République Algérienne Démocratique et Populaire
Ministère de L'enseignement Supérieur et de la Recherche Scientifique
Université 8 Mai 1945 - Guelma
Faculté des Sciences et de la Technologie
Département Electronique et Télécommunications

709

Mémoire de fin d'étude
Pour l'obtention du diplôme de Master Académique

Domaine : **Sciences et Techniques**
Filière : **Electronique**
Spécialité : **Systèmes Electroniques**

IDENTIFICATION DU LOCUTEUR EN MODE INDEPENDANT DU TEXTE Par la Méthode SVM

Présenté par :
CHETIBI Abderaouf
MESSAAD Touhami

Sous la direction de : Dr BOUROUBA Houcine

JUIN 2011

Remerciements

En préambule à ce mémoire, nous souhaitions adresser
Nos remerciements les plis sincères aux personnes qui nous
Ont apporté leur aide et qui ont contribué à l'élaboration de
Ce mémoire ainsi qu'à la réussite de cette formidable année
Universitaire.

Nous tenons à remercier sincèrement De BOUROUBA houcine
qui, étant l'encadreur de mémoire, s'est toujours montrée à
l'écoute et très disponible tout au long de la réalisation de ce
travail, ainsi pour l'inspiration, l'aide et le tems qu'elle a
bien voulu nous consacrer et sans elle ce mémoire n'aurait
jamais vu le jour.

Nous exprimons notre gratitude à tous les consultants et
internautes rencontrés lors des recherches effectuées et qui
Ont accepté de répondre à nos questions avec gentillesse
Nous n'oublions par nos parents pour leur contribution, leur
Soutien et leur patience.

Nos sincères remerciements vont à tous les membres du
jury de soutenance pour avoir accepté et pris le temps de
juger ce travail.

Enfin, nous adressons nos plus sincères remerciements à
Tous nos proches et amis, qui nous ont toujours soutenus et
encouragés au cours de la réalisation de ce mémoire
Merci à tous et à toutes.

CHETIBI.A et MESSAAD.T

ملخص :

يهدف هذا المشروع إلى دراسة إمكانية استعمال الآلات ذات الأشعة الحاملة لتصنيف يسمح بالتعرف على المتحدث (شخص مرخص أو شخص دخيل) حيث استعملت هذه التقنية في بداية الأمر للفصل بين المتحدث المرخص و المتحدث الدخيل وهو ما صغناه كمشكل تصنيف ذا مجموعتين, أما التعرف على هوية المتحدث بعد الجزم بأنه من الأشخاص المرخص لهم يتم عن طريق النموذج الحركي للتوافق الزمني (DTW).

كلمات مفتاحية : التعرف على المتحدث, الآلات ذات الأشعة الحاملة (SVM), التصنيف ,DTW , التحليل الصوتي.

Résumé :

Dans ce travail, nous avons étudié les SVM comme technique de classification pour la reconnaissance automatique du locuteur (RAL) en mode dépendant du texte. Nous avons testé cette technique pour discriminer entre locuteurs autorisés et imposteurs en faisant dans un premier temps une classification binaire entre individus. Par la suite, la méthode DTW a été utilisée pour identifier l'individu client.

Les études que nous avons réalisées constituent une des premières tentatives d'appliquer les SVM dans le domaine de la RAL. Les teste ont été faits en faisant varier plusieurs paramètres pour choisir la meilleure configuration.

Mots clés : Reconnaissance automatique du locuteur, SVM, Classification, DTW, analyse acoustique.

Abstract :

In this work, we study the feasibility of the SVM as classification technique for automatic speaker recognition (ASR) in text dependent mode. We firstly used this technique to discriminate between allowed speaker and impostors. We considered this problem as a binary classification. Secondly, we applied the dynamic time wrapping (DTW) method to decide about the identity of the client speaker.

Our study constitutes one of the first attempts to apply the SVM to the ASR field.

Key words : Speaker recognition, support vector machines (SVM), classification, DTW, acoustic analysis.

Liste des Abréviations

FFT : Fast Fourier Transform.

IAL : Identification Automatique du Locuteur.

VAL : Vérification Automatique du Locuteur.

RAL : Reconnaissance Automatique du Locuteur.

DTW : Dynamic Time Warping.

VQ : Vector Quantisation.

AAL : Authentification Automatique du Locuteur.

LPCC : Linear Predictive Cepstral Coefficients.

LPC : Linear Predictive Coefficients.

LFSC : Linear Frequency Spectral Coefficients.

MFSC : Mel Frequency Spectral Coefficients.

LFCC : Linear Frequency Cepstral Coefficients.

MFCC : Mel Frequency Cepstral Coefficients.

GMM : Gaussien Mixture Model.

EM : Estimation Maximisation.

RN : Réseaux de Neurones.

HMM : Hidden Markov Models.

MLP : Multi Layer Perceptrons.

RBF : Radial Basis Fonctions.

SVM : Support Vector Machine.

HO : Hyperplan Optimal.

M : Marge.

Listes des Figures

Liste des Tableaux

Sommaire

Introduction :

La parole est le moyen privilégié de communication de l'Homme. Le problème de la reconnaissance automatique de la parole consiste à extraire l'information lexicale contenue dans un signal de parole et éventuellement de l'interpréter. Depuis plus de quatre décennies, de nombreux laboratoires internationaux ont mené des recherches intensives dans ce domaine et des progrès importants ont été réalisés, notamment grâce au développement d'algorithmes puissants alliés aux technologies de traitement numérique du signal. Parallèlement à la RAP, les chercheurs se sont penchés sur le problème de la caractérisation du locuteur à l'aide de sa voix et, en particulier, de la Reconnaissance Automatique du Locuteur (RAL). L'expression vocale est une caractéristique propre d'un locuteur, ainsi est-il possible, dans des conditions normales de reconnaitre une personne à partir de sa voix.

La reconnaissance automatique du locuteur est interprétée comme une tâche particulière de reconnaissance de formes. Ce domaine regroupe les problèmes relatifs à l'identification ou à la vérification du locuteur sur base de l'information contenue dans le signal acoustique : Il s'agit de reconnaître une personne à partir de sa voix. Le champ d'application est très large, allant des applications domestiques aux applications militaires, en passant par des applications judiciaires.

Un système de reconnaissance automatique du locuteur est constitué généralement de trois modules un module pour l'extraction des coefficients acoustiques, un autre module pour la modélisation des locuteurs et enfin un module de classification et de décision.

Au cours de ce projet de fin d'étude, qui consiste à l'identification du locuteur en mode indépendant du texte, Nous nous intéressons essentiellement à l'information extralinguistique contenue dans le signal vocal. Pour extraire du signal vocal l'information relative à l'identité du locuteur, on utilise les coefficients cepstraux qui permettent une bonne séparation de la contribution du conduit vocal et celle de la source d'excitation glottique.

Pour la modélisation des locuteurs, plusieurs approches existent approche vectorielle, connexionniste, statistique et prédictives. Dans notre projet de fin étude nous avons basé sur la méthode de classification les plus utilisées, dans le domaine de reconnaissance du locuteur. En particulier, la méthode des machines à vecteur de support (SVM). Cette technique d'apprentissage statistique est relativement récente et est due à V. Vapink en 1995. Elle permet d'aborder des problèmes de diverses natures comme le classement qui une tâche de discrimination entre classes, la régression et la fusion. Dans notre contexte, le problème de RAL peut être considère comme un problème de classification.

Ce document s'articule autour de trois chapitres. Le premier chapitre constitue une introduction aux systèmes de reconnaissance automatique du locuteur.

Le deuxième chapitre expose l'approche de modélisation des locuteurs par la méthode SVM et leur détaille.

Le troisième chapitre décrit le contexte expérimental et expose les résultats des différents tests effectués. Pour cette dernière section, on a essayé d'examiner et de voir l'influence d'un certain nombre de paramètres (le paramètre de régularisation et le paramètre de noyau RBF Gamma, et la durée de segment test) sur le taux d'identification correcte et sélectionner par la suite l'ensemble des paramètres qui donne les meilleures performances, pour une éventuelle conception d'un système d'identification du locuteur.

Finalement, une conclusion générale conclue ce mémoire.

Chapitre I :

Système de Reconnaissance Automatique du Lecteur

I.1 Introduction :

La reconnaissance automatique du locuteur RAL, contrairement à la reconnaissance automatique de la parole RAP, s'intéresse tout particulièrement aux informations extra-linguistiques véhiculées par un signal vocal (signal de parole). Pourtant, la RAL a très souvent bénéficié des avancées de la RAP. Ainsi, de nombreuses techniques ont été appliquées en RAP avant d'être adaptées au domaine de la RAL. Finalement, les applications de la RAL sont principalement liées aux problèmes d'authentification ou de confidentialité.

Un système de reconnaissance automatique du locuteur est un système qui permet de décider à partir d'un signal de parole, appelé segment de test, sur l'identité du locuteur.

I.2 Caractéristiques du signal acoustique de la parole :

Le signal acoustique de la parole est un peu particulier, il présente des caractéristiques qui rendent l'interprétation très complexe. En effet, ce signal est très redondant (il véhicule beaucoup d'informations ce qui par ailleurs le rend très résistant aux bruits) il est aussi variable d'un locuteur à un autre (variabilité inter-locuteur), et pour le même locuteur (variabilité intra-locuteur). Nous ajoutons les variabilités dues aux conditions d'enregistrement et de l'environnement.

I.2.1 Variabilité intra-locuteur :

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

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

I.2.2 Variabilité inter-locuteur :

La variabilité inter-locuteur est un phénomène majeur en reconnaissance de la parole et du locuteur. Mais cela ne nous empêche pas de rappeler qu'un locuteur reste identifiable par le timbre de sa voix, malgré une variabilité qui peut être parfois importante.

La cause principale des différences interlocuteurs est de nature physiologique. La parole est produite par les vibrations des cordes vocales, qui déterminent l'importance et la forme du flux d'air s'échappant des poumons et elles sont amplifiées par les organes respiratoires, cette opération génère un son à une fréquence de base, le fondamental. Cette fréquence de base sera différente d'un individu à l'autre et plus généralement d'un genre à l'autre, une voix d'homme étant plus grave qu'une voix de femme, la fréquence du fondamental étant plus faible. Ce son est ensuite transformé par l'intermédiaire du conduit vocal, délimité à ses extrémités par le larynx et les lèvres. Cette transformation, par convolution, permet de générer des sons différents. Or le conduit vocal est de forme et de longueur variables selon les individus et, plus généralement, selon le genre et l'âge. Ainsi, le conduit vocal féminin adulte est, en moyenne, d'une longueur inférieure de 15% à celui d'un conduit vocal masculin adulte. Le conduit vocal d'un enfant en bas âge est bien sûr inférieur en longueur à celui d'un adulte.
Les convolutions possibles seront donc différentes et, le fondamental n'étant pas constant, un même phonème pourra avoir des réalisations acoustiques très différentes.

La variabilité interlocuteur peut se manifester encore dans les différences de prononciation qui existent au sein d'une même langue et qui constituent les accents régionaux.

I.2.3 Variabilité due à l'environnement :

Cette variabilité est due soit à un bruit qui peut provoquer une dégradation du signal parole sans que le locuteur ait modifié son mode d'élocution, soit à une déformation physiologique (comme le conduit vocale), dans ce dernier cas nous pouvons considérer cette variabilité comme étant une variabilité intra-locuteur.

Ainsi, un système mécanique provoquant une déformation du conduit vocal provoquera certainement une variation dans le signal de parole produit.

I.2.4 Variabilité due aux conditions d'enregistrement :

La transmission de la parole par un canal téléphonique entraîne une limitation dans la gamme de fréquence, de 300 Hz à 3400 Hz de la bande passante. Les spectres fournis par les lignes téléphoniques sont donc limités par la bande passante et également multipliés par une fonction de transfert de forme inconnue. Dans un premier stade, les études ont montré que la limitation des spectres de longue durée à la bande passante caractérisant la qualité du téléphone, n'affecte pas sensiblement le taux d'identification. Cependant, la pondération des spectres par des fonctions de transfert, détruit la fiabilité de l'identification parce que, dans certains cas, l'effet de la fonction de transfert sur les spectres est plus important que les caractéristiques des voix.

I.3 Introduction à la reconnaissance automatique du locuteur :

Comme l'illustre la figure I.1, la reconnaissance automatique du locuteur s'inscrit dans le domaine plus général du traitement de la parole. Elle exploite la variabilité inter-locuteurs et s'intéresse aux informations extralinguistiques du signal vocal.

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

Les mouvements des organes de production de la parole engendrent des variations de pression acoustique instantanée qui peuvent être captées par un transducteur (microphone) et transformées en variations de tension électrique.

Un enregistrement de la parole n'est ni un prélèvement direct ni une trace laissée sur une surface au contact d'une partie de son corps, il ne s'agit que de la capture indirecte de mouvements articulatoires complexes faisant intervenir les cordes vocales, la langue, le voile du palais, la mâchoire et les lèvres.

La reconnaissance vocale est considérée comme une des formes les moins intrusives de la technologie biométrique, car elle n'exige aucun contact physique avec le capteur (microphone) du système automatique de reconnaissance.

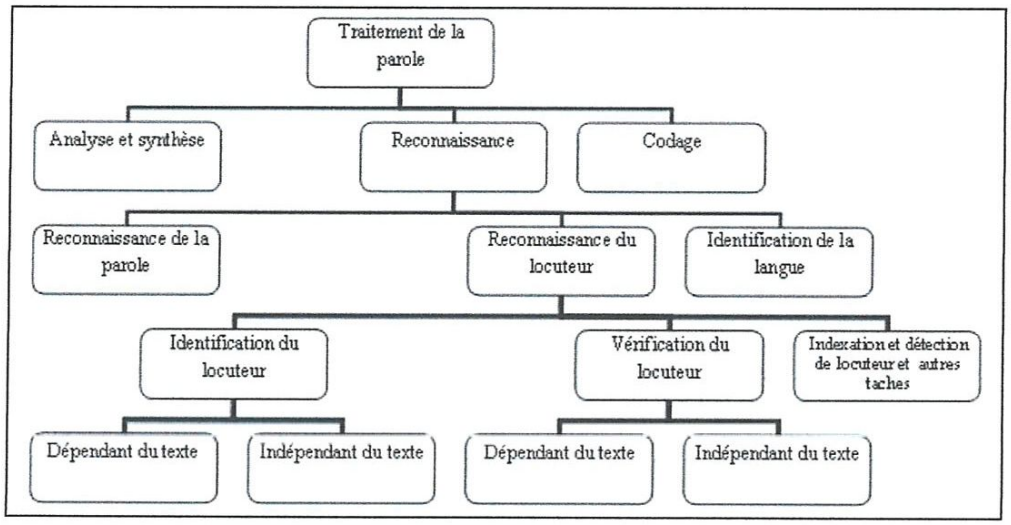

Figure I.1 : Traitement de la parole

I.4 Systèmes de reconnaissance automatique du locuteur :

La reconnaissance automatique consiste à extraire l'information contenue dans le signal acoustique de la parole et éventuellement de l'interpréter pour connaître automatiquement l'identité d'une personne prononçant une ou plusieurs phrases à l'aide d'un ordinateur qui joue aujourd'hui un grand rôle dans ce domaine. Les applications directes de la RAL concernent les problèmes de confidentialité et d'authentification. Nous distinguerons :

- Les applications (sur site) : Serrures vocales pour contrôle d'accès, cabines bancaires en libre service.
- Les applications liées aux télécommunications : Ces applications concernent l'identification du locuteur à travers le réseau téléphonique pour accéder à un service de transactions bancaires à distance ou pour interroger des bases de données en accès privé.
- Les applications judiciaires (forensic applications) : Recherche de suspects, orientations d'enquêtes.

La difficulté de la tâche de reconnaissance n'est pas la même d'une application à l'autre. Dans le cas des applications (sur site) l'environnement de prononciation de la phrase ou du mot de passe est plus facilement contrôlé que dans le cas des applications via le réseau téléphonique (distorsions dues au canal, différences entre les combinés téléphoniques, bande passante limitée). Les applications judiciaires présentent quand à elles des difficultés d'un autre ordre (locuteurs non coopératifs, enregistrements de mauvaise qualité).

Les systèmes de reconnaissance automatique du locuteur comportent plusieurs modules. Tout d'abord, un module d'acquisition qui capte le signal vocal et la convertie en un signal numérique. Ensuite vient le module d'analyse acoustique à l'issue duquel des vecteurs de coefficients pertinents, servants pour la modélisation du locuteur, sont extraits.

Dans l'étape d'apprentissage, un modèle est créé pour chaque locuteur. Dans l'étape de reconnaissance, un module de classification va mesurer la similarité entre les données de test et un ou tous les modèles de locuteurs présents dans la base. En dernier lieu, un module de décision, Basé sur une stratégie de décision donnée, fournit la réponse du système.

I.5 Les différentes tâches en RAL :

L'identification automatique du locuteur et la vérification automatique du locuteur sont les tâches pionnières du domaine de la RAL. Plus récemment, les besoins applicatifs ont fait naître de nouvelles tâches comme l'Indexation par Locuteur de flux audio ou le Suivi de locuteurs (ou speaker tracking) ou de nouvelles variantes telles que la détection d'un locuteur dans une conversation.

I.5.1 Identification automatique du locuteur (IAL) :

L'identification automatique du locuteur (IAL) est le processus qui consiste à déterminer, parmi une population de locuteurs connus, la personne ayant prononcé un message donné.

D'un point de vue schématique (figure I.2), une séquence de parole est donnée en entrée du système d'IAL. Pour chaque locuteur connu du système, la séquence de parole est comparée à une référence caractéristique du locuteur. L'identité du locuteur dont la référence est la plus proche de la séquence de parole est donnée en sortie du système d'IAL.

Deux modes sont proposés en IAL : l'identification en ensemble fermé pour laquelle on suppose que la séquence de parole est effectivement prononcée par un locuteur connu du système, dans ce cas il doit fournir en sortie un ensemble d'au moins un locuteur, et l'identification en ensemble ouvert pour laquelle le locuteur peut ne pas être connu et le système peut être amené à fournir un ensemble vide. En mode ensemble ouvert, le système d'IAL doit décider de la fiabilité de son jugement en acceptant ou rejetant l'identité qu'il a trouvée.

De par son principe déterminer une identité parmi les identités potentielles, les performances des systèmes d'IAL se dégradent généralement au fur et à mesure que la population de locuteurs augmente.

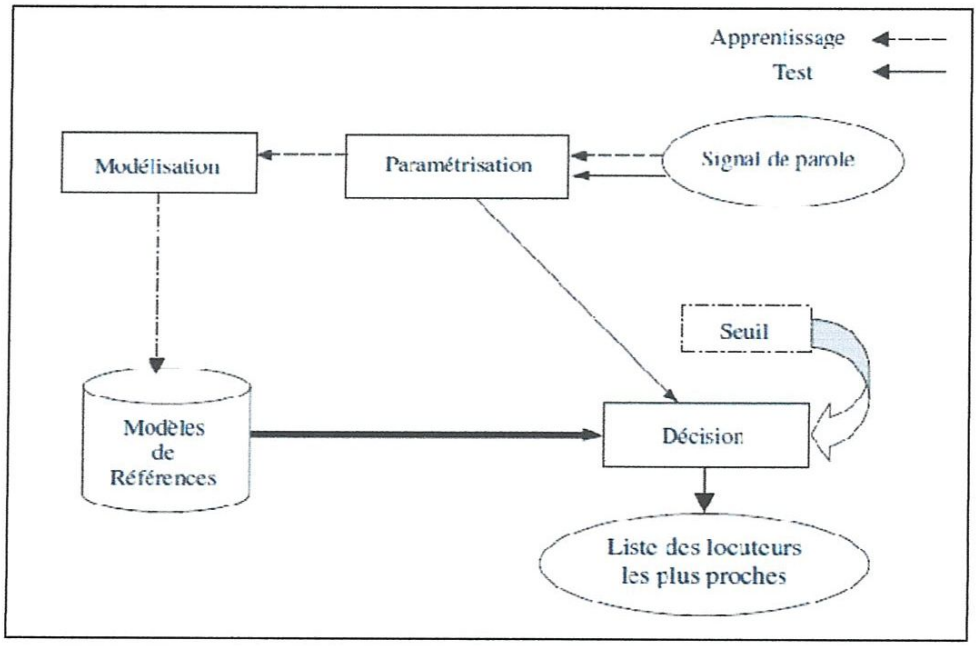

Figure I.2 : Système d'identification automatique du locuteur (IAL)

I.5.1.1 Applications :

En IAL, les applications sont peu nombreuses. On peut retenir, par exemple, l'utilisation d'un système d'IAL en vue de faciliter l'adaptation au locuteur des systèmes de RAP. Par ailleurs, il peut être intéressant pour des applications commerciales d'associer un même mot de passe pour une petite population de locuteurs (membres d'une famille, d'une société). Dans une telle situation, un système d'IAL en ensemble ouvert et dépendant du texte peut être utilisé pour contrôler l'accès a des données sensibles, a un réseau ou a un bâtiment.

Il est à noter que pour une identification en ensemble ouvert, la combinaison des deux tâches précédentes est nécessaire :

- Identification du locuteur le plus probable parmi les locuteurs de la base.
- Puis vérification que l'échantillon inconnu a bien été prononcé par le locuteur choisi dans l'étape d'identification.

I.5.2 Vérification automatique du locuteur (VAL) :

La vérification de l'identité est très importante dans la vie quotidienne. En effet, notre identité est vérifiée lorsque nous entrons dans notre lieu de travail, lorsque nous nous connectons au réseau informatique, lorsque nous exécutons des transactions bancaires, ... etc. Il y a deux manières classiques de vérifier l'identité d'un individu. L'une est basée sur une connaissance, par exemple un mot de passe, et l'autre est basée sur une possession, par exemple une pièce d'identité, une clé, un badge,... Par fois, ces deux manières sont utilisées en parallèle; c'est par exemple le cas d'une carte à puce avec un code confidentiel.

Pourtant, les possessions peuvent être volées ou perdues et les connaissances peuvent être oubliées. Ainsi, la biométrie représente une alternative à ces faiblesses, pour vérifier l'identité d'une personne. En effet, elle consiste à utiliser des caractéristiques physiques d'un individu comme son visage ou ses empreintes digitales, ou bien certaines caractéristiques comportementales comme sa signature manuscrite.

La biométrie est ainsi reliée à la personne, très difficile à mimer et ne peut jamais être perdue ou volée. Cependant, les caractéristiques biométriques dépendent beaucoup de l'environnement de capture, du stress de l'individu ou de son état général, ce qui gène le bon fonctionnement du système de vérification biométrique.

Un système de Vérification Automatique du Locuteur (VAL) doit vérifier à partir d'un signal de parole et d'une identité proclamée, qui appartient à la base de données, si le signal présenté provient de l'identité proclamée ou non. La figure I.3 représente un schéma modulaire d'un système de vérification du locuteur.

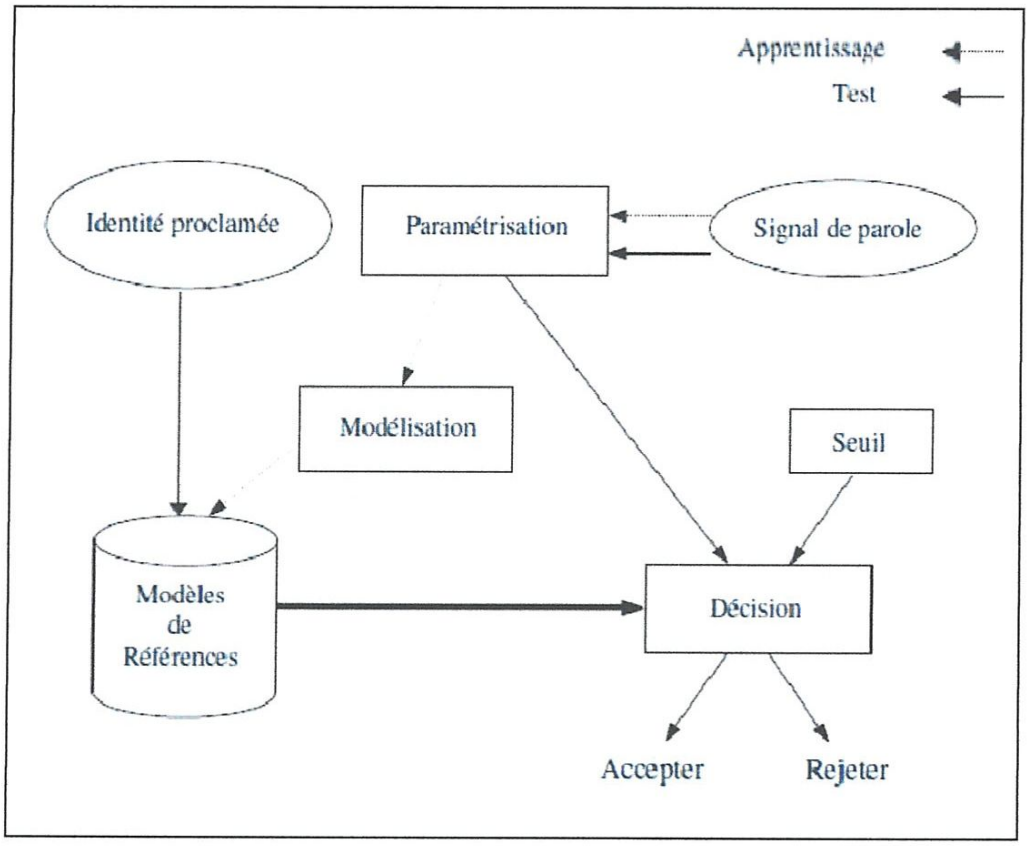

Figure I.3 : Système de vérification automatique du locuteur (VAL)

I.5.2.1 Applications :

Les applications de VAL sont multiples et principalement commerciales :

- Serrures vocales pour le contrôle d'accès à des locaux.
- Authentification pour l'accès à distance à des données sensibles ou à des services spécifiques à travers le réseau téléphonique (consultations ou transactions bancaires, consultations de bases de données à caractère confidentiel, consultations de boîtes vocales, etc.).
- Protection de matériel contre le vol (téléphones portables, voitures, etc.).

I.6 Modes dépendant et indépendant vis à vis du texte :

On peut classer les systèmes de reconnaissance automatique du locuteur en deux catégories, qui correspondent aux deux modes dépendant ou indépendant du texte. En mode dépendant du texte, le texte prononcé par le locuteur (pour être reconnu du système) est le même que celui qu'il a prononcé lors de l'apprentissage de sa voix.

En mode indépendant du texte, le locuteur peut prononcer n'importe quelle phrase pour être reconnu. Les niveaux de dépendance au texte sont classés suivant les applications :

- Systèmes à texte libre : Le locuteur est libre de prononcer ce qu'il veut, et les phrases d'apprentissage et de test sont différentes.
- Systèmes à texte suggéré : Un texte, différent à chaque session et pour chaque locuteur, est imposé par la machine. Les phrases d'apprentissage et de test peuvent être différentes.
- Systèmes dépendants du vocabulaire : Le locuteur prononce une séquence de mots issus d'un vocabulaire limité. Dans ce cas, l'apprentissage et le test sont réalisés sur des textes constitués à partir du même vocabulaire.
- Systèmes personnalisés dépendants du texte : Chaque locuteur a son propre mot de passe. Dans ce mode, l'apprentissage et le test sont réalisés sur le même texte.

I.7 Structure des systèmes de RAL et techniques associées :

Un système de RAL quelle que soit la tâche considérée, se résume à l'enchaînement de trois processus principaux que sont :

- La paramétrisation (analyse).
- La modélisation (classification).
- La décision.

Tout d'abord, le message vocal, capté par un microphone, est converti en signal numérique. Il est ensuite analysé dans un étage d'analyse acoustique. À l'issue de cette étape, le signal est représenté par des vecteurs de coefficients pertinents, ce qui permet de réduire l'information en quantité et en redondance pour la modélisation du locuteur. Dans l'étape d'apprentissage, on crée un modèle du locuteur. A la reconnaissance, un module de classification va mesurer la similarité entre les paramètres acoustiques du signal prononcé et les modèles de locuteurs présents dans la base. En dernier lieu, un module de décision, Basé sur une stratégie de décision donnée, fournit la réponse du système. On peut également introduire un module d'adaptation pour augmenter les performances du système de reconnaissance.

La structure d'un système d'identification du locuteur en ensemble fermé (qui constitue le cadre de notre travail) est représentée sur la figure I.4 :

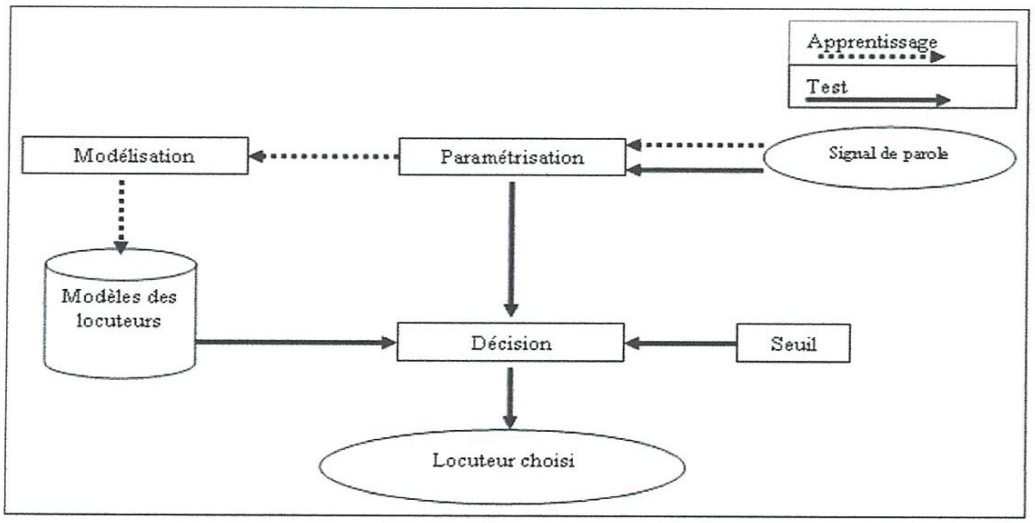

Figure I.4 : Schéma modulaire d'un système de VAL

I.7.1 Paramétrisation acoustique :

Le processus de paramétrisation consiste à extraire du signal de parole les informations pertinentes en vue de la reconnaissance. Le signal de parole, de par sa complexité (multitudes d'informations et redondance), ne peut être exploité directement. Une représentation simplifiée du signal de parole est par conséquent nécessaire. Cette représentation repose généralement sur des vecteurs de paramètres acoustiques, calculés périodiquement sur le signal de parole.

La première étape de la paramétrisation acoustique consiste à décomposer le signal de parole à cadence régulière, en trames de signal (d'une longueur variant généralement de 20 à 31ms). Un traitement particulier est ensuite appliqué à ces trames afin de produire les vecteurs de paramètres acoustiques.

La littérature propose un grand nombre de traitements selon la nature des informations à extraire du signal de parole. On considéré généralement trois grandes classes de paramètres les paramètres de l'analyse spectrale, les paramètres prosodiques et les paramètres dynamiques. Néanmoins, D'autres classifications sont envisageables.

I.7.1.1 Paramètres de l'analyse spectrale :

L'analyse spectrale est l'analyse la plus employée en RAL. Les paramètres qui en découlent sont généralement représentatifs des caractéristiques physiques de l'appareil phonatoire (forme du conduit vocal) de chaque individu.

De multiples paramètres ont été étudiés dans la littérature. Nous citons ici les plus pertinents en RAL :

- Coefficients issus d'une analyse par prédiction linéaire : LPCC (Linear Predictive Cepstral Coefficients) ou LPC (Linear Predictive Coefficients).
- Coefficients spectraux issus d'une analyse en banc de filtres : LFSC (Linear Frequency Spectral Coefficients) ou MFSC (Mel Frequency Spectral Coefficients).

- Coefficients cepstraux issus d'une analyse en banc de filtres : LFCC (Linear Frequency Cepstral Coefficients) ou MFCC (Mel Frequency Cepstral Coefficients).

I.7.1.2 Paramètres prosodiques :

Les paramètres prosodiques illustrent en grande partie le style d'élocution d'un locuteur :

- Vitesse d'élocution (débit).
- Durée.
- Fréquence des pauses.
- Fréquence fondamentale.
- Energie.
- Taux de voisement.

Néanmoins, ces paramètres caractéristiques du locuteur, notamment la fréquence fondamentale et ses variations, ne sont pas suffisamment discriminants pour être utilisés seuls dans un système de RAL. Ils sont généralement associés aux paramètres de l'analyse spectrale pour améliorer les performances des systèmes de RAL.

I.7.1.3 Paramètres dynamiques :

L'information dynamique véhiculée par le signal de parole est une source potentielle d'informations pour la caractérisation du locuteur, qui reste encore mal exploitée par les systèmes de RAL.

Les paramètres dynamiques les plus répandus demeurent les coefficients dérivés des vecteurs de paramètres instantanés, appelés coefficients Delta (première dérivée) et Delta-Delta (seconde dérivée).

I.7.2 Classification des vecteurs acoustiques :

Le processus de reconnaissance s'appuie généralement sur une modélisation des caractéristiques de chaque locuteur connu du système (modèles de locuteurs ou modèles clients). Cette modélisation est réalisée à partir des données d'apprentissage collectées au cours des sessions d'enrôlement.

Une mesure de similarité est ensuite calculée entre un modèle client et un signal de parole, puis transmise au processus de décision.

Dans le cadre de la reconnaissance vocale, nous modélisons les différentes prononciations qu'un locuteur peut avoir effectuées pour permettant de séparer les locuteurs les uns des autres (variabilités inter-locuteurs) et d'autres, Intrinsèques au locuteur (variabilités intra-locuteur) donc reconnaître un locuteur revient à essayer de le distinguer des autres.

Pour 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 du locuteur), tout en ayant la possibilité de s'adapter aux variations d'échelles fréquentielles et temporelles liées au signal de parole. Ces motifs peuvent être soit des segments de parole déterminés (mots, phonèmes) si nous travaillons en mode dépendant du texte, soit des segments de parole dont on ne connaît pas le contenu phonétique si l'application fonctionne en mode indépendant du texte. Ces algorithmes doivent être couplés avec une mesure qui permettra de donner une valeur de distance (ou de similitude) entre le modèle du locuteur et un motif inconnu dont on cherche à déterminer la provenance.

On peut distinguer quatre grandes approches pour la construction des modèles :

- Les approches vectorielles.
- Les approches statistiques.
- Les approches prédictives.
- Les approches connexionniste.

Nous présentons ici brièvement les fondements de chacune de ces approches, les techniques qui leur sont associées ainsi que les mesures de similarité utilisées.

I.7.2.1 L'approche vectorielle :

Dans l'approche vectorielle, un modèle de locuteur est un ensemble de vecteurs de paramètres représentatifs de l'espace acoustique construit lors de la phase de paramétrisation des signaux d'apprentissage. Lors de la reconnaissance, Une distance entre cet ensemble de vecteurs et les vecteurs de paramètres issus des signaux de test est calculée.

L'approche vectorielle compte deux grandes techniques, la programmation dynamique et la quantification vectorielle.

I.7.2.1.1 Programmation dynamique :

La programmation dynamique DTW (Dynamic Time Warping) consiste à aligner temporellement une séquence de vecteurs de paramètres de test avec une séquence de vecteurs d'apprentissage. Dans ce cas de figure, le modèle de locuteur est tout simplement l'ensemble des vecteurs de paramètres obtenus après paramétrisation des signaux d'apprentissage. Une distance est calculée entre vecteurs d'apprentissage et de test et moyenne sur l'ensemble de la séquence.

De par son principe, la programmation dynamique est utilisée exclusivement en mode dépendant du texte. Très rapide et montrant des performances relativement bonnes, la programmation dynamique est toutefois très sensible à la qualité d'alignement et notamment au choix du point de départ.

I.7.2.1.2 Quantification vectorielle :

La quantification vectorielle VQ (Vector Quantisation) repose sur un partitionnement de l'espace acoustique en sous-espaces. Chaque sous espace est associe a leur vecteur centroide i.e. à un vecteur de paramètres représentant l'ensemble des vecteurs composant le sous-espace. Dans ces conditions, un modèle de locuteur est composé d'un ensemble de vecteurs centroides, appelé dictionnaire de quantification (Code Book).

Lors de la phase de reconnaissance, une distance est calculée entre un vecteur de test et chaque vecteur centroide du dictionnaire. La distance minimale est assignée au vecteur de test, la distance d'une séquence de vecteurs de test est obtenue par moyenne des distances minimales attribuées à chacun des vecteurs de test.

I.7.2.2 L'approche statistique :

I.7.2.2.1 Modèles à mélange de distributions gaussiennes :

Le modèle de mélange de distributions gaussiennes GMM (Gaussien Mixture Model) consiste à supposer que la distribution des données peut être décrite comme une somme pondérée de densités gaussiennes multidimensionnelles.
Ce modèle de mélange est classique dans le domaine de la reconnaissance de forme car il correspond à une situation où les données appartiennent à un ensemble de classes distinctes, avec une probabilité d'appartenance propre à chaque classe. 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 est 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 (Estimation Maximisation) 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, le modèle à mélange de distributions gaussiennes permet d'obtenir de meilleures performances que celles des autres techniques comme quantification vectorielle).

Plusieurs points méritent d'être étudiés concernant ce modèle de mélange de densités gaussiennes.

Le premier concerne la structure des densités gaussiennes composant le mélange. Une simplification souvent utilisée consiste à considérer que les distributions gaussiennes composant le mélange possèdent toutes une matrice de covariance diagonale. Cette simplification est plus réaliste compte tenu de la difficulté posée par l'estimation complète des matrices de covariance.

De plus, l'algorithme EM est susceptible de fournir de multiples solutions. Le problème de l'initialisation de l'algorithme d'apprentissage est donc très important. Dans les applications de reconnaissance du locuteur, on trouve à la fois des méthodes d'initialisation très simples comme partition arbitraire et des méthodes plus performantes, comme détermination initiale des paramètres à l'aide d'une procédure de quantification vectorielle.

I.7.2.2.2 Modèles de markov cachés :

Un défaut commun à la plupart des techniques présentées précédemment (caractérisation par quantification vectorielle et modélisation par mélange de distributions gaussiennes) est le caractère global ces techniques ne tiennent aucun compte de l'ordre dans lesquelles sont présentées les fenêtres d'analyse de signal. Pour le modèle à mélange de distributions gaussiennes on suppose même que les paramètres mesurés dans des fenêtres distinctes sont statistiquement indépendants. 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 est le modèle de Markov caché (HMM).

Les propriétés statistiques des modèles de markov cachés HMMs (Hidden Markov Models) en font une des modélisations les plus efficaces actuellement en reconnaissance du locuteur dépendante du texte. Les HMMs 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 à états.

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.

I.7.2.3 Approche connexionniste :

Les systèmes connexionnistes ou RN (Réseaux de Neurones) qui furent redécouverts et développés dans la fin des années 80, ont suscité beaucoup d'intérêt dans plusieurs domaines. Cette approche comprend une grande famille de méthodes très différentes. Chaque méthode est représentée par un réseau qui implémente une fonction de transfert globale spécifiée par l'architecture et les fonctions élémentaires du réseau. Ainsi les réseaux de neurones peuvent être considérés comme un des modèles d'approximation de fonctions générales non paramétriques. La plupart des recherches dans le domaine reposent sur les perceptrons multicouches MLP (Multi Layer Perceptrons) ou sur des systèmes similaires comme RBF (Radial Basis Fonctions).

Dans cette approche, un locuteur est représenté par un ou plusieurs réseaux de neurones appris directement des trames obtenues en phase de paramétrisation et permettant de le discriminer par rapport un autre ensemble de locuteurs.

Malgré la capacité des RN d'implanter des techniques discriminantes très efficaces, et leurs performances de classement, ils restent incapables de résoudre leur principal problème qui est la durée d'apprentissage importante et nécessaire pour une grande population.

Le principal avantage de ces modèles est leur capacité discriminante qui n'exige pas beaucoup d'hypothèses ni beaucoup de connaissances sur l'application et qui leur permet d'être très efficaces. Par contre, il est toujours très difficile d'insérer de nouvelles données d'apprentissage sans refaire l'apprentissage d'une grande partie du système.

I.7.2.4 L'approche prédictive :

L'approche prédictive repose sur le principe qu'une trame de signal peut être prédite par la seule observation des trames précédentes. De par ce concept, cette approche est considérée dans la littérature comme une approche dynamique.

Une approche tenant compte des informations dynamiques véhiculées par le signal de parole. Elle s'appuie principalement sur l'estimation d'une fonction de prédiction, propre à chaque locuteur et apprise sur les signaux d'apprentissage. Lors de la reconnaissance, une erreur de prédiction peut être calculée entre une trame prédite (par la fonction de prédiction) et la trame réellement observée dans la séquence de test. L'erreur de prédiction moyenne constitue alors la mesure de similarité entre le signal de test et le modèle de locuteur (fonction de prédiction). Une autre solution envisagée est d'estimer une fonction de prédiction sur la séquence de test et de la comparer à l'aide d'une distance à la fonction de prédiction estimée lors de l'apprentissage.

I.7.3 Décision :

Qui consiste à utiliser des mesures de similarités entre le modèle de l'identité de référence et le segment de test, afin de décider si le signal de test est celui de référence ou non et déduire par la suite son identité.

I.8 Conclusion :

Ce chapitre est une introduction au domaine de la RAL. Il présente les différentes tâches liées à la RAL telles que l'identification et la vérification automatique du locuteur ou des tâches plus récentes comme les suivi de locuteur ou l'indexation par locuteur flux audio.

Les diverses applications et problème liés à l'exploitation de la RAL sont aussi exposés, comme la variabilité intra-locuteur et la variabilité due au matériel.

En résumé un système de reconnaissance automatique du locuteur, quelle que soit la tâche considérée, se résume à trois étapes principales qui sont :

- L'analyse acoustique du signal parole.
- La modélisation du locuteur.
- La décision.

Chapitre II :

Description d'un Système L'identification du Locuteur a Base SVM

II.1 Introduction :

Au cours de ce chapitre nous nous intéressons essentiellement à l'identification du locuteur en mode indépendant du texte, en utilisant un classifier SVM. Le principe général de la reconnaissance automatique peut être résumé par la figure II.1 :

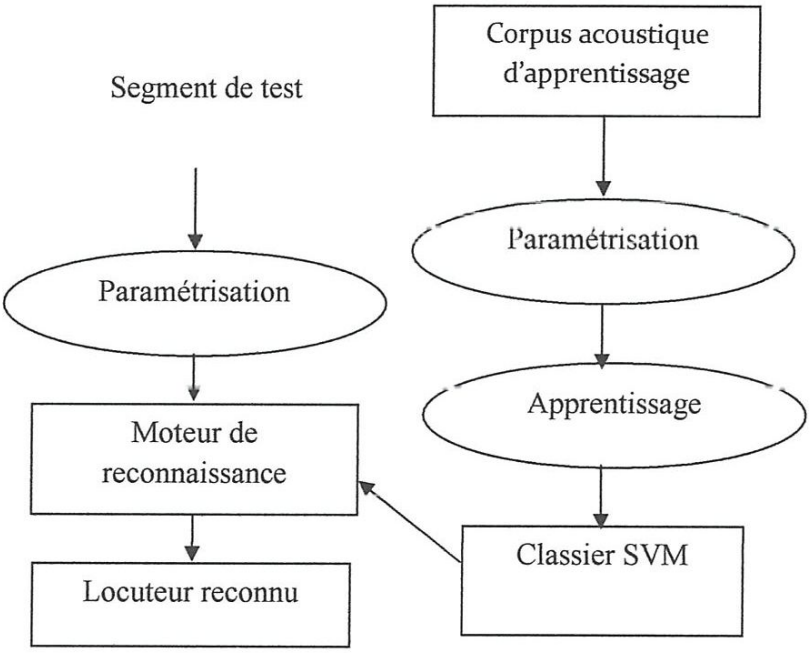

Figure II.1 : Schéma général de la reconnaissance automatique

II.2 Un module de traitement acoustique :

Un système de paramétrisation du signal, appelé aussi prétraitement acoustique, se décompose en deux blocs, le premier de mise en forme (numérisation, Préaccentuation, Décomposition en trames et fenêtrage figure II.2) et l'autre de calcul de coefficients.

Le signal analogique est fourni en entrée et une suite discrète de vecteurs, appelée trame acoustique est obtenue en sortie.

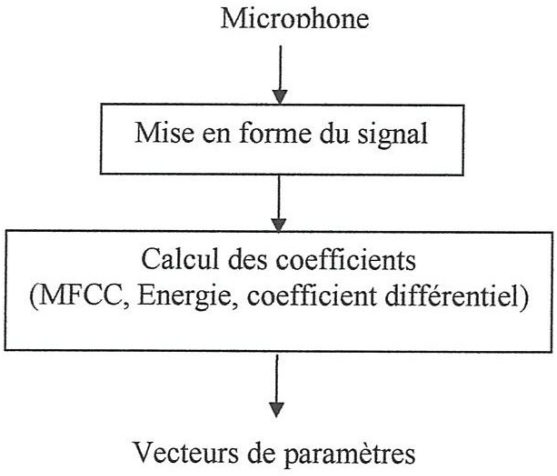

Figure II.2 : Différents étapes de traitement acoustique

II.2.1 Etape de mise en forme :

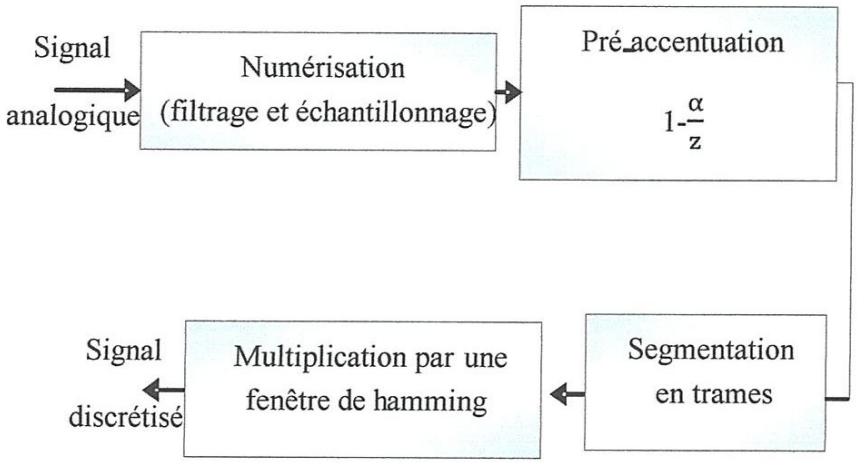

Figure II.3 : Mise en forme du signal

II.2.1.1 Numérisation :

Pour être utilisable par un ordinateur, un signal doit tout d'abord être numérisé. Cette opération tend à transformer un phénomène temporel analogique, le signal sonore dans notre cas, en une suite d'éléments discrets, les échantillons.

Ceux-ci sont obtenus avec une carte spécialisée courante de nos jours dans les ordinateurs depuis l'avènement du multimédia.

La numérisation sonore repose sur deux paramètres : la quantification et la fréquence d'échantillonnage.

La quantification définit le nombre de bits sur lesquels on veut réaliser la numérisation. Elle permet de mesurer l'amplitude de l'onde sonore à chaque pas de l'échantillonnage.

Le choix de la fréquence d'échantillonnage est aussi déterminant pour la définition de la bande passante représentée dans le signal numérisé. Le théorème de Shannon nous indique que la fréquence maximale $f_{\max}$ présente dans un signal échantillonné à une fréquence f_e est égale à la moitié de f_e. Un signal échantillonné à 16000 Hertz contient donc une bande de fréquences allant de 0 à 8000 Hertz.

II.2.1.2 Pré-accentuation :

L'étape de pré- accentuation (ou pré-emphase) consiste à accentuer les hautes fréquences. On fait généralement appel à un filtre de la forme :

$$H(z) = 1 - \alpha_{z^{-1}} \text{ Avec : } 0.9 < \alpha < 1.0$$

Où à est généralement égal à 0.95.

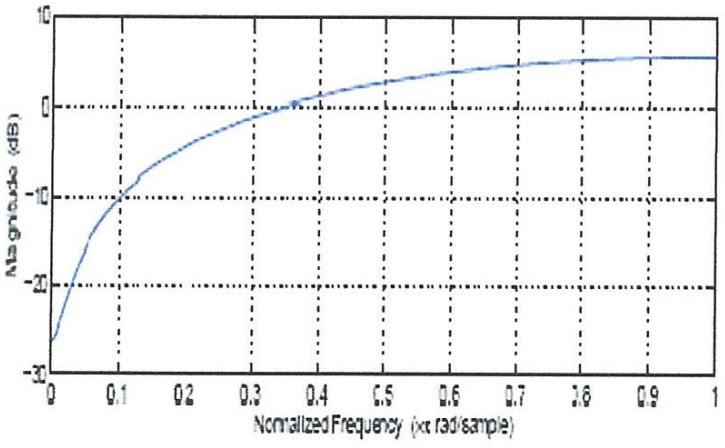

Figure II.4 : Le spectre d'un filtre de pré-accentuation

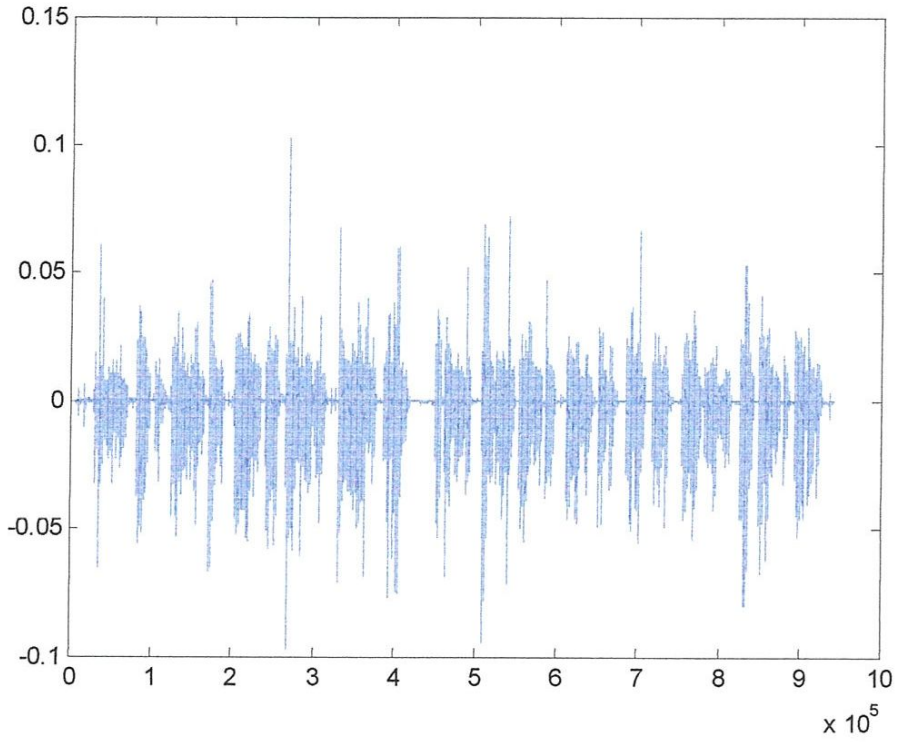

Figure II.5 : Accentuation du texte qui a été prononcée par un locuteur

II.2.1.3 Décomposition en trames et fenêtrage :

Le signal de parole est ensuite décomposé en trames dont la durée est proche de 30 ms. Chaque trame correspond à une portion sur laquelle le signal de parole peut être considéré comme stationnaire.

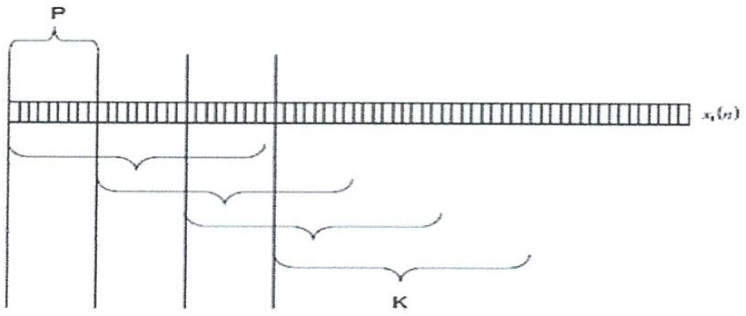

Figure II.6 : Décomposition en trames d'une séquence $x_1(n)$

Ensuite, on applique une fenêtre qui a pour fonction d'atténuer le signal au début et à la fin de chaque trame. Le choix se porte généralement sur les fenêtres de hamming :

$$Ham\min\ g(n) = 0.54 + 0.46.\cos[2\Pi\frac{n}{N-1}]$$

N étant la largeur de la fenêtre. Dans le domaine spectral, ce fenêtrage permet d'atténuer les lobes secondaires associés aux différentes composantes fréquentielles du signal.

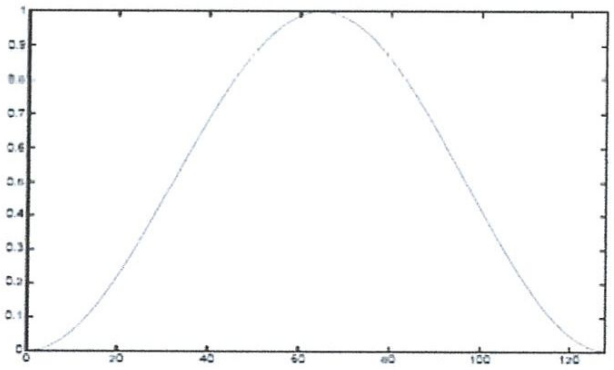

Figure II.7 : Fenêtre de hamming sur 128 points

II.2.2 Etape de paramétrisation :

Un système de paramétrisation du signal a pour rôle de fournir et d'extraire des informations caractéristiques et pertinentes du signal pour produire une représentation moins redondante de la parole. Le signal analogique est fourni en entrée et une suite discrète de vecteurs, appelée trame acoustique est obtenue en sortie.

Les coefficients les plus utilisés en RAL sont certainement les cepstres. Ils peuvent être extraits de deux façons soit par l'analyse paramétrique, à partir d'un modèle de production de type LPC, soit par l'analyse spectrale (modèle phénoménologique). Dans le premier cas, on parlera de LPCC (Lincar Prediction Cepstral Coefficient), et dans le deuxième de MFCC (Mel Frequency Cepstral Coefficients). Dans notre cas l'MFCC sera utilisé.

II.2.2.1 Analyse mel frequency cepstral coefficients (MFCC) :

L'extraction de coefficients MFCC est basée sur l'analyse par banc de filtres qui consiste à filtrer le signal par un ensemble de filtres passe-bande.

L'énergie en sortie de chaque filtre est attribuée à sa fréquence centrale. Pour simuler le fonctionnement du système auditif humain, les fréquences centrales sont réparties uniformément sur une échelle perceptive. Plus la fréquence centrale d'un filtre est élevée, plus sa bande passante est large. Cela permet d'augmenter la résolution dans les basses fréquences, zone qui contient le plus d'information utile dans le signal de parole. Les échelles perceptives les plus utilisées sont l'échelle Mel ou l'échelle Bark Du point de vue performance des systèmes de reconnaissance RAL, ces deux échelles sont quasiment identiques. Dans nos expériences, nous avons fait le choix d'utiliser l'échelle Mel :

$$Mel(f) = \frac{1000}{\log(2)}\left(1+\frac{f}{1000}\right)$$ Avec : f représente la fréquence

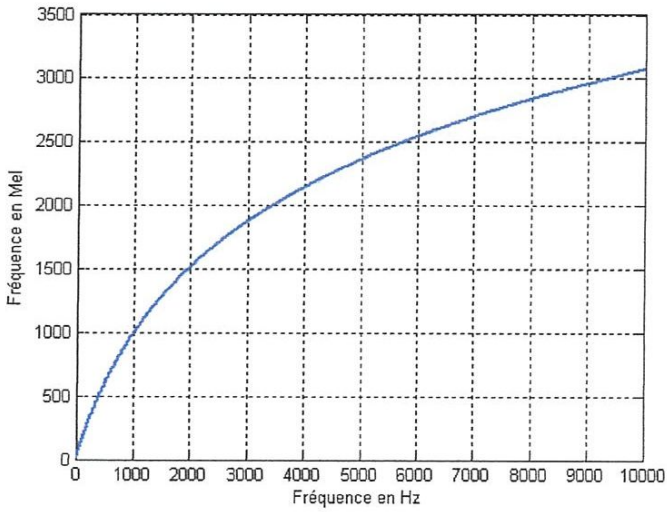

Figure II.8 : Transformation Hz en Mel

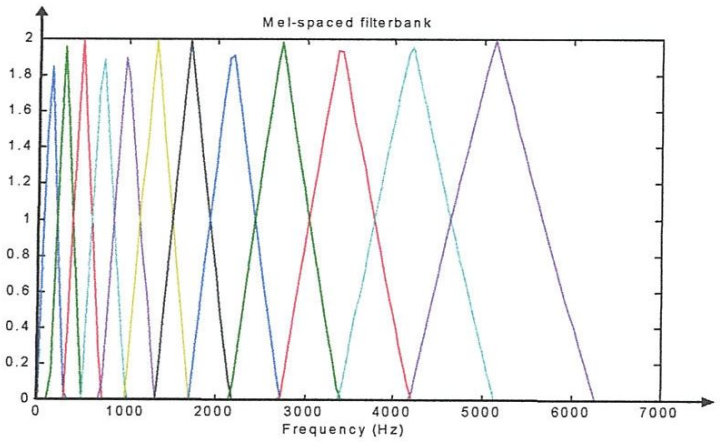

Figure II.9 : Bank de filtres triangulaires de Mel

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), une transformée de Fourier discrète DFT (Discret Fourier Transform), en particulier FFT (Fast Fourier Transform) transformée de fourier rapide, est appliquée pour passer dans le domaine fréquentiel et pour extraire le spectre du signal.

Ensuite le filtrage est effectué en multipliant le spectre obtenu par les gabarits des filtres. Ces filtres sont en général, soit triangulaires soit sinusoïdaux. Dans nos expériences, nous avons choisi d'utiliser des filtres triangulaires répartis sur une échelle Mel.

$$c_n = \sqrt{\frac{2}{k}} \sum_{f=1}^{N} [\log s_k \cos(n(k - \frac{1}{2}) \frac{\Pi}{k})]$$

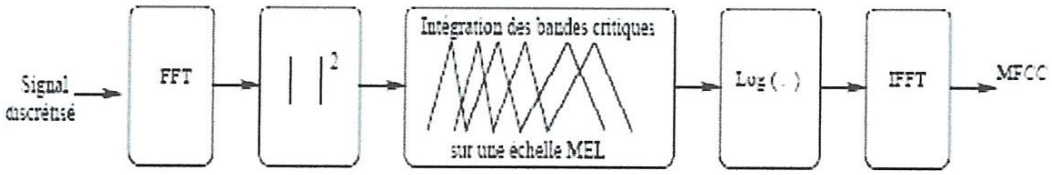

Figure II.10 : Calcul des coefficients MFCC

II.2.2.2 Paramètres dynamiques :

Le vecteur de paramètres issus des méthodes précédentes peut être complété par un vecteur correspondant aux dérivées temporelles premières et secondes de ces paramètres. Ces dérivées sont estimées sur base de plusieurs trames adjacentes. L'approche permet d'introduire une information concernant le contexte temporel de la trame courante.

Soit $c_k(t)$ le coefficient cepstral d'indice k de la trame t, alors le coefficient différentiel $c_k(t)$ correspondant est calculé sur $2n+1$ trames d'analyse par :

$$\Delta_{c_k}(t) = \frac{\sum_{i=-n_\Delta}^{n_\Delta} i.c_k(t+i)}{\sum_{i=-n_\Delta}^{n_\Delta} i^2}$$

On peut écrit :

$$\Delta_{c_k}(t) = \frac{\sum_{i=1}^{n_\Delta} i.(c_k(t+1)-c_k(t-1))}{\sum_{i=1}^{n_\Delta} i^2}$$

II.2.2.3 L'énergie :

L'énergie du signal est un indice qui peut par exemple contribuer à la détection du voisement d'un segment de parole. L'énergie totale E_0 est calculée directement dans le domaine temporel sur une trame de signal $(s_n)0 \le n \le N-1$ comme :

$$E_0 = \sum_{n=0}^{N-1} s_n^2$$

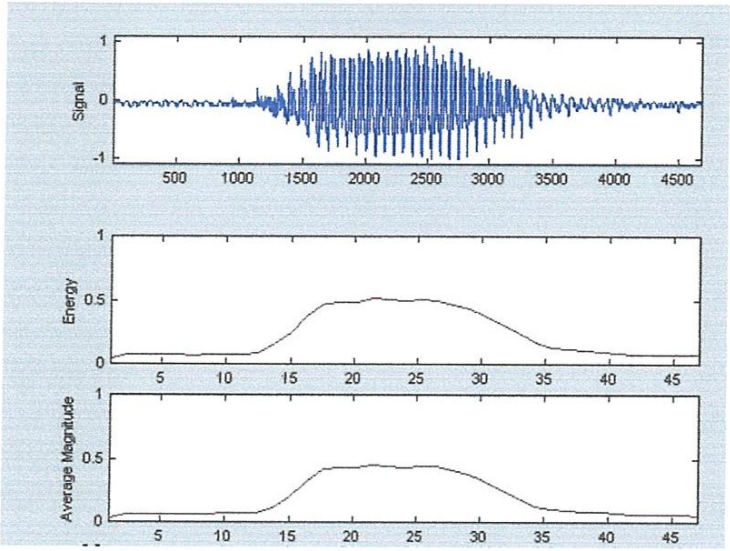

Figure II 11 : L'énergie

II.3 Un module de création des modèles :

Un problème dans le domaine de traitement de l'information est d'arriver à simplifier l'information. Une méthode simple, pour y remédier, est de regrouper en une même classe différents éléments porteurs d'une information à peu près similaire. Ainsi l'information totale est représentée par de grands ensembles (les classes) et non plus par les nombreux sous ensembles. Les méthodes de classification les plus utilisées sont la méthode SVM.

Les Machines à vecteurs de support SVM (Support Vector Machines) fournissent une approche très intéressante de l'approximation statistique. Souvent, le nombre d'exemples pour l'apprentissage est insuffisant pour que les estimateurs fournissent un modèle avec une bonne precision. D'un autre côté, l'acquisition d'un grand nombre d'exemples s'avère être souvent très coûteuse et peut même mener à des problèmes de sur apprentissage dans le cas où la capacité du modèle est très complexe. Pour ces deux raisons, il faut arriver à un compromis entre la taille des échantillons et la precision recherchée.

Dans ces cas spécifiques comme la reconnaissance de formes, il serait intéressant de trouver une mesure de la fiabilité de l'apprentissage, et d'avoir une mesure du taux d'erreur qui sera commis durant la phase de test.

Ces nouvelles techniques unifient deux théories, la minimisation du risque empirique et la capacité d'apprentissage d'une famille de fonctions.

Le principe des SVM consiste à projeter les données de l'espace d'entrée (appartenant à deux 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. Dans cet espace, on construit un hyperplan optimal séparant les classes tel 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.

La figure II.12 représente le principe de la technique SVM.

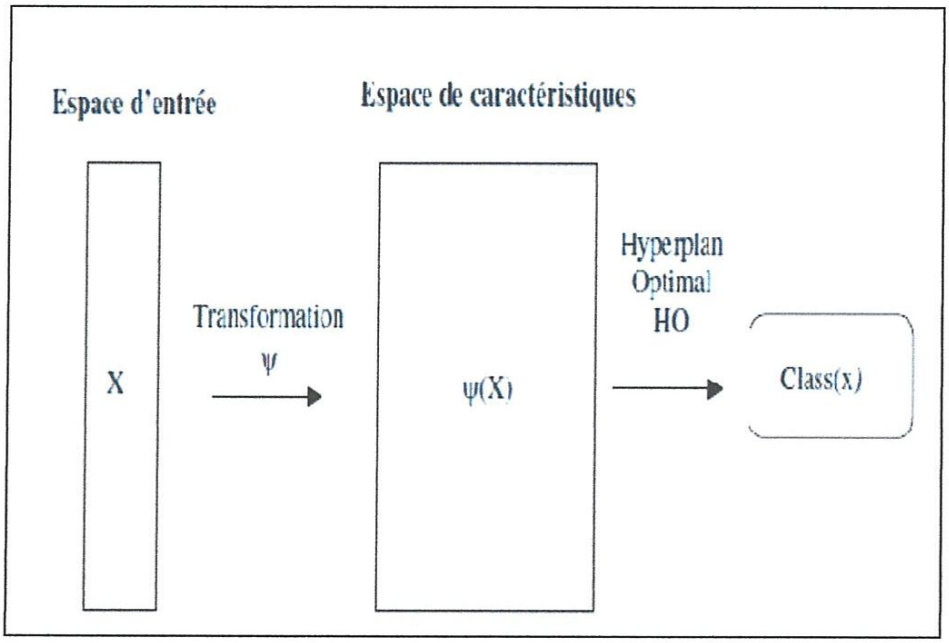

Figure II.12 : Principe des techniques SVM

II.4 Construction de l'hyperplan optimal :

Pour bien décrire la technique de construction de l'hyperplan optimal séparant des données appartenant à deux classes différentes dans deux cas différents. Le cas des données linéairement séparables et le cas des données non-linéairement séparables, nous utiliserons la notation suivante :

Soit l'ensemble D tel que :

$$D=(x_i,y_i)\in R^N X(-1,+1) \text{ Pour : } i=1,...,m$$

II.4.1 Cas des données linéairement séparables :

Dans ce paragraphe nous présentons la méthode générale de construction de l'Hyperplan Optimal (HO) qui sépare des données appartenant à deux classes différentes linéairement séparables. La figure II.13 donne une représentation visuelle de l'HO dans le cas des données linéairement séparables avec :

H : un hyperplan quelconque, HO : l'hyperplan optimal et M la marge qui représente la distance entre les différentes classes et HO (VS sont les Vecteurs Supports).

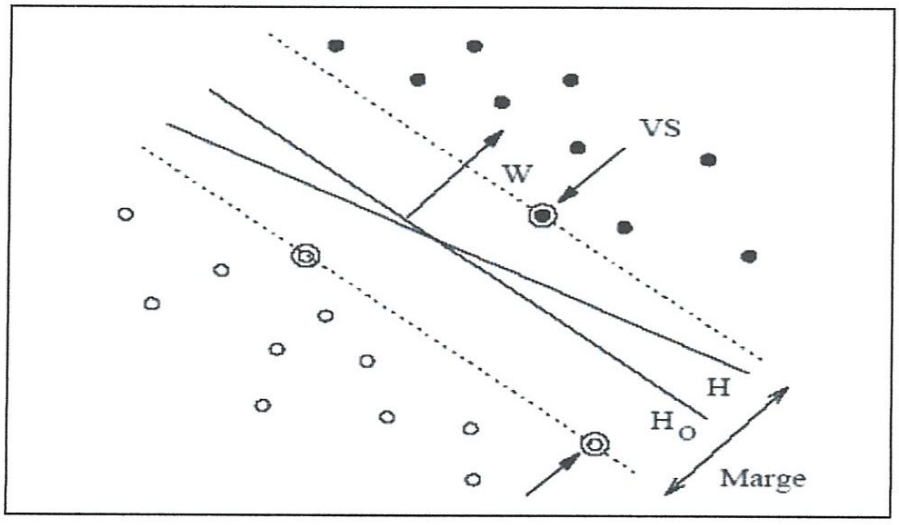

Figure II.13 : Données linéairement séparables

Soit H : l'hyperplan d'équation $(w.x)+b=0$ qui satisfait les conditions suivantes :

$$(w.x_i)+b\geq 1 \text{ Si : } y_i = 1$$

$$(w.x_i)+b\leq 1 \text{ Si : } y_i = -1$$

Ce qui est équivalent à :

$$y_i(w.x_i+b)\geq 1 \text{ Pour : } i=1,...,m \qquad \text{(II.1)}$$

Comme nous l'avons déjà mentionné, un HO est un hyperplan qui maximise la marge M qui représente la plus petite distance entre les différentes données des deux classes et l'hyperplan.

Maximiser la marge M est équivalent à maximiser la somme des distances des deux classes par rapport à l'hyperplan. Ainsi, la marge a l'expression mathématique suivante :

$$M=\min_{\substack{x_i \\ y_i=1}}\frac{w.x+b}{\|w\|}-\max_{\substack{x_i \\ y_i=1}}\frac{w.x+b}{\|w\|}=\frac{1}{\|w\|}-\frac{-1}{\|w\|}=\frac{2}{\|w\|}$$

Trouver l'hyperplan optimal revient donc à maximiser $\frac{2}{\|w\|}$. Ce qui est équivalent à minimiser $\frac{\|w\|^2}{2}$ sous les contraintes (II.1).

Ceci est un problème de minimisation d'une fonction objectif quadratique sous contraintes linéaires.

On définit les Vecteurs Supports VS tout vecteur x_i tel que :

$$y_i[(w.x_i)+b]=1$$

Ce qui est équivalent à : $VS=(x_i/\alpha_i)$ Pour : $i=1,...,m$

Ainsi, on peut facilement calculer w_i et b_i :

$$w_0 = \sum VS\alpha_i^0 y_i^0 x_i^0$$

$$b_0 = -\frac{1}{2}[(w_0.x*(1))]+[(w_0.x*(-1))]$$

La fonction de classement class (x) est définie par :

$$class(x) = [(w_{\grave{a}}.x) + b_i] = sign[\sum \alpha_1^0 y_i (x_i.x) + b_0] \qquad \text{(II.2)}$$

Avec : $x_i \in VS.$

Si class (x) est inférieur à 0, x est de classe –1 sinon il est de la classe 1.

II.4.2 Cas des données non-linéairement séparables :

Dans ce cas où les données sont non-linéairement séparables figure II.14, l'hyperplan optimal est celui qui satisfait les conditions suivantes :

- La distance entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale.

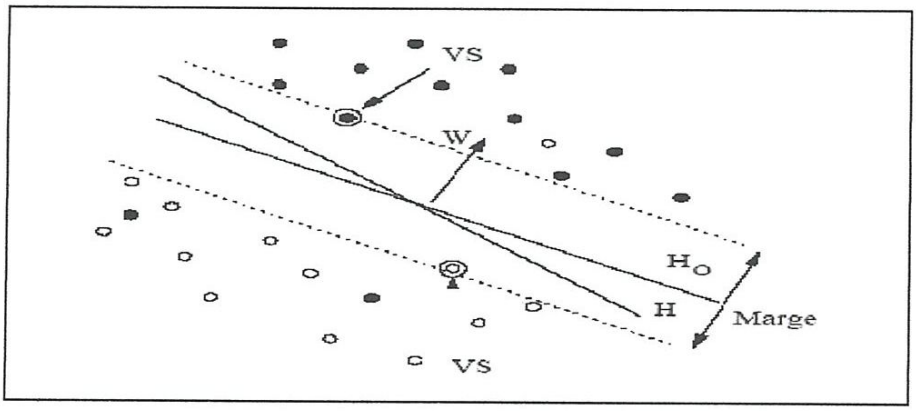

Figure II.14 : Données non-linéairement séparables

Pour formaliser tout cela, on introduit des variables de pénalité non-négatives δ_i appelées variables d'écart. Ces variables transforment l'inégalité (II.1) comme suit :

$$y_i(w.x_i + b) \geq 1 - \varepsilon_i \text{ Pour : } i =,...,m$$

L'objectif est de minimiser la fonction suivante :

$$\frac{1}{2} w.w + c\sum_{i=1}^{m} \delta_i$$

Où c est un paramètre de régularisation. Elle permet de concéder moins d'importance aux erreurs. Cela mène à un problème dual légèrement différent de celui du cas des données linéairement séparables. Maximiser le lagrangien par rapport à α_i sous les contraintes suivantes :

$$\sum_{i=1}^{m} \alpha_i y_i = 0 \text{ Avec : } 0 \leq \alpha_i \leq c \text{ Pour : } i = 1,...,m$$

Le calcul de la normale w_0, du biais b_0 et de la fonction de classification class (x) reste exactement le même que pour le cas des données linéairement séparables.

II.5 Principe des SVM :

Les classificateurs SVM utilisent l'idée de l'HO (Hyperplan Optimal) pour calculer une frontière entre des nuages de points. Elles projettent les données dans l'espace de caractéristiques en utilisant des fonctions non-linéaires. Dans cet espace on construit l'HO qui sépare les données transformées. L'idée principale est de construire une surface de séparation linéaire dans l'espace des caractéristiques qui correspond à une surface non-linéaire dans l'espace d'entrée.

Le problème principal à relever ici est comment bien manipuler la transformation de tous les vecteurs d'entrée dans l'espace des caractéristiques de façon à éviter une augmentation du coût en nombre de paramètres libres.
Soit l'ensemble D′ l'image de l'ensemble D, défini dans la section II.4, par la transformation ω.

$$D' = ((w(x_i), y_i) \in R^p x(-1,+1) \quad \text{Pour : } i = 1,...,m \quad p \geq n)$$

En construisant un HO dans l'espace des caractéristiques suivant la technique expliquée dans la section II.4. On aura la fonction de classement suivante :

$$class(x) = sign[\sum \alpha_i^0 y_i (\omega(x_i).\omega(x)) + b_0] \qquad \text{(II.3)}$$

Avec : $x_i \in VS$.

On peut remarquer que la fonction de classement dépend du produit scalaire dans l'espace des caractéristiques. Ainsi, pour que le coût de calcul reste pratiquement inchangé et le nombre de paramètres libres du système n'augmente pas, il faut que la fonction ω satisfasse la condition suivante : $\omega(u)\omega(v) = k(u,v)$.

C'est à dire le produit scalaire dans l'espace des caractéristiques va être représentable comme un noyau de l'espace d'entrée. Le classificateur est donc construit sans utiliser explicitement la fonction ω.

Plusieurs noyaux ont été utilisés par les chercheurs, en voici quelques uns :

- Le noyau linéaire : $k(u,v) = u.v$.
- Le noyau polynomial : $k(u,v) = [(u.v)+1]^d$ Ou $d \in N$ est le degré du polynôme à déterminer par l'utilisateur.
- Le noyau RBF (Radial basis function) : $k(u.v) = \exp(-\frac{\|u-v\|^2}{2\sigma^2})$ ou σ est à déterminer.
- Le noyau polynomial réel de Vovk : $k(u;v) = \frac{1-(u.v)^d}{1-(u.v)}$ avec :

  $-1 < u.v < 1$.
- Le noyau polynomial réel infini de Vovk : $k(u,v) = \frac{1}{1-(u.v)}$ avec :

  $-1 < u.v < 1$.

Maintenant que nous avons défini ce qu'est un noyau, la fonction de classement (II.3) devient :

$$class(x) = sign[\sum \alpha_i^0 y_i k(x_i, x) + b_0] \qquad (II.4)$$

Avec : $x_i \in VS$.

Reprenons l'exemple qu'on a évoqué au début de ce chapitre. Donc nous avons la transformation ω tel que :

$$\omega : R^2 \rightarrow R^3$$

$$(x_1, x_2) \rightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2$$

D'où :

$$k(u,v) = \omega(u).\omega(v) = ((u_1^2, \sqrt{2})u_1u_2, u_2^2)$$

$$\begin{pmatrix} v_1^2 \\ \sqrt{2v_1v_2} \\ v_2^2 \end{pmatrix}$$

$$= (u_1^2v_1^2 + 2u_1v_1u_2v_2 + u_2^2v_2^2) = (v_1u_1 + v_2u_2)^2$$

$$= \left((u_1, u_2).\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) = (u.v)^2$$

Comme on peut le constater, le noyau correspondant à la transformation ω de notre exemple proposé au début de ce chapitre n'est autre qu'un noyau polynomial de degré 2.

II.6 Extension du SVM binaire au cas multi-classes :

L'expression des SVM décrite précédemment peut seulement résoudre des problèmes de classification binaires. Or les problèmes les plus rencontrés en pratique sont ceux où l'on à plusieurs classes. L'extension du SVM aux problèmes de plus de deux classes présente un grand intérêt pratique.

II.6.1 Les SVM pour la classification de k classes :

Etant donnée un ensemble d'exemples étiquetés : $s=((x_1,y_1),(x_2,y_2),...,(x_1,y_1))$ .On s'intéresse cette fois ci au problème de classification où le nombre de classes est égal à k , les yi prennent alors leurs valeurs dans l'ensemble $y=(1,2,...,k)$. Dans ce cas, il ne s'agit plus de trouver une seule séparatrice entre deux classes mais il faudra que l'on soit en mesure de classer les exemples en plusieurs classes, ce qui revient à construire k hyperplans linéaires d'équations :

$$w_k^T x+b_k=0, k\in y$$

Un cas de figure que l'on rencontre souvent dans la pratique est le problème de la reconnaissance de formes. Où il y a plusieurs classes de formes.

Généraliser les SVM au cas multi-classes revient à résoudre le problème de programmation quadratique suivant (problème primal) :

$$\min_{w_i b_i \delta_i^k} \frac{1}{2}\sum_{k\in y}\|wk\|^2 + c\sum_{i\in I, k\in y/(y_i)}\delta_i^k$$

Sous contraintes $w_{y_i}^T x_i + b_{y_i} - (w_k^T x_i + b_k) \geq 1-\delta_i^k, \delta_i^k \geq 0, i\in I, k\in y/(y_i)$

La fonction de décision (classement) sera alors :

$$class(x)=\text{argmax} class_k(x) \text{ Ou } class_k(x)=w_k^T x+b_k \text{ Avec : } k \in y$$

Donc un nouvel individu aura comme classe, la classe y_i avec :

$$class_{y_i}(x) \geq class_k(x), k \in y/(y_i)$$

Etant donné que ce problème complexe est difficile à résoudre, plusieurs méthodes ont été proposées pour résoudre le cas Multi-classes.

II.6.1.2 Un contre tous (one versus all) :

Pour chaque classe on détermine un hyperplan séparant celle-ci de toutes les autres, en considérant cette dernière comme la classe +1 et les autres classes comme étant la classe -1, ce qui résulte en k SVM binaires (pour un problème à k classes).

II.6.1.3 Un contre un (one versus one) :

Dans ce cas on construit un SVM pour chaque paire de classes. Ainsi, on sera amené à calculer $k(k-1)/2$ SVM binaires.

II.7 Décision :

II.7.1 Identification automatique du locuteur :

En identification, un signal de test est comparé à toutes les références de locuteurs connus du système, résultant en un ensemble de mesures de similarité en entrée du processus de décision. Aussi, ce processus à pour tâche de rechercher la mesure de similarité maximale (ou minimale dans le cas de mesures de distance) et de designer l'identité du locuteur correspondant.

Dans ce contexte, la mesure des performances d'un système d'IAL est généralement donnée en termes de taux d'identification correcte. Ce taux s'obtient par la formulation suivante :

$$Taux\ d'identification\ correcte = \frac{\#\ tests\ ayant\ amené\ à\ une\ identification\ correcte}{\#\ tests\ total}$$

Il faut signaler que qu'il existe plusieurs facteurs qui peuvent augmenter la variabilité intra-locuteur et qui, par conséquent, influencent sur la décision du système de reconnaissance :

- L'état pathologique du locuteur (maladie, émotions, …).
- Vieillissement (la voix d'une personne change avec l'âge).
- Facteurs socioculturels (le locuteur peut changer d'accent).
- Locuteurs non coopératifs (notamment dans les applications judiciaires).
- Conditions de prise de son, bruit ambiant,…

II.8 Conclusion :

Dans ce chapitre, nous allons utiliser les SVM comme technique de classification qui exige des vecteurs d'entrée de taille fixe. Le choix des SVM est soutenu par leur souplesse et surtout par une grande gamme de fonctions que ces techniques proposent afin d'approcher au mieux la fonction de classement réelle. Cela n'empêche pas que ces techniques présentent d'autres avantages :

- Ce sont des techniques discriminantes qui permettent de construire des surfaces de décision non-linéaires alors que la plupart des autres techniques utilisées dans le domaine de la RAL sont limitées à des solutions linéaires.
- Ce sont des techniques adaptatives. Elles permettent aux systèmes d'évoluer en fonction des spécificités de la tâche qu'ils doivent réaliser, principalement dans le cas d'un apprentissage incrémental dans les applications réelles de la reconnaissance du locuteur.

Chapitre III :

Etude Expérimental

III.1 Introduction :

Nous allons aborder dans ce chapitre l'utilisation pratique de la méthode SVM vue précédemment et l'application de celles-ci dans un système réel de reconnaissance automatique de locuteur.

Nous allons présenter, une étude d'un système de reconnaissance de locuteur a base de SVM. Nous avons utilisé ici le logiciel de programmation MATLAB 7.4 pour réaliser notre implémentation, car nous pensons que, c'est un langage de programmation utile et flexible dans le domaine de traitement du signal.

III.2 Description de la base de données IViE corpus :

La base de données EIVE CORPUS est une base de données acoustique dédiée à la reconnaissance automatique de la parole, ainsi qu'au développement et à l'évaluation des systèmes de reconnaissance automatique de la parole. Elle contient les enregistrements de 110 locuteurs britanniques, prononçant chacun 5 phrases. Le texte est lu dans de bonnes conditions d'enregistrement et les données sont échantillonnées avec 16 KHz sur 16 bits.

Fréquence d'échantillonnage	16 KHZ
Résolution	16 bit
Nombre de locuteurs	110
Nombre de session par locuteur	5
Type de la parole	Lecture de phrases

Tableau III.1 : Description de la base de données EVIE corpus

III.3 Description de la base de données utilisée :

Dans le cadre de ce projet de fin d'études, on a utilisé une base de données composée de 12 locuteurs (6 hommes et 6 femmes) extraite exclusivement de la base de données Corpus IVIE. Pour chaque locuteur, on dispose de 5 phrases, chacune de 60 secondes en moyenne. On a concaténé une phrase pour l'apprentissage et toutes (5) phrases pour le test. Les fichiers d'apprentissage sont étiquetés de (locuteur test 1) à (locuteur test 12) et les fichiers de test de (locuteur test 01) (locuteur test 12).

III.4 Les étapes de construction du système :

III.4.1 Prétraitement :

- La préaccentuation du signal échantillonné par un filtre de transmit tance $(1-a.z^{-1})$.
- Découpage du signal en trame de 256 points, avec un décalage de 128 points.
- Pondération de chaque trame par une fenêtre de hamming.

III.4.2 Analyse acoustique par MFCC :

Après l'acquisition et le prétraitement, on passe à l'analyse acoustique de chaque mot, pour notre application on utilise la méthode MFFC. Le calcule des coefficients MFCC se fait pour chaque trames.

III.4.3 Création de classifier SVM :

Pour les expériences sur les SVM nous avons effectué les tests dans le cas binaire et multi-classe avec différentes configurations possibles et calculé à chaque fois l'erreur commise sur l'échantillon de test c'est à dire le pourcentage de vecteurs mal classés.

III.4.4 Phase de test :

Après la détection de frontière le mot de test subit le même traitement que les mots de la base d'apprentissage. Le mot sera comparé au modèle des locuteurs, pour la méthode de SVM une probabilité locale entre un vecteur acoustique et le modèle est calculé.

III.5 Résultats expérimentaux :

Les résultats obtenus dans ce premier cas sont représentés dans des tableaux :

III.5.1 Premier cas :

Le paramètre de régularisation C variable et le noyau RBF (G) a été fixée à 0.01

uteur	1	2	3	4	5	6	7	8	9	10	11	12	Moye
C=1	75.80	96.55	98.03	100	90.74	98.27	98.07	98.07	98.30	100	98.11	93.22	95.43
C=2	82.25	96.55	96.07	100	90.74	98.27	98.07	96.15	98.30	100	96.22	94.19	95.63
C=3	83.78	96.55	98.03	100	88.88	96.55	98.07	96.15	98.30	100	98.11	96.16	95.63
C=4	85.48	96.55	98.03	100	88.88	96.55	98.07	96.15	98.30	100	98.11	96.16	96.06
C=10	85.48	96.55	98.03	100	88.88	96.55	98.07	98.07	98.30	100	98.11	96.16	96.22
C=20	85.48	96.55	96.07	100	88.88	96.55	98.07	98.07	98.30	100	96.22	94.19	95.76
=100	98.38	96.55	96.07	100	100	100	98.07	98.07	98.30	100	100	100	98.78
=200	98.38	96.55	94.11	100	100	100	98.07	98.07	98.30	100	100	100	98.62
=300	98.38	96.55	94.11	100	98.14	100	98.07	98.07	98.30	100	100	100	98.47

Tableau III.2 : Taux de reconnaissance pour pp=125

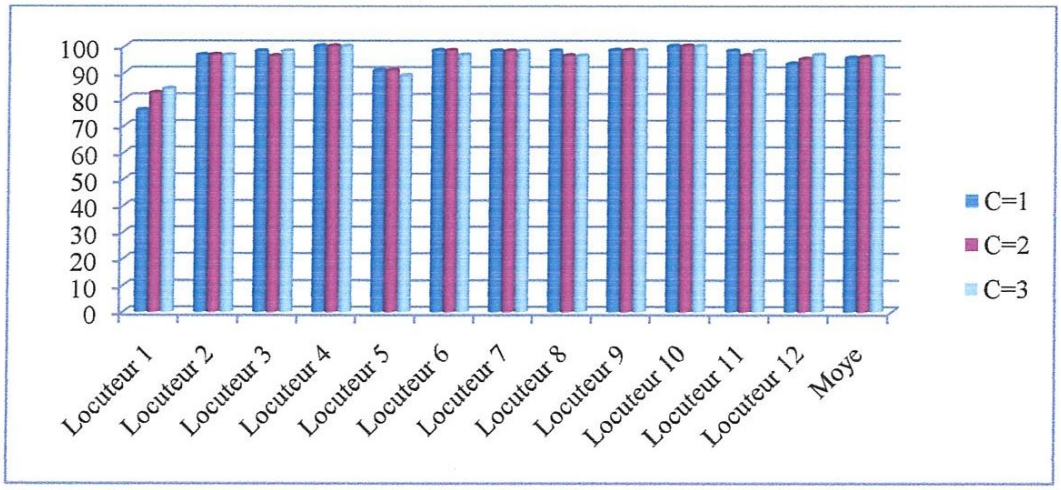

Figure III.1 : Taux de reconnaissance pour (C=1, C=2, C=3)

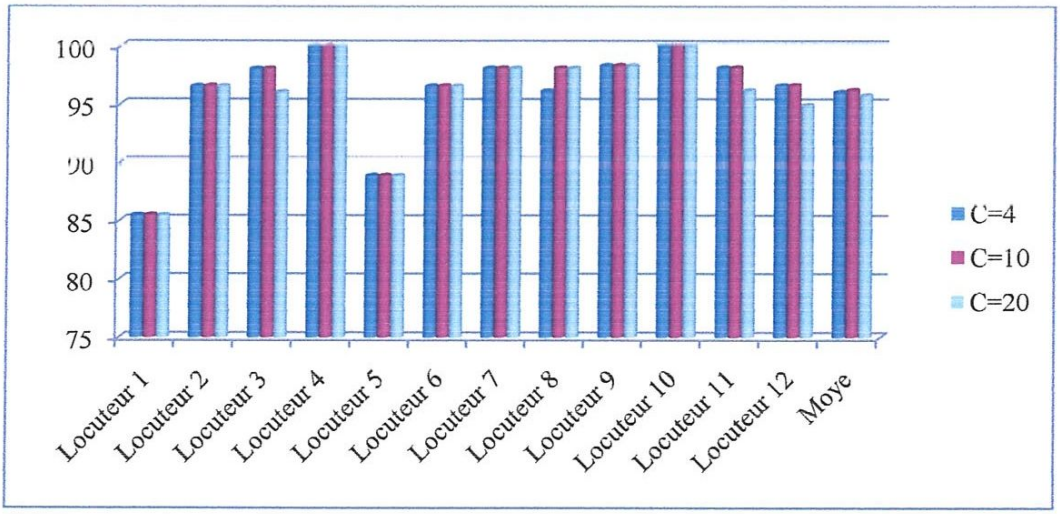

Figure III.2 : Taux de reconnaissance pour (C=4, C=10, C=20)

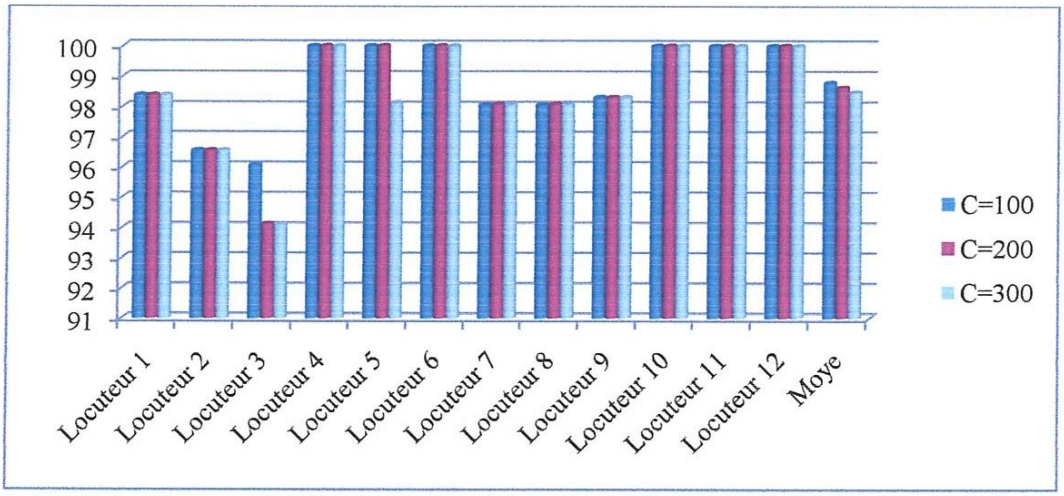

Figure III.3 : Taux de reconnaissance pour (C=100, C=200, C=300)

cuteur	1	2	3	4	5	6	7	8	9	10	11	12	Moye
C=1	76.03	98.26	88.88	95.04	95.65	93.38	96.26	86.66	96.36	98.03	90.00	83.33	91.49
C=2	76.03	98.26	86.86	95.04	94.56	93.38	96.26	85.71	96.36	99.01	80.00	83.33	90.49
C=3	77.68	96.52	86.86	95.04	94.56	92.56	96.26	84.76	94.54	99.01	76.66	83.33	89.82
C=4	77.68	97.39	87.87	94.05	94.56	91.73	96.26	84.76	94.54	98.03	80.00	83.33	90.02
C=10	78.51	96.52	87.87	93.06	93.47	94.21	96.26	85.71	92.72	98.72	76.66	83.33	89.70
C=20	76.85	96.52	88.88	93.06	93.47	94.21	96.26	83.80	92.72	98.03	80.00	83.33	89.76
C=100	95.93	94.78	86.27	99.00	96.26	96.52	97.08	97.08	93.16	99.02	100	96.61	95.98
C=200	95.93	93.91	85.29	99.00	96.26	96.52	97.08	97.08	93.16	98.05	100	96.61	95.74
C=300	95.93	93.94	86.27	99.00	97.19	96.52	97.08	97.08	93.16	99.02	99.05	96.61	95.90

Tableau III.3 : Taux de reconnaissance pour pp=125/2

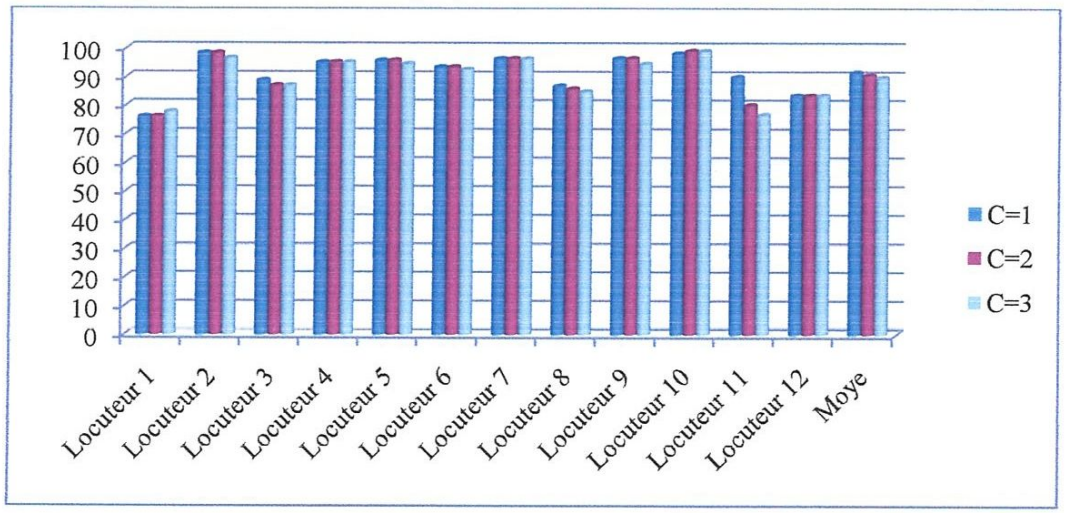

Figure III.4 : Taux de reconnaissance pour (C=1, C=2, C=3)

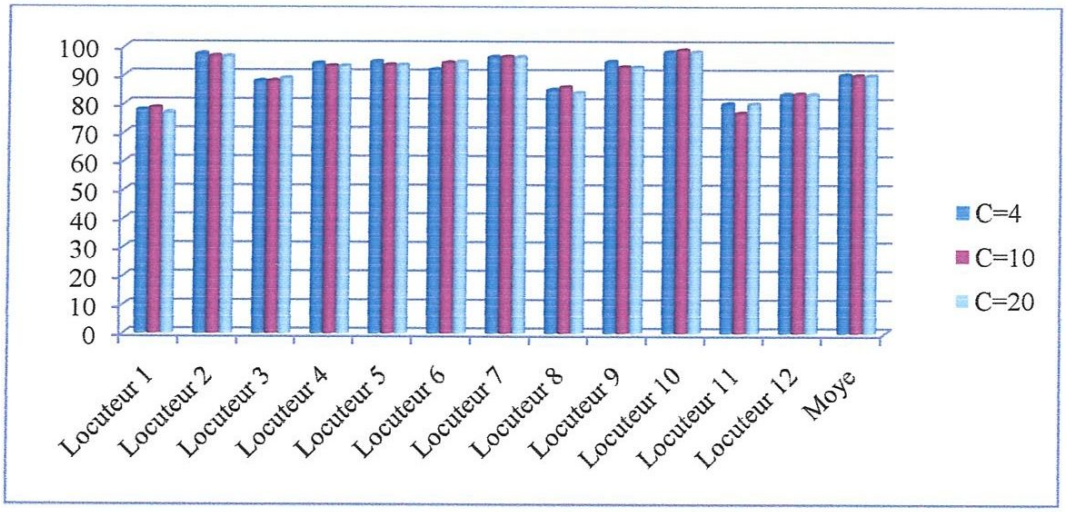

Figure III.5 : Taux de reconnaissance pour (C=4, C=10, C=20)

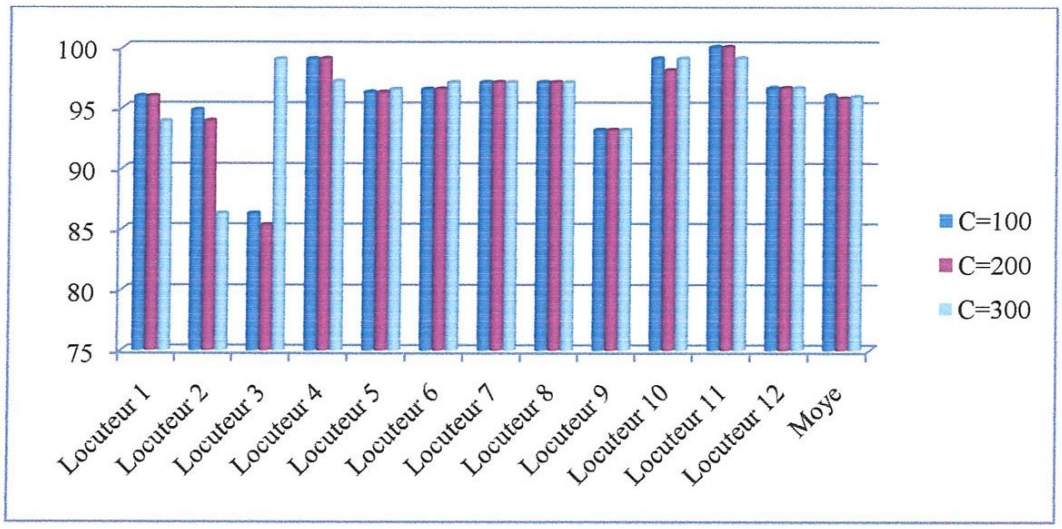

Figure III.6 : Taux de reconnaissance pour (C=100, C=200, C=300)

II.5.1.1 Discussion :

En comparant les résultats des deux tableaux (III.2 et III.3), dans le premier cas avec le paramètre de régularisation C variable et Gamma fixe et égale 0.01 pour un segment de test de taille 1 seconde (c-a-d 125 frames). Le meilleur résultat est obtenu pour une valeur C=100 puisque le taux de reconnaissance (98.78) plus approche a (100), et pour un segment de test de taille 0.5 seconde (c-a-d 125/2) frames), le meilleur résultat est obtenu pour C=100 puisque le taux de reconnaissance (95.98) plus approche a (100).

On conclure que le taux de reconnaissance est augmente avec la taille du segment de test.

III.5.2 Deuxième cas :

Gamma variable et le paramètre de régularisation C fixée à 100

cuteur	1	2	3	4	5	6	7	8	9	10	11	12	Moye
G=0.01	98.38	96.55	96.07	100	100	100	98.07	98.07	98.30	100	100	100	98.78
G=0.1	100	10.34	0	2.00	5.55	6.89	11.53	0	5.08	5.88	47.16	83.05	23.12
G=1	0	0	0	0	0	0	0	0	0	0	0	100	8.33
=0.001	100	100	100	100	100	100	100	100	100	100	100	98.19	99.84
G=0.02	100	100	100	100	100	100	100	100	100	100	100	100	100
=0.002	100	100	100	100	100	100	100	100	100	100	100	100	100

Tableau III.4 : Taux de reconnaissance pour pp=125

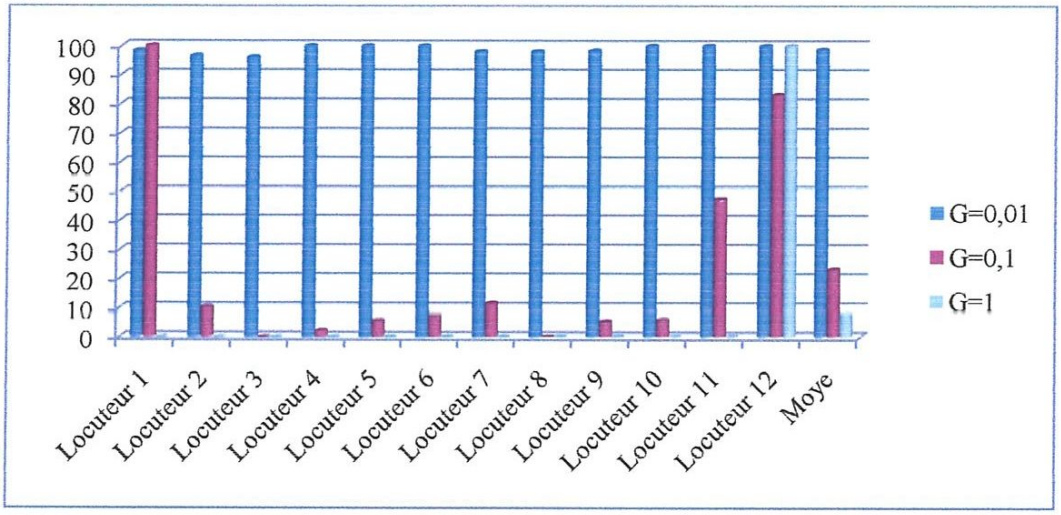

Figure III.7 : Taux de reconnaissance pour (C=0.01, C=0.1, C=1)

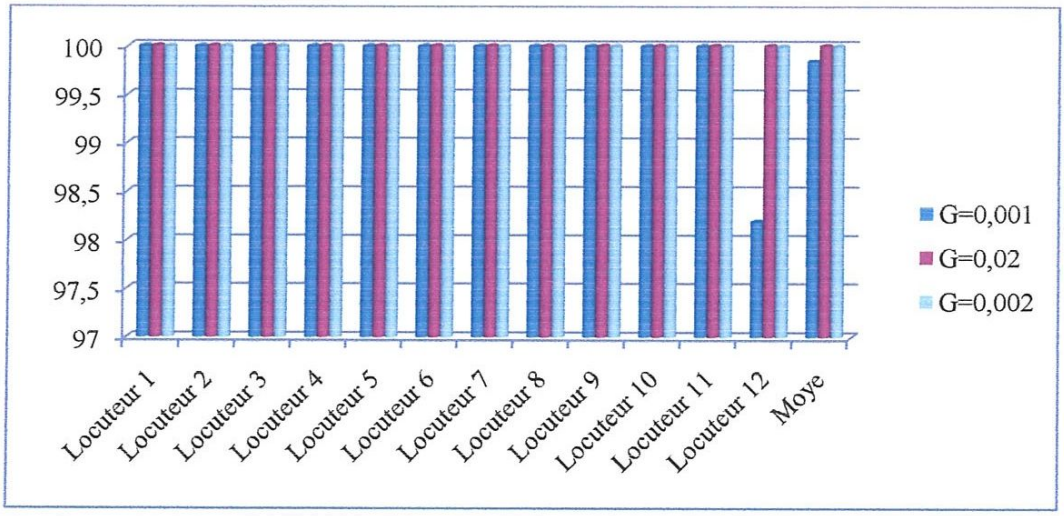

Figure III.8 : Taux de reconnaissance pour (C=0.001, C=0.2, C=0.002)

cuteur	1	2	3	4	5	6	7	8	9	10	11	12	Moye
=0.01	95.93	94.78	86.27	99.00	96.26	96.25	97.08	97.08	93.16	99.02	100	96.61	95.98
G=0.1	99.18	16.52	0.98	4.95	10.28	14.72	19.41	6.79	11.96	12.62	47.16	73.72	26.53
G=1	0	0	0	0	0	0	0	0	0	1.94	0.94	100	8.57
=0.001	100	99.13	100	98.27	100	100	100	97.29	100	100	98.86	95.28	99.07
=0.02	100	100	100	100	100	100	100	100	100	100	100	100	100
=0.002	100	99.13	100	98.27	100	100	100	100	100	100	100	99.05	99.70

Tableau III.5 : Taux de reconnaissance pour pp=125/2

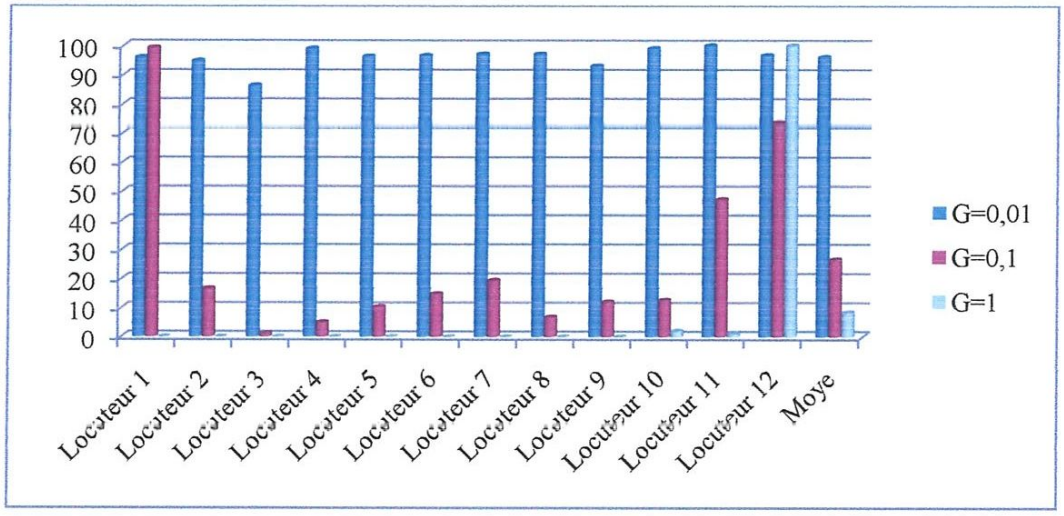

Figure III.9 : Taux de reconnaissance pour (C=0.01, C=0.1, C=1)

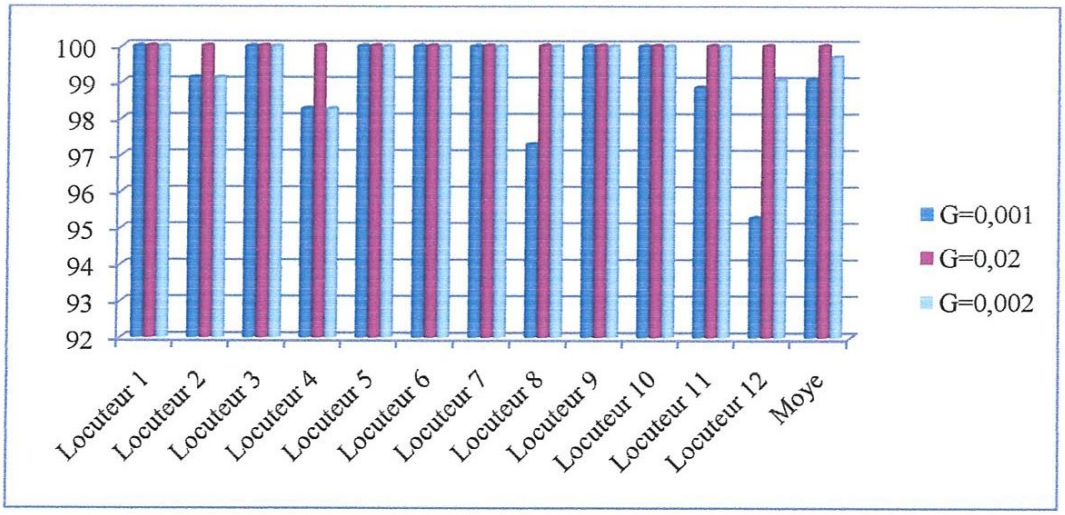

Figure III.10 : Taux de reconnaissance pour (C=0.001, C=0.2, C=0.002)

III.5.2.1 Discussion :

En comparant les résultats des deux tableaux (III.4 et III.5), de deuxième cas ou le noyau RBF (G) variable et le paramètre de régularisation C a été fixée à 100 pour pp=125 (voir III.4) le meilleur résultat G=0.02 et 0.002 puisque le taux de reconnaissance égale (100%), et pour pp=125/2 (voir III.5) le meilleur résulta G=0.02 a pour taux de reconnaissance égale 100%.
Donc on peut dire que si le pp est plus grand donc on obtient la meilleure valeur de taux de reconnaissance

En comparant les résultats des deux tableaux (III.4 et III.5), dans le deuxième cas avec le paramètre de régularisation C fixe et Gamma variable pour un segment de test de taille 1 seconde (c-a-d 125 frames). Le meilleur résultat est obtenu pour une valeur de Gamma =0.002 et Gamma =0.02 puisque le taux de reconnaissance est (100%), et pour un segment de test de taille 0.5 seconde (c-a-d 125/2) frames), le meilleur résultat est obtenu pour Gamma=0.02.

On conclure que le taux de reconnaissance est augmente avec la taille du segment de test avec une valeur de Gamma =0.02.

III.6 Conclusion :

Nous avons présenté dans ce chapitre les principaux résultats obtenues par les classificateurs SVM multi-classes et binaire pour la classification des locuteurs. Les performances atteintes par notre approche sont intéressantes. En effet avec l'utilisation des SVM binaire à base de noyau RBF, nous avons atteint un taux de réussite de 100% dans le cas d'un classifier SVM construite avec une valeur de Gamma =0.02 du noyau RBF et C=100 quelque soit la taille des segments de test.

Conclusion :

Dans le cadre de ce projet de fin d'études, nous avons étudié et évalué la technique vectorielle de modélisation du locuteur, SVM appliquées à l'identification du locuteur en mode indépendant du texte.

Nous avons testé la technique SVM sur notre base de données de deux façons : Dans la premier façon le paramètre de régularisation C variable et Gamma le paramètre du noyau RBF a été fixée à 0.01et dans la deuxième façon Gamma variable et le paramètre de régularisation C a été fixée à 100, pour des segments de test de taille 1 second et 0 .5 seconde. Les résultats que nous avons obtenus par notre système sont très intéressants pour un C=100, Gamma=0 .02 et Gamma −0.002.

Dans un processus de décision en temps réel, il faudra tenir compte du temps de calcul qui parfois peut devenir élevé. Il faudrait donc avoir une machine puissante pour permettre d'effectuer le calcul très rapidement. ou bien de construire un système qui donne un taux de reconnaissance très élevé avec des segments de test de taille plus court.

Dans l'absence d'une étude théorique sur le choix des noyaux à utiliser, la reprise de tous nos tests avec des noyaux autres que ceux utilisés dans ce travail (noyau linéaire et noyau RBF) peut être intéressante. Ces expériences vont nous permettre de déterminer le noyau le plus adapté à chacune des tâches traitées.

En fin, l'utilisation d'une classifier SVM pour une population de locuteurs assez grande n'est pas pratique en raison du temps gigantesque pris par la phase d'apprentissage. En effet, le temps d'apprentissage est proportionnel au nombre de locuteurs. Pour cela nous avons proposé comme perspectives d'introduire les techniques de réduction de paramètres comme PCA et LDA.

Bibliographie

1. Calliope, la parole et son traitement automatique. Edition Masson, pris, 1989.

2. C. Barras, reconnaissance de la parole continue : Adaptation au locuteur et contrôle temporel dans les modèles de Markov cachés, thèse de doctorat de l'université paris VI, mai 1996.

3. D.A.Reynolds, M.A. Zissman, T.F.Quatieri, G.C.O'leary and B.A. Carlson, the effects of telephone transmission degradation on speaker recognition performance, Massachusetts institute of technology, IEEE, USA, 1995, pp. 329-332.

4. G. Singh, A. Panda, S. Bhattacharyya and T. Srikanthan, vector quantization techniques for GMM based speaker verification, Indian Institute of technology, Kampur, india. IEEE, ICASSP, 2003, pp. 65-68.

5. H. Hadjali, M. Bouchmekh, identification du locuteur indèpendant du texte, thèse d'ingéniorat à l'école nationale polytechnique d'Alger, juin 2004.

6. J. Kharoubi, étude de techniques de classement (Machines à Vecteurs de Supports) pour la vérification automatique du locuteur, thèse de doctorat de l'école nationale supérieure de télécommunication de paris, juin 2002.

7. J. P. Campbell and D. A. Reynolde, corpora for the evaluation of speaker recognition systems, IEEE, 1999, pp. 829-832.

8. M. Bellanger traitement numérique du signal, thèse et pratique, edition Masson, 1987.

9. Y. Mami, reconnaissance du locuteur par localisation dans un espace de locuteurs de référence, thèse de doctorat d'état à l'école nationale supérieure de télécommunication de paris, octobre 2003.

10. Méthodes prédictives neuronales : Applications à l'indentification du locuteur, thèse de l'université de paris XI, orsay 1995.

11. Bimbot F., Paoloni A., Chollet G., Assessment methodology for speaker indentification and vérification systems. Technical report- Task 2500- report 19, SAM- A ESPRIT project 6819. 1993.

12. Grenier Y., identification de locuteur et adaptation au locuteur d'un système de reconnaissance phonétique, thèse de doctorat ingénieur : E. N. S. T. paris 1977.

13. Biométrie online, adresse internet : http: // biométrie. online. Fr /.

14. Wolf J., efficient acoustic parameters for speaker reconition. The journal of the acoustical society of america, pp 2044-2056, N516, 1972.

15. D. GENOUD, M. MOREIRA, and E. MAYORAZ. (text dependent speaker verification using binary classifiers). In PROC. ICASSP, 1998.

16. D. DAOUADI, M. NADIL, (proposition et mise en œuvre d'une approche pour l'amélioration des performances d'un classificateur SVM). PFE, école militaire polytechnique, 2004.

17. B. SCHOLKOPF, (support vector learning). Thesis, R. OLDENBOURG verlag, MUNICH, 1997.

18. T. MATSUI and S. FURUI. (comparison of text independent speaker recognition methods using VQ distortion and discrete / continuous HMM's). IEEE transaction on speaker and audio processing, vol.2, no.3, july 1994.