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

ECOLE NATIONALE POLYTECHNIQUE

Département Electronique

CENTRE DE DEVELOPPEMENT DES TECHNOLOGIES AVANCEES

Division Architecture des systèmes

Projet de fin d'études en vue de l'obtention du Diplôme d'Ingénieur d'Etat en Electronique

Authentification biométrique bimodale par le classificateur GMM Orthogonal.

Réalisé par : ***Melle HADJ BOUZID Amina Ikram***

Mr HAMROUCHE Mohamed El-Bachir

Soutenu le 03 Juillet 2011 devant le jury composé de :

President: Mr. H.BOUSBIA-SALAH

Promoteurs: Mr. M.BENGHERABI (CDTA)

Mme. L.HAMAMI (ENP)

Examinateurs: Mr. L. ABDELOUEL

Mlle. A. MOUSSAOUI

Ecole Nationale Polytechnique, 10 Av. Hassan Badi, El Harrach, Alger, Algérie.

ملخص

يندرج عملنا المقدم في إطار " مقاييس الحيوية" وأنظمة التعرف على الأشخاص. يمكن تقسيم هذه الدراسة التي قمنا بها إلى ثلاثة عناوين كبرى. بدايةً، دراسة كل من الصوت والسيماء، لحظة تسجيلهما إلى غاية استخراج محدداتهما وقولبتهما عن طريق مصنف " نموذج غوص GMM" الشائع الاستعمال في التعرف على المتكلم بغض النظر عن النص. وأتبعنا ذلك بدراسة "نظام تعامد غوص OGMM" الذي يعتبر تحسينا للخوارزمية بالقيام باختبارات مقارنة لكل كيفية. وأتممنا هذه الدراسة في الأخير بالاتجاه الرائج اليوم والمتمثل في تعدد الكيفيات، بغرض تقوية نظام " مقاييس الحيوية" دون تكاليف أو عوائق إضافية. ونختم عملنا هذا بدراسة مختلف الكيفيات وطرق مزج الصوت والوجه.

الكلمات المفتاحية : الصوت - السَيمَاء أو الوجه - نَسَق التعرف علي الأشخاص- اِستخراج- الإستولاء - يحفز - نظام تعامد غوص - نموذج غُوص - المعطيات القاعدية - تعدد الكيفيات - يتوافق – التحام.

RESUME

Notre travail se place dans le cadre de la biométrie et des systèmes de reconnaissance des individus. L'étude que nous avons menée peut être divisée en trois grandes parties. Elle commence par l'étude des deux modalités que sont le visage et la voix, depuis leur capture jusqu'à l'extraction de leurs paramètres et leur modélisation par le classificateur GMM (Gaussian Mixtures Models) qui est l'un des classificateurs les plus utilisés en reconnaissance du locuteur en mode indépendant du texte. Elle se poursuit par l'étude des OGMM (Orthogonal Gaussian Mixtures) qui est une amélioration de l'algorithme standard en effectuant des tests comparatifs pour chaque modalité. Elle se termine enfin par la tendance actuelle qui est la multimodalité. En effet, pour renforcer un système biométrique sans utiliser de modalité coûteuse ou intrusive.

On terminera donc par étudier différentes méthodes de fusion des modalités visage et voix, depuis les méthodes simples jusqu'aux méthodes basées sur les classificateurs GMM et OGMM.

Mots clés :*visage, voix, authentification, identification, extraction de paramètres, transformée en cosinus discrète, coefficients cepstraux de Mel, modélisation par mélange de gaussiennes, analyse en composante principale globale, analyse en composante principale, fusion.*

ABSTRACT

This work focuses on biometrics and individuals recognition systems. The study we carried out can be divided in three major parts. It starts with the work on the two modalities that are theface and the voice, from their acquisition until the extraction of their important parameters And their clustering with the classifier GMM (Gaussian Mixture Models), one of the most

Famous classifiers used in speech independent speaker recognition. It continues with the study of OGMM (Orthogonal Gaussian Mixture Models), an improvement of the original algorithm, by comparing it to the GMM in terms of precision and speed. It ends with the study of multimodality. Actually, nowadays, it is preferred to fusion simple, low cost modalities than to use expensive and intrusive ones. We will end our work with the study of various fusion technics, from simple methods to classifier based methods.
Key words : *face, speech, authentication, identification, extraction of parameters, discrete Cosine transform, mel frequency cepstral coefficients, gaussian mixture models, global Principal analysis component, principal analyses component, fusion.*

DEDICACES

Mon père, qui m'a permis d'en arriver là ou j'en suis aujourd'hui, en me poussant à dépasser mes limites sans jamais renoncer; sa droiture, sa générosité et sa conduite ont toujours été un modèle pour moi.

Ma mère, qui a toujours su m'écouter et pris le temps de me comprendre; sa douceur, sa tendresse et sa bonté ont toujours mérité mon plus profond respect.

Mon frère, que j'admire pour sa vivacité d'esprits et sa volonté sans faille et avec qui j'affiche une grande complicité.

Ma sœur, petit ange à l'éveil remarquable et à l'imagination débordante ; source de bonne humeur, grâce a qui j'arrive à avoir un regard neuf sur le monde qui m'entoure.

Mes grands parents, pour leur amour et leur soutient.

Ma famille, qui n'a jamais cessé de me soutenir et de croire en moi.

Mon binôme et ami doté de grandes qualités humaines, que je tiens à remercier tout particulièrement pour ces innombrables échanges scientifiques et ce travail très constructif.

Mes amis proches, qui m'ont apporté un grand soutient tout au long de mon cursus universitaire.

Je vous dédie ce modeste travail

Amina Ikram

DEDICACES

Je dédie ce travail à :

A **mes parents** qui ont toujours su me pousser à aller plus loin.

A ma sœur **Yasmine**, pour son aide précieuse et son soutien continu.

A ma sœur **Sihem**, pour ses encouragements et sa confiance.

A mon cousin **Zinou**, pour son humour et sa bonne humeur.

A mes neveux, **Manel et Nazim**, qui sont une source de bonheur pour moi.

A **ma binôme**, pour tous les moments heureux et difficile qu'on a eu.

A **mes amis**, pour les souvenirs que je garderai toute ma vie.

Bachir

Remerciements

Nous voilà arrivés à l'issu de cinq années d'études intensives et riches en enseignement. De ce fait, au doux sentiment d'accomplissement se mêlent les premiers élans de mélancolie. Afin d'éviter de sombrer dans ce qui pourrait être un brin de nostalgie abusive, nous tenons à rendre à toutes celles et tous ceux qui ont contribué de près ou de loin à parfaire notre formation au sein de l'ENP un hommage massif en leur soumettant nos remerciements les plus sincères.

En premier lieu, nous tenons à remercier nos promoteurs :

Mme. L. HAMAMI, qui a immédiatement accepté de nous encadrer en tant que copromotrice, et qui nous a encouragé à prendre ce sujet. Nous la remercions pour son aide et son suivi avisé.

Mr. M. BENGHERABI qui nous a permis de traiter un sujet aussi attrayant, tant en profondeur qu'en expérience acquises. Ses compétences techniques, son expérience, sa disponibilité et sa rigueur ont été d'une aide précieuse pour la concrétisation de ce travail.

Mr. F. HARIZI, pour ses explications et orientations ainsi que pour son soutien et ses précieux conseils qui nous ont permis d'intégrer et de maitriser les différentes approches biométriques dont nous avions besoin.

Nous remercions vivement Mr. H. BOUSBIA-SALAH, de nous avoir fait l'honneur de présider le jury de soutenance de ce projet de fin d'études.

Nous adressons également nos remerciements à Mr. L. ABDELOUEL ainsi que Mlle. A. Moussaoui pour nous avoir consacré une partie de leur précieux temps afin de juger et d'évaluer notre travail.

Enfin nous remercions l'ensemble de nos enseignants du département des sciences fondamentales ainsi que du département d'électronique qui nous ont guidé et enseigné durant notre cursus à l'ENP.

H. Bachir & HB. Ikram

Sommaire

Liste des figures

Liste des tableaux

Liste des abréviations

FRR: False Rejection Rate

FAR : False Acceptance Rate

IR : Identification Rate

FR : False Rejection

FA : False Acceptance

ROC : Receiver Operating Characteristic

EER : Equal Error Rate

HMM : Hidden Markov Model

EO : Eigen Objects

EBGM : Elastic Bunch Graph Matching

TM : Template Matching

PCA : Principal Compenent Analysis

LDA : Linear Discriminant Analysis

SVM : Support Vector Machines

GMM : Gaussian Mixtures Models

EM : Expectation Maximization

ANN : Artificial Neural Network

DCT : Discret Cosine Transform.

KLT : Karhunen-Loève Transform

FFT : Fast Fourrier Transform

DFT: Discret Fourrier Transform

VAD : Voice Activity Detection

ZCR : Zero Crossing Rate

MFCC : Mel Frequency Cepstral Coefficients

DTW : Dynamic Time Warping

VQ: Vector Quantization

LBG: Linde, Buzo & Gray

ML : Maximum Likelihood

MAP : Maximum A Posteriori

UBM : Universal Background Model

LLR : Log Likelihood Ratio

LRT : Likelihood Ratio Test

MM : Min-Max

ZS : Z-Score

TH : Tangente Hyperbolique

AD : Adaptative

QQ : Two-Quadrics

LG : Logistic

QLQ : Quadratic-Line-Quadric

MIN : Min-Score

MAX : Max-Score

SS : Simple-Sum

MW : Matcher-Weighting

TIMIT : Texas Instruments- Massachusetts Institute of Technology

ORL : Olivetti Research Laboratory

Introduction générale

Dans le monde d'aujourd'hui, la société connait un développement permanent et important, les besoins en sécurité se sont accrus considérablement. Pour maintenir un niveau de sureté suffisant, les méthodes de contrôle et de surveillance classiques ne sont plus efficientes. Il devient alors important de trouver des moyens de contrôles qui soient à la fois sûrs et faciles d'accès pour les usagers.

Pour répondre à ce problème, l'Homme a mis en place une nouvelle technique de reconnaissance qui a fait son apparition et ne cesse de croître depuis 1997 : il s'agit des contrôles d'accès par les systèmes biométriques.

La reconnaissance automatique des individus par leurs signatures biométriques spécifiques est une méthode sûre qui permet d'identifier une personne ou de vérifier son identité en se basant sur certaines de ses caractéristiques morphologiques, comportementales ou même biologiques.

Traditionnellement, les systèmes biométriques se basent sur une seule modalité telle que le visage, la rétine ou les empreintes digitales. Et bien que certaines modalités se soient montrées très utiles, elles restent coûteuses et difficiles à collecter telle que l'iris et la rétine. Pour remédier à cela, nous assistons aujourd'hui à l'apparition de systèmes dit multimodaux. Ceux-ci utilisent plusieurs modalités afin d'augmenter la précision du système sans pour autant atteindre des coûts trop élevés.

C'est dans ce cadre que se place notre projet de fin d'études qui traite de la fusion des modalités visage et voix après une modélisation de leurs paramètres par un mélange de gaussiennes puis par une amélioration de cette technique classique par l'application d'une transformation algébrique en amont.

Pour se faire nous procèderons à l'extraction des paramètres DCT (Discrete Cosine Transform) pour le visage ainsi que les paramètres MFCC (Mel Frequency Cepstral Coefficient) pour la voix. Ces derniers se verront modéliser en utilisant les GMM (Gaussian Mixture Models) ainsi que les OGMM (Orthogonal Mixture Models), ce qui permettra de mettre en évidence l'apport du classificateur orthogonal par rapport au classificateur diagonal classique. Viendra ensuite la génération des scores qui permettra en premier lieu d'évaluer les

deux systèmes uni modaux séparément pour ensuite combiner leurs apports en fusionnant les scores obtenus précédemment.

Le mémoire est organisé de la manière suivante :

Le mémoire traite d'abord des systèmes biométriques en général et de la reconnaissance du visage et de la voix en particulier. Il traite ensuite de la modélisation par les classificateurs GMM (Gaussian Mixture Models) et OGMM (Orthogonal Gaussian Mixture Models), puis entame les différentes techniques de fusion des scores avant d'examiner la programmation du système sous matlab et la mise en place d'une interface graphique aboutissant à une évaluation du système et à une comparaison des différentes approches abordées. Ceci sera relaté à travers les chapitres suivants :

Le chapitre 1, traite de généralités concernant la biométrie et les systèmes biométriques en mettant en évidence les avantages et les inconvénients de chacune des modalités ainsi que l'intérêt d'un système biométrique multimodal.

Le chapitre2, aborde l'état de l'art des systèmes de reconnaissance du visage en présentant les différents types de classificateurs utilisés dans ce cas de figure ainsi que les étapes à suivre pour aboutir à l'extraction des paramètres DCT .

Le chapitre3, aborde l'état de l'art des systèmes de reconnaissance du locuteur en présentant les différents types de classificateurs utilisés dans ce cas de figure ainsi que les étapes à suivre pour aboutir à l'extraction des paramètres MFCC.

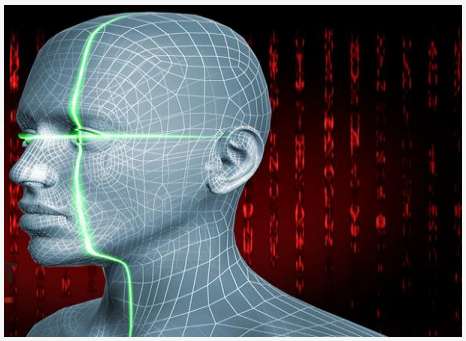
Le chapitre 4, présente la modélisation des paramètres biométriques par l'utilisation des GMM diagonales ainsi que des GMM orthogonales générées par deux différentes approches et la classification des paramètres modélisés dans le cas de la vérification d'une identité proclamée ou de l'identification d'un individu.

Le chapitre 5, présentera les techniques de normalisation de scores ainsi que la fusion de ces derniers.

Le chapitre 6, présente le cheminement suivi durant tout le travail accompli ainsi que la présentation du système à travers l'interface graphique sous Matlab.

Le chapitre 7, permet de prester les résultats obtenus à travers la multitude de tests effectués et sélection du classificateur optimal.

1 Biométrie et systèmes de reconnaissance des individus



I. INTRODUCTION

Depuis quelques décennies, nous vivons dans des sociétés où la circulation des données, de l'argent et même des individus est entièrement dépendante de l'outil informatique. Cette dépendance croissante a rendu nécessaire le développement de moyens de contrôle d'accès numérique pour protéger l'information sur les ordinateurs, les réseaux téléphoniques, internet et dans les zones sensibles (centres de recherches, bases militaires, centrales nucléaires…).

Ces moyens de contrôle doivent reconnaître une personne comme cliente avant de lui ouvrir l'accès physique ou numérique. Pour cela, ils se basent usuellement sur ce que celle-ci *possède* comme carte d'identité, carte à puce, badge magnétique ou bien sur ce qu'elle *sait* c'est-à-dire un mot de passe [Sta 09]. Néanmoins, ces méthodes posent un réel problème de fiabilité puisqu'il est facile pour quelqu'un d'oublier son code, ou pire, de se le faire voler au moyen de logiciels de décryptage.

Pour apporter une solution à ces problèmes, des efforts importants ont et sont fournis dans le domaine de la recherche depuis quelques décennies pour développer de nouveaux systèmes de sécurisation qui se basent, non pas sur ce que la personne possède ou sait, mais sur ce qu'elle *est* [Sta 09]. Ces derniers étant nommés ***systèmes de reconnaissance biométrique.*** Un système de reconnaissance biométrique se base sur les caractéristiques physiques et comportementales des individus afin de les reconnaître. Il devient alors pratiquement impossible pour une personne de voir son identité usurpée.

Aujourd'hui, l'intérêt pour ce domaine s'est accru du fait de la présence d'un contexte mondial dans lequel les besoins en sécurité deviennent de plus en plus importants et où les enjeux économiques sont colossaux.

C'est dans ce contexte que se place notre travail. Nous commencerons l'étude par une brève présentation de ce vaste domaine qu'est la biométrie ainsi que les diverses applications qui en découlent.

II. GENERALITES SUR LA BIOMETRIE

II.1 Définition

Reconnaître une personne familière, en utilisant certaines de ses caractéristiques, comme sa voix, son visage ou son écriture est une pratique naturelle chez les êtres humains qui s'en servent intuitivement. Concevoir un système de reconnaissance biométrique revient à apprendre à la machine à se comporter comme un être humain : mémoriser le modèle d'un individu après l'avoir « rencontré » une fois pour pouvoir le reconnaître ultérieurement.

Etymologiquement parlant, le terme ***« biométrie »*** est une combinaison des mots « bio » signifiant *vie* et « métrie » qu'on peut traduire par *mesure* en grec ancien. Ainsi, la biométrie signifie « l'application des méthodes statistiques modernes pour mesurer des objets biologiques » [Lar 07]. Cependant, par abus de langage, ce terme fait souvent référence aux technologies de mesure et d'analyse de caractéristiques biologiques et anthropologiques, comme les empreintes digitales, la rétine, l'iris, la voix ou le visage. Aussi, le terme ***«biométrie »*** se voit plus souvent associé avec le sens d'« identifier un individu en se basant sur ses caractères distinctifs » [Bol 03].

II.2 Historique

Le recours à la biométrie est né d'un très ancien besoin qui remonte aux premières civilisations : celui d'identifier une personne, ne serait-ce que pour valider la conformité d'un message transmis ou pour permettre l'accès d'une personne qu'on ne sait être alliée ou ennemie. Mais c'est la civilisation chinoise qui a été la première à utiliser, il y a 1000 ans, les empreintes digitales à des fins de signature de documents. Il fallut, cependant, attendre l'arrivé de l'anatomiste ***Marcello Malpighi (1628–1694)*** pour étudier plus en détails les empreintes avec un microscope. Et celle, plus tardive, du physiologiste tchèque ***Jan Evangelista Purkinge (1787–1869)*** pour les catégoriser selon certains critères.

Vers la fin du XIX siècle, le ***Dr Henry Faulds (1843–1930),*** chirurgien à Tokyo, a marqué le premier pas vers l'élaboration d'un système d'identification d'individus en se basant sur des méthodes statistiques pour la classification des empreintes.

Au même moment, un de ses contemporains, le français ***Alphonse Bertillon (1853-1914),*** testait une méthode d'identification des prisonniers nommée *anthropométrie judiciaire.* En effet, Il procédait à la prise de photographies de sujets humains, mesurait certaines parties de leurs corps (tête, membres, etc.) et notait les dimensions sur des fiches à des fins d'identifications ultérieures. Ce fut la naissance de la première base de données contenant des informations sur un groupe d'individus.

II .3 Les applications de la biométrie

De nos jours, les applications biométriques sont nombreuses et permettent d'apporter un niveau de sécurité supérieur à celui proposé par les techniques classiques, en ce qui concerne les *accès logiques* ou *physiques* [Jai 04].

La liste des applications pouvant utiliser la biométrie peut être très longue et n'est limitée que par l'imagination des concepteurs. On peut citer [Jai 04] :

- ✓ Le contrôle d'accès physique aux frontières, aux sites sensibles (service de recherche, site nucléaire, bases militaires etc.) ou juste aux salles informatiques.
- ✓ La gestion des titres d'identité (cartes nationales…etc.).
- ✓ La sécurisation des transactions financières et des transferts de données entre entreprises.
- ✓ Le contrôle d'accès logique aux systèmes d'informations lors du lancement d'un système d'exploitation ou lors de l'accès au réseau informatique.
- ✓ Verrouillage des équipements de communication comme les téléphones portables.

II.4 Le marché mondial de la biométrie

La croissance mondiale de la biométrie, portée par ses nombreuses applications, est incontestable. Bien qu'il existe peu d'informations publiques concernant ce marché, nous pouvons considérer certaines données relatant son évolution au fil des années, tant à l'échelle mondiale, qu'à l'échelle régionale (Américaine ou Européenne).

Régulièrement, un rapport sur le marché de la biométrie est édité par IBG (International Biometric Group). Ce rapport présente une analyse complète des chiffres d'affaires, des tendances de croissance et des développements industriels pour le marché de la biométrie actuel et futur. La lecture de ce rapport est essentielle pour des établissements

déployant la technologie biométrique, pour les investisseurs dans les entreprises biométriques et pour les développeurs de solutions biométriques.

Le chiffre d'affaires de l'industrie biométrique incluant les applications judiciaires et celles du secteur public, se développe rapidement comme l'indique la figure 1.1. Une grande partie de la croissance sera bientôt attribuable aux applications de contrôle d'accès aux systèmes d'information (ordinateur ou réseau) et au commerce électronique plutôt qu'aux applications du secteur public usuelles. On prévoit même que le chiffre d'affaires de ces marchés émergents dépasse celui de secteurs plus matures (identification de criminels et identification des citoyens) [Bio 08].

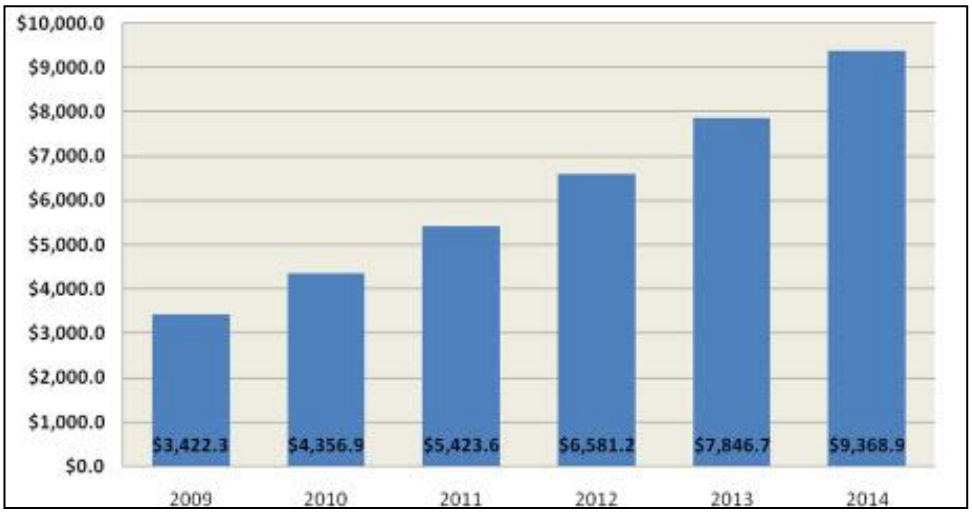


Figure 1.1 : Evolution du marché international de la biométrie [Bio 08]

III. LES SYSTEMES BIOMETRIQUES

III.1 Deux types de systèmes biométriques :

Bien qu'il existe de nombreuses applications à la biométrie, celle-ci ne présente, que deux modes de fonctionnement qui sont l'identification et l'authentification. [Fly 08]

III.1.1 L'identification

Un système d'identification permet de déterminer l'identité d'une personne parmi un ensemble de N personnes. L'utilisateur fournit un échantillon biométrique qui va subir des

prétraitements puis une extraction de paramètres pour en tirer l'information utile. Il sera ensuite comparé à tous les échantillons biométriques contenus dans la base de données du système. Celui-ci renvoie alors comme résultat la personne dans la base, dont l'échantillon testé se rapproche le plus comme le montre la figure 1.2.

En général, lorsque l'on parle d'identification, on suppose que le problème est fermé, c'est-à-dire que toute personne qui utilise le système possède un modèle dans la base de données.

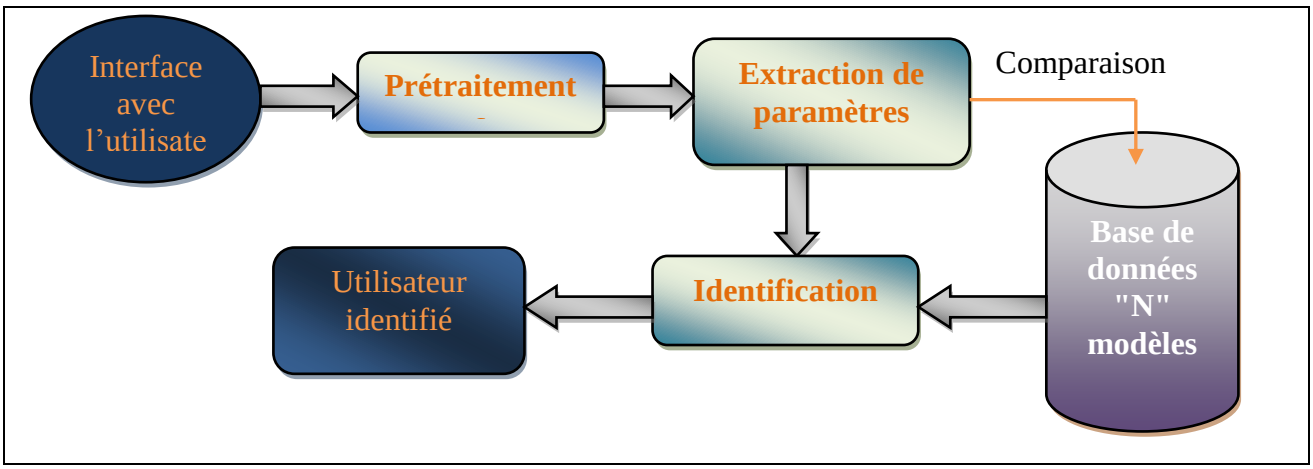


Figure 1.2 schéma bloc d'un système d'identification

III.1.2 L'authentification

Un Système d'authentification (ou de vérification) permet quant à lui de vérifier que l'identité proclamée par une personne est bien la sienne. Il comprend deux étapes :

Dans un premier temps l'utilisateur fournit un identifiant au système de reconnaissance (il proclame son identité). Ensuite, il fournit un échantillon biométrique qui va également subir des prétraitements et une extraction de paramètres avant d'être comparé à celui déjà présent dans la base de données et qui est associé à la personne correspondant à l'identifiant donné. Le système juge alors suivant le degré de rapprochement des deux échantillons si la personne est bien celle qu'elle prétend être (voir figure 1.3).

En mode vérification, on parle de problème ouvert puisque l'on suppose qu'un individu qui n'a pas de modèle dans la base de données (imposteur) peut chercher à être reconnu.

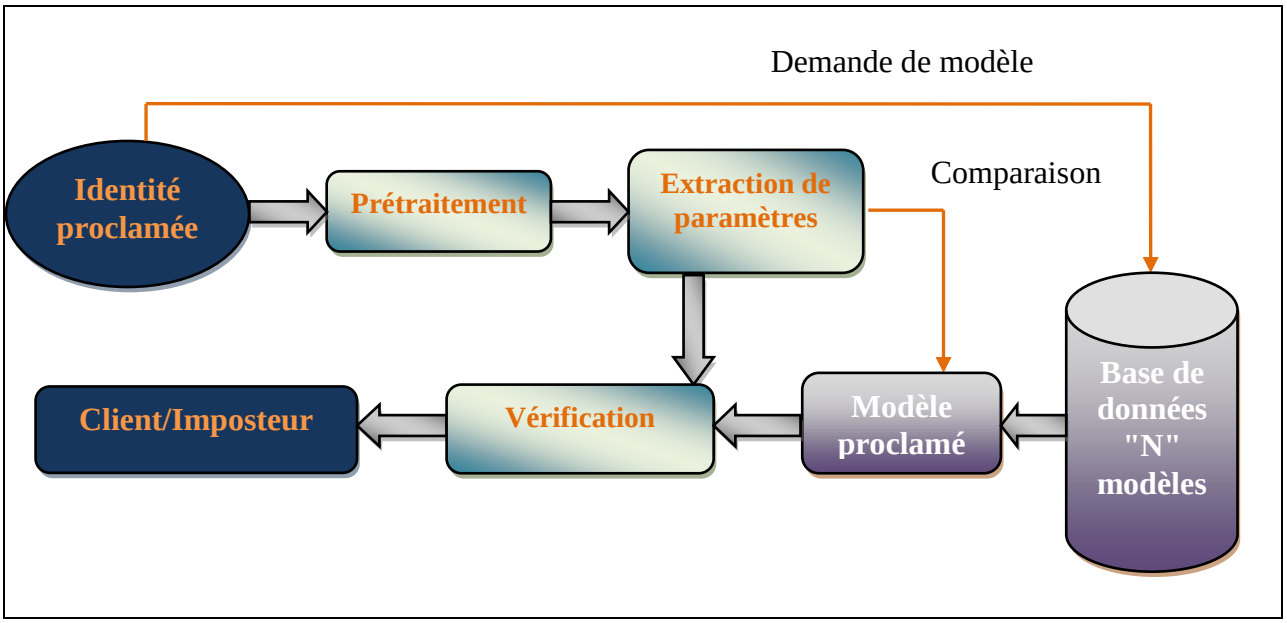


Figure 1.3 Schéma bloc d'un système de vérification

Nous avons donc deux systèmes qui ont des fonctions différentes, l'un d'eux répondant à la question « *qui suis-je ?* » et l'autre à la question *« suis-je bien la personne que je prétends être ? ».*

III.2 Architecture d'un système biométrique

L'architecture globale d'un système biométrique (voir figure 1.4) est généralement divisée en deux modules: un module d'apprentissage et un module de reconnaissance (un troisième module d'adaptation pouvant être ajouté).

Pendant la phase d'apprentissage, le système va traiter plusieurs données qui serviront à construire les modèles de chaque individu. Ceux-ci serviront de points de comparaison lors de l'étape de reconnaissance, qui varie selon le mode de fonctionnement, et pourront être réévalués après chaque utilisation grâce au module d'adaptation [Liu 01], [Jai 99].

III.2.1 Module d'apprentissage

L'apprentissage commence par l'acquisition des données relatives à chaque personne, en utilisant un capteur approprié. Ces données se voient appliquer une série de transformations qui ont pour but de faire ressortir *les informations utiles* (réorganisation des données) et *d'éliminer les informations inutiles* (compression des données), augmentant ainsi le pouvoir de discrimination et réduisant la complexité et donc le temps de traitement. Ces paramètres serviront par la suite à représenter chaque personne par un modèle mathématique, à l'aide d'un classificateur, qui sera stocké dans une base de données ou sur une carte à puce.

Il est à noter que la qualité du capteur peut grandement influencer sur les performances du système : meilleure est la qualité du système d'acquisition, moins il y aura de prétraitements à effectuer avant d'extraire les paramètres du signal.

III.2.2 Module de reconnaissance

Au cours de la reconnaissance, on renouvèle l'opération d'acquisition sur un échantillon à priori inconnu du système. De la même manière que précédemment, on procède à l'extraction des paramètres contenus dans cet échantillon. Le capteur utilisé doit, donc, avoir des propriétés aussi proches que possibles de celui utilisé durant la phase d'apprentissage.

La suite de la reconnaissance sera différente suivant le mode opératoire du système. Si le système procède à une identification, il devra chercher le modèle dont les nouveaux paramètres se rapprochent le plus. Indiquant ainsi la personne à laquelle l'échantillon appartient. Si, par contre, il travaille en mode vérification, Il devra estimer le degré de rapprochement entre les paramètres extraits et le modèle proclamé. Ce rapprochement sera généralement représenté par une distance ou une vraisemblance. Il devra alors comparer cette valeur à une valeur seuil, estimée par l'administrateur, pour vérifier si les paramètres extraits appartiennent à un client ou à un imposteur.

III.2.3 Module d'adaptation

Pendant la phase d'apprentissage, le système biométrique ne capture souvent que quelques instances d'un même attribut afin de limiter la gêne pour l'utilisateur. Il est donc difficile de construire un modèle assez général capable de décrire toutes les variations possibles de cet attribut. L'adaptation est donc nécessaire pour maintenir voir améliorer la performance d'un système après chaque utilisation. En effet, si un utilisateur est identifié par le module de reconnaissance, les paramètres extraits du signal serviront alors à ré-estimer son modèle.

Elle peut se faire en mode supervisé c'est-à-dire « manuellement » ou non-supervisé et donc « automatiquement ». Mais le second mode est de loin le plus utile en pratique. En général, la ré-estimation du modèle dépend du degré de confiance du module de reconnaissance dans les paramètres extraits.

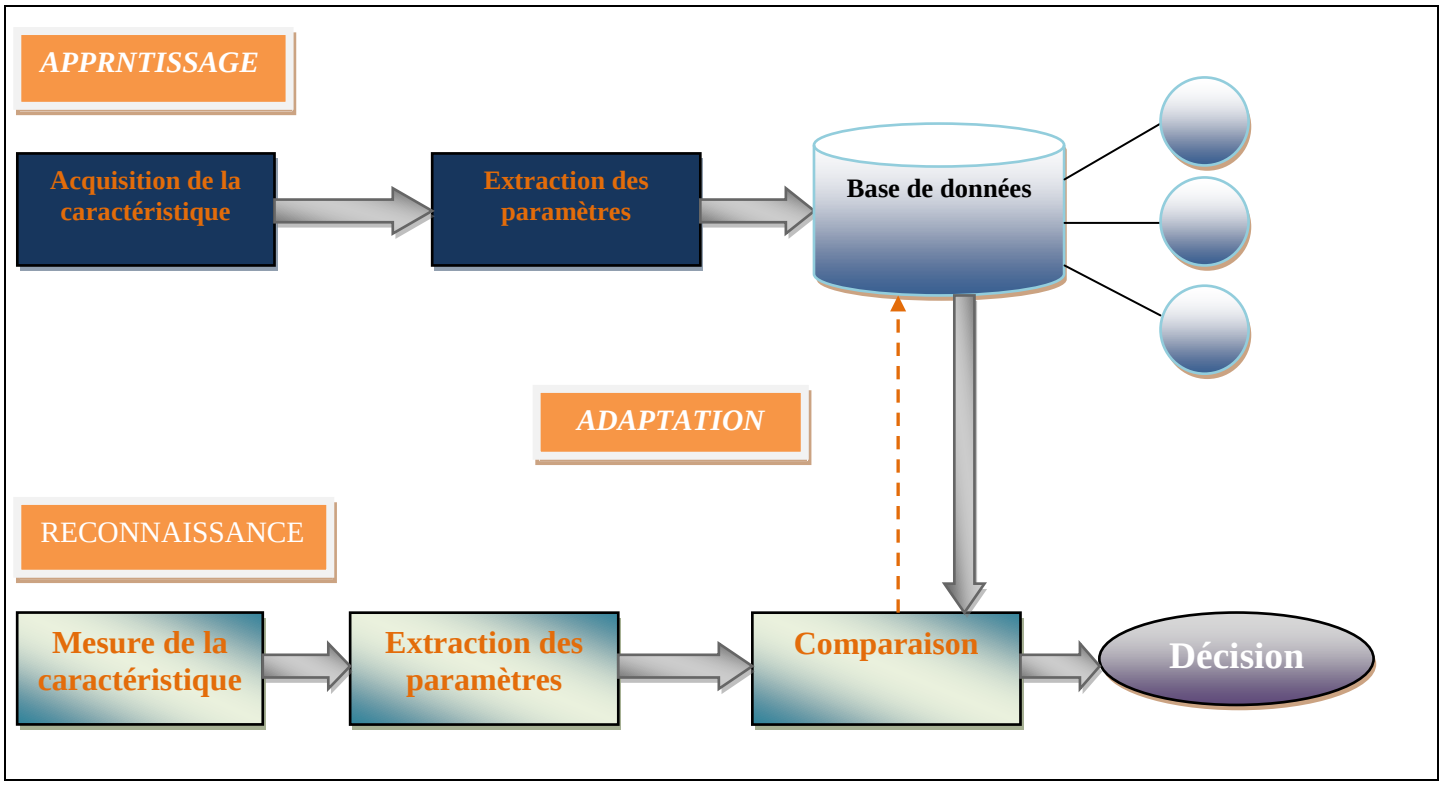


Figure 1.4 Architecture d'un système de reconnaissance biométrique

IV. EVALUATION DES PERFORMANCE D'UN SYSTEME BIOMETRIQUE

La performance d'un système d'identification se mesure principalement par sa précision. L'identification et la vérification étant des modes opératoires différents ils nécessitent donc des mesures de précision différentes.

IV.1 Evaluation de l'identification

L'erreur commise par ce genre de système est d'attribuer à l'individu présenté une identité autre que la sienne. Les performances de ce système sont mesurées à l'aide du taux d'identification (TID) suivant :

$$TID = \frac{nombre\ de\ tests\ ayant\ conduit\ \text{à}\ une\ identification\ correcte}{nombre\ total\ de\ tests} \qquad [1\text{-}1]$$

Ce paramètre dépend du nombre de personnes contenues dans la base de données. Plus la base est volumineuse (nombre de tests important), plus le taux d'erreurs risque d'être grand [Phi 00].

IV.2 Evaluation de la vérification

La vérification est un problème de décision similaire à la détection d'un signal dans le bruit en théorie de l'information. Si $H0$ est l'hypothèse que la capture C provienne d'un imposteur et $H1$ l'hypothèse que la capture C provienne de l'utilisateur légitime. On doit avoir :

$$\mathrm{P(H1|C)} > P(H0|C) \qquad [1\text{-}2]$$

Ce qui donne en appliquant la loi de Bayes (voir annexe 3):

$$\frac{P(C|H1)P(H1)}{P(C)} > \frac{P(C|H0)P(H0)}{P(C)} \qquad [1\text{-}3]$$

Qui devient :

$$\frac{P(C|H1)}{P(C|H0)} > \frac{P(H0)}{P(H1)} \qquad [1\text{-}4]$$

Le taux de vraisemblance (likelihood ratio) $\frac{P(C|H1)}{P(C|H0)}$ est comparé à un seuil θ appelé seuil de décision. Les valeurs $P(H0)$ et $P(H1)$ qui représentent respectivement la probabilité pour qu'un imposteur et un utilisateur légitime essayent d'accéder au système sont des valeurs difficiles à estimer. Le seuil est donc évalué expérimentalement.

Lorsqu'un système fonctionne en mode vérification il peut faire deux types d'erreurs. Il peut rejeter un utilisateur légitime et on parle de faux rejet (false rejection). Il peut aussi accepter un imposteur et on parle alors de fausse acceptation (false acceptance). La performance d'un système se mesure donc à son taux de faux rejets (False Rejection Rate ou FRR) et à son taux de fausses acceptations (False Acceptance Rate ou FAR). Ceux-ci sont donnés par :

$$\text{FAR} = \frac{nombre\ de\ fausses\ acceptations}{nombre\ d'imposteurs\ présentés} \qquad [1\text{-}5]$$

Ainsi que

$$\text{FRR} = \frac{nombre\ de\ faux\ rejets}{nombre\ de\ clients\ présentés} \qquad [1\text{-}6]$$

Nous avons représenté sur la figure 1.5 la distribution hypothétique des taux de vraisemblance qu'obtiendraient les utilisateurs légitimes et les imposteurs d'un système de vérification donné qui indique le nombre de fois où un taux de vraisemblance revient dans un ensemble de clients et d'imposteurs.

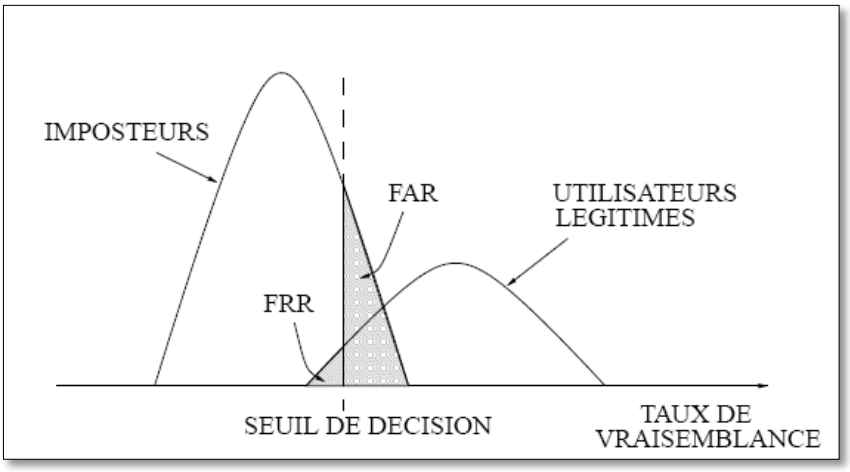


Figure 1.5-Distributions des taux de vraisemblance des utilisateurs légitimes et des imposteurs d'un système biométrique.

Les FAR et FRR sont représentés en hachuré. Idéalement, le système devrait avoir des FAR et FRR égaux à zéro. Comme ce n'est jamais le cas en pratique, il faut choisir un compromis entre FAR et FRR. Plus le seuil de décision θ est bas, plus le système acceptera d'utilisateurs légitimes mais plus il acceptera aussi d'imposteurs. Inversement, plus le seuil de décision θ (taux de vraisemblance dans la figure 1.5) est élevé, plus le système rejettera d'imposteurs mais plus il rejettera aussi d'utilisateurs légitimes.

La courbe dite ROC (Receiver Operating Characteristic), représentée à la figure 1.6, permet de représenter graphiquement la performance d'un système de vérification pour les différentes valeurs de θ. Le taux d'erreur égal EER (Equal Error Rate) correspond au point ou FAR = FRR, c'est-à-dire graphiquement à l'intersection de la courbe ROC avec la première bissectrice. Il est fréquemment utilisé pour donner un aperçu de la performance d'un système.

Le seuil θ doit donc être ajusté en fonction de l'application ciblée : haute sécurité, basse sécurité ou compromis entre les deux [Pra 02].

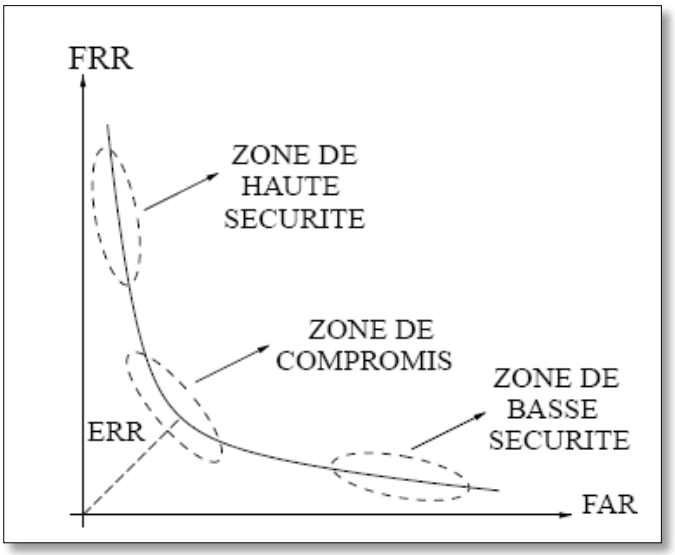


Figure 1.6-Courbe ROC

IV.3 Les différentes modalités

Le niveau de précision varie d'un système biométrique à l'autre, suivant les appareils et les méthodes utilisées. Mais c'est avant tout la modalité exploitée par un système qui influence ses performances.

Dans le cas idéal, les caractéristiques biométriques (ou modalités) utilisées pour une reconnaissance doivent être [Sui 04] :

- ✓ Robustes: elles ne doivent pas changer au cours du temps (permanence).
- ✓ Disponibles : elles doivent être communes a toute la population (Universalité).
- ✓ Distinctives pour la population : elles doivent varier considérablement d'une personne à l'autre (Unicité).
- ✓ Accessibles : elles doivent être faciles à acquérir.

Les modalités qui vérifient au mieux ces critères, peuvent être regroupées en trois catégories technologiques. La première est l'analyse biologique comme les tests portants sur le sang, l'ADN ou l'urine. La deuxième est l'analyse comportementale qui traite par exemple de la dynamique de la signature, de la façon d'utiliser un clavier ou encore de la manière de marcher. Et en dernier, l'analyse morphologique qui est la plus répandue maintenant et qui traite des empreintes digitales, de la forme de la main, des traits de visage, du dessin du réseau veineux de l'œil et de bien d'autres (voir figure 1.7) [Fly 08].

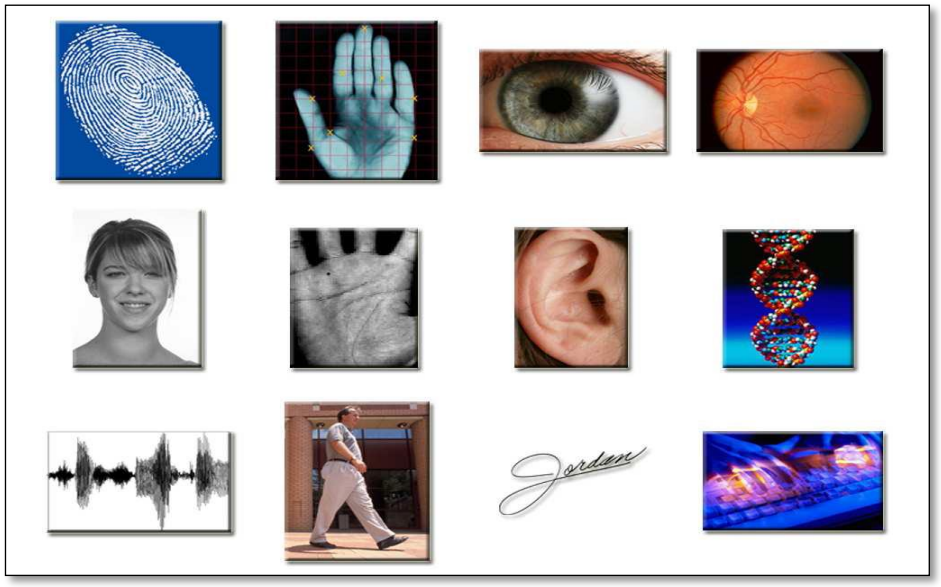


Figure 1.7 : les différentes modalités utilisées en biométrie

IV.4 Les systèmes portant sur analyse biologique

Ceux-ci portent sur l'analyse de caractéristiques génétiques difficiles à obtenir car nécessitant le prélèvement d'une partie du corps de la personne (morceau de peau, cheveux, sang…).

IV.4.1 L'analyse ADN (Acide Désoxyribonucléique)

L'analyse des empreintes génétiques est devenue en quelques années l'un des outils majeurs de la criminalistique (la science de l'identification des indices matériels dans un crime). L'analyse de l'ADN est couramment utilisée en criminologie pour identifier une personne à partir d'un morceau de peau, d'un cheveu ou d'une goutte de sang. Souvent les échantillons d'ADN trouvés sur le lieu du crime sont trop infimes pour être analysés. Mais il existe des appareils d'amplification en chaîne par polymérase (PCR) qui utilisent le même procédé naturel que l'ADN pour recopier et amplifier. Ils procurent ainsi aux criminalistes des brins d'ADN répliqués et exploitables.

IV.4.2 L'Analyse de l'ADN mitochondrial

Très proche de la technique précédente, cette technique identifie les personnes en utilisant l'ADN des mitochondries qui sont des organismes responsables de la production de

l'énergie dans nos cellules. Celui-ci n'est transmis que par la mère et est beaucoup plus unique et plus facile à avoir que l'ADN du noyau.

IV.5 Les systèmes portant sur analyse comportementale

Ces systèmes se basent sur des modalités en rapport avec le comportement des personnes, leur obtention est conditionnées par l'exécution d'un acte particulier (marcher, taper, signer…etc.).

IV.5.1 La signature dynamique

Il existe deux modes de reconnaissance de la signature : le mode statique et le mode dynamique.
Le mode statique n'utilise que l'information géométrique de la signature tandis que le mode dynamique utilise en plus l'information dynamique, c'est-à-dire les mesures des vitesses et des accélérations de la main lors de la signature.

Figure 1.8 Capture d'une signature

Le mode dynamique est plus riche en information que le mode statique et donc plus discriminant. De plus, il rend l'imitation beaucoup plus difficile à réaliser. L'utilisateur de cette technologie signe généralement avec un stylo électronique sur une tablette graphique (voir figure 1.6).

AVANTAGES	INCONVENIENTS
• Identification des personnes par cette méthode juridiquement acceptable. • Geste naturel accepté par le signataire.	• Difficulté à atteindre un niveau de performances élevé. • Dépendance de l'état émotionnel de la personne.

Tableau 1.1-Avantages et inconvénients des systèmes de reconnaissance basés sur la signature dynamique

IV.5.2 Dynamique de frappe au clavier

La dynamique de frappe au clavier est une caractéristique de chaque individu, c'est en quelque sorte la transposition de la graphologie aux moyens électroniques. Un système basé sur cette dynamique ne nécessite aucun équipement particulier, seulement un ordinateur disposant d'un clavier. Il s'agit d'un dispositif logiciel qui calcule le temps où un doigt effectue une pression sur une touche et le temps où un doigt est dans les airs (entre les frappes).

AVANTAGES	INCONVENIENTS
• Moyen non intrusif qui exploite un geste naturel.	• Dépendance de l'état physique et moral de la personne (âge, maladie, ..).

Tableau 1.2-Avantages et inconvénients des systèmes de reconnaissance basés sur la dynamique de frappe au clavier.

IV.5.3 Analyse de la démarche

Il s'agit de reconnaître un individu par sa façon de marcher et de bouger (vitesse, accélération, mouvements du corps…), en analysant des séquences d'images. La démarche serait en effet étroitement associée à la musculature naturelle et donc très personnelle. Son inconvénient majeur est qu'elle est sensible aux changements d'habits, chaussures et surface. Ce qui rend cette approche, actuellement, limitée au monde de la recherche seulement.

IV.6 Les systèmes portant sur analyse morphologique

Ces systèmes se basent sur des modalités liées à la morphologie des personnes. Elles sont toujours présentes indépendamment de la volonté de la personne testée.

IV.6.1 La forme de la main

Son principe consiste à placer la main de l'utilisateur sur un gabarit (voir figure 1.9) qui sera éclairé par une lumière infrarouge et l'image résultante sera ensuite captée par une caméra digitale. Ces données concernent la longueur, la largeur et l'épaisseur de la main, de même que la forme des articulations et longueur inter-articulations.

Figure 1.9-Scan de la forme de la main.

AVANTAGES	INCONVENIENTS
• Résultat indépendant de l'humidité des doigts et des souillures car il n'y a pas de contact direct avec le capteur. • Enrôlement facile et accepté. • Faible volume de stockage par fichier.	• Système encombrant. • Risque élevé d'erreurs [San 00]. • Technique qui n'a pas évolué depuis plusieurs années. • Lecteur plus cher que pour les autres types de captures.

Tableau 1.3-Avantages et inconvénients des systèmes de reconnaissance basés sur la forme de la main.

IV.6.2 Les empreintes digitales

Une empreinte digitale est constituée d'un ensemble de lignes localement parallèles formant un motif unique pour chaque individu. On distingue les stries ou crêtes qui sont les lignes en contact avec une surface au toucher et les sillons qui sont les creux entre ces stries. Les systèmes de capture des empreintes digitales peuvent être optiques, capacitifs ou à ultrasons. [Chu 01].

La technologie optique nécessite que l'utilisateur place un ou plusieurs doigts sur une vitre (voir figure 1.8), à travers laquelle l'image recherchée est mise sous éclairage et capturée par une caméra CMD (Charge Modulation Device) avec un dispositif de transfert de charge CCD (Charge-Coupled Device) qui convertit l'image, composée de crêtes foncées et de vallées claires, en un signal vidéo retraité afin d'obtenir une image utilisable.

AVANTAGES	INCONVENIENTS
• Ancienneté et mise à l'épreuve. • Résistance aux changements de température, jusqu'à un certain point. • Coût abordable. • Capacité à fournir de bonnes résolutions.	• Possibilité de dégradation de l'image par surimpression . • Apparition possible de rayures sur la fenêtre. • Usure du dispositif CCD avec le temps.

Tableau 1.4-Avantages et inconvénients des systèmes de reconnaissance basés sur les empreintes digitales dans le cas de la technologie optique.

La technologie capacitive, elle, consiste à effectuer l'analyse du champ électrique de l'empreinte digitale pour déterminer sa composition. L'utilisateur place ses doigts directement sur un microprocesseur spécialisé. L'image est transférée à un convertisseur analogique-numérique, l'intégration se faisant en une seule puce. Cette technique produit des images de meilleure qualité avec une surface de contact moindre par rapport à la technique optique. Les données fournies sont très détaillées. Elle possède une bonne résistance dans des conditions non-optimales.

AVANTAGES	INCONVENIENTS
• Cout assez bas.	• Capteur vulnérable aux attaques extérieures fortuites ou volontaires.

Tableau 1.5-Avantages et inconvénients des systèmes de reconnaissance basé sur les empreintes digitales dans le cas de la technologie capacitive.

Quant à la technologie ultrason, elle repose sur la transmission d'ondes acoustiques et mesure l'impédance entre le doigt, le capteur et l'air. Cette dernière permet de dépasser le problème lié aux résidus sur le doigt ou le capteur.

AVANTAGES	INCONVENIENTS
• Facilité d'usage avec de grandes plaques. • Capacité à surmonter des conditions de lecture non optimales.	• Cout élevé.

Tableau 1.6-Avantages et inconvénients des systèmes de reconnaissance basés sur les empreintes digitales dans le cas de la technologie ultrason.

À l'aide de l'un de ces mécanismes, plusieurs caractéristiques uniques à chaque individu que sont les boucles, les tourbillons, les lignes et les verticilles (cercle concentrique au centre d'un doigt) des empreintes sont localisées, situées les unes par rapport aux autres et enregistrées selon plusieurs modèles dans une base de données.

Figure 1.10-Capture d'une empreinte.

Ainsi une personne testée se verra extraire ses minuties à l'aide d'un AFIS, minuties qui seront comparées à celles de la base de données. Malgré son taux de précision très élevé, la reconnaissance d'individus par empreintes digitales est une méthode mal acceptée par les utilisateurs à cause de l'association qui est souvent faite avec la ciminalistique. [Jai 01].

IV.6.3 L'Iris

L'iris est la région annulaire située entre la pupille et le blanc de l'œil, ses motifs ne se forment qu'au cours des deux premières années de la vie et elles sont stables et non modifiables même par des interventions chirurgicales. Elle présente également l'avantage d'être unique puisque la probabilité de trouver deux iris identiques est inférieure à l'inverse du nombre d'humains ayant vécu sur terre.

Figure 1.11-Capture de l'image d'un iris.

L'acquisition de l'iris est effectuée au moyen d'une caméra (voir figure 1 .9) pour pallier aux mouvements inévitables de la pupille. Elle est très sensible (précision, reflet…) et relativement désagréable pour l'utilisateur car l'œil doit rester grand ouvert et il est éclairé par une source lumineuse pour assurer un contraste correct [Dau 93].

AVANTAGES	INCONVENIENTS
• Technique extrêmement fiable car il contient une infinité de points caractéristiques (ensemble fractal). • Pas de risque identifié pour la santé.	• Fraude possible en utilisant des lentilles. • Relativement désagréable pour l'utilisateur car l'œil doit rester grand ouvert. • Contrainte d'éclairage pour assurer un contraste correct.

Tableau 1.7-Avantages et inconvénients des systèmes de reconnaissance analysant l'iris.

IV.6.4 La rétine

La reconnaissance de la rétine est actuellement considérée comme l'une des méthodes biométriques les plus sûres. Elle se base sur les motifs que dessinent les veines sous sa surface qui sont uniques et stables dans le temps. Ils ne peuvent être affectés que par certaines maladies très rares.

Son principe consiste à placer l'œil de l'utilisateur à quelques centimètres d'un orifice de capture situé sur le lecteur de rétine. Il ne doit pas bouger et doit fixer un point vert lumineux qui effectue des rotations. À ce moment, un faisceau lumineux traverse l'œil jusqu'aux vaisseaux sanguins de la rétine. Le système localise et capture ainsi environ 400 points de référence. Ces points représentent le modèle de la personne [Dau 93].

AVANTAGES	INCONVENIENTS
• Résistance à la fraude. • Unicité même chez les vrais jumeaux. • Technique fiable. • Même cartographie de la rétine tout au long de la vie.	• Nécessité de placer ses yeux à très faible distance du capteur donc système intrusif mal accepté psychologiquement. • Cout élevé. • Difficulté d'utilisation en cas de contrôle d'une population importante (temps important). • Installation délicate.

Tableau 1.8-Avantages et inconvénients des systèmes de reconnaissance analysant la rétine.

IV.6.5 Le visage

Francis Galton (1822-1911) instaura les prémices de ce que devrait être la reconnaissance faciale dès 1888 dans son ouvrage « Personnal identification and description ». Quelques publications sont ensuite apparues au cours des années 60-70, mais ce n'est qu'à partir du milieu des années 80, lorsque la puissance des ordinateurs est devenue

suffisante, que les recherches les plus poussées ont commencé. Nous avons vu alors apparaître des systèmes fonctionnels et des sociétés qui se sont formées pour exploiter les algorithmes développés dans les laboratoires les plus prestigieux. Depuis le 11 septembre 2001, un intérêt tout particulier est porté sur cette technologie car elle présente de nombreux intérêts, dont celui de pouvoir effectuer une identification à distance. Il ne faut cependant pas oublier les limites technologiques des systèmes actuels qui nécessitent pour l'instant un environnement (éclairage, position de la caméra, etc.) bien contrôlé (voir figure 1.11) pour pouvoir pleinement exprimer leurs performances.

Cette technologie est employée dans des domaines aussi divers que le contrôle d'accès physique ou logique, la surveillance ou l'accès aux distributeurs automatiques de billets. Mais en veillant à utiliser le scénario le plus adapté pour ne pas gêner les utilisateurs, car le résultat restera toujours une identité probable [Zha 00].

AVANTAGES	INCONVENIENTS
• Modalité acceptée par le public. • Ne demande aucune action de l'usager. • N'est pas intrusive. • Ne nécessite pas de contact physique avec le capteur. • Peu coûteuse. • Peu encombrante.	• Sensible à l'environnement (éclairage, position, expression du visage...) et aux changements (barbe, moustache, lunette, chirurgie...).

Tableau 1.9-Avantages et inconvénients des systèmes de reconnaissance analysant le visage.

IV.7 le cas de la voix

C'est en 1962 que Lawrence Kersta, un ingénieur du Bell Laboratoires, a établit que la voix de chaque personne est unique et qu'il est possible de la représenter graphiquement, et c'est dans les années 80 que les premiers systèmes de reconnaissance vocale apparaissent.

La reconnaissance du locuteur vise à déterminer les caractéristiques uniques de la voix de chaque individu. Bien que généralement classée comme *caractéristique comportementale*, la voix se trouve à la frontière avec les *caractéristiques morphologiques*. En effet, une grande

partie de cette caractéristique est déterminée par le conduit vocal ainsi que par les cavités buccales et nasales. Mais aussi par la manière de parler des individus.

D'autre part, la voix n'est pas un attribut permanent. Elle change bien entendu avec l'âge mais peut être aussi affectée temporairement par l'état de santé ou émotionnel du locuteur.

De nos jours, tous les ordinateurs sont équipés en standard d'un microphone ce qui explique la popularité de la reconnaissance du locuteur pour les applications de type «desktop ». Il existe différentes modalités de reconnaissance du locuteur suivant que le texte soit fixé ou pas que nous verrons plus en détails dans le chapitre 3.

AVANTAGES	INCONVENIENTS
• Disponible via le réseau téléphonique. • Non intrusif. • Il est quasi impossible d'imiter la voix stockée dans la base de données car les imitateurs utilisent les caractéristiques vocales sensibles au système auditif humain mais ne sont pas capables de recréer les harmonique de la voix en question.	• Sensibilité à l'état physique et émotionnel d'un individu. • Sensibilité aux conditions d'enregistrement du signal de parole : bruit ambiant, parasites, qualité du microphone utilisé.

Tableau 1.10-Avantages et inconvénients des systèmes de reconnaissance analysant la voix.

IV.8 Comparaison entre les différents systèmes biométriques

Chaque technologie et procédé biométrique vu précédemment possède des avantages mais aussi des inconvénients, acceptables ou inacceptables suivant les applications. Ces technologies n'offrent pas les mêmes niveaux de sécurité ni les mêmes facilités d'emploi et nécessitent des couts différents pour être mis en place.

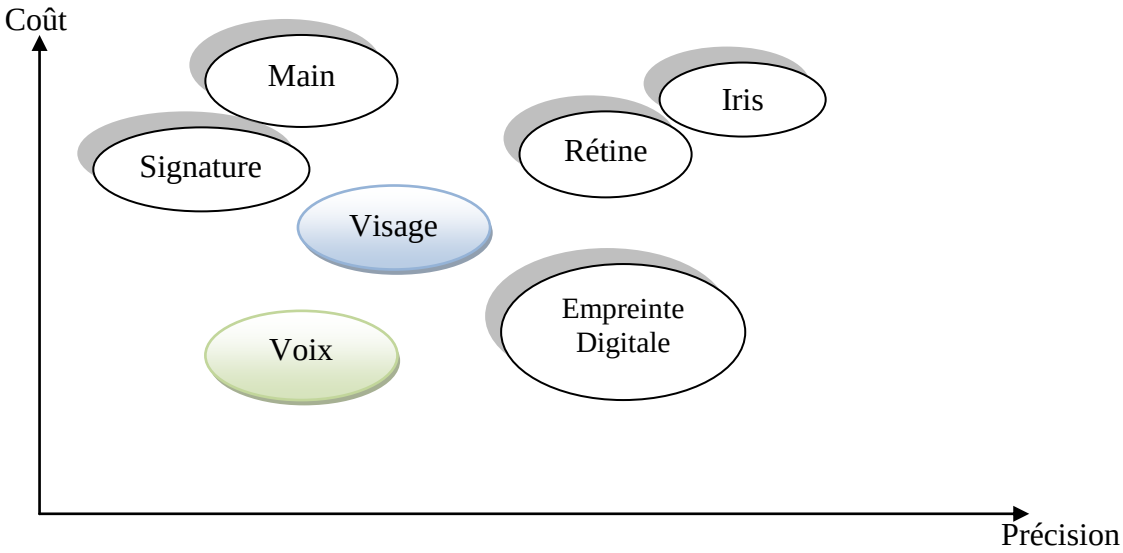


Figure 1.12-Comparaison des techniques biométriques les plus utilisées en fonction des coûts et de la précision [Mal 03]

Comme on peut le voir sur la figure 1.12, plus une modalité est précise, plus le coût de sa mise en place est élevé. Cela dit, certaines modalités comme la signature et la forme de la main nécessitent des couts élevés pour une précision faible, elles auront donc tendance à être écartées. Tandis que des modalités comme les empreintes digitales qui sont précises mais ont un faible cout auront tendance à être très utilisées.

Biométrie	*Visage*	*Empreinte*	*Main*	*Iris*	*Rétine*	*Signature*	*Voix*
Universalité	haute	moyenne	moyenne	haute	haute	faible	moyenne
Unicité	faible	haute	moyenne	haute	haute	faible	moyenne
Permanence	moyenne	haute	moyenne	haute	moyenne	faible	moyenne
Mesurabilité	haute	moyenne	haute	moyenne	faible	haute	haute
Performance	moyenne	haute	moyenne	haute	haute	faible	moyenne
Acceptabilité	haute	moyenne	moyenne	faible	faible	haute	haute
Circonvention	haute	moyenne	moyenne	faible	faible	haute	haute

Tableau 1.11- Comparaison des différentes technologies biométriques [Mal 03]

A partir de ces sept critères (Tableau 1.11), une première comparaison des principales technologies biométriques est proposée à travers le tableau. Parmi les techniques les plus matures, on distingue le visage, l'empreinte digitale, la géométrie de la main, la voix, l'iris et la rétine, qui présentent des bonnes caractéristiques. Aussi nous sommes tentés de dire que ces cinq solutions biométriques ne sont pas systématiquement en concurrence. Cela dit, pour des applications grand public, la reconnaissance rétinienne, qui nécessite un appareillage d'acquisition sophistiqué et coûteux, peut être d'ores et déjà écartée car trop intrusive

IV.9 Les parts de marché par technologie

Selon la précision qu'elles ont et les couts qu'elle nécessite, la fréquence d'utilisation diffère d'une modalité à l'autre. Les empreintes digitales continuent à être la principale technologie biométrique en termes de part de marché, près de 50% du chiffre d'affaires total (hors applications judiciaires), ce qui est du a leur une très bonne précision malgré un faible cout. La reconnaissance du visage, avec 12% du marché (hors applications judiciaires), dépasse la reconnaissance de la main, qui avait avant la deuxième place en termes de source de revenus après les empreintes digitales [Bio 08]. Et c'est ce que nous pouvons constater dans cette figure :

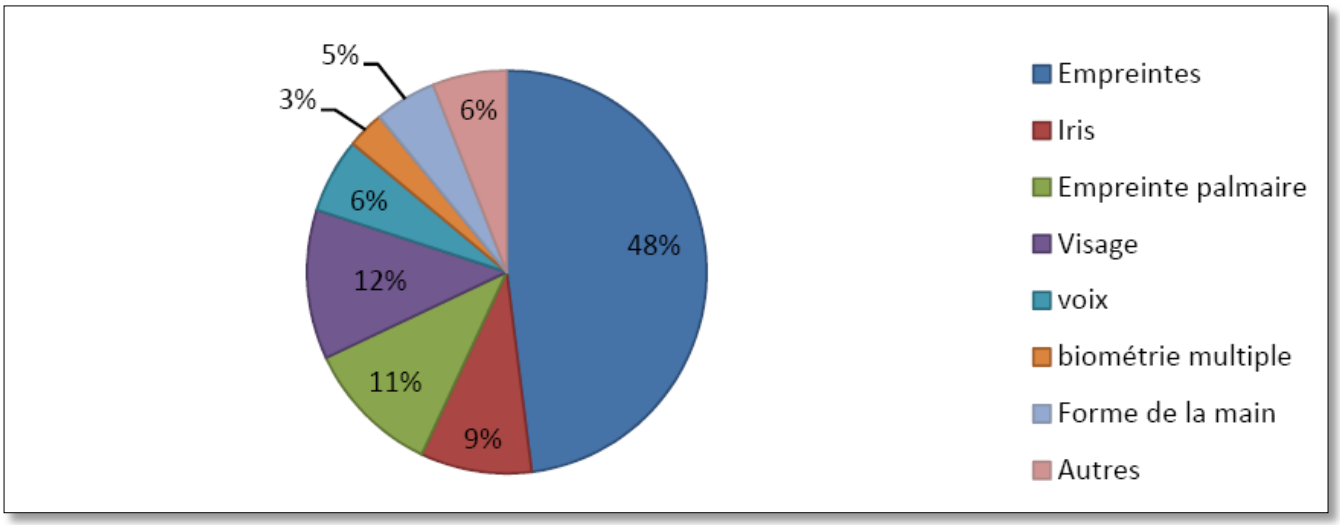


Figure 1.13-Parts de marché des différentes méthodes biométriques [Bio 08]

V. LA MULTI MODALITE DANS LA BIOMETRIE

Bien que de nos jours il existe des techniques biométriques extrêmement fiables telles que la reconnaissance de la rétine ou de l'iris, elles sont coûteuses et, en général, mal acceptées par le grand public et ne peuvent donc être réservées qu'à des applications de très haute sécurité. Pour les autres applications, des techniques telles que la reconnaissance du visage ou de la voix sont très bien acceptées par les utilisateurs et ont des performances satisfaisantes pour être déployées dans des conditions réelles [Ver 99].

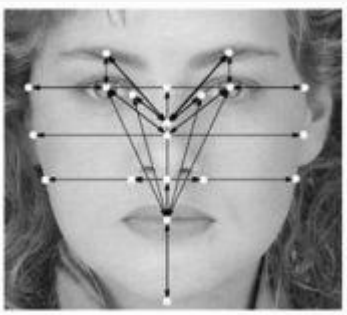
Afin d'améliorer la sécurité des systèmes précédents, une première solution consiste à intégrer la biométrie avec l'identification basée sur une connaissance ou une possession. Cette méthode permet d'améliorer la sécurité du système, mais elle possède les faiblesses inhérentes à l'identification basée sur une connaissance ou une possession.

La multimodalité, qui est la réalisation de systèmes de reconnaissance automatique en utilisant plusieurs modalités complémentaires simultanément, est une alternative qui permet d'améliorer de manière systématique la performance d'un système biométrique. Par performance, nous entendons à la fois la précision du système mais aussi son efficacité, plus particulièrement en mode identification [Kit 26]. En effet, des classificateurs différents font en général des erreurs différentes, et il est possible de tirer parti de cette complémentarité afin d'améliorer la performance globale du système [Hon 99]. Nous verrons la multimodalité plus en détail dans la suite de notre travail (voir chapitre 5).

VI. CONCLUSION :

A travers ce chapitre nous avons introduit le concept de système biométrique en présentant de manière globale les différentes modalités que ce dernier utilise. Cette approche inclue un passage par l'architecture du système ainsi que ces différentes applications en mettant en évidence ces performances et les inconvénients qui en résultent suivant la modalité choisie. A partir de ce dernier point, nous avons brièvement parlé de la ***multimodalité*** qui s'avère être une solution des plus appréciées pour palier aux différents problèmes que peuvent rencontrer les systèmes unimodaux.

2 Systèmes de reconnaissance du visage



I. INTRODUCTION

Par la fréquence à laquelle on le rencontre dans l'environnement et par son contenu riche en informations sociales de premier ordre, le visage humain constitue un stimulus visuel de classe à part. En effet, il suffit d'un clin d'œil porté sur le visage d'un individu pour en distinguer le sexe, l'état émotionnel ou l'identité.

Cette grande capacité à identifier les visages que possèdent les hommes a poussé les chercheurs à vouloir se rapprocher du cerveau humain dans sa rapidité, son exactitude et sa fiabilité par des systèmes de reconnaissance du visage.

Mais bien que tout observateur humain se montre capable d'identifier un nombre apparemment infini de visages, seules de fines discriminations visuelles permettent de les identifier par un système de reconnaissance automatique.

L'utilisation d'une telle modalité s'est avéré très intéressante et très rapide ce qui est d'autant plus surprenant que chaque visage est composé des mêmes attributs (yeux, nez, bouche) disposés selon une organisation similaire. Néanmoins, elle reste légèrement moins intéressante que l'utilisation des modalités telles que la rétine, l'iris ou les empreintes digitales même si elle est acceptée et facile à acquérir. Ceci est principalement du à plusieurs facteurs liés à l'environnement dans lequel la photo a été prise.

II. PROCESSUS DE RECONNAISSANCE AUTOMATIQUE DU VISAGE

Tout processus automatique de reconnaissance du visage doit prendre en compte plusieurs facteurs qui contribuent à la complexité de sa tâche, car le visage est une entité dynamique qui change constamment sous l'influence du monde extérieur.

La Figure 2.1 illustre la démarche générale adoptée pour réaliser de tels systèmes [Zia 01].

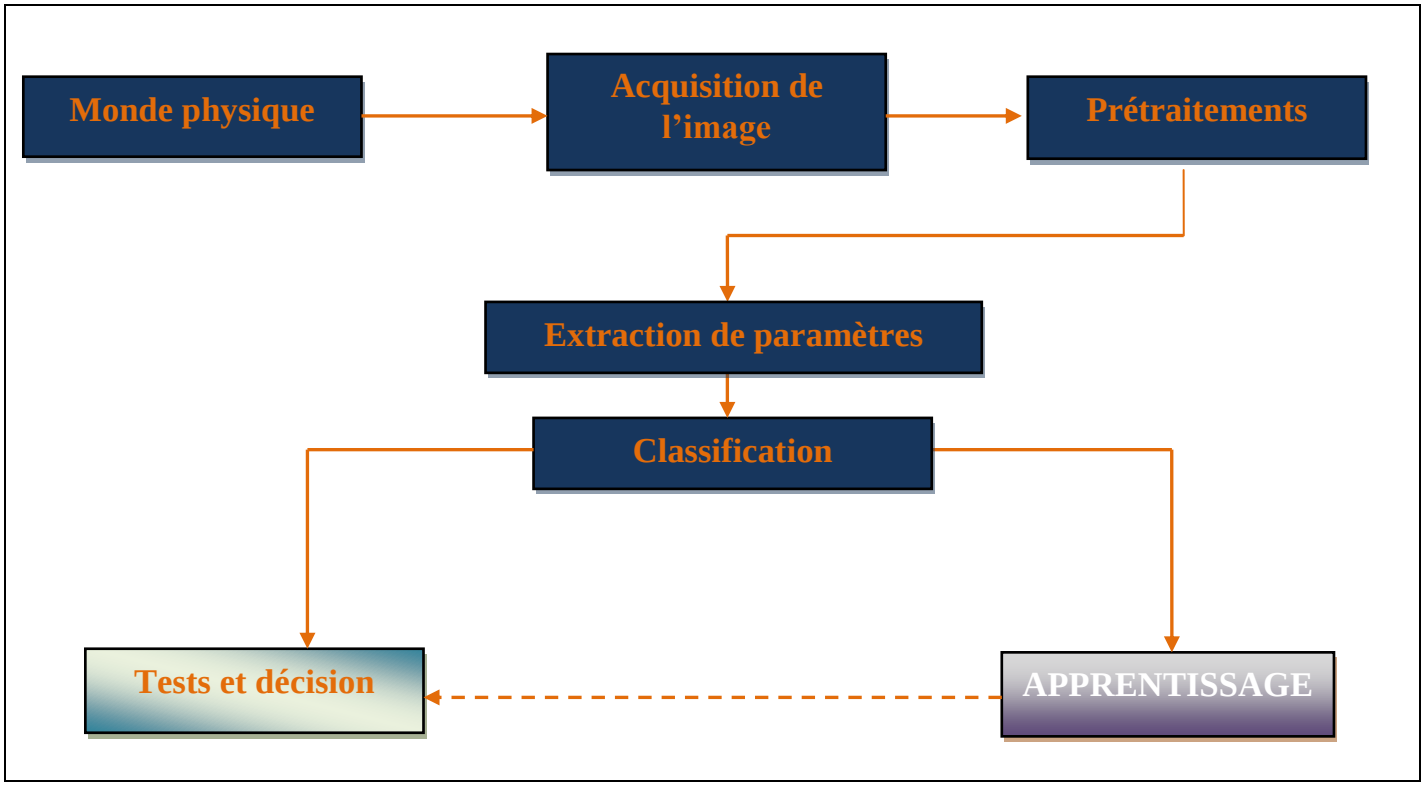


Figure 2.1-Le processus de reconnaissance de visage

Donc lorsqu'une personne se présente devant un système de reconnaissance automatique du visage, son image suit le processus suivant :

II.1 Le monde physique

C'est le monde réel en dehors du système avant l'acquisition de l'image. Dans cette étape, on tient compte généralement de trois paramètres essentiels : L'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents, et par conséquent une mauvaise décision.

II.2 L'Acquisition de l'image

Cette étape consiste à extraire l'image de l'utilisateur du monde extérieur dans un état statique à l'aide d'un appareil photo ou dynamique à l'aide d'une caméra. Après quoi, l'image extraite est digitalisée ce qui donne lieu à une représentation bidimensionnelle du visage, caractérisée par une matrice de niveaux de gris (Voir Annexe Traitement d'image). L'image

dans cette étape est dans un état brut ce qui engendre un risque de bruit qui peut dégrader les performances du système.

II.3 Les prétraitements

Le rôle de cette étape est d'éliminer les parasites causés par la qualité des dispositifs optiques ou électroniques lors de l'acquisition de l'image en entrée et ceci par des techniques de traitement et de restauration d'images dans le but d'homogénéiser les images des différentes personnes entre elles et avec celles utilisées lors du test.

Il existe plusieurs types de traitements et d'améliorations de la qualité de l'image, telle que : la normalisation, l'égalisation et le filtre médian (Annexe A).
Cette étape dépend fortement du contexte dans lequel le système est utilisé (cadre fixe ou variable, visages proches du dispositif ou éloignés). Elle peut même inclure, dans certains cas, la détection et la localisation du visage dans les images lorsque le décor est très complexe.

II.4 L'extraction de paramètres

L'extraction des paramètres consiste à extraire l'information utile des images. Ceci dans le but de faire ressortir les informations *utiles et discriminantes* et de réduire les dimensions à traiter en éliminant la redondance pour augmenter la rapidité de réponse du système.

II.5 Classification (Modélisation)

Les paramètres extraits sont ensuite regroupés par classes au sein d'un même individu, pour décrire au mieux ses variabilités à l'aide de ce qu'on appelle un classificateur. On obtient alors un modèle mathématique pour chaque personne.

II.6 La décision :

La décision, qui est l'aboutissement du processus, compare de nouvelles données recueillies par le système et d'origine inconnue et les compare aux modèles mémorisés pour prendre une décision (une identification de l'individu ou une vérification de son identité proclamée).

III. EXTRACTION DES PARAMETRES DCT

III.1 Introduction

Dans un système de reconnaissance de visages, l'image ne pouvant pas être traitée sous forme d'une matrice de niveaux de gris de milliers de pixels à cause de la complexité et de la lourdeur des calculs, cette dernière doit subir une série de transformations orthogonales afin d'éliminer les redondances et donc de ne conserver que l'information utile dans un nombre minimum de coefficients.

Plusieurs transformations sont utilisées dans ce genre de systèmes, les plus répandues sont la transformée de Fourier (FT), la transformée de Karhunen-Loève (KLT) et celle en Cosinus Discrète (DCT) qui sera utilisée dans le cadre de notre travail [Kha 03].

III.2 Présentation de la DCT (Discrete Cosine Transform)

La Transformée en Cosinus Discrète est une transformée linéaire qui a été appliquée la 1ère fois dans la publication des professeurs N. Ahmed, T. Natarajan et K. R. Rao [Ahm 74]. C'est une variante de la transformée de Fourier discrète, qui permet de garder uniquement les cosinus et d'éliminer les sinus dans le but d'obtenir une représentation fréquentielle purement réelle [Ben 09].

Il est à noter que cette transformée est très largement utilisée dans la compression audio et la compression d'images. En effet la DCT est utilisée dans le traitement de signal, le traitement d'image et incarne l'idée majeure sur laquelle est basée la compression JPEG.

III.3 Motivation quant au choix de la DCT

La transformée de Karhunen-Loève (KLT) est reconnue pour sa précision et son exactitude mais elle présente un inconvénient majeur causé par la complexité de ses calculs. La transformée de Fourier discrète (DFT) quand à elle est une méthode très rapide grâce à la simplicité de ses algorithmes mais sa périodicité horizontalement et verticalement cause son imprécision dans certain cas.

- ✓ comparaison entre la DCT et la KLT : Comme la DCT, la KLT est une transformation linéaire ou les fonctions de base sont tirées à partir des propriétés statistiques de l'image, elle est optimale dans sa manière de compactage d'énergie car elle introduit un maximum d'énergie dans un nombre réduit de coefficients (voir annexe). Cependant, la KLT présente généralement une dépendance directe avec les données en entrée, ce qui explique sa lourdeur (ses vecteurs de base doivent être recalculés à chaque fois que l'on change de données), contrairement à la DCT qui représente une indépendance totale de ses coefficients par rapport aux données en plus du fait que ses fonctions de base soient pré calculées ce qui réduit considérablement son temps de calcul.

- ✓ comparaison entre la DCT et la DFT : Chacune des deux transformées montrent de bonnes caractéristiques de concentration et de décorrélation d'énergie. Cependant, malgré le fait que la DFT soit une transformation linéaire séparable et symétrique, elle agit sur le domaine complexe ce qui représente des paramètres en plus à gérer. De plus elle rencontre un problème de discontinuité causé par un phénomène de Gibbs ce qui entraîne une déstabilisation lors du codage de l'image dû à l'effet de bord.

La transformée en cosinus discrète est donc venue pour équilibrer le compromis existant entre la précision et la vitesse, elle a allie la précision de la KLT et la performance de la DFT grâce à ses algorithmes simples et efficaces [Zia 01].

III.4 Les coefficients DCT

Dans une image naturelle la majorité des informations sont concentrées dans les basses fréquences. La transformation de l'image par la DCT va donc faire apparaître une occupation spectrale réduite, où une zone de taille relativement petite code une grande partie de l'information.

La DCT, appliquée à l'image graphique, transforme le signal discret bidimensionnel, d'amplitude donnée en niveau de gris, en une information bidimensionnelle de fréquences. Ce changement de domaine s'avère être très intéressant du fait que l'image classique admette une

grande continuité entre les valeurs des pixels, d'où un changement relativement lent de ces dernières.

De plus la sensibilité visuelle de l'être humain n'est pas linéaire mais logarithmique. Celle-ci est bien plus importante pour les basses fréquences que pour les hautes fréquences. Ainsi, nous parvenons à représenter l'intégralité de l'information de l'image sur très peu de coefficients ce qui permet une optimisation des calculs sans affecter de manière significative la perception que l'on a de l'image [Zia 01].

III.5 Principe et formulation mathématique de la DCT

Le principe de la DCT se base sur trois points essentiels : décorréler les informations portées par les pixels d'une image, les introduire dans un minimum de coefficients et les localiser dans une zone d'acuité visuelle minimale.

Cette dernière s'applique à une matrice carrée. Le résultat fournit est représenté dans une matrice de même dimension. Les coefficients correspondants aux basses fréquences se trouvant en haut à gauche de la matrice, et les hautes fréquences en bas à droite.

Il existe 8 variantes de la DCT, les plus connues sont DCT I, la DCT 2D ou DCT II et la transformée inverse IDCT ou DCT III [Pen 93].

III.5.1 La DCT à une dimension

La transformée en cosinus discrète à une dimension d'une séquence de données f(x) de longueur N est une suite d'éléments C(u) donnée par la formule :

$$c(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \qquad [2\text{-}1]$$

Avec $\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & pour\ u = 0 \\ \sqrt{\frac{2}{N}} & pour\ u \neq 0 \end{cases}$ et $u = 0, 1, 2, ..., N\text{-}1.$

On remarque que le premier élément de cette transformée représente la moyenne de f(x), il est connu sous le nom de coefficient DC. Les autres éléments sont nommés les coefficients AC. D'après cette écriture, on voit qu'on procède en réalité à une projection des données sur la base des cosinus.

III.5.2 La DCT à deux dimensions

C'est une extension directe de la DCT I dans deux dimensions 2D,N × N , qui seras notre image, sa formule mathématique est donnée par :

$$c(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \qquad [2\text{-}2]$$

Avec

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & pour\ u = 0 \\ \sqrt{\frac{2}{N}} & pour\ u \neq 0 \end{cases} ; \quad \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & pour\ v = 0 \\ \sqrt{\frac{2}{N}} & pour\ v \neq 0 \end{cases}$$

Pour $u, v = 0, 1, 2, ..., N\text{-}1.$

Notons que la DCT à deux dimensions peut être obtenue en multipliant le résultat de deux DCTs horizontale et verticale respectivement.

Comme dans la DCT à une dimension, le premier élément (u=0, v=0) est appelé le coefficient DC et représente la moyenne des intensités du bloc à transformer, et les autres (pour u≠0 et v≠0) sont appelés les coefficients AC.

III.5.3 La DCT inverse :

Désignée DCT III ou IDCT c'est l'inverse de la DCT II, sa formule est donnée par :

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c(u,v)\alpha(u)\alpha(v)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \qquad [2\text{-}3]$$

Pour $u, v = 0, 1, 2, \ldots, N\text{-}1$.

III.6 Propriétés de la DCT :

La DCT est caractérisée par plusieurs propriétés qui permettent sa rapidité et son efficacité de compression [Kha 03] :

III.6.1 Décorrélation

Le principal avantage de cette transformation d'image est d'enlever la redondance entre les pixels voisins. Redondance due au fait que les images naturelles soient fortement corrélées (les valeurs des pixels voisins sont presque identiques). Cette décorrélation est due à l'orthogonalité de ses fonctions bases ce qui, par conséquent, rend les composantes de la DCT indépendantes ce qui nous permet de les traiter individuellement.

III.6.2 Concentration des coefficients

La DCT est très efficace pour des images fortement corrélées du fait qu'elle permet de compacter les coefficients qui représentent les basses fréquences dans une seule partition (voir figure 2.7) de la matrice image, cela permet la séparation des fréquences basses des fréquences hautes sans présenter une déformation des caractéristiques de l'image et si c'est une image faiblement corrélée, les coefficients sont concentrés dans plusieurs partitions de la matrice image.

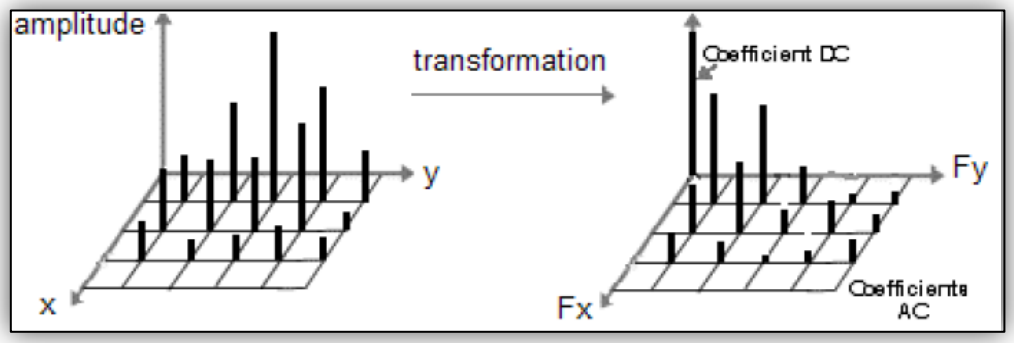


Figure 2.2-La concentration d'énergie par la DCT

III.6.3 Séparabilité :

Il est à noter que la formule de la DCT à deux dimensions peut être écrite de la façon suivante :

$$c(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \mathrm{f(x,y)} \cos\left[\frac{\pi(2x+1)u}{2N}\right] \sum_{y=0}^{N-1} \mathrm{f(x,y)} \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad [2\text{-}4]$$

Cette écriture permet de mettre en évidence la notion de séparabilité qui est une caractéristique très importante de la DCT. Cela veut dire qu'on peut calculer C(u, v) par deux exécutions successives de la DCT à une dimension, une sur les lignes et l'autre sur les colonnes d'une image comme indiqué sur la figure 2.7.

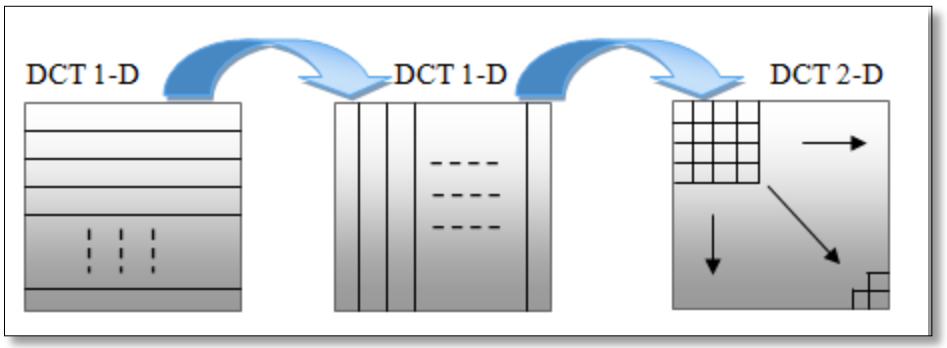


Figure 2.3 Séparabilité de la DCT II

III.6.4 Symétrie :

L'équation précédente permet aussi de remarquer que les opérations effectuées sur les lignes et les colonnes de la matrice de données sont identiques, ce qui fait de la DCT une transformation symétrique formulée par : $T = AfA$ avec :

- ✓ f est la matrice de niveaux de gris de l'image
- ✓ A est la matrice de transformation symétrique tel que :

$$a(i,j) = \alpha(j) \sum_{j=0}^{N-1} \cos\left[\frac{\pi(2j+1)i}{2N}\right] \qquad [2\text{-}5]$$

C'est une propriété extrêmement utile puisqu'elle implique que la matrice de transformation peut être précalculée et donc amélioration de l'efficacité de calcul.

III.6.5 Orthogonalité :

La DCT est une transformation orthogonale qui permet une et une seule représentation pour chaque image dans le domaine fréquentiel.

III.7 Sélection des coefficients DCT :

III.7.1 La DCT par bloc 8x8

L'image est découpée en bloc de (N x N) pixels. Cependant Pour des raisons de performance et de complexité de calcul, la DCT est souvent utilisée sur des blocs de taille 8x8.

En effet, en augmentant la taille de ces blocs, la compression serait meilleure, mais le coût en temps a été jugé trop grand. Le choix de la taille des blocs doit donner le meilleur compromis entre la complexité de calcul et le taux de compression et c'est pour cela que la valeur 8 a été choisie [Kha 03].

III.7.2 Le chevauchement des blocs 8x8

Pour éviter les effets de discontinuité entre les blocs DCT, on peut faire chevaucher les blocs dans l'image horizontalement et verticalement. Le degré de chevauchement est exprimé en pourcentage, si on a un chevauchement de 50%, alors 50% des pixels formant un bloc i vont construire 50% du bloc i+1, ainsi les blocs voisins horizontalement et verticalement se ressembleront à 50%.

Ce principe est utilisé surtout en compression, mais son principal inconvénient est que le temps de calcul augmente en augmentant le nombre de blocs, en plus il ajoute de la corrélation entre les coefficients qui proviennent des blocs voisins donc moins d'indépendance entre les données [Kha 03].

III.7.3 La sélection des coefficients :

L'information est essentiellement portée par les coefficients basses fréquences et donc elle peut être préservée par un nombre très petit de coefficients. Par conséquent, la plupart des coefficients de la DCT peuvent être ignorés ce qui réduit la dimensionnalité. Il existe plusieurs méthodes de sélections des coefficients DCT, la plus utilisée est la méthode appelée ZIGZAG.

Elle consiste à parcourir en zigzag les éléments de la matrice transformée dans un ordre bien précis à partir des fréquences les plus basses vers celles les plus hautes. Nous obtenons alors un vecteur de données classées selon la fréquence spatiale la première valeur étant le coefficient DC tandis que tous les autres sont des coefficients AC [Kha 03].

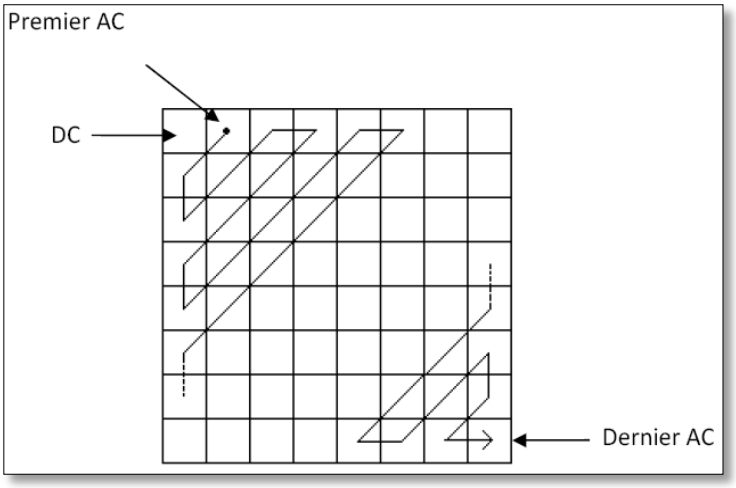


Figure 2.4-Sélection des coefficients DCT par la méthode ZigZag

IV. L'ETAT DE L'ART DES METHODES DE RECONNAISSANCES DE VISAGES :

L'authentification par le visage est la technique la plus commune et la plus populaire puisqu'elle correspond à ce que nous utilisons naturellement pour reconnaître une personne. Les caractéristiques qui servent à la reconnaissance du visage sont les yeux, la bouche, la forme du visage (contour), etc.

On peut diviser les méthodes de reconnaissance de visages en trois catégories : les méthodes globales, les méthodes locales, et les méthodes hybrides.

IV.1 Les méthodes locales :

Ce sont des méthodes à caractéristiques locales, ou analytiques. Elles consistent à appliquer des transformations en des endroits spécifiques de l'image, le plus souvent autour de points caractéristiques (coins des yeux, de la bouche, le nez, ...). Elles nécessitent donc une connaissance a priori sur les images.

L'avantage de ces méthodes est qu'elles prennent en compte la particularité du visage en temps que forme naturelle à reconnaître, en plus elles utilisent un nombre réduit de paramètres et elles sont plus robustes aux problèmes posés par les variations d'éclairement, de pose et d'expression faciale [Nic 06].

En utilisant une méthode locale, d'avantage d'énergie sera accorder aux petits détailles locaux évitant ainsi le bruit engendré par les cheveux, les lunettes, les chapeaux, la barbe, etc. De plus certaines parties du visage sont relativement invariantes pour une même personne malgré ses expressions faciales ; c'est le cas notamment des yeux et du nez. Ceci demeure vrai tant que ces caractéristiques du visage ne sont pas en occultation. Mais leur difficulté se présente lorsqu' il s'agit de prendre en considération plusieurs vues du visage ainsi que le manque de précision dans la phase "extraction" des points constituent leur inconvénient majeur [Moa 05].

Parmi ces approches on peut citer :

IV.1.1 Modèles de Markov Cachés (Hidden Markov Models):

La modélisation stochastique permet l'utilisation de modèles probabilistes pour traiter les problèmes à information incertaine ou incomplète. Ainsi, les modèles de Markov cachés connaissent un net regain d'intérêt tant dans ses aspects théoriques qu'appliqués.

Un modèle de Markov caché (HMM) est caractérisé par un modèle Markovien à état fini et un ensemble de distribution de sortie. Les paramètres de transition dans la chaise de Markov modélisent les variabilités temporelles, tandis que les paramètres de distributions de sortie modélisent les variabilités spectrales. Ces deux types de variabilités étant à la base de plusieurs processus physiques.

Les caractéristiques faciales les plus significatives d'une image de visage, à savoir les cheveux, le front, les yeux, le net et la bouche, se présentent dans un ordre naturel de haut en bas. On se basant sur cette observation, l'image d'un visage peut être modélisée en utilisant un HMM unidimensionnel en assignant à chacune de ces régions un état. Une séquence des valeurs des pixels forme une chaine de Markov, si la probabilité que le système à l'instant $n+1$ soit à l'état x_{n+1} dépend uniquement de la probabilité que le système soit à l'état x_n à l'instant n. Cette présentation est utilisée pour faire la comparaison entre deux images.

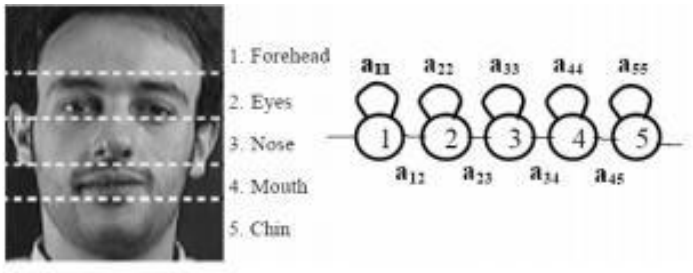


Figure 2.5-Les 5 états du HMM (de haut en bas)

Ces modèles de Markov cachés (HMM) sont utilisés depuis plusieurs années pour la détection et la reconnaissance du visage. Différentes variantes ont également été proposées mais celle des (Embedded HMM) génère des résultats supérieurs aux méthodes HMM de base. Reposant sur certains coefficients de la transformée en cosinus discrète (DCT) comme source d'observations, les Embedded HMM constituent un algorithme de reconnaissance très performant. Or, les temps d'exécution des phases d'apprentissage et de test sont relativement élevés, nuisant donc à son utilisation en temps réel sur d'immenses banques d'images [Nef 98].

IV.1.2 Eigen Objects (EO):

Cette méthode possède le même principe que les Eigen Faces, qu'on verra par la suite, mais appliquée à des parties précises du visage comme les yeux, le nez et la bouche. La personne peut par exemple être reconnue uniquement grâce à ses yeux. Cette méthode rencontre le problème de non précision lors de la localisation des points caractéristiques avant l'application de la méthode. Pour réaliser l'apprentissage, un module de ce type doit tout d'abord procéder à une analyse en composantes principales des yeux contenus dans la banque de visages. L'espace des yeux ainsi construit pourra alors servir au processus de reconnaissance qui est identique à celui utilisé pour les « EigenFaces » [Tur 91].

IV.1.3 Elastic Bunch Graph Matching (EBGM):

L'algorithme EBGM trouve ses fondements dans les neurosciences, en imitant le fonctionnement de certaines cellules spécialisées localisées dans le cortex visuel primaire. C'est un algorithme local (il ne traite pas directement les valeurs de niveaux de gris des pixels d'une image de visage), ce qui lui confère une plus grande robustesse aux changements

d'éclairement, de pose et d'expression faciale. Cette méthode représente le visage par un graphe étiqueté.

Un graphe est composé d'un ensemble de nœud connecté entre eux par des contours. Chaque nœuds peut correspondre à un point caractéristique du visage. Ce dernier est caractérisé par sa position et par un vecteur qui contient des informations de son voisinage.

Ceci peut être déterminé en appliquant un filtre de Gabor à l'image. Le résultat de cette analyse est enregistré dans un vecteur appelé Jets. Le modèle du visage sera constitué par les Jets et les positions relatives des nœuds. On a donc une information locale à travers les nœuds et une information globale à travers leur interconnexion. Cependant il est plus difficile à implémenter que les méthodes globales [Wiz 97][Hiz 09][Mor 09].

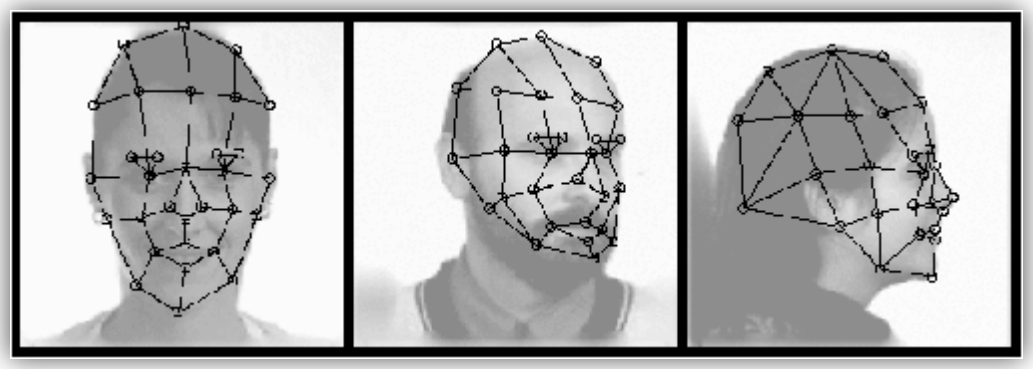


Figure 2.6-Sélection des points caractéristiques et leurs liaison,

IV.1.4 Template Matching :

L'appariement de gabarits est une technique de comparaison des images, son principe est simple. En effet, elle permet l'extraction et la construction des descripteurs des points d'intérêts de l'image, ces descripteurs sont très robustes et fiables et permettent une représentation fidèle de l'image en se basant sur son contenu. En plus on peut permettre une meilleure représentation à notre image par translation et rotation sans perte d'information [Las 05].

IV.2 Les méthodes globales:

Les méthodes globales basées sur des techniques d'analyse statistique bien connues. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) sont utilisées comme entrée à l'algorithme de reconnaissance et sont généralement transformées en vecteurs, plus faciles à manipuler. L'avantage principal des méthodes globales est qu'elles sont relativement rapides à mettre en œuvre. En revanche, elles sont très sensibles aux variations d'éclairement, de pose et d'expression faciale. Parmi les approches les plus importantes réunies au sein de cette classe on trouve:

IV.2.1 Analyse en Composantes Principales (PCA) ou Eigen Faces :

L'algorithme ACP appliqué au visage est né des travaux de MA. Turk et AP. Pentland [ref] au MIT Media Lab, en 1991. Il est aussi connu sous le nom de « Eigen Faces » car il utilise des vecteurs propres et des valeurs propres. Sa simplicité à mettre en œuvre contraste avec une forte sensibilité aux changements d'éclairement, de pose et d'expression faciale. Le principe selon lequel on peut construire un sous-espace vectoriel en ne retenant que les « meilleurs» vecteurs propres, tout en conservant beaucoup d'information utile, fait de l'ACP un algorithme efficace et couramment utilisé en réduction de dimensionnalité; il peut alors être utilisé en amont d'autres algorithmes (comme le LDA par exemple) [Kon 05].

L'analyse en composante principale est aussi une méthode d'analyse des données multi variées. Elle permet de décrire et d'explorer les relations qui existent entre plusieurs variables simultanément (à la différence des méthodes bi-variées qui étudient les relations supposées entre 2 variables), pour rapprocher au sein des "composantes" les variables les plus proches entre elles. Enfin, l'étude théorique de l'algorithme ACP est très pédagogique et permet d'acquérir de solides bases pour la reconnaissance 2D du visage [Mor 06].

Figure 2.7-Exemple d'Eigen Faces

IV.2.2 Analyse Discriminante Linéaire (LDA) :

Initiée par Fisher en 1936, elle a été introduite dans la reconnaissance de visage par Belhumeur et al, en 1997. L'analyse discriminante linéaire fait partie des techniques d'analyse discriminante prédictive. Il s'agit d'expliquer et de prédire l'appartenance d'un individu à une classe (groupe) prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives.

L'algorithme LDA est aussi connu sous le nom de « Fisherfaces ». Contrairement à l'ACP, il permet d'effectuer une véritable séparation de classes. Pour séparer en classes la base d'images d'apprentissage de sorte que chaque classe comporte plusieurs images de la même personne, on calcule les matrices de dispersion interclasses et intra classes puis on cherche une projection qui minimise la dispersion intra classes (variation des images d'une même personne) et maximise la dispersion interclasses (variation des images de personnes différentes). Lorsque le nombre d'individus à traiter est plus faible que la résolution de l'image, il est difficile d'appliquer le LDA qui peut alors faire apparaitre des matrices de dispersions singulières (non inversibles). Afin de contourner ce problème, certains algorithmes basés sur la LDA ont récemment été mis au point (les algorithmes ULDA, OLDA, NLDA) [Mor 06].

IV.2.3 Machine à Vecteurs de Support (SVM) :

C'est une technique qui a été proposée par V.Vapnik en 1995, elle est utilisée dans plusieurs domaines statistiques (classement, régression, fusion,… ect). L'idée essentielle de cette approche consiste à projeter les données de l'espace d'entrée (appartenant à des classes différentes) non linéairement séparables, dans un espace de plus grande dimension appelé

espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables [Guo 00].

Dans cet espace, la technique de construction de l'hyperplan optimal est utilisée pour calculer la fonction de classement séparant les classes tels que :

- ✓ Les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan.
- ✓ La plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale.

Depuis sont introduction dans le domaine de reconnaissance de formes, plusieurs travaux ont montré l'efficacité de cette technique, principalement en traitement d'images [Kha 02].

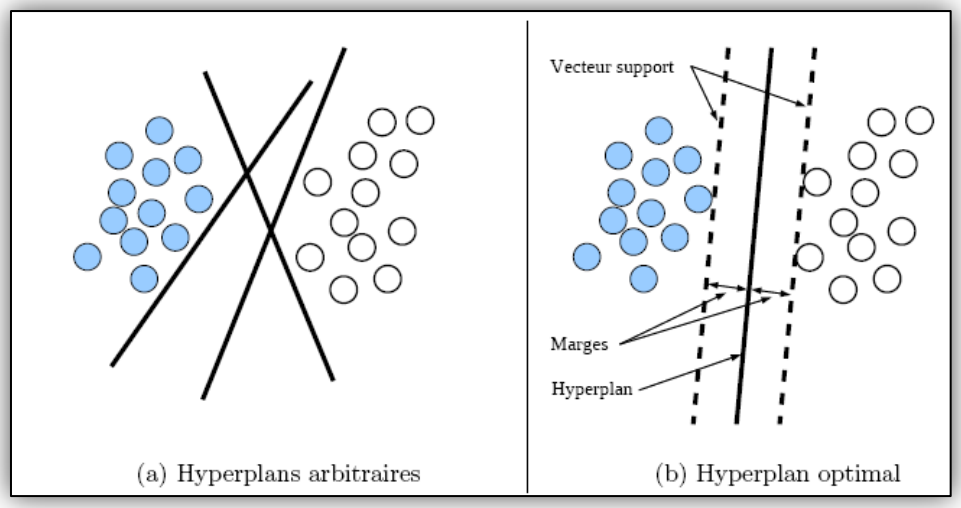


Figure 2.8-Séparation de deux classes de données

IV.2.4 Mélange de gaussiennes (GMM) :

Un modèle de mélange gaussien (GMM) est un *modèle statistique* exprimé selon une densité mélange. Elle sert usuellement à estimer paramétriquement la distribution de variables aléatoires en les modélisant comme une somme de **plusieurs gaussiennes** (appelées noyaux). Il s'agit alors de déterminer la variance, la moyenne et l'amplitude de chaque gaussienne. Ces paramètres sont optimisés selon un critère de maximum de vraisemblance pour approcher le

plus possible la distribution recherchée. Cette procédure se fait le plus souvent itérativement via l'algorithme espérance-maximisation (EM).

Ces approche, proposée par C.Sanderson et al [Fly 08], qui consiste à transformer les images de départ en plusieurs vecteurs de coefficients DCT (qu'on verra par la suite) puis modéliser leur distribution selon une combinaison linéaire de plusieurs gaussiennes qui vont représenter un modèle d'une personne.

Cette technique est venue pour améliorer les performances des HMM, elle a prouvé une efficacité surprenante surtout en matière de précision et de temps d'exécution [Ben 07]. Cette approche, ayant été choisie dans le cadre de notre sujet de fin d'études, sera présentée en détail dans le chapitre IV.

IV.2.5 Réseaux de neurones artificiels (RNA) :

Parmi les techniques non-linéaires d'extraction qui ont été largement utilisées pour la reconnaissance de visages, on trouve celles qui reposent sur un réseau de neurones artificiels (RNA). Ces derniers ont été initialement inspirés de la physiologie du système nerveux si parfaitement créée et conçue pour exécuter des tâches calculatoires. Elle a la particularité de s'adapter, d'apprendre, de généraliser pour classer les données en entrée. Le neurone formel, introduit par J. Mc Culloch et W. pitts dans les années quarante, constitue la base de l'architecture des RNA. [Hiz 09].

On classe généralement les réseaux de neurones en deux catégories: les réseaux faiblement connectés à couches que l'on appelle des réseaux « feedforward » ou réseaux directs et les réseaux fortement connectés que l'on appelle des réseaux récurrents. Dans ces deux configurations, on retrouve des connexions totales ou partielles entre les couches.

Les réseaux de neurones peuvent être utilisés tant pour la classification, la compression de données ou dans le contrôle de systèmes complexes en automatisme. Un réseau de neurone est un assemblage fortement connecté d'unités de calcul. On peut entraîner un réseau de neurone pour une tâche spécifique (reconnaissance de visage dans notre cas) en ajustant les valeurs des connections (ou poids) entre les unités de calcul. L'ajustement des

poids se fait par comparaison entre la réponse du réseau (ou sortie) et la cible, jusqu'à ce que la sortie corresponde (au mieux) à la cible. On utilise pour ce type d'apprentissage dit, supervisé un nombre conséquent de pair entrée/sortie.

L'avantage de ce modèle est le gain de temps considérable. Cependant, l'utilisation d'exemples pour apprendre apporte le risque de ne pouvoir résoudre que des situations déjà rencontrées, où un phénomène de sur-apprentissage qui spécialiserait le réseau uniquement sur les exemples connus sans généraliser [Ben 05].

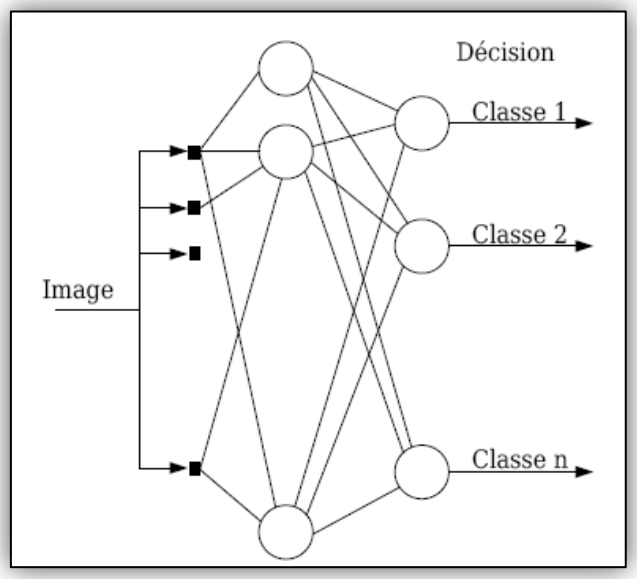


Figure 2.9-RNA discriminant pour la reconnaissance de visages

IV.3 Les méthodes hybrides :

La robustesse d'un système de reconnaissance peut être augmentée par la fusion de plusieurs méthodes. Chacune d'entre elles possède évidemment ses points forts et ses points faibles qui, dans la majorité des cas, dépendent des situations (pose, éclairage, expressions faciales,…etc.). Il est par ailleurs possible d'utiliser une combinaison de classificateurs basés sur des techniques variées dans le but d'unir les forces de chacun et ainsi pallier à leurs faiblesses.

V. Performances d'un système de reconnaissances de visages :

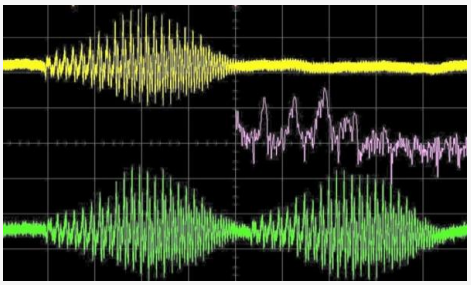
Les performances d'un système de reconnaissance de visage dépendent de plusieurs facteurs qui interviennent à plusieurs niveaux et qui peuvent limiter le degré de prédiction, parmi ces facteurs on peut citer :

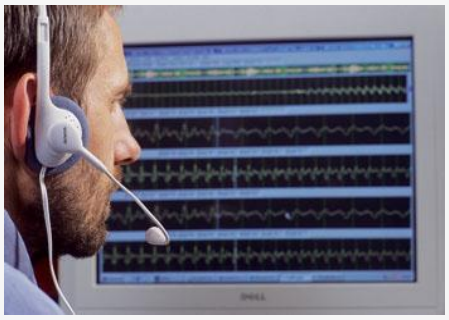
- ✓ L'environnement au moment de l'acquisition, luminosité de la pièce entre autre.
- ✓ La qualité des capteurs.
- ✓ Les différentes positions des capteurs.
- ✓ Les différentes positions du sujet par rapport au capteur.

VI. Conclusion :

Nous avons présenté, à travers ce chapitre, un résumé de l'état de l'art des systèmes de reconnaissance de visage en mettant en évidence le fait qu'il y ait plusieurs approches d'extraction de paramètres et de modélisation de ces derniers. Notre choix étant porté sur la DCT pour sa rapidité par rapport aux autres méthodes, sa capacité de décorrélation et sa concentration d'énergie ainsi que sur les GMM pour la modélisation des paramètres, GMM que nous étudierons en détail dans le chapitre IV.

3 Systèmes de reconnaissance du locuteur





I. INTRODUCTION

Le petit Robert définit la parole comme étant « la faculté de communiquer la pensée par un système de sons articulés émis par les organes de la phonation ». Pourtant, au-delà des idées qu'elle transporte, la parole est également porteuse d'une information sur celui qui l'émet.

La reconnaissance du locuteur est une technique biométrique qui exploite cette information et s'en sert pour identifier une personne. En effet, il n'existe pas deux personnes différentes qui possèdent des voix identiques. Celles-ci étant fonctions de leurs caractéristiques physiques telles que la largeur du larynx et de leurs caractéristiques comportementales comme leurs manières de parler.

Cette modalité, utilisée naturellement par les êtres humains pour se reconnaitre mutuellement avec une précision plus ou moins importante, présente les avantages d'être non intrusive en plus d'être l'une des techniques de biométrie les moins couteuses. Néanmoins, elle prend toute son importance dans les applications criminalistiques, puisqu'il s'agit bien de la seule modalité à laquelle on peut accéder via les réseaux de communications téléphoniques.

Un système de reconnaissance automatique du locuteur procède en trois étapes : l'analyse acoustique du signal de la parole, la modélisation du locuteur et une dernière étape de décision.

I. RECONNAISSANCE AUTOMATIQUE DU LOCUTEUR

Le message parlé véhicule en plus de son sens, une information sur le locuteur lui-même entre autre son identité c'est pour cela qu'il ne faut pas confondre reconnaissance du locuteur et reconnaissance de la parole. Dans le premier cas on cherche à déterminer l'identité de l'individu et dans le second cas on cherche à trouver le sens de ce que dit la personne sans se soucier de son identité.

En reconnaissance de la parole il nous faut neutraliser la variabilité entre les locuteurs. Par contre en reconnaissance du locuteur c'est aux caractéristiques de l'émetteur qu'il faut s'intéresser. Les facteurs de variabilité inter-locuteur deviennent alors des facteurs de caractérisation de la voix. En revanche, la variabilité intra-locuteur passe au premier plan des obstacles auxquels il faut faire face.

II.1 Variabilité de la voix

La variabilité inter-locuteurs démontre les différences du signal de parole en fonction du locuteur. Cette variabilité, utile pour différencier les locuteurs, est également mélangée à d'autres types de variabilité : variabilité intra-locuteur, variabilité due aux conditions d'enregistrement et de transmission du signal de parole (bruit ambiant, microphone utilisé, lignes de transmission) et variabilité due au contenu linguistique.

Les variabilités inter-locuteurs proviennent des différences physiologiques (différences dimensionnelles du conduit vocal, fréquence d'oscillation des cordes vocales) et de différences de style de prononciation. Certaines de ces différences qui influencent la représentation de chaque locuteur, nous permettent de les séparer.

Les variabilités intra-locuteur font que la voix dépend de l'état physique et émotionnel d'un individu. La voix humaine varie avec le temps ou les conditions physiologiques et psychologiques du locuteur. Cependant, ces variations intra-locuteur ne sont pas identiques pour tous les humains. En effet, hormis les variations lentes de la voix dues au vieillissement, certains phénomènes extérieurs tels que l'état de santé d'une personne ont une influence variable sur sa voix.

Enfin, il y a des variabilités qui ont un impact considérable sur les caractéristiques de la voix et qui résident dans l'ensemble des facteurs socioculturels (cellule familiale, l'école, milieu professionnel, origines géographiques) et ceci en influant sur la manière dont parlent les individus.

Ces facteurs sont les plus spontanément modifiés par le locuteur, que ce soit pour masquer certaines de ces caractéristiques ou bien en imiter d'autres. Notons que les imitateurs ne pouvant habituellement reproduire que les caractéristiques vocales les plus évidentes au système humain, il est impossible d'imiter la voix d'une personne inscrite dans la base de données du système automatisé.

Une dégradation croissante des performances a été observée au fur et à mesure que le temps qui sépare la session d'enregistrement de la session de test augmente. De plus, le comportement des locuteurs change lorsqu'ils ce s'habituent au système. Les modèles des locuteurs doivent donc être régulièrement mis à jour avec les nouvelles données d'exploitation du système. Les altérations de la voix dues à l'état physique (fatigue, rhume) ou émotionnel (stress), lorsqu'ils sont importants, peuvent mettre aussi en échec l'efficacité de certains systèmes [Cap 95].

II.2 Dépendance au texte

On parle de reconnaissance vocale en mode dépendant du texte lorsque le texte prononcé par le locuteur est fixé et connu à l'avance. A l'opposé, lorsque le texte prononcé par le locuteur n'est pas connu a priori, on parle de mode indépendant du texte. Mais cette terminologie ne rend pas bien compte des différentes dépendances au texte possible. Les différents systèmes peuvent être classés, selon le degré croissant d'indépendance au texte, de la façon suivante :

- ✓ Système à texte fixé dépendant du locuteur : pour un locuteur donné, le texte est toujours le même d'une session à l'autre. Mais chaque locuteur a un texte différent.
- ✓ Système dépendant du vocabulaire : l'utilisateur du système prononce une séquence issue d'un vocabulaire limité (des séquences de chiffres par exemple), mais dont l'ordre peut varier d'une session à l'autre.
- ✓ Système dépendant d'événements phonétiques : le vocabulaire n'est pas directement imposé, mais certains événements phonétiques doivent être présents dans la séquence de parole prononcée (p.ex. présence de certaines voyelles ou nasales). Les phrases à prononcer peuvent éventuellement être affichées sur l'écran à chaque session.

- ✓ Système à texte imposé par la machine : le texte est différent pour chaque session et pour chaque locuteur, mais affiché à chaque fois par la machine. Le texte est choisi de manière imprédictible pour éviter l'utilisation d'enregistrements par un imposteur.
- ✓ Système indépendant du texte : le locuteur est entièrement libre de ce qu'il dit à chaque session.

Malgré toutes ces difficultés apparentes, la voix reste un moyen biométrique intéressant à exploiter car pratique et disponible via le réseau téléphonique. L'authentification du locuteur, actuellement en plein essor, a une bonne acceptabilité. Elle offre en effet de nombreuses applications potentielles comme sécurisation accrue des téléphones portables, contrôle supplémentaire au niveau d'une application sur un site comme l'accès sécurisé à un bâtiment ou remplacement du mot de passe sur les ordinateurs.

Le principal avantage de cette technique est d'autoriser une authentification à distance. Ces applications concernent la vérification du locuteur à travers le réseau téléphonique pour accéder à un service (p.ex. validation de transactions par le téléphone) ou pour identifier un interlocuteur [Cap 95].

.

II.3 Sources d'erreurs

Le signal acoustique de la parole présente des caractéristiques qui rendent complexe son interprétation. L'information portée par ce signal peut être analysée de bien des façons à plusieurs niveaux (acoustique, phonologique, syntaxique, sémantique et pragmatique). Ce qui rend la tache de traitement de la parole complexe. Plus particulièrement, on a vu la variabilité inter-locuteurs est l'essence même de la reconnaissance [Cam 97]. Il existe, cependant plusieurs facteurs qui peuvent augmenter la variabilité intra-locuteur comme par exemple :

- ✓ Etat pathologique du locuteur (maladie, émotions, vieillissement..).
- ✓ Facteurs socioculturels (accent du locuteur).
- ✓ Locuteur non coopératifs (application judiciaire).
- ✓ Condition de prise de son, bruit ambiant…

II. PROCESSUS DE RECONNAISSANCE AUTOMATIQUE DU LOCUTEUR

La reconnaissance vocale peut être interprétée comme une tâche particulière de reconnaissance des formes. C'est une succession de modules dont l'étape finale est de reconnaître une forme particulière, c'est-à-dire le signal de parole que l'on met à l'entrée de cette chaîne. Un système automatique de reconnaissance vocale se divise généralement en quatre modules : un module de paramétrisation du signal de parole qui est généralement constitué d'une analyse spectrale vectorielle ; un module de modélisation qui détermine les caractéristiques d'un modèle à partir des paramètres extraits ; un module de comparaison qui consiste à utiliser des mesures de similarité entre modèles ou entre paramètres, voire entre paramètres et modèles ; et enfin un module de décision, qui fournit finalement la réponse du système [Zwi 81].

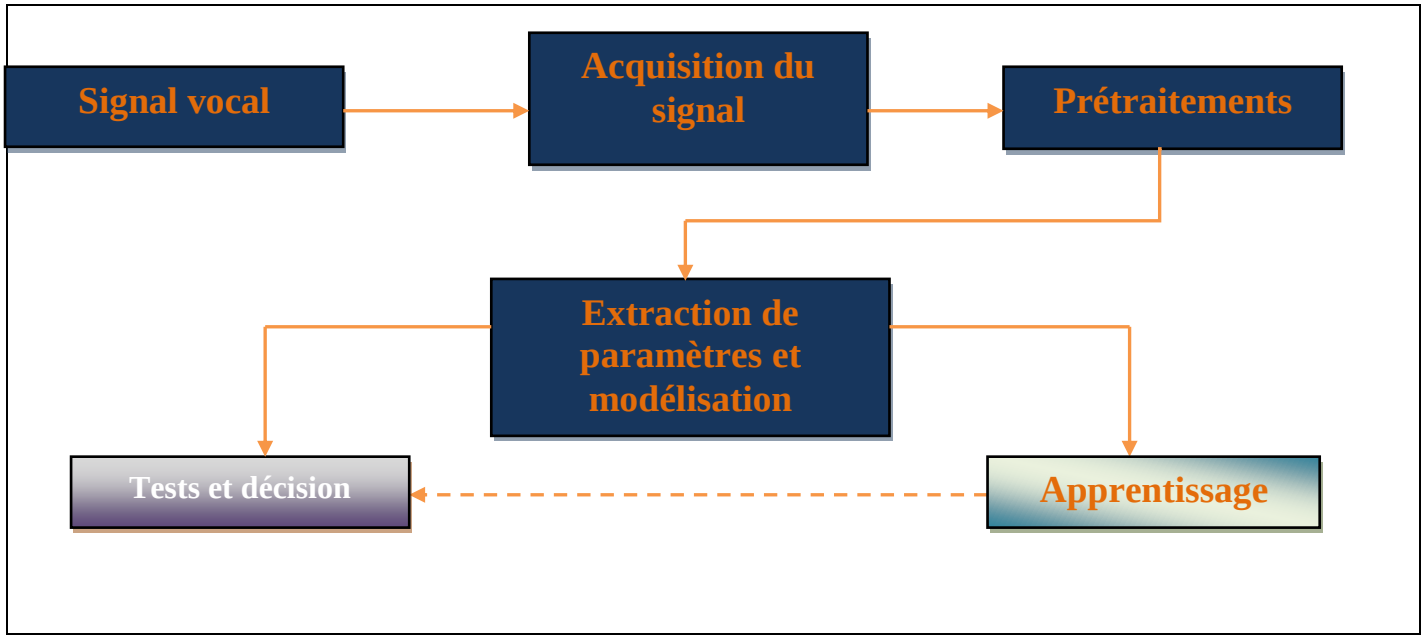


Figure 3.1-Processus de reconnaissance du locuteur

Donc le signal vocal de la personne à identifié devra suivre les étapes suivantes :

III.1 Acquisition du signal de la parole

Dans cette première étape, la parole prononcée est captée par un microphone et converti en signal numérique.

III.1.1 Production de la parole chez l'être humain

L'être humain est le seul être vivant à communiquer par la parole. Cette aptitude qui lui a permis de développer des langages complexes, n'est que le résultat de son anatomie particulière. En effet, l'appareil phonatoire humain est constitué de poumons qui sont la source d'air responsable de l'excitation initiale, de muscles (cage thoracique, abdomen…) , de cordes vocales situées dans le larynx et le pharynx qui, en vibrant sous l'effet de l'air émis par les poumons, produisent le son et de fosses nasales, du voile, du palais, de la langue et des lèvres qui agissent comme un guide d'onde acheminant les ondes sonores vers l'extérieur.

La forme du conduit vocal est le facteur de distinction physique le plus important et ceci est du au fait qu'il soit fondamentalement inhomogène puisqu'il est constitué de muscles, de ligaments et de structures rigides tels que les dents par exemple.

Néanmoins, le conduit vocal peut être vu comme étant un filtre ayant une fonction de transfert qui le caractérise et qu'on verra plus en détail par la suite [Mar 02].

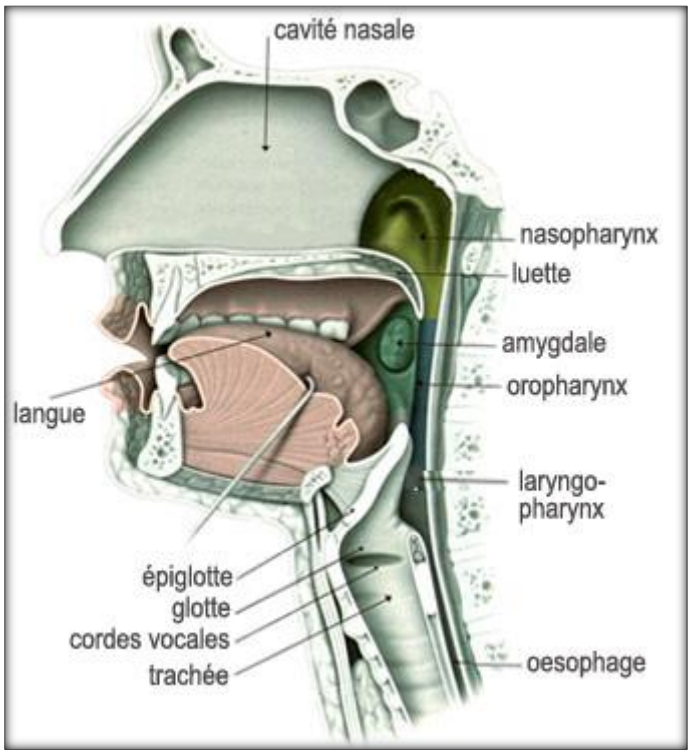


Figure 3.2-Schéma de l'appareil phonatoire

III.1.2 Acquisition de la parole

Le signal de parole est capté à l'aide d'un microphone qui transmettra ses caractéristiques à l'ordinateur. Le capteur ou microphone doit répondre à certains critères de justesse et de précision afin de limiter au maximum l'altération du signal. C'est en cela aussi que le bruit ambiant, la réverbération de la salle sont si influant quand à la qualité du signal reçu qui sera par la suite traité.

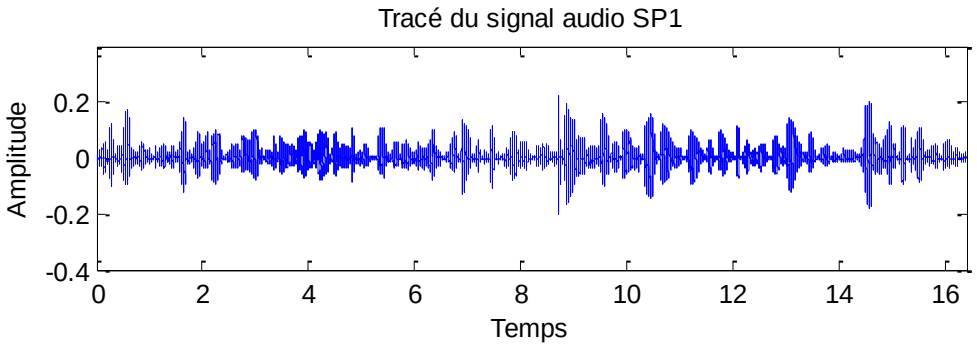


Figure 3.3-Tracé d'un signal audio après acquisition

III.1.3 Echantillonnage

L'échantillonnage consiste à représenter un signal fonction du temps $S(t)$ par ses valeurs $S(nT_e)$ à des instants multiples entiers d'une durée T_e, appelée période d'échantillonnage. Cette opération à pour but de créer un lien entre le signal vocal à temps continu avec le signal à temps discret représenté et traité par la machine.

Il est à noter que la fréquence d'échantillonnage F_e à un grand impact sur l'exactitude de la représentation du signal échantillonné par rapport au signal continu du fait qu'une fréquence trop petite par rapport à la fréquence maximale du signal engendrerait un recouvrement au niveau du spectre du signal échantillonné ce qui provoquerait une perte d'informations et par la suite une mauvaise représentation du signal continu, par ailleurs une fréquence d'échantillonnage plus grande que la fréquence maximale du signal donnerait naissance à un surplus d'informations. En réponse à ce problème la solution n'est autre que d'appliquer le théorème de Shannon, théorème qui stipule qu'une fréquence d'échantillonnage deux fois supérieur à la fréquence maximale du signal évite le repliement spectral du signal échantillonné tout en garantissant une représentation sans pertes d'information du signal traité.

III.2 Prétraitements

Le signal acquis est ensuite soumis à un certain nombre de traitements afin de garder l'information utile et ainsi ne fournir que les données nécessaires et convenables pour entamer la phase d'extraction de paramètres.

III.2.1 Découpage en trames

La déformation continue du conduit vocal fait que la nature du signal émis n'est pas stationnaire, ses paramètres sont donc variables dans le temps. Cependant il est possible de considérer le signal comme étant stationnaire sur un intervalle de 25 à 30 ms ceci étant du à la lenteur de la déformation de ce dernier. Afin de préserver au mieux les informations le découpage en trames est en réalité entrelacé. Dans notre cas, nous avons fait en sorte que deux trames successives aient les deux tiers de leurs informations en commun.

III.2.2 Préaccentuation

Ce traitement consiste à appliquer un filtre sur toutes les trames obtenues afin d'enlever la composante continue et d'amplifier le signal dans les hautes fréquences du fait que ces dernières contiennent les informations les plus pertinentes.

La transformée en Z de ce filtre est donnée par :

$$H(Z) = 1 - 0.95\, z^{-1} \qquad [3\text{-}1]$$

III.2.3 Elimination du silence

Tout signal de parole comprend des zones de silence. Ces zones ne contiennent aucune informations utiles ce qui affecte les performances du système en lui affectant un temps d'exécution trop long ou en altérant les paramètres lors de l'extraction. Le but donc de ce traitement est de localiser, dans le signal vocal, les segments de parole dépourvus de zones de silence.

Pour cette étape il est fréquent d'utiliser l'algorithme VAD (Voice Activity Detection) faisant appel à deux paramètres qui sont:

- ✓ Le taux de passage par zéro ou ZCR (Zero Crossing Rate) qui correspond au nombre de changements de signe d'un échantillon à son successeur et ceci dans une même trame « m » tel que ce dernier soit donné par :

$$\mathrm{ZCR_{S1}(m)} = \frac{1}{L}\sum_{m-L+1}^{m}\left|Sgn\left(S_1(n)\right) - Sgn\left(S_1(n-1)\right)\right| \qquad [3\text{-}2]$$

Ou : $Sgn(x) = \begin{cases} 1 & si\ x \geq 0 \\ 0 & si\ x < 0 \end{cases}$

- ✓ L'énergie est aussi un paramètre important lors de l'élimination de silence car une trame de silence à une énergie quasi nulle en comparaison avec une trame contenant des informations liées à la parole. cette dernière est définie par :

$$\mathrm{E_{S1}(m)} = \frac{1}{L}\sum_{m-L+1}^{m} {S_1}^2(n) \qquad [3\text{-}3]$$

L'algorithme VAD (Voice Activity Detection) utilise donc conjointement ces deux paramètres de la manière qui suit :

$$\mathrm{W_{S1}(m)} = \mathrm{E_{S1}(m)}[1 - \mathrm{ZCR_{S1}(m)}] \qquad [3\text{-}4]$$

Ainsi pour chaque trame « m », la valeur de cette fonction est comparée à un seuil afin de trancher sur sa nature (silence ou parole).

Le seuil de décision pour 5 trames successives, qui correspondant à 120 ms de silence, est donné par :

$$\mathrm{Seuil}_{\ silence} = \mu_{\mathrm{W}} + \alpha\ \delta_{\mathrm{W}} \qquad [3\text{-}5]$$

Avec : $\mu w : moyenne\ de\ \mathrm{W_{S1}}$

$\delta w : variance\ de\ \mathrm{W_{S1}}\ ;\ \alpha = 0.2\ {\delta_w}^{-8}$

III.2.4 Fenêtrage

Le but de traitement est de minimiser la déformation du spectre dans les hautes fréquences due au découpage en trame imposé au signal de parole, ainsi l'emploi d'une fenêtre dans le domaine temporel réduit progressivement l'amplitude du signal au commencement et à la fin de chaque trame.

Dans le domaine temporel la fenêtre rectangulaire interrompt brusquement le signal à ces extrémités générant artificiellement de hautes fréquences. Dans le domaine fréquentiel par contre sa fonction $Sinc$ a des lobes non négligeables loin de f=0 ce qui déforme le spectre.

La fenêtre de Hamming par contre présente dans le domaine fréquentiel des lobes secondaires qui deviennent vite négligeables ceci au prix d'un lobe principal plus large. Ainsi le spectre obtenu sera moins précis au voisinage de f_0 mais moins bruité dans les hautes fréquences. Les fenêtres de Hanning et Blackman-Harris ont dans le domaine fréquentiel des lobes secondaires négligeables mais ces derniers sont plus importants que ceux de la fenêtre de Hamming. D'où le choix d'utiliser la fenêtre de Hamming lors de notre traitement.

III.3 Extraction des paramètres MFCC

L'analyse acoustique du signal de parole consiste à extraire l'information pertinente et à réduire au maximum la redondance. Généralement, on calcule un jeu de coefficients acoustiques à des intervalles de temps réguliers, sur un des blocs du signal de longueur fixe. Ce jeu de coefficients constituera le vecteur acoustique recherché.

Il est toutefois commun de considérer l'analyse acoustique comme étant un changement de représentation dont l'objectif est de préserver au maximum l'information présente dans le signal d'origine tout en fournissant une description plus compacte de ce dernier. Il est à noter qu'en pratique la dimension des vecteurs acoustiques est de l'ordre de 15 à 20 [Ata 74].

III.3.1 Analyse spectrale

La parole est une combinaison entre une excitation des poumons et des interventions des différents organes vocaux qui caractérisent le locuteur et le différencie des autres. On peut donc voir ça comme une convolution tel que :

$$s(t) = e(t) * v(t) \qquad [3\text{-}6]$$

Ou :

$s(t)$ est le signal de la parole.

$e(t)$ est l'excitation des poumons.

$v(t)$ est la vibration et l'intervention des organes vocaux.

L'excitation étant un signal aléatoire qu'on peut négliger pour modéliser le locuteur.

Ainsi, la prochaine étape a pour but de diminuer l'influence de $e(t)$. Pour cela, il faut "déconvoluer" le signal $s(t)$. Pour cela il est plus simple de passer au domaine fréquentiel en appliquant la FFT (Fast Fourrier Transform) à chacune des trames constituant le signal ce qui nous donne :

$$S(f) = E(f) \times V(f) \qquad [3\text{-}7]$$

Par la suite, l'application du logarithme sur l'équation précédente transformera le produit en somme ce qui aura pour but de séparer l'influence de la source et celle du conduit vocal. On obtiendra donc :

$$\log|S(f)| = \log|E(f)| + \log|V(f)| \qquad [3\text{-}8]$$

Notons que la fonction logarithme permet d'avoir une meilleure appréciation du tracé du spectre d'énergie obtenu en appliquant la FFT. Cette distinction est montrée dans la figure suivante ou les zones contenant plus d'énergie sont clairement visible en rouge.

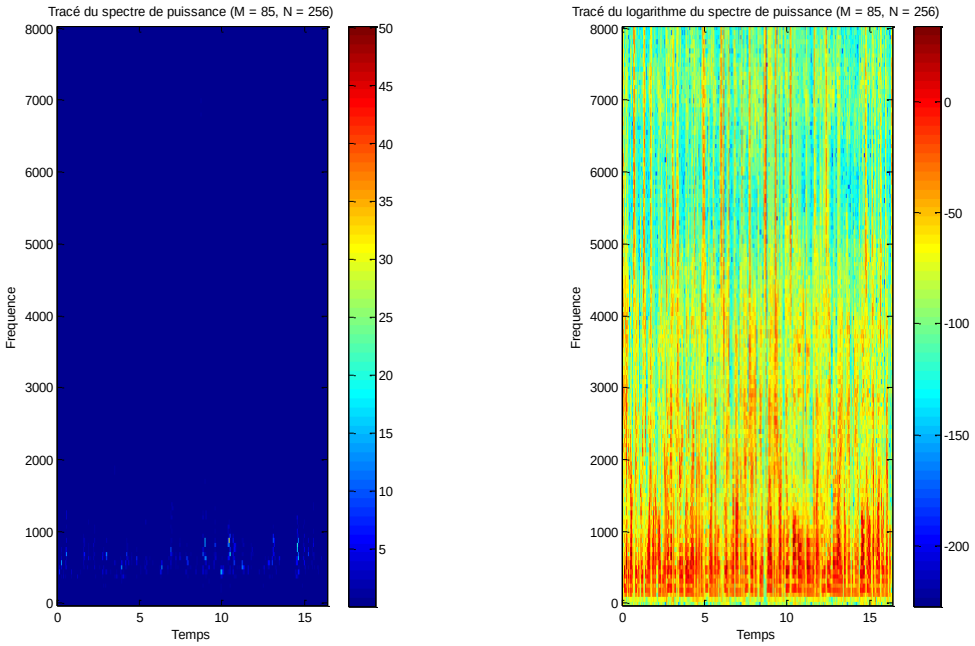


Figure 3.4-Tracé du spectre d'énergie du signal avant et après application du logarithme

Le retour au domaine temporel est réalisé grâce a la transformée de Fourier inverse IFFT, mais étant donné que les coefficients cepstraux sont de nature réelle il est plus intéressant d'utiliser la transformée en cosinus inverse IDCT.

Cette opération donne donc le « cepstral » du filtre du conduit vocal sous la forme :

$$\hat{s}(n) = \hat{e}(n) + \hat{h}(n) \qquad [3\text{-}9]$$

III.3.2 Filtrage Mel

De nombreuses études ont montré que la perception humaine des sons ne suit pas une échelle linéaire, en d'autre terme la sélectivité de l'oreille humain diminue avec l'accroissement des fréquences, d'où l'idée de définir pour chaque valeur de fréquence f une hauteur subjective qui est mesurée sur une échelle "Mel".

L'échelle "Mel" est caractérisée par le fait que l'espacement sur l'axe des fréquences est linéaire pour les fréquences basses alors qu'il est logarithmique pour celles supérieurs à 1KHz. Nous pouvons utiliser la formule approximative suivante afin de faire correspondre à chaque fréquence en Hz une fréquence sur l'échelle "Mel" :

$$f_{mel} = 2595 \times log_{10}(1 + \frac{f_{Hz}}{700}) \qquad [3\text{-}10]$$

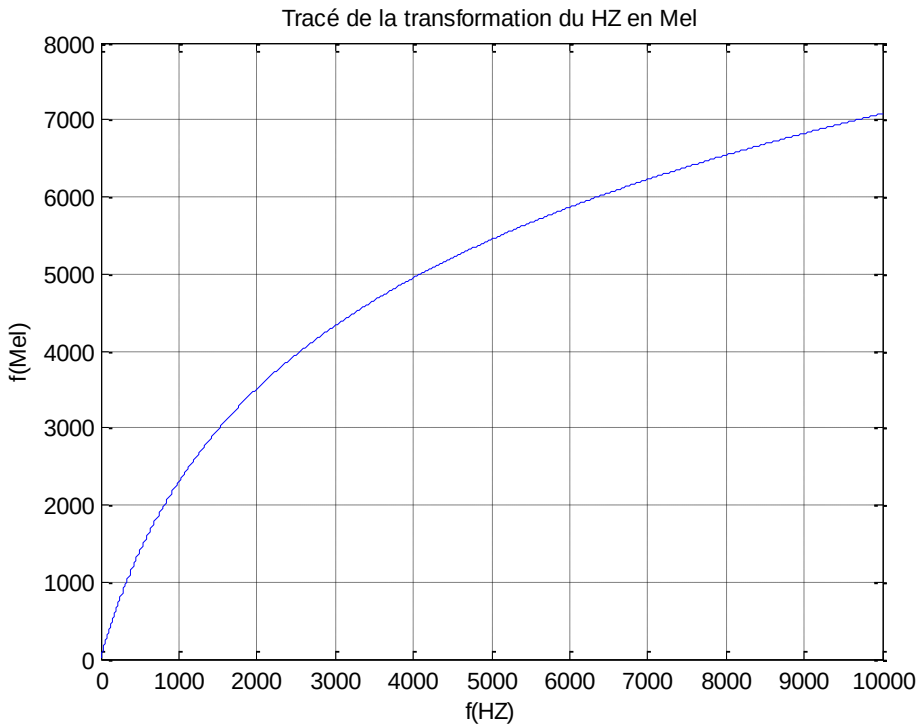


Figure 3.5-Transformation de l'échelle fréquentielle

Ce filtre reproduit donc la sélectivité de l'oreille qui diminue avec l'accroissement de la fréquence et ceci en appliquant un banc de K filtres triangulaires positionnés uniformément sur l'échelle Mel donc non uniformément sur l'échelle fréquentielle(Hz).

La distance entre deux triangles est d'environ 150 Mels et la largeur d'un triangle est d'environ 300 Mels. Ce qui donne :

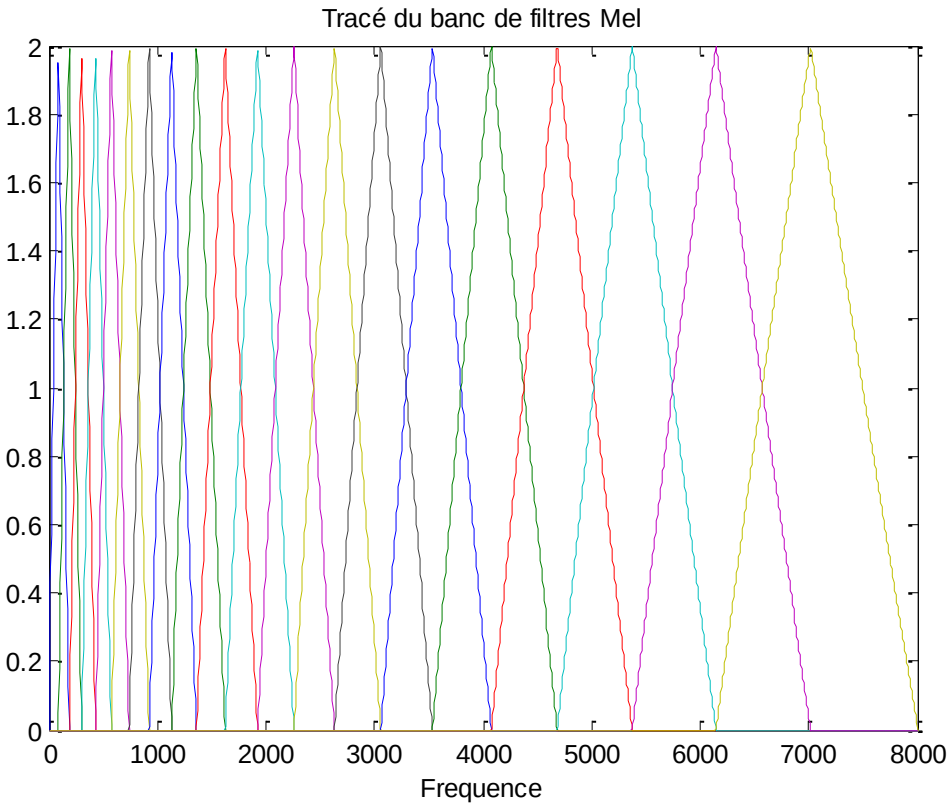


Figure 3.6-Tracé d'un banc de filtres Mel

III.3.3 Paramètres MFCC:

Apres la connaissance de la notion de « cepstre » et des avantages du filtrage Mel, la combinaison de ces deux concepts donne naissance aux paramètres MFCC qui viennent s'imposer sur l'ensemble des paramètres acoustiques en milieu bruité.

Les coefficients MFCC (Mel Frequency Cepstral Coefficients) d'une trame de parole sont calculés suivant le schéma suivant :

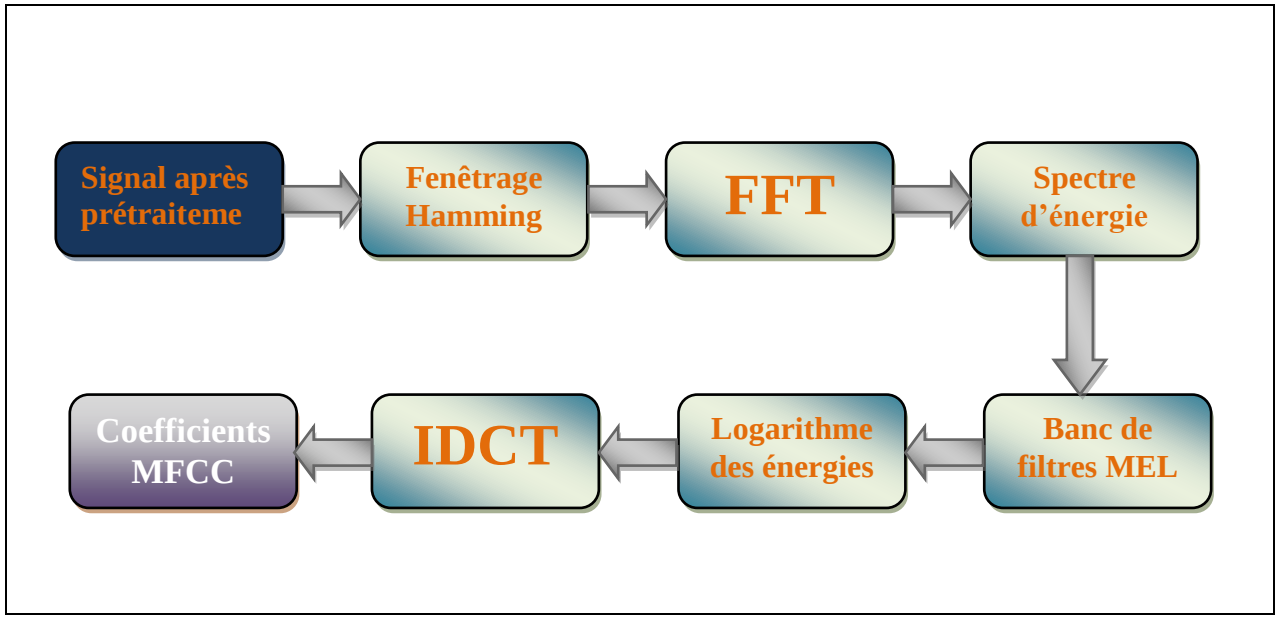


Figure 3.7-Calcul des coefficients cepstraux MFCC

En résumé :

- ✓ Le signal obtenu après prétraitement est découpé en trames.
- ✓ Chaque trame se voit appliquer la FFT afin d'obtenir le spectre d'énergie de chacune d'entre elles.
- ✓ Chaque spectre d'énergie passe ensuite à travers un banc de filtres Mel. Des études prouvent, et notre travail également comme le montreront les tests, que 20 filtres Mel sont largement suffisants pour représenter convenablement le locuteur.

Le logarithme de l'énergie résultante des filtres est calculé

- ✓ Enfin une transformation IDCT est ensuite appliquée au logarithme des coefficients d'énergie issus de l'analyse en banc de filtre afin d'obtenir les coefficients cepstraux.
- ✓ Cette transformation à pour effet de décorréler ces derniers ce qui amène une meilleure représentation du signal.

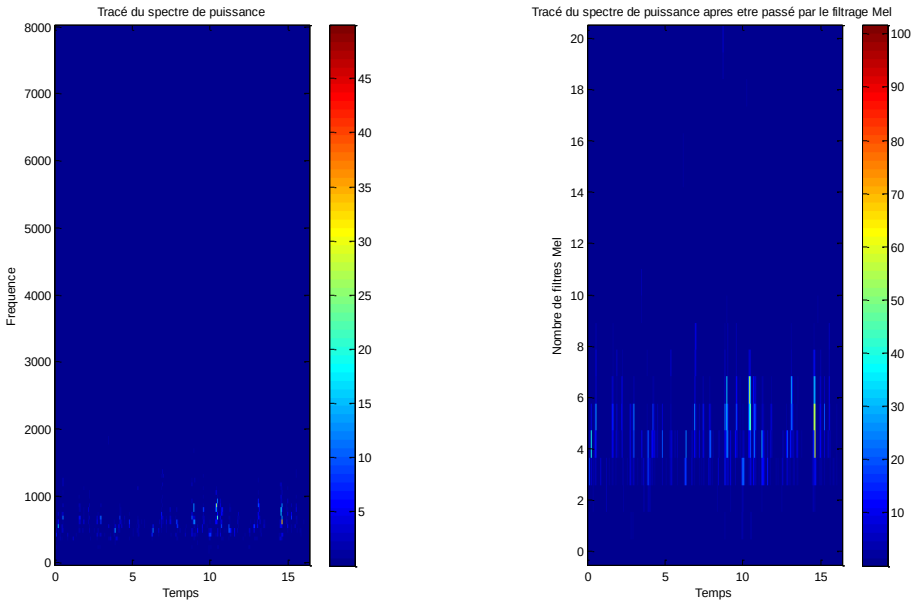


Figure 3.8-Tracé du spectre d'énergie avant et après filtrage Mel

III.MODELISATION DES PARAMETRES ACOUSTIQUES

De manière à modéliser des caractéristiques qui dépendent du locuteur, nous utilisons des algorithmes capables de capturer les points communs entre différentes représentations de motifs spectraux issus du même locuteur, constituant ainsi un modèle pour ce dernier, tout en ayant la possibilité de s'adapter aux variations d'échelles fréquentielles et temporelles liées à la parole. Ces algorithmes doivent être couplés avec une mesure qui permettra de donner une valeur de distance ou de similitude entre le modèle locuteur et un motif inconnu dont on cherche à déterminer la provenance.

Pour cela nous proposons ici un aperçu des méthodes déterministes et statistiques qui fournissent les meilleurs résultats en reconnaissance du locuteur.

IV.1 Comparaison temporelle dynamique DTW (Dynamic Time Warping)

La méthode de comparaison temporelle dynamique basée sur la programmation dynamique est utilisée intensivement en recherche opérationnelle pour résoudre les problèmes d'alignement séquentiel. Elle a été utilisée avec succès en reconnaissance de la parole puis en reconnaissance vocale [Rab 97].

En mode dépendant du texte, une idée assez naturelle consiste à considérer les paramètres dans l'ordre ou ils été mesurés et à calculer la distance, fenêtre par fenêtre, entre les paramètres correspondant aux différents locuteurs à comparer.

La difficulté posée par cette approche est qu'il est nécessaire de comparer les différences de vitesses d'élocution qui existent entre plusieurs répétitions d'un même segment de parole. C'est en partie pour cela que ce type de technique a été peu à peu abandonné laissant place aux modèles statistiques qui sont moins rigides et donc plus robustes vis-à-vis de la variabilité inhérente au signal de parole.

IV.2 Quantification vectorielle

Il s'agit de représenter l'espace acoustique par un nombre fini de vecteurs acoustiques.

Cela consiste à faire un partitionnement de cet espace en régions, qui seront représentées par leur vecteur centroïde. Pour déterminer la distance d'un vecteur acoustique à cet espace, on effectue une mesure de distance avec chacun des centroïdes des régions et on retient la distance minimale. Les plus petites distorsions sont observées dans le cas ou le vecteur acoustique provient du même locuteur pour qui nous avions établit le dictionnaire.

En général on considère que l'algorithme de quantification vectorielle ne fonctionne de manière satisfaisante que si on dispose d'au moins 20 à 50 fois plus de vecteurs de données que de vecteurs du dictionnaire. Notons qu'en pratique, on commence à observer une diminution très nette des performances lorsque la durée des enregistrements d'apprentissage devient inférieure à une dizaine de secondes [Cap 95].

IV.3 Modèles à mélange de distributions gaussiennes GMM (Gaussian Mixture Model)

Le modèle de mélange de distributions gaussiennes GMM consiste à supposer que la distribution des données peut être décrite comme une somme pondérée de densités gaussiennes multidimensionnelles [Rey 95].

Le cas particulier considéré ici est celui où dans chaque classe les données suivent une loi gaussienne. Ce choix tient essentiellement au fait que la loi gaussienne appartient à une famille de distributions dite exponentielles pour lesquelles le problème de l'identification des composantes du mélange se trouve simplifié. Pour le signal de parole, ce modèle ne paraît donc pas déraisonnable et il est d'autre part assez proche de la caractérisation fournie par la quantification vectorielle. La différence étant qu'avec la quantification vectorielle, on se contente de mettre en évidence un certain nombre de "points d'accumulation" des paramètres mesurés, alors qu'avec le modèle de mélange de distributions gaussiennes, on cherche en plus à décrire la distribution des paramètres mesurés autours de ces points d'accumulation.

Dans le cadre de la reconnaissance du locuteur, l'estimation des paramètres du modèle est toujours réalisée grâce à l'algorithme EM (que l'on expliquera par la suite) qui recherche de manière itérative les paramètres permettant de maximiser localement la vraisemblance des données d'apprentissage. La mesure de similarité est obtenue par calcul de la vraisemblance des vecteurs acoustiques à tester (en pratique on utilise plutôt le logarithme de la vraisemblance) compte tenu du modèle déterminé avec les données d'apprentissage.

En mode indépendant du texte et lorsque les données disponibles pour l'apprentissage sont suffisantes, le modèle à mélange de distributions gaussiennes permet d'obtenir de meilleures performances que celles des autres techniques comme quantification vectorielle). Cependant, lorsque la durée des enregistrements utilisés pour la phase d'apprentissage est faible (inférieure à 20 secondes) la méthode utilisant le modèle à mélange de distributions gaussiennes semble moins efficace compte tenu du nombre important de paramètres qu'il est nécessaire d'estimer [Dry 00].

IV.4 Modèles de Markov cachés HMM (Hidden Markov Models)

Les propriétés statistiques des HMMs en font une des modélisations les plus efficaces actuellement en reconnaissance du locuteur dépendante du texte. Ces chaines permettent de modéliser des processus stochastiques variant dans le temps. Pour cela, ils combinent les propriétés à la fois des distributions de probabilités et d'une machine d'état [Rab 97].

Un défaut commun à la plupart des techniques telle que la quantification vectorielle ou encore les GMMs est le caractère global. En effet ces techniques ne tiennent pas compte de l'ordre dans lesquelles sont présentées les fenêtres d'analyse de signal. Mais en pratique, cette hypothèse n'est pas vérifiée car les mesures effectuées dans des fenêtres voisines ne sont pas indépendantes. Une méthode permettant de prendre en compte certains aspects séquentiels et qui s'est avérée être très efficace en reconnaissance du locuteur dépendante du texte, consiste à utiliser un modèle de Markov caché (HMM).

Le modèle de Markov caché est un modèle statistique séquentiel qui suppose que les caractéristiques observées forment une succession d'états distincts. Un tel modèle est entièrement caractérisé par la donnée de trois jeux de paramètres :

Les probabilités initiales de se trouver dans chaque état.

- ✓ Les probabilités de transition qui décrivent les passages possibles entre les différents états.
- ✓ Les probabilités de sortie qui à proprement parler représentent les distributions. conditionnelles des caractéristiques observées en fonction de l'état du modèle.

Notons qu'en utilisant les chaines cachées de Markov en mode indépendant du texte, l'information supplémentaire apportée par la transition entre états n'améliore pas les performances de reconnaissance par rapport à l'utilisation des mélanges de gaussiennes.

IV.5 Tests et décision

La décision est l'aboutissement du processus de reconnaissance. Nous présenterons dans le chapitre IV les critères permettant de prendre une telle décision ainsi que les différentes méthodes pour y arriver.

IV.EVALUATION DES PERFORMANCES EN RECONNAISSANCE DU LOCUTEUR

Dans le domaine de la reconnaissance du locuteur, une des principales difficultés réside dans l'évaluation de l'efficacité des techniques employées. D'une manière générale, la phase d'évaluation est souvent plus coûteuse, en termes de moyens techniques et de quantité de travail nécessaires, que la phase de mise au point [G.R 00].

Il est possible d'évaluer la fiabilité d'une technique par une démarche empirique en constituant une base de données d'enregistrements de parole, puis en effectuant des tests systématiques. L'évaluation empirique constitue une méthode de validation très satisfaisante car elle permet d'obtenir directement une estimation de la fiabilité en situation réelle. Il faut bien avoir conscience du fait que l'évaluation empirique est, en général, une démarche très lourde car l'estimation des performances n'est significative que si le nombre d'enregistrements disponibles est très important. Le dimensionnement et la composition de la base de données utilisée pour l'évaluation empirique doivent en effet vérifier un ensemble de contraintes qui sont liées, soit à des considérations statistiques, soit à la nature du signal de parole.

Notons qu'une dégradation croissante des performances est observée au fur et à mesure que le temps qui sépare la session d'enregistrement de la session de test augmente. De plus, le comportement des locuteurs se modifie lorsque ceux-ci s'habituent au système. Les modèles des locuteurs doivent donc être régulièrement mis à jour avec les nouvelles données d'exploitation du système. Les altérations de la voix dues à l'état physique (fatigue, rhume) ou émotionnel (stress), lorsqu'ils sont importants, peuvent mettre aussi en échec l'efficacité de certains systèmes.

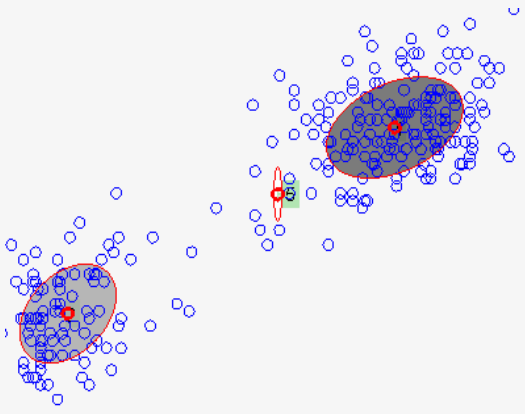
V.CONCLUSION

A travers ce chapitre, nous avons introduit le principe de la reconnaissance automatique du locuteur ainsi que les différentes étapes du système. La reconnaissance automatique du locuteur est probablement la méthode la plus ergonomique pour résoudre des problèmes d'accès. Actuellement en plein essor, et ayant une bonne acceptabilité, elle a de nombreuses applications potentielles comme la sécurisation accrue des téléphones portables, contrôle supplémentaire au niveau d'une application sur un site comme l'accès sécurisé à un bâtiment ou remplacement du mot de passe sur les ordinateurs.

Le principal avantage de cette technique est d'autoriser une authentification à distance. Ces applications concernent la vérification du locuteur à travers le réseau téléphonique pour accéder à un service ou pour le réseau internet pour sécuriser l'accès à un site web. Cependant la voix ne peut pas être considéré comme une caractéristique biométrique d'une personne compte tenu des la variabilité intra-locuteur.

4

Modélisation des paramètres biométriques: *classificateur O-GMM*



I. INTRODUCTION

Dans les chapitres précédents nous avons expliqué le principe des techniques d'extraction de paramètres que sont les coefficients DCT pour le visage et MFCC pour la voix et justifié leur choix pour chacune des modalités voix et visage. On obtient alors un nuage de points représentant la voix et le visage de chaque personne.

Nous devons à présent modéliser les personnes, a partir de leurs caractéristiques discriminantes, en leur associant une fonction mathématique à laquelle l'individu testé sera comparé.

La majorité des approches de modélisations utilisées en reconnaissance sont souvent présentées sous un formalisme probabiliste se basant généralement sur une ou plusieurs approches statistiques. Parmi les approches citées dans les chapitres II et III, nous avons choisis d'utiliser dans le cadre de notre projet les Modèles de Mélanges de Gaussiennes GMM.

II. MOTIVATION LIEE A LA MODELISATION PAR GMM

Les mélanges de gaussiennes ont été introduites pour la modélisation de paramètres biométrique, et tout particulièrement la modélisation du visage ainsi que de locuteur en mode indépendant du texte, et ceci à travers les travaux de thèse de Douglas Reynolds [Rey 95].

L'utilisation d'un mélange GMM se justifie essentiellement en faisant appel à l'interprétation des classes de mélange : il est certain que les vecteurs de paramètres biométriques, que ce soient les paramètres DCT ou MFCC, vont se répartir différemment selon les caractéristiques de l'image ou de la parole considérée. Chaque composante modélisera des ensembles sous-jacents de classes qui représenteront différentes variations. L'avantage principal de cette modélisation vient du cœur même de son élaboration qui est d'utiliser une combinaison linéaire de densités de probabilités gaussiennes multidimensionnelles pour représenter une distribution de probabilité aléatoire très complexe [Ben 04]

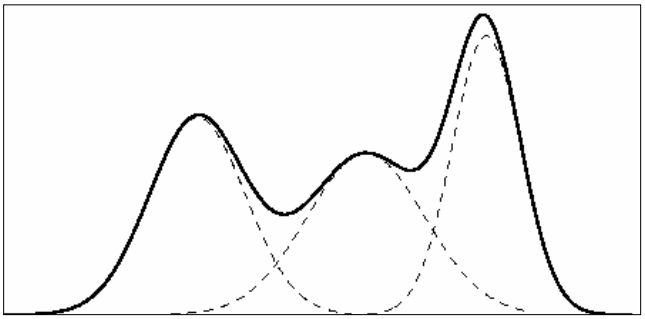


Figure 4.1- Approximation de la distribution d'un paramètre biométrique par une combinaison de gaussiennes.

Concernant la reconnaissance de visages, les GMM ont largement prouvé leur efficacité ainsi que leur rapidité. Par ailleurs, cette approche permet une approximation à large gamme des distributions complexes dans l'espace de représentation et ceci avec manière simple et efficace.

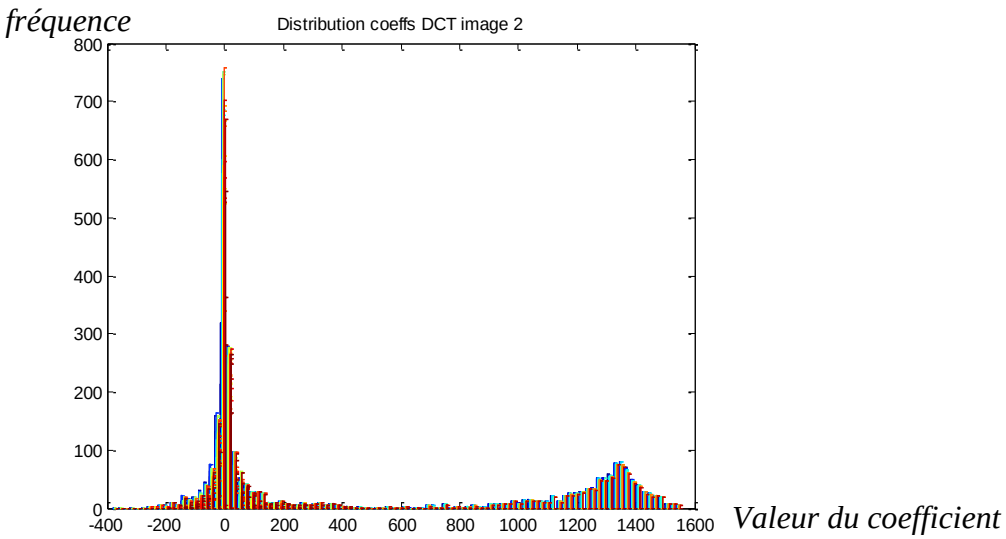


Figure 4.2-Distribution de l'ensemble des coefficients DCT de l'image n°2

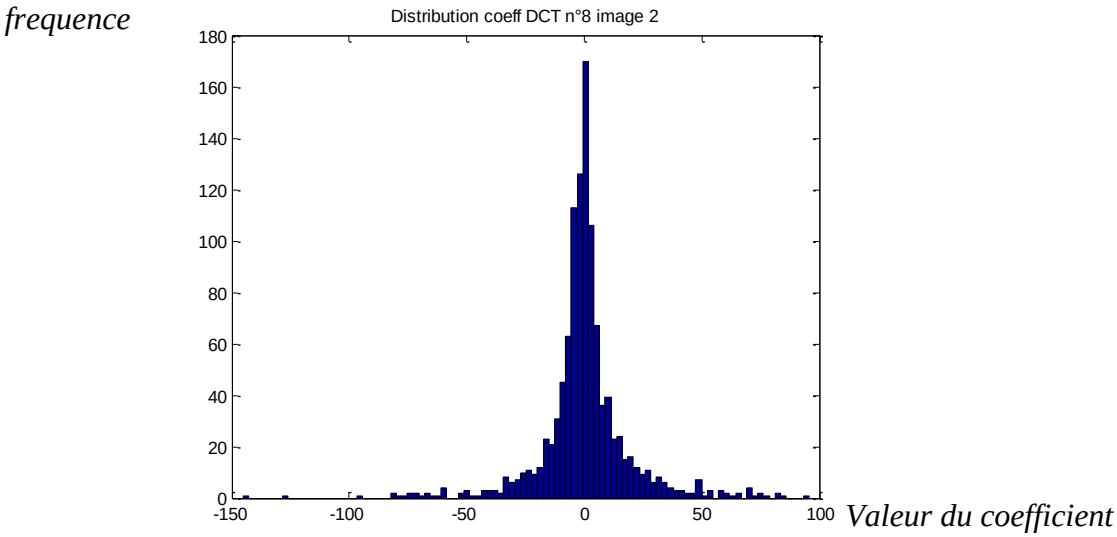


Figure 4.3-Distribution du 8eme coefficient DCT de l'image n°2

Concernant la reconnaissance du locuteur, les GMM ont largement contribué ces dix dernières années à améliorer les performances des systèmes leur correspondant. En effet, c'est l'approche statistique qui apparaît comme étant la plus adaptée pour capter les caractéristiques stochastiques du signal de parole. Ceci étant dû au fait que :

- ✓ Chaque distribution dans un modèle GMM est capable de représenter la structure spectrale d'une large classe phonétique. Ces classes représentent des configurations du conduit vocal spécifiques au locuteur et donc utiles pour la modélisation du locuteur.
- ✓ Une mixture de distributions gaussiennes donne une représentation approximative de la distribution à long terme de vecteurs acoustiques provenant des énoncés du même locuteur.

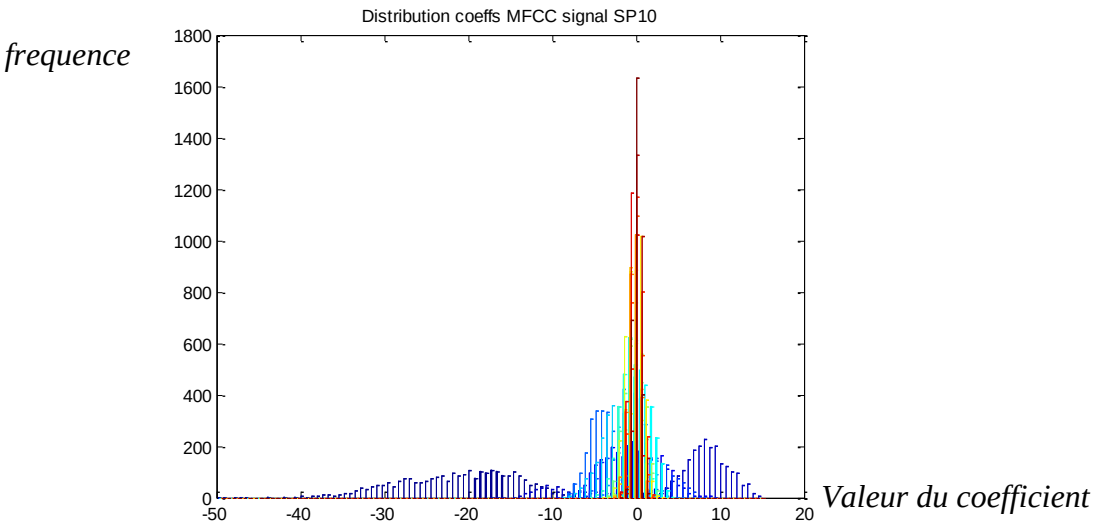


Figure 4.4-Distribution de l'ensemble des coefficients MFCC du signal SP10

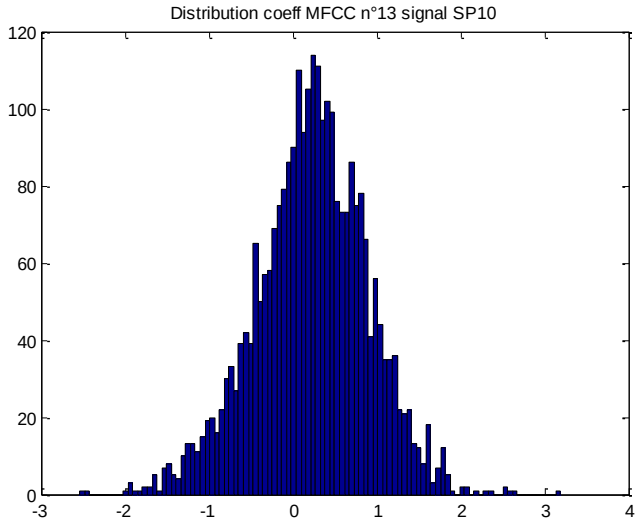


Figure 4.5-Distribution du 13eme coefficient MFCC du signal SP10

III. GENERALITES SUR LES STATISTIQUES GAUSSIENNES

III.1 Formule et définitions

La fonction de densité de probabilité gaussienne pour une variable x à d dimensions $x \cup \mathcal{N}(\mu, \Sigma)$ est donnée par :

$$f_{(\mu,\Sigma)}(x) = \frac{1}{\sqrt[d]{2\pi}\sqrt{\det(\Sigma)}}\, e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \qquad [4\text{-}1]$$

Ou μ est le vecteur moyen, Σ la matrice covariance et x représente le vecteur de paramètres (coefficients DCT ou MFCC dans notre cas) et $(x-\mu)^T$ la matrice transposée de $(x-\mu)$

III.2 Estimation du vecteur moyen

Le vecteur moyen μ représente la moyenne des échantillons suivant chaque dimension. Il représente l'espérance mathématique d'une variable aléatoire .

Il est estimé par la formule : $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$ [4-2]

III.3 Estimation de la covariance

La matrice covariance est une matrice carrée symétrique de dimension $(d \times d)$ de la forme :

$$\Sigma = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \qquad [4\text{-}3]$$

Dont les éléments a_{ij} représentent les coefficients de variance ou de covariance.

La diagonale de la matrice représente la variance de la variable sur chaque intervalle, c'est-à-dire son étalement.

$$a_{ii} = E[x_i^2] - E[x_i]^2 \qquad [4\text{-}4]$$

Tandis que les éléments du triangle supérieur, égaux à ceux du triangle inferieur représente les covariances, c'est-à-dire la corrélation entre chaque dimension du vecteur aléatoire (qui représentent chacune une variable aléatoire).

$$a_{ij} = E[x_i x_j] - E[x_i]E[x_j] \qquad [4\text{-}5]$$

La covariance est estimée par l'estimateur :

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \qquad [4\text{-}6]$$

III.4 L'effet de la matrice covariance sur la forme des gaussiennes

Prenons des gaussiennes a deux dimensions pour pouvoir les représenter dans le plan.

Suivant les valeurs de ses coefficients, nous obtenons différentes gaussiennes :

- Matrice covariance pleine (full) :

Les quatre coefficients sont non nuls. Nous obtenons donc une forme elliptique, dont la largeur suivant chaque dimension est fonction de l'étalement sur chaque variable aléatoire, et l'orientation est fonction de la covariance.

- Matrice covariance diagonale :

Les variances sont différentes mais la covariance est nulle, les deux dimensions sont décorrelées nous obtenons alors une sphère dont les axes sont parallèles aux axes des deux dimensions.

❖ **Remarque**

Dans le cas ou les variances sont égales, nous parlons d'une matrice covariance sphérique (processus sphérique) du fait que la forme a n dimension est une hyper sphère (les axes de l'ellipse étant égaux mais non parallèles aux axes des dimensions).

IV. Modélisation par Mélanges de Gaussiennes GMM

Cette modélisation est une approche statistique qui consiste à estimer une loi de probabilité inconnue à l'aide d'une combinaison de plusieurs Gaussiennes dont les paramètres sont à calculer [Rey 95].

IV.1 Présentation d'un modèle de mélange :

Une densité de mélange de gaussiennes est une somme pondérée de M densités Gaussiennes. Cette dernière est donnée par l'équation :

$$p(x|\lambda) = \sum_{m=1}^{M} \pi_m b_m(x) \qquad [4\text{-}7]$$

Ou :
x : représente un vecteur aléatoire de dimension D.
$b_m(x)$: représente les densités de probabilités Gaussiennes paramétrées par le vecteur moyenne μ_m et une matrice de Σ_m.
π_m : représente le poids des Gaussiennes.
En précisant que :

$$\sum_{m=1}^{M} \pi_m = 1 \qquad [4\text{-}8]$$

$$b_m(x) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_m|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu_m)^t (\Sigma_m)^{-1}(x-\mu_m)\right] \qquad [4\text{-}9]$$

Un modèle GMM est donc représenté par les vecteurs moyens, les matrices de covariance et les poids des gaussiennes, ceci est définit par la notation :

$$\lambda = \{\pi_m , \mu_m , \Sigma_m\} \quad [4\text{-}10] \qquad \text{Avec : m=1,…, M}$$

Cette famille de modèles est bien adaptée pour approximer les densités de probabilités réelles multidimensionnelles. En effet, en augmentant le nombre M de composantes Gaussiennes on pourra modéliser n'importe qu'elle loi probabiliste. Cependant l'augmentation du nombre de Gaussiennes intensifie le temps de calcul, à la fois pour l'étape d'estimation mais aussi pour le calcul de scores lors de l'authentification. D'autre part, il faut d'avantage de données pour estimer de façon fiable un nombre important de paramètres [Rey 95].

IV.2 Apprentissage du modèle GMM :

Il s'agit, lors de la phase d'apprentissage, d'estimer l'ensemble des paramètres d'un modèle GMM en utilisant des mesures de similarités entre un énoncé x et un modèle de l'individu λ, c'est à dire une approximation de la probabilité $p(x|\lambda)$ que l'énoncé observé ait été produit par la personne considérée.

La méthode conventionnelle est celle du Maximum de Vraisemblance dont le but est de déterminer les paramètres du modèle qui maximisent la vraisemblance des données d'apprentissage [Cam 97]. Pour une séquence de N vecteurs d'apprentissage $X = \{x_1, x_2, \dots , x_N\}$ suffisamment indépendants, la vraisemblance du modèle GMM est donnée par :

$$p(X|\lambda) = \prod_{n=1}^{N} p(x_n|\lambda) = \prod_{n=1}^{N} \sum_{m=1}^{M} p(x_n|\, \pi_m , \mu_m , \Sigma_m) \qquad [4-11]$$

Dans le but de simplifier cette équation, il est intéressant d'introduire l'opérateur logarithmique ce qui nous fournira un « log de vraisemblance » donné par :

$$V(X,\lambda) = \frac{1}{N} \log \prod_{n=1}^{N} p(x_n|\lambda) = \frac{1}{N} \sum_{n=1}^{N} \log p(x_n|\lambda) \qquad [4\text{-}12]$$

On obtient donc une expression complexe de la vraisemblance contenant le logarithme d'une somme et une fraction non linéaire des paramètres du modèle λ ce qui rend la maximisation directe assez difficile.

Cependant, la variable indicatrice « m » est une donnée constitutive du problème qui présente l'inconvénient de ne pouvoir être observé en pratique. En effet on observe des réalisations du vecteur aléatoire x_n sans savoir de manière certaine quelle est la classe du mélange associé à chaque observation. Au sens de l'algorithme EM (Expectation-Maximisation) (Annexe C), la variable « m » constitue une donnée manquante ou non-observée.

Nous verrons que l'introduction de ces données permet de résoudre de manière élégante un problème d'estimation relativement complexe et que ce type de problème est adapté à l'algorithme d'apprentissage EM. Notons que la fonction de vraisemblance, de par sa nature discriminante, servira comme outil d'aide à la décision.

IV.3 Estimation du modèle GMM par L'algorithme EM (Expectation- Maximisation) :

L'algorithme Expectation- Maximisation (Annexe C) qui peut être traduit par Estimation-Maximisation est un algorithme itératif de type sous optimal qui permet d'atteindre un maximum local dans l'espace de solutions des modèles GMM. Cet algorithme a pour objectif de maximiser, de manière itérative, la fonction de vraisemblance $p(X|\lambda)$ en faisant en sorte qu'elle atteigne un maximum local [Tom 04].

L'idée principale est, qu'à travers l'initialisation des paramètres initiaux, on estime de nouveaux paramètres $\bar{\lambda}$ de sorte que la vraisemblance du nouveau modèle soit supérieure ou égale à la vraisemblance du modèle initial, En d'autres termes, $p(X|\bar{\lambda}) \geq p(X|\lambda)$. Cet algorithme se compose de deux paliers. Le premier étant l'initialisation du modèle par quantification vectorielle. Le second étant l'optimisation de ce dernier qui se fait en deux étapes : estimation et maximisation [Bil 97].

IV.3.1 Initialisation :

Etant donné que l'algorithme EM est itératif et que la solution de ce dernier converge vers un maximum local, la phase d'initialisation joue alors un rôle prépondérant dans la détermination du résultat. L'étape d'initialisation se déroule comme suit :

Initialisation des moyennes μ_m à l'aide de l'algorithme LGB (Annexe B).

Initialisation équiprobable des poids π_m des M Gaussiennes tel que : $\pi_m = 1/M$.

Initialisation des matrices de covariance Σ_m à la matrice identité.

IV.3.2 Estimation :

Pour tout vecteur de données $= \{x_n;\ n = 1, \dots, N\}$, représentant les coefficients DCT ou MFCC, le calcul de la probabilité que ce dernier soit généré par la Gaussienne de classe « i » est donnée par :

$$P_{ni} = \frac{\pi_i\, b_i(X)}{\sum_{m=1}^{M} \pi_m b_m(X)} \qquad [4\text{-}13]$$

Cette étape est équivalente à avoir un ensemble Q de variables continues cachées, prenant des valeurs dans l'intervalle [0,1] qui donne la proportion qu'un vecteur X appartient à la gaussienne « i ».

IV.3.3 Maximisation :

Lors de cette étape, nous procédons à la réestimation des paramètres des modèles tel que :

$$\pi_i{}^* = \frac{1}{N}\sum_{n=1}^{N} P_{ni} \qquad [4\text{-}14]$$

$$\mu_i{}^* = \frac{\sum_{n=1}^{N} P_{ni}\ x_n}{\sum_{n=1}^{N} P_{ni}} \qquad [4-15]$$

$$\Sigma_i{}^* = \frac{\sum_{n=1}^{N} P_{ni}\ (x_n - \mu_i{}^*)(x_n - \mu_i{}^*)^t}{\sum_{n=1}^{N} P_{ni}} \qquad [4-16]$$

Les étapes d'estimation et de maximisation seront répétées jusqu'à atteindre un certain seuil de convergence.

IV.4 Autres algorithmes d'estimation : algorithme EM

Estimer un modèle par la méthode ML revient à proposer la valeur du modèle qui rend maximale la vrais semblance $p(X|\lambda)$ [Dem 77].

Il est dit que le modèle $\hat{\lambda}$ est une estimation par maximum de vraisemblance pour les données X si :

$$V(X,\hat{\lambda}) \geq V(X,\lambda) \qquad \forall\lambda$$

Ceci peut être traduit par :

$$\hat{\lambda} = \arg max\, V(X,\lambda)$$

Les paramètres $\hat{\lambda} = \{\hat{\pi}_m, \hat{\mu}_m, \hat{\Sigma}_m\}$ n'étant rien d'autres que les points qui annulent les dérivées partielles de la fonction de vraisemblance. La méthode d'estimation est théoriquement très efficace, cependant en pratique elle engendre des équations non linéaires complexes qui font leur apparition lorsque le nombre de Gaussiennes devient trop grand.

IV.5 Modélisation de l'imposteur par modèle UBM (Universal Background Model) :

Le modèle UBM (Universal background model) fait en sorte que le modèle des imposteurs soit un modèle unique, indépendant et commun à toutes les personnes de la base de données. Il vise à approximer la fonction de densité $p(X|\bar{Y})$de la personne X sous l'hypothèse qu'elle ait été produite par un imposteur $\bar{Y}$ [Gau 94].

Pour la construction du modèle UBM, plusieurs approches peuvent être employées. L'approche la plus simple est de collecter toutes les données d'apprentissage pour former un seul modèle (UBM) à l'aide de l'algorithme EM tout en équilibrant les sous populations pendant le choix des données. La combinaison est ensuite faite selon l'une des fonctions : minimum, maximum ou la moyenne entres les différentes sous populations du modèle [Car 92].

$$p(X|\bar{Y}) = f\{p(X|\bar{Y}_1), \dots, p(X|\bar{Y}_N)\} \qquad [4\text{-}17]$$

Il existe différentes approches de calcul du modèle UBM :

Combinaison des paramètres biométriques des clients, puis création du modèle UBM (μ) correspondant.

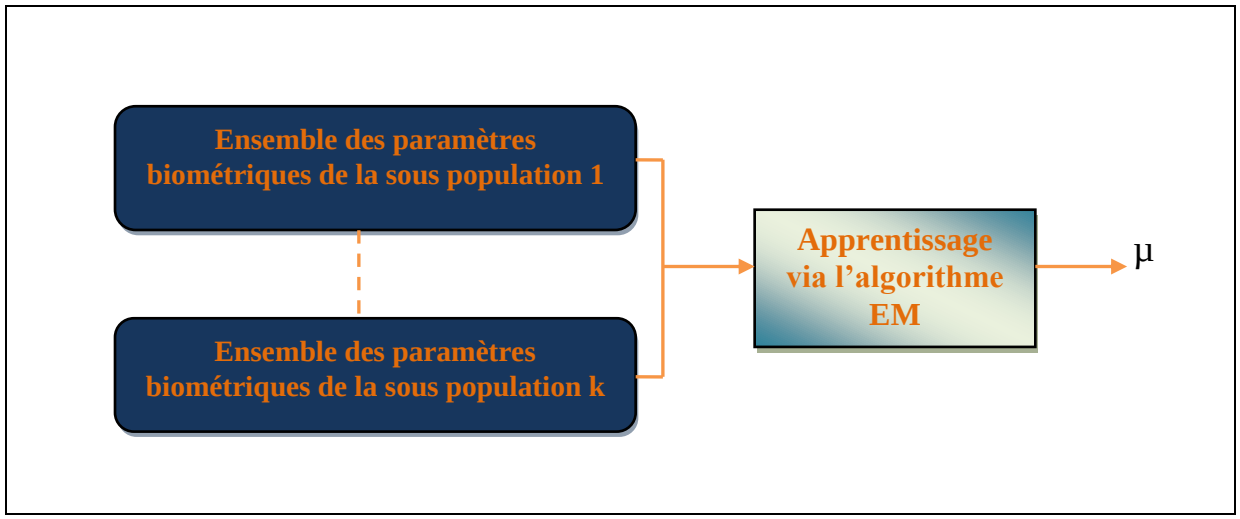


Figure 4.6-Première approche pour la génération du modèle UBM

Combinaison des différents modèles UBM émanant des différentes populations.

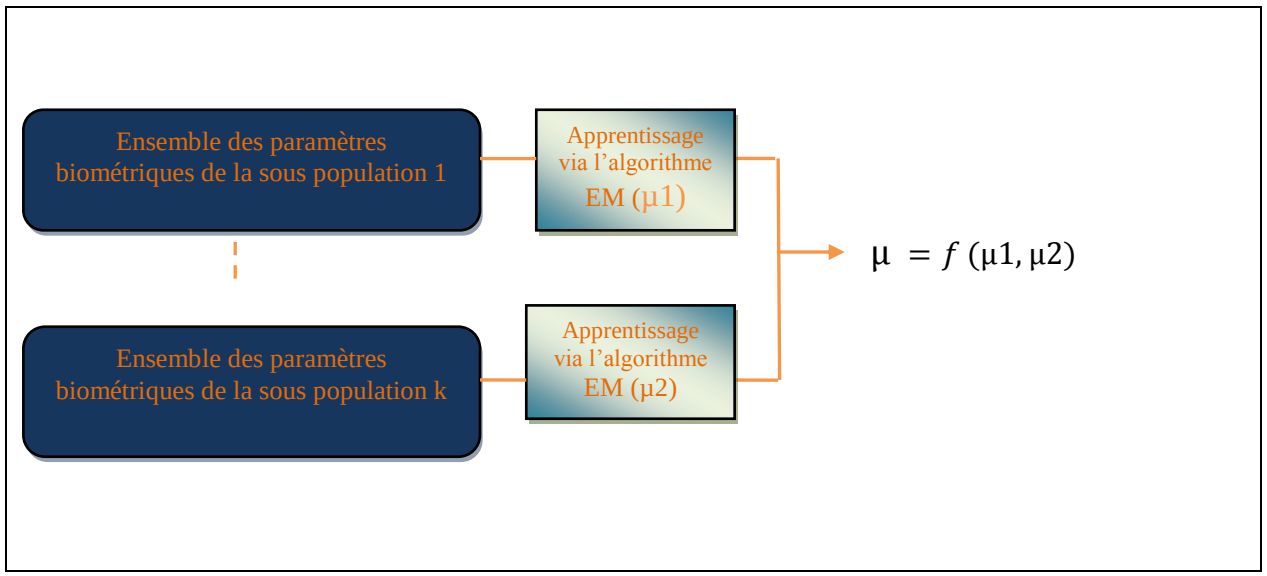


Figure 4.7-Seconde approche pour la génération du modèle UBM

f est une fonction qui peut être soit le maximum ou le minimum ou la moyenne.

IV.6 L'apport de l'orthogonalité :

IV.6.1 Les différents types de GMM

- ✓ GMM full.

Le classificateur GMM full, utilise des gaussiennes pleines lors de l'estimation du modèle d'une personne. Les dimensions ne sont pas supposées decorrélées et les matrices covariance ont une dimension de DxD, où D est la dimension des vecteurs de paramètres (nombre de coefficients DCT ou coefficients MFCC).
Bien qu'elle présente le cas optimal du point de vue des calculs, elle a un temps de calcul trop important pour pouvoir être utilisée dans la pratique.

- ✓ GMM diagonale.

Le classificateur GMM diagonal est une amélioration du classificateur GMM full, dont il allège les calculs. En effet, au lieu de considérer des matrices covariances pleines comme dans le cas précédent, le classificateur GMM diagonal néglige les covariances entre les dimensions et ne garde que la diagonale de la matrice (les variances), réduisant la taille de la covariance de DxD à Dx1. Mais cette simplification a pour conséquence une réduction de la précision lors de l'estimation des modèles et donc des performances moins intéressantes.

- ✓ GMM orthogonale.

Le classificateur GMM orthogonal essaye de rallier la rapidité des GMM diagonale avec la précision des GMM full. Son principe est d'appliquer une transformation de l'espace de travail pour travailler dans un espace ou les matrices covariance seront réellement diagonales. Il devient alors possible d'utiliser les propriétés de travailler de la même façon qu'avec les GMM diagonales, tout en faisant en sorte que l'information contenue dans les covariances de la matrice full soit contenue dans la diagonale de la nouvelle matrice covariance.

IV.6.2 Orthogonalisation par la KLT

Dans cette approche on applique une opération de diagonalisation à la matrice covariance décrivant tous les points d'apprentissage de chaque personne. On applique, alors une transformation similaire à la KLT.[LIU 99]

La KLT (Karnunen-Loeve-Transformation), comme on l'a vu dans le deuxième chapitre, est une transformation qui permet une decorrelation parfaite des données contrairement à la DCT. Son application après l'extraction des paramètres DCT et MFCC ne peux qu'améliorer l'effet de decorrelation que eux même possèdent.

Apres diagonalisation on obtient les vecteurs propres qui forment la base propre de chaque personne et qui seront associées à son modèle lors de la phase d'apprentissage. Lors de la phase de test, les nouvelles données extraites seront à chaque fois projetées dans la base de chaque personne pour pouvoir être comparée à son modèle.

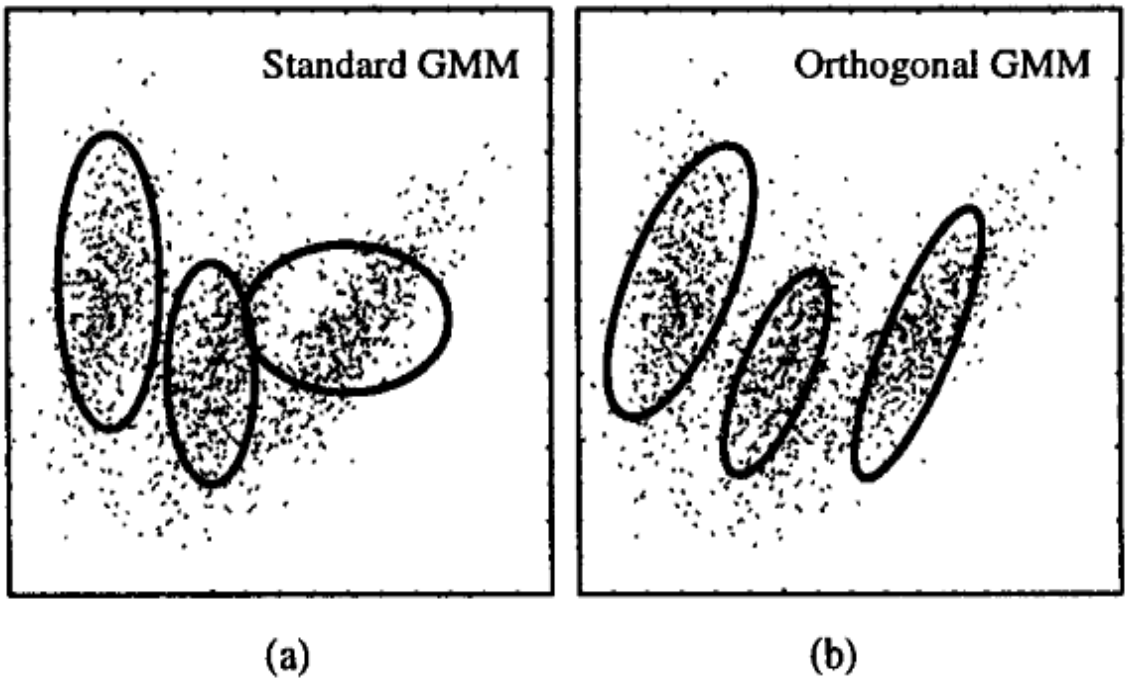


Figure 4.8-Estimation des données par GMM et OGMM (a) GMM et (b) OGMM

Ainsi l'allure dans l'espace des GMM diagonale de covariances d'OGMM tend vers celle des GMM full. L'observation (b) étant faite dans l'espace de travail des gaussiennes dans (a), on observe la présence des covariances, mais après projection on travaille dans un espace ou les gaussiennes ont des axes parallèles aux axes dimensionnels. L'algorithme EM est alors généré de la manière habituelle.

IV.6.3 Orthogonalisation par la PCA généralisée

Cette approche, bien que similaire à la précédente, nécessite beaucoup moins de calculs. En effet, au lieu de travailler dans l'espace propre associé à chaque personne, on travaille dans l'espace propre de toutes les données d'apprentissage en même temps.

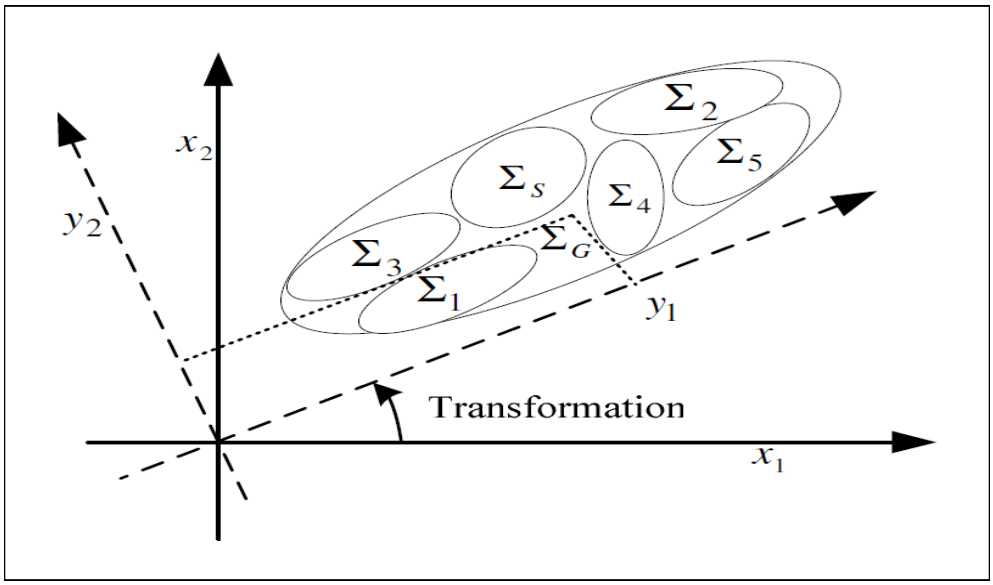


Figure 4.9-Principe de la GPCA

Comme on l'a vu précédemment, les coefficients DCT et MFCC permettant déjà une certaine corrélation, on s'intéresse plutôt à changer l'allure ou la disposition de tous les points de la base. L'algorithme EM travaille alors dans l'espace $(y1, y2)$ plutôt que $(x1, x2)$.

IV.7 Génération des scores et décision

IV.7.1 Génération des scores clients-imposteurs

Le score correspond à la probabilité d'un échantillon test sachant un modèle en mémoire. Son rôle varie en mode authentification et identification.

IV.7.2 Mode authentification

D'après la théorie Bayésienne de la décision (Annexe D), le rapport des vraisemblances $p(X|Y)$ ainsi que $(X|\bar{Y})$ comparé à un seuil permet d'obtenir des informations sur les performances du classificateur. C'est donc tout naturellement que cette statistique s'est imposée comme score d'authentification pour ce type de systèmes [Auc 00].

Pour des raisons liées à la précision des calculs et à la maniabilité des valeurs, c'est en général le logarithme de ce rapport de vraisemblance LLR (Log Likelihood Ratio) qui est utilisé.

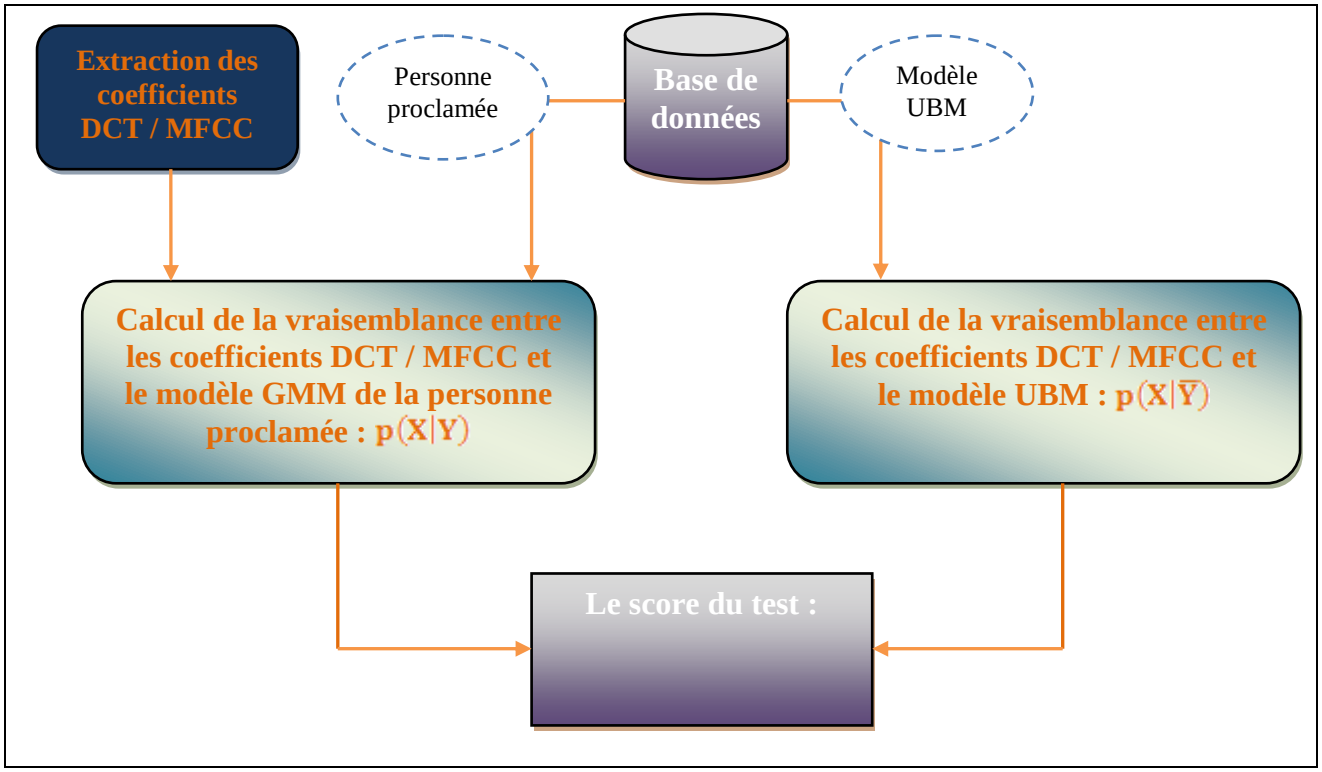


Figure 4.10-Génération des scores dans le mode d'authentification

✓ Mode identification :

Dans ce mode, le résultat est un ensemble de N scores où N est le nombre des personnes enregistrés dans la base de données et chaque score représente la vraisemblance entre les paramètres test et le modèle sauvegardé dans la base [Ros 92].

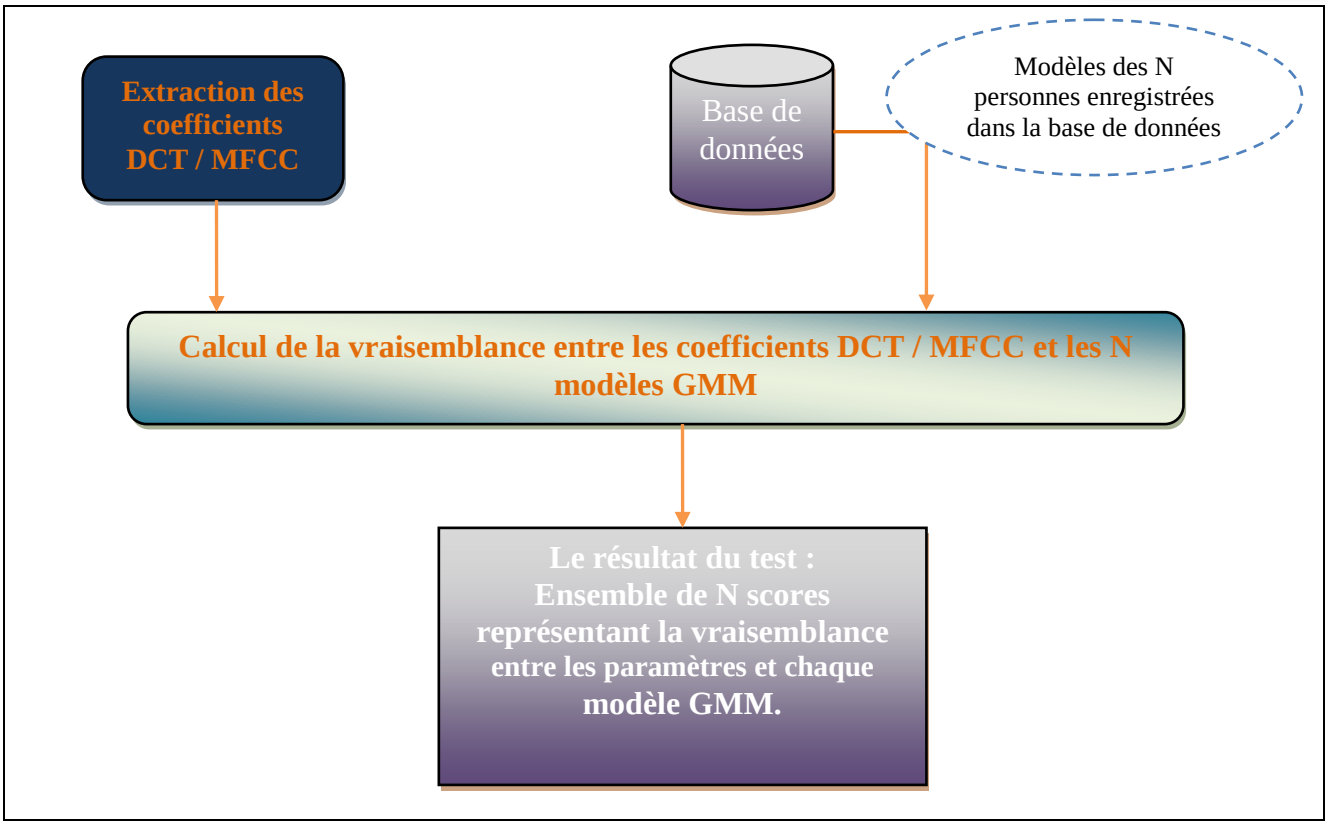


Figure 4.11-Génération des scores dans le mode d'identification

IV.7.2 Décision

Dans ce qui suit, nous expliquerons la manière de trancher lors de la prise de décision pour une application de reconnaissance, de visage ou du locuteur ce qui pourra être vu comme une déclinaison des processus de décision principaux que sont l'identification et l'authentification [Mam 03].

- ✓ Mode authentification :

Pour une personne testée X et une identité proclamée Y viennent deux hypothèses H_0 et H_1 avec :

H_0 : « L'identité de la personne X correspond bien à l'identité de la personne Y ».

H_1 : « L'identité de la personne X correspond à l'identité d'une autre personne $\bar{Y}$ ».

La décision doit se prendre en fonction de la vraisemblance des deux hypothèses concurrentes, mais aussi en fonction des coûts associés au choix à tort de chacune des deux (C_{FA} coût de fausses acceptations et C_{FR} coût de faux rejets). Le problème de décision se résout

dans le cadre de la théorie de la décision Bayésienne (Voir Annexe D) et cela par le test du rapport de vraisemblance LRT (Likelihood Ratio Test) donné comme suit :

$$LRT = \frac{p(X|H_0)}{p(X|H_1)} \lessgtr \theta \begin{cases} H_1 \ acceptée \\ H_0 \ acceptée \end{cases} \qquad [4\text{-}18]$$

Ou :

θ : est le seuil dépendant de la valeur de Y.

H_0 : correspond au modèle de Y (client).

H_1 : correspond au rejet de Y (imposteur).

La valeur théorique optimale du seuil de décision θ est donnée par :

$$\theta = \frac{p(\mathrm{X})C_{FA}}{p(\overline{\mathrm{X}})C_{FR}} \quad [4\text{-}19]$$

Ou $p(\mathrm{X})$ et $p(\overline{\mathrm{X}})$ représentent les probabilités à priori des deux hypothèses précédentes. Cependant, en pratique, le seuil de décision n'est jamais optimal, il doit être réajusté pour chaque personne considérée. Les fonctions de vraisemblance sont des densités de probabilité des modèles statistiques correspondant à la personne X. Ces modèles ne sont en effet qu'une estimation des modèles exacts et donc peuvent induire un biais au rapport de vraisemblance.

Afin de remédier à cela, le rapport obtenu lors de la phase de test est normalisé pour garder le seuil de décision fixe pour toutes les images des personnes. Le but est de stabiliser le plus possible le seuil et par conséquent avoir une procédure de recherche d'un unique seuil optimisant les performances du système pour l'ensemble des personnes.

- ✓ Mode identification :

considérons un groupe de K personnes, représentées par les modèles GMM $\lambda_1,\ldots,\lambda_K$. L'objectif de la phase d'identification est de trouver, à partir d'une séquence X, le modèle qui a la probabilité à posteriori maximale, c'est-à-dire :

$$\hat{S} = \arg\max p(\lambda_s|X) \qquad [4\text{-}20]$$

Ce qui d'après la loi de Bayes donne :

$$\hat{S} = \arg\max \frac{p(X|\lambda_s)}{p(X)} p(\lambda_s) \qquad [4\text{-}21]$$

En supposant l'équiprobabilité d'apparition des personnes, la loi devient :

$$\hat{S} = \arg\max p(X|\lambda_s) \qquad s = 1, \dots, K \qquad [4\text{-}22]$$

En utilisant le logarithme et l'indépendance entre les observations, le système d'identification calcule le score suivant :

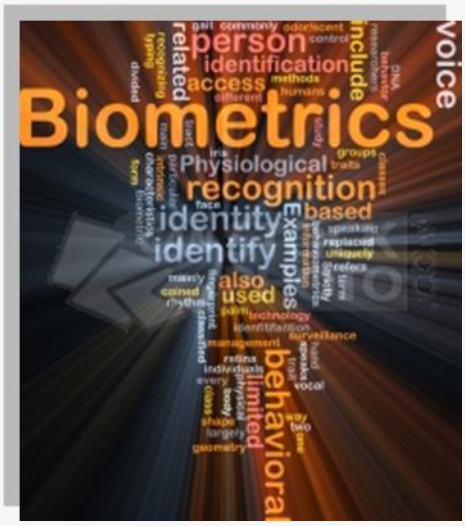
$$\hat{S} = \arg\max \sum_{n=1}^{N} log\, p(x_n|\lambda_s) \qquad [4\text{-}23]$$

V. Conclusion

Nous avons vu dans ce chapitre le fonctionnement global des GMM et plus principalement l'estimation de ses modèles par l'algorithme EM. Nous avons également pu voir l'importance d'introduire la notion d'imposteur par la normalisation UBM.

Nous nous sommes ensuite intéressé à deux approches qui permettent d'orthogonaliser nos gaussiennes. L'une d'elle est basée sur la transformée KLT (ou PCA) et l'autre sur ce qu'on appelle la GPCA.

5 *Fusion des scores et décision*



I. INTRODUCTION

De nos jours les systèmes à reconnaissance biométrique sont assez nombreux et sont déclarés plus ou moins fiables, mais les critères de performance de ces derniers ne sont pas les seuls à prendre en compte car, entrent en jeu également, les critères de coûts et d'acceptation par le public. Ainsi, selon les situations d'usage et les buts recherchés, chaque biométrie possède ses points forts ainsi que ses points faibles.

La multimodalité est une solution pour tenter de pallier ce problème du fait qu'a chaque modalité peut être associé un classificateur fournissant un score qui décidera de l'acceptation ou non d'une personne lors du processus de reconnaissance. En effet, l'utilisation de plusieurs systèmes a pour but premier d'améliorer les performances de reconnaissance. En augmentant la quantité d'informations discriminantes de chaque personne, on souhaite augmenter le pouvoir de reconnaissance du système. De plus, le fait d'utiliser plusieurs modalités biométriques réduit le risque d'impossibilité d'enregistrement ainsi que la robustesse aux fraudes. Le problème est alors de définir des stratégies pour combiner ces scores de décision, considérés indépendants [Ros 06].

II. LES DIFFERENTS SYSTEMES MULTIMODAUX

Les systèmes biométriques multimodaux diminuent les contraintes des systèmes biométriques monomodaux en combinant plusieurs systèmes [Ros 06]. On peut en différencier 5 types selon les systèmes qu'ils combinent :

II.1 Multi-capteurs :

Dans ces systèmes, un même trait biométrique est analysé à l'aide de plusieurs capteurs afin d'extraire diverses informations provenant de l'enregistrement des images. Par exemple, un système peut enregistrer le contenu de la texture 2D du visage d'une personne avec une caméra CCD et la forme de la surface 3D du visage avec une autre gamme de capteurs dans le but de procéder à la reconnaissance.

II.2 Multi-instances :

Ces systèmes utilisent tout simplement plusieurs instances d'un même trait biométrique. Par exemple l'acquisition de plusieurs images de visage avec des changements de pose, d'expression ou d'illumination peuvent être utilisée afin de vérifier l'identité d'une personne. Ces systèmes ne nécessitent généralement pas l'introduction de nouveaux capteurs, pas plus qu'ils n'entraînent le développement de nouveaux algorithmes d'extraction de caractéristiques ou de reconnaissance et sont, par conséquent, rentables.

II.3 Multi-algorithmes :

Dans ces systèmes, les mêmes données biométriques sont traitées à travers plusieurs algorithmes. Cette multiplicité d'algorithmes peut intervenir dans le module d'extraction en considérant plusieurs ensembles de caractéristiques et/ou dans le module de comparaison.

II.4 Multi-échantillons :

Un unique capteur peut être utilisé pour acquérir plusieurs échantillons du même trait biométrique dans le but de prendre en compte les variations qui peuvent se produire au sein de ce trait, ou pour obtenir une représentation plus complète du caractère sous-jacent. Par exemple, un système de reconnaissance faciale peut capturer (et enregistrer) le profil frontal du visage d'une personne ainsi que les profils gauches et droits afin de tenir compte des variations de la pose faciale. Dans ce cas les données sont traitées par le même algorithme mais nécessitent des références différentes à l'enregistrement contrairement aux systèmes multi-instances qui ne nécessitent qu'une seule référence.

II.5 Multi-biométries :

Au sens strict du terme, ces systèmes vont particulièrement attirer toute notre attention car ils permettent de combiner les preuves présentées par différentes modalités biométriques afin d'établir l'identité d'un individu. Par exemple, l'un des premiers systèmes biométriques multimodaux utilise les caractéristiques du visage de et la voix. On s'attend à ce que des traits biométriques décorrélés (comme les empreintes digitales et l'iris) fournissent une nette amélioration de la performance d'un système que des traits biométriques corrélés (comme la voix et les mouvements des lèvres). Le coût de déploiement de ce genre de systèmes est plus

dû à l'introduction de nouveaux capteurs et, par conséquent, au développement d'interfaces utilisateur appropriées.

La précision en reconnaissance peut significativement être améliorée en utilisant un nombre croissant de traits biométriques, bien que le phénomène problématique de la dimensionnalité grandissante devrait imposer une limite à ce nombre. Ce dernier limite le nombre d'attributs utilisés dans un système de classification de formes lorsque l'on possède seulement un faible nombre d'échantillons d'entraînement.

Le nombre de traits biométriques utilisés dans une application spécifique est également limité par des considérations pratiques comme le coût de déploiement, le temps d'enrôlement, le temps de retour ou encore le taux d'erreur attendu.

Tous ces types de systèmes peuvent pallier à des problèmes différents et ont chacun leurs avantages et inconvénients. Les quatre premiers systèmes combinent des informations issues d'une seule et même modalité ce qui ne permet pas de traiter le problème de la non-universalité de certaines biométries ainsi que la résistance aux fraudes, contrairement aux systèmes "multi-biométries".

En effet, les systèmes combinant plusieurs informations issues de la même biométrie permettent d'améliorer les performances en reconnaissance en réduisant l'effet de la variabilité intra-classe. Mais ils ne permettent pas de traiter efficacement tous les problèmes des systèmes monomodaux. C'est pour cette raison que les systèmes multi-biométries ont reçu beaucoup d'attention de la part des chercheurs.

De plus le gain en performance correspondant à un système multi-biométrique est affecté par la corrélation entre les scores issus des différents comparateurs biométriques. Ainsi la combinaison de deux faibles comparateurs biométriques qui ne sont pas corrélées peut entraîner une amélioration des performances plus importante que celle obtenue par la combinaison de deux fortes comparateurs biométriques positivement corrélées.

III. LES DIFFERENTS NIVEAUX DE FUSION BIOMETRIQUE

La combinaison de plusieurs systèmes biométriques peut se faire à quatre niveaux différents : au niveau des données, au niveau des caractéristiques extraites, au niveau des scores issus du module de comparaison ou au niveau des décisions du module de décision.

Ces quatre niveaux de fusion peuvent être classés en deux sous-ensembles : la fusion pré-classification (avant correspondance) et la fusion post-classification (après la correspondance) [Ros 06].

III.1 Fusion avant la correspondance ("matching") :

Avant le matching, l'intégration d'informations peut avoir lieu soit au niveau capteur, soit au niveau caractéristique.

III.1.1 Fusion au niveau des capteurs (Sensor Level) :

Les données brutes ("raw data") provenant des capteurs sont combinées par fusion au niveau capteur. La fusion au niveau capteur peut se faire uniquement si les diverses captures sont des instances du même trait biométrique obtenu à partir de plusieurs capteurs compatibles entre eux ou plusieurs instances du même trait biométrique obtenu à partir d'un seul capteur. De plus, les captures doivent être compatibles entre eux et la correspondance entre les points dans les données brutes doit être connue par avance.

Par exemple, les images de visage obtenues à partir de plusieurs caméras peuvent être combinées pour former un modèle 3D du visage. La fusion au niveau capteur n'est généralement pas possible si les instances des données sont incompatibles. Il est peut donc être difficile de fusionner des images de visages provenant de caméras ayant des résolutions différentes.

III.1.2 Fusion au niveau des caractéristiques (Feature Level) :

La fusion au niveau caractéristiques consiste à combiner différents vecteurs de caractéristiques qui sont obtenus à partir d'une des sources suivantes : plusieurs capteurs du même trait biométrique, plusieurs instances du même trait biométrique, plusieurs unités du même trait biométrique ou encore plusieurs traits biométriques. Quand les vecteurs de caractéristiques sont homogènes, un unique vecteur de caractéristiques résultant peut être calculé comme une somme pondérée des vecteurs de caractéristiques individuels. Lorsque les vecteurs de caractéristiques sont hétérogènes nous pouvons les concaténer pour former un seul vecteur de caractéristiques. Cependant, la concaténation n'est pas possible lorsque les ensembles de caractéristiques sont incompatibles.

Les systèmes biométriques qui intègrent l'information à une étape en amont du traitement sont censés être plus efficaces que les systèmes qui opèrent une fusion à un niveau plus abstrait. Puisque les caractéristiques issues d'une entrée biométrique sont supposées contenir une information plus riche qu'un score de correspondance ou la décision d'un module de reconnaissance biométrique, la fusion au niveau caractéristiques devrait fournir de meilleurs résultats de reconnaissance que les autres niveaux d'intégration. Cependant, la fusion aux niveaux caractéristiques est difficile à atteindre en pratique et ceci pour les raisons suivantes :

- ✓ La relation entre les espaces de caractéristiques de différents systèmes biométriques n'est pas forcément connue. Dans le cas ou la relation est connue par avance, on doit prendre soin d'éliminer les caractéristiques qui sont fortement corrélées. Cela requiert l'application d'algorithmes de sélection de caractéristiques avant l'étape de classification.

- ✓ La concaténation de deux vecteurs de caractéristiques peut engendrer un vecteur de caractéristiques ayant une grande dimension. Bien que ce soit un problème général dans la plupart des applications de reconnaissance de forme, cela est encore plus marquant dans les applications biométriques à cause du temps, de l'effort et du coût impliqués dans la collecte de grandes quantités de données biométriques.

- ✓ La plupart des systèmes biométriques commerciaux ne fournissent pas l'accès aux vecteurs de caractéristiques qui sont utilisés dans leurs produits. Ainsi, très peu de chercheurs ont étudié la fusion au niveau caractéristique et la plupart d'entre eux se tournent généralement vers les schémas de fusion après le matching.

III.2 Fusion après la correspondance:

Les schémas d'intégration de l'information après l'étape de la classification ou de correspondance peuvent être divisés en trois catégories : fusion au niveau décision, fusion au niveau rang et fusion au niveau score.

III.2.1 Fusion au niveau des décisions (Decision Level) :

L'intégration d'information au niveau abstrait ou au niveau décision peut être mis en place lorsque chaque "matcher" (module de reconnaissance) biométrique décide individuellement de la meilleure correspondance possible selon l'entrée qui lui est présentée.

La fusion au niveau des décisions est souvent utilisée pour sa simplicité. En effet, chaque système fournit une décision binaire sous la forme OUI ou NON que l'on peut représenter par 0 et 1, et le système de fusion de décisions consiste à prendre une décision finale en fonction de cette série de 0 et de 1. Ces méthodes de fusion au niveau des décisions sont très simples mais utilisent très peu d'information niveau des décisions sont très simples mais utilisent très peu d'information.

III.2.2 Fusion au niveau des rangs (Rank Level)

Quand la sortie de chaque matcher biométrique est un sous-ensemble de correspondances possibles triées dans un ordre décroissant de confiance, la fusion peut se faire au niveau rang. Pour cela nous trouvons trois méthodes pour combiner les rangs assignés par différents *matchers*. Dans la technique "highest rank method", on assigne à chaque correspondance possible le meilleur (minimum) rang calculé par différents matchers. En cas d'égalité, on en retient un seul au hasard afin d'arriver à un ordre de rang strict et la décision finale est prise selon les rangs combinés. La méthode "Borda count" utilise la somme des rangs assignés par les matchers individuels afin de calculer les rangs combinés.

La méthode de “régression logistique” est une généralisation de la méthode “Borda count” où une somme pondérée des rangs individuels est calculée et les poids sont déterminés par régression logistique.

III.2.3 Fusion au niveau score (Score Level)

Après les vecteurs de caractéristiques, les scores (de correspondance) donnés en sortie par les matchers contiennent l’information la plus riche à propos du modèle d’entrée. En fait, la fusion au niveau score donne le meilleur compromis entre la richesse d’information et la facilité d’implémentation. Aussi, il est relativement facile d’accéder et de combiner les scores générés par les différents matchers. En conséquence, l’intégration d’information au niveau score est l’approche la plus courante dans les systèmes biométriques multimodaux. Il existe deux approches pour combiner les scores obtenus par différents matchers.

La première approche est de voir cela comme un problème de classification, tandis que l’autre approche est de traiter le sujet comme un problème de combinaison. Il est important de noter que les approches par combinaison sont plus performantes que la plupart des méthodes de classification. Dans l’approche par classification, un vecteur de caractéristiques est construit en utilisant les scores de correspondance donnés en sortie par les matchers individuels ; ce vecteur est ensuite attribué à une des deux classes : "accepté" (utilisateur authentique ou “genuine user”) ou "rejeté" (utilisateur imposteur ou “impostor user”). En général, le classificateur utilisé pour cette opération est capable d’apprendre la frontière de décision sans tenir compte de la manière dont le vecteur de caractéristiques a été généré. Ainsi, les scores en sortie de différentes modalités peuvent être non-homogènes (mesure de distance ou de similarité, différents intervalles de valeurs prises, etc.) et aucun traitement n’est requis avant de les envoyer dans le classificateur.

Dans l’approche par combinaison, les scores de correspondance individuels sont combinés de manière à former un unique score qui est ensuite utilisé pour prendre la décision finale. Afin de s’assurer que la combinaison de scores provenant de différentes modalités soit cohérente, les scores doivent d’abord être transformés dans un domaine commun : on parle alors de *normalisation de score.*

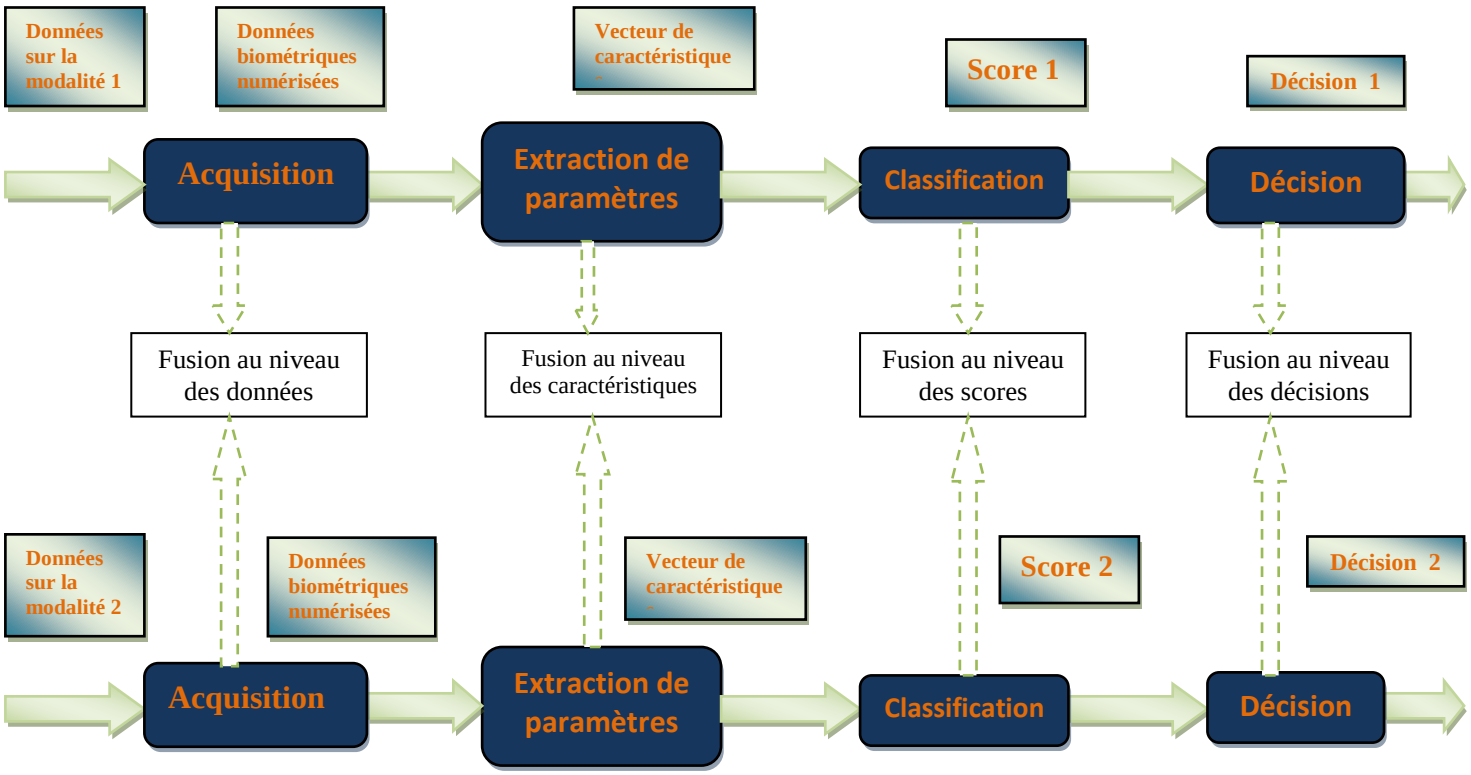


Figure 5.1-Les différents niveaux de fusion

IV. NORMALISATION DES SCORES

Les méthodes de normalisation de scores ont pour objectif de transformer individuellement chacun des scores issus des sous-systèmes pour les rendre homogènes avant de les combiner. En effet, trois problèmes importants ont besoin d'être considérés avant même de combiner les scores de correspondance en un seul et unique score.

Tout d'abord, les scores de correspondance au niveau des sorties des matchers individuels peuvent ne pas être homogènes. Par exemple, un matcher peut donner en sortie une mesure de distance (dissimilarité) pendant qu'un autre donne en sortie une mesure de proximité (similarité). Ensuite, les sorties des matchers individuels ne sont pas nécessairement inclus dans le même intervalle. Enfin, les scores de correspondance en sortie des matchers peuvent suivre différentes distributions statistiques. Pour toutes ces raisons, la normalisation de score est essentielle pour transformer les scores des matchers individuels dans un domaine

commun avant de les combiner. La normalisation de score est donc une étape critique dans la conception d'un schéma de combinaison pour la fusion au niveau score [Ros 06].

IV.1 Identification d'une technique de normalisation de scores :

La normalisation de score consiste à changer les paramètres de position (moyenne) et d'échelle (écart-type) des distributions de scores de correspondance en sortie des matchers individuels, de manière à ce que les scores de correspondance soient transformés dans un domaine commun. Quand les paramètres utilisés pour la normalisation sont déterminés en utilisant un ensemble de données d'entraînement fixé, on parle de normalisation de score fixée. Dans ce cas, la distribution des scores de correspondance de l'ensemble des données d'entraînement est examinée et un modèle cohérent est choisi pour s'adapter à la distribution. A partir de ce modèle, les paramètres de normalisation sont déterminés. Dans la normalisation de score adaptative, les paramètres de normalisation sont estimés en se basant sur le vecteur de caractéristiques actuel. Cette approche à la faculté de s'adapter aux variations de la donnée en entrée.

Pour avoir un bon schéma de normalisation, les estimateurs des paramètres de position et d'échelle de la distribution de score de correspondance doivent être robustes et efficaces. La robustesse se réfère à l'insensibilité à la présence de valeurs aberrantes quand à l'efficacité, elle se réfère à la proximité de l'estimateur obtenu par rapport à l'estimateur optimal lorsque la distribution des données est connue. Finalement, bien que de nombreuses techniques puissent être utilisées pour la normalisation de score, le défi réside dans l'identification d'une technique qui serait à la fois robuste et efficace.

IV.2 Les différentes techniques de normalisation de scores :

IV.2.1 Normalisation des scores par remise à l'échelle :

- ✓ Min-Max (MM):

C'est la plus adaptée dans le cas où les bornes des scores produits par un matchers sont connues. Dans ce cas, on peut facilement translater les scores minimums et maximums respectivement vers 0 et 1. Cependant, même si les scores de correspondance ne sont pas

bornés, on peut estimer les valeurs minimales et maximales pour un jeu de scores de correspondance donné et appliquer ensuite la normalisation Min-Max.
Cette méthode place donc les scores dans l'intervalle [0,1] tel que :

$$n_i = \frac{s_i - min_i}{max_i - min_i} \qquad [5\text{-}1]$$

Les paramètres min_i et max_i sont déterminés pour chaque sous-système sur une base de développement. La méthode du Min-Max met chaque score normalisé n_i dans l'intervalle [0,1] sous forme de score de similarité, c'est- à-dire, avec les clients proches de la borne supérieure (1) et les imposteurs proches de la borne inférieure (0).

Les valeurs minimales et maximales sont estimées à partir du jeu d'entraînement de scores donné d'où la fait que cette méthode ne soit pas robuste ce qui engendre un risque de débordement de données dans la phase opérationnelle du système si l'un des scores dépasse le maximum (ou soit inférieur au minimum). La normalisation Min-Max conserve la distribution de scores originale à un facteur d'échelle près et transforme tous les scores dans l'intervalle [0,1]. Les scores relatifs à des mesures de distance peuvent être transformés en des scores de similarité en soustrayant le score normalisé à 1.

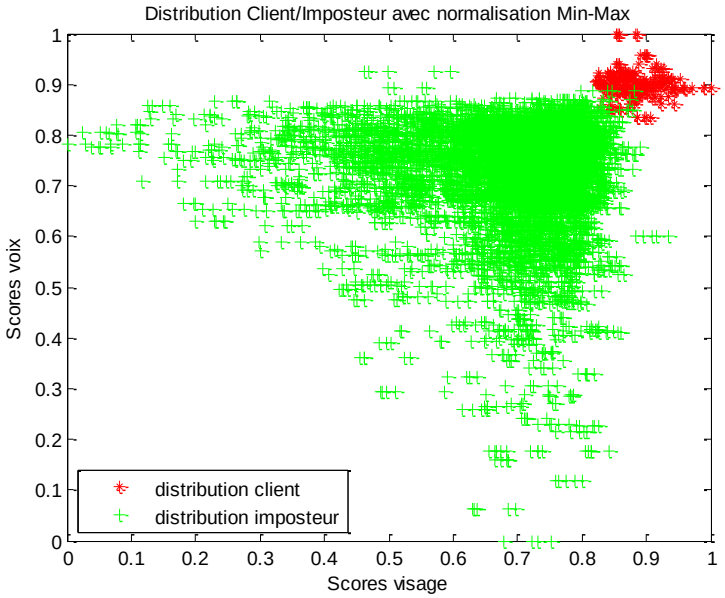


Figure 5.2-Distribution client-imposteur après normalisation Min-Max

✓ Z-Score (ZS):

La technique de normalisation de score la plus employée est certainement la Z-Score qui utilise la moyenne arithmétique et l'écart-type des données. On peut s'attendre à ce que cette méthode fonctionne bien si on a une connaissance a priori du score moyen et des variations de score d'un matcher. Si on n'a pas de connaissance a priori sur la nature de l'algorithme de reconnaissance, nous devons alors estimer la moyenne et l'écart-type des scores à partir d'un jeu de scores de correspondance donné. Les scores normalisés sont donnés par :

$$n_i = \frac{s_i - \mu_i}{\sigma_i} \qquad [5\text{-}2]$$

Les paramètres μ_i et σ_i (respectivement la moyenne et l'écart-type des scores) sont déterminés pour chaque sous-système sur une base de développement.

La normalisation "Znorm" consiste à centrer la distribution des scores en 0 et à en réduire la variance à 1. Les scores Client seront plutôt positifs et les scores Imposteur plutôt négatifs. L'effet produit par la normalisation "Znorm" ressemble à celui produit par la méthode du Min-Max car il s'agit d'une translation et un changement d'échelle, mais avec cette méthode les scores sont centrées en 0 et ne sont pas bornés.

Cependant, la moyenne et l'écart-type sont tous les deux sensibles aux valeurs aberrantes et donc cette méthode n'est pas robuste. De plus, la normalisation Z-Score ne garantit pas un intervalle commun pour les scores normalisés provenant de différents matchers. Si la distribution des scores n'est pas gaussienne, la normalisation Z-Score ne conserve pas la distribution d'entrée en sortie. Cela est simplement dû au fait que la moyenne et l'écart-type sont les paramètres de position et d'échelle optimaux seulement pour une distribution gaussienne. Pour une distribution arbitraire, la moyenne et l'écart-type sont respectivement des estimateurs raisonnables de position et d'échelle, mais ne sont pas optimaux

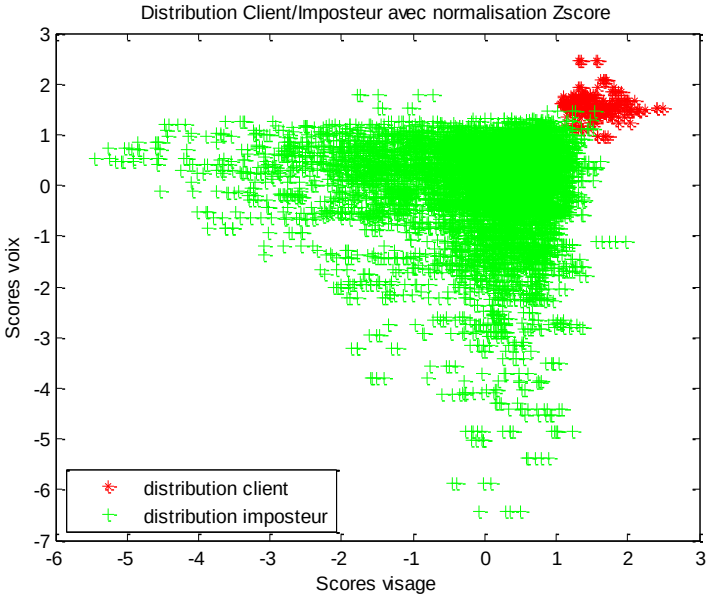


Figure 5.3-Distribution client-imposteur après normalisation Z-score

✓ Tanh (TH):

Les estimateurs Tanh (pour tangente hyperbolique), introduits par Hampel et Al,sont robustes et très efficaces rangent les scores n_i dans l'intervalle [0,1] via la normalisation donnée par :

$$n_i = \frac{1}{2}\left\{\tanh\left(0.001\,\frac{s_i-\mu_i}{\sigma_i}\right)+1\right\} \qquad [5\text{-}3]$$

Les paramètres μ_i et σ_i sont respectivement les estimateurs de la moyenne et de l'écart-type de la distribution des scores authentiques, tels qu'ils sont donnés par les estimateurs de Hampel. Ces derniers sont basés sur la fonction d'influence (ψ) suivante :

$$\psi(u) = \begin{cases} u & 0 \le |u| < a \\ a * sign(u) & a \le |u| < b \\ a * sign(u) * \left(\frac{c-|u|}{c-b}\right) & b \le |u| < c \end{cases} \qquad [5\text{-}4] \quad ; \psi(u) = 0 \quad ailleurs$$

Cette dernière réduit l'influence des points aux extrémités d'une distribution (identifiés par a, b et c) pendant l'estimation des paramètres de position et d'échelle. Ainsi, cette méthode n'est pas sensible aux valeurs aberrantes. Si plusieurs points constituant une extrémité d'une distribution ne sont plus pris en compte, l'estimateur est robuste mais pas efficace (optimal). D'autre part, si tous les points constituant l'extrémité d'une distribution sont considérés, l'estimateur n'est pas robuste mais son efficacité augmente. Par conséquent, les paramètres a, b, et c doivent être soigneusement choisis selon la quantité de robustesse exigée, ce qui dépend alternativement de l'évaluation de la quantité de bruit dans l'ensemble des données d'entraînement disponible.

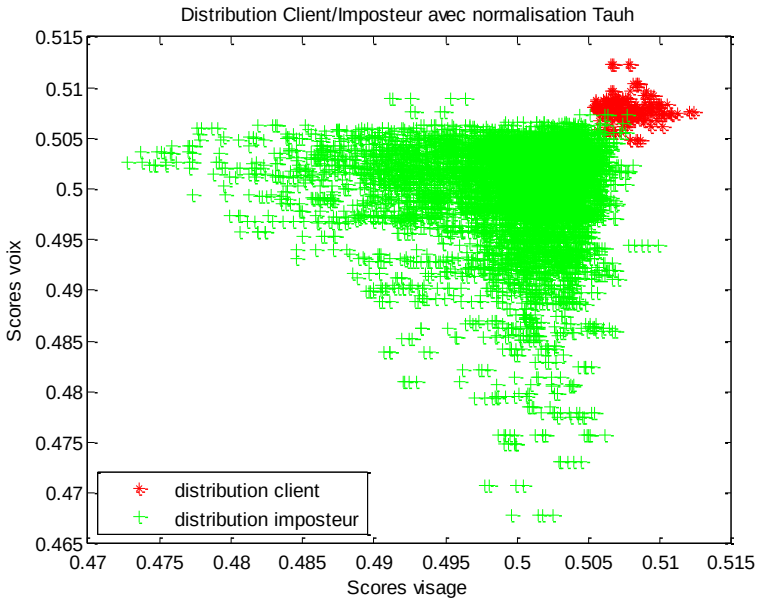


Figure 5.4-Distribution client-imposteur après normalisation Tauh

La région de chevauchement a pour largeur W tel que :

$$w = \max(s_{imposteurs}) - \min(s_{clients}) \quad [5\text{-}5]$$

Ou $s_{imposteurs}$ représente l'ensemble des scores imposteurs tant dis que $s_{clients}$ représente l'ensemble des scores clients.

Afin de minimiser au mieux l'influence de cette région sur l'algorithme de fusion, une normalisation dite adaptative est proposée dans le but de séparer au maximum les deux distributions tout en gardant les valeurs dans l'intervalle [0,1].

Pour ce faire, il existe trois fonctions dédiées à la normalisation des scores dans ce cas précis :

- **Two-quadrics(QQ) :**

Cette fonction, composée de deux segments quadratiques qui changent de cavité au point C, normalise les scores via la formulation suivante :

$$n_{AD} = \begin{cases} \frac{1}{c} {n_{MM}}^2 & si: n_{MM} \leq c \\ c + \sqrt{(1-c)(n_{MM}-c)} & sinon \end{cases} \quad [5\text{-}6]$$

Ce qui donnera la courbe suivante :

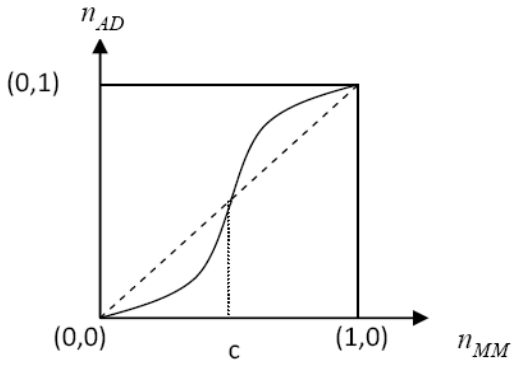


Figure 5.5-Fonction de Mapping de la méthode de normalisation QQ

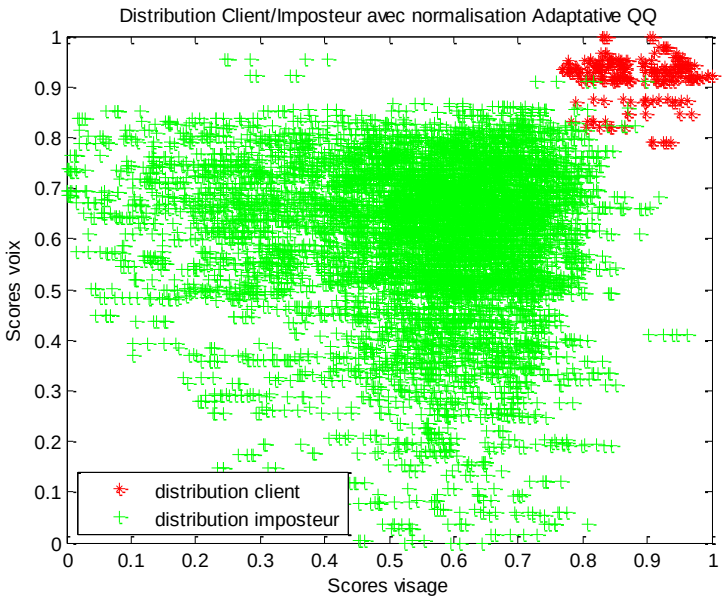


Figure 5.6-Distribution client-imposteur après normalisation Adaptative QQ

- **Logistic (LG) :**

Dans ce cas la formule de normalisation est présentée comme suit :

$$n_{AD} = \frac{1}{1+A.e^{-B.n_{MM}}} \qquad [5\text{-}7]$$

Avec : $A = \frac{1}{\Delta} - 1;\ B = \frac{lnA}{c};\ \Delta = 0.01$

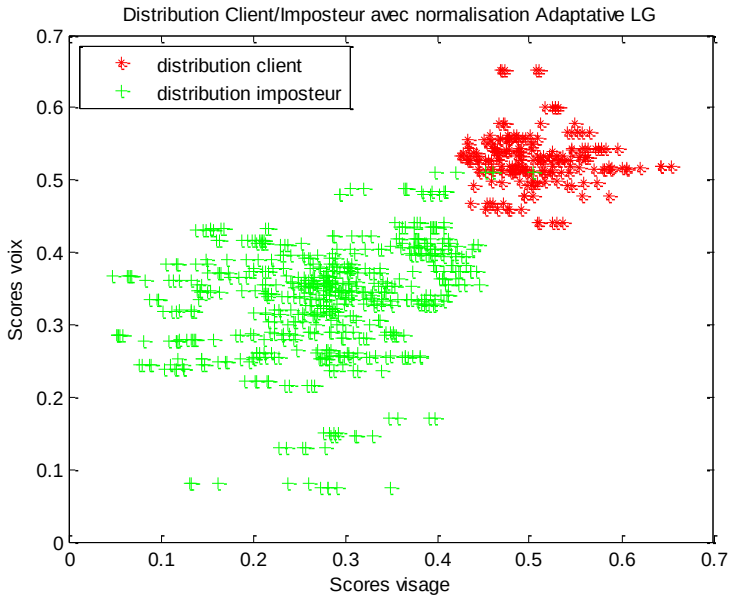


Figure 5.7-Distribution client-imposteur après normalisation Adaptative LG

- **Quadratic-Line-Quadric (QLQ) :**

La fonction quadratique-linéaire-quadratique (QLQ) normalise des scores n_{MM} préalablement transformés dans l'intervalle [0, 1]. Cette normalisation prend comme paramètres le centre C et la largeur W de la zone de recouvrement des distributions des scores clients et imposteurs. Quant aux régions restantes, elles seront tramées (projetées) avec deux segments de fonctions quadratiques. Dans ce cas la formule de normalisation est donnée par :

$$n_{AD} = \begin{cases} \frac{1}{\left(c-\frac{w}{2}\right)} {n_{MM}}^2 & si : n_{MM} \leq \left(c-\frac{w}{2}\right) \\ n_{MM} & si: \left(c-\frac{w}{2}\right) \leq n_{MM} \leq \left(c+\frac{w}{2}\right) \\ \left(c+\frac{w}{2}\right) + \sqrt{\left(1-c-\frac{w}{2}\right)\left(n_{MM}-\left(c-\frac{w}{2}\right)\right)} & sinon \end{cases} \qquad [5\text{-}8]$$

Ce qui donnera la courbe suivante :

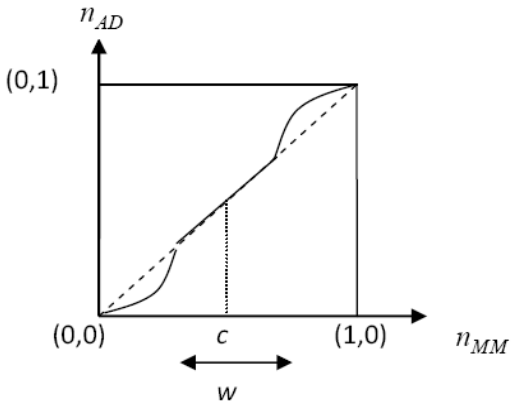


Figure 5.8-Fonction de Mapping de la méthode de normalisation QLQ

Les caractéristiques des ces différentes techniques de normalisation, en termes de robustesse et d'efficacité sont données dans le tableau suivant:

TECHNIQUE DE NORMALISATION	ROBUSTESSE	EFFICACITE
Min-Max	NON	Élevée
Z-Score	NON	Élevée
Tanh	OUI	Élevée
Adaptative QQ	OUI	Élevée
Adaptative LG	OUI	Élevée

Tableau 5.1- Résumé des caractéristiques des techniques de normalisation de scores

IV.2.2 Normalisation des scores par interprétation :

Ces méthodes de normalisation sont basées sur l'estimation des densités de probabilité des deux classes, Client et Imposteur. Dans une première partie nous expliquerons comment l'estimation de densités permet de normaliser les scores dans un espace commun par l'intermédiaire de critères tels que les probabilités a posteriori ou le rapport de vraisemblance.

Ensuite, dans une deuxième partie, nous ferons le lien entre ces méthodes de normalisation de scores basées sur les probabilités et les méthodes de fusion de scores qui seront présentées au prochain paragraphe [Ben 02].

- ✓ Les normalisations basées sur les estimations de densités :

Les normalisations de scores par remise à l'échelle ne traitent pas le problème des différences de nature entre les distributions des différents sous-systèmes. Par exemple, on constate sur la Figure 5.10, que si on combine les scores (normalisés par la méthode Min-Max), les résultats ne seront pas bons car les deux distributions ont des allures très différentes. Le Système 1 a une classe Client très étendue et une classe Imposteur très piquée, alors que le Système 2 a des classes de variances et de formes environ équivalentes. Le but des méthodes de normalisation de scores par interprétation basée sur l'estimation est donc de prendre en compte les différences de nature entre les densités de deux systèmes avant de les combiner.

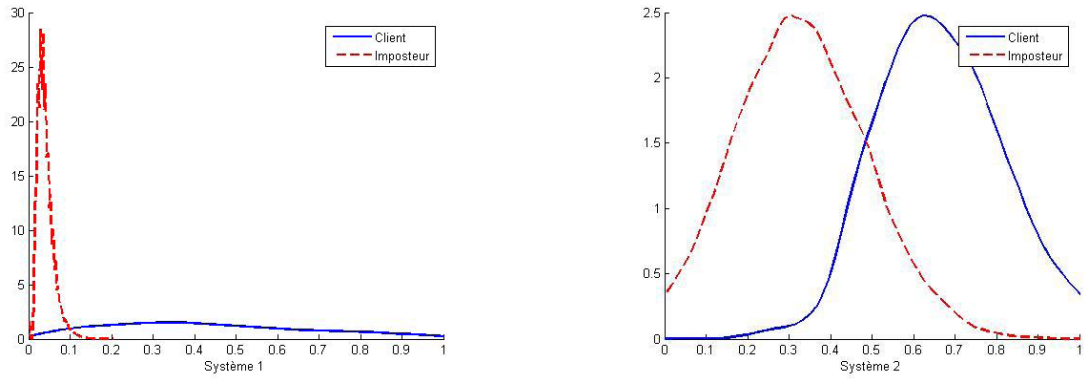


Figure 5.9 Densités de probabilités de deux systèmes différents après normalisation Min-Max

Les méthodes les plus utilisées pour interpréter les scores utilisent une estimation des distributions des deux classes. Elles transforment ensuite le score final en probabilité a posteriori d'appartenance à une classe ou en rapport de vraisemblance.

Transformer les scores issus de chaque système monomodal en probabilités a posteriori, avant de les fusionner, permet de se rapporter à la théorie de la décision

Bayésienne qui dit que pour un problème de classification à M classes $\{C_1, \dots, C_M\}$ d'un élément x, l'élément x doit être assigné à la classe qui maximise la probabilité a posteriori d'appartenance, c'est-à-dire :

$$x \in C_k \quad si \quad p(C_k|x) \geq p(C_i|x) \quad pour \quad i = 1, \dots, M$$

Pour pouvoir se ramener une décision Bayésienne il faut donc transformer les scores en probabilités a posteriori. Dans le cas d'une normalisation de chaque score issus des systèmes monomodaux avant de le combiner, il s'agit de transformer chaque score S_i issus des N systèmes monomodaux en probabilités a posteriori d'appartenance à la classe des clients $p(C|S_i)$, et d'appartenance à la classe des imposteurs $p(I|S_i)$.

Les probabilités a posteriori d'appartenance aux deux classes client et imposteur, sont estimées grâce à la règle de Bayes. Par exemple pour la probabilité a posteriori d'appartenance à la classe des Clients :

$$p(C|x) = \frac{p(C)\, p(x|C)}{p(x)} = \frac{p(C)\, p(x|C)}{p(C)\, p(x|C) + p(I)\, p(x|I)} \qquad [5\text{-}9]$$

Ou :

$p(C)$ et $p(I)$: sont les probabilités a priori des deux classes, respectivement Client et Imposteur.

$p(x|C)$ et $p(x|I)$: sont les densités de probabilités de la variable x conditionnellement aux deux classes, respectivement Client et Imposteur.

Pour suivre la théorie Bayésienne de décision dans notre problème à deux classes, Client et Imposteur, le score issu de chaque système devrait être transformé en quotient des deux probabilités a posteriori tel que :

$$S_i = \frac{p(C|S_i)}{p(I|S_i)} \qquad [5\text{-}10]$$

L'utilisateur est donc accepté comme étant un client si : $S_i \geq 1$.

Pour estimer les probabilités a posteriori il faut donc estimer les probabilités a priori des deux classes ainsi que les densités de probabilités conditionnelles aux deux classes.La théorie de la décision de Bayes revient donc à considérer comme score normalisé le rapport

des probabilités a posteriori S_i ou bien la probabilité a posteriori de l'une des classes car $p(C|S_i) + p(I|S_i) = 1$.

Le rapport des probabilités peut être relié aux densités de probabilités conditionnelles par les probabilités a priori. En effet :

$$\frac{p(C|S_i)}{p(I|S_i)} = \frac{p(C)\,p(S_i|C)}{p(I)\,p(S_i|I)} \qquad [5\text{-}11]$$

Le rapport des densités de probabilités conditionnelles, $\frac{p(S_i|C)}{p(S_i|I)}$ est, quand à lui, appelé rapport de vraisemblance (RV).

En pratique, comme les probabilités a priori $p(C)$ et $p(I)$ ne sont pas connues, on considère plutôt comme score normalisé le rapport de vraisemblance. Le seuil de décision pour le rapport de vraisemblance dans la théorie de la décision de Bayes serait alors $p(I)/p(C)$ car pour accepter un client il faut que :

$$\frac{p(C|S_i)}{p(I|S_i)} \geq 1 \;\leftrightarrow\; \frac{p(S_i|C)}{p(S_i|I)} \geq \frac{p(I)}{p(C)} \qquad [5\text{-}12]$$

Dans l'hypothèse où les probabilités a priori $p(C)$ et $p(I)$ sont égales, on peut considérer comme score normalisé, soit le rapport de vraisemblance soit une estimation de la probabilité a posteriori Client qui est :

$$p(C|S_i) \approx \frac{p(S_i|C)}{p(S_i|C)+p(S_i|I)} = \frac{1}{1+\frac{p(S_i|I)}{p(S_i|C)}} = \frac{1}{1+\frac{1}{RV}} \qquad [5\text{-}13]$$

- ✓ Application des méthodes de fusion aux probabilités :

La transformation de chaque score issu des systèmes monomodaux en probabilités a posteriori (ou en rapport de vraisemblance, ce qui est équivalent) a pour but de donner un sens à leur fusion. La méthode de fusion qui vient naturellement lorsque l'on manipule des probabilités est le produit.

En effet, le but de la fusion est de prendre une décision non plus sur un score mais sur un vecteur de scores à N dimensions si il y a N systèmes monomodaux. Prendre la décision selon la théorie Bayésienne revient à considérer les probabilités a posteriori des deux classes

sachant le vecteur à N dimensions, c'est à dire : $p(C|S_1,\dots,S_N)$. Dans le cas des méthodes de fusion, chaque score est traité séparément avant d'être combiné. Le fait de les traiter séparément consiste à faire une hypothèse d'indépendance des différents scores S_i. Cette hypothèse d'indépendance se traduit en terme de probabilité par :

$$p(C|S_1,\dots,S_N) = \prod_{i=1}^{N} p(C|S_i) \qquad [5\text{-}14]$$

On voit donc apparaître la méthode de fusion par le produit. En effet si l'on considère les probabilités a posteriori ou leur quotient comme scores normalisés, le produit des scores issus de cette normalisation sera une estimation des probabilités a posteriori du vecteur de N scores sous l'hypothèse d'indépendance. Mais le produit pose un problème pratique car il est très sensible aux erreurs d'estimations des probabilités et en particulier une valeur égale à 0 produit un effet non compensable par les autres systèmes. La fusion par la somme des probabilités (ou leur moyenne) est beaucoup moins sensible aux erreurs d'estimation des probabilités et est souvent préférée pour la fusion des probabilités [Kit 98].

IV.3 Les différentes méthodes de fusion des scores :

Chacun des deux classificateurs donnera sa décision sous forme de score. Ces deux scores pourront êtres fusionnés par plusieurs méthodes.

Notons par n_i^m le score fourni par le $m^{ème}$ classificateur à l'iéme test et par f_i le score résultant de la fusion [Sne 05].

IV.3.1 Les méthodes de fusion simples :

- ✓ Min-Score (MIN):

$$f_i = \min(n_i^1, n_i^2, \dots, n_i^M) \quad \forall i \qquad [5\text{-}16]$$

- ✓ Max-Score (MAX) :

$$f_i = \max(n_i^1, n_i^2, \dots, n_i^M) \quad \forall i \qquad [5\text{-}17]$$

✓ Simple-Sum (SS) :

$$f_i = \sum_{m=1}^{M} n_i^m \quad \forall i \qquad [5\text{-}18]$$

IV.3.2 Les méthodes dépendantes des classificateurs :

✓ Matcher Weighting (MW):

Les pondérations sont assignées aux modalités selon leur EER (Equal Error Rate). Soit le EER de la modalité m : $e^m, m = 1,2,\dots,M,$ et w^m la pondération qui lui est associée telle que :

$$w^m = \frac{\frac{1}{e^m}}{\sum_{m=1}^{M} \frac{1}{e^m}} \qquad avec : 0 \le w^m \le 1 \quad \forall m \quad et \quad \sum_{m=1}^{M} w^m = 1 \qquad [5\text{-}19]$$

Dans ce cas le score fusionné de l'utilisateur i est donné par :

$$f_i = \sum_{m=1}^{M} w^m \, n_i^m \quad \forall i \qquad [5\text{-}20]$$

V. CHOIX DE LA BIMODALITE VISAGE-VOIX

Dans le cadre de notre projet le choix s'est porté sur la fusion des deux modalités visage et voix et pour cause, ces dernières présentent de nombreux avantages.

En premier lieu ce sont des modalités acceptées par le public de par la simplicité d'acquisition de leurs paramètres caractéristiques, cela ne provoque donc aucune gène à l'utilisateur lors du test. Notons également qu'elles font partie des modalités moins contraignantes, des plus naturelles et sont de plus disponibles sur de nombreux systèmes (téléphones portables, PDA, ordinateurs...). De plus, ces dernières ne sont pas équivalentes mais complémentaires ce qui permet à l'une de compenser les déficits de l'autre. L'un des autres avantages majeur de la fusion de ces deux modalités est le fait qu'elles permettent toutes les deux une reconnaissance à distance, ce qui pour la criminel, représente un atout non négligeable.

VI.DECISION

Apres avoir procéder à l'extraction de paramètres et la modélisation vient l'étape d'apprentissage et ceci pour chacune des modalités. Chaque individu se verra attribué un modèle à base de GMM ainsi qu'un modèle UBM, qui lui, représentera les imposteurs. Les scores résultants pour chacune des modalités seront normalisés et pourront ainsi être fusionnés dans le but de fournir un modèle au système multimodal. Nous aurons donc le EER de ce dernier en plus du seuil de décision qui tranchera entre le choix client ou imposteur.

Le seuil sera fixé suivant l'application en question, en générale on choisi le point de fonctionnement neutre EER qui donnera un même degré d'erreur concernant la fausse acceptation ainsi que le faut rejet.

VI.1 Décision dans le cas de l'authentification :

Le score n^m obtenu pour la $m^{ème}$ modalité sera donc normalisé puis fusionné avec les scores des autres modalités afin d'obtenir un score final qui sera comparé à un seuil décisif tel que :

- ✓ Si $f > seuil$: l'individu présenté est l'individu proclamé.
- ✓ Sinon : l'individu présenté est considéré comme étant un imposteur.

VI.2 Décision dans le cas de l'identification :

Dans ce cas, les M classificateur donneront M ensembles de N scores ou N n'est autre que le nombre de clients enregistrés dans la base de données. Ces derniers seront donc normalisés puis fusionnés avec les scores des autres modalités. Nous obtiendrons ainsi un total de N scores et l'individu testé sera identifié au client ayant le plus grand score.

Nous pouvons maintenant présenter le processus de reconnaissance d'individu utilisant la biométrie bimodale, dans notre cas le visage et la voix, basée sur la fusion des scores, à travers le schéma suivant :

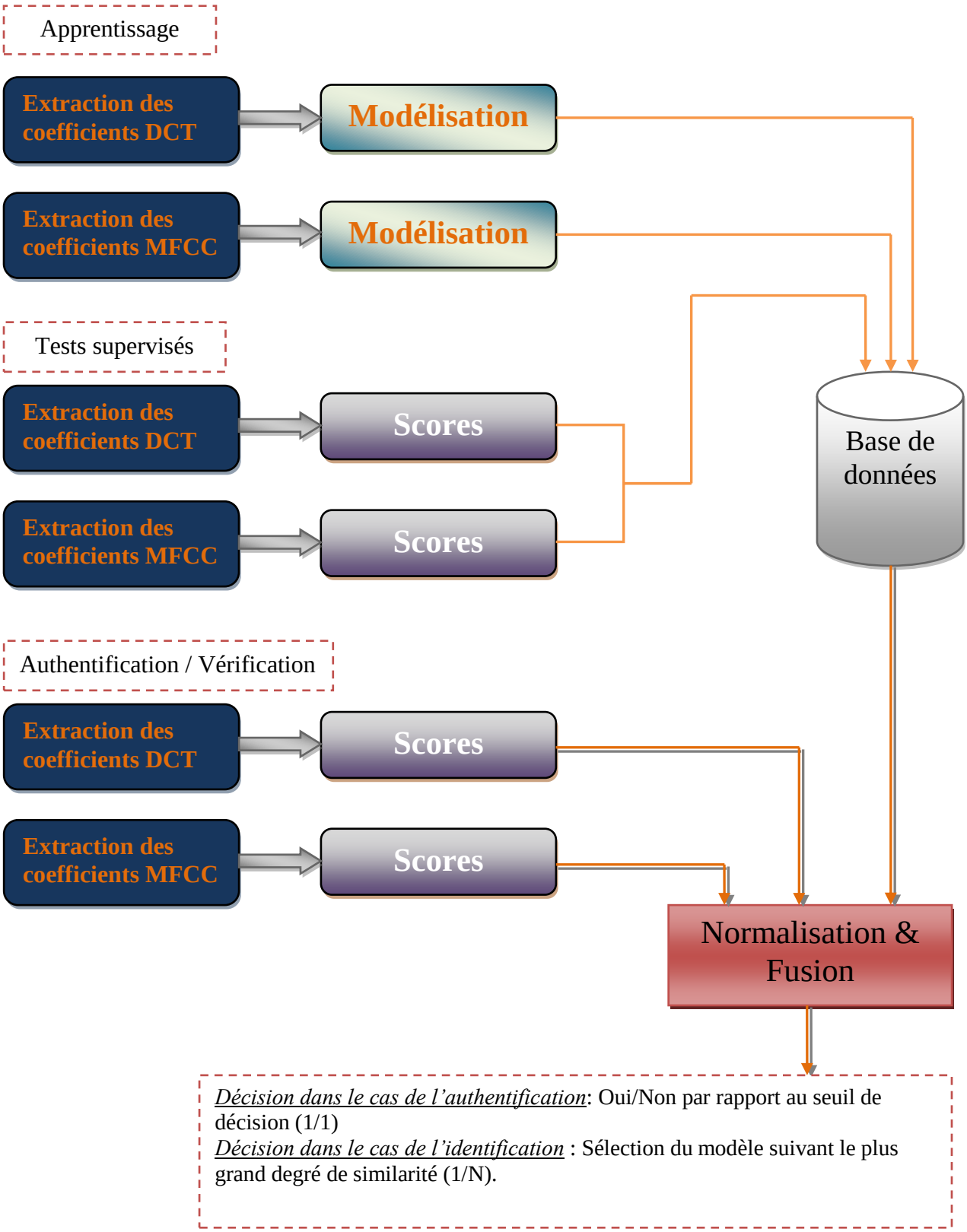


Figure 5.10-Schéma global du système biométrique bimodal (visage et voix) basé sur la fusion de scores

VII.CONCLUSION

La biométrie multimodale, qui consiste à combiner plusieurs systèmes biométriques, est de plus en plus étudiée. En effet elle permet de réduire certaines limitations des systèmes biométriques, comme l'impossibilité d'acquérir les données de certaines personnes ou la fraude intentionnelle, tout en améliorant les performances de reconnaissance. Ces avantages apportés par la multimodalité aux systèmes biométriques "monomodaux" sont obtenus en fusionnant plusieurs systèmes biométriques.

Les techniques de fusion peuvent soit être simples et appliquer une méthode simple sur les deux scores. Soit être basées sur des classificateurs et considérer à nouveau les modèles clients-imposteurs comme de nouveaux échantillons de deux « personnes » nouvelles.

6 Architecture du système et implémentation

I. INTRODUCTION

Nous avons présenté tout à travers les précédents chapitres les étapes à suivre permettant d'aboutir à l'extraction des paramètres DCT pour le visage et MFCC pour la voix ainsi que leur modélisation par le biais des mélanges de Gaussiennes et ceci dans le but de réussir au choix une identification ou une authentification selon le besoin de l'application.

Dans ce qui va suivre, nous présenterons le cheminement suivi lors de notre travail qui nous a permis d'aboutir à un système de reconnaissance efficace, robuste et fiable basé sur la fusion de scores ainsi qu'une présentation détaillée de l'interface graphique Matlab mise au point.

II. ARCHITECTURE DU SYSTEME

La structure générale du système de reconnaissance comporte trois phases que nous détaillerons au fur et à mesure et qui sont la phase d'apprentissage qui comporte l'extraction de paramètres et la génération de modèles, la phase de test qui traite les opérations d'identification et d'authentification et enfin la phase de décision qui tranchera sur l'issu du test et ceci en se référant au taux d'identification et du seuil de décision.

II.1 Phase d'apprentissage :

Lors de l'étape d'apprentissage le système a besoin de certaines données bien spécifiques pour ajouter une personne X et donc « apprendre » son modèle. Pour cela il nous faut introduire pour chaque personne voulant figurer dans la base de données 5 poses pour le visage et un enregistrement vocal de courte durée (entre 10 à 15 secondes).

II.1.1 Extraction de paramètres :

- Paramètres DCT :

✓ L'image est importée sous forme de niveaux de gris de taille112×92.

✓ Cette dernière se voit appliquer les prétraitements : normalisation, égalisation et filtre médian afin de pouvoir d'exploiter au mieux toutes ses caractériques.

- ✓ La matrice obtenue est découpée en bloc 8×8, notons que ce découpage est axé sur chevauchement de 50%.
- ✓ Chaque bloc obtenu se verra appliquer la DCT afin d'en extraire les paramètres recherchés.
- ✓ Par la suite ce dernier sera parcouru en ZIGZAG afin d'en extraire les paramètres DCT classés suivant un ordre bien précis allant des basses fréquences jusqu'aux hautes fréquences.
- ✓ L'ensemble des vecteurs colonnes obtenus formeront alors la matrice de coefficients DCT, dont la taille « p » des vecteurs pourra et sera variée selon les besoins de l'application.

- Paramètres MFCC :

- ✓ Le signal audio importé est échantillonné en respectant la condition de Shannon qui stipule que $f_{sh} \geq 2f_{max}$ ou f_{sh} est la fréquence d'échantillonnage tandis que f_{max} est la fréquence maximale du signal en question.
- ✓ Le signal échantillonné est soumis à un découpage en trames entrelacées, avec un chevauchement de 75% , afin d'obtenir un signal stationnaire par parties, ce qui impose à la trame d'avoir une taille maximale de $30ms$ ce qui nous donnera au final des trames dotées de 256 échantillons.
- ✓ Une fenêtre est ensuite appliquée à chaque trame afin d'éviter les déformations aux niveaux des hautes fréquences de par et d'autres de leurs extrémisées. Le choix de cette dernière a été portée sur la fenêtre Hamming du fait qu'elle possède un lobe principal large et des lobes secondaires très vite négligeables ce qui représenterait un avantage si elle venait à être comparée aux fenêtres Rectangulaire, Hann ou encore Blackman.
- ✓ La FFT est par la suite appliquée à chaque trame du signal ce qui nous permet d'obtenir le spectre d'énergie de ce dernier.
- ✓ Il nous faut ensuite définir une hauteur subjective de fréquence par le biais du filtrage MEL qui accentuera d'avantage les basses fréquences, afin d'interpréter au mieux le signal audio, car rappelons le la sélectivité de l'oreille humaine diminue avec l'accroissement de la fréquence.

- ✓ Le banc de M filtres MEL appliqué à chaque trame donne au final, pour chacune d'entre elles, M éléments qui se verront l'opérateur « log » ce qui fournira une représentation cepstrale de M éléments par trame.
- ✓ L'ensemble des vecteurs colonnes obtenus formeront alors la matrice de coefficients MFCC dont la taille « M » sera déterminée selon le nombre de filtres Mel utilisés.

II.1.2 Modélisation des paramètres :

Cette partie, il faut le préciser, représente le cœur même de notre système car elle a pour but de générer les modèles GMM, formés de « k » gaussiennes, des personnes à partir des matrices de coefficients DCT et MFCC.

Rappelons que pour modéliser une personne par un ensemble de gaussiennes il nous faudra déterminer pour chacune d'entre elles sa moyenne, sa matrice de covariance Σ ainsi que son coefficient de pondération c et ce comme suit :

- ✓ La moyenne μ représentant les coefficients DCT / MFCC est initialisée en utilisant l'algorithme « Kmeanlbg », cette dernière sera de dimension (p, k) pour le visage ou encore (M, k) pour la voix. Notons que pour le visage, une nouvelle matrice DCT sera formée à partir de la concaténation des matrices DCT des 5 images représentant la même personne.
- ✓ La matrice de covariance , elle, sera initialisée en tant que matrice identité de dimension (p, k) pour le visage ou encore (M, k) pour la voix.
- ✓ Le coefficient de pondération c quand à lui sera initialisé par un vecteur unité de taille.
- ✓ Le nombre d'itération par contre est pris à 10 par défaut et ceci en se référant à des travaux ultérieurs ayant prouvé que ce nombre est largement suffisant pour que l'algorithme converge.
- ✓ Apres la phase d'initialisation nous passerons à l'algorithme d'estimation-maximisation tel que ce dernier, pendant les 10 itérations, effectuera les taches suivantes :
  - Calculer la proportion de chaque vecteur DCT / MFCC par rapport à chaque gaussienne autrement dit le calcul de la probabilité que chaque vecteur soit généré l'une des k lois gaussiennes.
  - Réestimation des valeurs de , Σ et c.

Il est à noter qu'après la génération de modèles de chaque personne faisant partie de la base de données nous représenterons également les imposteurs en générant un modèle unique commun à tous les clients et qui n'est autre que le modèle UBM. Ce modèle sera utilisé par la suite dans la procédure d'authentification.

II.2 Phase de tests :

II.2.1 Identification :

Dans ce cas de figure l'identification se fait à travers les étapes suivantes :

- ✓ Acquisition de l'image ou de la voix de la personne testée par le système.
- ✓ Extraction des paramètres DCT ou MFCC de la même façon que pour l'apprentissage.
- ✓ Calcul de la probabilité de vraisemblance de l'image ou de la voix de la personne testée par rapport à l'ensemble des modèles enregistrés dans la base de données.
- ✓ Sélection du modèle ayant le plus grand degré de similarité et donc la plus grande vraisemblance.

II.2.2 Authentification :

Dans le cas unimodal, l'authentification se fait à travers les étapes suivantes :

- ✓ Acquisition de l'image ou de la voix de la personne testée par le système.
- ✓ Extraction des paramètres DCT ou MFCC de la même façon que pour l'apprentissage.
- ✓ Calcul de la probabilité de vraisemblance de l'image ou de la voix de la personne testée par rapport au modèle de la personne proclamée afin d'obtenir la probabilité client.
- ✓ Calcul de la probabilité de vrais semblance de l'image ou de la voix de la personne testée par rapport au modèle UBM afin d'obtenir la probabilité imposteur.

- ✓ La quantité « probabilité client- probabilité imposteur » sera comparée à un seuil de décision tel que :

$$\begin{cases} probabilité\ client - probabilité\ imposteur > seuil \rightarrow individu\ proclamé : client \\ probabilité\ client - probabilité\ imposteur < seuil \rightarrow individu\ proclamé : imposteur \end{cases}$$

Dans le cas bimodal par contre nous procédons aux même étapes que l'unimodal sauf qu'on aura les deux modalités visage et voix à traiter. Nous procèderons donc, en plus des étapes citées plus haut, à ce qui suit :

- ✓ Appliquer aux scores (probabilités de vraisemblance) visage et voix obtenus les normalisations suivantes au choix : Min-Max, Z-score, Tanh, Adaptative QQ et Adaptative LG.
- ✓ Fusionner les scores précédemment normalisés et ce de différentes manières qui sont : Min, Max, Somme simple et Somme pondérée.
- ✓ Le score final après fusion se verra soumis à la même comparaison par rapport au seuil de décision, on aura donc : $\begin{cases} score_{fusion} > seuil \rightarrow individu\,proclamé : client \\ score_{fusion} < seuil \rightarrow individu\,proclamé : imposteur \end{cases}$

II.3 Phase de décision :

Cette dernière phase nous permet de mesurer les performances du système et d'évaluer l'erreur engendrée par ce dernier pour ainsi effectuer une comparaison entre différents systèmes ou différentes configurations du même système.

Cette étude se fait dans les deux types de tests cités précédemment tel que pour :

II.3.1 Identification :

Dans ce mode de décision, le paramètre permettant de mesurer les performances du système est le taux d'identification (TID) qui n'est autre que la proportion du nombre de personnes correctement identifiées, ce dernier étant calculé comme suit :

$$TID = \frac{\text{nombre de personnes correctement identifiées}}{nombre\ total\ de\ personnes\ dans\ la\ base\ de\ données} \times 100 \quad [6\text{-}1]$$

Un système d'identification robuste doit donc avoir un taux d'identification proche de 100.

II.3.2 Authentification :

Dans ce mode de décision les paramètres qui contribuent à la mesure des performances du système sont définis comme suit :

- Le taux des fausses acceptations (TFA) :

La fausse acceptation parvient lorsque le système d'authentification accepte une personne qui a proclamé une identité autre que la sienne, elle est calculée comme suit :

$$TFA = \frac{\text{nombre de fausses acceptations}}{nombre\ total\ des\ accès\ imposteurs} \times 100 \qquad [6\text{-}2]$$

- Le taux des faux rejets (TFR) :

Le faux rejet parvient lorsque le système d'authentification rejette une personne qui a proclamé sa véritable identité, elle est calculée comme suit :

$$TFR = \frac{\text{nombre de fausses rejets}}{nombre\ total\ des\ accès\ clients} \times 100 \qquad [6\text{-}3]$$

- Le taux d'égales erreurs (EER) :

✓ C'est le point pour lequel les taux des fausses acceptations et des faux rejets sont égaux.

✓ C'est le choix raisonnable pour les systèmes d'authentification car il ne favorise ni la fausse acceptation ni le faux rejet.

✓ Si le seuil de décision est placé au dessous du EER le taux des faux rejets diminuera et le niveau de sécurité, donc, augmentera.

✓ Si le seuil de décision est placé en dessus du EER le taux de fausses acceptations augmentera et le niveau de sécurité, donc, diminuera.

III. PRESENTATION DE L'INTERFACE GRAPHIQUE SOUS MATLAB

Nous allons maintenant présenter l'interface élaborée sur MATLAB qui a pour rôle d'exécuter et de mettre en évidence toutes les étapes menant à la reconnaissance que ce soit pour l'identification ou pour l'authentification.

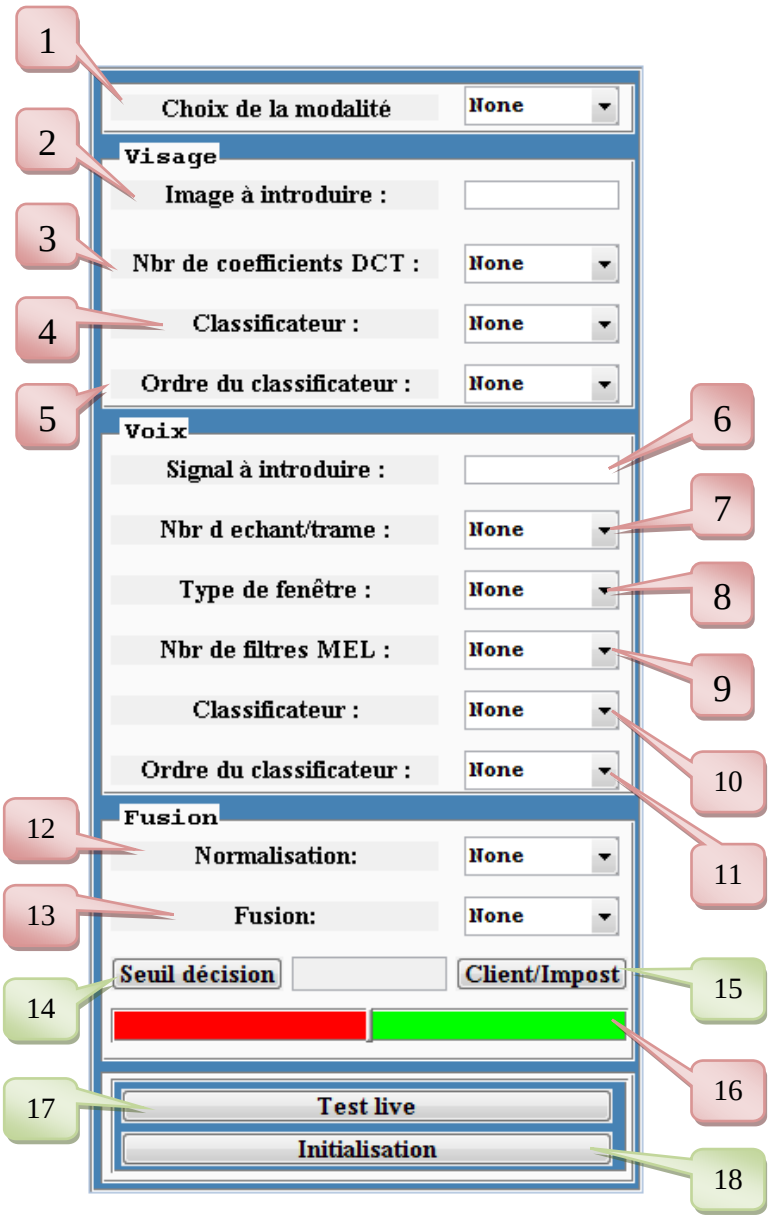


Figure 6.1-Module servant à configurer les paramètres d'entrée.

Figure 6.2-Module servant à configurer les fonctions à exécuter.

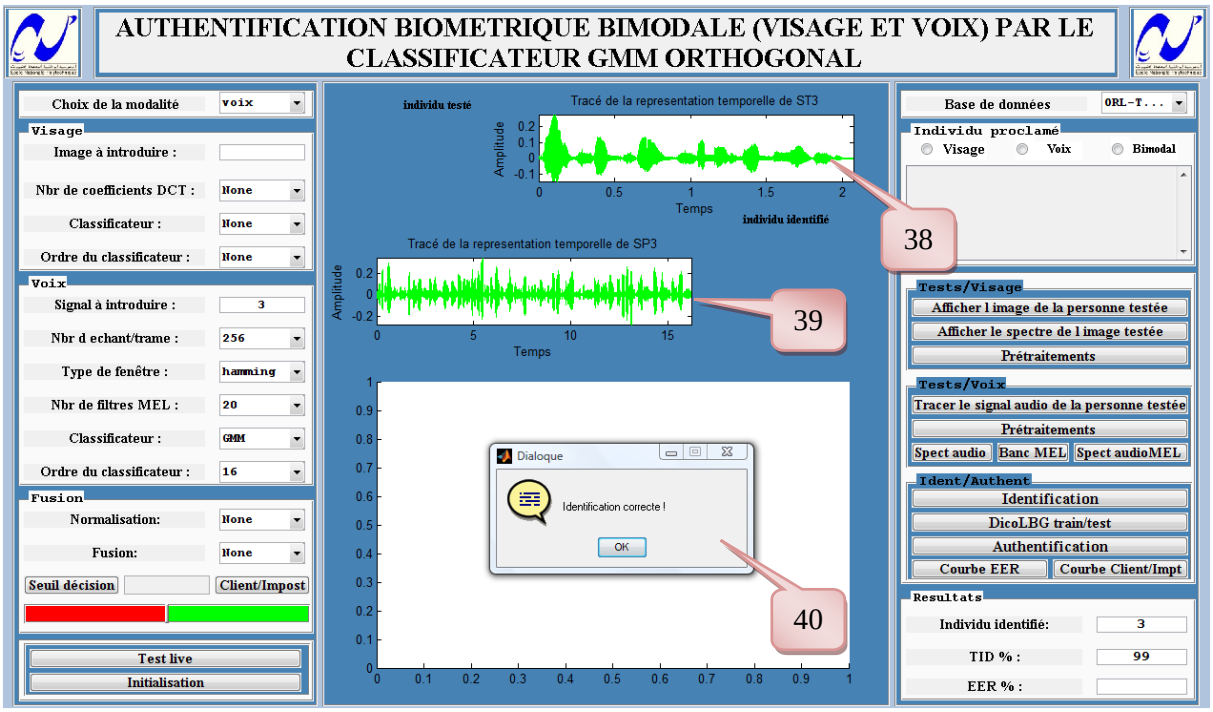


Figure 6.3-Traitement signal audio / Identification

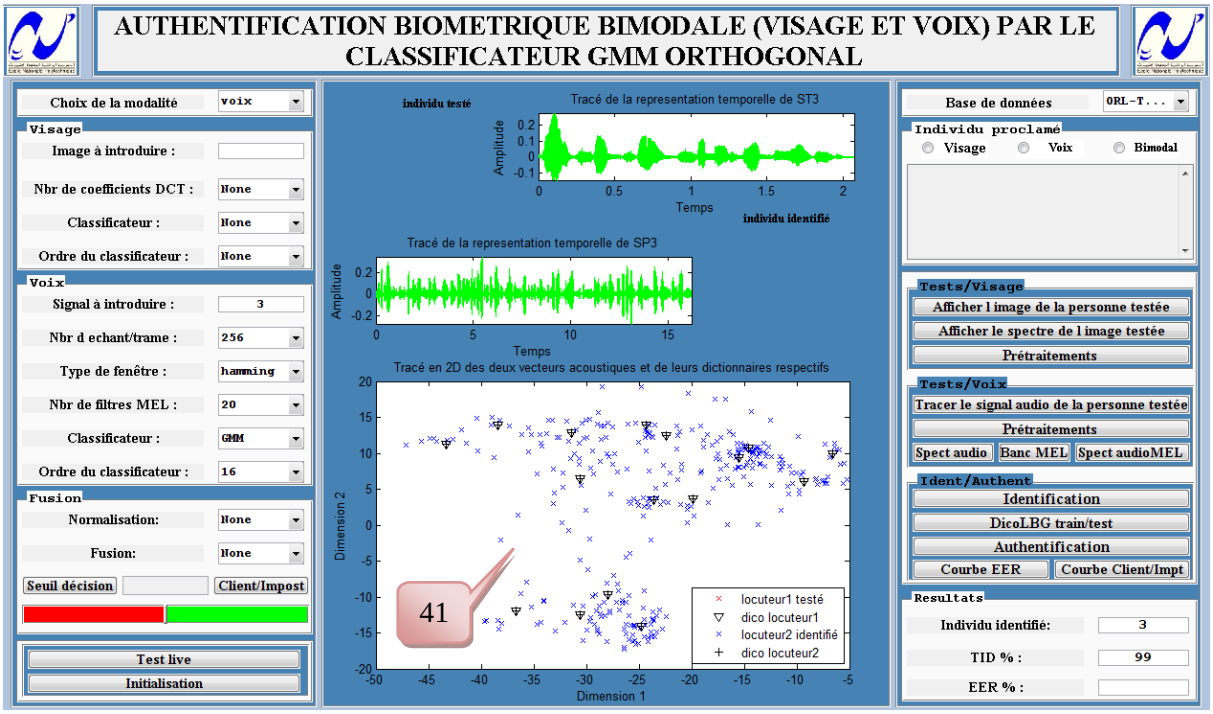


Figure 6.4-Traitement signal audio / modélisation des signaux des personnes (testée-identifiée) par leur dictionnaire VQLBG

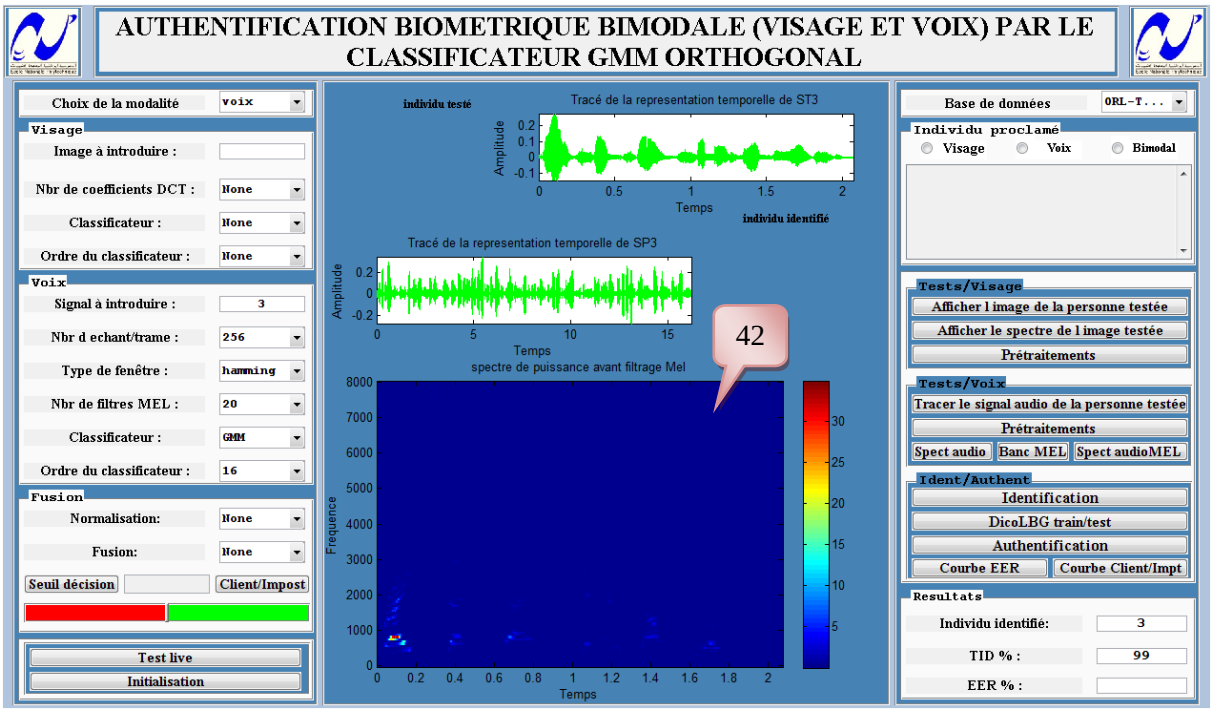


Figure 6.5-Traitement signal audio / tracé du spectre du signal testé avant filtrage MEL

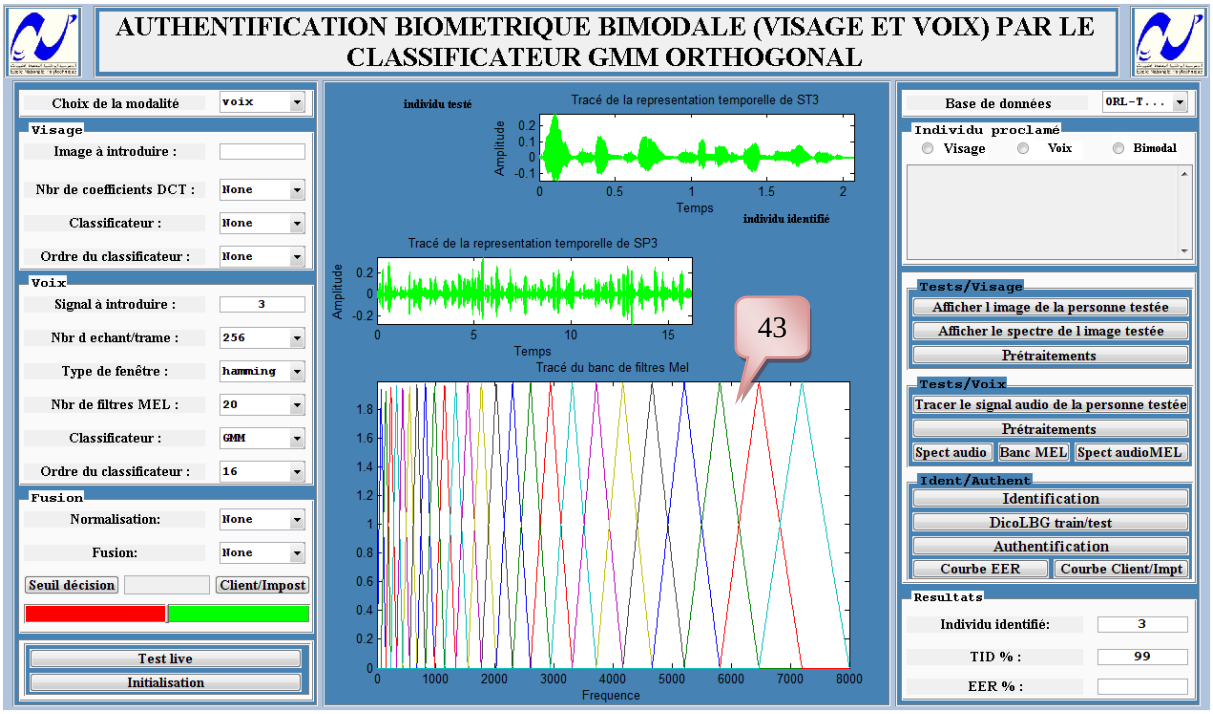


Figure 6.6-Traitement signal audio / tracé du banc de 20 filtres MEL

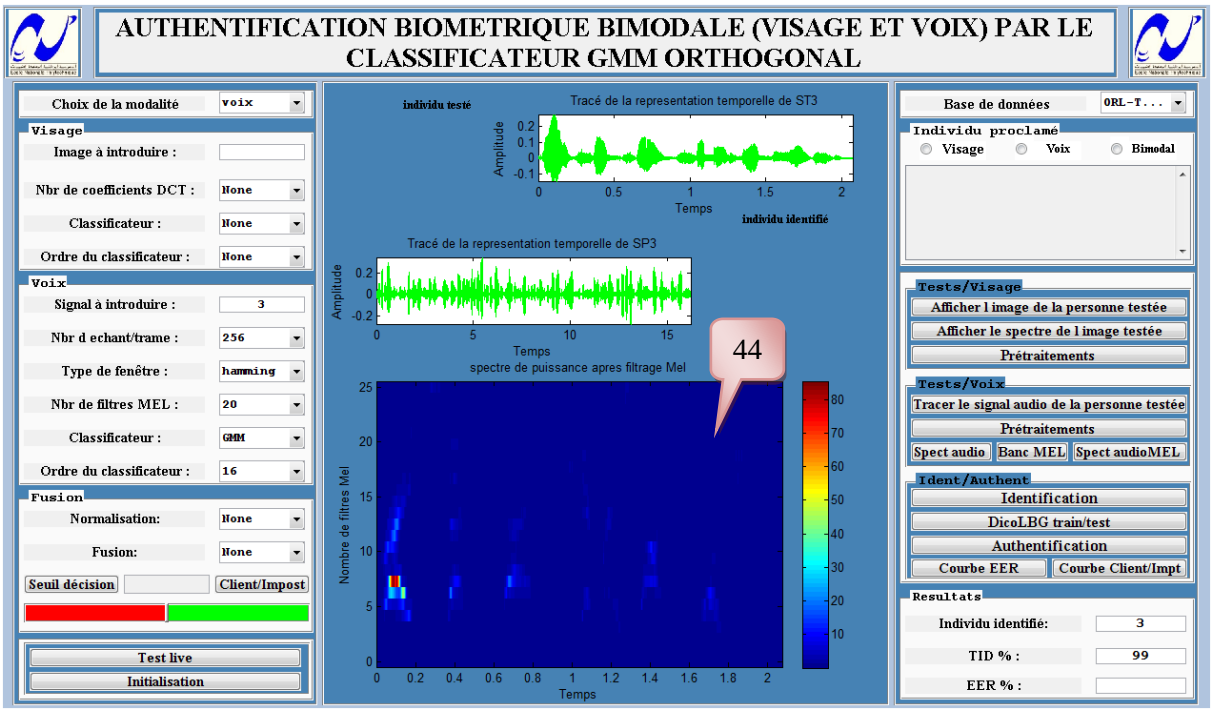


Figure 6.7-Traitement signal audio / tracé du spectre du signal testé après filtrage MEL

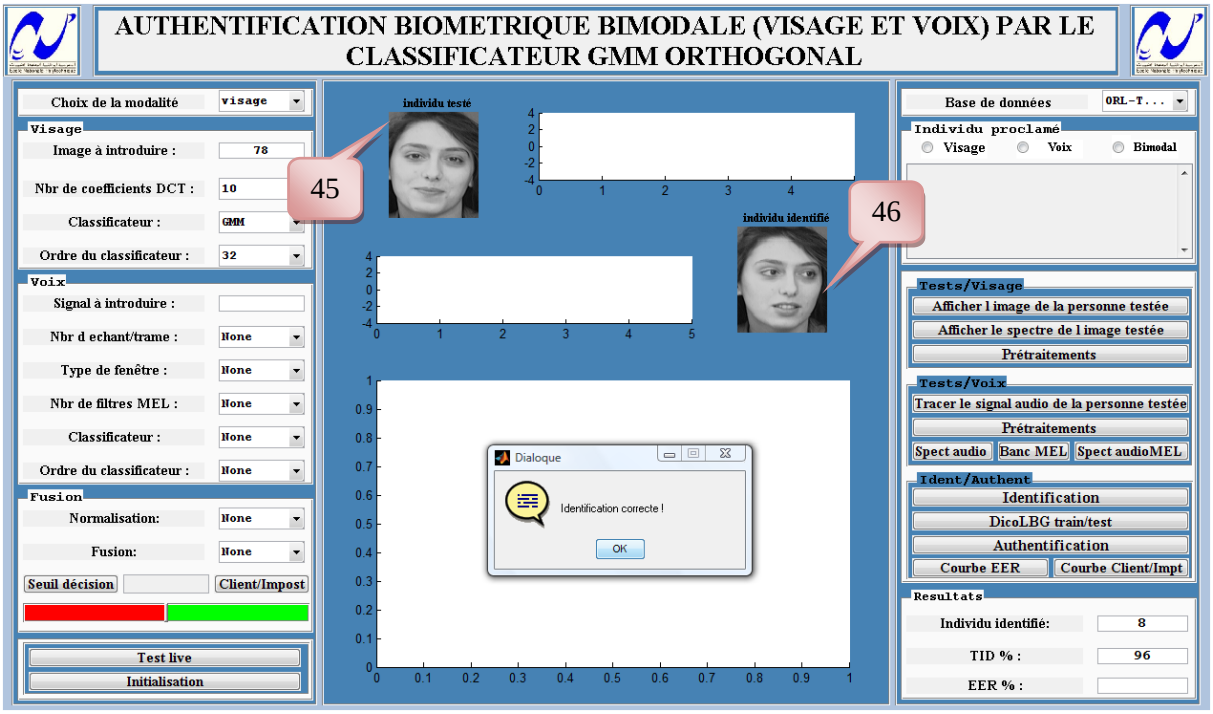


Figure 6.8-Traitement image / Identification

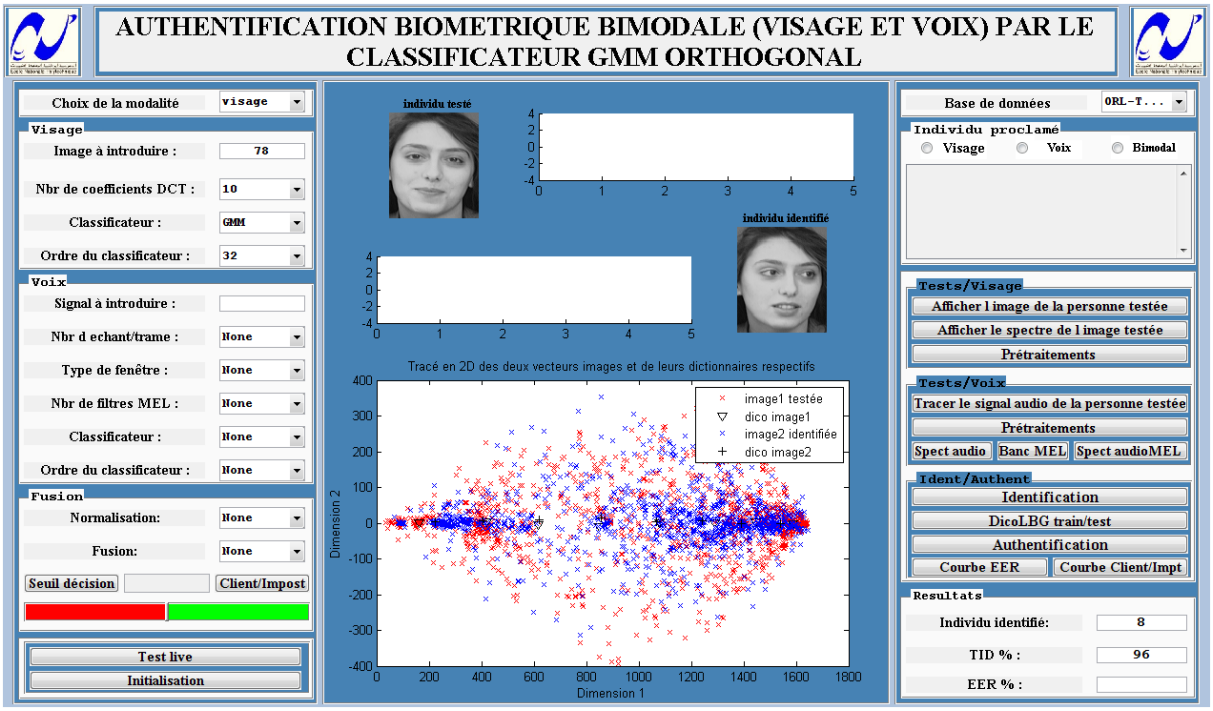


Figure 6.9-Traitement image / modélisation des images des personnes (testée-identifiée) par leur dictionnaire VQLBG

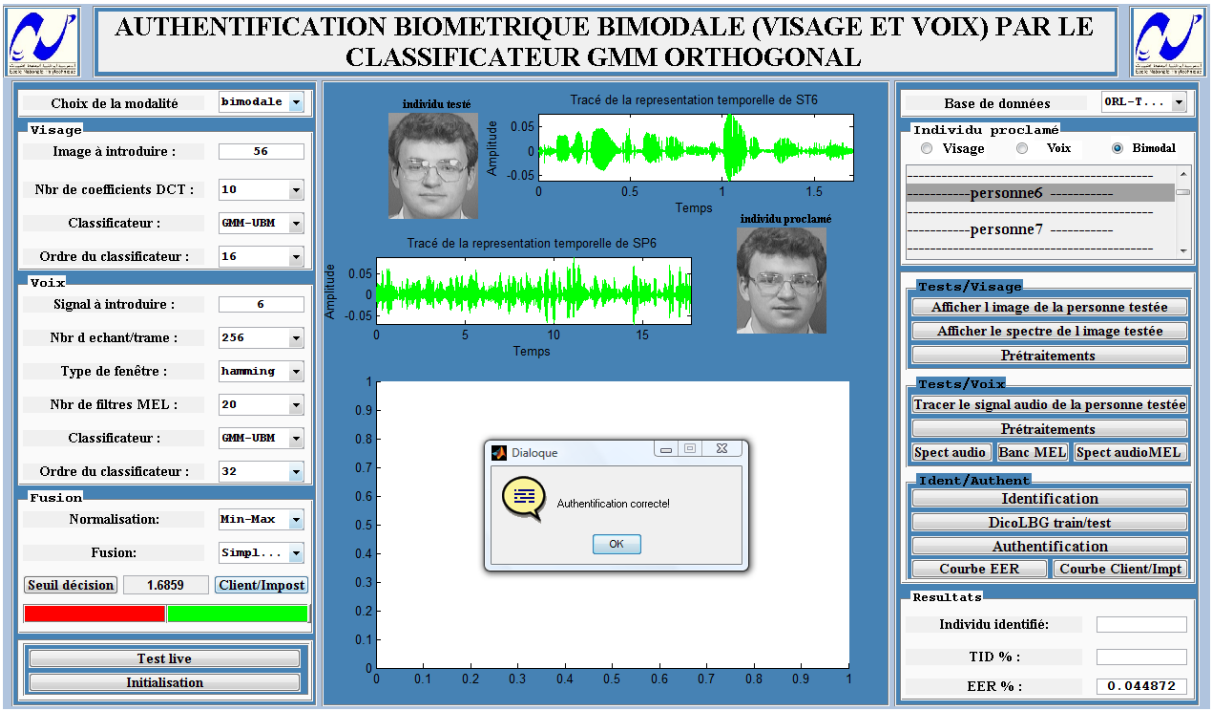


Figure 6.10-Traitement bimodal / Authentification

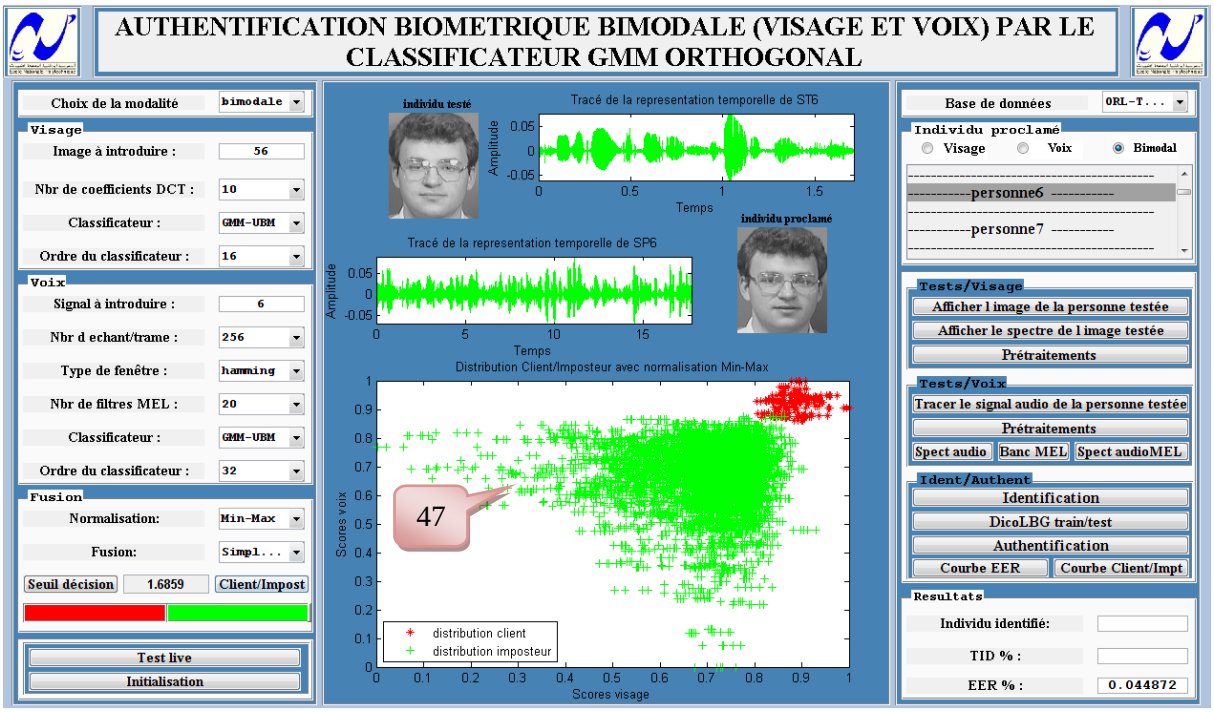


Figure 6.11-Traitement bimodal / tracé de la courbe client-imposteur suivant les paramètres de la fusion choisis

Clé des graphes précédents :

- 1 : choix de la modalité à traiter.
- 2 : indice de l'image à tester.
- 3 : choix du nombre de coefficients DCT délimitant la taille de l'image transformée.
- 4, 10: choix du classificateur.
- 5, 11: choix de l'ordre du classificateur.
- 6 : indice du signal audio à traiter.
- 7 : choix délimitant le nombre d'échantillons par trame.
- 8 : choix du type de fenêtre à utiliser.
- 9 : choix du nombre de filtres MEL à utiliser.
- 12 : choix du type de normalisation à utiliser.
- 13 : choix du type de fusion à adopter.
- 14 : fonction fixant le seuil de décision suivant les configurations choisies précédemment.

- 15 : fonction commandant un curseur dont la position a pour but de nous informer du degré d'appartenance de la personne testée à la classe « client » ou « imposteur ».
- 16 : zone de variation du curseur sachant que la zone *verte* est pour la classe « client » tant dis que la zone *rouge* est pour la classe « imposteur ».
- 17 : fonction enclenchant la seconde GUI pour lors du test réel.
- 18 : fonction initialisant la GUI.
- 19 : choix de la base à traiter.
- 20 : choix des modalités afin sélectionner la liste des personnes enregistrées dans la base de données.
- 21 : liste des identités fichées dans la base de données.
- 22 : fonction affichant l'image de la personne testée.
- 23 : fonction affichant le spectre de l'mage testée après avoir appliqué la DCT
- 24, 26 : fonction appliquant les prétraitements sur l'image ainsi que sur la voix.
- 25 : fonction trançant le signal audio de la personne testée.
- 27 : fonction traçant le spectre du signal audio avant filtrage MEL.
- 28 : fonction traçant le banc de filtres MEL.
- 29 : fonction traçant le spectre du signal audio après filtrage MEL.
- 30 : fonction enclenchant l'identification.
- 31 : fonction permettant de tracer les dictionnaires VQLBG en 2D lors de la modélisation.
- 32 : fonction enclenchant l'authentification.
- 33 : fonction permettant de tracer la courbe EER.
- 34 : fonction permettant de tracer la courbe client/imposteur.
- 35 : affichage de l'indice de la personne identifiée.
- 36 : affichage du TID.
- 37 : affichage de l'EER.
- 38 : tracé du signal de la personne testée.
- 39 : tracé du signal de la personne identifiée ou proclamée.
- 40 : affichage d'un message informant du résultat de l'opération encourue.
- 41 : tracé du dictionnaire VQLBG en 2D.
- 42, 44 : tracé du spectre audio avant et après filtrage MEL.

- 43 : tracé du banc de filtres MEL.
- 45 : affichage de l'image de la personne testée.
- 46 : affichage de la personne testée ou proclamée.
- 47 : tracé de la répartition client/imposteur.

7 Tests et évaluation des résultats

I. INTRODUCTION

A travers cet ultime chapitre nous illustrerons les résultats obtenus lors des tests sur l'identification et l'authentification qui représentent le cœur de notre projet. Différentes approches ont été utilisée telle que la quantification vectorielle VQ ou encore les mixtures de mélange de gaussiennes GMM et enfin le mélange de gaussiennes orthogonales.

Nous analyserons également les résultats d'un point de vue variation de quantité d'information requise et ceci pour chacune des modalités visage et voix. En d'autres termes la quantité d'élément que doivent contenir les vecteur DCT et MFCC qui nous permet d'avoir le meilleur résultat.

II. BASES DE DONNEES UTILISEES

Nous avons utilisé tout au long de notre étude deux bases différentes. La base ORL pour le visage couplées à la base TIMIT pour la voix.

Nous avons également utilisé une base composée d'étudiants de l'ENP.

II.1 La base TIMIT

La base TIMIT a été produite par l'effort commun des institutions Américaines suivante : Massachusetts Institute of Technology (MIT), Standford Research Institute (SRI), Texas Instruments (TI) et National Institute of Standards Adb and Technology (NIST).

La base TIMIT est une base de parole non bruitées enregistrées en utilisant des microphones de bonne qualité et avec une fréquence d'échantillonnage de 16KHZ, nous avons utilisé les paroles des 100 locuteurs de cette base ; chaque locuteur prononce 10 phrases, les 8 premières phases sont utilisées pour la phase d'apprentissage et les 2 dernières phrases sont utilisées lors de la phase de test.

II.2 La base ORL

Conçu par AT&T laboratoires de l'université de Cambridge en Angleterre, la base de données ORL (Olivetti Research Laboratory) est une base de données de référence pour les systèmes de reconnaissances automatique des visages. En effet, tous les systèmes de reconnaissances de visages trouvés dans la littérature ont été testés en utilisant la base ORL, cette popularité est dû aux nombre de contraintes existantes dans cette base car la plus part des changements possibles et prévisibles du visage ont été pris en compte, comme par exemple : le changement de coiffure, la barbe, les lunettes, les changements dans les expressions faciales, etc. Ainsi que les conditions d'acquisition telles que : le changement d'échelle dû à la distance entre le dispositif d'acquisition et l'individu.

La base de données ORL est constituée de 40 individus, chaque individu possède 10 poses, 5 pour l'apprentissage et les 5 autres pour le test (voir figure 7.2). Les poses ont étaient prises sur des intervalles de temps différents pouvant aller jusqu'à trois mois. L'extraction des visages à partir des images a été faite manuellement. Nous présenterons dans ce qui suit les figures montrant les spécificités de la base de données de référence ORL (voir figures 7.3, 7.4, 7.5, 7.6)

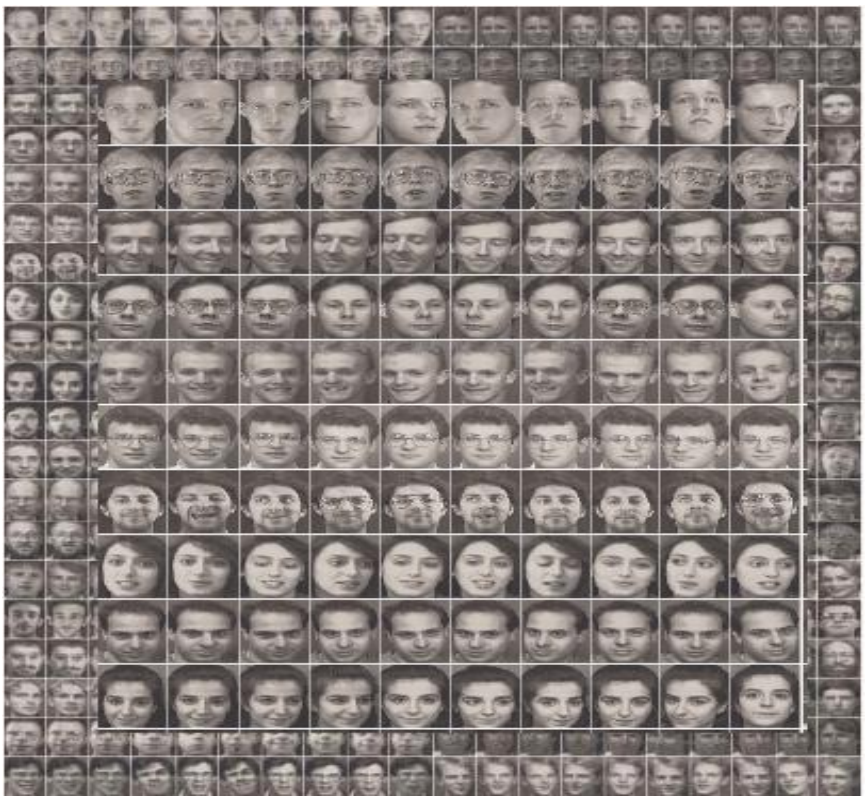


Figure 7.1-Base de données ORL

Figure 7.2-Exemple de changements d'orientations du visage

Figure 7.3-Exemple de changements d'éclairage

Figure 7.4-Exemple de changements d'échelle

Figure 7.5-Exemple de changements des expressions faciales

Figure 7.6-Exemple de port de lunettes

II.3 La base bimodale réelle

Cette base comporte 27 Etudiants de l'ENP, chacun d'entre eux ayant mis à notre disposition 13 photos ainsi qu'un enregistrement de 35 secondes environ. Les données récoltées que ce soit pour la voix ou pour le visage seront répertoriées dans trois groupes. Le premier pour l'apprentissage (5 photos), le second pour le test supervisé (4 photos) et le troisième pour l'évaluation (4 photos), que ce soit pour l'identification ou pour l'authentification.

Cette dernière met le système face aux conditions réelles(voir figure 7.7). En effet une telle base de données est importante pour tester le bon fonctionnement du système car elle a été conçue dans des milieux bruités avec un matériel un peu modeste.

Figure 7.7 : échantillons de la base réelle de l'ENP

III. PROTOCOLE D'EVALUATION

Notre système a été soumis à une batterie de tests en jouant sur la variations des différents paramètres de configuration des deux approches monomodales ainsi que sur l'ordre des classificateurs qui lui a permis de comparer les différentes techniques de modélisation utilisées (diagonale et orthogonale).

Ces tests ont été établis suivant les configurations paramétriques suivantes :

III.1 Paramètres de modélisation fixes:

- ✓ L'algorithme EM : initialisation par l'algorithme Kmean-LBG avec un nombre d'itérations fixé à 10.
- ✓ Le nombre d'images pour l'apprentissage : fixé à 5 .
- ✓ Le nombre d'images pour le test supervisé : fixé à 5 pour ORL, 4 pour la base réelle.

- ✓ La taille de l'enregistrement audio pour l'apprentissage : 16 secondes.
- ✓ La taille de l'enregistrement audio pour le test supervisé : 6 secondes.
- ✓ Le modèle UBM : ce modèle qui servira à modéliser les imposteurs lors du test d'authentification sera généré par les données ayant servies lors de l'apprentissage que ce soit pour le visage ou pour la voix. Il peut donc être vu comme une concaténation des données d'apprentissage.

III.2 Paramètres de modélisation variables

- ✓ Nombre d'échantillons par trame : 256.
- ✓ Type de fenêtre utilisée : fenêtre de Hamming.
- ✓ Taille de la fenêtre de glissement lors du traitement d'image : 8x8.
- ✓ Indice de chevauchement pour le traitement de l'image : 50%.
- ✓ Indice de chevauchement pour le traitement de la parole : 67% soit 2 tiers de données communes entre deux trames successives.
- ✓ Nombre de coefficients DCT : 10, 16, 32.
- ✓ Nombre de coefficients MFCC : 10, 20, 25.
- ✓ Type de classificateurs (cas identification) : VQ, GMM, OGMM-GPCA, OGMM-KLT.
- ✓ Type de classificateurs (cas authentification) : GMM, OGMM-GPCA, OGMM-KLT.
- ✓ Nombre de gaussiennes utilisée lors la génération de modèles (cas diagonal) : 16, 32, 64.
- ✓ Nombre de gaussiennes utilisés lors de la génération de modèles (cas orthogonal) : 4, 8, 16, 32, 64.
- ✓ Type de normalisations utilisées : Min-Max, Z-score, Tauh, Adaptative QQ, Adaptative LG.
- ✓ Types de fusions utilisées : Min, Max, Somme simple, Somme pondérée, GMM, OGMM-KLT.

Suivant les différentes combinaisons résultantes de ces configurations, les résultats des tests sont détaillés comme suit:

IV. RESULTATS DES TESTS SUR LES GMM Diagonales

IV.1 Mode identification - ORL

IV.1.1 Quantification vectorielle (VQ) :

	10 DCT	*16 DCT*	*32 DCT*
16 VQ	70%	68%	67.5%
32 VQ	69.5%	70.5%	70%
64 VQ	73.5%	74%	71.5%

Tableau 7.1-Taux d'identification (%) avec modélisation par VQ/visage ORL

IV.1.2 Mélanges de Gaussiennes (GMM)

	10 DCT	*16 DCT*	*32 DCT*
16 GMM	94%	92.5%	87.5%
32 GMM	93.5%	94.5%	93%
64 GMM	97.5%	98.5%	94%

Tableau 7.2-Taux d'identification (%) avec modélisation par GMM/visage ORL

Nous observons dans un premier temps que le classificateur GMM est beaucoup plus performant que la quantification vectorielle.

D'autre part, nous remarquons que le nombre de coefficients idéal, que ca soit pour la VQ ou pour les GMM est de l'ordre de dix DCT. Ceci s'explique par la compression appliquée par celle-ci. En effet, à partir d'un certain ordre les coefficients ne représentent plus l'information utile mais le bruit. D'où une dégradation du système.

IV.2 Mode identification - TIMIT

IV.2.1 Quantification vectorielle (VQ) :

	10 MFCC	*20 MFCC*	*25 MFCC*
16 VQ	79%	92%	97%
32 VQ	87%	97%	98%
64 VQ	92%	99%	100%

Tableau 7.3-Taux d'identification (%) avec modélisation par VQ/voix TIMIT

IV.2.2 Mélanges de Gaussiennes (GMM) :

	10 MFCC	*20 MFCC*	*25 MFCC*
16 GMM	93%	100%	99%
32 GMM	97%	100%	99%
64 GMM	98%	100%	99%

Tableau 7.4-Taux d'identification (%) avec modélisation par GMM/voix TIMIT

La même conclusion peut être tirée à partir des tests sur la voix. Le mélange de gaussiennes reste toujours préférable à la quantification vectorielle.

Nous observons également que le nombre de coefficients MFCC idéal est de l'ordre de vingt. Enfin, il faut noter l'efficacité des GMM pour l'identification en mode indépendant du texte. Même sur une base de cent personnes, on arrive à avoir des taux d'identification de 100%.

IV.3 Mode identification- base réelle

IV.3.1 Visages réelles

	10 DCT	*16 DCT*	*32 DCT*
16 GMM	71.29%	70.37%	72.22%
32 GMM	75%	74.07%	75.92%
64 GMM	84.2%	79.62%	78.7%

Tableau 7.5- TI (%) avec modélisation par GMM /visage ORL

IV.3.2 Voix réelles

	10 MFCC	*20 MFCC*	*25 MFCC*
16 GMM-UBM	89%	100%	100%
32 GMM-UBM	92.59%	100%	100%
64 GMM-UBM	92.6%	100%	100%

Tableau 7.6- TI(%) avec modélisation par GMM-UBM /visage ORL

Les résultats obtenus pour la voix et le visage extraits de la base réelle confirment les conclusions déduites du travail sur les bases « académiques » que sont ORL et TMIT. Les résultats particulièrement élevés pour la voix se justifient par le nombre réduit de personnes testées.

IV.4 Mode authentification - ORL

IV4.1 Mélanges de Gaussiennes (GMM) sans UBM

	10 DCT	*16 DCT*	*32 DCT*
16 GMM	24.89%	33.00%	40.50%
32 GMM	19.5%	28.01%	38.52%
64 GMM	15.04%	23.00%	34.99%

Tableau 7.7- EER (%) avec modélisation par GMM /visage ORL

IV4.2 Mélanges de Gaussiennes (GMM) avec UBM

	10 DCT	*16 DCT*	*32 DCT*
16 GMM-UBM	3.46%	3.98%	5.99%
32 GMM-UBM	2.53%	2.54%	4.08%
64 GMM-UBM	2.50%	2.40%	3.38%

Tableau 7.8- EER (%) avec modélisation par GMM-UBM /visage ORL

Nous observons, avant tout, que les conclusions tirées pour l'identification sont toujours valables en authentification. Les meilleurs résultats restent associés à un nombre de coefficients DCT de dix et à un nombre de gaussiennes de 32. Un ordre de 64 étant trop lourd à générer pour une différence aussi infime.

D'autre part, nous observons une nette amélioration des résultats après application de la norme UBM. Celle-ci permet d'une part d'éliminer ce qui est commun à la population, et donc leurs similitudes. D'autre part, elle permet de représenter l'imposteur qui n'est autre que l'individus moyen de la population.

IV.5 Mode authentification –TIMIT

IV.5.1 Mélanges de Gaussiennes (GMM) sans UBM

	10 MFCC	20 MFCC	25MFCC
16 GMM	37.76%	35.01%	36.93%
32 GMM	36.01%	34.88%	36.85%
64 GMM	33.01%	33.00%	34.25%

Tableau 7.9- EER (%) avec modélisation par GMM /voix TMIT

IV.5.2 Mélanges de Gaussiennes (GMM) avec UBM

	10 MFCC	20 MFCC	25 MFCC
16 GMM-UBM	4.97%	2.80%	2.96%
32 GMM-UBM	4.00%	1.00%	2.00%
64 GMM-UBM	4.22%	2.99%	2 .23%

Tableau 7.10- EER (%) avec modélisation par GMM /voix TMIT

On remarque que l'ordre optimal pour la génération des modèles de voix est toujours de 32 GMM. Ceci, pour un nombre de coefficients MFCC de 20.

Cette fois encore, la normalisation UBM a apporté une amélioration particulièrement intéressante.

IV.6 Mode authentification –Base réelle

IV.6.1 Mode authentification – visage ENP

	10 DCT	16 DCT	32 DCT
16 GMM	19.76%	22.71%	23.04%
32 GMM	16.46%	18.52%	20.63%
64 GMM	14.81%	15.64%	18.94%

Tableau 7.11- EER (%) avec modélisation par GMM /visage ENP

	10 DCT	16 DCT	32 DCT
16 GMM-UBM	4.61%	7.34%	8.33%
32 GMM-UBM	0.94%	0.95%	0.96%
64 GMM-UBM	0.93%	0.93%	0.96%

Tableau 7.12- EER (%) avec modélisation par GMM-UBM /visage ENP

IV.6.2 Mode authentification – voix ENP

	10 MFCC	20 MFCC	25 MFCC
16 GMM	43.59%	41.60%	40.79%
32 GMM	40.60%	40.50%	40.60%
64 GMM	40.67%	40.60%	40.74%

Tableau 7.13- EER (%) avec modélisation par GMM /voix ENP

	10 MFCC	20 MFCC	25MFCC
16 GMM-UBM	6.75%	6.55%	3.7%
32 GMM-UBM	3.77%	3.63%	4.56%
64 GMM-UBM	3.70%	3.7%	4.7%

Tableau 7.14- EER (%) avec modélisation par GMM-UBM /voix ENP

Les résultats de la base réelle, en plus de nous permettre de voir la robustesse du système pour des données réelles, nous permet de généraliser les conclusions obtenues sur les bases ORL\TIMIT. Ainsi, on observe que les meilleurs résultats sont obtenus pour 10 et 20 coefficients DCT et MFCC. Et que généralement, le meilleur ordre du classificateur est de 16 ou 32 gaussiennes. Un choix de 64 gaussiennes n'étant pas justifié, étant donné l'apport minimal pour un temps consommé important.

Nous traçons dans les figures 7.8 et 7.9 les courbes ROC associées à ces différents systèmes pour ORL\TMIT.

Elles indiquent que 10 DCT et 20 MFCC sont les meilleurs choix pour un ordre de 32 gaussiennes.

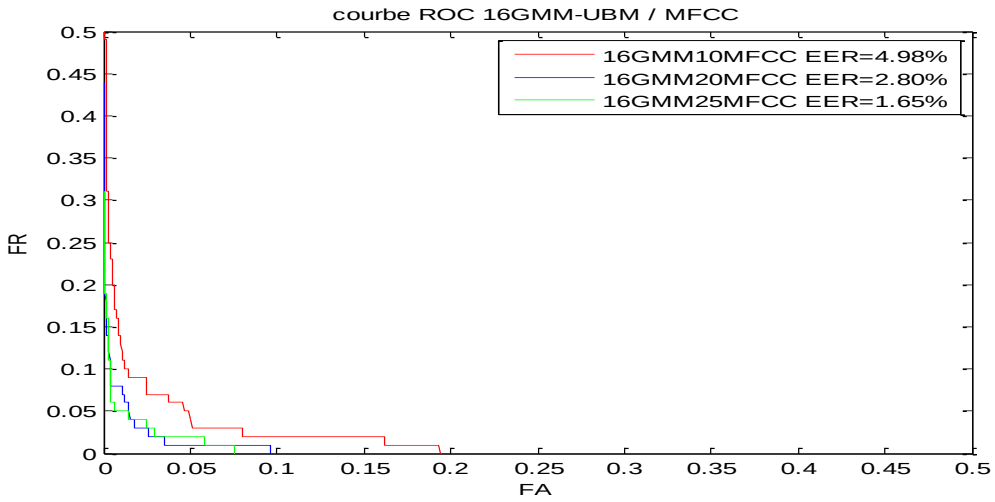


Figure 7.8 : Courbe ROC Pour 16GMM et différents nombres de coefficients MFCC

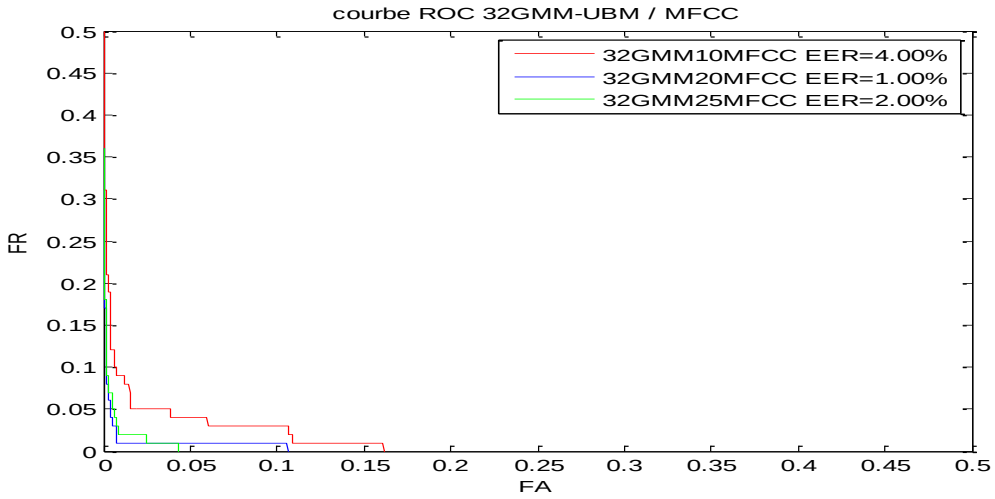


Figure 7.9 : Courbe ROC Pour 32 GMM et différents nombres de coefficients MFCC

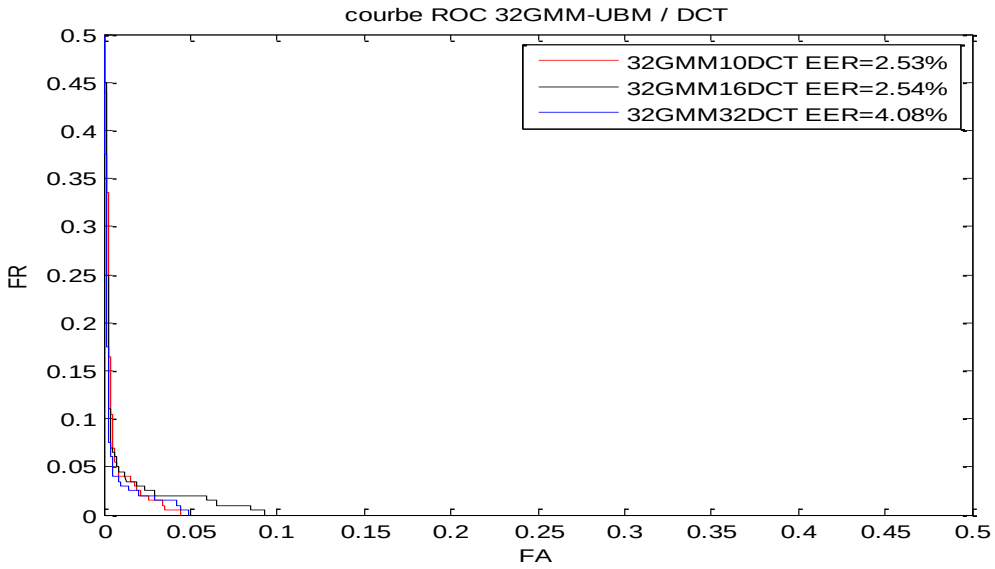


Figure 7.10: Courbe ROC Pour 32 GMM et différents nombres de coefficients DCT

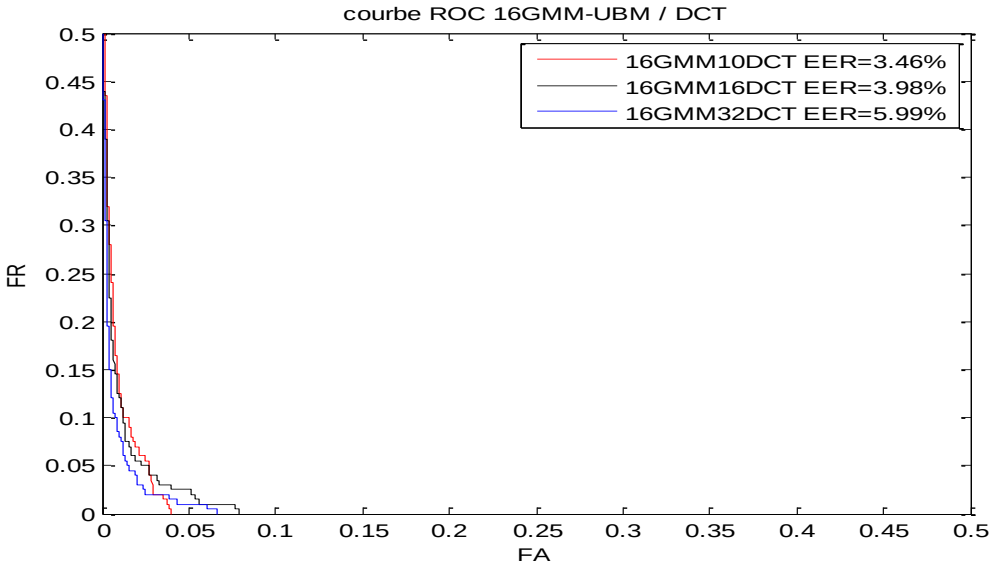


Figure 7.11 : Courbe ROC Pour 16 GMM et différents nombres de coefficients MFCC

V. RESULTAT DES TESTS POUR LES GMMO

La suite de notre travail est axé sur le classificateur GMM orthogonale, mis en oeuvre de deux manières. D'abord en appliquant une PCA sur chaque modèle, et en projetant à chaque traitement les données dans l'espace propre associé. Puis, en appliquant une PCA généralisée

sur toutes les données de toutes les populations en amont de la génération de modèle. Le traitement se faisant dans un même et unique espace propre.

V.1 *Mode authentification « diagonal /orthogonal » -ORL*

V.1.1 Sans modèle UBM

	4	*8*	*16*	*32*	*64*
GMM	**34.58%**	**31.43%**	**24.89%**	**19.50%**	**15.04%**
OGMM-GPCA	**26%**	**20.96%**	**18.50%**	**14.00%**	**10.5%**
OGMM-KLT	**26.10%**	**22.50%**	**19.50%**	**13.80%**	**11.90%**

Tableau 7.15- EER (%) config : 10DCT classificateurs GMM/OGMM sans UBM

V.1.2 Avec modèle UBM :

	4	*8*	*16*	*32*	*64*
GMM-UBM	**16.92%**	**8.57%**	**3.46%**	**2.53%**	**2.5%**
OGMM-GPCA-UBM	**11%**	**7.4%**	**3.5%**	**2.4%**	**2.5%**
OGMM-KLT-UBM	**14.5%**	**8.56%**	**3.53%**	**2.99%**	**2.2%**

Tableau 7.16- EER (%) config : 10DCT classificateurs GMM/OGMM avec UBM

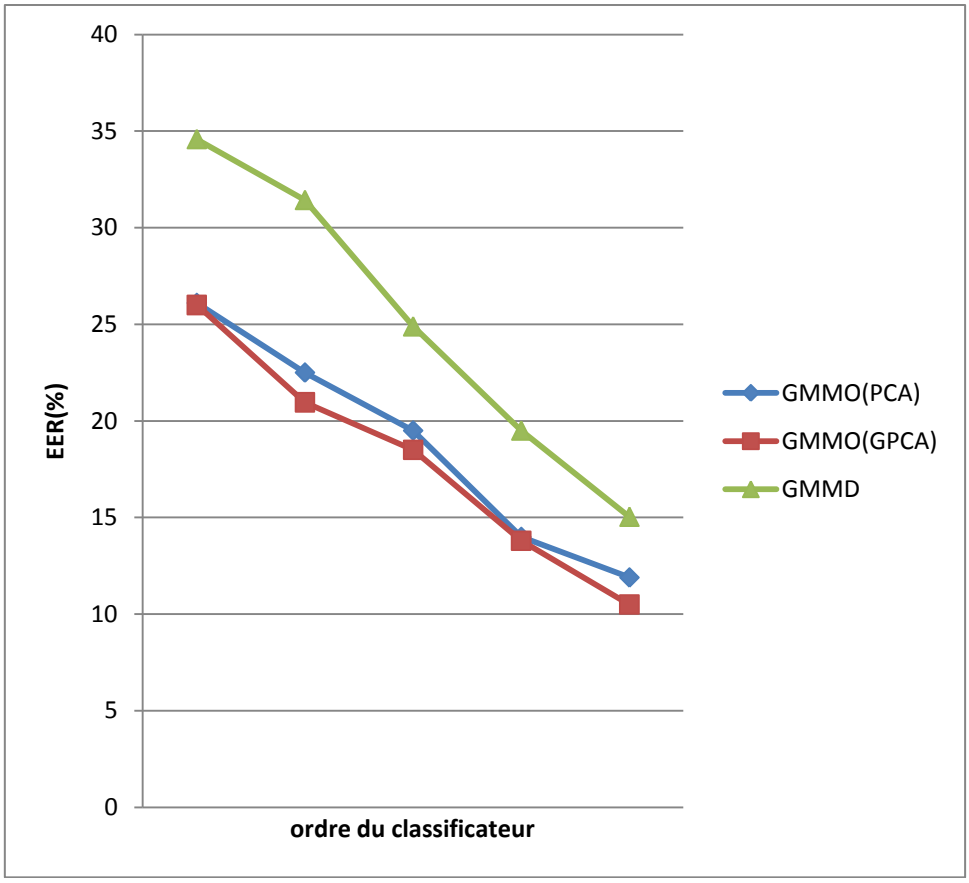


Figure 7.12 : comparaison des différentes variantes de GMM pour différents ordres pour le visage (10 DCT)

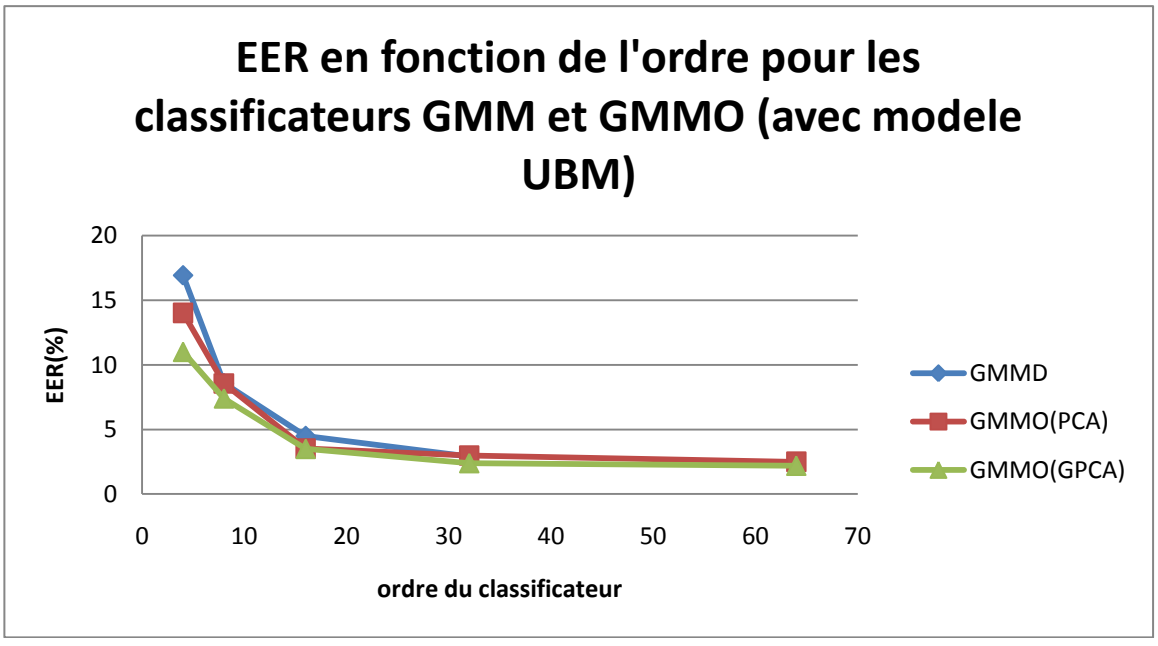


Figure 7.13 : comparaison des différentes variantes de GMM pour différents ordres pour le visage (10 DCT) avec modèle UBM

Nous observons une nette amélioration des différents modes, principalement au niveau des ordres faibles (voir figure 7.9 et 7.10).

Remarque-2 : Bien que n'entrant pas directement dans le cadre de notre projet, nous soulignons le fait que l'orthogonalisation a apporté également une amélioration à l'identification (voir figure 7.11).

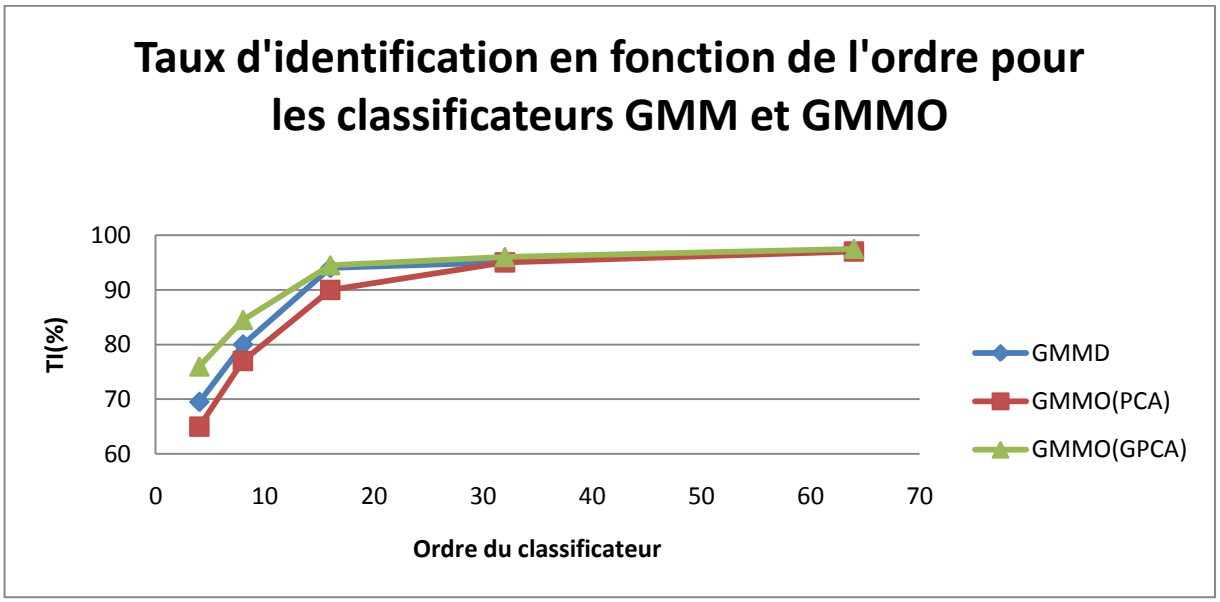


Figure 7.14: comparaison des différentes TID de differentes GMM pour différents ordres pour le visage (10 DCT) avec modèle UBM

V.2 *Mode authentification « diagonal /orthogonal » -TIMIT:*

V.2.1 Sans modèle UBM :

	4	*8*	*16*	*32*	*64*
GMM	**40%**	**36%**	**35.01%**	**34.88%**	**33.00%**
OGMM-GPCA	**38.2%**	**35.80%**	**35.00%**	**34.00%**	**34.20%**
OGMM-KLT	**36.87%**	**36%**	**35.20%**	**34.80%**	**34.33%**

Tableau 7.17- EER (%) config : 20MFCC classificateurs GMM/OGMM sans UBM

V.2.2 Avec modèle UBM :

	4	8	16	32	64
GMM-UBM	11%	4.01%	2.8%	1.0%	2.99%
OGMM-GPCA-UBM	7.88%	3.50%	1.77%	1.1%	1.9%
OGMM-KLT-UBM	8.25%	3.7%	2.7%	1.31%	2.0%

Tableau 7.18- EER (%) config : 20MFCC classificateurs GMM/OGMM avec UBM

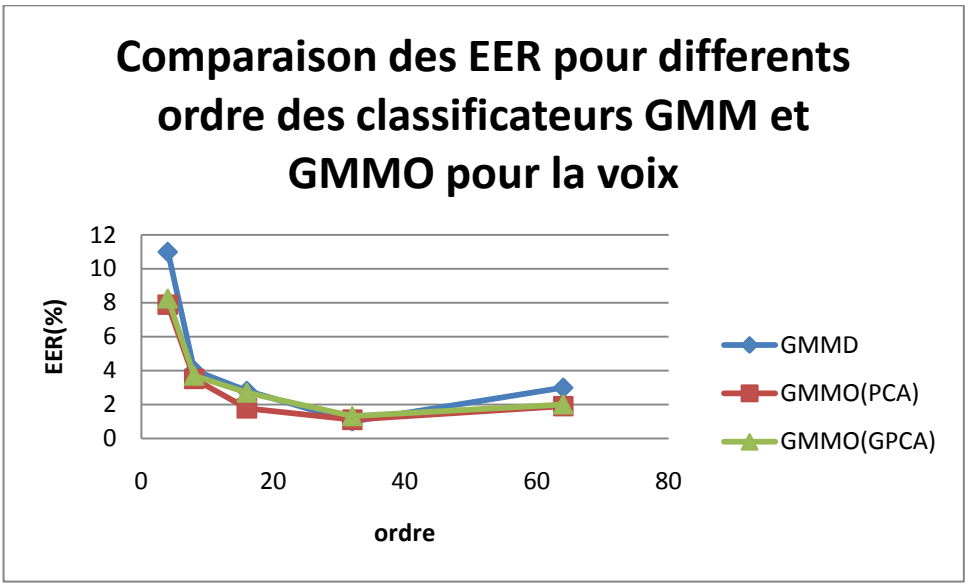


Figure 7.15: comparaison des différentes TID de differentes GMM pour différents ordres pour le visage (10 DCT) avec modèle UBM

Nous observons une amélioration du taux d'authentification que ca soit pour le visage ou pour la voix. Cela dit, celle-ci est beaucoup plus notable pour le visage (voir figure 7.10 et 7.12) ceci étant probablement dû à une plus forte corrélation des coefficients DCT causée par un chevauchement plus important.

D'autre part, nous observons un bien plus grand effet de l'opération d'orthogonalisation sur les ordres bas. En effet, les GMMO étant plus proches des GMM full que les GMM diagonales, elles ont tendance a apporter une meilleure estimation. Cela dit, dépassé un

certain ordre, toutes les deux convergent vers les GMM full qui sont le cas optimal du point de vue resultat.

Une autre remarque que nous pouvons faire est que la PCA globale a apporté une meilleure amélioration que la PCA appliquées à chaque personne. Ceux-ci peut être lié au fait que la transformée en cosinus (présente également dans l'extraction des paramètres MFCC) ait déjà un pouvoir de décorrelation remarquable. Il devient alors plus intéressant d'appliquer une transformation globale qui transforme l'allure générale des points, que de l'appliquer localement et n'apporter qu'une légère amélioration à des classes qui sont déjà pratiquement orthogonales.

Remarque-1 : Notons que l'application de la PCA a permis de réduire la dimension des vecteurs de données de moitié. On peut justifier cela par le caractère même de la PCA. En effet les vecteurs propres associés aux valeurs propres les plus importantes sont toujours les plus significatifs. Car ils mettent en valeurs les dimensions dont l'étalement est le plus important. Il ne reste alors que du bruit qui n'apporte rien au système et peut même le détériorer.

V.3 Mode authentification « diagonal /orthogonal » - Base réelle

	4	*8*	*16*	*32*	*64*
GMM-UBM	**17.50%**	**8.1%**	**4.61%**	**0.94%**	**0.95%**
OGMM-GPCA-UBM	**10.90%**	**4.56%**	**1.69%**	**0.94%**	**0.93%**
OGMM-KLT-UBM	**12.04%**	**4.56%**	**2.78%**	**1.75%**	**0.94%**

Tableau 7.19- EER (%) config : 10 DCT classificateurs GMM/OGMM avec UBM

	4	*8*	*16*	*32*	*64*
GMM-UBM	**9.00%**	**8.4%**	**6.55%**	**3.77%**	**3.7%**
OGMM-GPCA-UBM	**6.56%**	**4.8%**	**3.6%**	**3.28%**	**3.7%**
OGMM-KLT-UBM	**8.12%**	**7.26%**	**6.67%**	**3.7%**	**3.7%**

Tableau 7.20- EER (%) config : 20MFCC classificateurs GMM/OGMM avec UBM

Les tests sur la base réelle nous permettent non seulement de confirmer les conclusions tirées à partir des bases ORL/TIMIT mais aussi de confirmer l'intérêt de l'approche GPCA par rapport à la KLT appliquée à chaque personne, celle-ci étant instable et n'apportant pas des améliorations à tous les coups.

VI. RESULTATS DES TESTS POUR LA FUSION

Nous avons pris les cas qui nous paraissaient le plus favorable pour le classificateur GMM diagonal et lui avons appliqué différentes techniques de fusion.

ORL/TIMIT : - 32 GMM et 20 coefficients MFCC pour la voix (EER=1.00%)
- 32 GMM et 10 DCT pour le visage (EER=2.53%)

Base reelle : - 32 GMM et 20 MFCC pour la voix (EER=2.92%)
- 32 GMM et 10 DCT pour le visage (EER=0.94%)

	Min-Max	*Z-score*	*Tanh*	*AdaptativeQQ*	*AdaptativeLG*
MIN	1.71%	1.40%	1.24%	1.25%	1.17%
MAX	0.96%	0.24%	0.24%	0.28%	0.67%
Somme Simple	0.15%	0.37%	0.19%	0.13%	0.50%
Somme Pondérée	0.25%	0.19%	0.16	0.1%	0.2%

Tableau 7.21- EER (%) avec fusion config : ORL/TIMIT

	Min-Max	*Z-score*	*Tanh*	*AdaptativeQQ*	*AdaptativeLG*
MIN	1.9%	2.1%	0.9%	1.2%	1.1%
MAX	1.0%	1.3%	1.2%	0.3%	0.5%
Somme Simple	0.7%	0.27%	0.7%	0.3%	0.1%
Somme Pondérée	0.09%	0.13%	0.22%	0.2%	0.4%

Tableau 7.22- EER (%) avec fusion config : Base réelle

Nous avons également effectuer une fusion à l'aide du classificateur GMMO. Celui-ci nous à donné le meilleur résultat qui est un EER de 0.07% pour ORL/TIMIT et 0.08% pour la base de l'ENP.

L'amélioration de la fusion est tout d'abord dûe au passage à une dimension supérieure. Le passage à une dimension supérieure, lorsque les dimension sont toutes importantes apporte toujours des améliorations.

La classification apporte une amélioration supérieure car elle permet de complètement distinguer les classes client et imposteur, facilitant ainsi la tache de les départager. Le partage se faisant par des lois probabilistes, identiques aux phases d'apprentissage et de tests sur les personnes, plutôt que par des lois d'ordre.

Conclusion et perspectives

Durant ce travail, nous avons pu aborder de manière générale les systèmes biométriques et plus particulièrement ceux basés sur le visage et la voix.

Dans un premier temps, notre travail nous a permis de vérifier l'efficacité du classificateur GMM standard par rapport à d'autres techniques que ce soit pour le visage ou la voix. Nous avons également pu tirer une conclusion par rapport aux combinaisons DCT-GMM et MFCC-GMM donnant les meilleurs résultats, et voir l'apport de certaines techniques de normalisation telles que UBM sur l'authentification.

Suite à cela, nous avons entamé une étude comparative de ce classificateur avec le concept d'OGMM qui lui apporte une amélioration. Nous avons d'abord essayé une technique qui applique une KLT à chaque personne séparément. Puis, une technique plus récente, nommée GPCA, qui traite les données de toutes les personnes simultanément. Nous avons ainsi pu comparer les résultats de deux approches différentes déjà appliquées sur la voix. Et nous avons même pu vérifier la validité de ces résultats pour une autre modalité telle que le visage. On a conclut que l'approche par GPCA était la plus intéressante en précision, et en temps, que ce soit pour le visage, la voix ou après fusion. Puisqu'elle ne nécessite qu'un seul calcul avant le traitement.

Pour terminer, nous avons testé une panoplie de techniques de fusion, qu'elles soient simples ou basées sur les classificateurs GMM et OGMM.

Comme perspective, nous pensons qu'il serait intéressant de faire une réalisation logicielle ou matérielle pour mieux évaluer l'apport de la nouvelle méthode en temps et en complexité.

Bibliographie :

[Ahm 74] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform", IEEE Trans. Computers, 90-93, Jan 1974.

[And 87] M.Andre, « Introduction aux techniques de traitement d'images », Eyrolles 1987.

[Atl 74] B.Atal, «Effectiveness of linear prediction of speech wave for automatic speaker identification and verification», Journal of statistical society of America (JASA), vol. 55, 1974.

[Auc 00] M.Auckenthar, R.Carey, L.Thomas, « Score normalization for text independant speaker verification systems », Digital signal processing, 2000.

[Ben 02] S.Bengio, C.Marcel, S.Marcel, J.Mariéthoz, «Confidence measures for multimodal identity verification», Information Fusion, 2002.

[Ben 04] M.Ben, «Approche robust pour la vérification automatique du locuteur par normalization et adaptation hiérarchique», Thèse de doctorat Université de Rennes, 2004.

[Ben 05] M.Benkiniouar, Mohamed Benmohamed, «Méthodes d'identification et de reconnaissance de visages en temps réel basées sur AdaBoost» Article ,2005.

[Ber 93] A. Bertillon, « Identification Anthropométrique et Instructions Signalétiques », Melun: Imprimerie administrative, 1893.

[Bil 97] J. Bilmes, «A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models», Report, University of Berkeley, ICSI-TR-97-021, 1997.

[Bio 08] http://www.biometrie-online.net/index.php (janvier 2008).

[Bol 03] D. Bolme, J. Beveridge, M. Teixeira, and B. Draper, «The CSU Face Identification Evaluation System : Its Purpose, Features, and Structure » in Proceedings of the 3rd International Conference on Computer Vision Systems (ICVS), 2003.

[Bou 05] A.Boucher, « Traitement d'images » Cours de Traitement d'images, Institut de la francophonie pour l'informatique (IFI), 2005.

[Cam 97] J.Campbell, « Speaker recognition : A tutorial », Proc. IEEE, vol. 85, 1997.

[Cap 95] O.Cappé, « Etat actuel de la recherche en reconnaissance du locuteur et des applications en criminalistique », 1995.

[Car 92] M. J. Carey, E. S. Parris, «Speaker verification using connected words», Proceedings of Institute of Acoustics, 1992.

[Cha 02] M. Chassé, «La biométrie au Québec : Les enjeux», Analyste en informatique de la Commission d'accès à l'information, Québec, 2002.

[Chu 01] L. Chung Ern, G. Sulong, « Fingerprint Classification Approaches », International Symposiom on Signal Processing and its Applications, Vol. 1, Kuala Lumpur, Malaisie, 13-16 Août 2001.

[Dau 93] J. Daugman, « High Confidence Visual Recognition of Persons by a Test of Statistical Independence », IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, 1993.

[Dem 77] A. P. Dempster, N. M. Laird, D. B. Rubin, «Maximum-likelihood from incomplete data via the EM algorithm», Journal of Acoustical Society of America JASA, 1977.

[Dod 00] G.R. Doddington, M.A. Przybocki, « The NIST speaker recognition evaluation - Overview, methodology, systems, results, perspective », Speech Communication, 2000.

[Dry 00] A.Drygajlo, M.Eel-Maliki, « Integration and Imputation Methods for Unreliable Feature Compensation in GMM Based Speaker Verification », June 2000.

[Enc 97] Encyclopédie ENCARTA, 1997.

[Fly 08] P. Flynn, A. Ross, K. Jain, «Handbook of Biometrics », Springer, 2008.

[Gau 94] J. L. Gauvain, C. H. Lee, «Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains», IEEE Transactions on Speech and Audio Processing, 1994.

[Gon 77] R.C.Gonzales, P.Wintz, « Digital Image Processing », Addison Wessley, 1977.

[Guo 00] G. Guo, S.Z. Li, K. Chan, «Face Recognition by Support Vector Machines, Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition», Grenoble, France March 2000.

[Hiz 09] W.Hizem, « Capteur Intelligent pour la Reconnaissance de Visage », Thèse de doctorat à l'Institut National des Télécommunications et l'Université Pierre et Marie Curie - Paris 6, France ; 2009.

[Hon 99] L. Hong, A. Jain, S. Pankanti, « Can Multibiometrics Improve Performance ? », Proceedings AutoID'99, Summit, NJ, Oct 1999.

[Jai 99] A.K. Jain, A. Ross, and S. Pankanti, « A prototype hand geometry-based verification system », Proc. of 2nd International Conf. on Audio- and Video-based Biometric Person Authentication, March 1999.

[Jai 00] A. K. Jain, L. Hong, S. Pankanti, « Biometrics : Promising Frontiers for Emerging Identification Market », Communications of the ACM, February 2000.

[Jai 01] A. Jain, S. Pankanti, « Advances in Fingerprint Technology », 2^{nd} édition, Elsevier Science, New York, 2001.

[Jai 04] A.K. Jain, R. Arun, P. Salil, « An Introduction to Biometric Recognition, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics », Vol. 14, No. 1, Janvier 2004.

[Kha 02] J.Kharroubi « Etude de Techniques de Classement : Machines à Vecteurs Supports pour la Vérification Automatique du Locuteur »; Thèse de doctorat à l'Ecole Nationale Supérieure des Télécommunications ; 2002.

[Kha 03] S-A.Khayam, « The Discrete Cosine Transform (DCT) », ECE 802 – 602, Information Theory and Coding, Seminar 1 – The Discrete Cosine Transform: Theory and Application ; Mars 2003.

[Kit 98] J.Kittler, M.Hatef, W. Duin, J.Matas, «On combining classifiers», IEEE Trans,Pattern Anal, Mach, Intell, 1998.

[Kon 05] H.Kong, X.Li, L.Wang, E.Khwang-Tho, « Generalized 2D Principal Component Analysis », School of Electrical and Electronic Enineering Nanyang Technological University; 2005.

[Kun 93] M.Kunt, « Traitement numérique des images », 1993.

[Lar 07] Larousse, 2007.

[Las 05] M. T. Laskri, S.Yessaadi, «Un modèle basé Templates Matching/Réseau de neurones pour la reconnaissance des visages humains» P2. Groupe de recherche en intelligence artificielle, Département d'informatique, Université d'Annaba, 2005.

[Liu 01] S. Liu, M. Silverman, « A Practical Guide to Biometric Security Technology », IEEE Computer Society, IT Pro-Security, Janvier-Février 2001.

[Mal 03] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, « Handbook of Fingerprint Recognition», Springer, 2003.

[Mam 03] Y.Mami , «Reconnaissance de locuteurs par localization dans un espace de locuteurs de référence », Thèse de doctorat, Ecole supérieur des télécommunication. TelecomParis, 2003.

[Mor 06] N.Morizet, T.EA, F.Rossant, F.Amiel, A.Amara, « Revue des algorithmes PCA, LDA et EBGM utilisés en reconnaissance 2D du visage pour la biométrie » ; Institut Supérieur d'Electronique de Paris (ISEP), Département d'Electronique, 2006.

[Mor 09] N.Morizet, « reconnaissance biométrique par fusion multimodale du visage et de l'iris » ; thèse de doctorat à l'Ecole National Supérieure de télécommunication, France, 2009.

[Nef 98] A.V. Nefian, M.H. Hayes III, «Hidden Markov Models for Face Recognition, Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing», ICASSP'98, 12-15 May 1998, Seattle, Washington, USA.

[Pen 93] W.B. Pennebaker, J.l. Mitchell, «JPEG - Still Image Data Compression Standard, Newyork: International Thomsan Pulishing», 1993.

[Phi 00] P. Phillips, A. Martin, C. Wilson, M. Przybocki, « An Introduction to Evaluating Biometric Systems », Computer,Vol. 33, n°2, Février 2000.

[Phi 00] P. Phillips, H. Hyeonjoon, S. Rizvi, P. Rauss, « The FERET Evaluation Methodology for Face-Recognition Algorithms », IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, n°10, Octobre 2000.

[Pla 89] R. Plamondon, G. Lorette, « Automatic Signature Verification and Written Identification: The State of the Art », Pattern Recognition, Vol. 22, 1989.

[Pra 02] S. Prabhakar, A. Jain, « Decision-Level Fusion in Biometric Verification », Pattern Recognition, Vol. 35, 2002.

[Rab 97] L.Rabiner, B.Juang, «Fundamentals of speech recognition», Prentice Hall, 1997.

[Rey 95] D. A. Reynolds, «Speaker identification and verification using gaussian mixture speaker models», Speech Communication, 1995.

[Ros 92] Rosenberg, A.Delong, J.Lee, C.Juang, B.Soong, «The use of cohort normalized scores for speaker recognition», ICSLP, 1992.

[Ros 06] A.Ross, K.Nandakumar, K. Jain, «Handbook of Multibiometrics» **,** Springer, 2006.

[San 00] R. Sanchez-Reillo, C. Sanchez-Avila, A. Gonzalez-Marcos, « Biometric Identification through Hand Geometry Measurements », IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol. 22, Octobre 2000.

[Sne 05] R. Snelick, U. Uludag, Alan Mink, M. Indovina and A. Jain « Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems » 2005.

[Sui 04] R. Suikerbuik, H. Tangelder, H. Daanen, and A. Oudenhuijzen, «Automatic feature detection in 3D human body scans » in Proceedings of the Conference SAE Digital Human Modelling for Design and Engineering, 2004.

[Tom 04] Carlo Tomasi, « Estimating Gaussian Mixtures Densities with EM», Tutorial, Duke University, 2004.

[Tur 91] M. Turk, A. Pentland, «Eigenfaces for Recognition, Journal of Cognitive Neurosicence», Vol. 3, No. 1, 1991.

[Ver 99] P.Verlinde, « Une Contribution à la Vérification Multimodale de l'Identité en Utilisant la Fusion de Décision », École Nationale Supérieure des Télécommunications, Paris, France, Septembre 1999.

[Wis 97] L.Wiskott, J-M.Fellous, N.Krüger, C-V-D.Malsburg, « Face Recognition by Elastic Bunch Graph Matching »; Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 19, NO. 7, Juliet 1997.

[Zha 02] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, « Face Recognition: A Literature Survey », UMD CAR-TR-948, 2000.

[Zia 01] Ziad M.Hafed, Martin D.Levine «Face Recognition Using the Discrete Cosine Transform" International Journal of Computer Vision » Volume 43 , Issue 3 Juillet/Aout, 2001.

[Zwi 81] E.Zwicker, R.Feldtkller, «Psuchoacoustique », Collection technique et scientifique de télécommunications, 1981.

ANNEXES

Sommaire

Annexe A
Généralités sur le traitement d'images

A-1 Introduction :

Avec la parole, l'image constitue l'un des moyens les plus importants qu'utilise l'homme pour communiquer avec autrui. C'est un moyen de communication universel dont la richesse du contenu permet aux êtres humains de tout âge et de toute culture de se comprendre.

C'est aussi le moyen le plus efficace pour communiquer, chacun peut analyser l'image à sa manière, pour en dégager une impression et d'en extraire des informations précises.

De ce fait, le traitement d'images est l'ensemble des méthodes et techniques opérant sur celles-ci, dans le but de rendre cette opération possible, plus simple, plus efficace et plus agréable, d'améliorer l'aspect visuel de l'image et d'en extraire des informations jugées pertinentes [Ala 05].

A-2 Définition de l'image :

L'image est une représentation d'une personne ou d'un objet par la peinture, la sculpture, le dessin, la photographie, le film, etc.

C'est aussi un ensemble structuré d'informations qui, après affichage sur l'écran, ont une signification pour l'œil humain.

Elle peut être décrite sous la forme d'une fonction I(x,y) de brillance analogique continue, définie dans un domaine borné, tel que x et y sont les coordonnées spatiales d'un point de l'image et I est une fonction d'intensité lumineuse et de couleur. Sous cet aspect, l'image est inexploitable par la machine, ce qui nécessite sa numérisation [And 87].

A-3 Image numérique :

Contrairement aux images obtenues à l'aide d'un appareil photo, ou dessinées sur du papier, les images manipulées par un ordinateur sont numériques (représentées par une série de bits).

L'image numérique est l'image dont la surface est divisée en éléments de tailles fixes appelés cellules ou pixels. La numérisation d'une image est la conversion de celle-ci de son état analogique (distribution continue d'intensités lumineuses dans un plan xOy) en une image numérique représentée par une matrice bidimentionnelle de valeurs numériques f(x,y) où :

x, y : coordonnées cartésiennes d'un point de l'image ; f(x, y) : niveau de gris en ce point.

Pour des raisons de commodité de représentation pour l'affichage et l'adressage, les données images sont généralement rangées sous formes de tableau I de n lignes et p colonnes. Chaque élément I(x, y) représente un pixel de l'image et à sa valeur est associé un niveau de gris codé sur m bits (2^m niveaux de gris ; 0 = noir ; 2^m-1 = blanc). La valeur en chaque point exprime la mesure d'intensité lumineuse perçue par le capteur.

A-4 Caractéristiques d'une image numérique :

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants:

II.1 1-Pixel :

Contraction de l'expression anglaise " picture elements ": éléments d'image, le pixel est le plus petit point de l'image, c'est une entité calculable qui peut recevoir une structure et une quantification. Si le bit est la plus petite unité d'information que peut traiter un ordinateur, le pixel est le plus petit élément que peuvent manipuler les matériels et logiciels d'affichage ou d'impression [Enc 97].

La quantité d'information que véhicule chaque pixel donne des nuances entre images monochromes et images couleurs. Dans le cas d'une image monochrome, chaque pixel est codé sur un octet, et la taille mémoire nécessaire pour afficher une telle image est directement liée à la taille de l'image.

Dans une image couleur (R.V.B.), un pixel peut être représenté sur trois octets : un octet pour chacune des couleurs : rouge (R), vert (V) et bleu (B).

III.1 2-Dimension :

C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image [Enc 97].

IV.1 3-Résolution :

C'est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. Sur les moniteurs d'ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution pour

désigner le nombre total de pixels affichables horizontalement ou verticalement sur un moniteur; plus grand est ce nombre, meilleure est la résolution [Enc 97].

V.1 4-Bruit :

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur [Gon 77].

VI.1 5-Histogramme :

L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image. Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme correspondant.

Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est repartie la majorité des niveaux de gris (couleur) dans les cas d'une image trop claire ou d'une image trop foncée.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci [Kun 93] [Gon 77].

VII.1 6-Contours et textures :

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative [GRA 91]. Les textures décrivent la structure de ceux-ci. L'extraction de contour consiste à identifier dans l'image les points qui séparent deux textures différentes [Kun 93].

VIII.1 7-Luminance :

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet [Kun 93].

IX.1 8-Contraste :

C’est l’opposition marquée entre deux régions d’une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d’images [Kun 93].

Si L1 et L2 sont les degrés de luminosité respectivement de deux zones voisines A1 et A2 d’une image, le contraste C est défini par le rapport :

$$C = \frac{L1-L2}{L1+L2} \qquad \text{[A-1]}$$

X.1 9-Images à niveaux de gris :

Le niveau de gris est la valeur de l’intensité lumineuse en un point. La couleur du pixel peut prendre des valeurs allant du noir au blanc en passant par un nombre fini de niveaux intermédiaires. Donc pour représenter les images à niveaux de gris, on peut attribuer à chaque pixel de l’image une valeur correspondant à la quantité de lumière renvoyée. Cette valeur peut être comprise par exemple entre 0 et 255. Chaque pixel n’est donc plus représenté par un bit, mais par un octet. Pour cela, il faut que le matériel utilisé pour afficher l’image soit capable de produire les différents niveaux de gris correspondant.

Le nombre de niveaux de gris dépend du nombre de bits utilisés pour décrire la " couleur " de chaque pixel de l’image. Plus ce nombre est important, plus les niveaux possibles sont nombreux.

XI.1 10-Images en couleurs :

Même s’il est parfois utile de pouvoir représenter des images en noir et blanc, les applications multimédias utilisent le plus souvent des images en couleurs. La représentation des couleurs s’effectue de la même manière que les images monochromes avec cependant quelques particularités. En effet, il faut tout d’abord choisir un modèle de représentation. On peut représenter les couleurs à l’aide de leurs composantes primaires. Les systèmes émettant de la lumière (écrans d’ordinateurs,...) sont basés sur le principe de la synthèse additive : les couleurs sont composées d’un mélange de rouge, vert et bleu (modèle R.V.B.) [And 87].

a- La représentation en couleurs réelles :

Elle consiste à utiliser 24 bits pour chaque point de l’image. Huit bits sont employés pour décrire la composante rouge (R), huit pour le vert (V) et huit pour le bleu (B). Il est ainsi possible de représenter environ 16,7 millions de couleurs différentes simultanément. Cela est cependant théorique, car aucun écran n’est capable d’afficher 16 millions de points. Dans la

plus haute résolution (1600 x 1200), l'écran n'affiche que 1 920 000 points. Par ailleurs, l'œil humain n'est pas capable de distinguer autant de couleurs.

b- La représentation en couleurs indexées :

Afin de diminuer la charge de travail nécessaire pour manipuler des images en 24 bits, on peut utiliser le mode de représentation en couleurs indexée. Le principe consiste à déterminer le nombre de couleurs différentes utilisées dans l'image, puis à créer une table de ces couleurs en attribuant à chacune une valeur numérique correspondant à sa position dans la table. La table, appelée palette, comporte également la description de chacune des couleurs, sur 24 bits.

c- Autres modèles de représentation :

Le modèle R.V.B. représentant toutes les couleurs par l'addition de trois composantes fondamentales, n'est pas le seul possible. Il en existe de nombreux autres. L'un d'eux est particulièrement important. Il consiste à séparer les informations de couleurs (chrominance) et les informations d'intensité lumineuse (luminance). Il s'agit du principe employé pour les enregistrements vidéo. La chrominance est représentée par deux valeurs (selon des modèles divers) et la luminance par une valeur.

XII.1 11-Convolution :

C'est l'opérateur de base du traitement linéaire des images. Soit I une image numérique.

Soit h une fonction de $(x_1, x_2)\times(y_1, y_2)$ à valeurs réelles.

La convolution de I par h est définie par :

$$(I * h)(x,y) = \sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} h(i,j) * I(x-i, y-j) \qquad \text{[A-2]}$$

La fonction h est dite noyau de la convolution, les nouvelles valeurs du pixel sont calculées par produit scalaire entre le noyau de convolution et le voisinage correspondant du pixel [Mor 06].

A-5 Prétraitements :

XIII.1 1-Traitement à base d'histogrammes :

Présentons quelques traitements d'analyses effectuées uniquement à partir de l'histogramme. Retenons que certains de ces traitements sont souvent calculés au niveau des

capteurs, et qu'en général leur pertinence est très intimement liée aux conditions d'acquisition [Mor 06].

a- Normalisation :

La normalisation d'histogramme, ou expansion de dynamique, est une transformation affine du niveau de gris des pixels de telle sorte que l'image utilise toute la dynamique de représentation.

$$f_{new}(x,y) = (f(x,y) - N_{min})\frac{2^D - 1}{N_{max} - N_{min}} \qquad [A\text{-}3]$$

D : Dynamique ou nombre de bits.

N_{max} : Plus grande valeur dans l'image.

N_{min} : Plus petite valeur dans l'image.

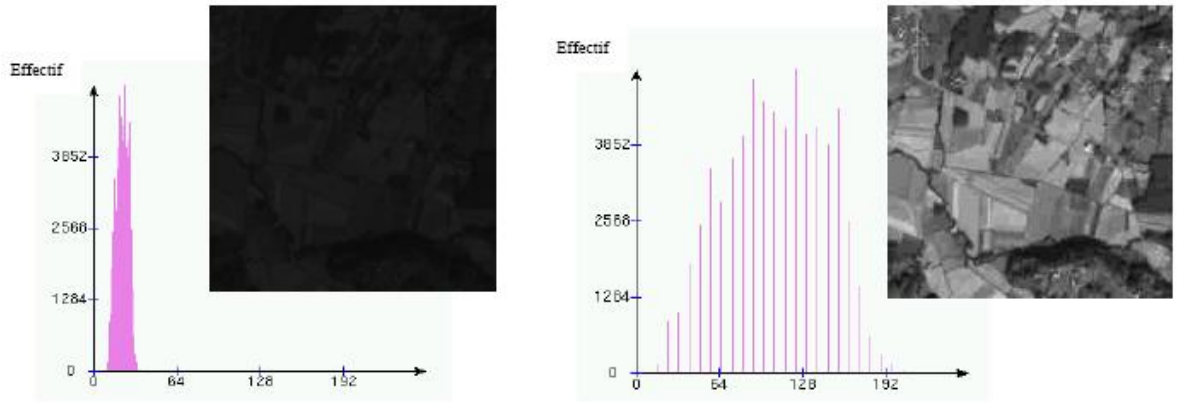


Figure A.1-Exemple de normalisation d'histogramme

b- Egalisation :

L'égalisation d'histogramme est une transformation des niveaux de gris dont le principe est d'équilibrer le mieux possible la distribution des pixels dans la dynamique (Idéalement, on cherche à obtenir un histogramme plat).

La technique classique consiste à rendre « le plus linéaire possible » l'histogramme cumulé de l'image en utilisant la transformation suivante :

$$f_{new}(x,y) = (2^D - 1)\frac{HC(f(x,y))}{wh} \qquad [A\text{-}4]$$

D : Dynamique (nombre de bits).

(w,h) : Dimension de l'image.

HC(.) : Histogramme cumulé.

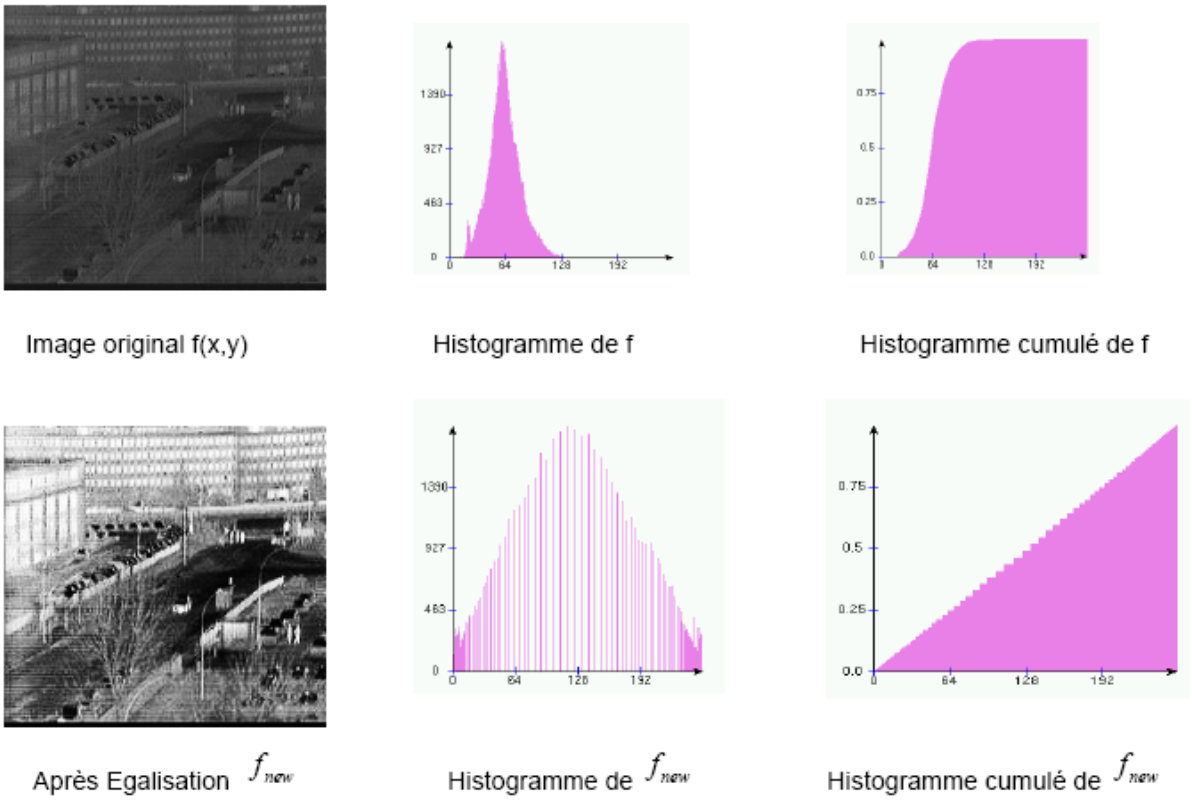


Figure A.2-Exemple d'égaliseur d'histogramme

XIV.12-Filtrage :

a- Filtres linéaires :

C'est le résultat d'une combinaison linéaire des plus proches voisins d'un pixel, parmi les filtres linéaires existant on peut citer :

- *Filtre moyen* :

C'est un filtre qui, comme son nom l'indique, fait la moyenne entre toutes les valeurs de pixels avoisinant un point. Il permet de lisser l'image, réduit le bruit, réduit les détails inutiles et brouille ou rend floue l'image.

Son intérêt est qu'il ne change pas trop le contour. On peut mentionner aussi que plus le filtre est grand, plus le lissage devient important et plus le floue s'accentue ce qui pourrait engendrer des conséquences néfastes.

Figure A.3-Exemple de filtre moyen

Filtre Gaussien :

C'est un filtre qui s'appuie sur la version échantillonnée normalisée de la fonction gaussienne donnée par :

$$h(x,y) = \frac{1}{2\pi\sigma^2} \times e^{\frac{-(x^2+y^2)}{2\pi\sigma^2}} \qquad [A\text{-}5]$$

Le filtre gaussien donnera un meilleur lissage et une meilleure réduction du bruit que le filtre moyenne.

Figure A.4-Exemple de filtre gaussien

b- Filtres non linéaires :

Contrairement aux filtres linéaires, les filtres non linéaires ne sont pas le résultat d'une combinaison linéaire de leurs voisins qui ne peuvent pas s'implémenter comme un produit de convolution. Deux aspects du lissage sont concernés par le filtrage non linéaire :

- Le bruit impulsionnel : les filtres linéaires éliminent mal les valeurs aberrantes.
- L'intégrité des frontières : on souhaiterait éliminer le bruit sans rendre flous les frontières des objets ;

On peut citer et définir parmi les filtres linéaires les plus utilisés :

Filtre Médian :

Permet d'éliminer certains types de bruits (poivre et sel), son principe est de remplacer la valeur d'un pixel par la valeur médiane de la suite mathématique constituée des valeurs des pixels avoisinants à ce point. Pour une meilleure performance de ce filtre, on commence par trier les valeurs des pixels du voisinage, suivra ensuite la détermination de la médiane et enfin l'affectation de cette valeur au pixel. La principale fonction du filtre médian est de forcer des points avec des intensités très distinctes pour être comme leurs voisins, ainsi éliminer réellement les intensités transitoires qui apparaissent isolées dans la zone de masque.

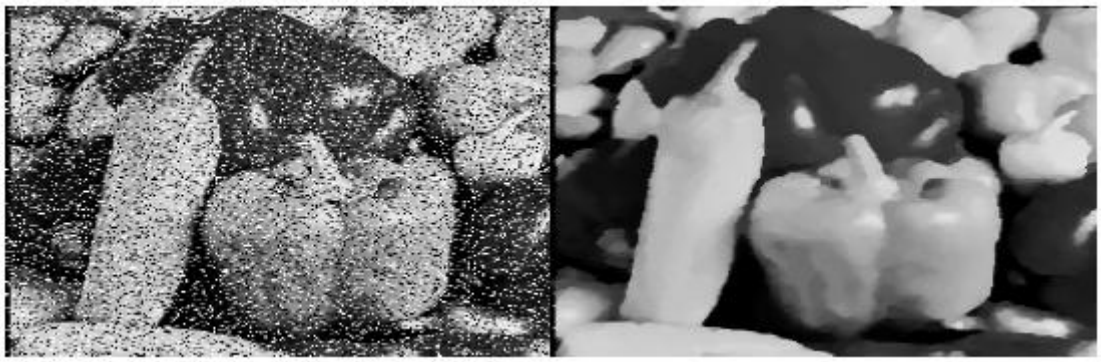


Figure A.5-Exemple de filtre médian

Filtre de Nagao :

Utilisé fréquemment dans les images comportant de très fortes structures artificielles. Découpe d'une fenêtre 5x5 centrée sur le pixel en 9 fenêtres de 9 pixels, mesure sur chacune de ces fenêtres d'une valeur de l'homogénéité (variance par exemple). Le résultat de l'opérateur est la moyenne du domaine qui présente la plus faible variance.

Figure A.6-Exemple de filtre Nagao

c- Filtres de détection de contours :

La détection de contour est une étape préliminaire à de nombreuses applications de l'analyse d'images. Les contours constituent en effet des indices riches, au même titre que les

points d'intérêts, pour toute interprétation ultérieure de l'image. Les contours dans une image proviennent des :

- Discontinuités de la fonction de réflectance (texture, ombre)
- Discontinuités de profondeur (bords de l'objet)

Il existe plusieurs méthodes de détection de contours, on citera à cet effet trois classes suivant la manière d'estimer les dérivées de la fonction d'intensité :

Différences finies :

Une image est discrète par nature. Les premières approches ont donc consisté à approximer les dérivées par différence, ces dérivées sont calculées par convolution de l'image avec un masque de différences. On citera à cet effet les filtres de Roberts, Prewitt, Sobel, Kirsh et Robinson.

Filtrage optimal :

Les dérivations présentées consistent à convoluer l'image par des masques de petites dimensions. Ces approches sont donc dépendantes de la taille des objets traités, elles sont aussi très sensibles au bruit. Un autre type d'approche plus récente repose sur la définition de critères d'optimalité de la détection de contours; ces critères débouchant sur des filtres de lissage optimaux. On citera parmi les filtres correspondants : Canny, Shen-Castan, Deriche, Marr.

Modélisation de la fonction d'intensité:

Les différents filtres cités ci-dessus permettent de calculer le gradient ou le Laplacien d'une image mais ne donnent pas des points de contours. Un traitement ultérieur est nécessaire, ce traitement étant dépendant du type d'approche choisi, approche par Gradient ou approche par le Laplacien.

Annexe B
Quantification vectorielle : Algorithmes LBG et K-means

Introduction :

La quantification vectorielle, méthode de compression de données, a pris une place très importante dans le domaine de la communication, que ce soit dans le but de transmettre ou d'archiver des informations. L'objectif de la compression étant de d'extraire un maximum d'informations avec un minimum de distorsion. Cette méthode s'applique essentiellement dans deux domaines, l'image et la parole, car ces deux forment de signaux contiennent des informations redondantes.

Au cours des dernières années, les travaux de recherche en compression et codage des informations se sont intensément focalisés sur la quantification vectorielle.

Quantification vectorielle :

La quantification vectorielle ou VQ consiste à représenter tout vecteur *x* de dimension *k* par un vecteur *y* de dimension analogue appartenant à un ensemble fini appelé dictionnaire *D*. Lors du codage par VQ, l'étape primordiale réside dans l'élaboration du dictionnaire *D*. Sa création s'effectue à partir d'une séquence d'apprentissage.

B-2-1 Description d'un quantificateur vectoriel :

Un quantificateur vectoriel se décompose en deux applications :

- Codeur
- Décodeur

B-2-2-1 Le codeur :

Le rôle du codeur consiste, pour tout vecteur x du signal d'entrée, à rechercher dans Y le code vecteur y le plus proche du vecteur source x. la notion de proximité a été modélisée dans notre mise en œuvre par la distance euclidienne entre vecteurs.

B-2-2-2 Le décodeur :

Le décodeur dispose d'une réplique du dictionnaire et consulte celui-ci pour fournir le code-vecteur correspondant.

B-2-2-3 Le dictionnaire et ses caractéristiques :

L'élaboration du dictionnaire se fait à partir d'une séquence d'apprentissage. Dans notre cas ce dernier sera déterminé en utilisant l'algorithme LBG.

B-2-2 Algorithme LBG :

Cet algorithme itératif proposé par Linde, Buzo et Gray est utilisé pour la création de dictionnaire dans le cas de la quantification vectorielle. Pour un dictionnaire initial donné, cet algorithme optimise le codeur et le décodeur de façon à obtenir la meilleure partition possible du signal.

Le codeur est défini selon la règle du plus proche voisin :

$$\forall j; \text{si } d(x;z_i) < d(x;z_j) \rightarrow c(x) = z_i$$

Où C représente une région de Voronoï associée à l'espace R^k.

Cette règle détermine donc la partition de R^ken cellules de Voronoï.

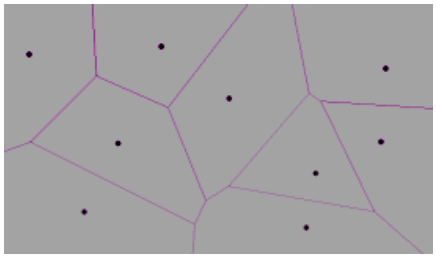


Figure B.1-Un exemple simple de découpage de l'espace en régions de Voronoï.

Lors de la sélection d'un vecteur *y* destiné à représenter un vecteur *x*, la distance séparant les vecteurs est calculée au moyen de la distance Euclidienne d :

$$d(x,y) = \sum_{i=1}^{k} |x_i - y_i|^2$$

La partie la plus sensible de cette technique réside donc dans le choix et la construction du dit dictionnaire. Cet algorithme nécessite un dictionnaire initial. Tout utilisateur averti de la méthode LBG à remarqué qu'une initialisation judicieuse contribuait fortement aux performances de cette technique. En effet chaque itération ne provoquant qu'un changement local du dictionnaire, l'algorithme peut converger vers un minimum local.

Il existe diverses techniques pour choisir le dictionnaire initial. La méthode proposée dans la version initiale de l'algorithme LBG est une méthode par dichotomie vectorielle plus connue sous le nom de « splitting ». Cette technique consiste à diviser chaque vecteur

représentant y en deux nouveaux vecteurs $y+\varepsilon$ et $y-\varepsilon$ où ε est un vecteur aléatoire de perturbation de faible énergie. On applique ensuite les itérations de LBG sur ce nouveau dictionnaire. Le dictionnaire initial est alors le barycentre de la séquence d'apprentissage.

L'algorithme génère ensuite une succession de dictionnaires dans lesquels le nombre de vecteur est multiplié par 2 à chacune des itérations.

En résumé, l'organigramme de cette méthode est donné par :

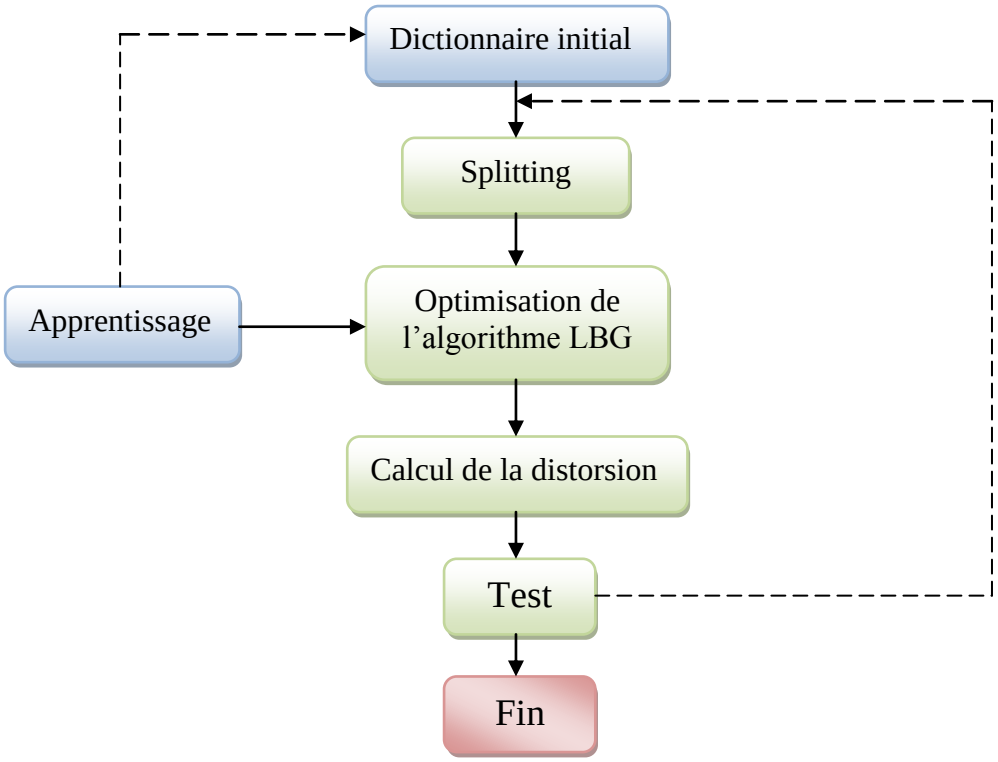


Figure B.2-Organigramme de l'algorithme LBG.

L'algorithme LBG procède donc comme suit :

1) Choix du premier centroïde y constituant le dictionnaire. Dans notre étude, l'initialisation se fera en utilisant l'algorithme K-means que nous verrons par la suite.
2) Le vecteur obtenu est dédoublé en partant du principe de « splitting » en introduisant une marge d'erreur ε de sorte à obtenir deux nouveaux vecteurs $y+\varepsilon$ ainsi que $y-\varepsilon$.
3) Vient ensuite l'association de chaque vecteur de la base au centroïde le plus proche en termes de distance euclidienne.
4) Il en résulte la création de deux groupes distincts affectés à chacun des deux centroïdes, ce qui nous amène à déterminer de nouveaux centroïdes pour chacune des deux régions précédemment obtenues.
5) Itération 1 : répéter les étapes 3 et 4 jusqu'à ce que la distances moyenne « vecteur-centroïde » soit inférieur au seuil précédemment fixé ε.

6) Itération 2 : répéter les étapes 2, 3 et 4 jusqu'à l'obtention du nombre de centroïde désiré.

Notons que le nombre d'itérations peut se déduire de la formule suivante :

$$nbr\ d'itérations = log_2 nbr\ centroïdes\ désirés$$

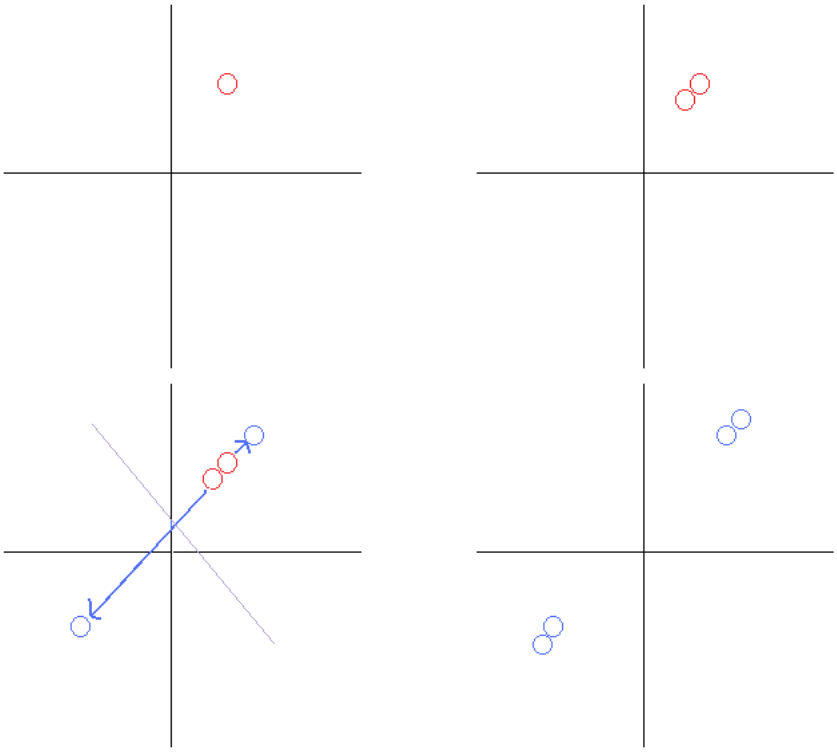


Figure B.3- Premières étapes du « splitting » dans le cas d'une séquence d'entrée de dimension 2.

En haut à gauche : Le centroïde de la séquence d'apprentissage. ***En haut à droite*** : Séparation du centroïde pour obtenir un code à deux mots. ***En bas à gauche :*** Application de l'algorithme de Lloyd sur le code de deux mots pour obtenir un code optimal. La droite violette correspond à la séparation des deux cellules de Voronoï. ***En bas à droite :*** Séparation en deux des deux mots.

B-2-3 Algorithme K-means :

C'est un algorithme qui permet de découvrir les K groupes ou « clusters » d'individus faisant parti de la même communauté en se basant sur une mesure de similarité ou distance afin de grouper les données.

L'approche de la quantification vectorielle fait appel à l'algorithme des K-means ou K-moyennes car ce dernier est basé sur la règle du plus proche voisin.

Comme point de départ, admettons qu'on ait un groupe de X vecteurs tel que :

$X = \{x_t; t = 1 \dots T\}$ qu'on veut diviser en K groupes, chacun d'entre eux étant représenté par un centroïde noté $\{\mu_j; j = 1 \dots K\}$; centroïde auxquels seront attribués les x_t suivant la règle du plus proche voisin.

L'objectif de cet algorithme étant de minimiser l'erreur par le biais de la distance euclidienne tel que :

$$E(X) = \sum_t \|x_t - \mu_t\|^2$$

Avec :

X_k : groupe de données relatif au centroïde μ_k

N_k : nombre d'éléments dans chaque groupe X_k

La règle d'apprentissage du K-mens est basée sur deux principes fondamentaux qui sont :

- Détermination des éléments appartenant à chaque groupe tel que :

$$x \in X_k \text{ si } \|x - \mu_k\| < \|x - \mu_j\| \qquad \forall j \neq k$$

- Adaptation des centroïdes μ_k de chaque groupe car la règle de métrique impose le retrait ou l'ajout de certains éléments. Le centroïde de chaque groupe doit alors se déplacer de sorte à avoir

$$\mu_j = \frac{1}{N_j} \sum X$$

En faisant en sorte d'avoir la plus petite distance euclidienne possible entre élément et centroïde.

Notons que parfois la variance du groupe de données a aussi un grand intérêt comme pour le cas de mixtures de gaussiennes, dans ce cas cette dernière est donnée par :

$$\Sigma_j = \frac{1}{N_j} \sum (X - \mu_j)(X - \mu_j)^T$$

On construit donc K partitions que l'on adapte en ajustant le positionnement des K centroïdes jusqu'à obtenir une similarité satisfaisante.

Pour l'obtention de K centroïdes ce dernier procède donc comme suit :

1) K éléments choisis au hasard sont initialement placés comme étant les centres des K groupes.
2) Chaque élément de l'ensemble que l'on désire partitionner est assigné au centroïde le plus proche, en ce basant sur la mesure de la distance euclidienne, ce qui formera une première version des K groupes.
3) Les centres de gravité de ces groupes sont par la suite recalculés.

4) De ce fait les éléments de l'ensemble sont réaffectés au nouveaux centroïdes obtenus ce qui peut mener a un changement de groupe pour certains d'entre eux.

5) Apres l'obtention des nouveaux groupes, de nouveaux centroïdes sont calculés.
6) Les étapes 2,3 et 4 sont reproduites jusqu'à ce les éléments ne changent plus groupes.

Pour mieux comprendre cet algorithme, voici un exemple pour le cas K=3 :

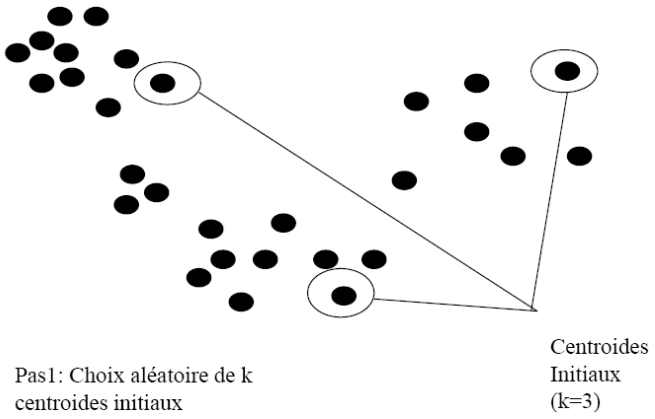


Figure B.4- Choix aléatoire des K centroïdes initiaux.

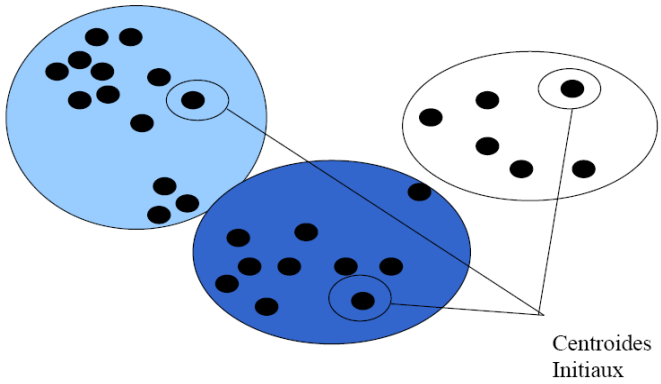


Figure B.5- Formation des premiers K groupes suivant l'affectation des éléments aux centroïdes.

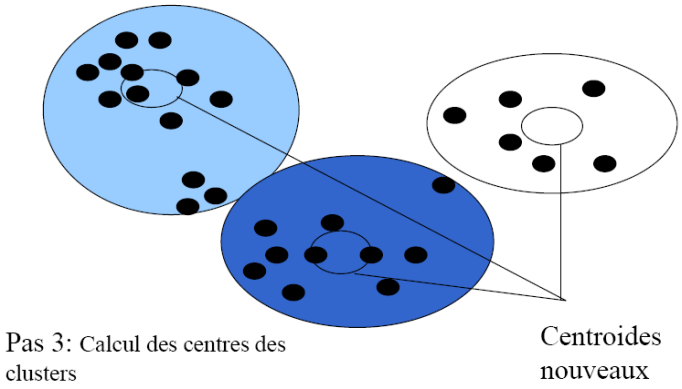


Figure B.6- Calcul des nouveaux centroïdes des K groupes.

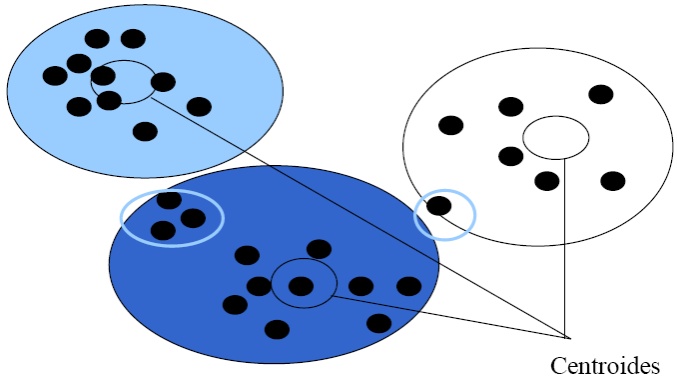


Figure B.7- Réaffectation des éléments par rapports au changement de position des centroïdes.

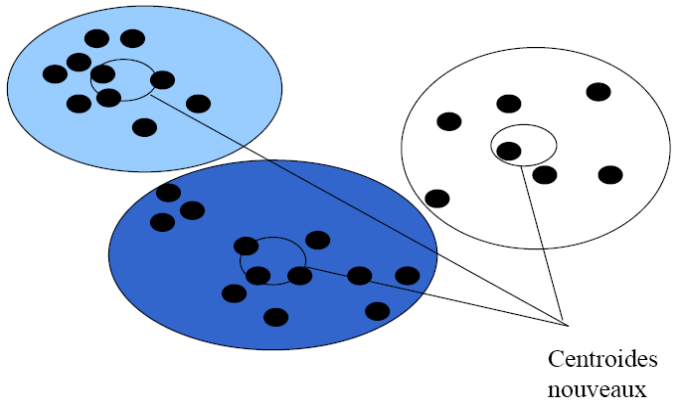


Figure B.8- Formation des K nouveaux groupes suivant le changement de position des K centroïdes.

Annexe C
L'algorithme EM (Expectation-Maximisation)

C-1 Introduction :

L'algorithme EM (Expectation-Maximisation) est un algorithme itératif du à Dempster, Laird et Rubin (1977). Il s'agit d'une méthode d'estimation paramétrique s'inscrivant dans le cadre général du maximum de vraisemblance.

Lorsque les seules données dont on dispose ne permettent pas l'estimation des paramètres, et/ou que l'expression de la vraisemblance est analytiquement impossible à maximiser, l'algorithme EM peut être une solution. De manière grossière et vague, il vise à fournir un estimateur lorsque cette impossibilité provient de la présence de données cachées ou manquantes ou plutôt, lorsque la connaissance de ces données rendrait possible l'estimation des paramètres.

L'algorithme EM tire son nom du fait qu'à chaque itération il opère deux étapes distinctes :

- la phase « Expectation », souvent désignée comme « l'étape E », procède comme son nom le laisse supposer à l'estimation des données inconnues, sachant les données observées et la valeur des paramètres déterminée à l'itération précédente ;
- la phase « Maximisation », ou « étape M », procède donc à la maximisation de la vraisemblance, rendue désormais possible en utilisant l'estimation des données inconnues effectuée à l'étape précédente, et met à jour la valeur du ou des paramètre(s) pour la prochaine itération.

L'algorithme garantit que la vraisemblance augmente à chaque itération, ce qui conduit donc à des estimateurs de plus en plus corrects.

C-2 L'algorithme EM et son principe de convergence :

C-2-1 Formalisation d'une itération de l'algorithme :

- Nous disposons d'observations X de vraisemblance notée $P(X|\theta)$.
- Maximiser le log de vraisemblance n'est par contre impossible.

- Considérons les données cachées Z dont la connaissance rendrait possible de maximisation de la « vraisemblance des données complètes ».
- Le fait ne pas connaitre l'ensemble Z nous conduit à estimer la vraisemblance des données complètes en prenant en compte toutes les informations connues.
- On maximise enfin cette vraisemblance estimée pour déterminer la nouvelle valeur du paramètre.

C-2-2 Convergence de l'algorithme EM :

Les notations adoptées lors de cette étude sont les suivantes :

- $X = \{x_t; t = 1 \dots T\}$ est la séquence d'observation.
- $Z = \{z_t; t = 1 \dots T\}$ est le jeu de données manquantes spécifiant l'information d'état caché.
- $C = \{C^{(j)}; j = 1 \dots J\}$ est le jeu d'étiquettes de chaque groupe de la mixture et où J est le nombre de groupes de la mixture.
- $\theta = \{\theta^{(j)}; j = 1 \dots J\}$ est le jeu de paramètres inconnus qui définissent la fonction de densité de probabilité.
- $\theta^{(j)} = \{\pi^{(j)}, \Phi^{(j)}\}$ ou $\pi^{(j)}$ est la probabilité antérieure de la densité de la $j^{ème}$ composante et $\Phi^{(j)}$ représente la densité de la $j^{ème}$ composante.

On définit :

- $log\ p(X|\theta_n)$ comme étant la vraisemblance de la donnée complète étant donné la donnée courante de θ_n , ou n représente l'index d'itération.
- $log\ p(Z, X|\theta_n)$ comme la vraisemblance de la donnée complète.
- $L(X|\theta) \equiv log\ p(X|\theta_n)$.

De ce fait, la théorie des probabilités nous permet d'écrire :

$$p(X|\theta_n) = \frac{p(Z,X|\theta_n)}{p(Z|X,\theta_n)} \qquad [C\text{-}1]$$

La vraisemblance des données complètes s'écrit donc comme suit :

$$\begin{aligned} L(X|\theta) &\equiv \log\ p(X|\theta_n) \\ &= [\log(X|\theta_n)]\ \textstyle\sum_z p(Z|X,\theta_n) \\ &= \textstyle\sum_z p(Z|X,\theta_n) \log p\ (X|\theta_n) \end{aligned}$$

$$= \sum_z p(Z|X,\theta_n) \log \frac{p(Z,X|\theta_n)}{p(Z|X,\theta_n)}$$

$$= \sum_z p(Z|X,\theta_n) \log\ p(Z,X|\theta_n) - \sum_z p(Z|X,\theta_n) \log\ p(Z|X,\theta_n)$$

On obtiendra donc un résultat sous forme d'espérance tel que :

$$L(X|\theta) = E\{ \log\ p(Z,X|\theta_n)|X,\theta_n \} - \mathrm{E}\{ \log\ p(Z|X,\theta_n) \mid X,\theta_n \} \qquad \text{[C-2]}$$

A l'itération *n*, nous disposons d'une valeur θ_n du vecteur de paramètres. Le but est de mettre à jour avec une meilleure valeur , augmentant la vraisemblance, donc telle que :

$$\Delta(\theta,\theta_n) = \log\ p(X|\theta) - \log\ p(X|\theta_n) \geq 0$$

On souhaite évidement que cette différence soit la plus grande possible.

Un moyen de maximiser cette différence consiste à chercher une fonction qu'on est capable de traiter et qui n'est autre que :

$$\begin{cases} \Delta(\theta,\theta_n) \geq \delta(\theta|\theta_n) \\ \delta(\theta|\theta_n) = 0 \end{cases}$$

Ainsi, $\delta(\theta|\theta_n)$ borne inférieurement $\Delta(\theta,\theta_n)$, et son maximum est au moins égal à 0. Il faut donc trouver un θ' capable de maximiser cette borne afin d'obtenir $\Delta(\theta',\theta_n)$.

C'est dans ce but que nous en somme venu à l'utilisation marginale des données cachées Z tel que :

$$p(X|\theta) = \sum_z p(X|Z,\theta)\, p(Z|\theta) \qquad \text{[C-3]}$$

Il vient alors :

$$\Delta(\theta,\theta_n) = \log p(X|\theta) - \log p(X|\theta_n)$$

$$\Delta(\theta,\theta_n) = \log\left(\sum_z p(X|Z,\theta)p(Z|\theta)\right) - \sum_z p(Z|X,\theta_n) \qquad \text{[C-4]}$$

Cette expression utilise les logarithme d'une somme : en se souvenant de l'inégalité de Jensen, on commence à voir apparaître clairement une façon de minorer $\Delta(\theta,\theta_n)$.

On aura alors :

$$\Delta(\theta,\theta_n) = \log\left(\sum_z \frac{p(X|Z,\theta)p(Z|\theta)}{p(Z|X,\theta_n)}\ p(Z|X,\theta_n)\right) - \sum_z p(Z|X,\theta_n) \log p(X|\theta_n)$$

Et en remarquant que $\sum_z p(Z|X,\theta_n) = 1$, nous appliquons l'inégalité de Jensen :

$$\Delta(\theta,\theta_n) \geq \sum_z p(Z|X,\theta_n)\log\frac{p(X|Z,\theta)p(Z|\theta)}{p(Z|X,\theta_n)} - \sum_z p(Z|X,\theta_n)\log p(X|\theta_n)$$

$$\Delta(\theta,\theta_n) = \sum_z p(Z|X,\theta_n)\log\frac{p(X|Z,\theta)p(Z|\theta)}{p(Z|X,\theta_n)p(X|\theta_n)}$$

$$\Delta(\theta,\theta_n) = \sum_z p(Z|X,\theta_n)\log\frac{p(X|Z,\theta)}{p(X,Z|\theta_n)} \qquad \text{[C-5]}$$

$$\Delta(\theta,\theta_n) = \delta(\theta|\theta_n)$$

Nous avons maintenant obtenu une fonction $\delta(\theta|\theta_n)$ vérifiant les conditions précédentes, finalement nous poserons que :

$$\theta_{n+1} = \arg\max \delta(\theta|\theta_n)$$

$$\theta_{n+1} = \arg\max\{\sum_z p(Z|X,\theta_n)\log\frac{p(X|Z,\theta)}{p(X,Z|\theta_n)}\}$$

$$\theta_{n+1} = \arg\max\{\sum_z p(Z|X,\theta_n)\log p(X,Z|\theta)\}$$

$$\theta_{n+1} = \arg\max\{E_{Z|X,\theta_n}[\log p(X,Z|\theta)]\} \qquad \text{[C-6]}$$

On détermine ainsi une valeur θ_{n+1} plus vraisemblable que θ_n puisque :

$$\log p(X|\theta_{n+1}) - \log p(X|\theta_n) = \Delta(\theta_{n+1},\theta_n) \geq \delta(\theta_{n+1}|\theta_n) \geq \delta(\theta_n|\theta_n) \geq 0$$

La mécanique itérative de cet algorithme est très astucieuse et débouche sur une amélioration progressive réciproque des données cachées Z et de la valeur du vecteur de paramètre .

En effet, on démarre l'algorithme avec une ignorance absolue des données cachées Z et en initialisant θ à une valeur θ_0 totalement arbitraire, potentiellement très loin de la réalité. L'algorithme se sert de θ_0 pour estimer Z, puis se sert de $\hat{Z}$ pour réestimer les paramètres en une valeur θ_1 plus pertinente.

A l'itération suivante, on améliore donc l'estimation des données cachées Z puisque cette nouvelle estimation se base cette fois sur θ_1 et cette nouvelle précision sur $\hat{Z}$ conduit à son tour une meilleure précision θ_2 etc.

Au final l'algorithme EM fournit donc non seulement une estimation de plus en plus pertinente de θ, mais aussi une estimation de plus en plus pertinente de Z. Si l'algorithme est couramment utilisé pour l'estimation paramétrique, rien n'empêche de le considérer dualement comme une façon d'estimer les données cachées, si tel est notre but. Une autre utilisation de l'algorithme EM peut donc être la complétion de données manquantes.

Il faut noter que, dans certains cas, l'algorithme peut ne converger que vers un point-selle ou un maximum local de la vraisemblance, si toute fois elle en possède un. La dépendance en la condition initiale θ_0 choisie arbitrairement est forte : pour certaines mauvaises valeurs, l'algorithme peut rester gelé en un point selle, alors qu'il convergera vers le maximum global pour d'autres valeurs initiales plus pertinentes.
L'algorithme EM peut donc parfois nécessiter plusieurs initialisations différentes.

C-3 L'algorithme EM appliqué aux GMM :

Afin de bien illustrer les étapes de l'algorithme EM, nous allons l'appliquer à un modèle de mélange de gaussiennes (GMM).
Soit un modèle de mélange de gaussiennes :

$$\theta = \{\pi_j, \mu_j, \Sigma_j; j = 1 \dots J\}$$

Avec :

π_j : Poids de la $j^{ème}$ densité du modèle.

μ_j : Vecteur moyen de la $j^{ème}$ densité du modèle.

Σ_j : Matrice de covariance de la $j^{ème}$ densité du modèle.

Comme vu précédemment, le mélange de gaussienne est donné par :

$$p\left(x_t \middle| \delta_t^{(j)} = 1, \Phi^{(j)}\right) = \sum_{j=1}^{J} \pi_j (2\pi)^{-D} \left|\Sigma_j\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_t - \mu_j)^T (\Sigma_j)^{-1} (x_t - \mu_j)\right\}$$

[C-7]

Après l'initialisation de θ_0 les étapes de l'algorithme EM sont les suivantes :

- L'étape E :

A l'itération « n », nous calculons p_n^j tel que :

$$p_n^j(x_t|\theta) = \frac{\pi_j(2\pi)^{-D}|\Sigma_j|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x_t-\mu_j)^T(\Sigma_j)^{-1}(x_t-\mu_j)\}}{\sum_{j'=1}^{J}\pi_{j'}(2\pi)^{-D}|\Sigma_{j'}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x_t-\mu_{j'})^T(\Sigma_{j'})^{-1}(x_t-\mu_{j'})\}} \quad [C\text{-}8]$$

Donc pour chaque vecteur x_t nous calculons la probabilité qu'il soit généré par la gaussienne « j ».

- L'étape M :

Dans ce cas nous procédons à la réestimations, ce qui donne :

$$\pi_j{}^* = \frac{1}{T}\sum_{t=1}^{T} p_n^j(x_t|\theta) \quad [C\text{-}9]$$

$$\mu_j{}^* = \frac{\sum_{t=1}^{T} p_n^j(x_t|\theta)\ x_t}{\sum_{t=1}^{T} p_n^j(x_t|\theta)} \quad [C\text{-}10]$$

$$\Sigma_j{}^* = \frac{\sum_{t=1}^{T} p_n^j(x_t|\theta)\ (x_t-\mu_j{}^*)(x_t-\mu_j{}^*)^T}{\sum_{t=1}^{T} p_n^j(x_t|\theta)}$$

Cette itération continue donc jusqu'à atteindre le seuil de convergence.

Annexe D
Théorie de la décision Bayèsienne

D-1 Théorie de la décision de Bayès :

La théorie de la décision se base sur le schéma suivant :

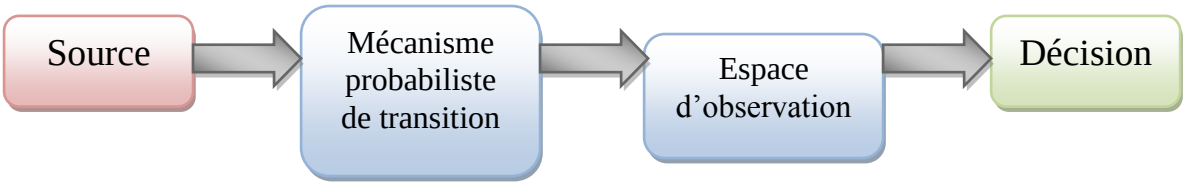


Figure D.1- Composants de la théorie de décision

- **La source** : il génère un objet en sortie. Dans le cas général, la sortie peut être une hypothèse parmi M hypothèses. Dans notre cas, cet objet ne sera pas défini que sous deux types (M=2). Nous les désignons comme étant des hypothèses et les notons H_0 et H_1.
- **Le mécanisme probabiliste de la transition** : il peut être vu comme un 'dispositif' qui, connaissant l'hypothèse 'vrai' générée par la source (parmi les deux hypothèses), génère un point dans l'espace d'observation suivant une certaine loi de probabilité.
- **L'espace d'observation** : ce dernier est à dimensions finies. En d'autres termes, les observations sont représentées par un ensemble de N valeurs qui peut être représenté par un point dans un espace à N dimensions.
- **La règle de décision** : associe chaque point d'observation à une des deux hypothèses de sortie.

D-2 Tests d'hypothèses binaires selon le critère de décision Bayésien :

Pour commencer, considérons un problème de décision où chaque sortie de la source correspond à une hypothèse. Chaque ensemble d'observation sera donc considéré comme un point à N dimensions dans l'espace d'observation et qui peut être noté r.

Le mécanisme probabiliste de transition génère des points suivant les deux densités de probabilités conditionnelles connues: pr/ H_1 (R/ H_1) et pr/ H_0 (R/ H_0). Le but étant d'utiliser cette information pour développer une règle appropriée de décision.

Dans le test d'hypothèses binaires, l'une ou l'autre des hypothèses H_0 et H_1 est exclusivement vraie ce qui impose à la règle de décision de faire un choix.

A chaque expérience l'une de ces quatre situations sera donc produite :

- H_0 vrai → H_0 choisie.
- H_0 vrai → H_1 choisie.
- H_1 vrai → H_1 choisie.
- H_1 vrai → H_0 choisie.

La première et la troisième alternative correspondent à des choix corrects. La deuxième et la quatrième alternatives quand à elles correspondent à des erreurs. La considération relative de ces alternatives est l'une des bases des critères Bayes.

Le critère de Bayes pour le problème de décision est basé sur deux suppositions. La première est que les sorties engendrées par la source suivent les probabilités P_0 et R, appelées les probabilités a priori. Ces probabilités représentent l'information observée sur la source avant la conduite de l'expérience. La deuxième supposition est l'association d'un coût pour chacune des possibilités, notés respectivement C_{00}, C_{10}, C_{11}, C_{01}. Le premier indice indique l'hypothèse choisie, le deuxième indique celle qui est vraie. Chaque fois que l'expérience est menée, un certain coût doit être 'dépensé'. Nous devons donc élaborer une règle de décision de telle sorte à ce que le coût moyen soit minimum.

Voici la formule du coût noté ici $\Re$ pour représenter le risque :

$$\Re = C_{00}P_0P_r\,(\text{choisir } H_0|H_0 \text{ est vrai}) + C_{10}P_0P_r\,(\text{choisir } H_1|H_0 \text{ est vrai}) + C_{11}P_1P_r\,(\text{choisir } H_1|H_1 \text{ est vrai}) + C_{01}P_0P_r\,(\text{choisir } H_0|H_1 \text{ est vrai})$$

Du fait que la règle de décision doit choisir H_0 ou H_1, l'espace d'observation Z se trouve divisé en deux parties, Z_0 et Z_1.

L'expression du coût total, en fonction des régions de décision et des probabilités de transition, devient alors :

$$\Re = C_{00}P_0\int P_{r/H_0}\,(R|H_0)\,dR$$

$$+C_{10}P_0\int P_{r/H_0}\,(R|H_0)\,dR$$

$$+C_{11}P_1\int P_{r/H_1}\,(R|H_1)\,dR$$

$+C_{01}P_0\int P_{r/H_1}(R|H_1)\,dR$

Dans un espace d'observation à N dimensions, les intégrales dans l'expression précédente sont à N dimensions. Nous supposons que le coût d'une mauvaise décision est plus élevé que celui d'une bonne décision. En d'autres termes :

$C_{10} < C_{00}$ et $C_{01} < C_{11}$

Pour trouver le bon choix selon Bayes il faut choisir les régions Z_0et Z_1 d'une telle façon que le risque (coût) soit minimisé. Autrement dit, nous devons assigner exclusivement chaque point R dans l'espace d'observation à une ou deux régions Z_0ou Z_1. D'où Z=Z_0+ Z_1.

En considérant le fait que :

$\int P_{r/H_0}(R|H_0)\,dR = \int P_{r/H_1}(R|H_1)\,dR$

On obtient :

$$\Re = C_{10}P_0 + C_{11}P_1 + \int \{[P_1(C_{01} - C_{11})P_{r/H_1}(R|H_1)] - [P_0(C_{10} - C_{00})P_{r/H_0}(R|H_0)]\}dR$$

Les deux premiers termes représentent la partie fixe du coût. L'intégrale représente la partie du coût contrôlée par ces points R 'rangés' dans Z_0.

La supposition $C_{10} < C_{00}$ ainsi que $< C_{11}$ implique que les deux termes entre parenthèses dans l'équation précédente sont positifs, notons ces termes respectivement par A et B.

Ainsi toutes les valeurs de R doivent être inclues dans Z_0 quand le seconds terme B est supérieur au premier terme A, parce qu'elles ajoutent une quantité négative à l'intégrale (diminuer le coût). Identiquement, toutes les valeurs de R, quand le premier terme A est supérieur au second terme B, doivent être exclues de Z_0 (inclues dans Z_0), parce qu'elles ajoutent une quantité positive à l'intégrale donc au coût total. Par conséquent, les régions de décision sont définies comme suit :

Si : $P_1(C_{01} - C_{11})P_{r/H_1}(R|H_1) > P_0(C_{10} - C_{00})P_{r/H_0}(R|H_0)$

Alors : Ranger R dans Z1 et ainsi dire H1 est vrai, sinon : Ranger R dans Z0 et ainsi dire que H0 est vrai. Ceci peut être réécris sous la forme suivante :

$$\frac{P_{r/H_1}(R|H_1)}{P_{r/H_0}(R|H_0)} \underset{<H_0}{\overset{>H_1}{}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

- La quantité à gauche de l'inégalité est appelée ***le rapport de vraisemblance*** et notée par Λ(R).
- La quantité à droite de l'inégalité est appelée ***seuil du test*** et est notée par η.

En conséquence, le critère de Bayes nous conduit au *Test de rapport de Vraisemblance* ***LTR*** (pour Likelihood Ratio Test) :

$$\Lambda(R) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Il est à noter que le traitement des données qui utilisé pour le calcul de $\Lambda(R)$ est indépendant des probabilités a priori et les coûts assignés. Cette invariance des traitements est d'une importance pratique et considérable, car les coûts et les probabilités a priori sont initiés par des connaissances a priori (utilisateurs du système basé sur ce test) ou simplement par des suppositions.

Le résultat de l'inégalité nous permet de construire un processus entier de décision, et cela en considérant η comme un seuil variable et paramétrable, qui peut être adapté selon les coûts, les probabilités a priori et notre estimation des probabilités de transitions.

Enfin, vu que la fonction logarithmique est monotone et que $\Lambda(R)$ et η sont positifs (par construction), un test équivalent est :

$$\ln(\Lambda(R)) \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\eta)$$

Annexe E
Caractéristiques des fenêtres d'analyse courantes en traitement du signal

En traitement du signal, le fenêtrage est utilisé dès que l'on s'intéresse à un signal de longueur volontairement limité. En effet, un signal réel ne peut qu'avoir une durée limitée dans le temps ; de plus, un calcul ne peut se faire que sur un nombre de points fini. Pour observer un signal sur une durée finie, on le multiplie par une fonction fenêtre d'observation (également appelée fenêtre de pondération) [Tis 08].

Voici les principales caractéristiques des fenêtres d'analyse les plus courantes :

Type de fenêtre	**Équations (N échantillons)**	**Lobes secondaires**		**Lobe principal**	
		Niveau dB	Pente dB/oct	Largeur à –3 dB	Largeur à –6 dB
Rectangulaire	$\omega_k = 1$ $k = 0, \dots, N-1$ $\omega_k = 0 \;\; ailleurs$	-13	-6	0.89/N	1.21/N
Hamming	$\omega_k = 0.54 - 0.46\, cos\left(\frac{2\pi k}{N}\right)$ $k = 0, \dots, N-1$ $\omega_k = 0 \;\; ailleurs$	-43	-6	1.30/N	1.81/N
Hanning (Hann)	$\omega_k = 0.50 - 0.50\, cos\left(\frac{2\pi k}{N}\right)$ $k = 0, \dots, N-1$ $\omega_k = 0 \;\; ailleurs$	-32	-18	1.44/N	2.00/N
Blackman-Harris	$\omega_k = 0.42 - 0.50\, cos\left(\frac{2\pi k}{N}\right) + 0.08\, cos\left(\frac{4}{}\right.$ $k = 0, \dots, N-1$ $\omega_k = 0 \;\; ailleurs$	-58	-18	1.68/N	2.35/N

Tableau 3.1- Caractéristiques des fenêtres d'analyse courantes.

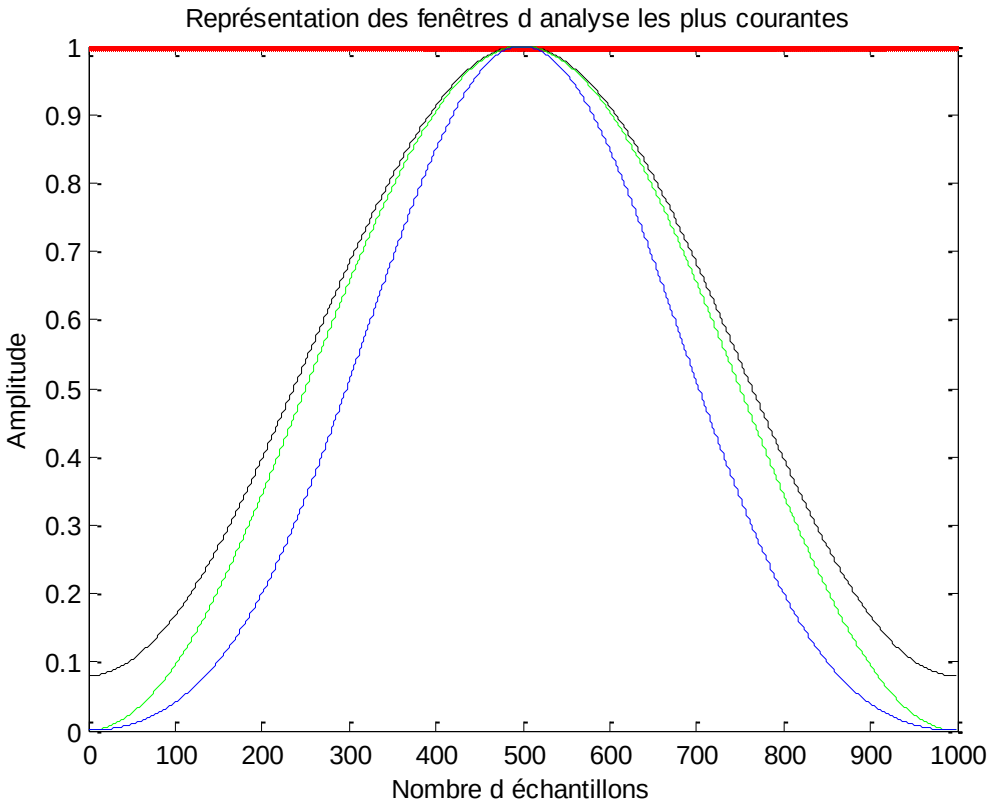


Figure E- Représentation temporelle des fenêtres les plus courantes : Rectangulaire (rouge), Hamming (noir), Hanning (vert), Blackman (bleu).