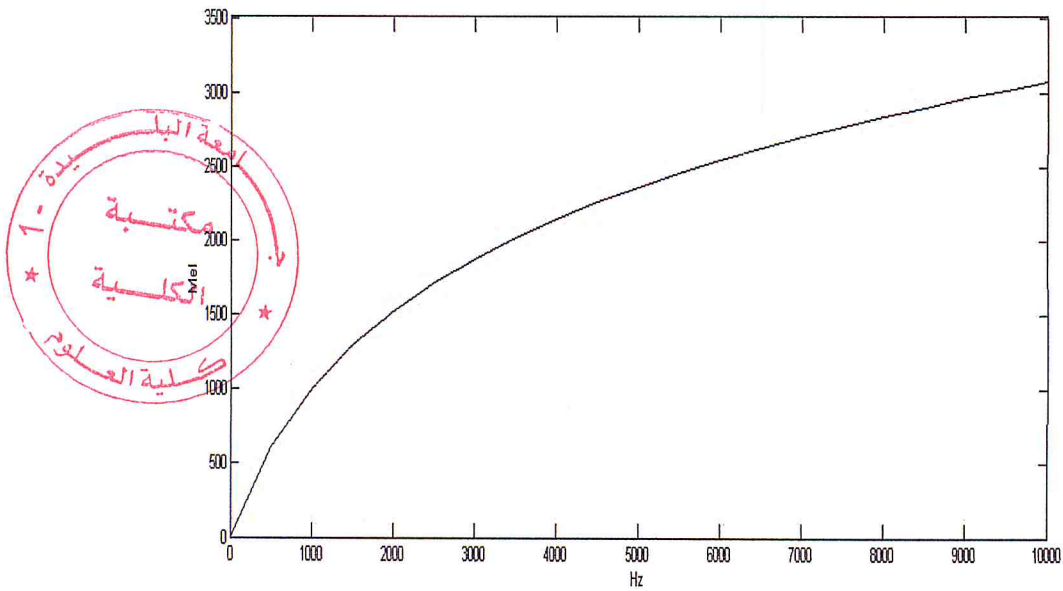


Figure II.14 : L'allure de la sélectivité de l'oreille humaine.

Pour des raisons pratiques, le spectre Mel est simulé en utilisant un banc de K filtres (R_i) triangulaires positionné uniformément sur l'échelle Mel, tels que chaque triangle couvre une fréquence. Ces triangles ont de mêmes tailles dans la zone inférieure à 1KHz pour assurer la linéarité et ils ont des tailles croissantes dans la zone supérieure à 1KHz pour assurer l'allure logarithmique. (Figure II.15).

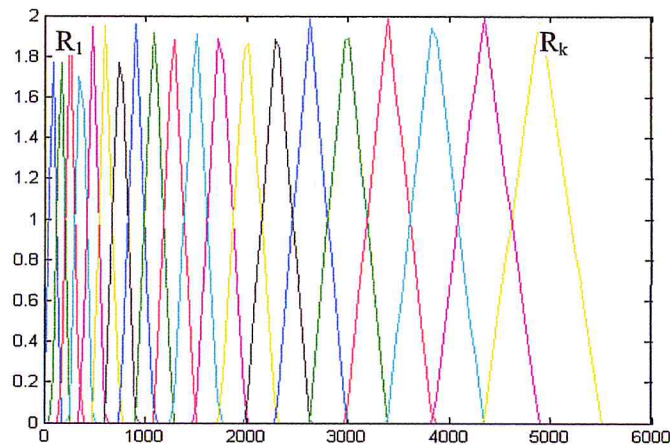


Figure II.15 : Répartition des filtres sur les échelles Mels.

Finalement les coefficients sont calculés consiste à appliquer une transformée en cosinus discret(DCT) d'où résultent les coefficients cepstraux (MFCC). Avec la FFT nous avons passé au fréquentiel et avec la DCT nous revenons vers le temporel, nous avons utilisé DCT au lieu de FFT^{-1} car DCT a l'avantage de dé corrélacion. (i.e. une matrice de covariances diagonale) :

$$mfcc[n] = \sum S[m] \cos\left(\frac{\pi n(m-\frac{1}{2})}{M}\right), \quad 0 \leq n \leq M \quad \text{II.13}$$

Enfin, nous avons extrait les N coefficients MFCCs de chaque trame, les données que nous allons travailler avec sont une matrice de N lignes et du nombre de trames colonnes pour chaque locuteur.

III.4. Modélisation

La modélisation de locuteur est l'étape que permet de générer un modèle compact et représentatif à partir des données d'apprentissage. Il existe plusieurs approches utilisées pour modéliser les données extraites de la voix et du visage, on cite parmi eux les approches suivantes:

III.4.1. Quantification vectorielle

La quantification vectorielle (Vecteur Quantization : VQ) repose sur un partitionnement de l'espace acoustique en sous-espaces (Figure.10). Chaque sous-espace est associé à un vecteur centroïde. Dans ces conditions, un modèle de locuteur est composé d'un ensemble de vecteurs centroïde, appelé dictionnaire de quantification (codebook). Lors de la phase de test, on fait la comparaison des trames extraites d'un signal vocal avec le dictionnaire des vecteurs. Cette comparaison est faite en calculant la distance qui sépare les trames en entrée avec les centres des partitions. La distance mesure le degré de similarité entre la voix de test et le modèle d'un locuteur. La quantification vectorielle s'applique en mode dépendant ou indépendant du texte [26], [27]. Cependant, cette méthode élimine beaucoup d'information sur les locuteurs et elle nécessite des paroles très longues pour avoir des informations statistiques stables.

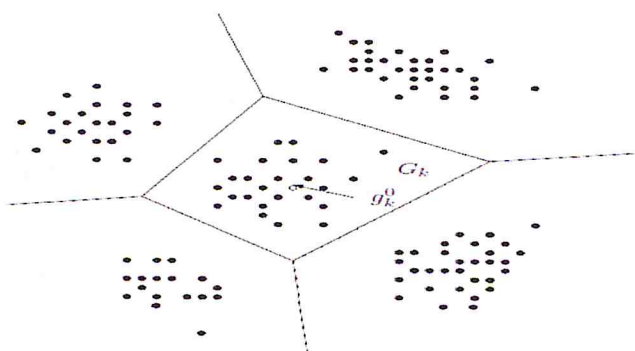


Figure II.16 : Définition des classes de données en quantification vectorielle.

CHAPITRE II :

Modélisation de Locuteur

I. Introduction

Dans ce chapitre, nous présentons un module important de la Reconnaissance Automatique du Locuteur. Comme nous l'avons déjà mentionné dans le chapitre précédent, les systèmes de RAL nécessitent une étape de modélisation statistiques des paramètres (ou vecteurs) acoustiques, extraits du signal de parole du locuteur. En effet, plusieurs méthodes de modélisation statistique ont été utilisées et chacune d'elle présente des avantages et des inconvénients. Nous commençons d'abord par détailler le modèle statistique le plus utilisé dans les systèmes RAL en mode indépendant du texte, à savoir, le modèle **GMM** (Gaussian Mixture Model), le modèle **GMM-UBM**, le modèle **JFA** et le modèle **I-Vecteur**. Par la suite, nous mettons l'accent sur les deux modèles **GMM-UBM** et **I-vecteur**, qui sont utilisés comme classifieur discriminatif.

II. Approche GMM (Gaussians Mixture Model)

Les mélanges de gaussiennes sont utilisés pour modéliser un locuteur donné par une somme pondérée de gaussiennes. On peut assimiler un modèle GMM à un HMM (Hidden Markov Model) à un seul état. On ne modélise donc pas les aspects temporels du signal. Cette méthode est la plus utilisée en ce qui concerne la reconnaissance de locuteur en mode indépendant du texte. Les travaux de D.A.Reynolds [28] consistent l'état de l'art en la matière. Le modèle GMM est une segmentation en plusieurs classes des paramètres acoustiques représentant l'identité d'un locuteur (Figure III.1).

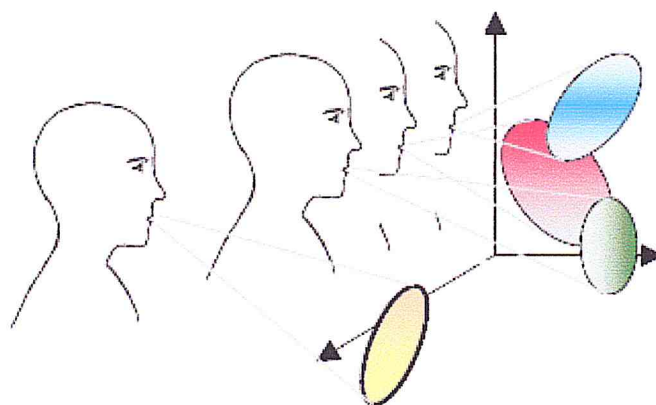


Figure.III.1 : Illustration de nuages acoustiques représentant l'identité d'un locuteur.

Un mélange de gaussiennes (Figure.III.2) est une somme pondérée de M distributions gaussiennes [32].

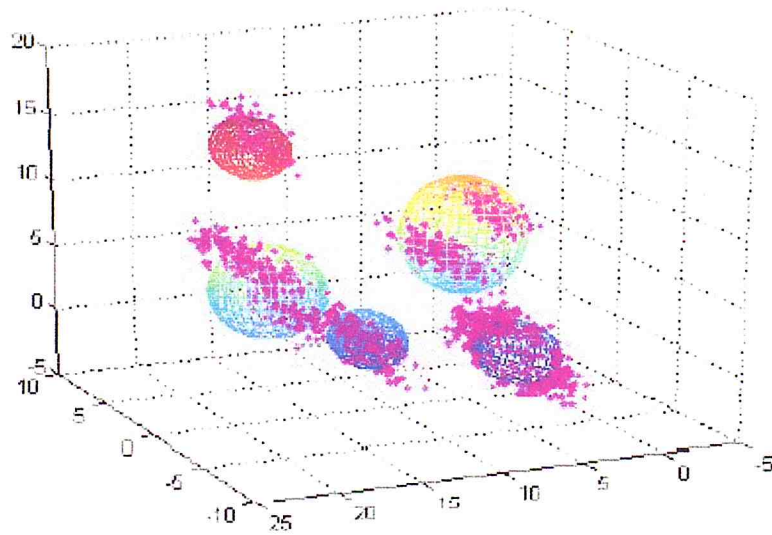


Figure.III.2 : Mélanges des gaussiennes en trois dimensions.

Cela nous amène à considérer que les vecteurs acoustiques (étudiés dans le chapitre précédent) sont des réalisations de variables aléatoires mutuellement indépendantes de densité de probabilité $b_m(\vec{x})$ gaussienne.

Le mélange de gaussiennes $p(\vec{x}|\theta)$ est donné par la formule (III.1) qui est une somme pondérée par les coefficients π_m des distributions gaussiennes $b_m(\vec{x})$.

$$p(\vec{x}|\theta) = \sum_{m=1}^M \pi_m b_m(\vec{x}) \quad \text{III.1}$$

Où :

- \vec{x} est un vecteur aléatoire de dimension d qui représente un vecteur acoustique.
- $b_m(\vec{x})$, $m=1,2,\dots, M$ sont les densités de probabilités gaussiennes qui composent le modèle.
- π_m , $m=1,2,\dots, M$, sont les poids des densités $b_m(\vec{x})$.

Chaque densité de probabilité $b_m(\vec{x})$ est définie par :

$$b_m(\vec{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma_m|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}_m)' \Sigma_m^{-1} (\vec{x} - \vec{\mu}_m)^T \right\} \quad \text{III.2}$$

Tel que :

$\overline{\mu}_m$ est le vecteur moyen et Σ_m est la matrice de covariance. Les poids π_m vérifient la contrainte donnée par la formule suivante :

$$\sum_{m=1}^M \pi_m = 1 \quad \text{III.3}$$

Ce modèle statistique est composé par des vecteurs moyens $\overline{\mu}_m$, des matrices de covariances Σ_m et des coefficients π_m de toutes les densités de probabilités $b_m(\vec{x})$. Ces paramètres sont représentés par la notation suivante :

$$\theta = \{\pi_m, \overline{\mu}_m, \Sigma_m\}, m = 1, \dots, M \quad \text{III.4}$$

III. Approche GMM-UBM (Modèle de Monde)

La modélisation des locuteurs par l'approche GMM-UBM passe par les étapes suivantes :

III.1. Estimation du modèle UBM

La modélisation des locuteurs par l'approche GMM-UBM nécessite un modèle global (modèle UBM) qui représente théoriquement l'ensemble des paramètres de la population en question. La création de ce modèle passe par les étapes suivantes :

III.1.1. Concaténation des paramètres MFCC

Le modèle UBM est estimé à partir de la matrice des paramètres acoustique.

III.1.2. Initialisation de l'algorithme EM (*Expectation Maximisation*)

La modélisation des paramètres par l'algorithme EM nécessite une bonne initialisation de ce dernier. Pour cela, nous avons utilisé l'algorithme K-means [41].

III.1.3. Application de l'algorithme EM

L'objectif de l'algorithme EM est de maximiser d'une façon itérative la vraisemblance entre les paramètres et leur modèle associé. Les deux étapes d'estimation et de maximisation de cet algorithme sont données par :

- **Estimation :**

L'estimation consiste à calculer pour chaque vecteur X_i de la matrice des paramètres acoustique, et chacune des gaussiennes j du modèle UBM, l'appartenance P_{ij} du vecteur X_i à la gaussienne j

$$P_{ij}(j/x_i) = \frac{\omega_j p_j(X_i)}{\sum_{l=1}^L \omega_l p_l(X_i)} \quad \text{III.5}$$

Ou

$$p_n(x_m) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{D}{2}}} \exp(-1/2(x_m - \mu_n)^T \Sigma_n^{-1}(x_m - \mu_n)) \quad \text{III.6}$$

• **Maximisation :**

La maximisation consiste à mettre à jour les poids, les moyens et les matrices de variance-covariance suite aux résultats obtenus lors de l'estimation.

$$w_j = \sum_{t=1}^T P_r(j/X_t, \lambda) \quad \text{III.7}$$

$$\mu_j = \frac{\sum_{t=1}^T P_r(j/X_t, \lambda) X_t}{\sum_{t=1}^T P_r(j/X_t, \lambda)} \quad \text{III.8}$$

$$\sigma_j^2 = \frac{\sum_{t=1}^T P_r(j/X_t, \lambda) X_t^2}{\sum_{t=1}^T P_r(j/X_t, \lambda)} - \mu_j^2 \quad \text{III.9}$$

T est le nombre des vecteurs dans la matrice des paramètres acoustique.

On répète les deux étapes d'estimation et de maximisation jusqu'à ce que la variation de la vraisemblance (équation III.24) entre l'itération i et l'itération $i + 1$ soit inférieur à certain seuil ϵ .

III.1.4 Estimation des modèles des locuteurs par l'adaptation MAP

Etant donnée le modèle UBM, la création des modèles des locuteurs par l'adaptation MAP est donnée par [40]:

Pour chaque vecteur X_t de la matrice des paramètres acoustique d'un client, et chacune des gaussiennes i du modèle UBM, nous calculons les statistiques suffisantes :

$$n_i = \sum_{t=1}^T Pr(i/x_t, \lambda_{proir}) \quad \text{Poid,} \quad \text{III.10}$$

$$E_i(x) = \frac{1}{n_i} \sum_{t=1}^T Pr(\frac{i}{x_t}, \lambda_{proir}) x_t, \quad \text{Moyen,} \quad \text{III.11}$$

$$E_i(X^2) = \frac{1}{n_i} \sum_{t=1}^T Pr(i/\lambda_{proir}) x_t^2 \quad \text{Variance,} \quad \text{III.12}$$

En utilisant ces statistiques, on met à jour les paramètres de la gaussienne i :

$$\widehat{w}_i = [\alpha_i^w n_i / T + (1 - \alpha_j^w) w_i] \gamma \quad \text{Adapté le poids} \quad \text{III.13}$$

$$\widehat{\mu}_i = \alpha_i^m E_i(x) + (1 - \alpha_j^w) \mu_i \quad \text{Adapté la moyen} \quad \text{III.14}$$

$$\widehat{\sigma}_i^2 = \alpha_i^v E_i(x^2) + (1 - \alpha_i^v)(\sigma_i^2 + \mu_i^2) - \widehat{\mu}_i^2 \quad \text{Adapté la variance} \quad \text{III.15}$$

Ou $\alpha = \frac{n_i}{n_i + r^p} = \{m, v, w\}$ et r^p est un facteur de confiance constant. Dans notre travail nous avons pris $r_w = r_m = r_v = 16$.

Finalement, le modèle du client i est défini par les nouveaux paramètres du modèle UBM après l'adaptation.

III.2 Calcule des scores

La similarité entre la donnée test $X_t = \{X_1; X_2; \dots; X_T\}$ et le modèle du client λ_i est donnée par

$$LLR(X_t/\lambda_i) = \frac{1}{T} \sum_{t=1}^T (\log P(X_t/\lambda_i) - \log P(X_t/\lambda_{UBM})) \quad \text{III.16}$$

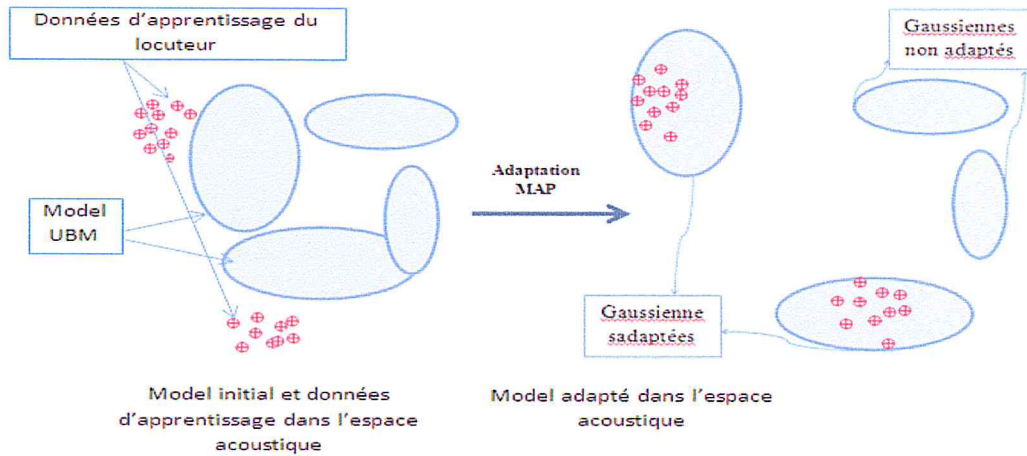


Figure III.3 : Adaptation MAP d'un modèle GMM-UBM.

L'architecture globale du système à base de la méthode GMM-UBM est donnée par la figure III.4

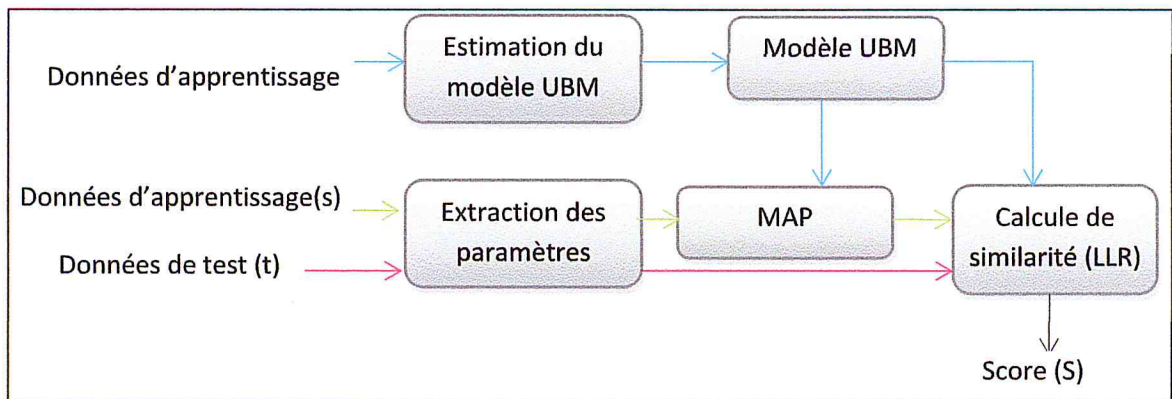


Figure III.4 : Architecture du système à base de la méthode GMM-UBM.

IV. Approche JFA L'analyse jointe de facteur (Joint Factor Analysis)

Dans l'espace des super vecteurs, le supervecteur M_s (de taille $C * F$, C c'est le nombre des composantes GMM et F est la dimension des vecteurs des paramètres) du client s est décomposé en deux composantes une composante S dépend du locuteur et une autre composante C dépend du canal.

$$M_s = S + C \quad \text{III.17}$$

Où

$$S = M + Vy_s + Dz_s \quad \text{III.18}$$

$$C = Ux_{h,s} \quad \text{III.19}$$

M est le supervecteur indépendant du canal et du locuteur ; généralement c'est le supervecteur de modèle UBM; V est la matrice de variabilité du locuteur de taille $CF \times R_s$ ($R_s \ll CF$), D est une matrice diagonale de taille $CF \times CF$ qui représente l'espace de variabilité restante, et U est la matrice de variabilité du canal (session) de taille $CF \times R_c$ ($R_c \ll CF$). y_s , z_s et $x_{h,s}$ sont des vecteurs aléatoires indépendants qui suivent une distribution normale de moyenne nulle et de variance unité. y_s est le vecteur des facteurs du locuteur, il représente les coordonnées du locuteur s dans l'espace formé par la matrice VV^T . z_s est le vecteur des facteurs communs. Le terme Dz_s décrit les variabilités restantes des locuteurs qui ne sont pas capturées par l'espace du locuteur ; $x_{h,s}$ est le vecteur des facteurs du canal qui représente les coordonnées de la session h dans l'espace de variabilité du canal formé par la matrice UU^T

Les matrices V , D et U sont appelées les hyperparamètres du modèle JFA. Elles sont estimées à partir d'un grand ensemble de locuteurs ou chaque locuteur disposé de plusieurs enregistrements. L'estimation est faite par l'algorithme EM, on commence tout d'abord par estimer la matrice V en supposant que les deux autres matrices sont nulles. Etant donné l'estimation de la matrice V , on estime ensuite la matrice U en gardant toujours la matrice D nulle, à la fin on estime la matrice D en considérant l'estimation des deux autres matrices.

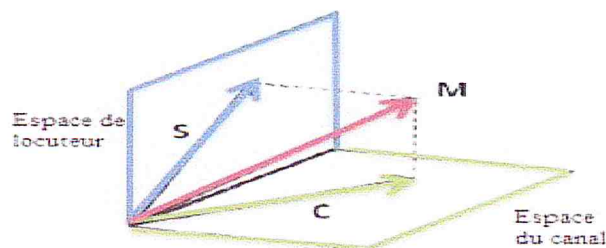


Figure III.5 Décomposition du supervecteur du locuteur

Enfin on peut représenter la forme générale de la technique JFA sous la forme :

$$m(s) = M + Vy(s) + Dz(s) + Ux_h(s) \quad \text{III.20}$$

Avec M est le supervecteur indépendant du canal et de locuteur.

V est la matrice de variabilité du locuteur.

D la matrice de variabilité restante.

U la matrice de variabilité du canal.

Et $y(s)$, $z(s)$ et $xh(s)$ sont respectivement les facteurs du locuteur, les facteurs communs et les facteurs du canal.

V. Approche (I-Vecteur)

Motivé par l'utilisation des facteurs du locuteur directement dans la modélisation SVM que a été proposée par Dehake et al [33] ; nous avons utilisé une nouvelle approche de modélisation du locuteur nommée I-Vecteur (I-Vecteur) pour Identity vecteur. Contrairement à la méthode JFA, la méthode I-Vecteur représente les sous-espaces du locuteur et du canal avec un seul espace de variabilité totale. Cette représentation est motivée par le fait que l'espace du canal de la représentation JFA contient des informations qui peuvent aider à la discrimination entre les locuteurs. La nouvelle représentation sera donc :

$$m(s) = M + Tw(s) \quad \text{III.21}$$

Où M est le supervecteur indépendant du canal et de locuteur, T est la matrice de variabilité totale, et w est le vecteur d'identité (I-Vecteur).

On peut résumer les étapes de la vérification de locuteur à base de I-Vecteur par le chemin suivant :

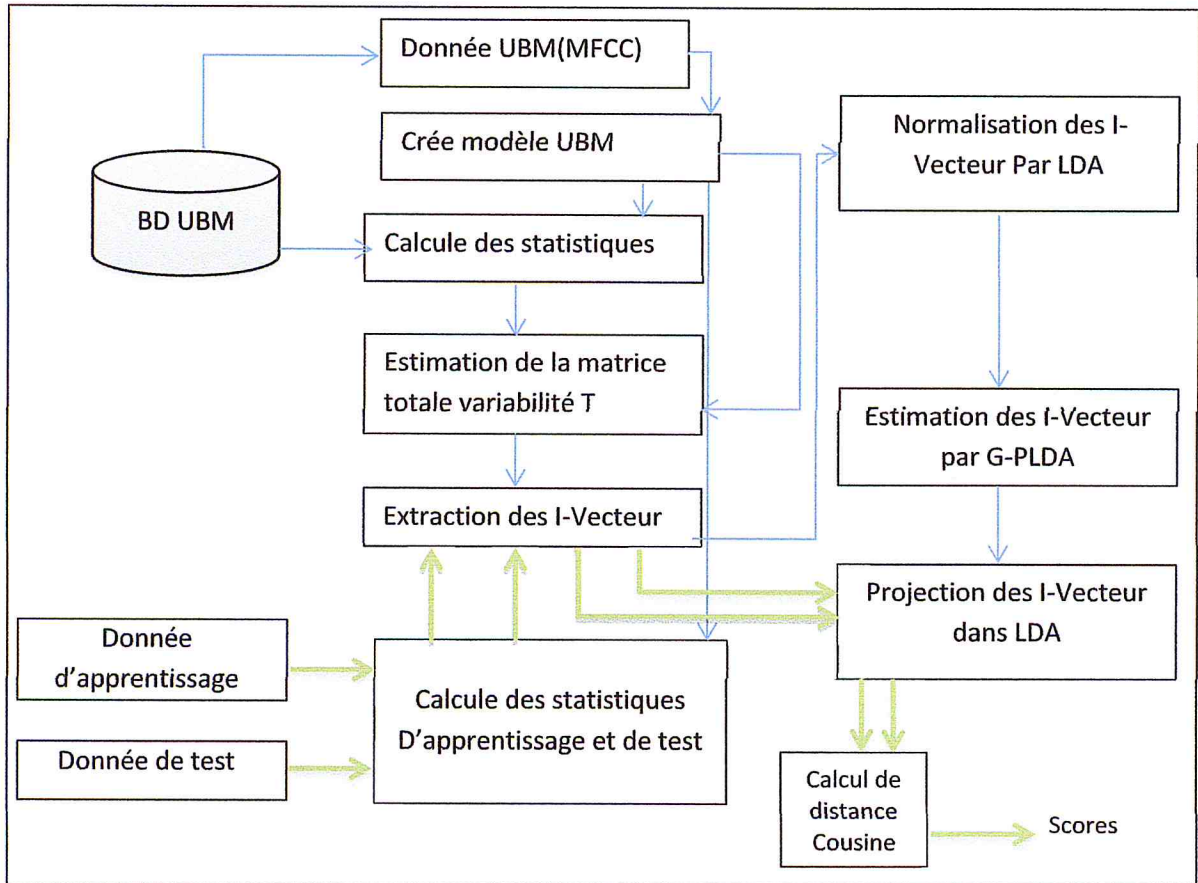


Figure III.6 : Architecture du système à base de la méthode I-Vecteur.

A partir de la figure III.6 on peut citer les étapes vérification de locuteur à basse de I-Vecteur de manier générale :

V.1. Estimation de la matrice de variabilité totale T

Dans l'estimation de la matrice T nous traitons les séquences de parole du même locuteur comme si elles étaient produites par différents locuteurs. L'algorithme d'estimation de cette matrice est donné par:

Algorithme 1 : Algorithme d'estimation de la matrice T

- Pour chaque enregistrement u :
 - Calculer les statistiques d'ordre zéro et d'ordre un $N(u)$ et $F(u)$
 - Centrer le premier ordre statistique $\tilde{F}(u) = F(u) - N(u)M$
- Initialiser T aléatoirement

- Répéter pour un certain nombre d'itérations (environ 10 itérations)
- Pour chaque enregistrement u
 - Calculer la matrice $G(u) = (I + T^T \Sigma^{-1} N(u) T)^{-1}$
 - Calculer les statistiques du vecteur $w(u)$
 - $E_1(u) = E(w) = G(u) T^T \Sigma^{-1} \tilde{F}(u)$
 - $E_2(u) = E(ww^T) = E_1(u) E_1(u)^T + G(u)$
- Résoudre le système d'équations $\sum_u N(u) \tilde{T} E_2(u) = \sum_s \tilde{F}(u) E_1(u)^T$
- Remplacer T par \tilde{T}

On généralise le but de la matrice des variabilités totale T et reprissent une maximum d'espace pour compte toutes les paramètres ou les variables du locuteur et de canal, pour éviter des pertes d'information.

V.2 L'extraction des I-Vecteur

L'extraction des I-Vecteur qui a été proposée par N. Dehak et al [34], est une représentation vectorielle pour les paramètres acoustiques d'un locuteur UBM à partir des statistiques, la matrice de variabilité totale (T) et les MFCC UBM.

V.3. Compensation de l'effet de canal

La normalisation de dimension de totale variabilité (TvDim).

V.3.1 LDA (Linear Discrimination Analysis)

La motivation pour l'utilisation de LDA est de maximiser la variance entre-locuteur et de minimiser la variance intra-locuteur, qui est le point important dans la vérification du locuteur. [2]. La matrice de projection (A) de cette méthode est obtenue en résolvant le problème des valeurs propres suivant :

$$S_b v = \lambda S_w v \quad \text{III.22}$$

Où S_b et S_w sont respectivement la matrice de variabilité inter-classes et la matrice de variabilité intra-classe.

$$S_b = \sum_{i=1}^L (w_i - \bar{w})(w_i - \bar{w})^t \quad \text{III.23}$$

$$S_w = \sum_{i=1}^L \frac{1}{n_i} \sum_{j=1}^{n_i} (w_i^j - \bar{w}_i)(w_i^j - \bar{w}_i)^t \quad \text{III.24}$$

L est le nombre des locuteurs, n_l est le nombre des I-Vecteurs pour chaque locuteur l (locuteur), \overline{w}_l est le moyen des I-Vecteur pour chaque locuteur l et w est le moyen des tous les I-Vecteurs.

V.4 G-PLDA (Gaussian Probabilistic LDA)

L'analyse discriminante linéaire probabiliste (Probabilistic Linear Discriminant Analysis, PLDA) [35] est une technique qui a été initialement proposée en reconnaissance de visages. En reconnaissance du locuteur [36], elle peut être vue comme un équivalent de la JFA (Joint Factor Analysis) dans l'espace des I-Vecteurs, c'est à dire dans un espace mono-gaussien. Ce modèle est génératif. Il ignore le mécanisme d'extraction des I-Vecteurs pour les considérer comme des observations d'un modèle génératif probabiliste. Il fait l'hypothèse que tout I-Vecteur w de dimension p peut être décomposé comme :

$$w = \mu + \Phi\gamma_s + \Gamma z + \epsilon \quad \text{III.25}$$

Ce modèle est formé de deux parties :

- (i) La composante $\mu + \theta\gamma_s$ ne dépend que du locuteur.
- (ii) La composante $\Gamma z + \epsilon$ diffère pour chaque session de voix et représente la "variabilité" intra-locuteur.

V.5. Calcul des scores

Pour calculer la similarité entre la donnée test représentée par son I-vecteur (w_T) et la donnée d'enroulement représentée par son I-vecteur (w_E), une simple mesure de distance scoring est utilisée le log-vraisemblance (log-likelihood) est défini par la formule suivante :

$$score(w_1, w_2) = \log \frac{P(w_1, w_2 / \phi_{tar})}{P(w_1, w_2 / \phi_{non})} \quad \text{III.26}$$

Ce logarithme du rapport des vraisemblances se calcule facilement dans le cas d'une PLDA Gaussienne (les vraisemblances marginales sont gaussiennes).

VI. Conclusion

Nous avons exposé à travers ce chapitre, les quatre techniques de modélisation GMM, GMM-UBM, JFA et la nouvel technique I-Vecteur. En mettant l'accent sur la technique visée par notre travail à savoir I-Vecteur qui se base sur l'approche GMM-UBM.

Chapitre IV : Analyse et Conception

I. Introduction

Dans ce chapitre nous présentons la conception de notre système. Pour cela, nous avons adopté comme démarche de développement, le processus unifié UP et nous avons utilisé le langage de modélisation UML.

II. Présentation de la démarche utilisée

Pour le développement de notre système de RAL basé sur les techniques de modélisation de locuteur GMM-UBM et I-Vecteur, nous avons adopté le processus unifié UP. En effet, nous avons établi une analyse et conception suivie d'une étape d'implémentation pour une des deux techniques GMM-UBM / I-Vecteur, pour refaire la même démarche pour la mise en œuvre de l'autre technique de modélisation, et aboutir enfin à réaliser tous les besoins auquel notre système est censé répondre.

Dans ce qui suit, nous présentons une brève description du processus en cascade.

II.1 Le Processus en Cascade

- Problèmes [37]
 - Il est difficile de séparer les étapes
 - On peut l'utiliser quand les besoins sont bien définis et ils sont stables.
- Avantages [37]
 - Bien documenté à chaque phase
- Désavantages [37]
 - Rigide (on ne peut pas de répondre aux besoins nouveaux ou modifiés des clients)

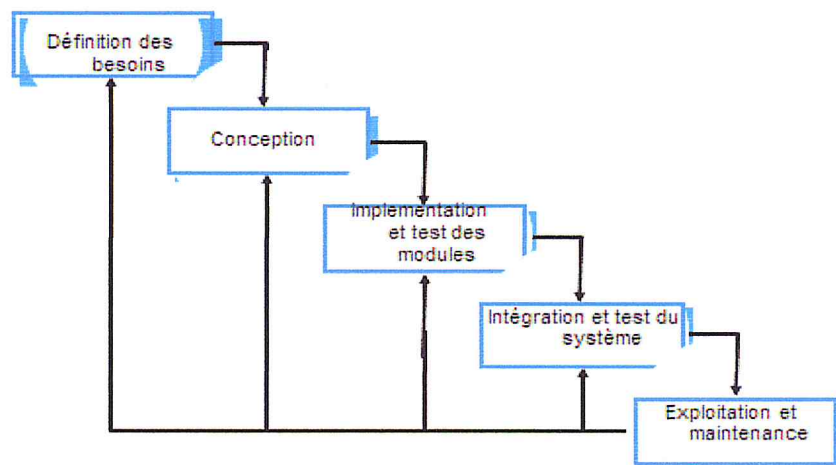


Figure IV.1: Cycle de vie de Processus en cascade [37].

III. Le langage UML

UML ou Unified Modeling Language, est un langage de modélisation qui est né au milieu des années 90 Défini par l'OMG (Object Management Group). Définit un méta-modèle et des types de diagrammes [38]. Il est hérite principalement des méthodes objets de Booch (Booch), OMT (Rumbaugh) et OOSE (Jacobson), Mais intègre également d'autres approches, comme les machines à états de Harel [39].

UML comporte treize types de diagrammes représentant autant de vues distinctes pour représenter des concepts particuliers du système d'information. Ils se répartissent en quatre grands groupes [39]:

- Diagrammes fonctionnels.
- Diagrammes statiques.
- Diagrammes dynamiques.
- Diagrammes d'implémentation.

Dans notre cas nous utilisons les :

- ✓ Diagramme de cas d'utilisation. (fonctionnel)
- ✓ Diagramme de classe. (statique)
- ✓ Diagramme de séquence. (interaction)

IV. Expression des besoins

IV.1 Identification des acteurs

Un Acteur représente l'abstraction d'un rôle joué par des entités externes (utilisateur, dispositif matériel ou autre système), qui interagissent directement avec le système étudié. Le tableau suivant présente l'utilisateur de notre système :

Acteurs	Description
Utilisateur	C'est l'acteur qui peut prendre des décisions à partir de résultats générés par le système, et qui choisit le type de modélisation et fait un test d'identification ou de vérification.

Tableau IV.1 : Tableau descriptif d'acteur.

IV. 2 Identification des cas d'utilisation

Les cas d'utilisation font apparaître les besoins fonctionnels et leur ensemble constitue le modèle des cas d'utilisation qui décrit les fonctionnalités complètes du système.

IV.3 Diagrammes des cas d'utilisation

IV.3 .1 Diagramme des cas d'utilisation global

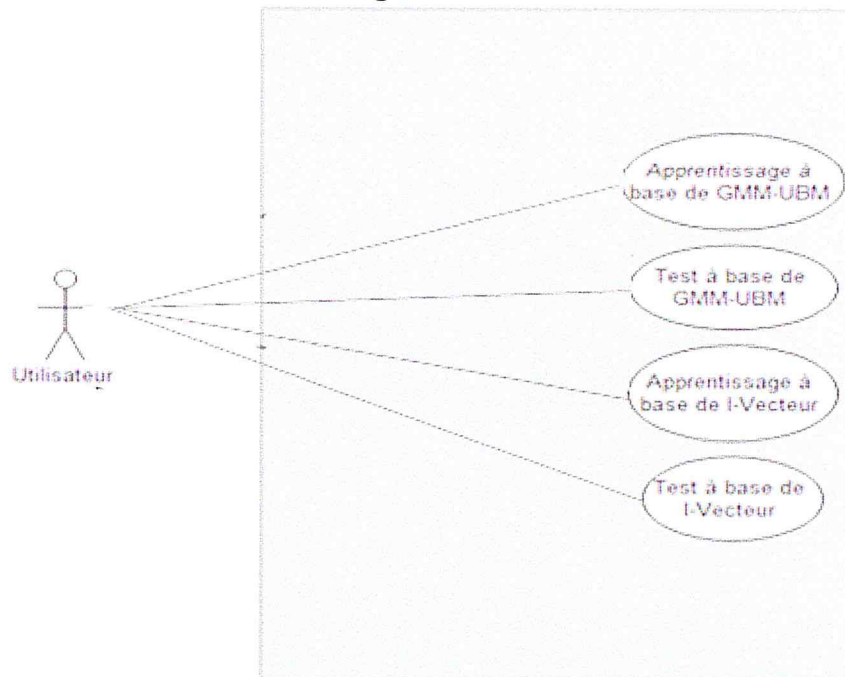


Figure IV.2 : diagramme de cas d'utilisation global.

IV .3.2 Description du diagramme de cas d'utilisation global

Reconnaissance de locuteur à base de GMM-UBM	
Acteur	Utilisateur
But	Modélisation de locuteur à base de GMM-UBM afin de créer des modèles de locuteurs et reconnaissances de locuteur à base de GMM-UBM.
Description	<ul style="list-style-type: none"> - choisir le type de modélisation GMM-UBM. - créer des modèles de locuteur client (Train) et pour les locuteurs test.
Elément déclencheur	- L'arrivée d'un nouveaux Locuteur client (Train) ou d'une identité proclamée.
Reconnaissance de locuteur à base I-Vecteur	
Acteur	utilisateur

But	Choix le type de Modélisation I-Vecteur pour créer des modèles des locuteurs et pour le test
Description	<ul style="list-style-type: none"> - choisir le type de modélisation I-Vecteur. - créer des modèles de locuteur client (Train) et pour les locuteurs test.
Elément déclencheur	- L'arrivée d'un nouveaux client ou d'une identité proclamée

Table IV.2 : Table descriptive pour la Reconnaissance de locuteur à bas GMM-UBM et à bas I-Vecteur.

IV.3.3 Diagrammes des cas d'utilisation détaillés

IV.3.3.1 Diagramme de cas d'utilisation d'apprentissage à base de GMM-UBM

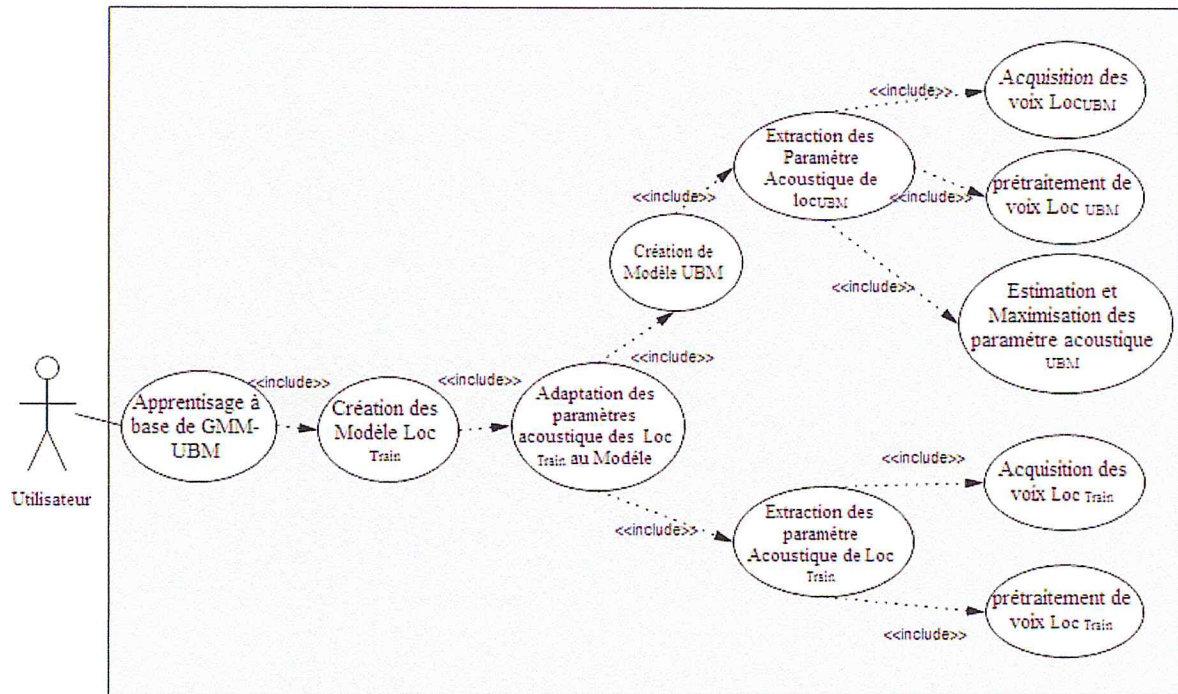


Figure IV.3 : diagramme de cas d'utilisation d'apprentissage à base de GMM-UBM

IV.3.3.1 Description de diagramme de cas d'utilisation d'apprentissage à base de GMM-UBM

cas d'utilisation d'apprentissage à base de GMM-UBM	
Acteur	Administrateur de société.

But	Création des Modèles de Locuteurs Train (Client).
Description	<ul style="list-style-type: none"> - Enregistrement de voix d'un locuteur train. - prétraitements. - l'extraction des paramètres acoustique des locuteurs train. - Adaptation des paramètres acoustique des locuteurs train au modèle UBM.
Pré-condition	- L'arrivé d'un nouveaux locuteur train.
Pos-condition	- Création d'un modèle de locuteur _{train}

Table IV.3 : Table descriptive pour l'apprentissage à bas de GMM-UBM.

IV.3.3.3 Diagramme de cas d'utilisation de test à base GMM-UBM

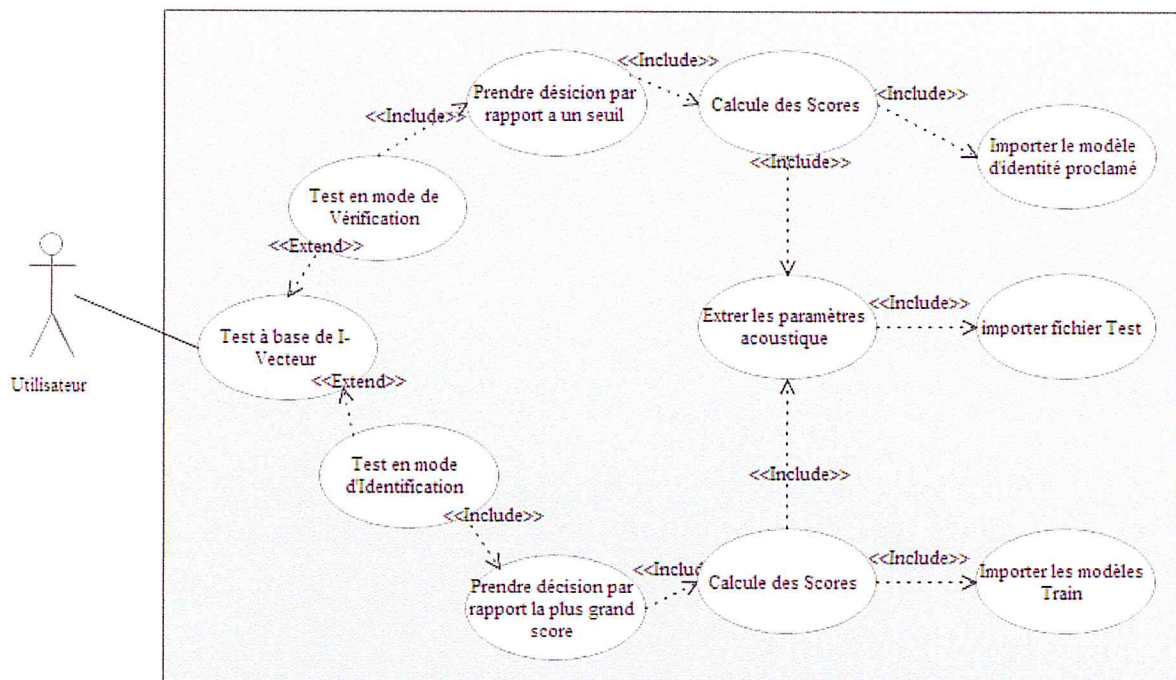


Figure IV.4 : Diagramme de cas d'utilisation de test à bas GMM-UBM.

IV.3.3.4 Description de diagramme de cas d'utilisation de test à base de GMM-UBM

Test de reconnaissance de locuteur à base de GMM-UBM	
Acteur	L'administrateur
But	Etablir un test de vérification ou d'identification à base de GMM-UBM.
Description	<ul style="list-style-type: none"> - Enregistrement de la voix d'un locuteur test. - Les prétraitements. - L'extraction des paramètres acoustique du locuteur test. - Comparé par rapport aux modèles des locuteurs clients.
Pré-condition	- L'arrivé d'un nouveaux locuteur proclamé.
Post-condition	<ul style="list-style-type: none"> ✓ En mode vérification {loc_{test} =Client ou imposteur} ✓ En mode d'identification {identifie un loc_{test}}

Table IV.4 : Table descriptive pour le Test à bas de GMM-UBM.

IV.3.3.5 Diagramme de cas d'utilisation d'apprentissage à base I-Vecteur

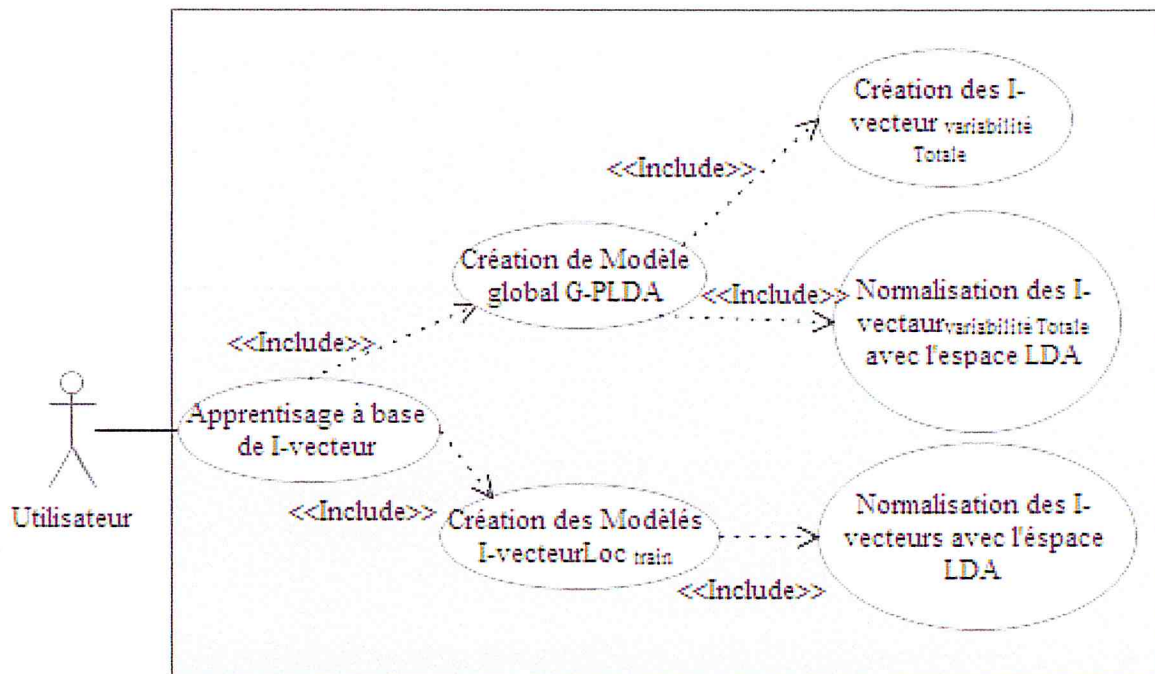


Figure IV.5 : Diagramme de cas d'utilisation d'apprentissage à base I-Vecteur.

IV.3.3.6 Diagramme de cas d'utilisation de création de modèle de modèle global (GPLDA) à base I-Vecteur.

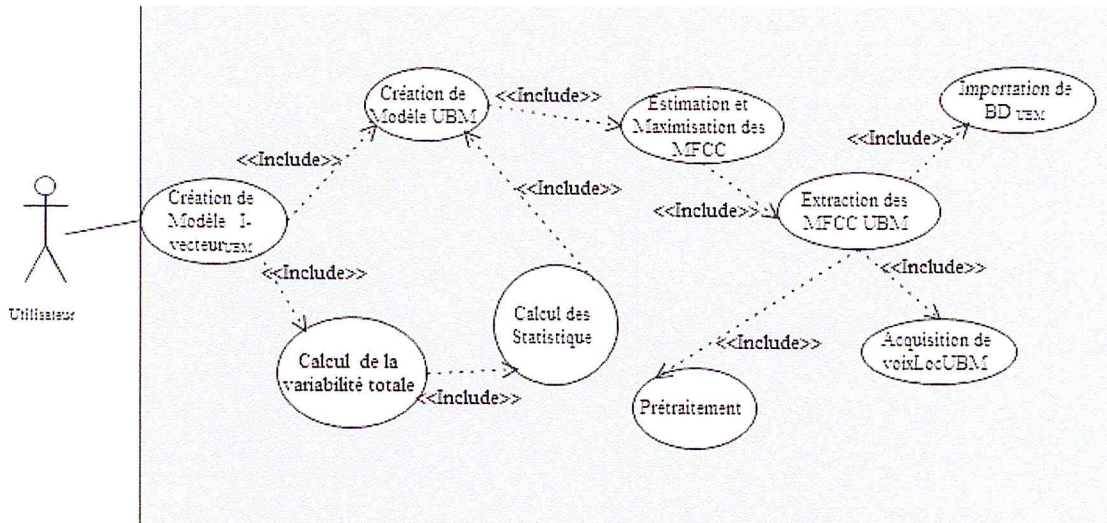


Figure IV.6 : Diagramme de cas d'utilisation de création de modèle global à base I-Vecteur.

IV.3.3.7 Diagramme de cas d'utilisation de création des modèles train à base I-Vecteur

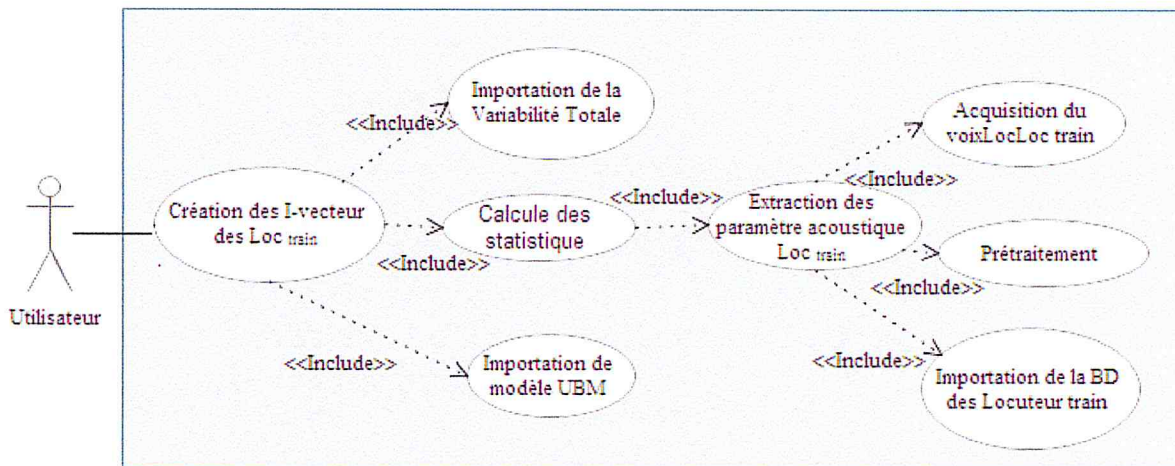


Figure IV.7 : Diagramme de cas d'utilisation de création des modèles train à base I-Vecteur.

IV.3.3.8 Description de Diagramme de cas d'utilisation de création de modèle train à base I-Vecteur

création de modèle des locuteur _{train} à base I-vecteur.	
Acteur	Utilisateur
But	Création des modèles des locuteurs train.
Description	<ul style="list-style-type: none"> - Acquisition de la voix des Locuteurs train. - Extraction des MFCC des locuteurs train. - Création les I-vecteur pour les Locuteurs train. - Calcule les stat d'ordre un et zéro des locuteurs train. - Création les I-vecteur pour les Locuteurs train. - Projection des I-vecteur des locuteurs train dans l'espace LDA (Normalisation).
Pré-condition	✓ L'arrivé d'une nouveaux Locuteur train.
Post-condition	✓ Enregistrement de modèle des locuteurs train

Table IV.5: Table descriptive pour la création des modèles des locuteurs train à base I-Vecteur.

IV.3.3.9 Diagramme de cas d'utilisation de test à base I-Vecteur

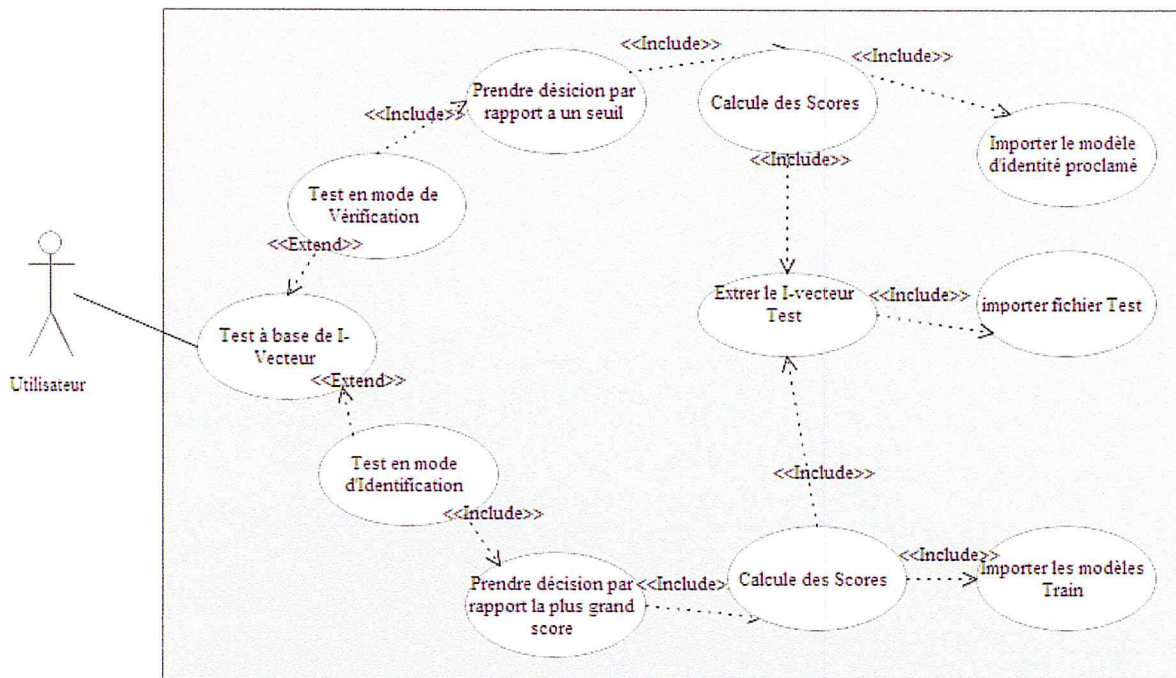


Figure IV.8 : Diagramme de cas d'utilisation de test à base I-Vecteur.

IV.3.3.10 Description de diagramme de cas d'utilisation de test à base I-Vecteur

test à base I-Vecteur	
Acteur	Utilisateur
But	Le Test de Mode Identification ou Vérification.
Test en mode d'identification	
Description	<ul style="list-style-type: none"> ✓ Création des I-Vecteur pour les locuteurs test : <ul style="list-style-type: none"> - Acquisition de la voix des Locuteurs test. - Extraction des MFCC des locuteurs test. - Création les I-Vecteur pour les Locuteurs test. - Projection des I-Vecteur des locuteurs test dans l'espace LDA(Normalisation). ✓ Calcul des Score, et prend la désicion
Pré-condition	- L'arrivé d'un nouveaux locuteur test.
Post-Condition	- Résultat de test (accepte ou refusé le locuteur test)
Test en mode de vérification	
Description	<ul style="list-style-type: none"> - Importation des modèles des locuteurs Proclamé. ✓ Création des I-Vecteur pour les locuteurs test : <ul style="list-style-type: none"> - Acquisition de la voix des Locuteurs test. - Extraction des MFCC des locuteurs test. - Création les I-Vecteur pour les Locuteurs test. - Projection des I-Vecteur des locuteurs test dans l'espace LDA(Normalisation). ✓ Calcul des Score, et prend une décision.
Pré-condition	<ul style="list-style-type: none"> - Disponibilité d'un enregistrement vocal d'un locuteur Proclamé. - et l'arrivé d'un locuteur test.
Post-Condition	- Résultat de test (locuteur test= locuteur proclamé 'Client' ou Non 'Imposteur').

Table IV.6 : Table descriptive pour le test à bas I-Vecteur.

IV.4 Méta modèles

IV.4.1 Méta-modèle relatif aux modèle GMM-UBM :

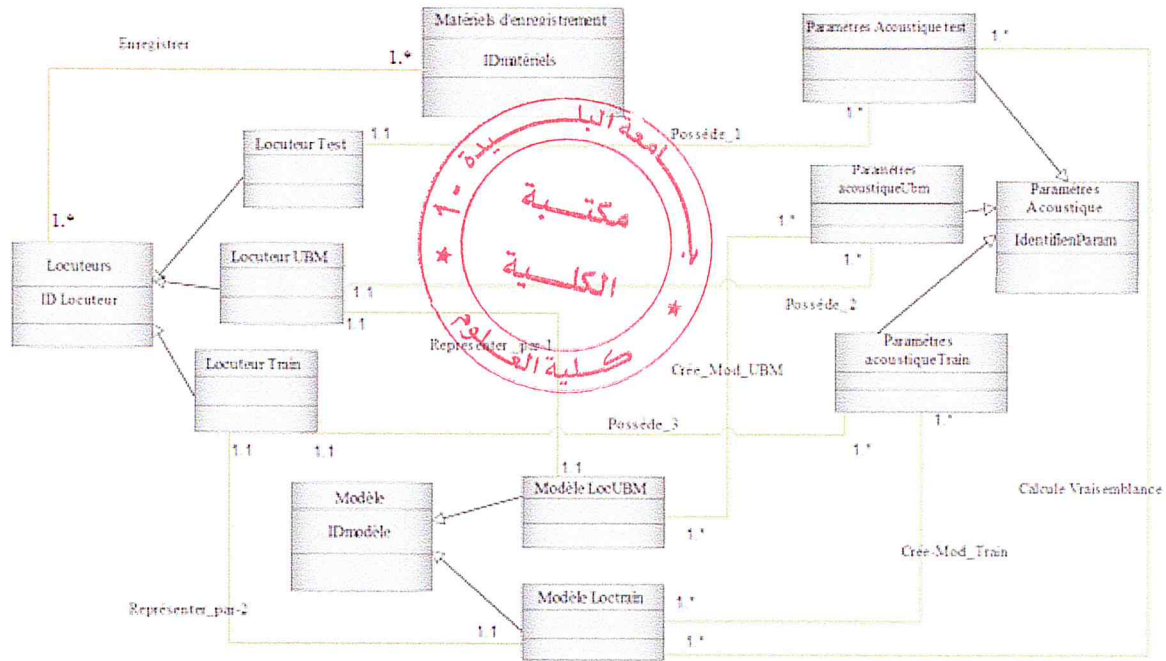


Figure IV.9: Méta-modèle relatif aux modèle GMM-UBM

IV.4.2 Dictionnaire de données relatif aux GMM-UBM

Clase	ATTRIBUTS	TYPE	Identifiant	DESCRIPTION
Locuteur	ID Locuteur	VARCHAR	Oui	Identifiant de Locuteur
	NomLocuteur	VARCHAR	Non	Le nom de Locuteur
	Prénomlocuteur	VARCHAR	Non	Le prénomme de Locuteur
Modèle	ID modèle	VARCHAR	Oui	Identifiant de Modèle
	Moyenne	VARCHAR	Non	La Moyenne que représente le modèle
	Covariance	VARCHAR	Non	La covariance que représente le modèle
	Le poids	VARCHAR	Non	Le poids que représente le modèle
Paramètres	IdentifienParam	VARCHAR	Oui	Identifiant des paramètres acoustique
	Nbrcoefficient	VARCHAR	Non	Nombre des coefficients extraire pour un Locuteur
	Energie	VARCHAR	Non	Paramètres Energie

Acoustique	PremierDérivé	VARCHAR	Non	Le Premier dérivé des paramètre acoustique pour avoir des informations temporelles (vitesse)
	DeuxièmDérive	VARCHAR	Non	Le deuxième dérivé des MFCC pour avoir des informations temporelles (accélération)

Table IV.7 : Dictionnaire de données relatif aux GMM-UBM

IV.4.3 Liste des méthodes relatives aux GMM-UBM

CLASSES	METHODE	DESCRIPTION
Locuteur	Ajouter locuteur ()	Ajouter une voix d'un locuteur
	supprimer ()	Supprimer une voix d'un locuteur
Matériels d'enregistrement	Ajouter ()	Ajouter un nouveau matériel d'enregistrement de la voix.
	Supprimer ()	Supprimer un type de matériel d'enregistrement
Paramètres acoustique	Extraction ()	Extraction des paramètres MFCC
Modèles	Crée ()	Crée des modèles
	Supprimer ()	Supprimer un Modèle

Table IV.8 : Liste de méthodes relatives aux GMM-UBM

Paramètres acoustique	PremierDérivé	VARCHAR	Non	Le Premier dérivé des acoustiques pour avoir des informations temporelles (vites)
	DeuxièmDérive	VARCHAR	Non	Le deuxième dérivé des MFCC pour avoir des informations temporelles (Accélération)
Totale variabilité	<u>ID_T</u>	VARCHAR	Oui	L'identifiant de la totale variabilité
	tv_dim	VARCHAR	Non	La dimension Proposé pour l'espace de totale variabilité
	niter	VARCHAR	Non	Le nombre des itérations
I-Vecteur	<u>Id-vecteur</u>	VARCHAR	Oui	Identifiant d' I-Vecteur
Stat	Id-stat	VARCHAR	Oui	Identifiant des States
	Ordre	VARCHAR	Non	Ordre des states (zéro ou un)

Table IV.9 : Dictionnaire de données relatif aux I-Vecteur.

IV.4.5 Liste des méthodes relatives aux I-Vecteur

CLASSES	METHODE	DESCRIPTION
Locuteur	Ajouter locuteur ()	Ajouter une voix d'un locuteur
	supprimer ()	Supprimer une voix d'un locuteur
Matériels d'enregistrement	Ajouter ()	Ajouter un nouveau type de matériel d'enregistrement.
	Supprimer ()	Supprimer par un type de matériel d'enregistrement
Paramètres acoustique	Extraire ()	Extraction des paramètres acoustique
Modèle UBM	Crée ()	Crée des modèles
	Supprimer ()	Supprimer
Stats	Calcul ()	Calcul des statistique d'ordre un et d'ordre zéro
variabilité Totale	Calcul ()	Calculer l'espace de la variabilité total

Remarque : pour l'Identification à base de I-Vecteur se port dans le même sens, l'extraction des paramètres acoustique jusqu'à la projection dans l'espace LDA en fin la comparaison avec tous les modèles existe dans la base de données et on lui associe une identité la plus proche.

IV.5.4 Diagramme de séquence d'apprentissage de modèle globale à base I-Vecteur

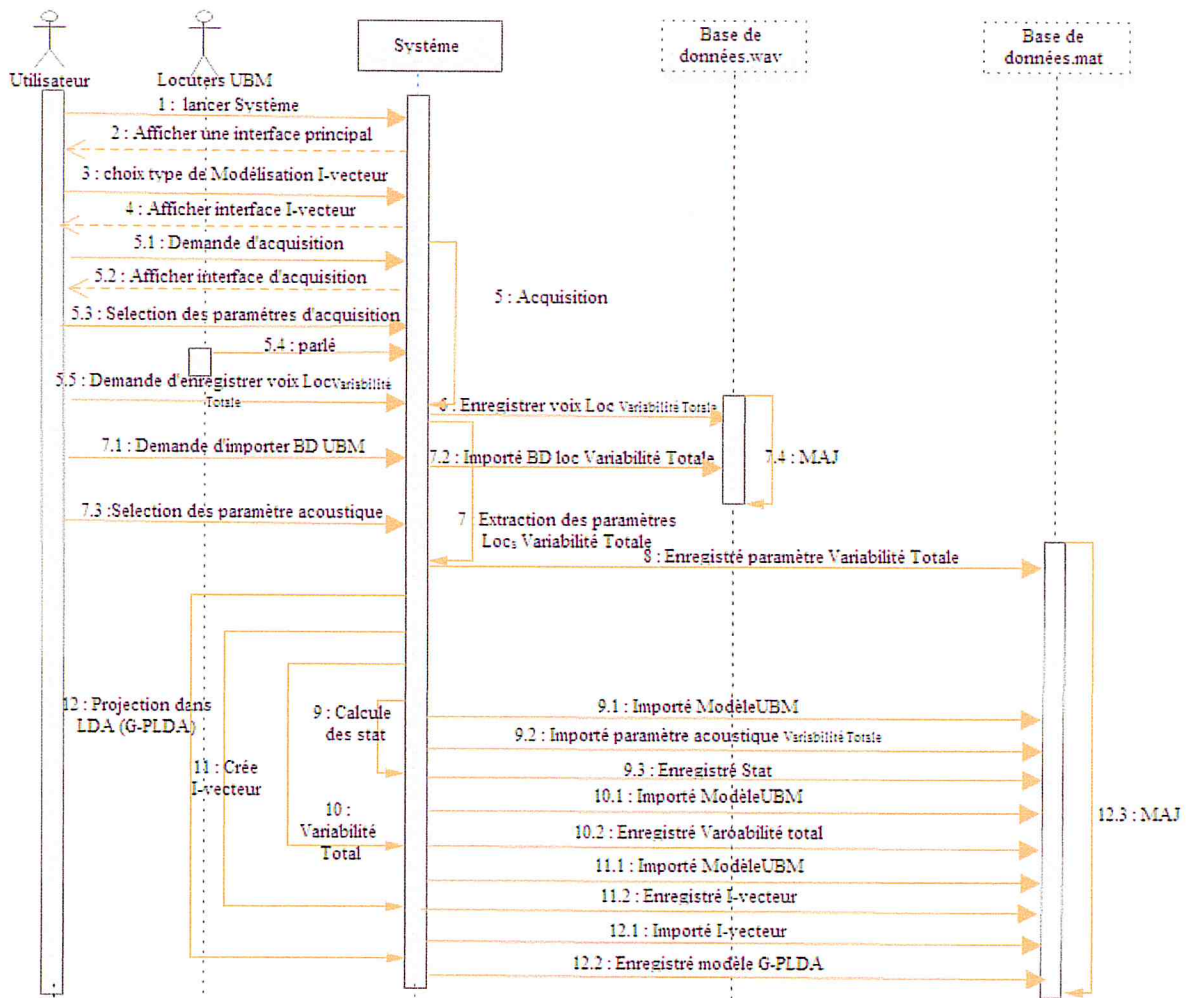


Figure IV.14 : Diagramme de séquence de création de modèle globale (G-PLDA) à base I-Vecteur

Remarque : après l'affichage de l'interface principale on peut choisi la modélisation avec concaténation des paramètres ou bien sans concaténation, cette choix n'as pas une influence sur les diagrammes de conception.

IV.5.5 Diagramme de séquence d'utilisation d'apprentissage des locuteurs Train à base I-Vecteur

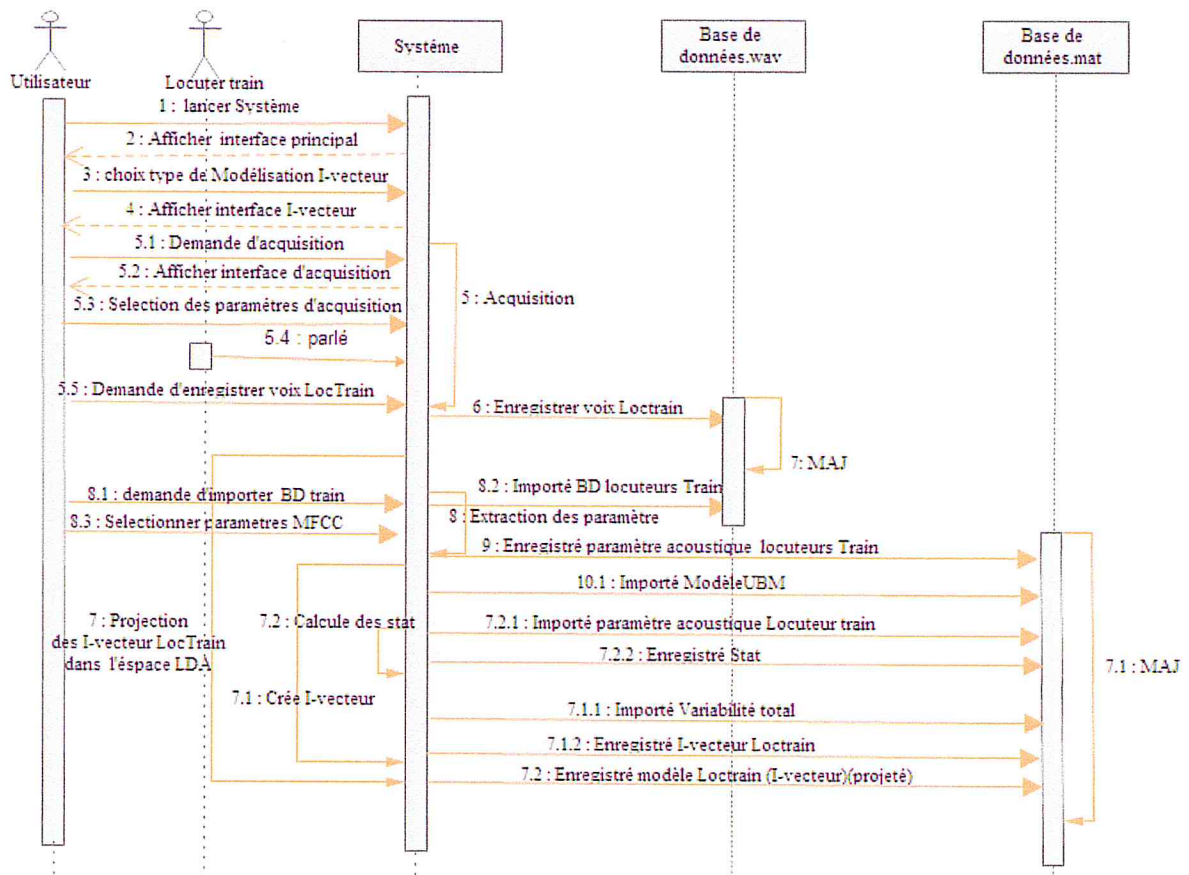


Figure IV.15 : Diagramme de séquence de création des Modèle de Client à bas I-Vecteur.

IV.5.6 Diagramme de séquence de Test en mode de vérification à base I-Vecteur.

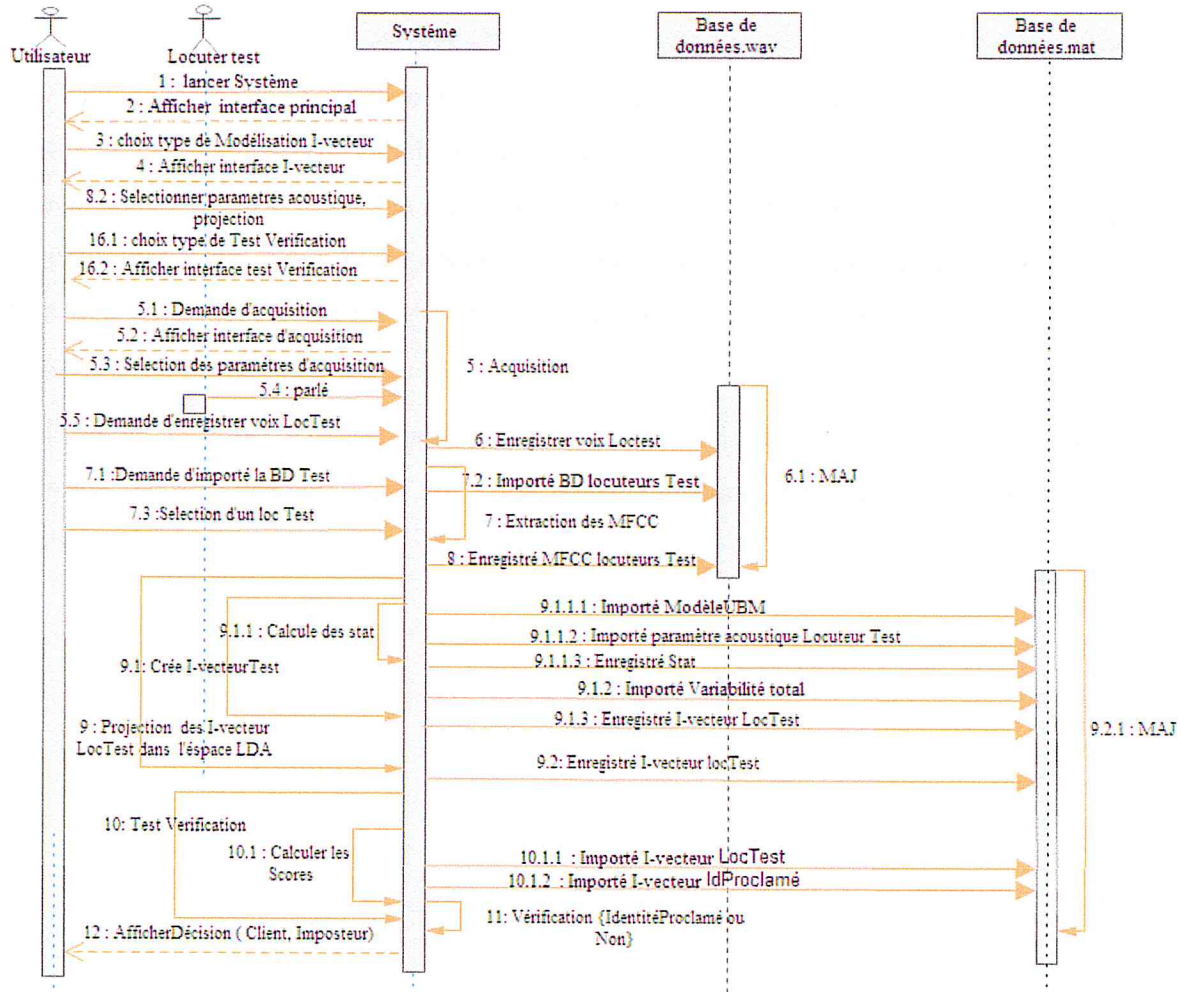


Figure IV.16 : Diagramme de séquence de Test en mode de vérification à base I-Vecteur

Remarque : la Vérification ce fait de même principe avec la technique GMM-UBM : l'extraction des paramètres acoustique de Locuteur Test ; importé le modèle d'identité proclamé et en fin calcule les score et affichez la décision au utilisateur.

Chapitre V : Test et Implémentation

I. Introduction

Dans ce chapitre, nous allons étudier l'effet de la variation du canal sur les performances du système de vérification du locuteur. Nous présentons l'implémentation de notre système de RAL qui est basé sur deux approches de modélisation à savoir le GMMUBM et I-Vecteur. Pour tester les performances de notre système, nous avons utilisé une base de données algérienne composées de 39 locuteurs.

II. Description de la base de donnée utilisée

La base de données utilisée lors du test contient des dialectes de différentes régions algériennes. Elle a été construite par le Centre du Développement des Technologies Avancées CDTA en coopération avec la société espagnole Agnitio ([http : www.agnitio.es](http://www.agnitio.es)), qui est un leader mondial dans la reconnaissance de la voix et ses applications. La base de données contient 39 locuteurs choisis à partir de l'ensemble des personnels du CDTA. Le critère de choix de ces locuteurs vise à couvrir les différents dialectes qui existent en Algérie. Les conditions d'enregistrement sont très proches de la réalité, et les enregistrements vocaux n'ont été soumis à aucun traitement. Pour chaque locuteur, trois enregistrements vocaux ont été créés. Un enregistrement en utilisant le microphone, un autre en utilisant le réseau téléphonique fixe, et un dernier en utilisant le réseau mobile GSM. Chacun d'eux contient environ 3 minutes de paroles spontanées. A partir de ces enregistrements, nous avons créé deux sous-ensembles de fichiers vocaux. Pour chaque type d'enregistrement (Microphone, Mobile et Fixe), nous avons construit un ensemble d'apprentissage et un ensemble de test.

III. L'architecture du Logiciel

Le logiciel que nous avons développé a une architecture Desktop, exécuté sur un seul Ordinateur, utilisé par un seul utilisateur à la fois.

IV. Prétraitement et extraction des paramètres

IV.1 Prétraitement

Pour améliorer la qualité des signaux de parole et garder que les informations utiles à la reconnaissance, nous avons utilisé la méthode d'élimination de silence [31].

IV.2 Extraction des paramètres MFCC

Dans nos expériences une analyse est appliquée toutes les 10ms sur des fenêtres d'analyse de 20ms (par glissement et recouvrement des fenêtres d'analyse). A chaque trame, on associe un vecteur de représentation acoustique. 19 MFCC sont calculés pour chaque trame. Des coefficients différentiels du premier ordre et du second ordre plus le log énergie sont ensuite calculés pour former un vecteur de dimension 60.

V. Protocole expérimental

V.1 Système de vérification de locuteur

A partir des enregistrements, nous avons créé deux sous-ensembles de fichiers vocaux :

- La base d'apprentissage : elle est composée de 117 fichiers vocaux prononcés par 39 locuteurs (Hommes) avec 3 types d'enregistrement (Microphone, Mobile et Fixe). La durée de chaque enregistrement est de 150 secondes environ.
- La base de test : elle est composée de 117 fichiers vocaux prononcés par 39 locuteurs (Hommes) avec 3 types d'enregistrement (Microphone, Mobile et Fixe) de durée de 2 secondes environ.

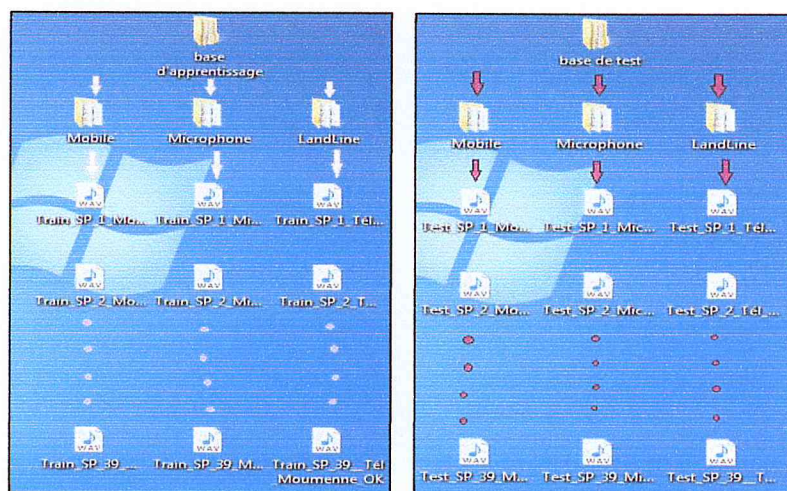


Figure V.1 : Base de données d'apprentissage et base de données de test.

L'évaluation de la robustesse de notre système est exprimée en terme de taux d'erreurs égales (EER). Le EER est le point de fonctionnement qui correspond au seuil pour lequel les deux taux d'erreurs, le taux de fausses acceptations (False Acceptance Rate FAR) et le taux de faux rejets (False Rejection Rate FRR) sont égaux. De plus nous avons tracé la courbe DET

(Detection Error Tradeoff) qui décrit le taux de fausses acceptations en fonction du taux de faux rejets.

V.2 Vérification du locuteur à base de GMMUBM

Le model du monde est estimé sur 3 sessions (Fixe, Microphone et Mobile) de 39 locuteurs. Les épreuves (trials) de vérification sont constitués de toutes les combinaisons possibles modèle-test, pour un total de 1521 essais (39 essaies client test par rapport à 1482 essais imposteurs).

- *Expérience 1 : Influence du nombre de GMM*

Cette première expérience a pour but de choisir l'ordre du modèle GMM nécessaires à la mise en œuvre du notre systèmes. Les EERs du notre système de vérification sont représentés par l'histogramme de la figure V.2.

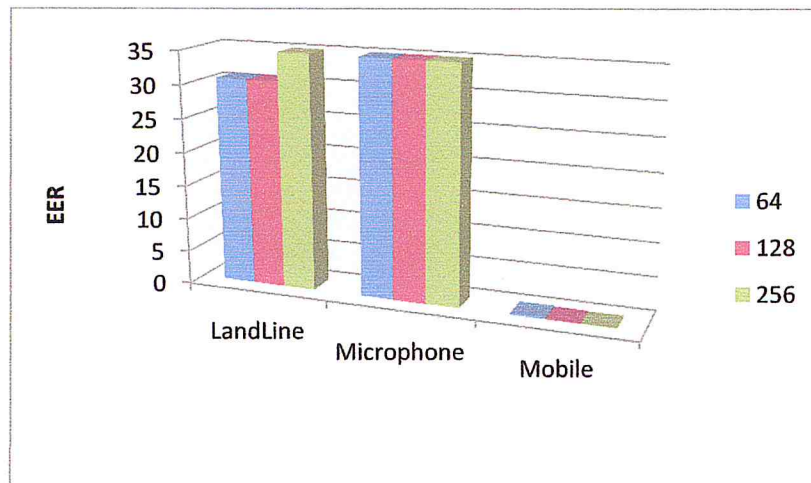


Figure V.2 : Les EER de système de vérification à base de GMM-UBM.

Le choix de l'ordre du modèle GMM dépend de la quantité de données et la durée de l'enregistrement vocal utilisée pour l'apprentissage et le test. Choisir un ordre trop peu va nuire à la précision du modèle. Choisir trop de composante engendra une charge de calcul plus importante. En plus, les paramètres acoustiques extraits à partir d'un enregistrement vocal, qui contient l'intervention de plusieurs locuteurs seront dispersés dans plusieurs régions de l'espace de vecteurs acoustiques ; ici il est nécessaire d'augmenter l'ordre du modèle GMM pour couvrir ces différentes régions.

Dans notre cas, on dispose de 39 locuteurs et la durée de l'enregistrement vocal lors la phase d'apprentissage est 150 seconde environ et pour la phase de test est 10 seconde, et

d'après les résultats trouvés (voir l'histogramme de la figure V.2), le choix d'un modèle avec 128 gaussiennes suffit pour représenter convenablement le locuteur.

- *Expérience 2 : Influence du canal*

Dans cette deuxième expérience, nous avons étudié l'effet de la variation du canal entre les données de test et les données d'apprentissage sur la robustesse des systèmes de vérification du locuteur. Les taux d'erreurs égales (Equal Error Rate EER) sont présentés dans le tableau V.1.

Tableau V.1 : L'EER du système de vérification du locuteur dans les quatre cas de tests : Microphone vs à (Fixe, microphone et Mobile), Microphone vs à (Microphone, Fixe et Mob), Mobile vs à (Microphone, Fixe et Mobil) et 3Sessions vs à (Microphone, Fixe et Mobil).

Test \ Client	Fixe	Microphone	Mobile
Fixe	7.48	48.71	30.76
Microphone	48.71	17.94	35.89
Mobile	38.79	43.85	0.13
3 Sessions	10.25	17.88	2.56

A partir du tableau V.I, nous observons que le système basé sur la méthode GMMUBM donne un bon résultat dans le cas où la séquence de test et la séquence d'apprentissage ont le même canal (Mobile vs Mobile). Cependant, ces performances dégradent significativement lorsque les deux séquences sont issues de deux canaux différents (Mobile vs Fixe et Mobile vs Microphone) qu'on appelle le Mismatch. Le EER dégrade de 0.13% pour le cas Mobile vs Mobile à 38.79% pour le cas Mobile vs Fixe et à 43.85% pour le cas Mobile vs Microphone.

Nous remarquons aussi que le type du canal utilisé influe sur la robustesse du notre système par exemple l'utilisation de Mobile vs Mobile donne de meilleurs résultats par rapport Microphone vs Microphone et Fixe vs Fixe.

V.3 Vérification du locuteur à base de I-Vecteur

La matrice T de variabilité totale utilisée a été apprise à partir d'énoncés composé de la concaténation de 39 locuteurs avec 3 session pour chacun (ou 3 types d'enregistrement) à savoir Mobile, Microphone et Fixe. Les I-Vecteurs issus de ces énoncés sont utilisés comme apprentissage pour la technique de compensation LDA et la PLDA.

- *Expérience 1 : Influence du l'ordre de La matrice de variabilité totale T*

L'analyse de facteur joue dans cette technique de modélisation, le rôle d'un extracteur de paramètres. Une série d'expérimentations a été faite pour évaluer l'influence du l'ordre T sur les performances du notre système de vérification du locuteur. Au premier lieu nous avons fixé l'ordre de LDA à 38. L'histogramme de la figure V.3 présente les résultats trouvés.

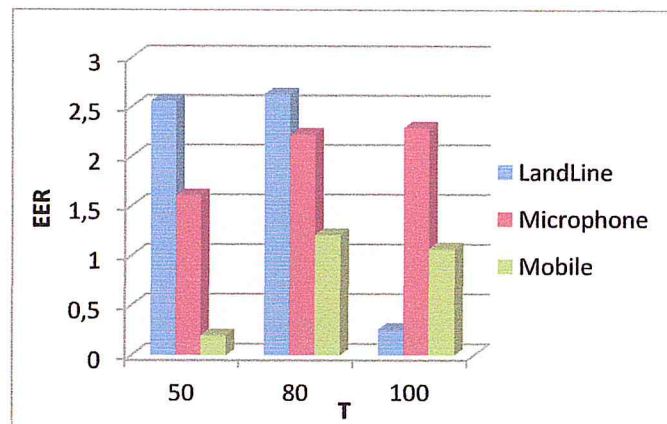


Figure V.3 : Les résultats de test (EER) selon l'espace de variabilité totale(T) de système de vérification à base d'I-Vecteur.

En regardant les résultats, on constate que le système avec $T=50$ donne des meilleurs résultats. Par la suite nous fixons T à 50.

- *Expérience 2 : La compensation de canal et l'influence de l'ordre LDA*

Dans l'approche I-Vecteur originale la compensation du canal est faite par la méthode LDA. La compensation de canal ne se fait plus dans l'espace des supervecteurs GMM, mais plutôt dans l'espace (à faible dimension) des facteurs de la variabilité totale. Dans cette expérience nous allons effectuer une analyse discriminante linéaire (Linear Discriminant Analysis) LDA pour définir de nouveaux axes minimisant la variance intra-classe causée par les effets du canal, et maximisant par la même occasion la variance entre locuteurs. Nous

allons aussi voir l'influence de l'ordre LDA. Les résultats sont représentés dans la figure V.4 ci-dessous.

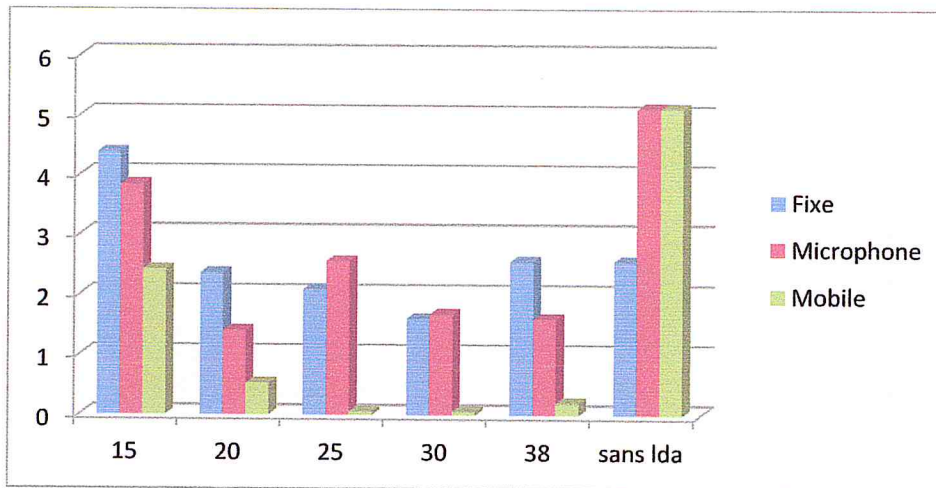


Figure V.4 : test d'influence de l'ordre LDA.

Comme prévu, les techniques de compensation de la variabilité inter-sessions permettent d'améliorer considérablement les performances des systèmes, le EER diminue de 5.12% sans LDA jusqu'à **0.06% avec LDA (cas mobile/mobile)**. Notons aussi qu'un ordre LDA de 30 donne des meilleures performances. Nous allons fixer par la suite l'ordre de LDA à 30.

V.4 Comparaison entre les modélisations à base GMM-UBM et I-Vecteur

Tableau V.2 : table comparative des valeurs de test du système de vérification données par les deux méthodes de modélisation GMM-UBM et I-Vecteur.

Test Client	Fixe		Microphone		Mobile	
	<i>GMMUBM</i>	<i>I-Vecteur</i>	<i>GMMUBM</i>	<i>I-Vecteur</i>	<i>GMMUBM</i>	<i>I-Vecteur</i>
Fixe	7.48	0,13	48.71	2.02	30.76	0.4
Microphone	48.71	2,56	17.94	0,67	35.89	1,28
Mobile	38.79	1,61	43.85	1,68	0.13	0,06
3 Sessions	10.25	5.12	17.88	12.82	2.56	11.80

En comparant les valeurs EER des deux systèmes et d'après les courbes DET de la figure, nous remarquons que le système à base de I-Vecteur donne les meilleures performances dans tous les cas **Fixe vs Fixe**\ Microphone\ Mobile, **Microphone vs Fixe**\ Microphone\ Mobile et **Mobile vs Fixe**\ Microphone\ Mobile.

Courbe DET Mobile

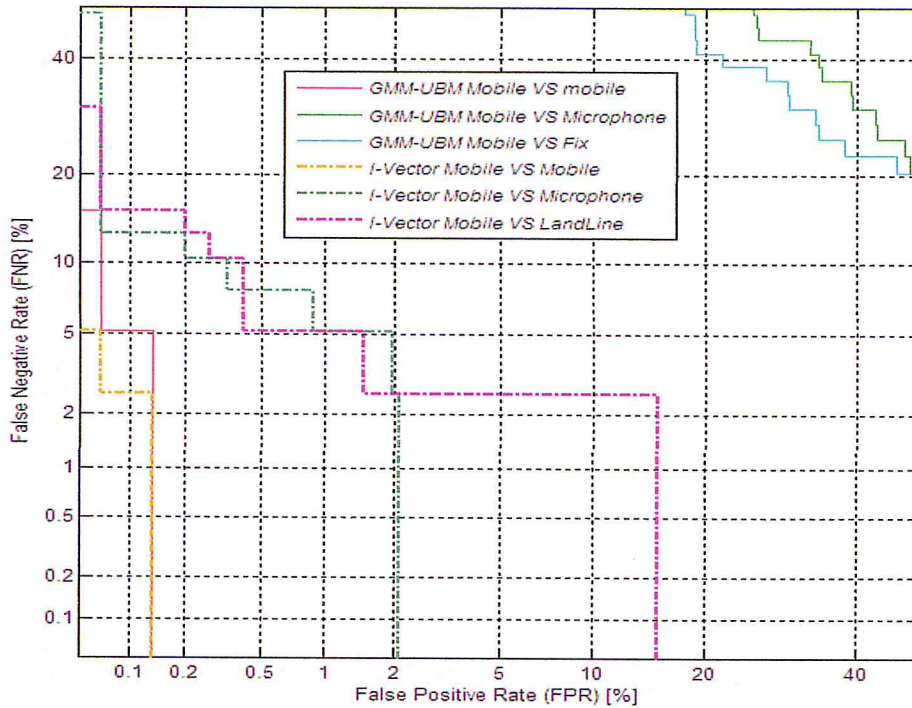


Figure V.5 Les Courbes DET de GMM-UBM VS I-Vecteur.

V.5 Nouvelle technique de concaténation de caractéristiques MFCC

Les I-Vecteurs qui correspondent aux résultats des évaluations affichés dans la dernière case du tableau (l'utilisation de 3 sessions/ou 3 types d'enregistrement) sont calculé en utilisant la formule suivante :

$$I_{\text{vecteur}}_{3 \text{ sessions}} = \frac{I_{\text{vecteur}}_{\text{LandLine}} + I_{\text{vecteur}}_{\text{Microphone}} + I_{\text{vecteur}}_{\text{Mobile}}}{3}$$

Par ailleurs, nous avons pensé de concaténer les coefficients MFCC des trois sessions et par la suite calculer un seul modèle I-Vecteur. L'algorithme détaillant cette méthode est décrit ci-dessous :

```

MFCC= [] ; matrice vide
Pour i=1 : n (n nombre de session, i.e n=3) Faire
    MFCCi ← extraire_coefficient(signal)
    MFCC ← [MFCC MFCCi]
Fin
Statistiques ← stats (MFCC, T, UBM)
I-Vecteur ← ivecteur (statistiques, T, UBM)
    
```

Les résultats obtenus de comparaisons entre les deux méthodes (modélisation avec I-Vecteur) sont mis sous forme des courbes DET et de tableau ci-dessous

Les résultats confirment que la technique de concaténation des paramètres MFCC permet d'améliorer les résultats.

Tableau V.3 : table comparative des valeurs de test entre les deux méthodes (la moyenne de I-Vecteurs, Concaténation) selon les trois sessions du système de vérification à base de I-Vecteur.

	Fixe	Microphone	Mobile
3 sessions (la moyenne de I-Vecteurs)	5.12	12.82	11.80
Concaténation (Nouvelle méthode)	0.67	2.56	0.67

VI. L'interface de l'application

VI.1 Environnement du travail

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

VI.2 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 pavillon g serie avec les caractéristiques suivantes :
- Processeur : intel Core i5.
- RAM 4GO.
- Disque Dur : 550 GO.
- OS : Microsoft Windows Seven.
- Outils de développement : Nous avons en recours lors de l'élaboration de notre système à Matlab 13 que nous présenterons ci-dessous.

❖ Matlab (R2013a)

Matlab et son environnement interactif est un langage de haut niveau qui permet l'exécution de tâches 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++.

VI.3 Présentation de l'interface de l'application

L'interface assure une communication aisée entre l'utilisateur et l'application, elle facilite à l'utilisateur la configuration du système en lui proposant une liste de choix préétablie. Elle facilite la lecture et la compréhension des résultats en optimisent la manière dont ils sont présentés par le système.

Notre interface, destinée aux utilisateurs, est simple et permet d'illustrer les principaux processus du système de vérification de locuteur, ainsi que d'autres fonctions supplémentaires telles que la gestion de l'enregistrement, la paramétrisation du signal vocal, Normalisation... etc.

VI.3.1 La fenêtre principale de l'application

La fenêtre principale de l'application (figure V.1) est celle qui s'affiche en premier lieu lors du lancement du programme. Elle permet l'accès aux différentes fonctionnalités du programme, présentées sous forme de composants visuels (boutons, menus,...). Dans cette fenêtre ce trouve deux types de modélisation : modélisation par GMM-UBM et modélisation par I-vecteur avec le choix de Concaténation des paramètres MFCC pour la création des modèle des clients pour les deux types de modélisation. Cette interface donne le droit à l'utilisateur pour choisit le type de modélisation.

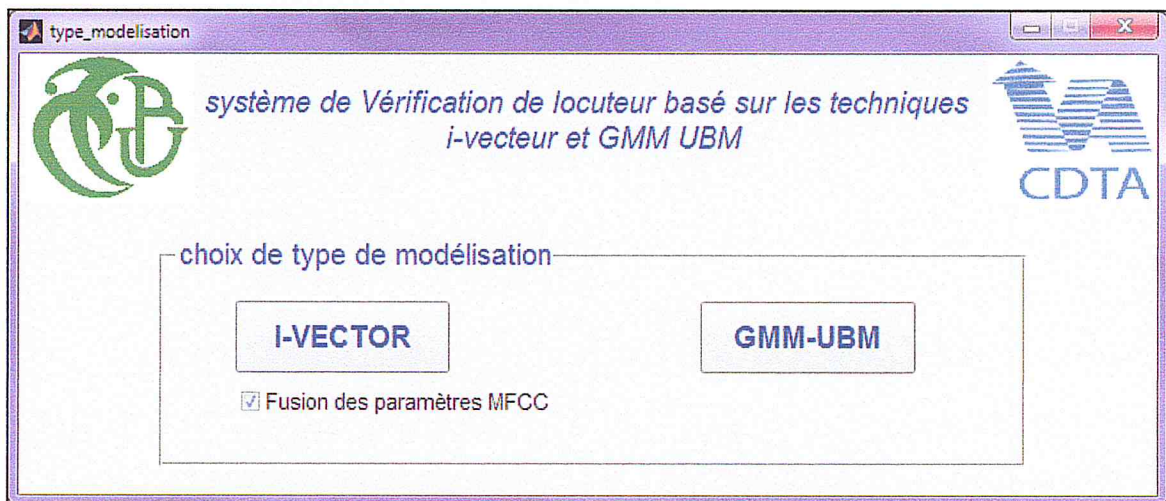


Figure V.6 : la fenêtre principale de l'application.

VI.3.2 Modélisation par GMM-UBM

Après avoir choisir la Modélisation par GMM-UBM, une fenêtre apparaîtra (voir figure 2) qui nous permettra de calculer le taux d'erreur EER du notre système ainsi de vérifier l'identité proclamée. Nous allons décrire dans ce qui suit les différents événements qui s'y passent lors de ce processus.

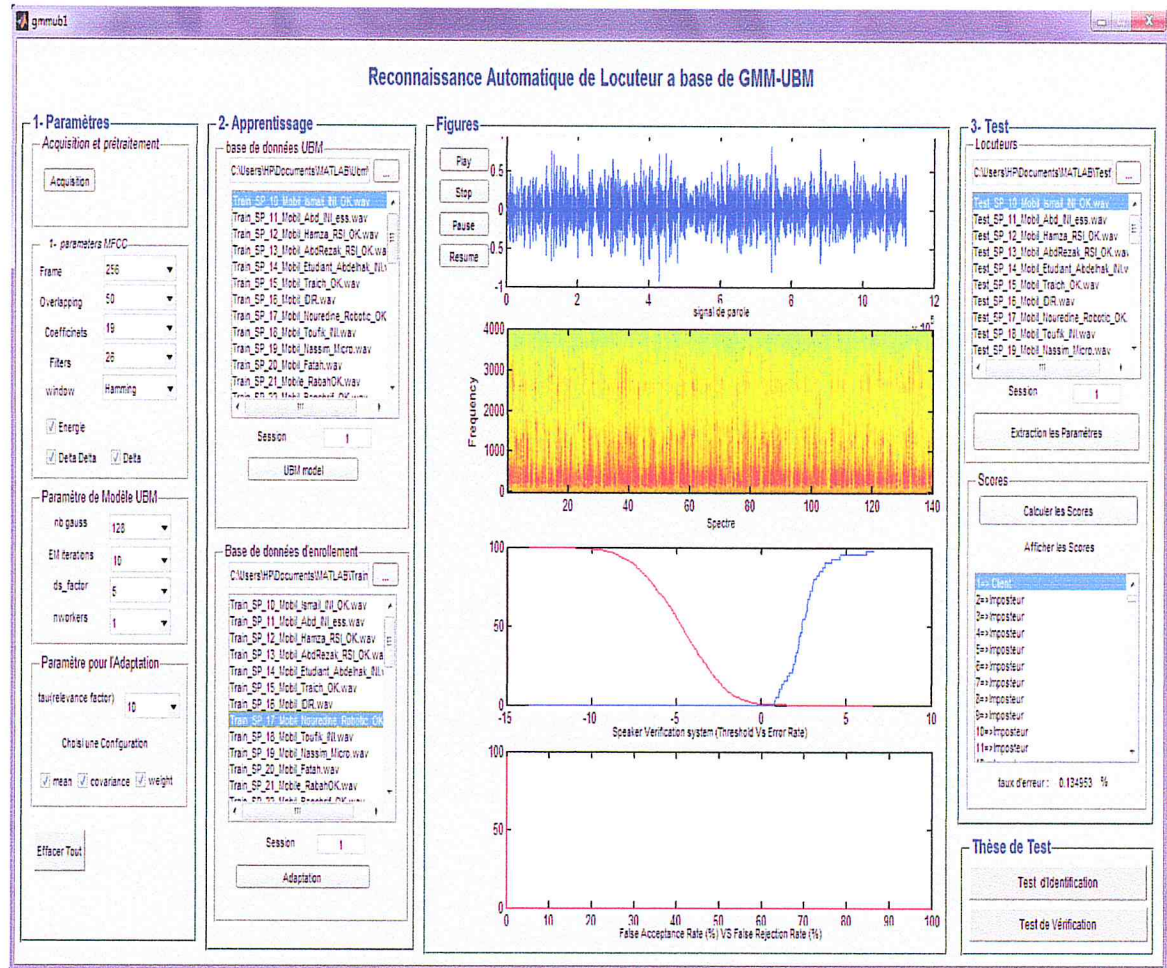


Figure V.7 : la fenêtre principale de modélisation par GMM-UBM.

Le panel **Reconnaissance Automatique de locuteur à base de GMM-UBM** comprend 4 sous-panels :

- ✓ Paramètres.
- ✓ Apprentissage.
- ✓ Test.
- ✓ Figure.

V.3.2.1 Paramètres

Ce panel comprend plusieurs sous-panels :

- Le panel acquisition permet d'enregistrer une voix d'un locuteur (création de base de données).
- Les panels Paramètres MFCC, Paramètres de modèle UBM et Paramètres pour l'adaptation permettent le réglage des paramètres nécessaires pour l'implémentation d'un système de vérification de locuteur.

1- Paramètres

Acquisition et prétraitement

1- paramètres MFCC

Frame	256	▼
Overlapping	50	▼
Coefficients	19	▼
Filters	26	▼
window	Hamming	▼

Energie

Delta Delta Delta

Paramètre de Modèle UBM

nb gauss	128	▼
EM iterations	10	▼
ds_factor	5	▼
nworkers	1	▼

Paramètre pour l'Adaptation

tau(relevance factor) 10 ▼

Choisi une Configuration

mean covariance weight

Figure V.8 : Formulaire des Paramètres GMM-UBM

V.3.2.2 Création du modèle globale (UBM)

Une fois que l'on possède une base de données (données d'apprentissage), on peut passer à l'étape de la création de modèle du monde ou l'UBM.

Utilisez le panel de la figure V.3 pour régler les paramètres (nombre de coefficients, type de fenêtre, nombre de filtre Mel, nombre de gaussiennes, nombre d'itération...etc).

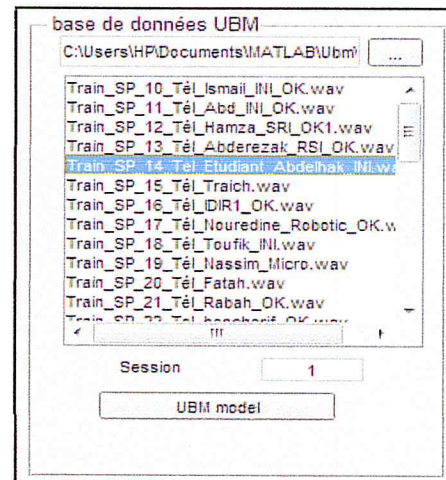


Figure V.9 : Charger Basse de donnée UBM

V.3.2.3 Création des modèles Client (modèle pour chaque locuteur)

Tout d'abord, allez dans le répertoire de base de données d'enroulement (données Clients). Puis calculez les paramètres MFCC. A la fin créez les modèles clients en cliquant sur le bouton 'adaptation' (adapter les données client au modèle UBM en utilisant l'adaptation MAP).

Utilisez le panel de la figure V.2 pour régler les paramètres à utiliser (nombre de coefficient, nombre d'itération, nombre de gaussiennes facteur de relevance...etc).

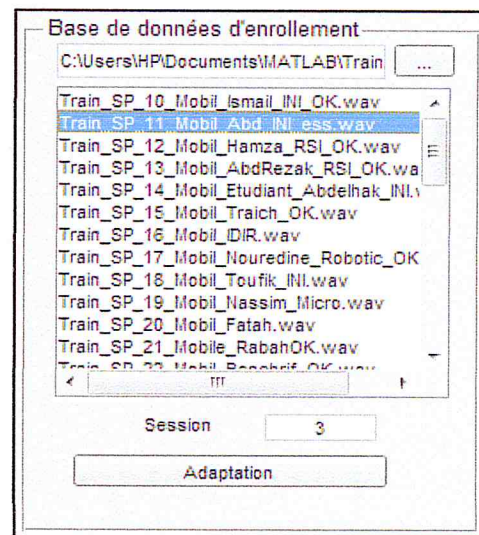


Figure V.10 : Charger Basse de donnée Train

V.3.2.4 Test

Cette section décrit ce qui se passe pendant la phase de test.

1. Choisissez la base de données qui correspond aux clients test (figure V.6).
2. Par la suite calculez les coefficients MFCC (le choix des paramètres est déjà fait dans la phase précédente).
3. Ensuite calculez les scores client et imposteur en utilisant l'équation (III.14) décrite dans le chapitre trois.

4. Affichez les scores et EER correspond à notre système de vérification (figure V.7)

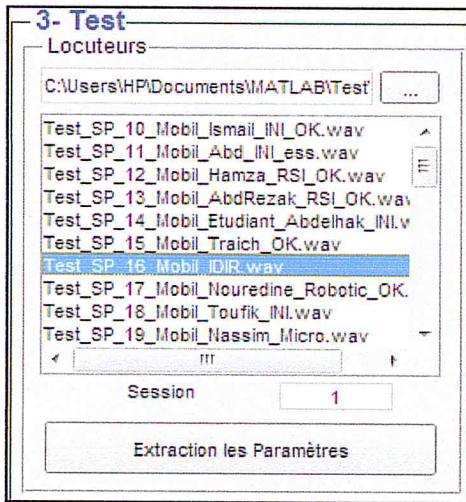


Figure V.11 : Charger BD Test

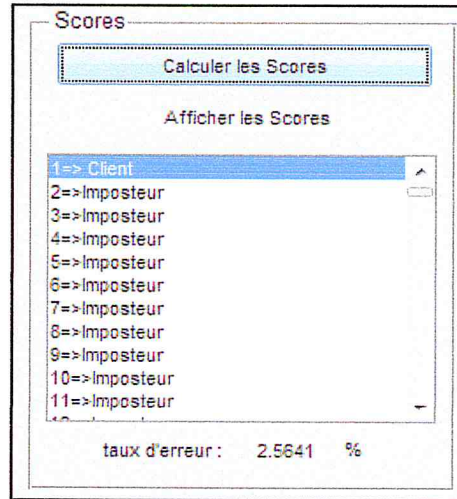


Figure V.12 : Affichez les Scores

V.3.3 Modélisation à base de I-Vecteur

Le panel Reconnaissance Automatique de locuteur à base de I-Vecteur comprend 4 sous-panels :

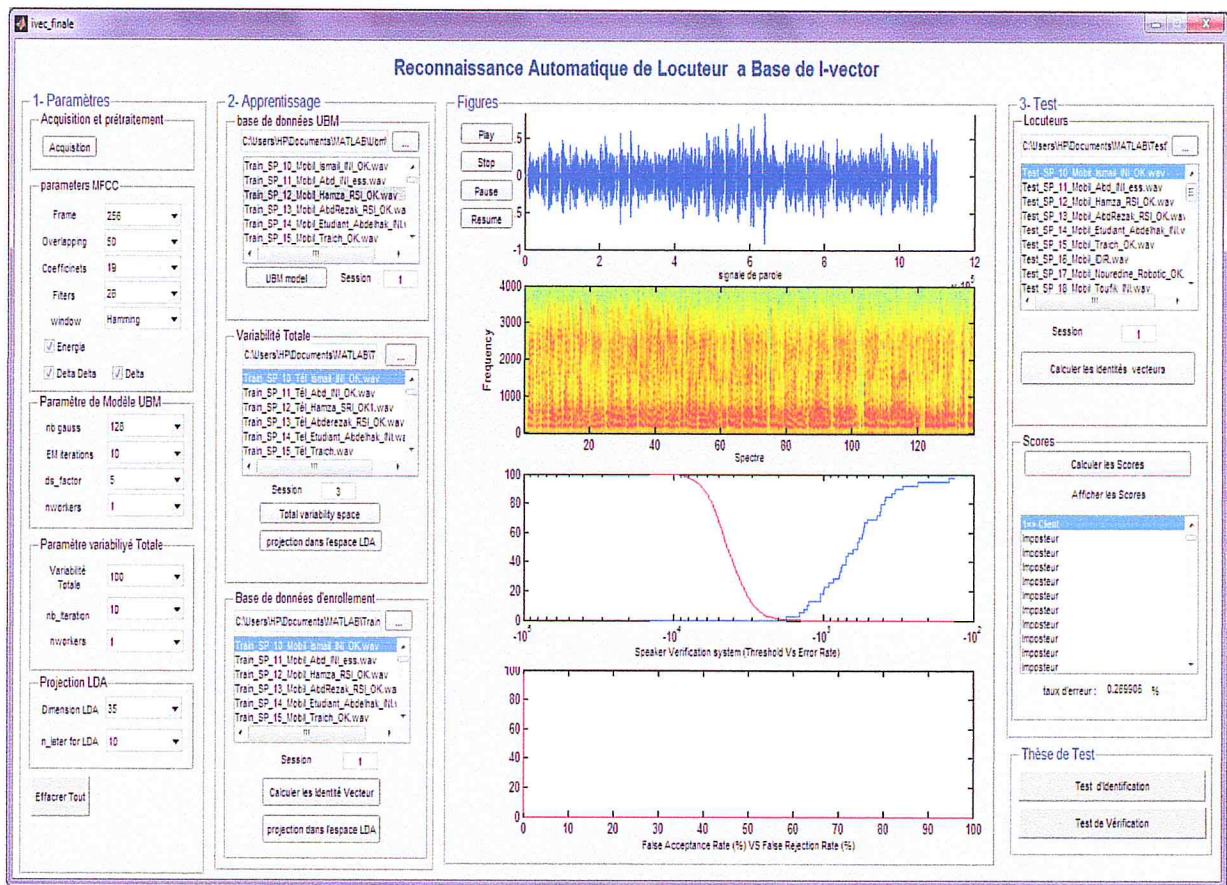


Figure V.13 : interface principale de modélisation par I-Vecteur.

V.3.3.1 Panel paramètre

Ce panel comprend plusieurs sous-panels :

- Le panel acquisition permet d'enregistrer une voix d'un locuteur (création de base de données).
- Le panel Paramètres MFCC: le choix du nombre de coefficients, l'utilisation des dérivées première et secondaire, le type de la fenêtre...etc.
- Paramètres de modèle UBM: le choix de nombre de gaussiennes, nombre d'itérations,..., etc.
- Le panel Paramètres variabilité Totale : le choix de l'ordre T, nombre d'itérations,..., etc.
- Le panel Projection LDA : le choix de l'ordre LDA

1- Paramètres

Acquisition et prétraitement

Acquisition

parameters MFCC

Frame 256

Overlapping 50

Coefficinetns 12

Filters 26

window Hamming

Energie

Delta Delta Delta

Paramètre de Modèle UBM

nb gauss 128

EM iterations 10

ds_factor 5

nworkers 1

Paramètre variabilité Totale

Variabilité Totale 80

nb_iteration 10

nworkers 1

Projection LDA

Dimension LDA 15

n_ieter for LDA 10

Effacer Tout

Figure V.14 : Formulaire des Paramètres I-Vecteur.

V.3.3.2 Panel Apprentissage

Comprend plusieurs sous-panels :

A. Sous-panel Base de données UBM

- B. choisissez la base de données,
- C. calculez les MFCC,
- D. créez le modèle UBM.

E. Sous-panel Variabilité Total

1. Allez dans le répertoire de base de données, calculez les paramètres MFCC, et puis Calculez les statistiques d'ordre 1 et 2,
2. Choisissez le nombre de sessions (dans cette phase, on utilise toutes les sessions, i.e. 3 sessions).
3. Calculez la variabilité Total T
4. Extraire les I-vecteur-UBM
5. Projetez les I-vecteur-UBM dans l'espace LDA et créez le modèle globale (G-PLDA).

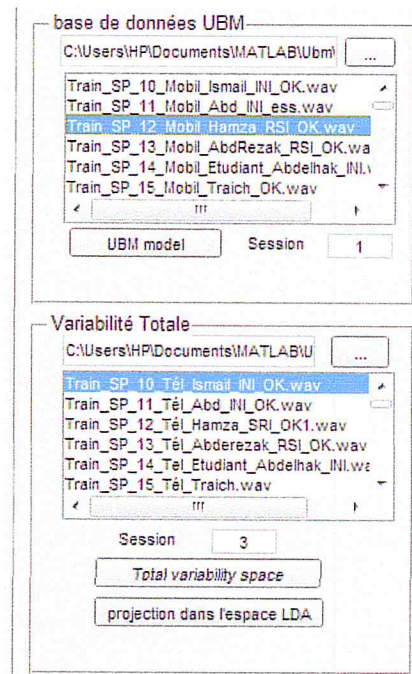


Figure V.15 : chargement des BD UBM et Variabilité totale.

C. Panel Enrôlement

1. Choisissez le nombre de sessions.
2. Allez dans le répertoire de base de données d'enroulement (données Clients), Choisissez la session et/ou les sessions, et puis calculez les statistiques d'ordre 1 et 2,
3. Extraire les I-vecteur-Clients
4. Projetez les I-vecteur- Clients dans l'espace LDA.

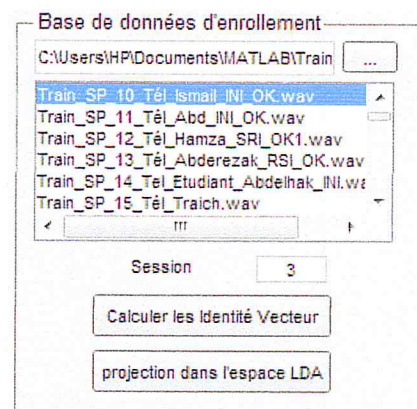


Figure V.16 : Charger BD Train.

V.3.3.3 Test

A. Sous-panel Base

1. Choisissez le nombre de sessions.
2. Choisissez la base de données qui correspond aux clients test, Choisissez la session et puis calculez les statistiques d'ordre 1 et 2,
3. Extraire les I-vecteur-Test
4. Projetez les I-vecteur-Tests dans l'espace LDA.

B. Sous-panel Scores

1. calculez les Scores clients et imposteurs
2. Affichez les scores et EER correspond à notre système de vérification.

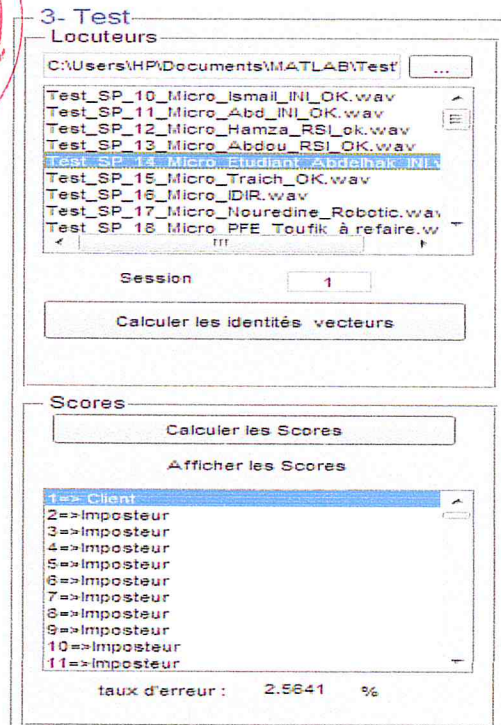
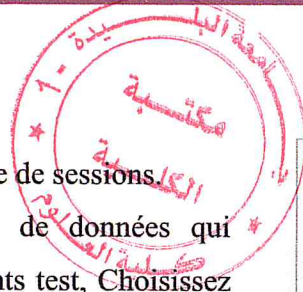


Figure V.17 : calcule des scores.

V.3.3.4 Panel des Figure

Ce panel commun entre les deux approche GMM-UBM et I-Vecteur. Aide à visualiser les Résultat calcule de performance.

- Sélectionné un fichier Voix (ubm, train, test, variabilité totale)
- Ecouter le signal et arrête, pause et résumé.
- Voir le % de la fausse acceptance par rapport le % du faux rejet.
- affichage la gamme du FAR (La fausse acceptance) et la gamme du FRR(Le faux rejet).

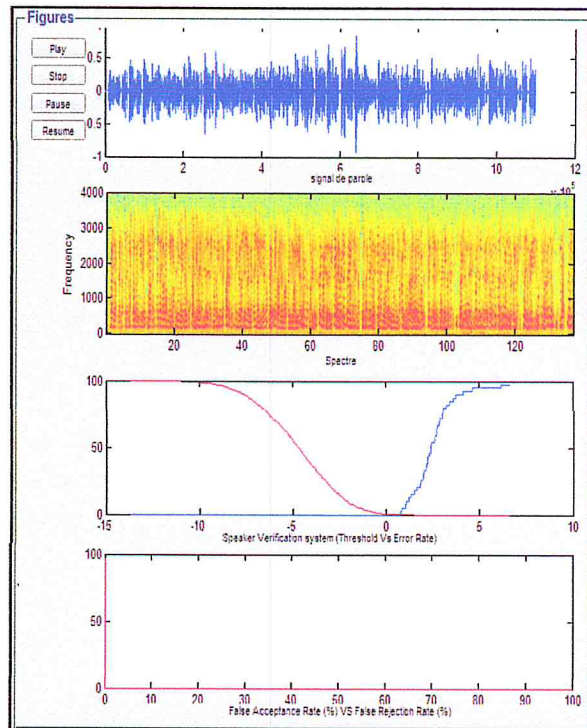


Figure V.18 : affichage les signaux et les résultats.

VI.3.3.5 Panel d'acquisition

Permet l'acquisition en temps réel d'un voix pour enrichie les Base de données (Test,UBM,Train,Variabilité Totale),

1. remplit les paramètres d'acquisition ensuite appuie sur le bouton 'Start recording'.
2. A la fin de l'acquisition choisi l'emplacement de ce nouveau fichier WAV, écoute le fichier ou bien ouvrir les figure indépendamment de l'interface.

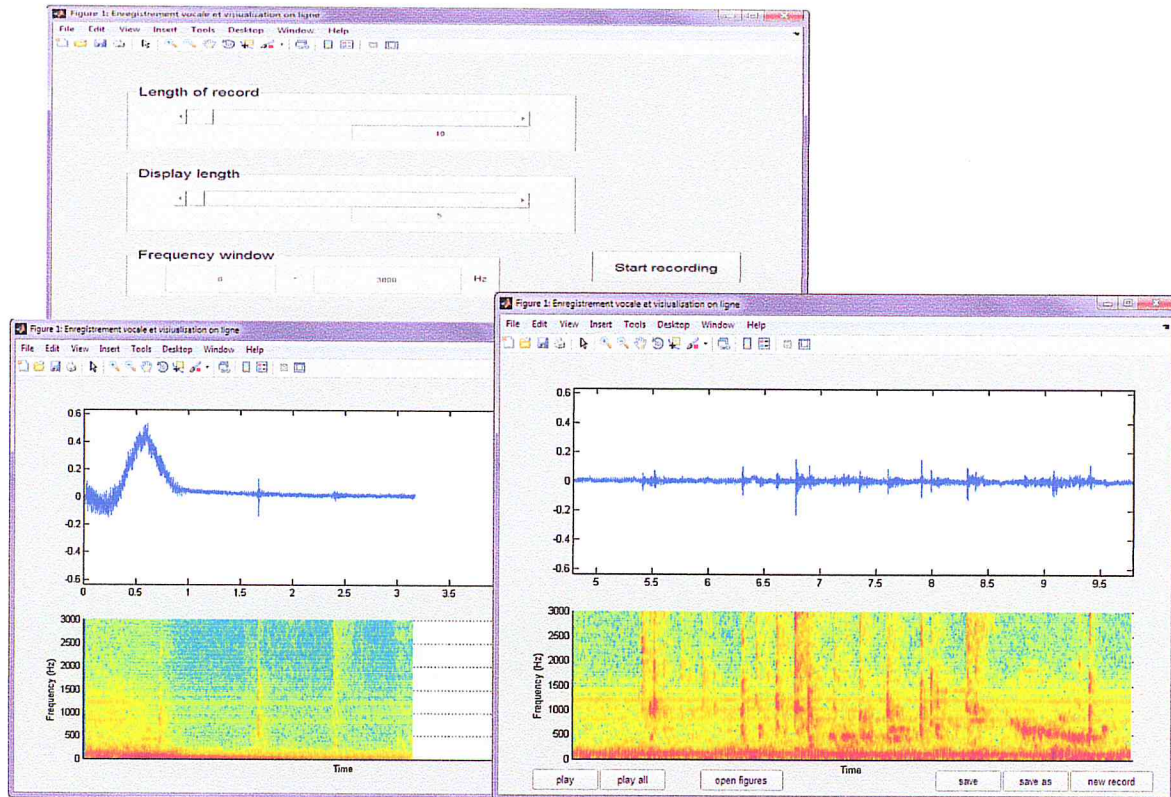


Figure V.19 Acquisition d'une voix en temps réel

VII. Application en temps réel

Cette interface (figure V.13) permet d'illustrer le processus d'identification/vérification à base GMM-UBM et/ou I-Vecteur.

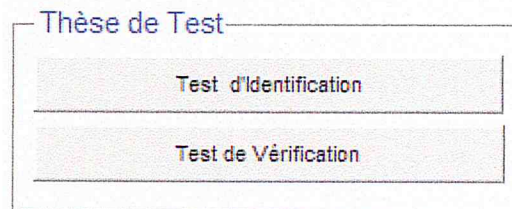


Figure V.20: Thèse de Test.

VI.1 En mode de Vérification

L'interface de la figure V.14 conclure les étapes de la Vérification comme suit :

- ✓ Chargez les Modèles clients et choisissez l'identité proclamée.
- ✓ Choisissez une identité-Test
- ✓ Calculé le score et le comparé à un seuil et vérifiez l'identité.
- ✓ Accepté ou rejeté l'identité (Client/Imposteur).

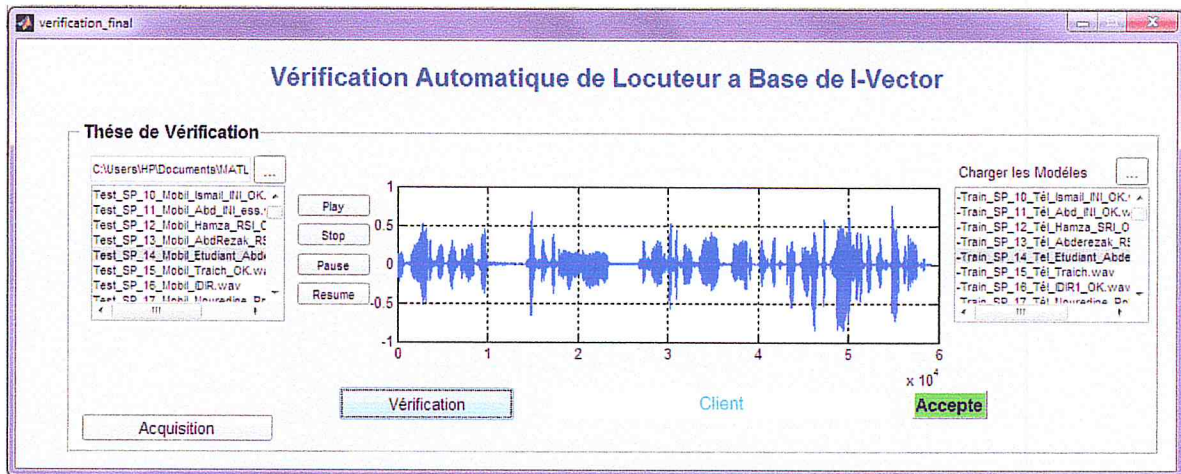


Figure V.21: Thèse de Test en mode de Vérification.

VI.2 En mode d'identification

Le système d'identification associé à la voix à identifier, qu'elle que soit d'une personne connue ou inconnue, une voix de la base d'apprentissage.

L'interface de la figure V.15 conclue les étapes d'Identification comme suit :

- ✓ Chargez les Modèles clients.
- ✓ Choisissez une identité-Test à identifier.
- ✓ Calculez le score.

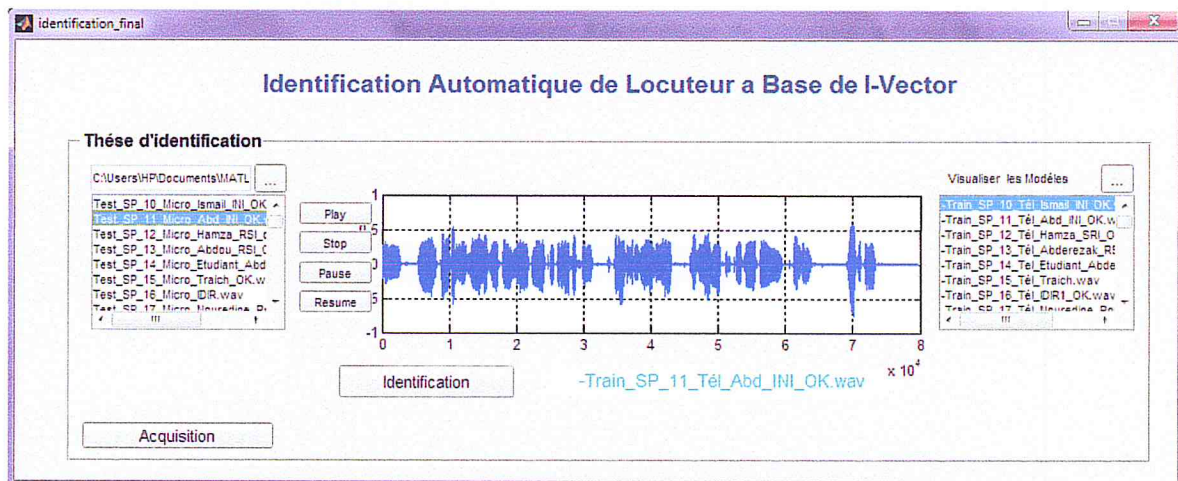


Figure V.22: Thèse de Test en mode d'Identification.

VIII. Conclusion

Ce chapitre était dédié à l'implémentation de notre système de RAL basé sur les deux techniques de modélisation GMM-UBM et I-Vecteur. Nous avons présenté les environnements matériels et logiciel utilisés ainsi que ses principales fonctionnalités en exposant ses différentes interfaces. Pour tester notre système, nous avons utilisé une base de données algérienne de 39 locuteurs. Les performances de notre système sont données par les valeurs EER et représentées par les graphes DET. Ces derniers ont montré une grande robustesse de la méthode I-vecteur par rapport à la méthode GMM-UBM.

Conclusion générale

La majorité des systèmes de reconnaissance de locuteur utilisent la variabilité inter-locuteurs 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 et intersessions d'enregistrement ou le "mismatch".

De plus, les systèmes de reconnaissance du locuteur s'appuient depuis plus d'une dizaine d'années sur l'approche probabiliste par mixtures de gaussiennes. Cette approche est basée sur la modélisation probabiliste de l'espace acoustique par un modèle unique (Gaussian Mixture Model-Universal Background Model ou GMM-UBM) et la représentation sous forme vectorielle des énoncés de voix, les supervecteurs, constitués par les statistiques adaptées du GMM-UBM.

Dans ce cadre de recherche, des avancées récentes ont révélé le pouvoir discriminant d'une nouvelle représentation vectorielle des énoncés vocaux, les i-vectors. Ce type de représentation présente la particularité de se réduire numériquement à un vecteur de dimension faible.

Le travail présenté dans ce mémoire s'inscrit dans le cadre de ce nouveau paradigme. Nous avons réalisé une plateforme logicielle de reconnaissance du locuteur englobant les deux approches de modélisation GMM-UBM et I-vector.

Nous avons effectué une étude sur l'effet de variations des canaux de transmission sur les performances du système de vérification du locuteur en utilisant un *corpus* d'un *dialecte algérien* de 39 locuteurs. Les résultats trouvés ont montré que l'approche I-vecteur est plus performante que celle basée sur la modélisation par GMM-UBM.

Comme perspectives et afin de donner plus de valeur à notre travail :

- ✓ Nous envisageons de tester notre système sur d'autres bases de données, comme la base de données NIST et MOBIO.
- ✓ Essayer d'intégrer des méthodes de calibration des scores afin d'améliorer les performances de notre système.
- ✓ Intégrer d'autres applications à savoir les applications judiciaires (identification forensique du locuteur).

BIBLIOGRAPHIE

- [1] D.A. Reynolds, T.F. Quatieri, R.B. Dunn: Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing* **10**(1) (2000) P 19–41
- [2] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, et P. Ouellet. "Front-End analyse factorielle pour la vérification Président". *IEEE transactions on audio, speech, and language processing*, VOL. 19, NO. 4, MAY 2011.
- [3] A. Lanitis, Nicolas Tsapatsoulis and Anastasios Maronidis, "Age Factors in Biometric Processing", Chapitre 2 "Review of ageing with respect to biometrics and diverse modalities". The institution of Engineering and Technology , London, United Kingdom 2014.
- [4] A. LARCHER. "Modèles acoustiques à structure temporelle renforcée pour la vérification du locuteur embarquée", THÈSE Doctorat, École Doctorale 166 « Information Structures Systèmes» Laboratoire d'Informatique (EA 4128), ACADÉMIE D'AIX-MARSEILLE UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE. 2009.
- [5] M. FAIRHURST "Age Factors in Biometric Processing ", Chapter 1 "Ageing and biometrics: an introduction" .p 8. University of Kent, UK. 2014.
- [6] F. MASSICOTTE " LA BIOMÉTRIE, SA FIABILITÉ ET SES IMPACTS SUR LA PRATIQUE DE LA DÉMOCRATIE LIBÉRALE ". Mémoire présenté comme exigence partielle de la maîtrise en science politique. 2007.
- [7] J. S. Mason, J. Oglesby, L. Xu, "Codebooks to optimise speaker recognition", European Conference on Speech Communication and Technology (Eurospeech), pages 267-270, Paris (France), 1989
- [8] F. Perronnin et Jean-Luc, "Introduction à la biométrie Authentification des individus par traitement Audio-Video", *Revue traitement signal*, volume 19, numéro 4, 2002.
- [9] P. BARTHELEMY "Le corps humain, clé d'accès aux systèmes de haute sécurité- La main comme passeport " *journal Le Monde* p 24. 1997

- [10] G. FLORIN, S. NATKIN "LES TECHNIQUES DE CRYPTOGRAPHIE. Unité de valeur Systèmes et applications répartis. P 106.
- [11] F. Monrose et A. D. Rubin. "Keystroke dynamics as a biometric for authentication". *Future Gener Comput Syst* 16(4), 351–359. 26, 2000.
- [12] C. Agency "Commonwealth of Australia, *Key Dates in the History of DNA Profiling*", [en ligne], <http://www.crimtrac.gov.au/dnahistory.htm>. (consulté le 14 septembre 2006).
- [13] A. Morris, D.Wu, J. Koreman, "GMM based clustering and speaker separability in the Timit speech database", "*Special Section/Issue on Corpus-Based Speech Technologies IEICE TRANS, VOL. E85-A/B/C/D, No. 1 pp 1-13, March 2005.*
- [14] B. Tounsi. " Inférence d'identité dans le domaine forensique en utilisant un système de reconnaissance automatique du locuteur adapté au dialecte Algérien". Mémoire présenté pour obtenir le grade de MAGISTER EN INFORMATIQUE. Spécialité Informatique Industrielle (II). Institut National de Formation en Informatique (I.N.I) Oued-Smar Alger. Direction de la Post-Graduation et de la Recherche.2007-2008.
- [15] T. Dutoit.2000. "Introduction au Traitement Automatique de la Parole". Notes de cours / DEC2. Première édition.Copyright © 2000 Faculté Polytechnique de Mons – T. Dutoit . Faculté Polytechnique de Mons.
- [16] Club de la sécurité des SI. "www.clusif.asso.fr" 2003.
- [17] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, W.Worek, N. I. of Standards, et T. Preliminary Face Recognition Grand Challenge Results. Dans les actes de "International Conference on Automatic Face and Gesture Recognition (FGR)". US Dept. of Commerce, National Institute of Standards and Technology. 2006.
- [18]N. Morizet, "Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris", *Thèse de doctorat,Ecole Nationale Supérieure des Télécommunications, Paris, Mars 2009.*
- [19] C.Champod, D.Meuwly, "The inference of identity in forensic speaker recognition", *Speech Communication*, vol. 31, pp. 193-203, 2000.

- [20] X. Tan, S. Chen, Z. Zhou, et F. Zhang. Face recognition from a single image per person : A survey. *Pattern Recognition* 39(9). 2006.
- [21] B. S. Atal, "Automatic recognition of speakers from their voices", *Proceedings of the IEEE*, 64(4). p460–475 , 1976.
- [22] G. R. Doddington, *Speaker recognition. identifying people by their voices*, *Proceedings of the IEEE*, 73(11) :1651–1664, 1985.
- [23] F. K. Soong, A. E. Rosenberg, L. R. Rabiner, B. H. Juang, "A vector quantization approach to speaker recognition", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 387-390, Tampa (USA), 1992.
- [24] R.G. Bachu, S. Kopparthi, B. Adapa, B.D. Barkana "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal". *Electrical Engineering Department School of Engineering, University of Bridgeport*.
- [25] G. RICHARD. Février, "Reconnaissance de la parole (suite): Paramétrisation " livre de cours p4. Master IAD Module PS. Publiée par "TELECOM ParisTech", 2008.
- [26] F. PERRONNIN et J. DUGELAY, "An Introduction to Biometrics Audio and Video-Based Person Authentication", *Traitement du Signal*, 2002, Vol. 19, n°4, pp 253-265.
- [27] D. Matrouf, J.-F. Bonastre, C. Fredouille, A. Larcher, S. Mezaache, M. McLaren, et F. Huenupan. LIA GMM-SVM system description : NIST SRE08. Dans les actes de NIST "Speaker Recognition Evaluation Workshop", Montreal (Canada). P 29, 60, 61. 2008.
- [28] D. A. Reynolds, T. F. Quatieri, et R. B. Dunn, 2000. "Speaker Verification Using Adapted Gaussian Mixture Models". *Digital Signal Processing* 10, 19–41. 112, 116.
- [29] N. Cristianini et J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", *Cambridge University Press*, 2000.
- [30] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 3, pp. 345-354, 2005.

[31] BOULKENAFET Zinelabidine." Vers un système robuste d'identification du locuteur en criminalistique" MEMOIRE DE MAGISTERE. ECOLE NATIONALE SUPERIEURE D'INFORMATIQUE (ESI) OUED-SMAR ALGER. 2012-2013.

[32] A. Amehraye "Traitement du Signal et Télécommunications ". Thèse' Docteur de Télécom Bretagne'. L'école Nationale Supérieure des Télécommunications de Bretagne en habilitation conjointe avec l'université de bretagne sud en cotutelle avec l'université mohamed-v-agdal de rabat. 2009

[33] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front end factor analysis for speaker verification," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 788-798, 2011

[34] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-End Factor Analysis For Speaker Verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 788 - 798, May 2010.

[35] S. J. Prince, and J. H. Elder, (2007). Probabilistic linear discriminant analysis for inferences about identity. In *International Conference on Computer Vision*, pages 1–8. IEEE.

[36] P. Kenny (2010). Bayesian speaker verification with heavy-tailed priors. In *Speaker and Language Recognition Workshop (IEEE Odyssey)*.

[37] P. Roques. " Les Cahiers du Programmeur. UML 2". Modéliser une application web 4^{ème} édition. Livre. P(10).

[38] Y. Prié, "UML Unified Modeling Language." M1 MIAGE – SIMA. UFR Informatique - Université Claude Bernard Lyon 1. (2006-2007).

[39] E. Cariou " Introduction à UML 2 ". Master Technologies de l'Internet 1^{ère} année. Université de Pau et des Pays de l'Adour. UFR Sciences Pau – Département Informatique : Eric.Cariou@univ-pau.fr.

[49] D. Reynolds. "Gaussian Mixture Models *". MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA: dar@ll.mit.edu.

[41] J. MacQueen. (1967). Some methods for classification and analysis of multivariate observations. In *Proc. of the 7th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281-297.

ANNEXE

L'algorithme EM (Expectation-Maximisation)

L'algorithme EM (Expectation Maximisation) est un moyen idéal pour résoudre les problèmes d'estimation des paramètres d'un GMM. En particulier, l'algorithme EM est applicable aux problèmes dont les données observées fournissent, seulement, des informations partielles ou s'il y a des informations manquantes. Voir la Fig a1. Prenant le modèle de mélange de gaussiennes présenté dans la Fig a2 comme un exemple, où π_j désigne la probabilité a priori d'appartenance au groupe j et $\varphi^{(j)} = \{\mu^{(j)}, \Sigma^{(j)}\}$ désigne les paramètres (moyenne et covariance) du groupe. Dans ce qui suit, on va répondre à la question suivante : pourquoi l'algorithme EM peut servir comme un outil puissant pour l'estimation de ces paramètres ?

L'algorithme EM est applicable à plusieurs modèles statistiques, par exemple, le modèle de mélange de gaussiennes (GMM) et la quantification vectorielle.

Chaque itération de l'algorithme EM est composée de deux étapes : Estimation (E) et Maximisation (M) . l'étape M maximise une fonction de vraisemblance qui est raffinée à chaque itération par l'algorithme E.

Un paramètre important de l'algorithme EM est qu'il peut être appliqué aux problèmes dont les données observées fournissent une information partielle. La Fig B(2) illustre le concept de données cachées et de données partielles.

Supposons qu'il y a deux groupes de données observées. Bien que toutes les données qui constituent les deux groupes soient observables, on ne peut pas savoir exactement à quel groupe les données appartiennent. La procédure d'estimation peut être simplifiée si l'information d'appartenance est supposée connue. Par exemple, si l'identité des groupes des variables x_1 à x_7 dans la figure Fig B(1) est connue, pour trouver la moyenne des groupes, on calcule la moyenne de chaque groupe séparément. La figure B(2) , le problème des données partielles. Contrairement à la figure B(1), le problème des données partielles dans la Figure B(2) contient une donnée incertaine y car y peut être égale à 5.0 ou à 6.0. par conséquent, la vraie valeur de y est inobservable, mais les valeurs de x_1 à x_4 sont observables. L'algorithme EM peut résoudre le problème de donnée partielle en calculant la valeur de y attendue. La figure B(3) illustre le cas où l'information d'appartenance à un groupe est cachée et seulement une information partielle est disponible. Ce

problème peut être vu comme une généralisation des problèmes illustrés dans les figure B(1) et B(2).

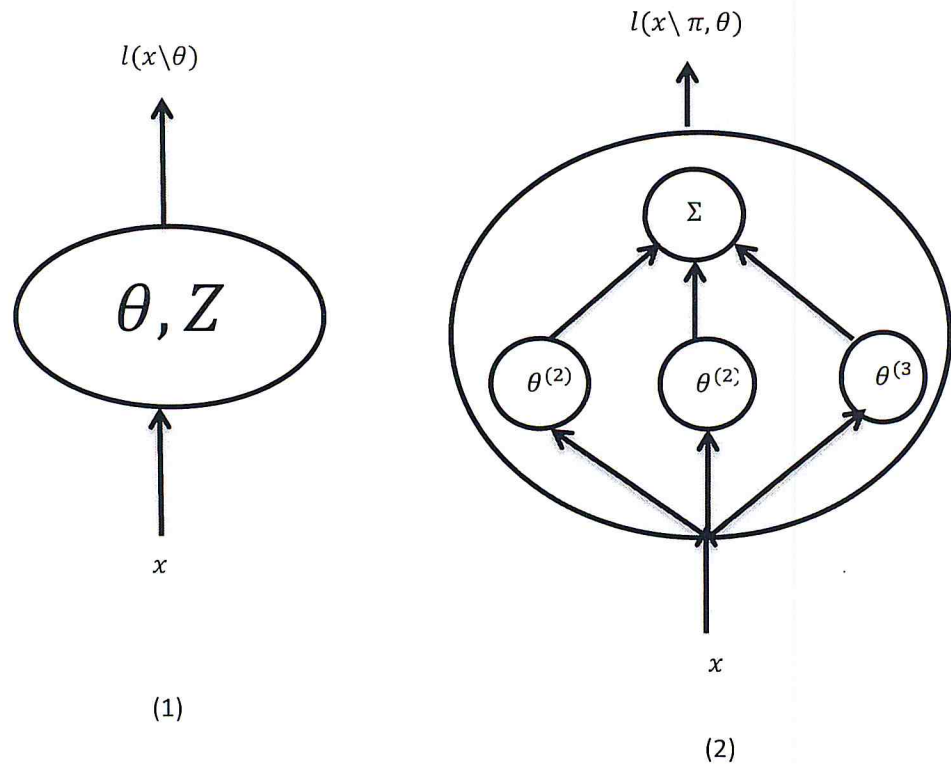
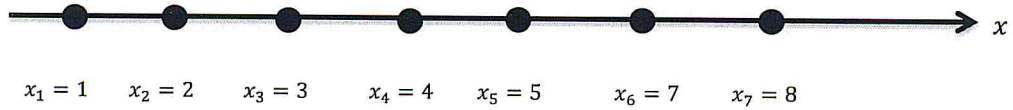
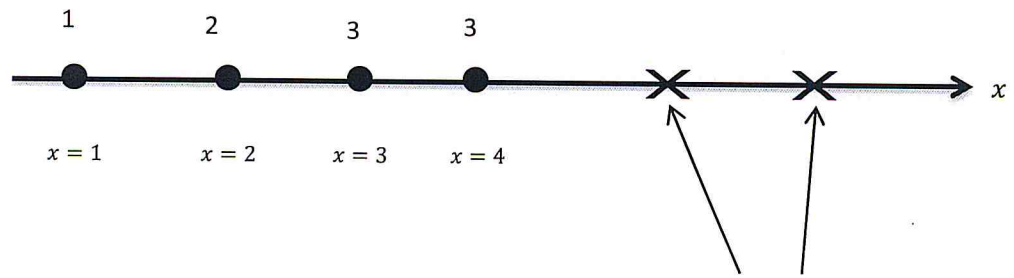


Figure a : l'estimation de paramètres par l'algorithme EM. (1) l'algorithme EM pour les problèmes des données générales. (2) pour les problèmes d'état caché ou le paramètre θ peut être divisé en deux groupes : $\{\pi^{(j)}\}_{j=1}^3$ et $\{\theta^{(j)}\}_{j=1}^3$, tel que $\pi^{(j)}$ représenté de données cachées et données partielles a été appliqué dans plusieurs applications scientifique. Par exemple la communication numérique, le récepteur reçoit une séquence de +1 et -1 sans savoir quel bit représente l'information manquante. Dans les applications de la biométrie, un GMM est typiquement appliqué pour modéliser les paramètres d'un individu. Chaque groupe est désigné pour modéliser quelques paramètres d'un individu. Dans ce cas, l'affectation des groupes constitue l'information cachée.

Exemple 1 : le problème d'état caché



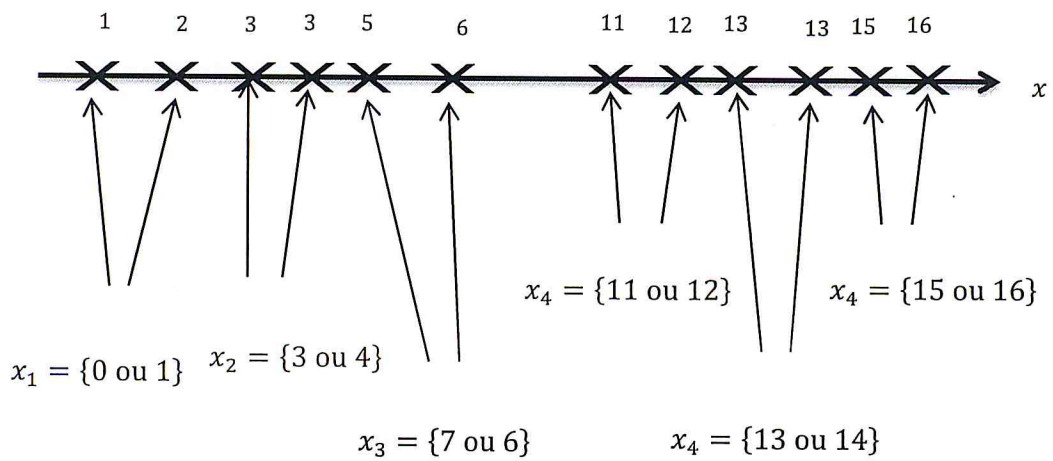
Exemple 2 : le problème de donr (a) rtielle



$Y = 5 \text{ ou } 6$

(b)

Exemple 3 : données partielle et état caché



(c)

Figure 2 : un exemple illustre le concept : (a) d'état cachée, (b) données partielles et (c) combinaison des deux problèmes, état cachée et données partielles. Dans (a), l'information d'appartenance de x_i à un groupe est cachée. Dans (b), y est

fournissent une information partielle car aucune de ces valeurs exactes n'est connue, at l'information d'apprentissage à un groupe est cachée.

2. la convergence de l'algorithme EM

Les notations suivantes sont adoptées

- $X = \{x_t \in \mathbb{R}^D; t = 1, \dots, T\}$ est la séquence d'observation, où T est le nombre d'observation et D est la dimension de x_t .
- $C = \{C^{(1)}, \dots, C^{(J)}\}$ est le jeu des données manquantes (spécifiant l'information d'état caché).
- $\theta = \{\theta^{(j)}, j = 1, \dots, J\}$ est le jeu de paramètres inconnus qui définissent la fonction de probabilité.
- $\theta^{(j)} = \{\pi^{(j)}, \phi^{(j)}\}$, où $\pi^{(j)}$ est la probabilité a priori de la densité de la $j^{\text{ième}}$ composante, et $\phi^{(j)}$ représente la densité de la $j^{\text{ième}}$ composante.

Notons que la combinaison des observations X et les états cachés Z constituent les données complètes. On définit :

$$L(X|\theta) = p(X|\theta_n) \quad (\text{a.1})$$

Comme la vraisemblance de la donnée complète étant donné l'estimation courante de θ_n , où n représente l'indice de l'itération. Définissons, ainsi, $p(Z, X|\theta_n)$ comme la vraisemblance de la donnée complète.

D'après la théorie des probabilités on peut écrire $p(X|\theta_n)$ comme suit :

$$p(X|\theta_n) = p(Z, X|\theta_n) / p(Z|X, \theta_n) \quad (\text{a.2})$$

En utilisant les équations (1) et (2), on peut écrire la vraisemblance des données complètes comme suit :

$$L(X|\theta_n) \equiv \log p(X|\theta_n) = [\log p(X|\theta_n)] \sum_Z p(Z|X, \theta_n) \quad (\text{car } \sum_Z p(Z|X, \theta_n) = 1)$$

Le P majuscule représente une probabilité, et le p minuscule représente une fonction de densité.

$$= \sum_Z P(Z|X, \theta_n) \log p(X|\theta_n)$$

$$\begin{aligned}
&= \sum_Z P(Z/X, \theta_n) \log \frac{p(Z, X|\theta_n)}{P(Z|X, \theta_n)} \quad (\text{d'apri la formule (a.2)}) \\
&= \sum_Z P(Z/X, \theta_n) \log p(Z, X|\theta_n) - \sum_Z P(Z/X, \theta_n) \log P(Z|X, \theta_n) \\
&= E_Z\{\log p(Z, X|\theta_n)/X, \theta_n\} - E_Z\{\log P(Z|X, \theta_n)/X, \theta_n\} \quad (\text{par difinition de l'espérance}) \\
&= Q(\theta_n|\theta_n) - R(\theta_n|\theta_n)
\end{aligned}$$

Ou

$$Q(\theta_n|\theta_n) = E_Z\{\log p(Z, X|\theta_n)/X, \theta_n\}$$

$$R(\theta_n|\theta_n) = E_Z\{\log P(Z|X, \theta_n)/X, \theta_n\}$$

Dans ce qui suit, on va démontrer pour quoi l'algorithme EM a une propriété de convergence générale. L'idée principale est basée sur l'inégalité de **Jensen**.

$$L(X|\theta) = \log p(X|\theta) = \log \left\{ \sum_Z p(Z, X|\theta) \right\} = \log \left\{ \sum_Z P(Z/X, \theta_n) \log \frac{p(Z, X|\theta)}{P(Z|X, \theta_n)} \right\}$$

$$L(X|\theta) = \log p(X|\theta) = \log \left\{ \sum_Z P(Z/X, \theta_n) \log \frac{p(Z, X|\theta)}{P(Z|X, \theta_n)} \right\}$$

$$= \log \left\{ E_Z \left[\frac{p(Z, X|\theta)}{P(Z|X, \theta_n)} \right] \middle| X, \theta_n \right\}$$

$$\geq E_Z \left\{ \log \left[\frac{p(Z, X|\theta)}{P(Z|X, \theta_n)} \right] \middle| X, \theta_n \right\} \quad (\text{l'intégration de Jensen})$$

$$= \sum_Z P(Z|X, \theta_n) \log \frac{p(Z, X|\theta)}{P(Z|X, \theta_n)} \quad (\text{Par définition de l'espérance})$$

$$= \sum_Z P(Z|X, \theta_n) \log p(Z, X|\theta) - \sum_Z P(Z|X, \theta_n) \log P(Z|X, \theta_n)$$

$$= \sum_Z P(Z/X, \theta_n) \log p(Z, X|\theta_n) - \sum_Z P(Z/X, \theta_n) \log P(Z|X, \theta_n)$$

$$= E_Z\{\log p(Z, X|\theta_n)/X, \theta_n\} - E_Z\{\log P(Z|X, \theta_n)/X, \theta_n\}$$

$$= Q(\theta_n|\theta_n) - R(\theta_n|\theta_n)$$

Dans l'étape M de la n^{ieme} itération, θ^* est calculé comme suit :

$$\theta^* = \arg \max_{\theta} Q(\theta|\theta_n)$$

On peut toujours choisir θ^* à l'itération n tel que :

$$Q(\theta^*|\theta_n) \geq Q(\theta_n|\theta_n)$$

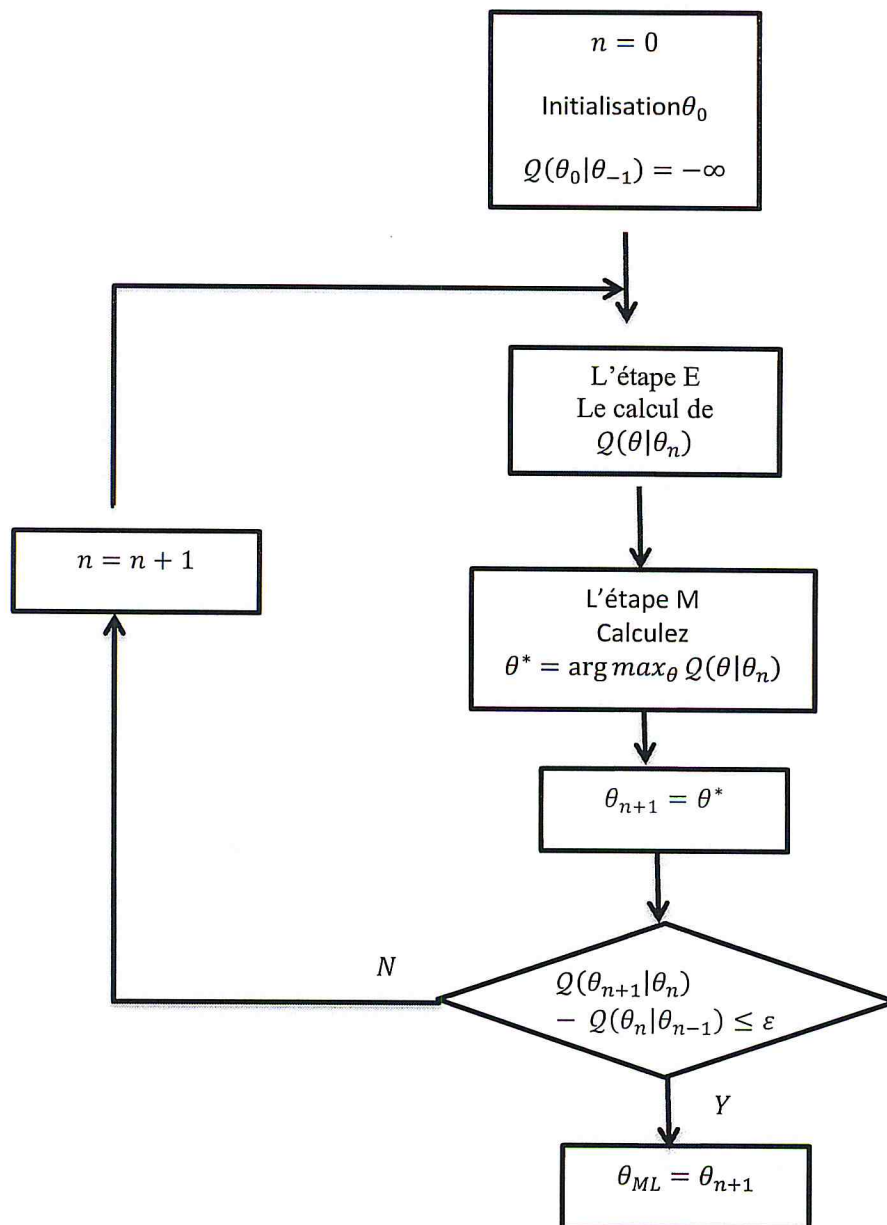
Notons que cette équation constitue une condition suffisante pour la convergence de l'algorithme EM, en effet, on a :

$$\begin{aligned} L(X|\theta_n) &\geq Q(\theta^*|\theta_n) - R(\theta_n|\theta_n) \\ &\geq Q(\theta_n|\theta_n) - R(\theta_n|\theta_n) \\ &= L(X|\theta_n) \end{aligned}$$

Au lieu de maximiser $L(X|\theta_n)$ directement, l'algorithme EM divise le problème d'optimisation en deux sous problèmes : Estimation et maximisation.

Dans chaque itération de l'algorithme EM, l'étape E calcule $Q(\theta|\theta_n)$ en utilisant un ensemble de paramètres θ_n . l'étape M détermine la valeur de θ (disent θ^*) qui maximise $Q(\theta|\theta_n)$; tel que

$$\theta^* = \max_{\theta} \sum_z P(Z|X, \theta_n) \log p(Z, X|\theta)$$



θ_{ML} : Estimation de maximisation de vraisemblance.

ε : Seuil d'arrêt.

Figure a. 3 : L'écoulement de l'algorithme EM.

A.3 La vraisemblance des données complètes

Si $X = \{x_t \in \mathcal{R}^D; t = 1, \dots, T\}$ contient T vecteurs indépendants et $Z = \{z_t \in \mathcal{C}; t = 1, \dots, T\}$ ou $z_t = C^0$ veut dire que z_t est généré par la J^{ieme} mixture, on peut écrire $p(Z, X|\theta)$ comme suit :

$$p(Z, X|\theta) = \prod_{t=1}^T p(z_t, x_t|\theta)$$

Soit $\Delta = \{\delta_t^{(j)}; j = 1, \dots, J; t = 1, \dots, T\}$ tel que :

$$\delta_t^{(j)} \equiv \delta(z_t, C^{(j)}) = \begin{cases} 1 & \text{si } x_t \text{ est généré par la mixture } C^{(j)} \\ 0 & \text{si non} \end{cases}$$

Comme pour toutes, seulement, un terme de $\{\delta_t^{(j)}; j = 1, \dots, J\}$ est à un et tous les autres sont à zéro, on peut exprimer $p(Z, X|\theta)$ comme suit :

$$p(Z, X|\theta) = \prod_{t=1}^T \sum_{j=1}^J \delta_t^{(j)} p(z_t, x_t|\theta)$$

$$= \prod_{t=1}^T \sum_{j=1}^J \delta_t^{(j)} p(x_t, z_t = C^{(j)}|\theta)$$

$$= \prod_{t=1}^T \sum_{j=1}^J \delta_t^{(j)} p(x_t, \delta_t^{(j)} = 1|\theta)$$

Par conséquent, la vraisemblance des données complètes est donnée par :

$$\log p(Z, X|\theta) = \sum_{t=1}^T \log \left\{ \sum_{j=1}^J \delta_t^{(j)} p(x_t, \delta_t^{(j)} = 1|\theta) \right\}$$

$$= \sum_{t=1}^T \log \left\{ \sum_{j=1}^J \delta_t^{(j)} p(x_t | \delta_t^{(j)} = 1, \theta^{(j)}) P(\delta_t^{(j)} = 1|\theta) \right\}$$

$$= \sum_{t=1}^T \log \left\{ \sum_{j=1}^J \delta_t^{(j)} p(x_t, | \delta_t^{(j)} = 1, \theta^{(j)}) P(\delta_t^{(j)} = 1|\theta) \right\}$$

$$= \sum_{t=1}^T \sum_{j=1}^J \delta_t^{(j)} \log(p(x_t | z_t = C^{(j)}, \phi^{(j)}) \pi^{(j)}) \quad (a.3)$$

Ou $\pi^{(j)}$ est le poids de la J^{ieme} mixture. La formule (a.3) utilise le fait que $p(x_t | \delta_t^{(j)} = 1, \theta) = p(x_t | \delta_t^{(j)} = 1, \theta^{(j)})$ et $P(\delta_t^{(j)} = 1 | \theta) = \pi^{(j)}$. Comme il y a un seul terme à un dans la sommation $\sum_{j=1}^J$, alors on peut extraire $\delta_t^{(j)}$ de la log

Sans affecter le résultat.

Dans l'étape E, on prend l'espace de la formule (a.3), on aura :

$$\begin{aligned} Q(\theta | \theta_n) &= E_z \{ \log p(Z, X | \theta) | X, \theta_n \} \\ &= \sum_{t=1}^T \sum_{j=1}^J p\{\delta_t^{(j)} | x_t, \theta_n\} \log [p(x_t | \delta_t^{(j)} = 1, \theta^{(j)}) \pi^{(j)}] \end{aligned}$$

Définissons $h_t^{(j)}(x_t)$ comme suit :

$$h_t^{(j)}(x_t) \equiv E\{\delta_t^j | x_t, \theta_n\} = P(\delta_t^j = 1 | x_t, \theta_n)$$

On définit, ainsi, $\pi_n^{(j)}$ comme le poids de la J^{ieme} mixture à l'itération n .

En utilisant le théorème de Bayes, on a :

$$\begin{aligned} h_t^{(j)}(x_t) &= P(\delta_t^j = 1 | x_t, \theta_n) \\ &= \frac{p(x_t | \delta_t^j = 1, \theta_n) P(\delta_t^j = 1 | \theta_n)}{p(x_t | \theta_n)} \\ &= \frac{p(x_t | \delta_t^j = 1, \theta_n^j) P(\delta_t^j = 1 | \theta_n)}{p(x_t | \theta_n)} \\ &= \frac{p(x_t | \delta_t^j = 1, \theta_n^j) \pi_n^{(j)}}{\sum_{k=1}^J p(x_t | \delta_t^k = 1, \theta_n^k) \pi_n^{(k)}} \end{aligned} \tag{a.4}$$

a.4 l'algorithme EM pou GMM

Pour bien illustrer les étapes de l'algorithme EM, on va appliquer ce dernier à un modèle de mélange de gaussiennes (GMM).

Soit un modèle, de mélange de gaussiennes :

$$\theta = \{ \pi^{(j)}, \mu^{(j)}, \Sigma^{(j)}, j = 1, \dots, J \}$$

Où $\pi^{(j)}$, $\mu^{(j)}$ et $\Sigma^{(j)}$ représentent, respectivement, le poids de la densité, le vecteur moyen et la matrice de covariance de la j^{ieme} densité du modèle.

Comme il est mentionné précédemment, le mélange de gaussienne est donné par :

$$p(x_t|\theta) = \sum_{j=1}^J \pi^{(j)} p(x_t | \delta_t^{(j)} = 1, \phi^{(j)})$$

Où

$$p(x_t | \delta_t^{(j)} = 1, \phi^{(j)}) = (2\pi)^{-D} |\Sigma^{(j)}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_t - \mu^{(j)})^t (\Sigma^{(j)})^{-1} (x_t - \mu^{(j)})\right\}$$

(a.5)

Est la j^{ieme} densité du modèle.

Après l'initialisation de θ_0 , les étapes de l'algorithme EM sont comme suit :

L'étape E : dans l'itération n, on calcule $h_n^j(x_t)$ pour chaque j et t en utilisant les formules (a.4) et (a.5).

L'étape M : on Maximise $Q(\theta|\theta_n)$ pour trouver θ^* . On remplace θ_n par θ^* . puis, on incrémente n par 1 et on répète E jusqu'à la convergence.

Pour trouver $\mu^{(k)*}$, on met $\frac{\partial Q(\theta|\theta_n)}{\partial \mu^{(k)}} = 0$, qui donne :

$$\mu^{(k)*} = \frac{\sum_{t=1}^T h_n^{(k)}(x_t) x_t}{\sum_{t=1}^T h_n^{(k)}(x_t)}$$

Pour trouver $\Sigma^{(k)*}$, on met $\frac{\partial Q(\theta|\theta_n)}{\partial \Sigma^{(k)}} = 0$, qui donne :

$$\Sigma^{(k)*} = \frac{\sum_{t=1}^T h_n^{(k)}(x_t) (x_t - \mu^{(k)*})(x_t - \mu^{(k)*})^T}{\sum_{t=1}^T h_n^{(k)}(x_t)}$$

Pour détermine $\pi^{(k)*}$, on maximise $Q(\theta|\theta_n)$ par rapport $\pi^{(k)}$, en respectant la contrainte suivante :

$$\sum_{j=1}^J \pi^{(j)} = 1$$

Cela nous donne le résultat suivant :

$$\pi^{(k)*} = \frac{1}{T} \sum_{t=1}^T h_n^{(k)}(x_t)$$

a.5 Complexité de l'algorithme EM

soient T le nombre de vecteur observés, J le nombre de mixtures et D la dimension des vecteur./ ce qui suit, est une évaluation approximative de la complexité du calcul en utilisant l'algorithme EM pour faire l'apprentissage d'un modèle GMM.

- L'étape E : $o(TJD + TJ)$ pour chaque itération.
- L'étape M : $o(2TJD)$ pour chaque itération.