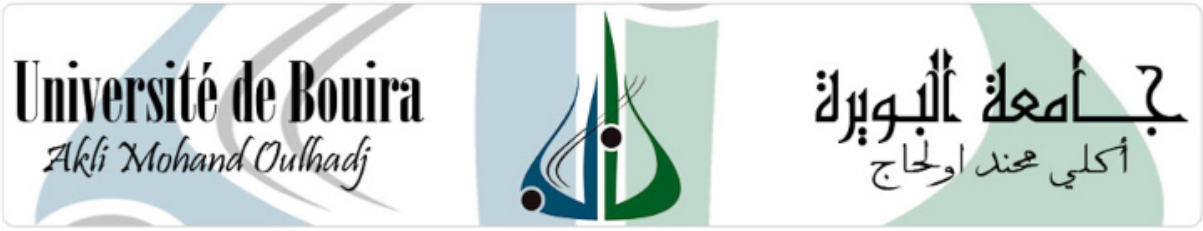
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Akli Mohand Oulhadj (Bouira)



Faculté des Sciences et des Sciences Appliquées
Département de Génie Electrique
Filière de télécommunication

PROJET DE FIN D'ETUDES

En vue de l'obtention du diplôme de master
en systèmes des télécommunications

THEME

Utilisation des paramètres i-vecteur pour la reconnaissance des marques de téléphone

Réaliser par :

- **BOUZIANE Toufik**
- **TOUMI Yacine**

Encadré par :

Dr. ALIMOHAD Abdennour

Année Universitaire : 2017/2018

Remerciements

Au terme de ce travail :

*Nous tenons à remercier tout d'abord **ALLAH** tout puissant et maitre de l'univers qui nous a donné la force nécessaire, la forte volonté et la patience afin d'accomplir ce travail.*

Nous tenons à remercier vivement nos chers parents.

Nous remercions également notre promoteur :

***DR. ALIMOHAD ABDENNOUR** pour l'aide qu'il nous a apporté, pour sa disponibilité, ainsi pour ses précieux conseils.*

Nous tenons à exprimer nos gratitudes à nos enseignants de département Génie électrique de l'université de bouira.

Nous tenons à exprimer nos profonds remerciements aux membres de jury pour leurs participations dans le jugement de notre travail.

Enfin nous adressons nos reconnaissances à tous ceux qui nous ont aidés de prés ou de loin.

Dédicaces

A mes parents :

Grâce à leurs tendres encouragements et leurs grands sacrifices, ils ont pu créer le climat affectueux et propice à la poursuite de mes études.
Aucune dédicace ne pourrait exprimer mon respect, ma considération et mes profonds sentiments envers eux.
Je prie le bon Dieu de les bénir, de veiller sur eux, en espérant qu'ils seront toujours fiers de moi.

A mes frères.
A la famille TOUMI.
Ils vont trouver ici l'expression de mes sentiments de respect et de reconnaissance pour le soutien qu'ils n'ont cessé de me porter.
A tous mes professeurs :

Leurs générosité et leurs soutien m'oblige de leurs témoigner mon profond respect et ma loyale considération.
A ma chère amie qui ma beaucoup encourager.

Tous mes amis et mes collègues :
Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie.

Yacine

Dédicaces

A mes parents :

Grâce à leurs tendres encouragements et leurs grands sacrifices, ils ont pu créer le climat affectueux et propice à la poursuite de mes études.
Aucune dédicace ne pourrait exprimer mon respect, ma considération et mes profonds sentiments envers eux.
Je prie le bon Dieu de les bénir, de veiller sur eux, en espérant qu'ils seront toujours fiers de moi.

A mes frères.
A ma sœur qui ma beaucoup encouraegr
A la famille BOUZIANE
Ils vont trouver ici l'expression de mes sentiments de respect et de reconnaissance pour le soutien qu'ils n'ont cessé de me porter.
A tous mes professeurs :

Leurs générosité et leurs soutien m'oblige de leurs témoigner mon profond respect et ma loyale considération.

Tous mes amis et mes collègues :
Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie.

Toufik

Résumé

La reconnaissance des marques de téléphones est un domaine récent qui traite le problème d'identification et de vérification des marques de téléphones à partir de l'enregistrement. Comme tout système de reconnaissance, il est composé essentiellement d'une phase d'apprentissage et d'une phase test. Notre système comporte les coefficients MFCC comme extracteur de paramètres principaux des téléphones, ensuite la modélisation de chaque type de téléphone par une technique qui s'appelle I-Vecteur, et qui est basée sur les modèles GMM. Enfin, un comparateur réalisé par un algorithme de correspondance qui effectue le calcul de la mesure de similarité entre les signaux de test et les modèles. Les résultats obtenus montrent l'efficacité du système de reconnaissance des marques de téléphones à identifier ces derniers avec des performances acceptables.

Abstract

Cellphone brand recognition is a recent field that addresses the problem of identification and verification of cellphones brands from recorded signals. Like any recognition system, it consists essentially of a training phase and a test phase. Our system has the MFCC coefficients as the main parameter extractor of the phones, then the modeling of each type of phone by a technique called I-Vector, which is based on the GMM models. Finally, a comparator realized by a matching algorithm that calculates of the similarity measure between the test signals and the models. The obtained results show the effectiveness of the brand cellphones recognition system with acceptable performance.

TABLE DES MATIERES

LISTES DES FIGURES

Liste des tableaux

ABRÉVIATIONS

RAP Reconnaissance Automatique de la Parole

RAL La reconnaissance automatique du locuteur

VAL La vérification automatique du locuteur

IAL L'identification Automatique du Locuteur

LPC Les paramètres linéaires prédictifs (linear predictive coding)

PLP Prédiction Linéaire Perceptuel

VAD la détection d'activité vocale

MFCC coefficients cepstral de frequance mel

DFT transformée de Fourier discrète

FFT transformée de Fourier rapide

QV La quantification vectorielle

HMM modèles de Markov cachés

GMM modele de melange de gaussiennes

UBM Universel Background Model

EM Expectation Maximisation

WCCN Within-Class Covariance normalisation

INTRODUCTION GENERALE

La parole est l'un des moyens, les plus directs d'échange de l'information, utilisés par l'homme. Le signal de parole est porteur de plusieurs types d'informations comme le message, la langue, les émotions, ou même l'environnement. Cet avantage a donné naissance à plusieurs travaux de recherche dont l'objectif est la conception des systèmes de reconnaissances. Nous citons à titre d'exemple, la reconnaissance des téléphones portables, la reconnaissance de la parole, la reconnaissance du langage naturel, et la reconnaissance du locuteur. Toutes ces méthodes utilisent l'information véhiculée par le signal vocal. Dans ce projet, nous nous intéressons à la reconnaissance des téléphones portables à partir de leur discours enregistré. En utilisant des signaux vocaux, nous essayons d'identifier la marque et le modèle d'un téléphone portable, par lequel ils sont enregistrés. Le terme marque d'un téléphone portable désigne le fabricant, par ex. Nokia, Condor, etc. et le terme modèle d'un téléphone portable est utilisé pour désigner le type de produit du même fabricant, par ex. Nokia 302, Condor P6, etc [1].

Dans la littérature, il existe des travaux qui tentent d'identifier les caméras sources (appareils photos numériques, téléphones portables) utilisant leurs images enregistrés. D'autres chercheurs se sont penchés à la reconnaissance des marques et modèles de téléphones, qui peuvent être utilisés comme une bonne preuve par le tribunal ou les agents d'application de la loi [1].

Le système de reconnaissance du téléphone constitué de trois parties principales. Tout d'abord, 1' extraction des paramètres acoustiques de chaque téléphone. Ensuite, nous avons le module de modélisation qui va créer un modèle de chaque téléphone à partir des vecteurs de caractéristiques obtenus précédemment. Et enfin, on trouve le module de reconnaissance qui va tester les nouveaux sujets sur les modèles existants et stockés dans une base de donnée, afin de vérifier ou d'identifier le téléphone [2].

Ce mémoire est organisé selon trois chapitres. La définition des systèmes et les étapes fondamentales de la reconnaissance du téléphone, la reconnaissance de locuteur, parole, langage, sont décrits dans le chapitre I. L'identification du téléphone portable et la procédure d'extraction MFCC (coefficients cepstral de frequance mel) et la modélisation i-vecteur sont donnés en détail dans le chapitre II. Enfin, les résultats expérimentaux sont fournis dans le chapitre III.

Chapitre I

Généralité

I.1 INTRODUCTION

La science de la parole a été l'un des domaines de recherche les plus difficiles au cours de cesdernières décennies. La reconnaissance de la parole, la reconnaissance du locuteur, la reconnaissance du langage, la reconnaissance des émotions et la reconnaissance du genre sont les applications les plus populaires. Toutes ces méthodes utilisent l'information véhiculée par le signal vocal. Les systèmes de reconnaissance des locuteurs, par exemple, utilisent les fonctionnalités qui représentent l'identité des locuteurs tandis que dans la reconnaissance de la parole, les informations qui paramètrent le texte sont utilisées. Les deux types de caractéristiques sont extraits des signaux vocaux. Puisque la parole est un signal naturel, la voix peut être caractérisée en tant que biométrique [1].

Dans ce travail, nous abordons leproblème de reconnaissance des téléphones cellulaires à partir des discours et sons enregistrés. Pour être plus précis, en utilisant des signaux vocaux, nous essayons d'identifier la marque et le modèle d'un téléphone portable, par lequel il estenregistré [1].

I.2 LA PAROLE

La parole est le mode de communication le plus naturel. Grâce à elle nous pouvons donner une voix à notre volonté et à nos pensées. Nous pouvons l'utiliser pour exprimer des opinions, des idées, des sentiments, des désirs ou pour échanger, transmettre, et demander des informations. Aujourd'hui, nous ne l'utilisons pas uniquement pour communiquer avec d'autres humains, mais aussi avec des machines [3].Cet avantage a donné naissance à plusieurs travaux de recherche dont l'objectif est la conception de système permettant de reconnaître la séquence des mots parlés [4].

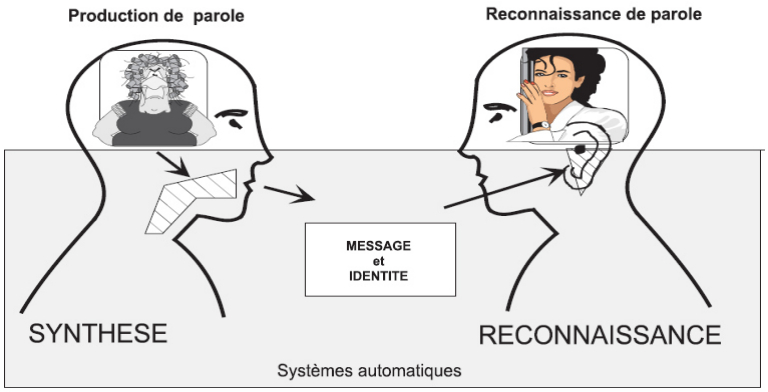


Figure I.1 : Production et reconnaissance de la parole.

I.2.1 SEGMENTATION DE LA PAROLE

La segmentation de la parole est un ensemble de procédés d'identification d'unités variées dans un signal de parole selon la nature du segment considéré. Selon cette nature, on distingue les formes de segmentation suivantes :

- la segmentation en sons voisés ou non.
- la segmentation en phonèmes.
- la segmentation en syllabes.
- la segmentation en mots.
- la segmentation en locuteurs.

La tâche de segmentation de la parole est indispensable pour l'apprentissage des modèles acoustiques d'un système de reconnaissance de la parole et de synthèse vocale.

Il existe deux formes de segmentation : **la segmentation manuelle** (dépendant du corpus) et **la segmentation automatique** (indépendant du texte). Les grandes bases de données des langues moyennement et très bien dotées (Français, anglais ou l'italien) disposent déjà d'un étiquetage phonétique des signaux audio : ce qui n'est pas le cas des langues peu dotées. Il est donc bien indiqué une segmentation automatique de bonne précision pour des bases de données non annotés car elle accélère les procédures de vérification manuelle qui prend suffisamment de temps et coûtent chères. Malgré le coût élevé en temps et en argent, la segmentation manuelle est beaucoup plus précise que la segmentation automatique [5].

1.2.2 PRODUCTION DU SIGNAL DE PAROLE

Le signal de parole est le résultat de l'excitation du conduit vocal par un train d'impulsions ou par un bruit.Ces deux excitations donnent lieu respectivement aux sons voisés et non voisés. Dans le cas des sons voisés, l'excitation est une vibration périodique des cordes vocales suite à la pression exercée par l'air provenant de l'appareil respiratoire. Ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale F_0. Quant aux signaux non-voisés, l'air passe librement à travers la glotte (du moins pas dans tout le conduit vocal) sans provoquer de vibration des cordes vocales [6].

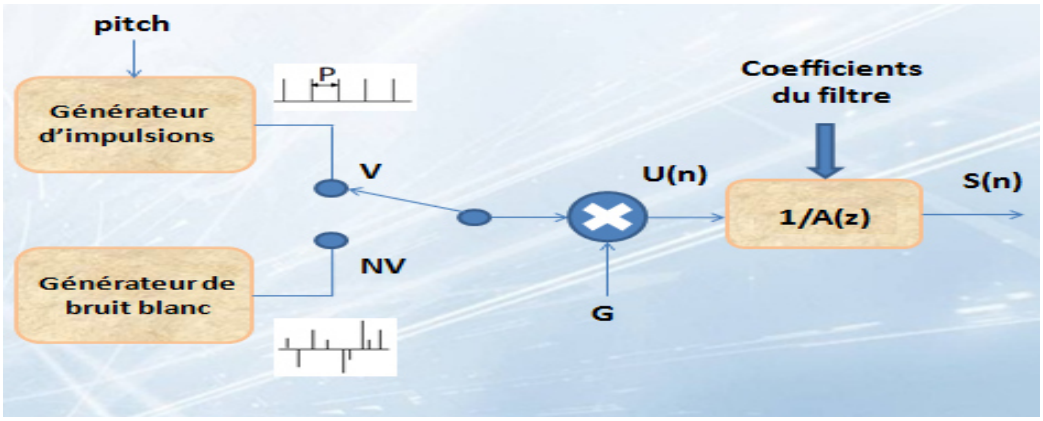


Figure I.2 : Modèle simple de production de la parole.

1.3. LE SIGNAL DE PAROLE

Le signal de parole peut être aussi considéré comme un vecteur acoustique porteur d'informations d'une grande complexité.

1.3.1. PARAMETRES DU SIGNAL DE PAROLE

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

- **Fréquence fondamentale :**Elle représente la fréquence du cycle d'ouverture et de fermeture des cordes vocales. Cette fréquence caractérise seulement les sons voisés.

- **Energie :** Elle est représentée par l'intensité du son qui est liée à la pression de l'air en amont du larynx. L'amplitude du signal de la parole varie au cours du temps selon le type de son.
- **Spectre :**L'enveloppe spectrale ou spectre représente l'intensité de la voix selon la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme.

La quasi stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal vocal sur des fenêtres de durée généralement comprise entre *20ms* et *30ms* appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse[7].

1.4 SYSTEMES DE RECONNAISSANCES BASES SUR LA PAROLE

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage. L'importance particulière du traitement de la parole dans ce cadre plus général s'explique par la position privilégiée de la parole comme vecteur d'information [8]. Il regroupe l'analyse / synthèse de la parole, la reconnaissance et le codage (figure I.3)

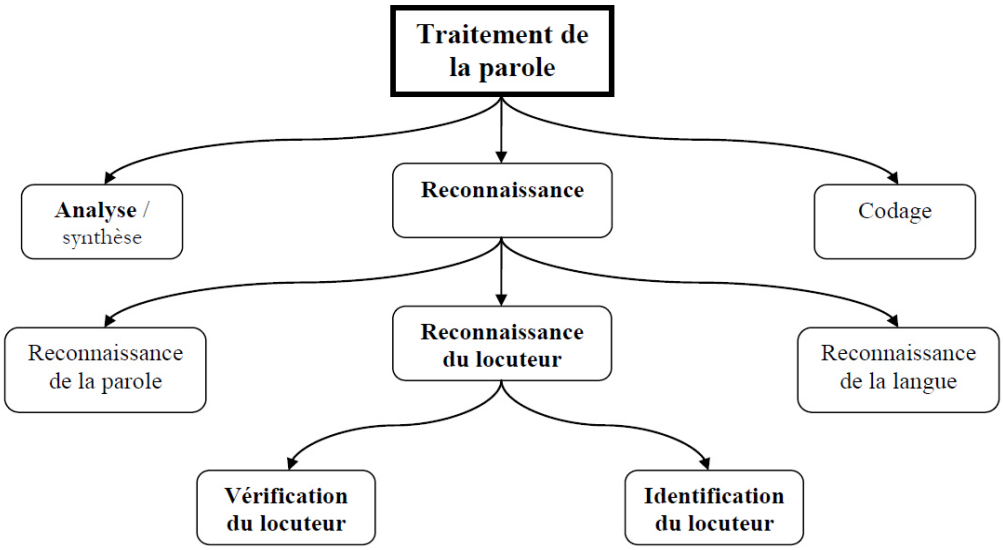


Figure I.3 : Domaines de traitement de la parole.

I.5 SYSTEMES DE RECONNAISSANCE DE LA PAROLE

La reconnaissance de la parole est la technique qui permet l'analyse des sons captés par un microphone pour les transcrire sous forme d'une suite de mots.Ceci, permet à la machine de comprendre et de traiter les informations fournies oralement par un utilisateur humain [3]. Depuis son apparition dans les années 1950, la reconnaissance automatique de la parole a été constamment améliorée avec l'aide des phonéticiens, linguistes, mathématiciens et ingénieurs, qui ont défini les connaissances acoustiques et linguistiques nécessaires pour bien comprendre la parole d'un humain.

Un système de Reconnaissance Automatique de la Parole (RAP) est un système qui a lacapacité de détecter la parole et de l'analyser dans le but de générer une chaîne de mots ouphonèmes représentant ce que la personne a prononcé. Cette analyse se fonde sur l'extractiondes paramètres descriptifs de la parole [4].

Il existe des systèmes de reconnaissance de parole indépendants du locuteur pour des vocabulaires limités et qui offrent des performances intéressantes à condition que l'environnement ne soit pas trop corrompu par le bruit et les interférences. Toutefois, l'interrogation de bases de données à distance, (téléphone, radio, etc.) nécessite de concevoir des systèmes polyvalents et peu sensibles aux conditions et aux environnements d'opérations ainsi qu'aux bruits ambiants (téléphone cellulaire, cabine téléphonique, voitures, avions, camions, etc.). Or, le traitement de la parole en milieu bruyant est loin d'être résolu et les systèmes de reconnaissance actuels ne peuvent fonctionner dans de tels environnements sans observer des dégradations très significatives des performances.Ce qui les rend à toute fin pratique non utilisables par le grand public [9].

I.5.1 LES MODELES MISES EN ŒUVRE DANS LE SYSTEME DE RECONNAISSANCE DE LA PAROLE

- **Analyse du signal audio**

Le signal audio est enregistré à l'aide d'un ou plusieurs microphones, dont la position et la qualité sont cruciales pour la performance du système de reconnaissance vocal. La présence de plusieurs microphones placés à différents endroits peut servir à

localiser le locuteur, à d'ébruiter le signal, à améliorer la performance de reconnaissance [3].

- **Modèle de langage**

Les systèmes de contrôle vocal, dont la taille du vocabulaire est petite, utilisent souvent la reconnaissance des mots isolés, ou des formes phonétiques simples. Dans ce cas, les systèmes de reconnaissance automatique de la parole dépendent peu du modèle de langage. Un décodeuracoustique-phonétique peut alors atteindre un taux de reconnaissance très élevé [10].

- **Modèle de prononciation**

Le lexique d'un système de reconnaissance vocale précise une ou plusieurs prononciations pour chaque mot. Pour le français, les prononciations multiples sont en partie dues aux événements de liaison ou de réduction, dans le cadre desquels un locuteur peut prononcer ou pas un certain phonème dans un certain contexte. Les accents et les dialectes peuvent aussi générer diverses variantes de prononciations.

- **Modèle acoustique**

Le modèle acoustique est un modèle statistique qui estime la probabilité qu'un phonèmeait généré une certaine séquence de paramètres acoustiques. Une grande variété deséquences de paramètres acoustiques sont observées pour chaque phonème en raison de toutes les variations liées à la diversité des locuteurs, à leur âge, leur sexe, leur dialecte, leur état de santé, leur état émotionnel, etc.[3].

I.6 LA RECONNAISSANCE AUTOMATIQUE DU LOCUTEUR

La reconnaissance automatique du locuteur (RAL) est généralement décrite comme l'identification d'une personne à partir de sa voix. Il est possible de classifier les systèmes de reconnaissance automatique du locuteur selon leur dépendance au texte prononcé, la tâche ciblée ou leur domaine d'application [7]. La reconnaissance du locuteur est probablement la plus naturelle et économique pour les systèmes de communication homme-machine parce que d'une part la collecte de données de parole est beaucoup plus pratique que les autres motifs, et

d'autre part, la parole est le mode dominant d'échange d'information pour les êtres humains et tend à être le mode dominant pour l'échange d'information pour les systèmes de communication homme-machine [11].

I.6.1 STRUCTURE D'UN SYSTEME DE RAL

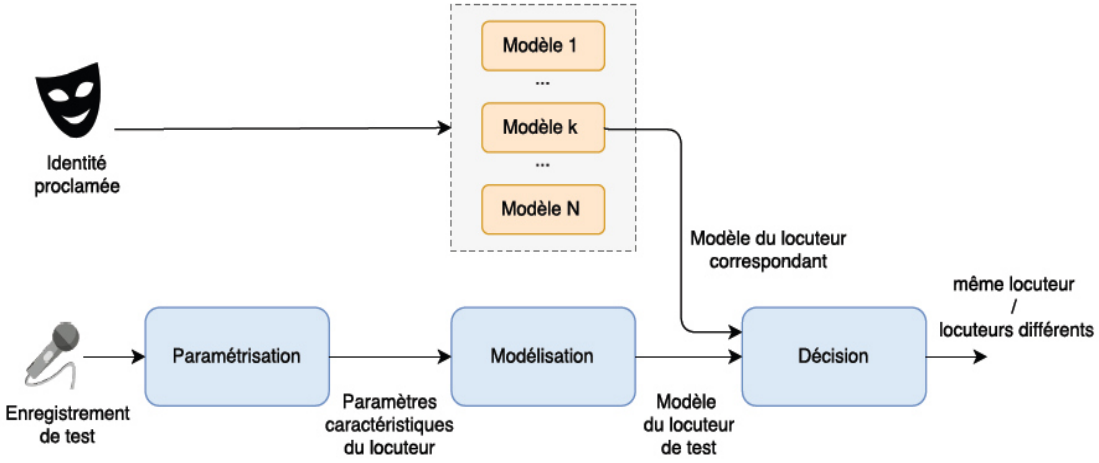


Figure I.4 : Structure d'un système de vérification du locuteur.

- **Paramétrisation :** Cette étape vise à capturer des paramètres caractéristiques dela parole d'une personne donnée. Suite à de nombreux travaux de recherche, il s'est avéré que les paramètres basés sur la représentation spectrale de la parole sont les plus pertinents pour la RAL.Ces paramètres sont corrélés à la forme du conduit vocal et sont les plus utilisés dans les systèmes de RAL modernes. Cependant, les paramètres prosodiques qui décrivent le style de parole du locuteur sont aussi utilisés en pratique.
- **Modélisation :** Les paramètres acoustiques extraits d'un enregistrement donné sont utilisés pour construire un modèle qui résume l'information acoustique correspondante.
- **Décision :** La phase de décision désigne l'identité du locuteur reconnu. Dans le cas de la vérification, cette décision est binaire et consiste à confirmer ou infirmer la correspondance de la session de test à une identité proclamée. Vu qu'il est impossible d'avoir une similarité de 100% entre le signal du locuteurde test et celui des locuteurs clients, les modèles sont conçus de telle sorte qu'une telle comparaison fournisse un score (une valeur scalaire) indiquant si les deux énoncés correspondent au même locuteur. Si ce score est

supérieur (ou inférieur) àun seuil prédéfini, le système accepte (ou rejette) le locuteur de test. [7]

I.6.2 IDENTIFICATION ET VERIFICATION AUTOMATIQUE DE LOCUTEUR

La vérification automatique du locuteur (*VAL*) permet de décider si l'identité revendiquée par un locuteur est compatible avec sa voix. Il s'agit donc de trancher entre deux hypothèses : soit le locuteur est bien le locuteur autorisé (on l'appelle aussi locuteur client), c'est-à-dire que son identité correspond à celle revendiqué, soit le locuteur est un imposteur qui cherche à se faire passer pour la personne qu'il n'est pas. À partir d'un échantillon de voix de référence, et d'un échantillon de voix de test, le système va donc devoir dire si oui ou non les deux locuteurs correspondent. Les systèmes de *VAL* sont très dépendants des différences entre les échantillons de voix de référence et les échantillons de tests. Accepter un locuteur qui devrait être rejeté peut avoir de lourdes conséquences, en particulier dans les applications où un haut niveau de sécurité est demandé (contrôle aux frontières, système bancaire, identification judiciaire, etc.) [7].

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. Cela est fait en calculant des mesures de similarité entre le signal en entrée et tous les modèles des locuteurs de la base [12].Deux modes d'identification sont possibles :

- **Identification en ensemble fermé :** c'est le cas où le système doit fournir comme sortie un ensemble d'au moins un Locuteur. En d'autres termes, la séquence fournie en entrée doit être en fait prononcée par un Locuteur connu du système.

- **Identification en ensemble ouvert :** le système dans ce cas peut être amené à fournir un ensemble vide, car le Locuteur peut ne pas être connu.

Dans ce mode, le système d'IAL doit décider de la fiabilité de son jugement en acceptant ou rejetant l'identité qu'il a trouvée. [12]

I.7 SYSTEME D'IDENTIFICATION AUTOMATIQUE DES LANGUES

Les sources d'information discriminantes pour l'identification des langues sont liées à quatre domaines linguistiques :

- ✓ **La phonologie :** les espaces acoustiques des langues sont différents, les inventaires phonétiques sont distincts suivant les langues. Même s'il existe des recoupements pour certaines langues, les fréquences d'apparition des phonèmes peuvent être utilisées comme des caractéristiques. De plus, les règles d'enchaînements (phonotactiques) de ces unités varient d'une langue à l'autre.
- ✓ **La morphologie:** les lexiques sont différents suivant les langues, chaque langue à son propre vocabulaire et sa propre manière de former les mots.
- ✓ **La syntaxe :** les phrases sont structurées différemment selon les langues.
- ✓ **La prosodie :** le rythme, l'intonation et l'accentuation varient suivant les langues.

Dans une perspective de l'augmentation du nombre de langues à reconnaitre par les systèmes d'identification, il est important de prendre en compte le maximum de sources possible [13]. La structure générale d'un système d'identification automatique des langues est illustrée à la figure I.5 **:**

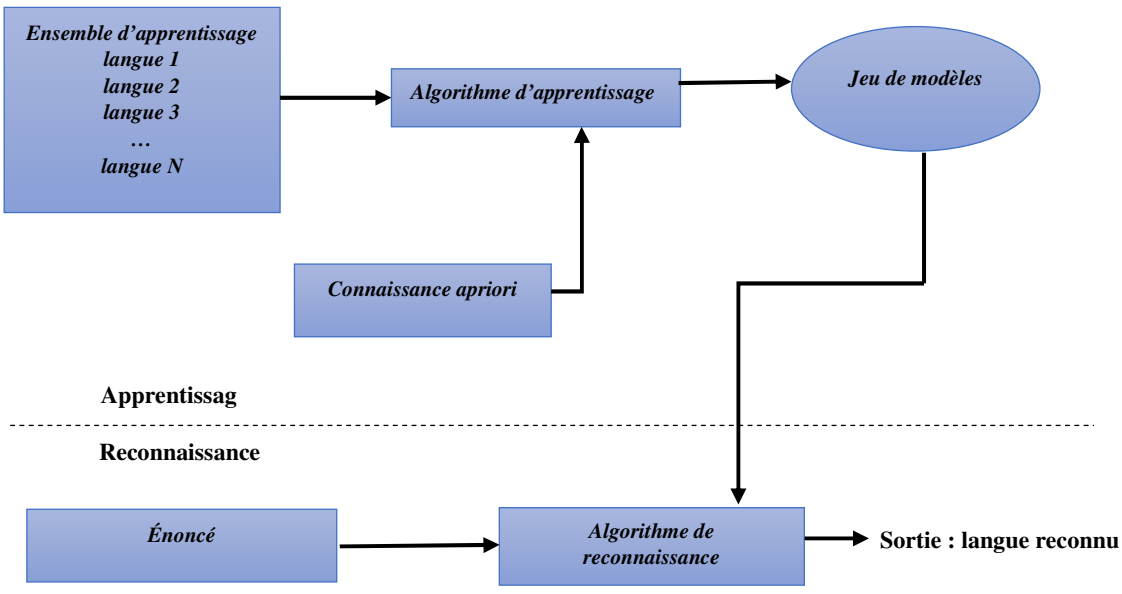


Figure I.5 : Structure générale d'un système d'identification automatique des langues [13].

I.8 LA RECONNAISSANCE DU TELEPHONE PORTABLE

I.8.1 CONTEXTE DE TRAVAIL

Le développement de la technologie numérique a conduit au développement d'outils portables à bas prix, tels que les caméras de poche, les machines à dicter, les téléphones cellulaires et les téléphones intelligents qui font partie intégrante de notre vie quotidienne. Ces outils sont utilisés pour l'enregistrement et la transmission de données multimédias qui jouent un rôle de plus en plus important en tant qu'éléments de preuve dans l'investigation médico-légale. Ainsi, il existe un besoin croissant d'une analyse et d'une classification précise des données multimédias médico-légales. L'analyse multimédia s'est largement concentrée sur les images numériques [14].

La détermination de l'intégrité et de l'authenticité d'une image, l'identification de la caméra source avec laquelle l'image a été prise, le tatouage numérique d'images et de vidéos et le problème de détection de la présence de messages cachés dans les images sont des applications courantes. Parmi ceux-ci, la reconnaissance de la source de caméra est l'un des problèmes les plus difficiles. Étant donné une image, la tâche consiste à déterminer le périphérique source avec lequel l'image a été prise. Ceci est possible parce que les imperfections dans les dispositifs d'acquisition laissent leurs empreintes spécifiques à l'appareil aux images. Par exemple, l'identification des caméras numériques basées sur le bruit de modèle de capteur et les impuretés a été proposée. Une approche similaire a été utilisée pour identifier les sources de scanners à partir d'images numérisées en fonction de la poussière, de la saleté et des rayures sur le plateau du scanner. Lorsqu'un échantillon vocal enregistré apparaît comme preuve médico-légale, il est souvent nécessaire de tracer le dispositif d'enregistrement ou l'environnement [14].

I.8.2 SYSTEME DE RECONNAISSANCE DE TELEPHONE PORTABLE :

Comme dans tous les types de systèmes de reconnaissance, la reconnaissance de téléphone portable peut être utilisée comme un terme générique qui se réfère à deux tâches différentes : l'identification et vérification de téléphone portable. Dans la tâche de vérification, une revendication d'identité est donné au système comme une entrée et le système accepte ou rejette la revendication d'identité donnée (figure I.6.b). Cependant, l'identification englobe deux types d'applications appelées identification fermée et identification ouverte. Dans

l'identification en circuit fermé, le but est de faire correspondre l'entrée inconnue un des échantillons vocaux enregistré à partir d'un ensemble de N téléphones cellulaires a priori connus (voir figure I.6.a). Dans ce cas l'entrée est supposée correspondre forcément à l'un des échantillons enregistrés. Le système d'identification dans un ensemble ouvert vise à détecter d'abord si l'entrée vient ou non de l'ensemble de téléphone portable connu a priori du système. Puis établir la correspondance de cette entrée avec l'un des téléphones de l'ensemble. Le système décidera d'accepter ou de rejeter l'entrée. Dans notre travail nous considérerons le problème d'identification dans un ensemble fermé, dans lequel nous identifions la marque et le modèle d'un téléphone cellulaire inconnu en utilisant les échantillons de la voix enregistrés à partir d'un ensemble de N téléphones portables [14].

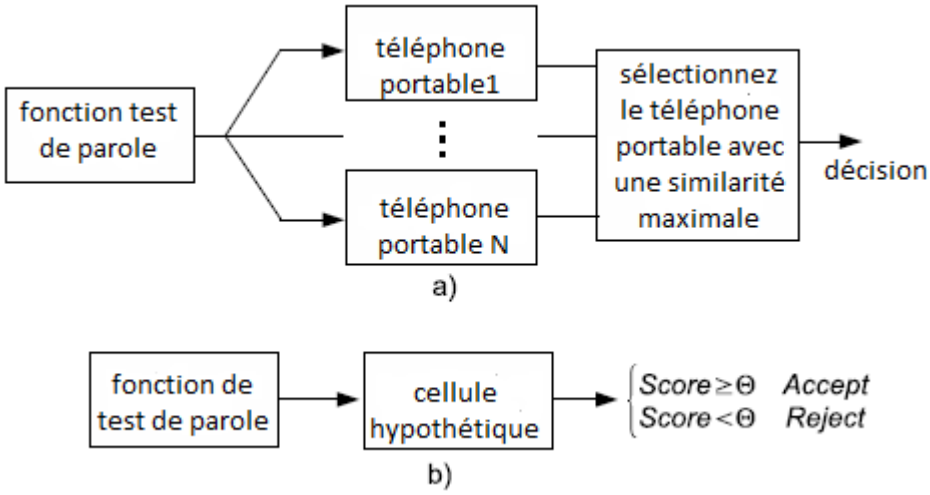


Figure I.6 : Logique de décision pour (a) l'identification du téléphone portable et (b) la vérification du téléphone portable**.**

I.8.3 ETAPES D'UN SYSTEME DE RECONNAISSANCE DE TELEPHONE PORTABLE

Comme tout système de reconnaissance de formes, un système de reconnaissance de la marque du téléphone portable peut travailler en deux modes : appelés l'apprentissage et le test(figure I.7). La représentation et le prétraitement des signaux de parole sont des étapes communes aux deux modes de travail. Dans le mode d'apprentissage, le vecteur de paramètres représentant le signal de parole d'un téléphone sert à produire le modèle du téléphone en utilisant des algorithmes de classification appropriés. Les modèles résultants de l'apprentissage seront ensuite enregistrés dans une base de données. Dans le mode de test, le

vecteur de paramètres correspondant à un téléphone inconnu sera comparé à un ou plusieurs modèles et marque de la base de données. Ainsi, sur la base des scores de comparaisons faites, le système de reconnaissance décidera pour que le téléphone appartienne cette voix [15].

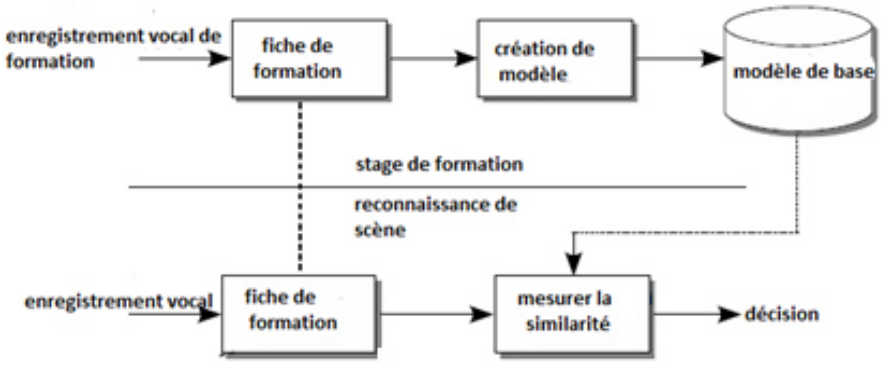


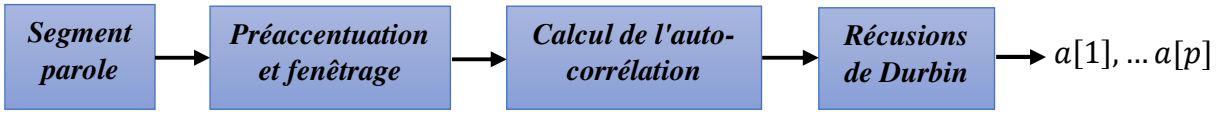
Figure I.7 : Système de reconnaissance générique de téléphone cellulaire.

La fiche de formation est l'élément clé du système. Plusieurs paramètres ont été utilisés dans ce domaine. Les paramètres linéaires prédictifs (linear predictive coding LPC), autrement appelés vocodeurs. Sont des codeurs paramétriques modélisant le principe de production de la parole humaine.

Il s'agit d'un codeur bas débit basé sur le modèle de la production de la parole présenté par Fant (Fant1960). Selon ce modèle, le signal de parole$s(n)$ peut être modélisé comme suit :

$$s(n) = \sum_{k=1}^{p} a_k \, s(n-k) + Gu(n) \qquad (I.1)$$

OùGest le gain du filtre LPC et $u(n)$un signal d'excitation. Celui-ci diffère suivant le type de signaux, sons voisés ou non voisés. Un son est en effet dit voisé lorsque sa production nécessite la vibration des cordes vocales, dans le cas contraire, il est qualifié de non voisé. [16]. Une technique de calcul des coefficients LPC utilise la méthode d'auto-corrélation comme montré dans la figure I.8.



FigureI.8 : Calcul des coefficients LPC par la méthode d'auto-corrélation.

La méthode de coefficients de prédiction est basée sur les notions psycho acoustiques, elle est connue sous le nom de Prédiction Linéaire Perceptuel PLP(PerceptuallybasedLinearPrédiction), est une méthode inspirée du principe de prédiction linéaire (LPC). Elle combine ce principe à une représentation du signal qui suit l'échelle humaine de l'audition. La figure I.9 résume le principe de la méthode dont une analyse spectrale est effectuée au signal parole afin d'obtenir un spectre suivant une échelle d'audition. Ce spectre est ensuite modifié par une interpolation et une transformée de Fourier inverse, le signal obtenu étant passé dans un filtre pour réduire la dimension du spectre et augmenter la résolution fréquentielle [17].

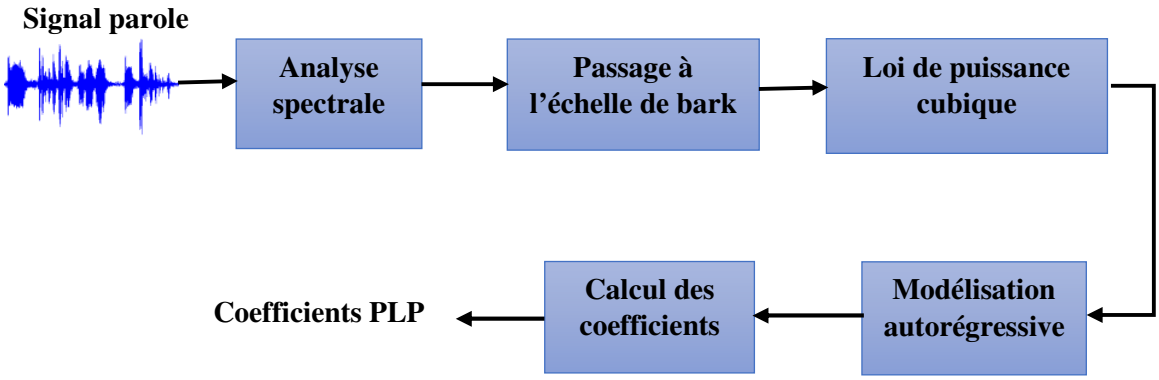


Figure I.9 : Méthode de calcul des coefficients PLP.

Une autre techniques d'extraction de paramètre appelée MFCC (Mel Frequency Cepstral Coefficient) consiste à extraire les informations utiles qui caractérisent le mieux un téléphone portable à partir d'un ensemble d'échantillons vocaux donnés sera détaillée dans le chapitre suivant.

En plus de l'extraction de paramètres une autre étape consiste à générer le modèle de téléphone portable qui sera utilisécomme référence.

L'algorithme de correspondance effectue le calcul de la similarité mesuré et fait des comparaisons entre les modèles.

I.9 CONCLUSION :

Dans ce chapitre nous avons commencé par passé en revue des notions de base de la parole. Puis nous avons décrit les principesgénérauxdes systèmes de reconnaissances basées sur la parole, à savoir, la reconnaissance automatique de la parole, la reconnaissance du locuteur, la reconnaissance du langage.Aussi,nous avons abordé un problème récent lié à la reconnaissance de la marque et du modèle d'un téléphone portable.Le chapitre suivant sera,exclusivement, consacré à lareconnaissance de téléphone portable dans lequel nous détaillerons les parties constituant ce système de reconnaissance.

Chapitre II

Système de reconnaissance du téléphone

II.1. INTRODUCTION

Le système de reconnaissance de la marque du téléphone s'intéresse dans son fonctionnement aux caractéristiques particulières du signal de parole. Cette discipline s'inscrit dans le cadre général de la reconnaissance des formes, c'est un terme générique qui regroupe les problèmes relatifs à l'identification ou à la vérification de la marque du téléphone sur la base de l'information contenue dans le signal acoustique, où il est question de reconnaître la marque de téléphone à partir de leur discours enregistré. Le champ d'application est très vaste, il va du simple contrôle d'accès, aux applications militaires passant par des applications judiciaires. Un système de Reconnaissance de la marque du téléphone opère en trois étapes : l'analyse acoustique du signal de parole, la modélisation et une dernière étape de décision [12]. La figure II.1 illustre ces étapes.

Dans ce chapitre on va explorer les procédures nécessaires à la reconnaissance de la marque du téléphone portable dans le sens large, comme le prétraitement de signal parole, les techniques d'extraction de paramètres et les différents modèles utilisés (La quantification vectorielle (QV),modele de melange de gaussiennes (GMM), modèles de Markov cachés (HMM), i-vecteur). On se basera ensuite dans ce travail sur le domaine des i-vecteurs qui se base sur les GMM et les super-vecteurs.

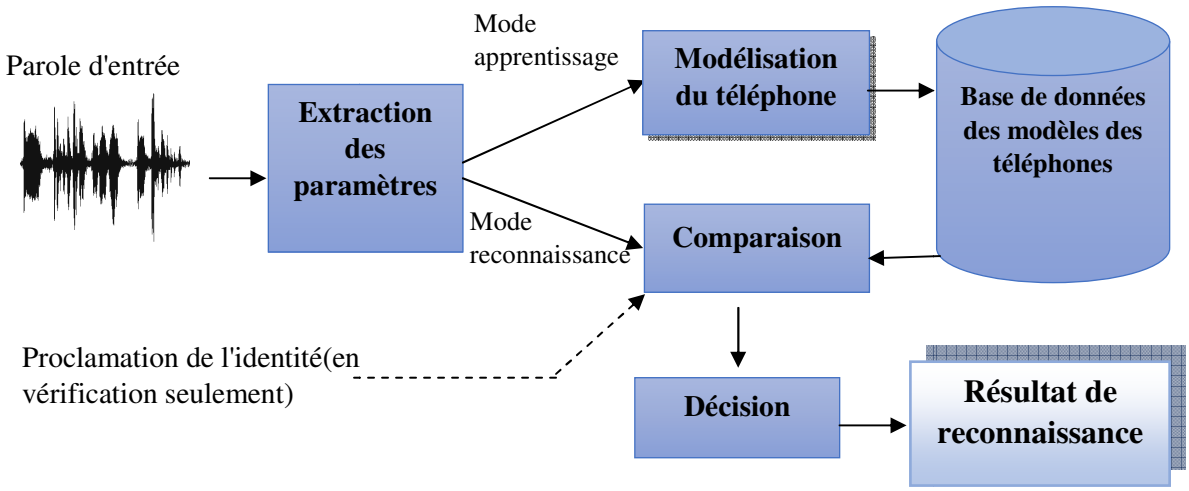


Figure II.1 : Système de reconnaissance du téléphone.

II.2.RECONNAISSANCE DE LA MARQUE DU TELEPHONE PORTABLE

Dans ce paragraphe nous détaillerons les différentes fonctions d'un système de reconnaissance de la marque du téléphone.

II.2.1.EXTRACTION DES PARAMETRES

De par sa production, la parole est variable dans le temps. Les modèles de paramétrisation sont, par conséquent, eux aussi variables dans le temps. L'estimation de ces paramètres nécessite, dans ce cas, une analyse à court terme. Le signal de parole change tous les 10 à 30 millisecondes. Dans cet intervalle, le signal est supposé rester stationnaire, un vecteur de paramètres est extrait durant des segments de temps courts appelées trames. En général, avant de procéder à d'autres étapes d'extraction de paramètres, le signal de parole subit des prétraitements qui incluent :

a. Détection de la parole

Le signal de parole peut contenir du silence dans différentes positions comme le début du signal, entre les mots d'une phrase, à la fin du signal, etc. Ce silence n'apporte aucune contribution à la reconnaissance du téléphone. Il doit être supprimé avant la poursuite des traitements. Plusieurs méthodes sont utilisées pour réaliser cette objective. La plus connue est la détection d'activité vocale (Voice activitydétection VAD), surtout dans les transmissions téléphoniques où le silence constitue la grande partie du temps de transmission, dans chaque direction. Typiquement, la VAD exploite deux types de caractéristiques : (a) la différence spectrale entre le bruit et la parole et (b) les variations temporelles de l'énergie en court terme. Dans la technique VAD, la première étape consiste à calculer les énergies de toutes les trames, puis à sélectionner la valeur maximale. Le seuil de détection est, alors, fixer au-dessous de ce maximum. Un autre seuil est nécessaire pour annuler les trames ayant une énergie absolue faible [15].

b. Segmentation et chevauchement

Le but de la segmentation est de découper le signal de parole en petites tranches (chacune de durée 10 à 30 ms) où il peut être considéré localement comme quasi-stationnaire.

En outre, et pour profiter de l'évolution lente du signal vocal, la segmentation permet le traitement en temps réel et facilite l'analyse des signaux sur la machine. Les ressources d'une machine étant limitées, le signal ne peut pas être traité dans sa globalité [15].

Si la trame est de longueur N échantillons, telles que les trames adjacentes sont séparées par un nombre d'échantillons (de 5 à 15 millisecondes environ), cela est illustré à la figure II.2.

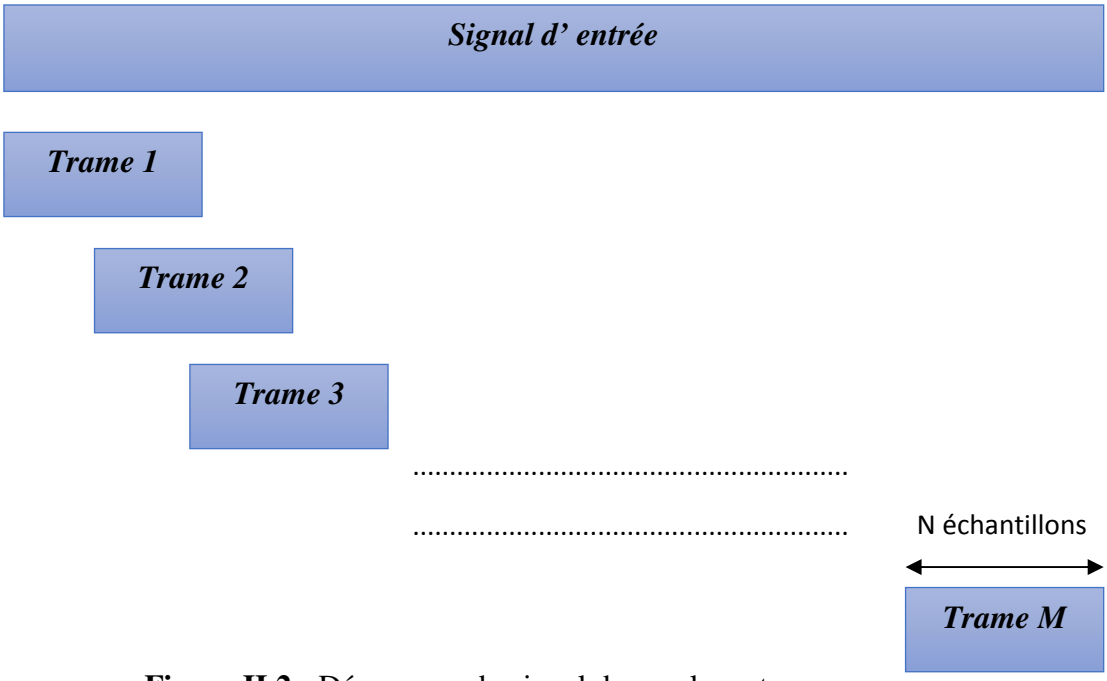


Figure II.2 : Découpage du signal de parole en trames.

II.2.2. COEFFICIENTS CEPSTRAL DE FREQUANCE MEL (MFCC)

Le coefficient MFCC est l'une des techniques d'extraction de caractéristiques la plus populaire utilisée dans la reconnaissance vocale. Cette technique est basée sur le domaine fréquentiel en utilisant l'échelle Mel. Cette échelle reproduit l'échelle de l'oreille humaine. Les MFCC sont considérés comme des caractéristiques du domaine fréquentiel [18]. Ils sont beaucoup plus précis que les caractéristiques de domaine temporel. Les paramètres extraits du discours sont similaires à ceux utilisés par les humains pour entendre un discours, tout en désaccentuant les autres informations. Le signal vocal est d'abord divisé en intervalles de temps constitués d'un nombre arbitraire d'échantillons [18]. Les principales étapes impliquées dans les coefficients MFCC sont montrées dans la figure II.3.

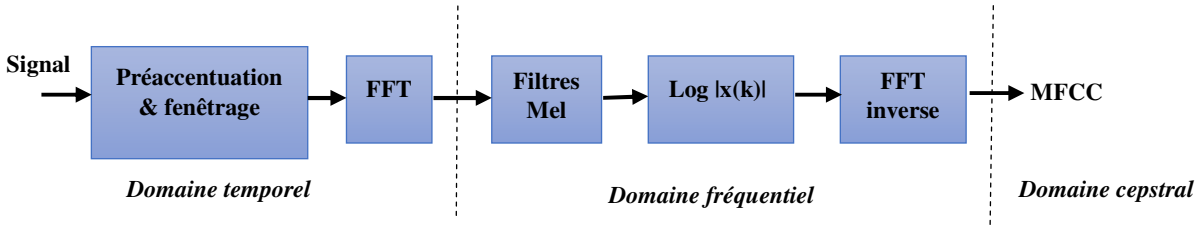


Figure II.3 : Les étapes d'une paramétrisation MFCC.

L'acquisition du signal de parole constitue la première étape à franchir. Il s'agit de numériser un signal analogique (la parole) pour qu'il soit prêt à des traitements numériques ultérieurs. Cette étape est généralement réalisée à l'aide d'une carte d'acquisition spécialisée.
Une fois capté par un microphone, le signal est tout d'abord filtré, ensuite échantillonné et enfin quantifié. Ces opérations successives permettent de transformer un signal continue x(t) (où t désigne le temps) en un signal numérique x(n) où n correspond à des instants discrets [19].
Le signal est segmenté en trames où chaque trame est constituée d'un nombre fixe K d'échantillons de parole. En général, K est fixé de telle manière que chaque trame correspond à environ 30 ms de parole. Cette segmentation est réalisée à l'aide de fenêtres temporelles glissantes, de taille 256 ou 512 points. En général, les fenêtres successives se recouvrent de moitié de leur taille soit de 128 ou 256 points respectivement.

Le découpage du signal en trames produit des discontinuités aux frontières des trames. Pour réduire ces problèmes, des fenêtres de pondération sont appliquées. Ce sont des fonctions que l'on applique à l'ensemble des échantillons prélevés dans la fenêtre du signal original de façon à diminuer les effets de bord. Parmi les fenêtres les plus courantes, nous pouvons citer la fenêtre de Hamming (figure II.4) [19] :

$$w[k] = \begin{cases} 0.54 + 0.46\cos\left(\frac{2\pi k}{N-1}\right) si & 0 \le k \le N-1 \\ 0 & sinon \end{cases} \qquad \text{(II. 1)}$$

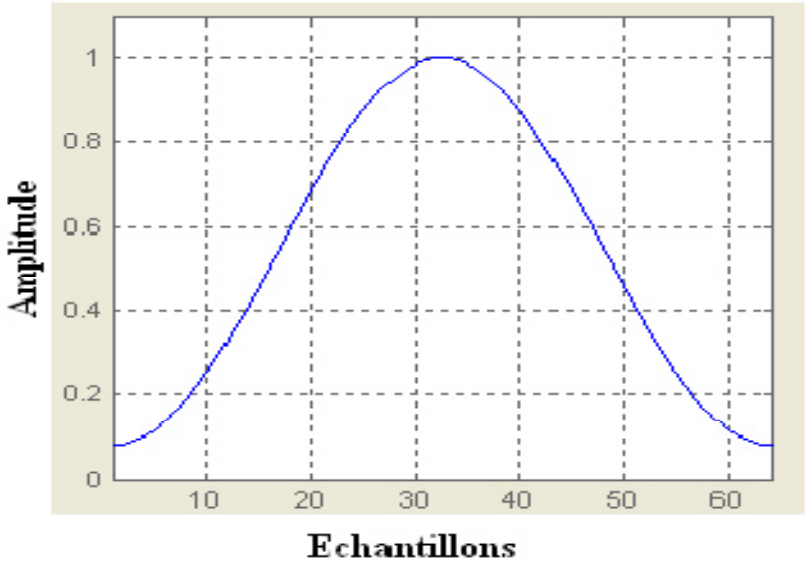


Figure II.4 : Fenêtre de Hamming

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), des analyses temporelles ou spectrales peuvent être appliquées au signal. L'énergie du signal est le paramètre le plus intuitif utilisé pour caractériser le signal de parole. Elle correspond à la puissance du signal et elle est calculée directement dans le domaine temporel sur une trame de parole [19].

L'analyse spectrale reste cependant le moyen le plus utilisé pour caractériser le signal de parole. Elle permet de mettre en évidence certains phénomènes caractéristiques de la production de ce dernier. Les spectrogrammes ont été utilisés pour représenter la parole dès les années 40 en utilisant des filtres analogiques. Actuellement, les spectres sont obtenus numériquement par une transformée de Fourier discrète (DFT : Discret Fourier Transform), en particulier grâce à l'algorithme de transformée de Fourier rapide (FFT : Fast Fourier Transform). Une fois appliquée, cette transformée nous permet de passer du domaine temporel au domaine fréquentiel. La formule de la DFT est donnée par l'équation suivante [19] :

$$X(n) = \frac{1}{N}\sum_{K=0}^{N-1} x(k)e^{-jk2\pi(\frac{n}{N})} \qquad \text{(II. 2)}$$

Le nombre de paramètres spectraux calculés sur une trame par la FFT reste trop élevé pour un traitement automatique ultérieur.

L'énergie du spectre est donc calculée à travers un banc de filtres numériques couvrant la bande passante, ce qui permet de ne conserver qu'un sous ensemble de ces paramètres. Les filtres triangulaires sont les plus utilisés. Ils sont préférés pour leur simplicité et leur effet de lissage sur le spectre. Ces filtres sont les plus souvent répartis sur l'échelle Mel qui est non linéaire. La relation entre la fréquence en échelle Hertz et sa correspondance en Mels est la suivante :

$$Mel(f) = x.\log\left(1 + \frac{f_{Hz}}{y}\right) \text{(II. 3)}$$

Où f_{Hz} est la fréquence, x = 2595 et y = 700. L'intérêt de l'échelle Mel est qu'elle est assez proche d'échelles issues d'études sur la perception sonore et sur les bandes passantes critiques de l'oreille [19].

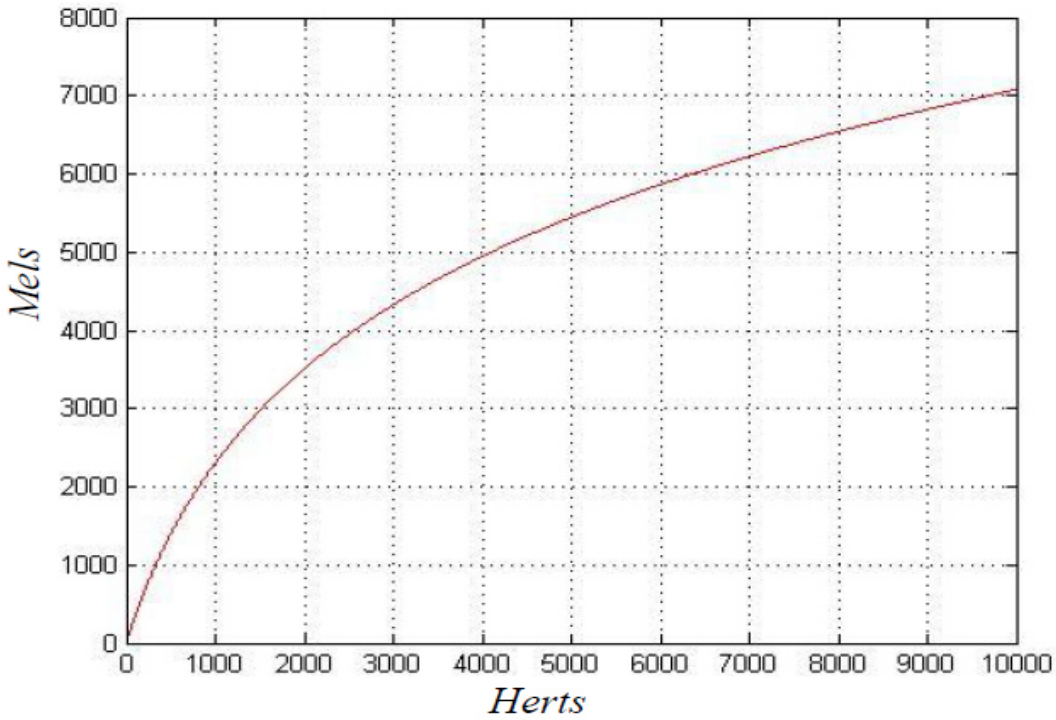


Figure II.5 : Echelle de Mel.

II.2.3 AVANTAGES ET INCONVENIENTS DES MFCC

a. AVANTAGES

Les coefficients de fréquence de Mel sont basés sur les variations connues des largeurs de bande critiques de l'oreille humaine avec des fréquences inférieures à 1000 Hz. Le principal objectif du traitement MFCC est de reproduire le comportement des oreilles humaines, et de capturé les principales caractéristiques des téléphones dans la parole avec faible complexité [20].

Les coefficientsMFCC sontutilisé pour les tâches de traitement de la parole, afin d'obtenir une précision de reconnaissance élevée. Cela signifie que le taux de performance des coefficients MFCC dans le système de reconnaissance basés sur la parole est meilleur [20].

b. INCONVENIENTS

Le bruit est assimilé à une perturbation sonore, dans ces conditions les coefficients ne donnent pas des résultats précises et les performances du système de reconnaissance sont sévèrement influencées. Comme la bande passante du filtre n'est pas un paramètre de conception indépendant.Les performances peuvent être affectées par le nombre de filtres [20].

II.2.4 APPROCHES DE MODELISATION

La modélisation vise à créer des références représentant chaque téléphone. Dans tous les enregistrements on tient compte des dépendances temporelles entre les vecteurs paramétriques extraits. On peut ainsi envisager d'aligner temporellement les séquences de vecteurs d'apprentissage et de test, car elles doivent contenir la même séquence. Néanmoins dans les applications indépendantes du texte, la modélisation tient compte de la seule distribution des paramètres acoustiques. Les techniques de modélisations peuvent dérivées de différentes grandes approches, comme l'approche vectorielle, connexionniste, prédictive et statistique [21].

II.2.4.1 LA QUANTIFICATION VECTORIELLE (QV)

La quantification vectorielle (QV) est une méthode d'approximation d'un signal d'amplitude continue par un signal d'amplitude discrète [22]. Elle est également connue sous le nom de modèle centroïde. Cette approche consiste à construire des modèles en partitionnant les vecteurs caractéristiques en K groupes non-chevauchants qui représentent individuellement différentes classes acoustiques [23].

II.2.4.2 LES MODELES DE MARKOV CACHES (HMM).

Selon le formalisme des modèles de Markov cachés *(Hidden Markov Models* ou HMM), le signal de parole est supposé être produit par un automate d'états finis stochastique. Il est donc, construit à partir d'un ensemble d'états stationnaire régis par les lois statistiques. Le formalisme des modèles HMM suppose que le signal de parole est formé d'une séquence de segments stationnaires, tous les vecteurs associés à un même segment stationnaire étant supposés avoir été générés par le même état HMM. Chaque état de cet automate est caractérisé par une distribution de probabilité décrivant la probabilité d'observation des différents vecteurs acoustiques. Les transitions entre les états sont instantanées, elles sont caractérisées par une probabilité de transition. Si chaque état du modèle permet de modéliser un segment de parole stationnaire, la séquence d'état permet quant à elle de modéliser la structure temporelle de la parole comme une succession d'états stationnaires [24].

Les modèles de Markov cachés sont devenus la solution par excellence au paradigme de la reconnaissance de la parole. Ils s'inscrivent dans une approche statistique de la reconnaissance vocale assimilant les unités acoustiques de la parole à un processus aléatoire [25].

II.2.4.3 MODELE DE MELANGE DE GAUSSIENNES (GMM)

Le modèle de mélange de Gaussiennes est un modèle statistique où la distribution des données est un mélange de plusieurs lois Gaussiennes. Le GMM est le modèle de référence en reconnaissance du locuteur.

La même loi gaussienne d'un mélange λ à M composantes est paramétrée par un vecteur de moyennes μ_m de dimension D (D étant la dimension de l'espace des données), une

matrice de covariance $\sum_m$de dimension D × D et un poids $w_m \geq 0$. La fonction de densité de probabilité s'écrit sous forme de :

$$P(\boldsymbol{x}/\lambda) = \sum_{m=1}^{M} \omega_m N(\boldsymbol{x}/\mu_m, \sum_m) \text{ (II. 4)}$$

$$N(\boldsymbol{x}/\mu_m, \Sigma_m) = \frac{1}{\sqrt{(2\pi)^D \mid \Sigma_m \mid}} exp\left(-\frac{1}{2}(x-\mu_m)^T \Sigma_m^{-1}(x-\mu_m)\right) \quad \text{(II. 5)}$$

Et

$$\sum_{m=1}^{M} \omega_m = 1 \quad \text{(II. 6)}$$

L'apprentissage d'un modèle GMM consiste en l'estimation de l'ensemble des paramètres $\lambda = \{\mu_m \quad, \sum_m, \omega_m\}$, en utilisant un ensemble de données d'apprentissage $X = \{X_1, X_2, \ldots X_T\}(x_t \in R^D)$. Cet apprentissage fait souvent appel à la technique d'estimation par maximum de vraisemblance (Maximum Likelihood Estimation) MLE [21].

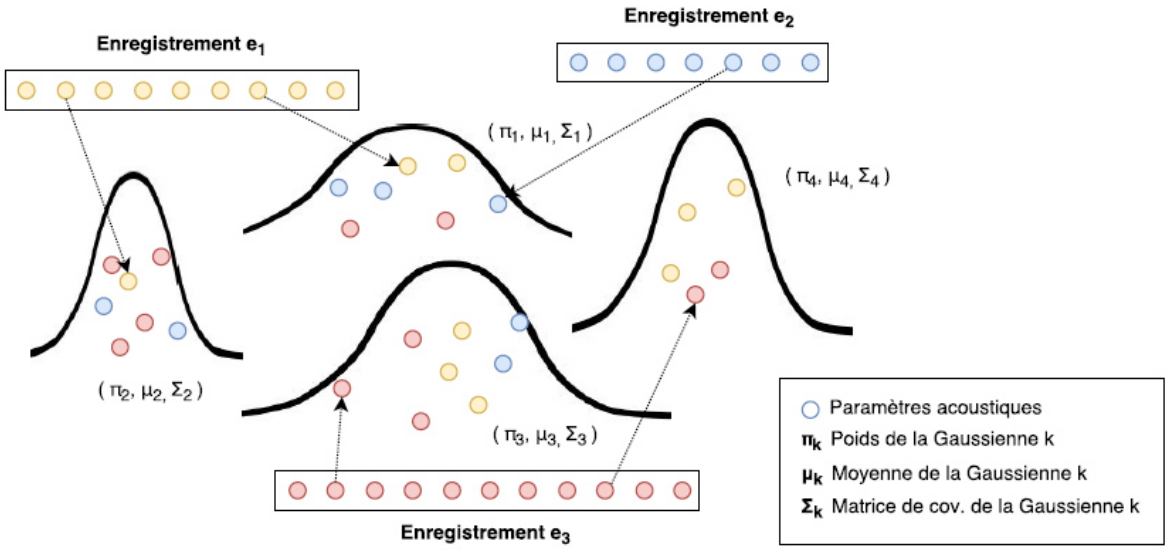


Figure II.6 : Un mélange de Gaussiennes (GMM) construit en utilisant des paramètres acoustiques issus de plusieurs enregistrements.

Une version améliorée de ce modèle GMM a été proposée pour palier au problème d'insuffisance de données est appelée GMM-UBM (Gaussien Mixture Model - Universel Background Model). On peut résumer cette approche par les étapes suivantes :

1. un seul modèle indépendant des téléphone, appelé modèle du monde (Universel Background Model UBM), est utilisé pour représenter$P(\boldsymbol{X}/\lambda_{UBM})$. L'apprentissage du modèle UBM se fait en utilisant un ensemble large de données de parole obtenu par la concaténation d'un large épouvantail de téléphones. C'est un large modèle GMM appris pour représenter la distribution de paramètres indépendants d'une marque de téléphone. Plus précisément, on désire sélectionner des échantillons de parole enregistrés par un téléphone qui reflètent l'enregistrement provenant d'une partie alternative nécessaire à la reconnaissance. Ceci s'applique au type et à la qualité de l'enregistrement, ainsi qu'à la composition des téléphones.

2. Un algorithme itératif appelé espérance-maximisation (Expectation Maximisation EM) est utilisé pour estimer le maximum de vraisemblance du modèle pour les vecteurs de paramètres d'apprentissage. L'estimation des paramètres du modèle de mélange des gaussiennes $\lambda_{UBM} = \{\omega_{UBM}, \mu_{UBM}, \Sigma_{UBM}\}$ est améliorée d'une itération a l'autre jusqu'a atteindre la convergence. L'apprentissage peut être arrêté lorsque le changement de vraisemblance entre deux itérations est au-dessous d'un certain seuil ou après avoir fait un nombre prédéterminé d'itérations [15].

3. La dérivation des marques du téléphone se fait, dans le système GMM-UBM de façon adaptative. Les signaux d'enregistrement d'apprentissage de chaque téléphone serrent a adapter les paramètres du modèle du monde (UBM) en utilisant l'algorithme d'estimation du maximum a posteriori (MAP) [15]. Cet algorithme adapte itérativement le modèle UBM pour arriver au modèle du téléphone comme illustré dans la figureII.7.

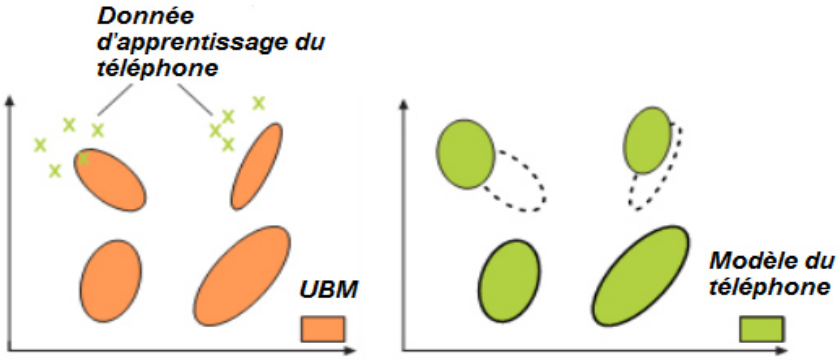


Figure II.7 : Obtention du modèle du téléphone par la méthode d'adaptation MAP.

II.2.5 L'APPROCHE I-VECTEUR

Le paradigme i-vecteur a été proposé comme extension des modèles d'analyse factorielle dans le domaine de la reconnaissance du locuteur. Pour l'expliquer, nous considérons le même contexte de la reconnaissance du locuteur, puis nous l'appliquerons dans notre cas qui est les reconnaissances des marques de téléphones portables.

Cette approche propose d'apprendre un seul sous espace qui contient à la fois la variabilité de locuteur et de session. Cet espace, appelé espace de **variabilité totale**, a permis d'avoir une représentation à faible dimension qui capture l'ensemble des variabilités acoustiques existant dans un enregistrement donné. Le modèle équivalent peut être représenté par [7] :

$$M = m + Tw \qquad \text{(II. 7)}$$

Où m est le super-vecteur UBM, T est une matrice rectangulaire de bas rang couvrant un sous-espace comportant la majeure partie de la variabilité dans l'espace superviseur, et w est un vecteur de dimension M ayant une distribution normale standard. Pour chaque session, le i-vecteur est l'estimation ponctuelle MAP de la variable latente$w$ [26].

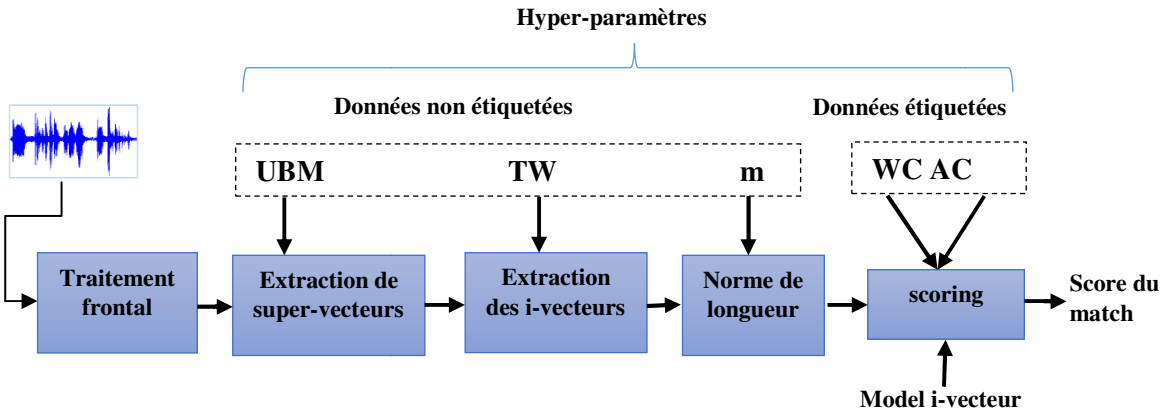


Figure II.8 : Diagramme simplifié de reconnaissance par i-vecteurs.

Les composants du vecteur $\boldsymbol{w}$ sont les facteurs totaux. Nous nous référons à ces nouveaux vecteurs en tant que vecteurs d'identité ou i-vecteurs pour faire court. Dans cette modélisation, $\boldsymbol{M}$est supposé être normalement distribué avec le vecteur moyen $\boldsymbol{m}$et la matrice de covariance est$\boldsymbol{TT^t}$**.** Le processus d'entraînement de la matrice de variabilité totale $\boldsymbol{T}$est exactement la même chose que l'apprentissage de la matrice de vecteur propre$V$, à la seule différence que dans l'apprentissage de la matrice V, tous les enregistrements d'un téléphone donné sont considérés comme appartenant à la même marque alors que dans le cas de la matrice de variabilité totale, l'ensemble des enregistrements d'un téléphone donné est considéré comme étant produit par différentes marque. Le nouveau modèle que nous proposons peut être vu comme une simple analyse factorielle qui nous permet de projeter un enregistrement de parole sur l'espace de variabilité totale de faible dimension [27].

Le facteur total $\boldsymbol{w}$est une variable cachée, qui peut être définie par sa distribution postérieure conditionnée aux statistiques de Baum-Welch pour un enregistrement donné. Cette distribution postérieure est une distribution gaussienne, et la moyenne de cette distribution correspond exactement à notre i-vecteur.

Supposons que nous ayons une suite de $\boldsymbol{L}$trames $\{\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_L}\}$et l'$UBM\Omega$composé de mélanges $\boldsymbol{C}$ définis dans un espace de caractéristiques de dimension$F$.

Les statistiques de Baum-Welch nécessaires pour estimer le i-vecteur pour un enregistrement de parole $\boldsymbol{u}$sont obtenues par :

$$N_c = \sum_{t=1}^{L} P(c/y_t, \Omega) \qquad \text{(II. 8)}$$

$$F_c = \sum_{t=1}^{L} P(c/y_t, \Omega) y_t \text{(II. 9)}$$

Où$c = 1, \dots C$est l'indice gaussien et $P(c/x_t, \Omega)$ correspond à la probabilité a posteriori du composant de mélange c générant le vecteury_t. Afin d'estimer le i-vecteur, nous devons également calculer les statistiques centralisées de Baum-Welch du premier ordre sur la base des composants du mélange moyen UBM :

$$\tilde{F}_c = \sum_{t=1}^{L} P(c/y_t, \Omega)(y_t - m_c) \quad \text{(II. 10)}$$

Où$\boldsymbol{m_c}$est la moyenne du composant $\boldsymbol{c}$ du mélange UBM. Le i-vecteur pour un enregistrement donné peut être obtenu en utilisant l'équation suivante :

$$w = (I + T^t \sum^{-1} N(u) T)^{-1} . T^t \sum^{-1} \tilde{F}(u) \quad \text{(II. 11)}$$

Nous définissons $N(u)$ comme une matrice diagonale de dimension $CF \times CF$ dont les blocs diagonaux sont$N_C I(c = 1, \dots, C)$. $\tilde{F}(u)$Est un super-vecteur de dimension $CF \times 1$obtenu en concaténant toutes les statistiques de Baum-Welch de premier ordre $\tilde{F}_c$pour un enregistrement donné u. $\sum$Est une matrice de covariance diagonale de dimension $CF \times CF$estimée lors de la formation à l'analyse factorielle. Cette matrice modélise la variabilité résiduelle non captée par la matrice de variabilité totale$T$.

II.2.5.1 SCORE DE DISTANCE EN COSINUS

En phase de test d'un système de reconnaissance comporte aussi l'étape de comparaison qui permet de donner un score reflétant la relation entre le modèle et l'enregistrement de test de l'appareil téléphonique en question. La technique de distance en cosinus utilise directement la valeur du noyau cosinus entre le vecteur-cible du téléphone cible et le vecteur-test comme score de décision. Il est donné par l'équation suivante :

$$score(w_{target}, w_{test}) = \frac{\langle w_{target}, w_{test} \rangle}{\| w_{target} \| \| w_{test} \|} \gtrless \theta \quad (II.14)$$

La valeur de ce noyau est ensuite comparée au seuil θpour prendre la décision finale. L'avantage de cette technique est qu'aucune inscription de locuteur cible n'est requise, contrairement aux machines à vecteurs de support et à l'analyse factorielle classique, où le superviseur cible dépendant du téléphone doit être estimé dans une étape d'inscription. Notez aussi que les i-vecteurs cible et test sont estimés exactement de la même manière (il n'y a pas de processus supplémentaire entre l'estimation de la cible et des i-vecteurs de test). Donc, les i-vecteurs peuvent être considérés comme de nouvelles caractéristiques de reconnaissance.
Dans cette nouvelle modélisation, l'analyse factorielle joue le rôle d'extracteur de caractéristiques plutôt que de modélisation des effets de téléphones et de canaux. L'utilisation du noyau cosinus comme score de décision pour la vérification rend le processus plus rapide et moins complexe que les autres méthodes d'analyse factorielle. [27]

II.2.5.2 NORMALISATION DE LA LONGUEUR DES I-VECTEURS

Une analyse de la longueur des i-vecteurs a révélé un décalage significatif entre les distributions de longueurs des i-vecteurs d'apprentissage et de test. Afin de corriger ce décalage, un prétraitement couramment utilisée en reconnaissance consiste à normaliser la longueur des i-vecteurs. Cette transformation non-linéaire permet de Gaussianiser la distribution des i-vecteurs et d'améliorer les performances de reconnaissance. Ce processus divise simplement chaque i-vecteur par sa norme [7]. La forme normalisée d'un i-vecteur w est donnée par :

$$w_{norm} = \frac{1}{\| w \|} \times w (II.15)$$

II.2.5.3 COMPENSATION DE LA VARIABILITE SESSION DANS L'ESPACEDES I-VECTEURS

L'approche i-vecteur ne fournit qu'une représentation à dimension réduite (~400-800) d'un enregistrement donné et n'effectue pas de compensation de la variabilité session / canal. Ainsi, les méthodes discriminatives qui étaient peu pratiques dans le paradigme GMM-UBM

et GMM-SVM (comme l'analyse discriminante linéaire) pourraient être appliquées dans cet espace à dimension réduite. Certaines techniques de normalisation (comme la WCCN et la projection NAP) [5], avaient des racines dans la modélisation super-vecteur et seront discutés dans ce qui suit dans le contexte des i-vecteurs [7].

II.2.5.4 NORMALISATION WCCN

La technique WCCN (Within-Class Covariance normalisation) a été proposée pour améliorer la robustesse dans le cadre de la reconnaissance du locuteur basée sur les SVM. La projection WCCN vise à améliorer les performances en minimisant le taux de fausses alertes (FA) lors de l'apprentissage des SVM. L'espace de projection construit est défini par la racine carrée de l'inverse de la matrice W_{WCCN}définie par [7] :

$$W_{WCCN} = \frac{1}{S} S_W = \frac{1}{S} \sum_{S=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} (w_{s,i} - \mu_s)(w_{s,i} - \mu_s)^T \qquad (\text{II. } 16)$$

La matrice de projection **Q** est trouvée par la décomposition de Cholesky de l'inverse de la matrice $\mathbf{W}_{\text{WCCN}}$ définie par [7] :

$$W_{WCCN}^{-1} = QQ^T \qquad (\text{II. } 17)$$

II.2.5.5 MESURE DES SCORES DANS L'ESPACE DES I-VECTEURS

Suite à l'introduction du paradigme de la variabilité totale, de nombreuses méthodes ont été proposées pour comparer d'une manière efficace deux i-vecteurs correspondant à deux enregistrements donnés.

La comparaison des i-vecteurs calcule un quotient de vraisemblances qui teste la validité de deux hypothèses (les hypothèses client et imposteur). Étant donnés deux i-vecteurs $W1$ et $W2$ à comparer, l'opération de calcul des scores est définie par :

$$score = \log \frac{P(w_1, w_2 / H_{client})}{P(w_1, w_2 / H_{impost})} \qquad (\text{II. } 18)$$

L'hypothèseH_{client}indique que les vecteurs w_1 et w_2 sont issus du même téléphone

Et l'hypothèse H_{impost} indique qu'ils correspondent à deux téléphones différents.

II.2.6 PHASE DE DECISION

Enfin, nous abordons le dernier module constituant le système de reconnaissance à savoir la prise de décision. Au bout de la phase test une décision concernant la marque de téléphone portable sera prose.

Dans un système d'identification, cette phase est purement et simplement un classificateur en maximum de vraisemblance. Pour une base de données de R téléphones r = {1, 2, ..., R} représentée par les GMM$\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_R$, l'objectif est de trouver la marque de téléphone portable qui a la plus grande probabilité a posteriori pour la séquence $X = \{X_1, X_2, \dots X_L\}$donnée en entrée [28].

En appliquant le minimum d'erreur pour la loi de décision de Bayes, on peut estimer :

$$\tilde{S}_c = arg \max_{1 \le r \le R} P_{priori}(\lambda_r / X) = arg \max_{1 \le r \le R} \frac{P(X/\lambda_r) P_{priori}(\lambda_r)}{P(X)} \quad (II.19)$$

En assumant l'égale probabilité a priori de tous les téléphone portable :

$$P_{priori}(\lambda_r) = 1/R \quad (II.20)$$

Et en remarquant que les termes $P_{Priori}(\lambda_r)$et *p(X)* sont constants pour tous les marques de téléphones portable, on peut alors simplifier la loi de classification précédente et écrire que :

$$Sc = arg \max_{1 \le r \le R} p(X/\lambda_r) \quad (II.21)$$

En considérant que les données sont indépendantes, et en appliquant le logarithme, on a :

$$\log[p(X/\lambda_r)] = \sum_{l=1}^{L} \log p(X_l/\lambda_r) \quad (II.22)$$

Et par conséquent, le système d'identification va établir un score correspondant au logarithme du maximum de vraisemblance de la séquence de vecteurs de caractéristique test pour chaque téléphone portable, à savoir :

$$S\tilde{c}_{Log} = arg \max_{1 \le i \le R} \sum_{l=1}^{L} \log P(X_l/\lambda_r) \quad (II.23)$$

$$P(X\_l/\lambda) = \sum_{n=1}^{N} \prod_n P\left(X_l/_{\lambda_n}\right) = \sum_{n=1}^{N} \prod_n g_n(X_l) \quad (II.24)$$

Le GMM qui génère le plus grand score SC̃Log est identifié comme étant le modèle du téléphone testé [28].

En vérification, l'objective est de déterminer si la séquence de test X a été prononcée par le téléphone L. Pour cela, nous avons besoin de deux hypothèses [15] :

- $H0$: X a été parlée par L

  Et

- $H1$: X n'a pas été parlée par L.

Le rapport de vraisemblance (Likelihood ratio LR) entre les deux hypothèses H_0 et H_1 s'écrit par :

$$LR = \frac{P(H_0/L)}{P(H_1/L)}. \quad (II.25)$$

D'après le théorème de Bayes, on a :

$$P(H_0/L) = \frac{P(L/H_0)P(H_0)}{P(X)}. \quad (II.26)$$

En remplaçant $P(H_0/L)$ par sa valeur dans l'équation$(II.25)$ nous obtenons :

$$LR = \frac{P(L/H_0)P(H_0)}{P(L/H_1)P(H_1)}. \quad (II.27)$$

Cette expression est utilisée en pratique car il est plus facile de calculer $P(L/H_0)$ que de calculer$P(H_0/L)$.

Dans la technique GMM-UBM, L'évaluation du rapport LR est réalisée en calculant la différence du logarithme de vraisemblance$\log P(X/\lambda_i) - \log P(X/\lambda_{UBM})$. Si cette différence est supérieure à un seuil donné, le téléphone est accepté, sinon il est rejeté [15].
Ceci peut se formuler comme :

$$\begin{cases} \log P(X/\lambda_i) - \log P(X/\lambda_{UBM}) \geq \theta, & Acceptation \\ \log P(X/\lambda_i) - \log P(X/\lambda_{UBM}) < \theta, & rejet \end{cases} \quad (II.28)$$

II.3. CONCLUSION

Dans ce chapitre, nous avons introduit le principe de la reconnaissance de la marque du téléphone portable, ainsi que les différentes étapes du système de reconnaissance du téléphone portable qui se compose généralement de trois étapes : l'analyse acoustique de signal de parole, la modélisation et en dernier lieu l'étape de décision.
En analyse acoustique, les MFCC sont les coefficients les plus répondus. Quant à la modélisation, l'approche i-vecteur permis d'avoir une représentation à faible dimension qui capture l'ensemble des variabilités acoustiques existant dans un enregistrement donné. La décision est basée sur deux processus d'identification et de vérification.

Chapitre III

Résultat et simulation

III.1 INTRODUCTION :

Dans ce chapitre, nousallons mettre en expérience les bases théoriques développées jusque-là. Pour réaliser notre système de reconnaissance de la marque et du modèle d'un téléphone portable, nous utilisons des échantillons vocaux enregistrés provenant d'un ensemble de 15 téléphones portables.

Nous avons construitcette base de données qui nous servira pour étudier les performances de notre système de reconnaissance de téléphone portable.

Pour cela, nous avons utilisé la base de données TIMIT qui est une base de données de reconnaissance vocale utiliséepour la parole ou la reconnaissance des locuteurs.Elle est composée de 630 locuteurs de différents dialectes d'anglais Américain. Nous avons choisi seulement 50 locuteurs pour faire nos simulations.

III.2 BASE DE DONNEES :

Dans les expériences, qu'on va faire, nous utilisons 15 téléphones portables, la collection des marques comprend Alcatel, quatre Condor, Huawei, Nokia, Oppo, deux Samsung, Stream, Volt 4, et deux wiko. Ont été utilisés dans les expériences. Les marques et les modèles de téléphones portables sont répertoriés dans le tableau III.1.

Marque et Modèle
Alcatel Onetouche
CONDOR A8
CONDOR C6
CONDOR P8
Condor PGN528
HUAWEI Y6PRO
Nokia 302
oppo A1601
samsung galaxy S3 GT_19300
samsung galaxy S5 SM_G900T
stream b1
VOLT 4 GLTE
wiko LENNY 2
Wikoroby
ZTE V795

Tableau III. 1 : Les marques et modèles de téléphones portables utilisés.

III.3 LES ETAPES DE L'ENREGISTREMENT ET DE SIMULATIONS

Dans notre étude on a rassemblé 15 téléphones portablesdifférents pour effectuerun enregistrement de durée 25 min via un pc portable.Ces enregistrements ont été pris dans une salle de lecture au sein de notre université.Afin de réaliser nos enregistrements dans les mêmes conditions pour tous les téléphones, nous avons placéla source du son qui est pc portable sur une table et les 15 téléphones portablesà environ la même distancede son haut-parleur. Nous avons met tous les téléphones en marche pour que ces derniers puissent enregistrer le son. Après avoir stocké nos données sur le PC, Nous avons convertis tous les enregistrements en fichiers wav. Nous lesavons, ensuite,divisé en trois intervalles.La première partie va être utilisée pour modéliser le modèle du monde (UBM). La deuxième partie seraà la

création des modèles des téléphones. Enfin la troisième partie sera utilisée pour effectuer les tests.

A l'aide du logiciel de programmation MATLAB, nous avons commencé à la construction de notre système de reconnaissance tel qu'expliqué dans le chapitre II. La première étape étant l'extraction des paramètres des téléphones en utilisant les coefficients MFCC.Ces paramètres sont par la suite utilisés pour la création de chaque modèle de téléphone parla technique i-Vecteur.Les modèles obtenus seront stockés dans une base de données. Toutes ces étapes constituent la phase d'apprentissage.

La phase test comprend l'extraction de caractéristiques, et l'étape de correspondance. Comme pour la phase d'apprentissage l'extraction de paramètres est le processus d'extraction des informations utiles qui caractérisent le téléphone portable à partir d'un ensemble d'échantillons vocaux donnés. L'algorithme de correspondance effectue le calcul de la mesure de similarité et effectue des comparaisons entre les modèles.

Les résultats obtenus sont évalués à travers l'erreur de reconnaissance qui correspond au nombre de fois où le système commet une erreur de test sur le nombre total de tests.

Dans la reconnaissance du système d'identification considéré dans cette étude, les vecteurs de caractéristiques pour les téléphones portables sont calculés à partir d'échantillons vocaux et comparés à chaque modèle stocké dans la base de données et au modèle produisant la similarité maximale est attribuée comme identité de téléphone portable comparé.

III.4 RESULTATS DE SIMULATIONS

Essentiellement, Dans toutes nos simulations nous nous basons sur trois facteurs principaux qui sont le nombre de paramètres (n_MFCC), le nombre de gaussiennes (n mix), et le nombre d'itération des algorithmes d'estimation des GMM, de la matrice T et de l'analyse LDA (n itérations).Aussi, on a noté l'énergie du signal par la lettre e, alors que filtre correspond à utiliser ou non le filtre de préaccentuation.

III.4.1 INFLUENCE DE NOMBRE DE COEFFICIENT MFCC

Dans la première simulation, nous avons fixé deux facteurs qui sont n mix, et le nombre d'itération.Les résultats obtenus sont illustrés dans le tableau suivant :

N_MFCC	10	12	14	16	18	20
N MIX	8	8	8	8	8	8
Nombre d'itération	10	10	10	10	10	10
Erreur avec 'e' avec filtre	37.7713%	40.4185%	40.0951%	38.6665%	41.7130%	41.5237%
Erreur2 sans 'e' Avec filtre	45.5237%	40.5713%	42.9523%	37.1427%	43.0476%	38.5332%
Erreur 2 avec 'e' sans filtre	36.5714%	37.9999%	40.0000%	35.99991 %	28.0952%	36.4761%
Erreur 2 sans 'e' et sans filtre	37.2380%	39.4285%	37.5237%	38.0953%	35.8095%	35.6190%

Tableau III. 2 : Influence du nombre de coefficient MFCC sur le système.

Dans ce tableau nous avons effectué l'opération avec différents valeurs de coefficient MFCC qui correspond au nombre de paramètre de GMM ,et les résultats obtenu sont en fonction de l'énergie et le filtre pour la 1er ligne, et sans énergie et avec filtre pour la 2eme ligne, la 3eme ligne correspond aux résultats avec énergie et sans filtre et pour la 4eme ligne correspond aux résultats sans énergie et sans filtre. Nous avons constaté que le meilleur résultat est obtenu pour un nombre de coefficients MFCC égal à 18 paramètres avec énergie et sans filtre. L'erreur dans ce cas est de 28.0952%.

III.4.2 INFLUENCE DU NOMBRE DE GMM

Dans cette simulation nous avons fixé le nombre de coefficients MFCC à 18 pour sa meilleure performance, et un nombre d'itérations de 10. Différentes valeurs ont été choisies pour le nombre de GMM et les résultats obtenus sont illustrées dans le tableau suivant :

MFCC	18	18	18	18
N MIX (nbre de GMM)	8	16	32	64
Nombre d'itération	10	10	10	10
Erreur avec 'e' sans filtre	28.0952%	38.4761%	40.4761%	37.7142%

Tableau III. 3 : Influence du nombre de GMM sur le système de reconnaissance.

Dans notre cas nous avons effectué notre expérience avec quatre valeurs pour GMM pour voir à quel nombre de GMM le système est le plus performant.Nous avons constaté que le meilleur résultat est obtenu pour un nombre de GMM de 10 l'erreur est égaleà 28.0952 %, pour les autres valeurs les performances du système sont moins bonnes.

III.4.3 INFLUENCE DU NOMBRE D'ITERATIONS

Enfin, dans cette dernière simulation, nous avons fixé n mix, et le n_MFCC sur les valeurs qui donne le meilleur résultat et nous avons fait varier le nombre d'itérations pour le modèle du monde UBM, la matrice T, et la technique LDA. En fixant deux paramètres et en variant le troisième nous avons obtenules résultats obtenus illustrés dans les tableaux ci-dessous :

MFCC	18	18	18	18	18	18
N MIX	8	8	8	8	8	8
Nombre d'itération	UBM : 2 T : 10 LDA : 10	UBM : 4 T : 10 LDA : 10	UBM : 6 T : 10 LDA : 10	UBM : 8 T : 10 LDA : 10	UBM : 10 T : 10 LDA : 10	UBM : 12 T : 10 LDA : 10
Erreur avec 'e' sans filtre	39.0475%	44.5719%	46.6666%	42.0417%	28.0952%	47.8095%

Tableau III. 4 : Influence de nombre d'itération pour UBM.

D'après les résultats obtenu dans le tableau précédent (III.4), réalisé pour en faisant varierle nombre d'itération du modèle ubm, nous avons constaté que les meilleurs résultats de notre simulation est obtenu pour un nombre d'itération égal à 10.

MFCC	18	18	18	18	18	18
N MIX	8	8	8	8	8	
Nombre d'itération	UBM : 10 T : 2 LDA : 10	UBM : 10 T : 4 LDA : 10	UBM : 10 T : 6 LDA : 10	UBM : 10 T : 8 LDA : 10	UBM : 10 T : 10 LDA : 10	UBM : 10 T : 12 LDA : 10
Erreur avec 'e' sans filtre	38.9523%	36.1904%	37.8095%	37.8095%	28.0952%	37.0476%

Tableau III. 5 : Influence de nombre d'itération pour total variabilité.

De même dans cette simulation les résultats obtenusmontrent que pour le nombre d'itération concernant total variabilité égal à 10 nous avons obtenu le meilleur résultat.

MFCC	18	18	18	18	18	18
N MIX	8	8	8	8	8	8
Nombre d'itération	UBM : 10 T : 10 LDA : 2	UBM : 10 T : 10 LDA : 4	UBM : 10 T : 10 LDA : 6	UBM : 10 T : 10 LDA : 8	UBM : 10 T : 10 LDA : 10	UBM : 10 T : 10 LDA : 12
Erreur avec 'e' sans filtre	38.9523%	38.0000%	37.8095%	33.8095%	28.0952%	29.5238%

Tableau III. 6 : Influence de nombre d'itération pour LDA.

D’après les résultats obtenus dans ce tableau (III.6). La variation du nombre d’itération concernant la technique LDA donne le meilleur résultatde reconnaissance toujours pour un nombre d'itération de 10.

III.5 CONCLUSION

Dans ce chapitre, nous avons mis en simulation les différentes étapes du système de reconnaissance. Les paramètres caractéristiques des téléphones portables sont calculés à partir d'échantillons vocaux obtenus à partir d'une base de données que nous avons construit au sein de notre université. Ces paramètres sont comparés à chaque modèle stocké dans une base de données. L'enregistrement produisant la similarité maximale est attribué comme identité de téléphone portable. Nous avons étudié différentes situations pour montrer l'efficacité de notre système.Nous avons constaté que dans notre cas de reconnaissance des marques de téléphone le meilleur résultat est obtenu avec un pourcentage d’erreur égal à 28.0952 %.

CONCLUSION GENERALE

Dans ce mémoire, nous avons étudié les systèmes de reconnaissances des téléphones portables, et nous avons utilisé l'une des méthodes de paramétrisation les plus utilisées en reconnaissance qui est la méthode MFCC. Les coefficients MFCC capturent les caractéristiques de l'appareil source et peuvent être utilisées pour reconnaître les téléphones portables à partir de leurs enregistrements. Nous avons, aussi, utilisé une approche pour la reconnaissance du téléphone qui fait appel aux modèles i-vecteur à base de GMM et qui sont devenus populaires pour les systèmes de traitement de la parole.

Toutes les connaissances théoriques impliquées dans la réalisation du système de reconnaissance ont été testées dans la dernière partie de ce mémoire.

Pour cela nous avons commencé par effectuer des enregistrements réels sur plusieurs types de téléphones afin de valider nos résultats. Les données obtenus ont été par la suite utilisé pour réaliser le système de reconnaissance de base.

Différents paramètres ont été testés comme le nombre de paramètres MFCC, le nombre de GMM et le nombre d'itérations. Ces simulations visaient à obtenir le système le plus performant possible. Les résultats obtenus montrent qu'un nombre de 18 paramètres avec 8 GMM et 10 itérations réalisaient le plus faible taux d'erreur. Nous avons donc, arrivé à identifier notre liste de téléphones portables, où chaque enregistrement correspond au téléphone avec lequel il a été enregistré.

Cette expérience nous a permis d'apprendre et surtout de toucher à plusieurs domaines tels que le traitement de signal, la programmation, etc…

Bibliographie

[1] CemalHanilc¸i_, FigenErtas¸, TuncayErtas¸ and ¨Omer Eskidere « Recognition of Brand and Models of Cell-Phones From Recorded Speech Signals ».June 06, 2011,pp.1-10

[2] Filipe Velho, « la reconnaissance du locuteur à l'aide de la transformée en ondelettes continue ».thèse de doctorat, école de technologie supérieure,université du québec,8 mars 2006.

[3] Luiza. OROSANU ,« Reconnaissance de la parole pour l'aide à la communication pour les sourds et malentendants ».Thèse de doctorat, Université de Lorraine, soutenue le 11 Décembre 2015.

[4] Abdenour.HACINE-GHARBI , « sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole ». Thèse en cotutelle internationale, école doctorale sciences et Technologies (orleans) et faculté de technologie(sétif),Algérie, soutenue le 09 décembre 2012.

[5] Fréjus A. A. LALEYE , « Contributions à l'étude et `a la reconnaissance automatique de la parole en Fongbe ». Thèse de doctorat, l'Université d'Abomey-Calavi et l'Université du Littoral Côte d'Opale, soutenue le 10 décembre 2016.

[6] Asmaa.AMEHRAYE , « Débruitage perceptuel de la parole ». Thèse de doctorat, l'Ecole Nationale Supérieure des Télécommunications de Bretagne, Soutenue le 15 mai 2009.

[7] waad.BEN KHEDER , « reconnaissance du locuteur en milieux difficiles ». thèse de doctorat, l'université d'avignon et des pays de vaucluse, soutenue le 18 juillet 2017.

[8] R. Boite, H. Bourlard, T. Dutoit, J. Hancq et H.Leich, « Traitement de la Parole ».Presses Polytechniques Universitaires Romandes, Lausanne, 2000.

[9] Hassan.EZZAIDI, Jean.ROUAT et Ivan.BOURMEYSTER, « Reconnaissance Automatique de Parole en Français pour Milieu Difficile: Exemplede Détection de Double Parole pour le Radiotéléphone en Mains Libres. ». Université du Québec à chicoutimi, Canada, Alcatel Mobile Phones, Paris, France.

[10] LÊ.VIET BAC, « Reconnaissance automatique de la parole pour des langues peu dotées ». Thèse de doctorat, UNIVERSITÉ JOSEPH FOURIER - GRENOBLE 1, soutenue le 1er juin 2006.

[11]Houda. KADI, « La reconnaissance automatique du locuteur par la voix IP ».mémoire de master, université sidi mohamed ben abdallahfes (maroc), Soutenu le : 18/06/2014.

[12] Feriel.DEBBECHE-GUERID, « conception d'un systèmeacoustico-anatomique pour l'identification du locuteur : architecture et parametrisation ». Mémoire de magister, option : texte, image et parole, Université Badji Mokhtar Annaba (Algérie), 2008.

[13] Jean-Luc.ROUAS , « Caract´erisation et identification automatique des langues ». Thèse de doctorat, l'Universit´e Toulouse III – Paul Sabatier, soutenue publiquement le 11 Mars 2005.

[14] Cemal. HANIL, TomiKinnunen , « Source Cell-Phone Recognition from Recorded Speech Using Non-Speech Segments ». Department of Electrical-Electronic Engineering, Bursa Technical University, 16190, FI-80101, Joensuu , Finland, August 21, 2014.

[15] Abdennour ALIMOHAD : « contribution a l'inférence d'identité en utilisant un système de reconnaissance du locuteur gmm-ubm ». Thèse de doctorat, Université Blida-1, soutenue septembre 2015.

[16] Tiraogo Abdoulaye Yves ZANGO, « Évaluation subjective de la qualité : proposition d'un système de référence pour les codecs en bande élargie ». Thèse de doctorat, Université de Rennes 1, soutenue le 6 février 2013.

[17] AJGOU.Riadh , « Reconnaissance Automatique du Locuteur à Travers les Canaux Digitaux ». Thèse de doctorat, Université Mohamed Khider – Biskra, Soutenue le 14/02/2016.

[18] NamrataDAVE , « feature extraction methods lpc, plp and mfcc in speech recognition ». International journal for advance research in engineering and technology, volume 1, issue VI, july 2013.

[19] Salma. JAMOUSSI, « Méthodes statistiques pour la compréhension automatique de la parole ». Thèse de doctorat, l'université Henri Poincaré-Nancy 1, soutenue le 6 décembre 2004.

[20] Shreya Narang, Ms. DivyaGupta, « Speech Feature Extraction Techniques : A Review ». International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March-2015, pg. 107-114.

[21] Reda.JOURANI ,« econnaissance automatique du locuteur par des GMM à grande marge ». Thèse de doctorat, Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier), soutenue le 6 septembre 2012.

[22] Y. Ait khouya, M. Ait OUSSOUS , N. Alaa , B. Ait Es Said , « Construction d'un dictionnaire pour la Quantification Vectorielle : Application à la compression d'images fixes ». TELECO2011 & 7ème JFMMA, Mars 16-18, 2011 – Tanger MAROC.

[23] Surosh G. Pillay : « voice biometrics under mismatched noise conditions ». Thèse de doctorat, University of Hertfordshire, September 2010.

[24] Khenfer-koummichfatima, Mesbahi Larbi, Hendel Fatiha, « Reconnaissance des commandes vocales d'un robot mentor dans un environnement bruité à base HMM ». Université des Sciences et de Technologie d'Oran Mohamed Boudiaf (USTO-MB), Oran, Algérie.

[25] Rémi PREISS, « étude du canal téléphonique dans un système de reconnaissance robuste de la parole ». Thèse de doctorat, université du Québec, soutenue le 29 mars 2006.

[26] Hagai.ARONOWITZ, Oren Barkan, « EFFICIENT APPROXIMATED I-VECTOR EXTRACTION ».978-1-4673-0046-9/12/$26.00 ©2012 IEEE

[27] NajimDEHAK, patrick j. kenny, RédaDEHAK, Pierre DUMOUCHEL, and Pierre OUELLET, « front-end factor analysis for speaker verification ».ieee transactions on audio, speech, and language processing, vol. 19, no. 4, may 2011.

[28] Filipe.VELHO , « la reconnaissance du locuteur à l'aide de la transformée en ondelettes continue ». thèse de doctorat, université du Québec, soutenue le 8 février 2006.