Ordre ...../F.S.S.A/UAMOB/2019

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE AKLI MOUHAND OULHADJ-BOUIRA

Faculté des Sciences et des Sciences Appliquées
Département : **Génie Electrique**

Mémoire de fin d'étude

Présenté par :

Alouache Lyes
Louggani Yanis

En vue de l'obtention du diplôme de **Master** en **:**

Filière : Télécommunication
Option : Système des télécommunications

Thème :

Compensation de la variabilité du canal en reconnaissance du locuteur.

Devant le jury composé de :

H.A.Benghenia	MAA	UAMOB	Président
A. Alimohad	MCB	UAMOB	Encadreur
A.Bougharouat	MCB	UAMOB	Examinateur
Z.Asradj	MAA	UAMOB	Examinateur

Année Universitaire 2018/2019

Rire souvent et beaucoup, mériter le respect des gens intelligents et l'affection des enfants, gagner l'estime des critiques honnêtes et endurer les trahisons de ceux qui ne sont pas de vrais amis, apprécier la beauté, trouver ce qu'il y a de mieux dans les autres, laisser derrière soi un monde un peu meilleur, par un bel enfant, un jardin fleuri ou une condition sociale moins dure, savoir qu'une vie seulement a respiré plus facilement grâce à vous, voilà ce qu'est la réussite

Anthony Robbins

Remerciements

Nous tenons à remercier notre encadreur qui a eu confiance en nous dès le premier contact et nous a soutenu durant tous ces mois.

Tous nos remercîments au chef de département génie électrique pour toute l'aide matériel qu'il a mis à notre disposition.

Merci aussi aux membres du jury qui ont accepté de lire ce document et assister à notre présentation pour juger ce travail de fin de cycle.

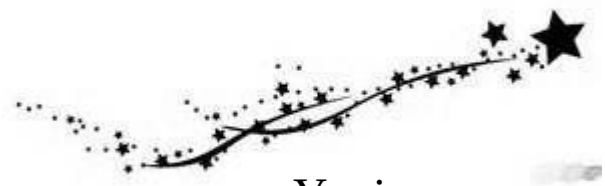
Un grand merci à toute personne qui nous a aidé et soutenu de prêt ou de loin.

Dédicaces

je dédie ce travail à mes chers parents, source de ma vie qui sont toujours à mes coté pour me soutenir et m'encourager, à mon cher frère Adel et cheres sœurs Amira,Naouel,Sihem,Amel à toute ma famille et mes amis soit en Algérie ou bien à l'extérieur , Sans oublier mon ami et mon binôme Yanis malgré la difficulté des jours passés, c'était amusant "SOLY"

Lyes

A mes parents qui ont toujours été là durant ces longues années d'études offrants un énorme soutien moral, affectif et financier, à Yilda qui a toujours été là au petit soin avec moi en supportant mes hauts et mes bas, à mon Yan qui n'apporte que de la joie et qui me rend fier chaque jour un peu plus.



Yanis

Sommaire

Liste des figure.....iv

Liste des tableaux.....v

Liste des abréviations.....vii

Introduction Générale 1

Chapitre I Généralités sur le traitement de la parole.....3

Généralités sur le traitement de la parole.....3

I.1) Introduction :.....4

I.2) Production de la parole :.....5

I.2.1) Les différents organes de production de la parole :.....5

I.2.2) Mécanisme de production de la parole :.....6

I.2.3) Modélisation du mécanisme de production de la parole :.....7

I.3) La perception de la parole :.....8

I.3.1) Le système auditif :.....8

I.3.2) Analyse fréquentielle et échelles de modélisation :.....9

I.3.2.1) L'échelle de Bark :.....9

I.3.2.2) Echelle de Mel :.....10

I.4) La Reconnaissance :.....11

I.4.1) La Reconnaissance automatique de la parole :.....11

I.4.2) La reconnaissance automatique de langage :.....12

I.4.3) La reconnaissance automatique du locuteur (RAL) :.....13

I.4.3.1) Deux phases de la reconnaissance du locuteur :.....14

I.4.3.2) Taches de la reconnaissance du locuteur :.....14

I.4.3.3) Méthodes de reconnaissance du locuteur :.....15

I.5) Variabilité du signal parole :.....15

I.5.1) Variabilité intra-locuteur :.....16

I.5.2) Variabilité interlocuteurs :.....16

I.5.3) Variabilité intersession :.....16

I.6) Conclusion :.....16
Chapitre II Le système de reconnaissance automatique du locuteur.....18
Le système de reconnaissance automatique du locuteur.....18
II.1) Introduction :.....19
II.2) Parameterisation :.....19
II.2.1) Propriété liée au signal parole :.....19
II.2.2) les différents types de paramètres :.....19
II.2.2.1) les paramètres prosodiques :.....19
II.2.2.2) Paramètre de l'analyse spectrale :.....20
II.3) Modélisation :.....25
II.3.1) le modèle de mélange de Gaussiennes (Gaussian Mixture Model GMM):.....25
II.3.2) La quantification vectorielle :.....25
II.3.3) Modèle de Markov-caché (Hidden Markov Modeling HMM):.....26
II.4) Comparaison et décision :.....26
II.4 .1) En quantification vectorielle :.....27
II.4 .2) Dans le modèle de mélange gaussien :.....27
II.5) Evaluation des performances du système :.....27
II.6) Comparaison entre systèmes :.....28
II.7) Compensation de la variabilité du canal :.....28
II.7.1) Intervention sur le bloc parameterisation (prétraitement) :.....28
II.7.1.1) préaccentuation :.....28
II.7.1.2) Suppression des zones des silences :.....28
II.7.2) Intervention sur le bloc de modélisation (GMM-UBM):.....30
II.8) Conclusion :.....30
Chapitre III Compensation de la variabilité du canal.....31
Compensation de la variabilité du canal.....31
III.1) Introduction.....32
III.2) Le contexte expérimental :.....32

III .2.1) Le système de base :.....32
III .2.2) La base de données : 33
III.2.3) Le logiciel utilisé :.....33
III.3) Fixation des paramètres optimaux.....33
III.4) Compensation de la variabilité du canal :.....34
III.4.1) Intervention sur le bloc parameterisation :.....35
III.4.1.1) Le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe :.....35
III.4.1.2) Le cas d'apprentissage avec téléphone mobile et test avec microphone : 35
III.4.1.3) Le cas d'apprentissage avec microphone et test avec téléphone fixe : 36
III.4.1.4) Le cas d'apprentissage avec microphone et test avec téléphone mobile : 36
III.4.1.5) Le cas d'apprentissage avec téléphone fixe et test avec microphone : 37
III.4.1.6) Le cas d'apprentissage avec téléphone fixe et test avec microphone : 37
III.4.2) Intervention sur le bloc modélisation :.....38
III.4.2.1) Le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe :.....39
III.4.2.2) Le cas d'apprentissage avec téléphone mobile et test avec microphone : 39
III.4.2.3) Le cas d'apprentissage avec microphone et test avec téléphone fixe : 40
III.4.2.4) Le cas d'apprentissage avec microphone et test avec téléphone mobile : 40
III.4.2.5) Le cas d'apprentissage avec téléphone fixe et test avec microphone : 41
III.4.2.6) Le cas d'apprentissage avec téléphone fixe et test avec téléphone mobile :.....42
III.5) Comparaison du taux d'amélioration entre le cas de modélisation GMM et le cas modélisation GMM-UBM : 42
III.6) Conclusion :.....43
Conclusion Générale.....44
bibliographies.....47

Liste des figure

Figure I.1 : Traitement de la parole…….4

Figure I.2 : Schéma de simplifié de l'appareil phonatoire……...……5

Figure I.3 : Schéma simplifié de production de la parole……..……...…..…….6

Figure I.4 : Comparaison d'un son voisé et d'un son non-voisé……...……....……...7

Figure I.5 : Schéma source-filtre de production de la parole…...……..…...……...7

Figure I.6 : Observation spectrale du conduit vocal……..…….…….……...…….……8

Figure I.7 : Système auditif……...……..……...…….……..8

Figure I.8 : Représentation de la dépendance de la fréquence résonnance par rapport à la position de la cellule cilié…….……....……..9

Figure I.9 : Tracé de l'échelle de Bark en fonction de la fréquence (en Hz)…….....……10

Figure I.10 : L'échelle des Mel……..……...…..……10

Figure I.11 : Blocs de base composant le système RAP……...…….……11

Figure I.12: Structure générale d'un système d'identification automatique des langues...….....….....13

Figure I.13 : schéma simplifié le système de reconnaissance du locuteur…….....…….…...…….13

Figure I.14 : Structure de base des systèmes de vérification du locuteur……..……..…..…...14

Figure I.15 : Structure de base des systèmes d'identification du locuteur……..……..……..15

Figure II.1 : Evolution de la fréquence de vibration des cordes vocales dans la phrase « je me suis enrhumé » ……...19

Figure II.2 : Extraction des paramètres MFCC…….…..……19

Figure II.3 : Fenêtrage du signal de parole pour l'obtention de vecteurs paramétriques…..….....20

Figure II.4 : Banc de filtres Triangulaires équidistance en échelle Mel…….……...….....21

Figure II.5 : Exemple d'un spectre LPC……...…….…….22

Figure II.6 : Modèle simplifié de la production de la parole…….....……..23

Figure II.7 : Signal de parole après le passage par un détecteur de silence……..….28

Figure II.8 : Structure générale d'un système RAL à base GMM-UBM……..….....29

Figure III.1 : Structure du système de base……..……..…….……31

Liste des tableaux

Tableau III.1 : Tableau explicatif des différentes variations du canal……..……...….……32

Tableau III.2 : Paramètres optimaux pour chaque expérience accompagnée de la performance atteinte……33

Tableau III.3 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe avec l'intervention sur le bloc parameterisation…….34

Tableau III.4 : performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec microphone avec l'intervention sur le bloc parameterisation…….35

Tableau III.5 : performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone fixe avec l'intervention sur le bloc paramétrisation……...35

Tableau III.6 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone mobile avec l'intervention sur le bloc parameterisation…….36

TableauIII.7 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc parameterisation…….….36

Tableau III.8 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc parameterisation…….….37

Tableau III.9 : Performance du système avec les paramètres optimaux pour 30 personnes……..37

Tableau III.10 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe avec l'intervention sur le bloc modélisation……...38

Tableau III.11 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec microphone avec l'intervention sur le bloc modélisation……...38

Tableau III.12 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone fixe avec l'intervention sur le bloc modélisation….....39

Tableau III.13 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone mobile avec l'intervention sur le bloc modélisation……...40

Tableau III.14 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc modélisation……...40
Introduction générale

Tableau III.15 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec téléphone mobile avec l'intervention sur le bloc modélisation...……....41

Tableau III.16 : Comparaison entre le taux d'amélioration obtenu avec GMM et celui obtenu avec GMM-UBM……...…….....…...……41

Liste des abréviations

RAL : Reconnaissance Automatique du Locuteur.

PIN: Personal Identification Number.

MFCC: Mel Frequency Cepstral Coefficients.

DCT: Discrete Cosine Transform.

AR : Auto Régressifs.

LPC: Linear Predectiction coefficients.

LPCC: Linear Prediction Cepstral Coefficients.

GMM: Gaussian Mixture Model.

MLE: Maximum Liklihood Estimation.

EM: Expectation Maximization.

VQ: Vector Quantization.

HMM: Hidden Markov Modelling.

UBM : Universal Background Model.

CDTA : Centre de Développement des Technologies Avancées.

MATLAB : MAtrix LABoratory.

PC : Personal Computer.

VAD : Vaccal Activity detection

Introduction Générale

Savoir déterminer de manière à la fois efficace et exacte l'identité d'un individu est devenu un problème critique dans notre société. En effet, bien que nous ne nous en rendions pas toujours compte, notre identité est vérifiée quotidiennement par de multiples organisations : lorsque nous utilisons notre carte bancaire, lorsque nous accédons à notre lieu de travail, lorsque nous nous connectons à un réseau informatique, etc.

Il existe traditionnellement deux manières d'identifier un individu. La première méthode est basée sur une connaissance. Cette connaissance correspond par exemple au mot de passe utilisé au démarrage d'une session Unix ou au code qui permet d'activer un téléphone portable. La seconde méthode est basée sur une possession. Il peut s'agir d'une pièce d'identité, une clef, un badge, etc. Ces deux modes d'identification peuvent être utilisés de manière complémentaire afin d'obtenir une sécurité accrue. Cependant, elles ont leurs faiblesses respectives. Dans le premier cas, le mot de passe peut-être oublié par son utilisateur ou bien deviné par une autre personne. Dans le second cas, le badge (ou la pièce d'identité ou la clef) peut être perdu ou volé. [1]

A la différence de ces deux manières traditionnelles, la biométrie consiste à identifier une personne à partir de ses caractéristiques physiques ou comportementales. Le visage, les empreintes digitales, etc. sont des exemples de caractéristiques physiques. La voix, l'écriture, le rythme de frappe sur un clavier, etc. sont des caractéristiques comportementales. Ces caractéristiques sont uniques à un l'individu et il y a peu de possibilités que d'autres personnes puissent les remplacer et elles ne peuvent pas être oubliés ou perdus. D'autre part, le domaine judiciaire a considérablement évolué en résolvant des crimes se basant sur l'identification des individus en utilisant ces caractéristiques prélevées sur les scènes de crimes. [2]

C'est dans ce contexte que nous nous intéressons à la reconnaissance du locuteur qui est l'objet des recherches les plus récentes en terme d'amélioration de performances. En pratique , de nombreux facteurs liés aux conditions d'enregistrement peuvent dégrader significativement les performances de ces systèmes. Ces facteurs peuvent être en relation avec l'environnement (bruit additif, réverbération, etc.), le dispositif d'enregistrement (variabilité du canal) ou le locuteur luimême (état psychologique, effort vocal, changement de voix, etc.). Souvent, ces facteurs ne peuvent pas être connus à l'avance, ce qui représente un défi pour les applications réelles. [3]

Par conséquent notre mémoire est organisé comme suite :

Dans le premier chapitre de ce travail, nous allons faire un état de l'art du traitement de la parole pour comprendre les différents concepts qui tournent au tour de la parole.

Dans le deuxième chapitre de ce travail, nous nous intéresserons au système de reconnaissance du locuteur et détailler ses différentes étapes.

Enfin dans le troisième chapitre, nous allons essayer de compenser pratiquement la variabilité du canal qui, nous allons le voir, cause énormément de dégradations non négligeables.

Chapitre I

Généralités sur le traitement de la parole

I.1) Introduction :

La parole est le principal moyen de communication dans toute société humaine. Son apparition peut être considérée comme concomitante à l'apparition des outils, l'homme ayant alors besoin de raisonner et de communiquer pour les façonner. L'importance de la parole fait que toute interaction homme-machine devrait plus ou moins passer par elle. D'un point de vue humain, la parole permet de se dégager de toute obligation de contact physique avec la machine, libérant ainsi l'utilisateur qui peut alors effectuer d'autres tâches. [4]

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur, son importance s'explique par la position privilégiée de la parole comme vecteur d'information dans notre société humaine.

Cette science a pour rôle d'essayer de reproduire la fonction de certaines taches de la perception de la parole ainsi que la fonction de production de la parole, qui sont en générale des fonctions spécifique au cerveau humain.

Les systèmes automatiques issus du traitement de la parole remplient des taches très importantes qui sont résumées dans le schéma suivant[4] :

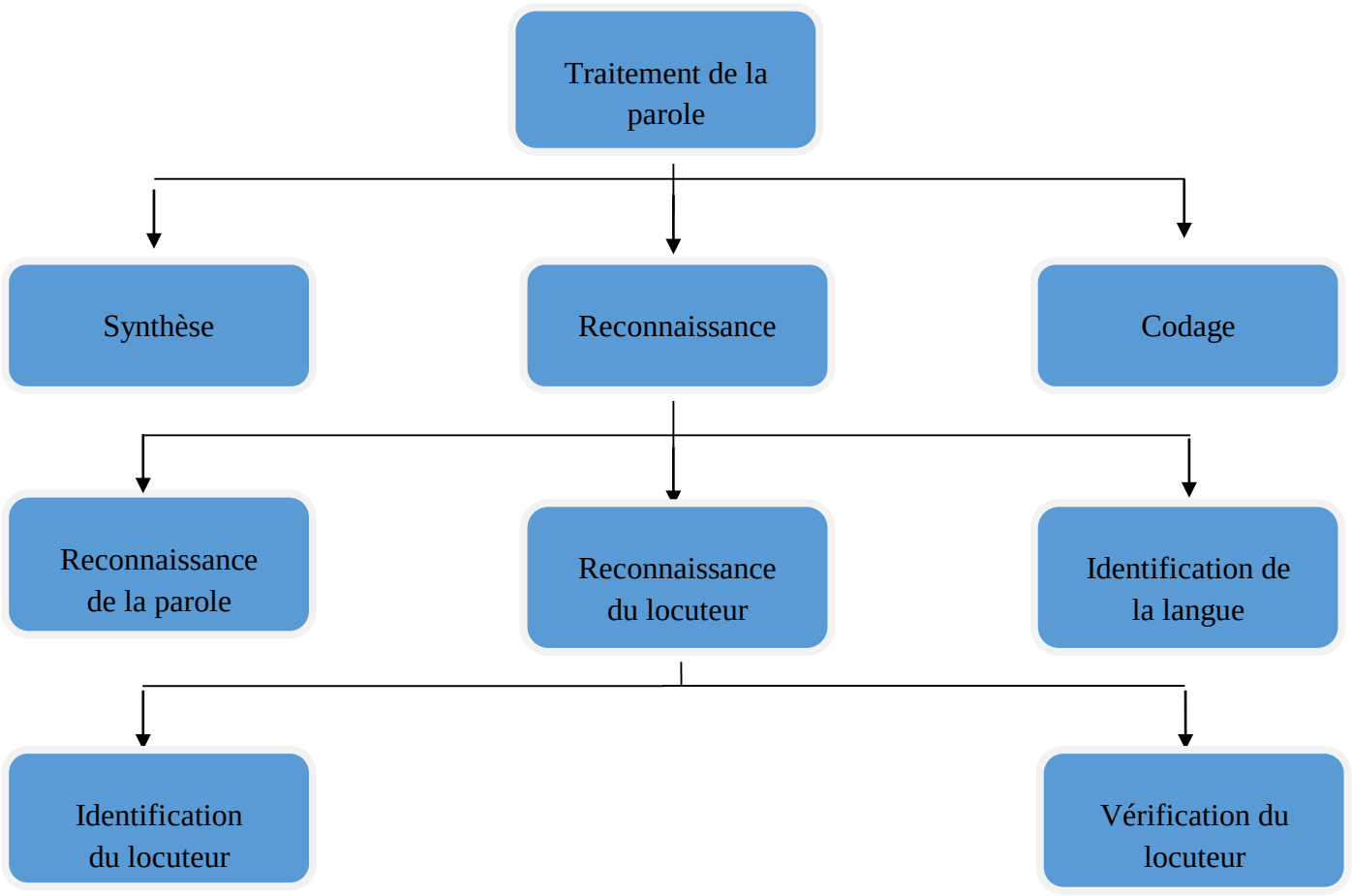


Figure I.1. Traitement de la parole.

Dans le cas de la synthèse, nous disposons d'une entrée texte (ou mots-clés) à partir de laquelle nous devons produire de la parole de synthèse (faire parler l'ordinateur) [4].

Dans le cas du codage, on a pour but de réduire la taille de l'information pour la transformation et le stockage. [4]

Dans le cas de la reconnaissance, nous disposons d'un signal de parole dont nous devons déduire une information [4].

Parmi les systèmes de reconnaissance, nous pouvons citer :

- ✓ Les systèmes de reconnaissance de la parole
- ✓ Les systèmes de reconnaissance de la langue.
- ✓ Les systèmes de reconnaissance du locuteur.

I.2) Production de la parole :

Dans notre étude, la parole entre comme le principal sujet à mettre en analyse. Cependant pour pouvoir avancer, il est nécessaire de connaitre les paramètres caractérisant la parole humaine. Pour cela il faut comprendre le phénomène de production de cette dernière. Dans ce qui suit nous allons aborder d'une façon simple ce mécanisme :

I.2.1) Les différents organes de production de la parole :

Trois groupes d'organes assument les fonctions essentielles dans l'acte de parole, ou phonation [5] (FigureI.2) :

- ➢ **Partie subglottique « appareil respiratoire » :** composé de :
  - ✓ Diaphragme
  - ✓ Poumons
  - ✓ Trachées
- ➢ **Partie glottique « larynx » :** composé des cordes vocales
- ➢ **Partie supraglottique « Le conduit vocal » :** composé de :
  - ✓ Pharynx
  - ✓ Cavité buccale
  - ✓ Cavité nasale

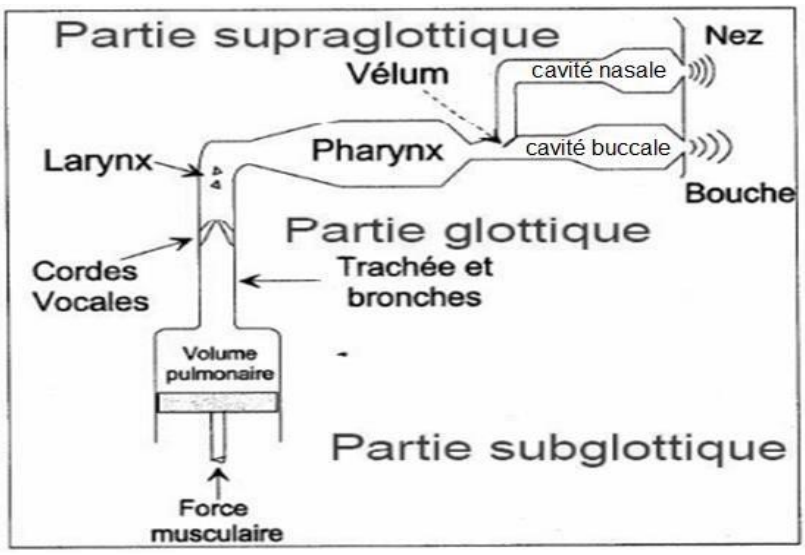


Figure I.2 : Schéma simplifié de l'appareil phonatoire

I.2.2) Mécanisme de production de la parole :

Le mécanisme de production de la parole convertit l'énergie musculaire en énergie acoustique, le tout, engendré et contrôlé par le système nerveux central. Le schéma suivant résume les différentes étapes de la production de la parole allant de la planification au mouvement puis à la parole [6] :

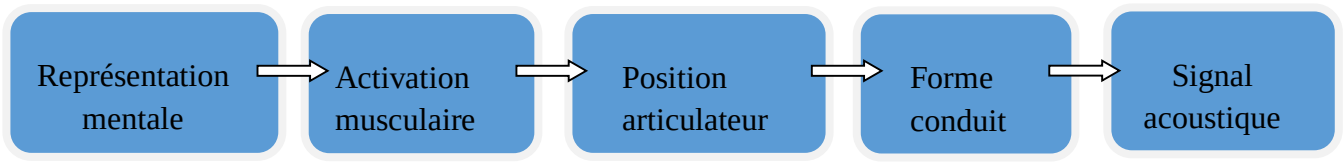


Figure I.3 : Schéma simplifié de production de la parole.

- ✓ **Représentation mentale :** C'est l'étape de préparation conceptuelle. Elle implique la préparation d'une intention de communication qui génère une commande motrice.
- ✓ **Activation musculaire :** La commande motrice arrive vers les différents muscles de l'appareil phonatoire. Les poumons fournissent l'énergie nécessaire en insufflant l'air vers la partie glottique.
- ✓ **Position articulateur :** Cette activation musculaire entraine la musculation de l'état initial des articulateurs (langue, mâchoire, lèvres…) en fonction de l'ordre établit.
- ✓ **Forme du conduit vocal :** Les mouvements des articulateurs entrainent des modifications de forme du conduit vocal, qui cause des transformations de la cavité de résonance.
- ✓ **Signaux acoustiques :** Ces transformations se traduiront par des changements du signal acoustique on obtient alors une variation de la pression d'air en extérieur qui est le signal parole en sortie. La parole naît de l'excitation de la cavité résonante. L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant l'air à travers l'appareil phonatoire, vers la source du résonateur [7].

La source du résonateur est en fait décomposable en deux émissions distinctes et d'origines différentes [8] :

- ✓ Les cordes vocales, qui possèdent la particularité de produire, en plus de leur fréquence fondamentale, un spectre riche en harmoniques ; elles produisent les sons voisés.
- ✓ Le bruit d'écoulement de l'air en provenance des poumons, dont le spectre est similaire à un bruit blanc; il crée les sons non-voisés.

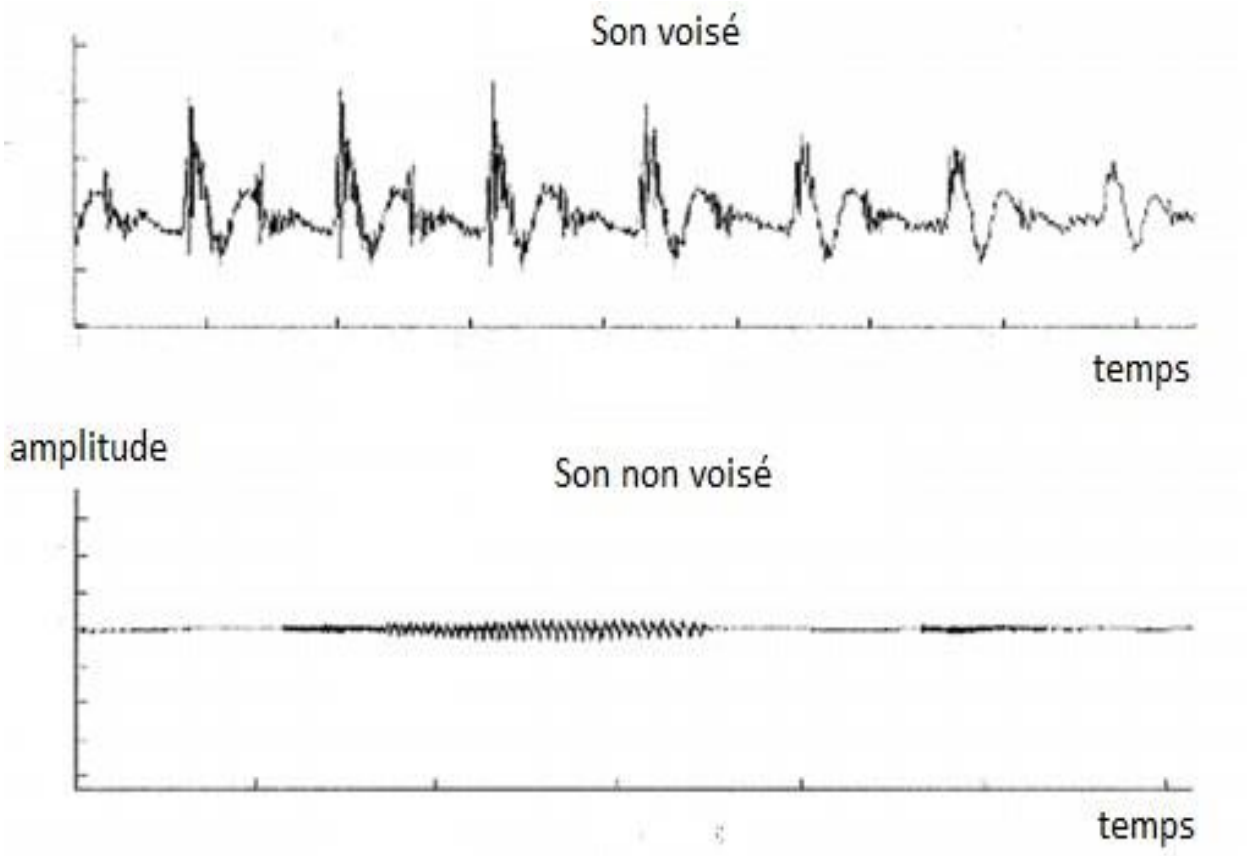


Figure I.4 : Comparaison d'un son voisé et d'un son non-voisé.

I.2.3) Modélisation du mécanisme de production de la parole :

L'appareil phonatoire humain est formé de différentes parties qui peuvent nous sembler complexes. Cependant, il peut facilement être assimilé, et même souvent représenté (figure I.5) comme un système composé simplement d'une source vibrante et d'un filtre (résultant du conduit vocal qui est formé d'une cavité résonante complexe) [9], [10], [11].

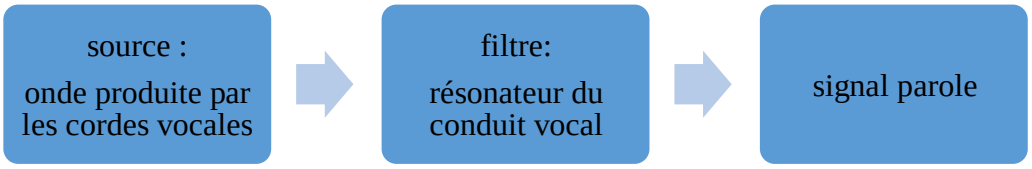


Figure I.5 : Schéma source-filtre de production de la parole

En effet, lorsque le signal provenant des cordes vocales passe à travers le conduit vocal son contenu sera altéré. L'observation spectrale du conduit vocal (figure I.6) laisse apparaître une enveloppe spectrale formée de pics de résonance, appelés formants, et d'affaiblissements nommés anti-formants (introduits par les sons nasalisés) [8].

Ainsi, chaque locuteur possédant une cavité résonante qui lui est propre, va produire un signal vocal qui aura été altéré par les caractéristiques de cette cavité et qui contiendra par conséquent de l'information sur ce dernier.

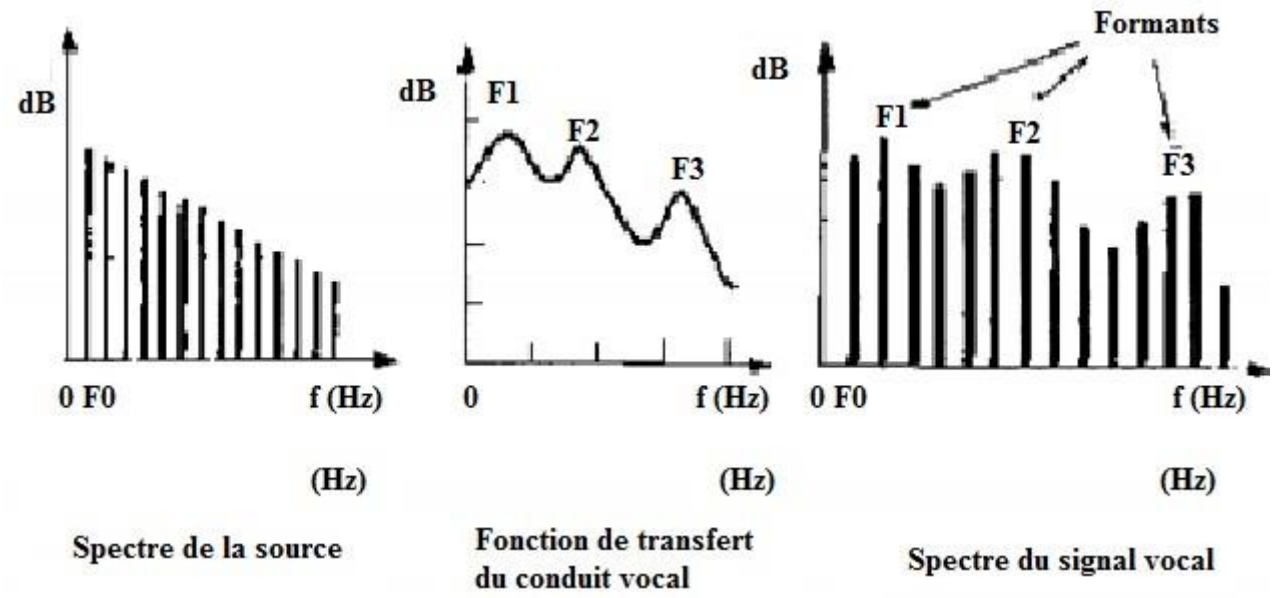


Figure I.6 : Observation spectrale du conduit vocal.

I.3) La perception de la parole :

En traitement de la parole, il est essentiel d'avoir aussi une bonne connaissance des mécanismes de l'audition et des propriétés perceptuelles de l'oreille si on désire extraire l'information de façon pertinente. Tout ce qui peut être mesuré acoustiquement ou observé par la phonétique articulatoire n'est pas nécessairement perçu et par conséquent il y a une partie du signal qui est inutile d'analyser. Ceci nous permet alors une réduction des données, et un gain en vitesse de traitement.

I.3.1) Le système auditif :

Nous présentons dans la (figure I.7) suivante les principales constituantes du système auditif [12]

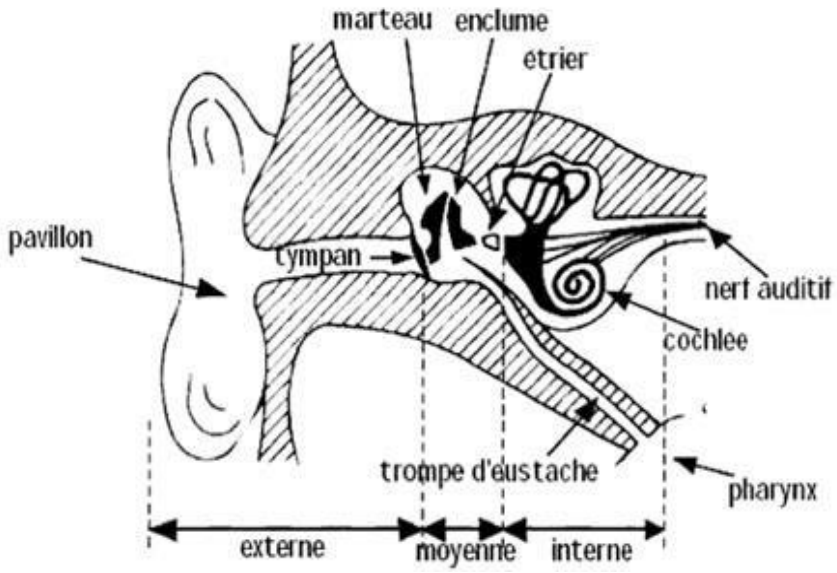


Figure I.7 : Système auditif.

Ce qui nous intéresse dans notre étude est la cochlée, Elle est constituée de cellule ciliée qui ont pour fonction la conversion du signal mécanique de la parole (son en général) vers des impulsions nerveuses.

I.3.2) Analyse fréquentielle et échelles de modélisation :

L'oreille effectue une sorte d'analyse fréquentielle du signal acoustique. Ainsi, Il existe environ 25 000 cellules ciliées qui sont réparties au niveau de la cochlée. Chaque cellule ciliée vibre à une certaine fréquence dite de résonance (fréquence qui dépend de la position de la cellule) [13].

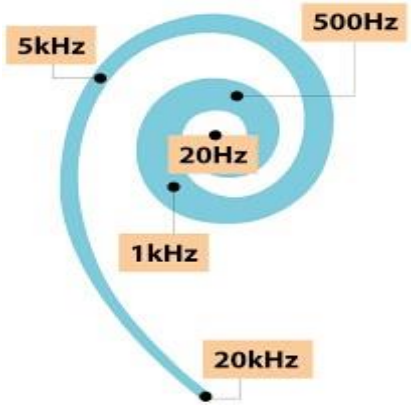


Figure I.8 : Représentation de la dépendance de la fréquence résonnance par rapport à la position de la cellule ciliée.

Les travaux de Fletcher ont démontré l'existence de bandes critiques dans la réponse de la cochlée [14]. Cette dernière agit comme si elle était constituée d'un banc de filtres dont les bandes, appelées "bandes critiques", se chevauchent, et dont les fréquences centrales s'échelonnent continûment. Cette bande critique correspond à l'écartement en fréquence nécessaire pour que deux harmoniques soient discriminées dans un son complexe périodique [15], [16]. Une des modélisations de ces bandes critiques est appelée 1' échelle de fréquences de Bark [17]. Une autre échelle de perception des fréquences fut proposée par Mel [18].

I.3.2.1) L'échelle de Bark :

En traitant l'énergie spectrale avec l'échelle de Bark, les chercheurs espéraient qu'on puisse aboutir à un traitement de l'information spectrale aussi proche que celui effectué au niveau de 1'oreille. L'échelle de Bark s'étend de 1 à 24 Barks, ce qui correspond aux 24 bandes critiques de l'audition. Notons que les bandes critiques de l'oreille sont continues, et un son de n'importe quelle fréquence audible se trouve toujours centré sur une bande critique. La fréquence de Bark B peut être exprimée en termes de fréquence linéaire (en Hz) par [19] :

$$B(f) = \frac{26.81\mathrm{f}}{1960+\mathrm{f}} - 0.53 \text{ (Exprimée en Barks) (B)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(I.1)}$$

Cette approximation s'accompagne de deux fonctions de correction : une pour les basses fréquences, qui assure que la courbe approche au mieux les valeurs standard arrondies ; et l'autre pour les hautes fréquences, qui a pour but de rectifier les mauvaises valeurs obtenues par calcul pour B(f) > 20.1:

$$B' = \begin{cases} B + 0.15(2 - B) SiB\ (f) < 2 \\ B + 0.22(B - 20.1) SiB(f) > 20.1 \end{cases} \quad \text{..... (I.2)}$$

Sur la (figure I.9), nous avons la représentation de l'échelle de Bark (approximée par Traunmuller[19]) en fonction de la fréquence exprimée en Hz :

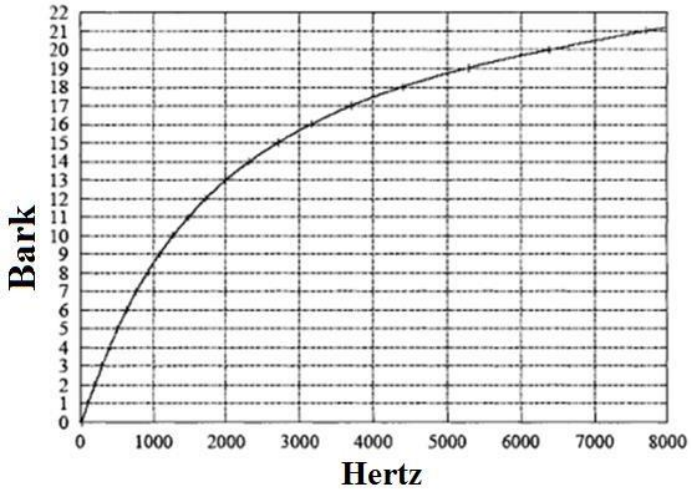


Figure I.9 : Tracé de l'échelle de Bark en fonction de la fréquence (en Hz).

I.3.2.2) Echelle de Mel :

L'audition humaine a tendance à percevoir comme identique deux sons séparés par une octave. En dessus de 1000 Hz, l'oreille humaine perçoit une octave comme un doublement de fréquence. Au-delà de 1000 Hz, ce ne sera plus le cas. Des expériences psycho acoustiques ont alors permis d'établir la loi qui relie la fréquence et la hauteur perçue : l'échelle des Mels où le « Mel » est une unité représentative de la hauteur perçue d'un son (figure I.10) [11], [16].

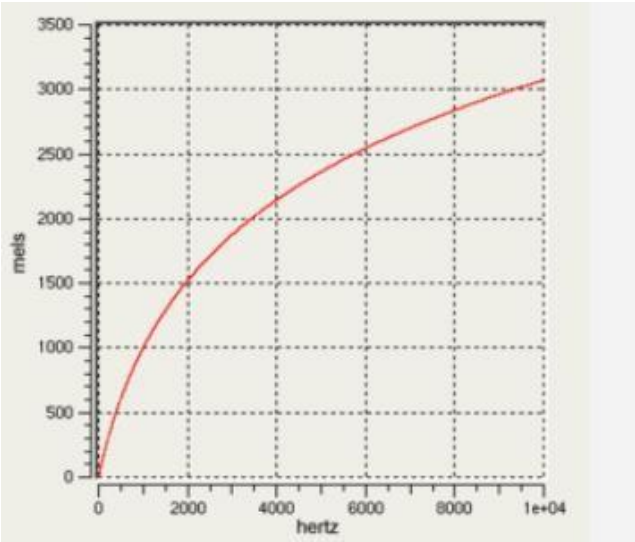


Figure I.10 : L'échelle des Mels [11]

Si on considère une fréquence de base de 1000Hz, alors la fréquence à l'octave supérieure est 2000 Hz. Cependant, un son à 2000 Hz ne sera pas perçu par un auditeur comme étant 2 fois supérieur à la fréquence de base. Par contre, si on effectue le changement d'octave dans le domaine

de Mel, 1000 Hz correspondent à 1000 mels. L'octave supérieure est donc à 2000 mels, ce qui équivaut à environ 3428 Hz. Un son ayant une fréquence de 3428 Hz sera donc perçu comme étant deux fois plus aigu qu'un son à 1000 Hz.

La conversion d'Hertz en Mels se fait à l'aide d'une des formules suivantes :

$$m = 1127.01048 \ln\left(1 - \frac{f}{700}\right) = 2595\ log_{10}\left(1 + \frac{f}{700}\right) \quad \text{(I.3)}$$

La formules réciproques (pour passer de Mels en hertz) est donc :

$$f = 700\ \left(e^{\left(\frac{m}{1127.01084}\right)} - 1\right) = 700\ (10^{\frac{m}{2595}} - 1) \quad \text{(I.4)}$$

I.4) La Reconnaissance :

Doter la machine des capacités de compréhension des comportements humains : tel est le défi scientifique autour duquel se rassemblent différentes communautés scientifiques (traitement du signal, traitement automatique du langage, intelligence artificielle, interaction homme-machine, etc.). L'un des signaux fréquemment utilisé est le signal de parole. La parole est en effet l'une des modalités fondamentales que l'homme utilise pour communiquer, ainsi la reconnaissance automatique au tour de la parole se voit diversifié selon ses applications. Dans ce qui suit nous allons voir les différents types de la reconnaissance automatique en traitement de la parole.

I.4.1) La Reconnaissance automatique de la parole :

La reconnaissance de la parole est une technique informatique qui permet d'analyser la parole humaine captée au moyen d'un microphone pour la transcrire sous la forme d'un texte exploitable par une machine. L'objectif de la Reconnaissance Automatique de la Parole (RAP), est d'extraire l'information textuelle contenue dans un signal de la parole à l'aide d'un logiciel informatique, de commander un ordinateur ou un appareil mobile, De faire de la traduction simultanée sur le Web…

Le schéma suivant résume les étapes de base d'un système RAP [20] :

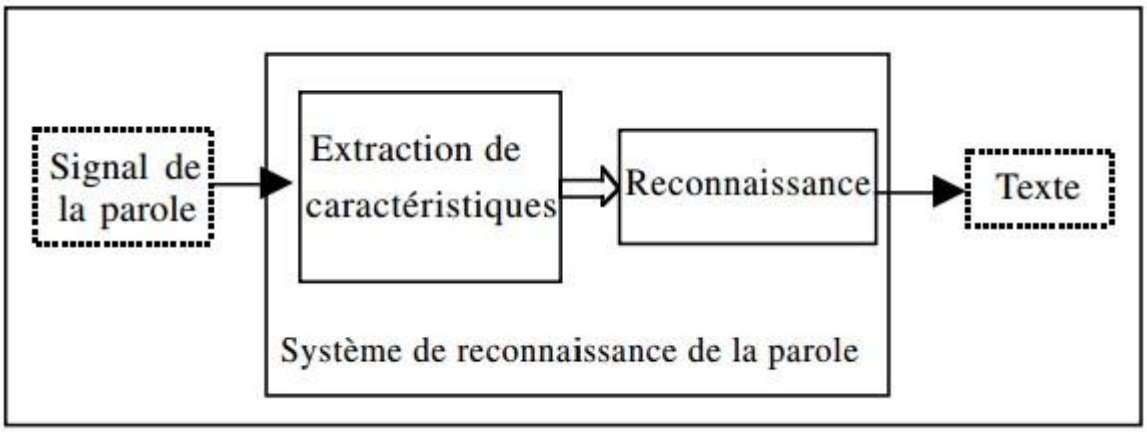


Figure I.11: Blocs de base composant le système RAP.

- **Extraction de caractéristique** : permet d'extraire l'information utile à la caractérisation de son contenu linguistique en réduisant la redondance du signal de la parole. Le signal sonore brut est converti en une séquence de vecteurs acoustiques adaptée à la reconnaissance.

- **Reconnaissance :** Dans la reconnaissance plusieurs modules entre en jeu :
  - ✓ **Les modèles acoustiques :** Un ensemble réduit d'unités de sons élémentaires d'une langue donnée.
  - ✓ **Un module lexical :** fourni la transcription de mots de la langue modélisée par un simple dictionnaire phonétique.
  - ✓ **Un module de langage :** introduit la notion de contraintes linguistiques par un modèle statistique utilisant une grande base de données textuelles pour estimer les probabilités d'une suite de phonèmes, de manière automatique. Il permet de guider le décodeur vers les suites de mots les plus probables.
  - ✓ **Un module de décodage :** consiste à sélectionner, parmi l'ensemble des phrases possibles, celle qui correspond le mieux à la phrase prononcée. Le décodage de la parole s'effectue à l'aide de tous les modules déjà présentés.

I.4.2) La reconnaissance automatique de langage :

Une langue est définie comme un regroupement de dialectes partageant un vocabulaire similaire et ayant des systèmes phonologiques et grammaticaux similaires. Par exemple, la langue française est constituée de dialectes comme le français méridional ou provençal (accent de Marseille), le français parisien, le français du sud-ouest (accent toulousain), etc. Il existerait actuellement plus de 6000 langues parlées `à travers le monde et plus de 10000 dialectes [21].

L'identification automatique des langues a pour rôle de reconnaitre la langue parlée par un locuteur inconnu, parmi un ensemble fini de langues. Ainsi cette tache joue un rôle très important dans la consultation de bases de données documentaires multilingues (notamment sur internet), l'enseignement ou encore la traduction automatique et autant d'autres applications issus de l'ouverture mondiale des télécommunications. [21]

Quatre domaines linguistiques [22] représentent la source d'informations discriminantes pour la reconnaissance des langues :

- **La phonologie :** Les fréquences d'apparition des phonèmes (la plus petite unité qu'on peut isoler par segmentation dans la chaine parlée) peuvent être caractéristiques. De plus, les règles d'enchainements de ces unités varient d'une langue à l'autre.
- **La morphologie :** Chaque langue a 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.

Le fonctionnement chaque système de reconnaissance, le système d'identification de la langue se décompose en deux phases (figure I.12) :

- **Phase d'apprentissage :** Des paramètres sont extraits pour les signaux de parole de chaque langue. Pour chaque source d'information prise en compte, un modèle spécifique à chaque langue est appris à partir de ces paramètres.
- **Reconnaissance :** Les paramètres sont extraits pour un signal de parole d'une langue inconnue. La langue la plus vraisemblable est déterminée en fonction des modèles issus de la phase d'apprentissage.

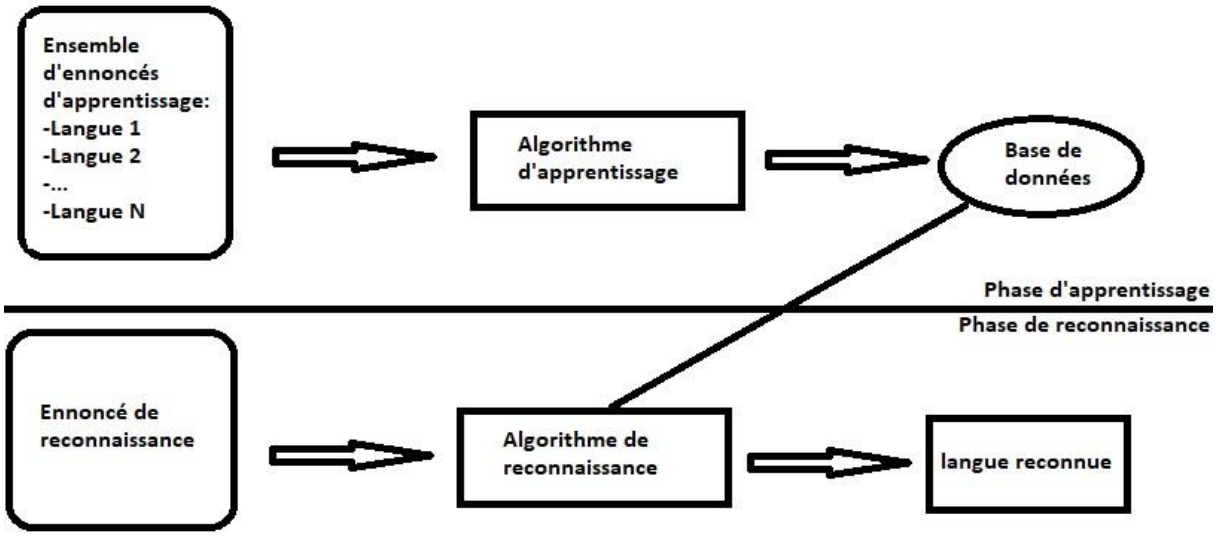


Figure I.12: Structure générale d'un système d'identification automatique des langues.

I.4.3) La reconnaissance automatique du locuteur (RAL) :

La reconnaissance du locuteur est un processus qui consiste à reconnaître automatiquement qui a parlé à partir de l'information qui est contenue dans le signal de parole de chaque individu. Grâce à cette technique, il est possible d'utiliser la voix des personnes pour vérifier leur identité et contrôler l'accès à des services tels que les achats par téléphone, les messageries, le démarrage automatique d'une voiture sécurisée par l'empreinte vocale du propriétaire [23], ou encore, pour résoudre des investigations criminelles en trouvant lequel des suspects correspond aux enregistrements trouvés sur la scène du crime par exemple [24].

Les trois modules principaux qui constituent le fondement de base d'un système de reconnaissance du locuteur [16], [25] sont résumés sur la (figure I.13) :

Figure I.13 : schéma simplifié le système d'avoir reconnaissance du locuteur.

- **Prétraitement et parameterisation :** Le prétraitement permet quelque conditionnement du signal tel que la normalisation de l'amplitude, la suppression des silences …ceci rendra le signal propice à la phase de parameterisation qui consiste à extraire le paramètre acoustique de chaque locuteur.
- **Modélisation :** ce module va créer un modèle de chaque personne à partir des vecteurs de caractéristiques obtenue précédemment
- **Reconnaissance :** Ce module pendant la phase de test du système va tester les nouveaux sujets sur les modèles existants et stockés dans une base de données lors de la phase d'apprentissage du système afin de vérifier au d'identifier le locuteur qui à parler.

I.4.3.1) Deux phases de la reconnaissance du locuteur :

Les systèmes RAL doivent suivre les deux phases suivantes [26] :

- **Phase d'apprentissage :**
  Dans cette phase, chaque locuteur qui sera susceptible d'être testé dans le futur doit fournir des échantillons de parole afin que le système puisse créer un modèle de référence pour chacun d'entre eux.
- **La phase de test :**
  Dans cette phase, le locuteur qu'on cherche à reconnaitre va fournir un signal parole qui sera associé aux modèles de référence collectés après l'apprentissage, ainsi une décision sera prise.

I.4.3.2) Taches de la reconnaissance du locuteur :

Le domaine de la reconnaissance du locuteur offre la réalisation de deux grandes taches [27], [28], [29], [30], [31], [32] :

- **Vérification :**
  Cette tâche consiste à confirmer ou infirmer qu'un locuteur est bien celui qu'il prétend être [33], [34].

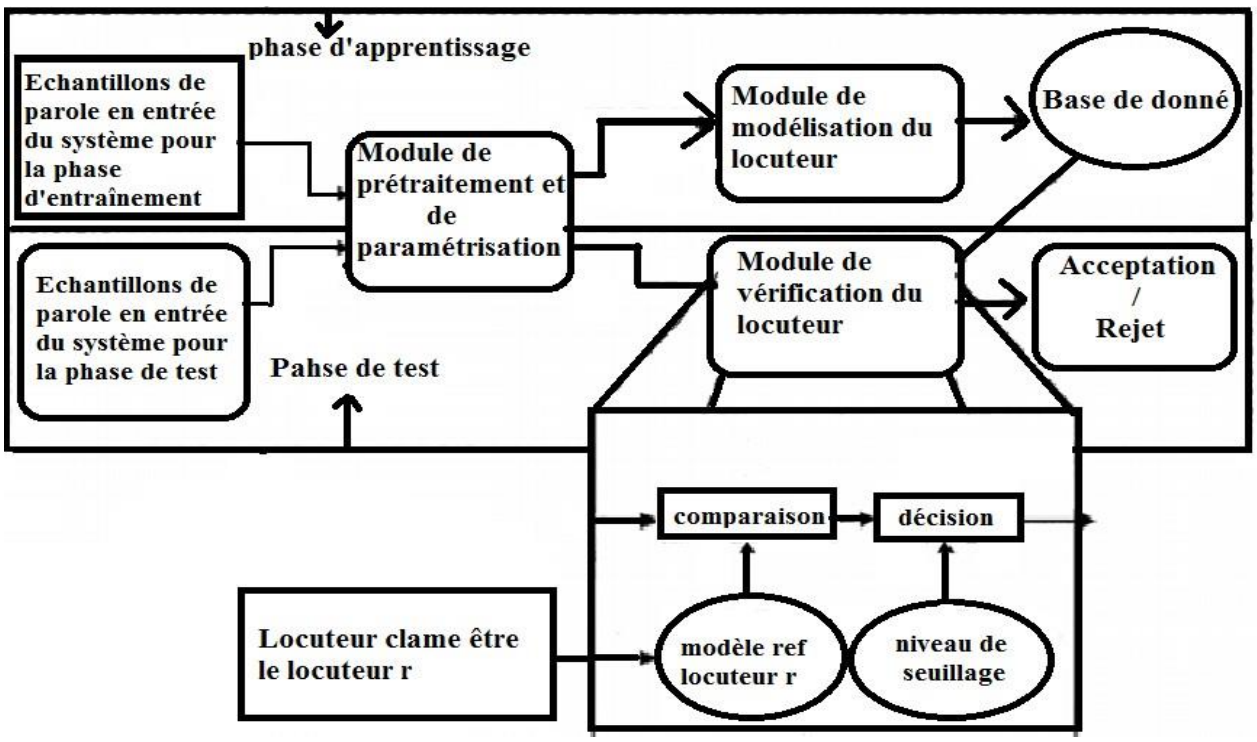


Figure I.14 : Structure de base des systèmes de vérification du locuteur.

- **Identification :** cette tâche consiste à indiquer qui est le locuteur. (Figure I.15)

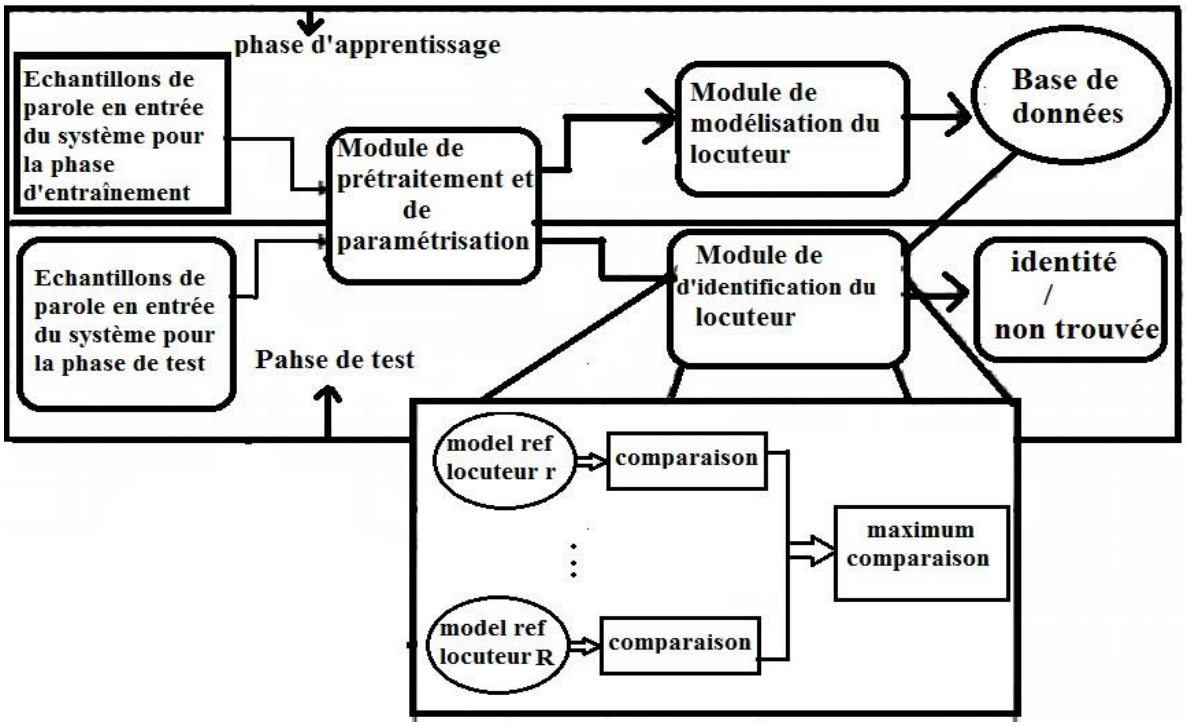


Figure I.15 : Structure de base des systèmes d'identification du locuteur.

I.4.3.3) Méthodes de reconnaissance du locuteur :

Selon leur application, les méthodes de reconnaissance du locuteur se divisent en deux :

- **La méthode indépendante du texte :** dans laquelle la modélisation des locuteurs se fait à partir de caractéristiques indépendantes de ce qui est dit.[23]
- **La méthode dépendante du texte :** qui est basé sur les caractéristiques obtenue lors de la répétition d'une ou plusieurs phrases spécifiques telles que des codes PIN (personal identification Number), des mots de passe ou numéro de carte bancaire. [23],[35],[36]

I.5) Variabilité du signal parole :

Le signal de parole est très complexe où se mêlent informations linguistiques, informations caractéristiques du locuteur, informations relatives au matériel utilisé pour la transmission ou l'enregistrement du signal, etc... Cette caractéristique est d'ailleurs reconnue pour faciliter la communication entre deux personnes dans un environnement très bruyant. Par ces différents aspects, le signal de parole présente une très grande variabilité.

Ainsi on trouve trois types de variabilité :

I.5.1) Variabilité intra-locuteur :

La variabilité intra-locuteur est une variabilité propre au locuteur qui ne peut pas reproduire exactement le même signal. Cette variabilité intra-locuteur est dépendante d'état physique et psychologique pour un même individu, ainsi les facteurs de variabilités sont multiples, on cite :

✓ L'état pathologique : fatigue, rhume et stress, les variations émotionnelles, la voix peut changer entre le début et la fin de la journée etc.

- ✓ Dans le cas d'une interaction volontaire et consciente avec un système de reconnaissance, le comportement d'un individu se modifie au fur et à mesure de son utilisation du système. L'individu devient de plus en plus confiant ainsi sa voix évolue dans ce sens et s'en trouve modifiée.
- ✓ Enfin, à plus long terme, la voix change avec le vieillissement d'une personne.

I.5.2) Variabilité interlocuteurs :

Les signaux de parole véhiculent plusieurs types d'informations. Parmi eux, la signification du message prononcé est d'importance primordiale. Cependant, d'autres informations telles que le style d'élocution ou l'identité du locuteur jouent un rôle important dans la communication orale. Ecouter un interlocuteur permet d'avoir des indications concernant son sexe, son état émotionnel et bien souvent de l'identifier si on l'a déjà entendu.

La grande variabilité entre les locuteurs est due aux différences physiologiques des organes responsables de la production vocale. L'expression acoustique de ces différences peut être traduite par une variation de la fréquence fondamentale, dans l'échelle des formants (plus haute chez les femmes et les enfants que chez les hommes) et dans le timbre de la voix (richesse en harmoniques due à la morphologie du locuteur et au mode de fermeture des cordes vocales).

I.5.3) Variabilité intersession :

La transmission du signal de parole au système de reconnaissance chargé de l'analyser nécessite plusieurs étapes et emprunte divers types de supports. A chacune de ces étapes, le media utilisé (ex : microphone, combiné téléphonique) pour transporter ce signal y imprime sa marque. Ces empreintes apparaissent le plus souvent sous la forme de déformations/dégradations du signal de parole. Ces déformations sont différentes selon le type de matériel utilisé.

I.6) Conclusion :

Dans ce chapitre nous avons introduit la notion de traitement de la parole et définir ses différents aspects pour enfin comprendre l'emplacement de la reconnaissance du locuteur dans ce domaine.

Nous avons pu comprendre aussi les différents défis imposés par cette science pour aboutir à la reconnaissance :

- Le signal de la parole véhicule plusieurs types d'informations. Par conséquent, il est nécessaire aux systèmes de reconnaissance de n'extraire que l'information nécessaire à son application.
- La nécessité de s'inspirer du système auditif humain pour aboutir à des résultats pertinents.

- La nécessité de d'utiliser des outils mathématiques tels que les transformées, afin de passer dans le domaine fréquentiel et faire apparaître les caractéristiques.

Nous avons enfin abordé de façon générale les différentes étapes qui constituent un système de reconnaissance du locuteur afin de pouvoir se pencher sur les différentes techniques qui constituent chaque étape dans la suite de durant la suite de notre travail.

Chapitre II

Le système de reconnaissance automatique du locuteur

II.1) Introduction :

La reconnaissance automatique du locuteur est interprétée comme une tache particulière de reconnaissance de formes. Ses applications sont principalement liées aux problèmes d'authentification ou de confidentialité. La variabilité de la parole entre locuteurs (variabilité interlocuteur) est l'essence même de la RAL. Sans cette variabilité, il serait impossible de reconnaître une voix parmi plusieurs voix possibles. [37]

Comme nous l'avons vu précédemment un système RAL est basiquement composé des modules suivants : extraction des paramètres, la modélisation, la comparaison, et la décision, dans ce qui suit nous allons détailler ces modules en donnant les principales techniques de traitement de chaque étape

II.2) Parameterisation :

Cette phase a pour but de créer des éléments vectoriels riches en informations sur ce qui est dit ou sur qui le dit. Ces éléments seront dépourvus de redondance. Ils représentent de fortes caractéristiques discriminatoires entre locuteurs.

II.2.1) Propriété liée au signal parole :

Le signal de parole est un signal qui varie lentement dans le temps, il est non-stationnaire, cependant il peut être considère comme quasi-stationnaire sur de courtes intervalles de temps (entre 10 ms et 30 ms). C'est cette propriété qui va nous permettre d'appliquer les différentes techniques d'analyse propre aux signaux stationnaires. [7], [16]

II.2.2) les différents types de paramètres :

II.2.2.1) les paramètres prosodiques :

Le terme "paramètres prosodiques" réunit l'énergie, la durée et la fréquence fondamentale (ou pitch) et le timbre.

L'énergie :

L'énergie du son est liée à la pression de l'air en amont du larynx. L'énergie ou l'intensité est une sensation auditive basée sur la perception de la force du signal acoustique. Elle correspond à la variation de l'amplitude du signal de parole causée par une force plus ou moins forte provenant du pharynx et provoquant une variation de la pression de l'air sous la glotte. Ce paramètre permet de fournir une mesure de la force sonore de la voix (faible ou forte). L'énergie à court terme d'un signal échantillonné sur une fenêtre de longueur T, $(S_t)_{t=1,T}$, est définie par [16]:

$$E = \frac{1}{T}\sum_{t=1}^{T} S_t^2 \text{ [38]} \quad \text{(II.1)}$$

La fréquence fondamentale :

L'évolution temporelle de la fréquence fondamentale (F0) ou pitch est une information spécifique à chaque locuteur, qui varie en fonction des phonèmes qu'il prononce au cours d'une phrase. Sur la

(figure.II.1), nous pouvons observer l'évolution temporelle de la fréquence fondamentale de la phrase «je me suis enrhumé ». Les sons non-voisés sont associés à une fréquence nulle. Ces résultats sont obtenus avec WaveSurfer [39].

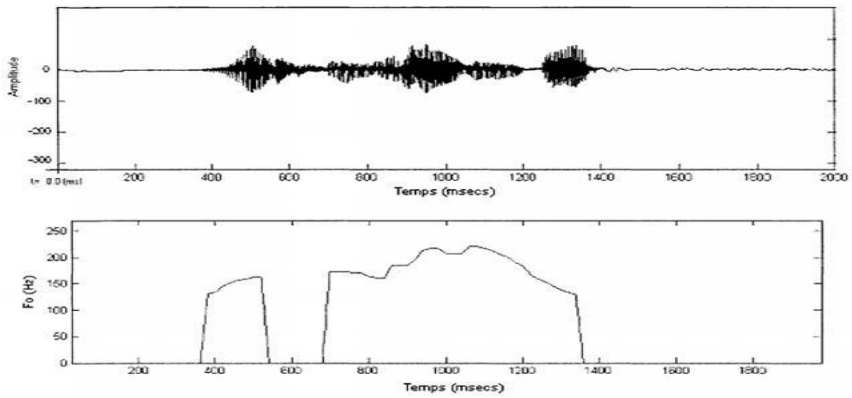


Figure II.1 : Evolution de la fréquence de vibration des cordes vocales dans la phrase « je me suis enrhumé ».

II.2.2.2) Paramètre de l'analyse spectrale :

Plusieurs paramètres peuvent être calculés, dans cette partie nous allons présenter les plus utilisées :

Les coefficients cepstraux de fréquence de Mel :

Les coefficients MFCC (Mel Frequency cepstral coefficients) sont basés sur une échelle de perception non-linéaire qui correspond à la distribution fréquentielle de l'oreille humaine. Le modèle de l'échelle de Mel défini précédemment sera utilisé dans cette méthode. Les principales étapes de calcul de ces coefficients sont illustrées dans la figure suivante [40]:

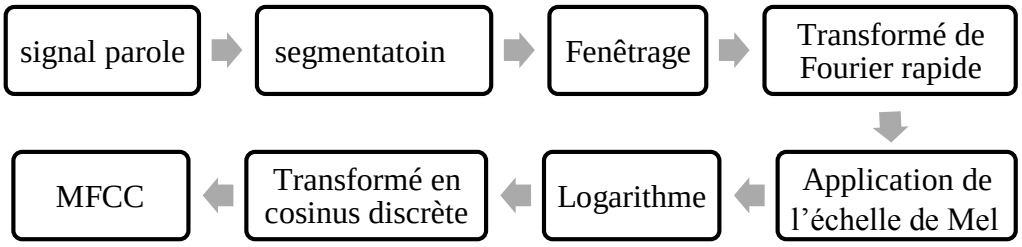


Figure II.2 : Extraction des paramètres MFCC.

- **Parameterisation :**

Comme nous l'avons vu précédemment les signaux de parole sont des signaux non stationnaires. Cependant sur un court intervalle ils peuvent être considérés comme quasi stationnaires, c'est pourquoi il est utile de diviser le signal en segments stationnaire.

Dans cette étape, le signal de parole est divisé sous forme de trames de 'E' échantillons, chaque trame se chevauche à la suivante après « C » échantillons (E◌>C) (figure II.3) :

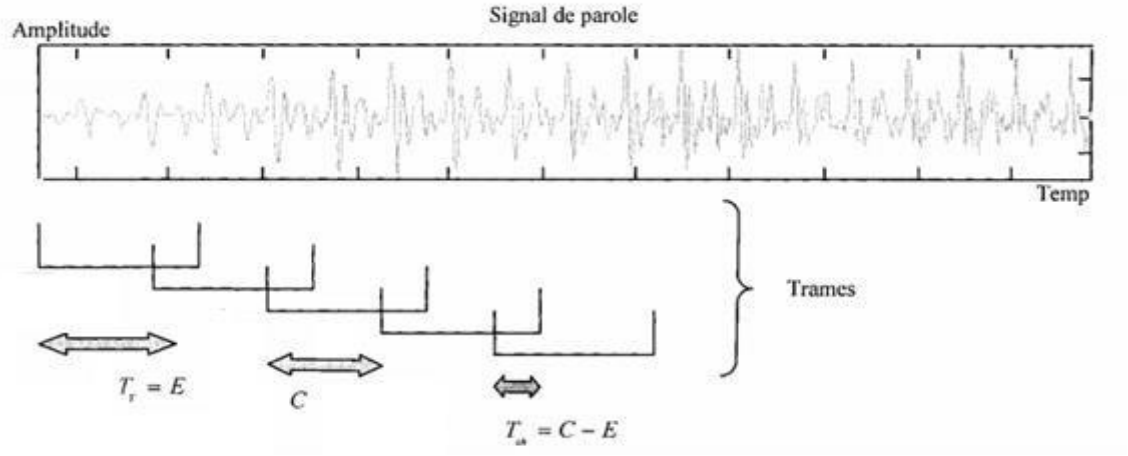


Figure II.3 : Fenêtrage du signal de parole pour l'obtention de vecteurs paramétriques.

- **Fenêtrage :**

Afin de minimiser la distorsion spectrale lors de la transformation du domaine temporelle vers le domaine fréquentiel, on effectue un fenêtrage qui tend à rendre le signal nul au début et à la fin de chaque trame. La fenêtre la plus utilisée est la fenêtre de Hamming :

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(2\,pi * \frac{n}{N}\right) & 0 >= n >= N-1 \\ 0 & sinon \end{cases} \quad \text{(II.2)}$$

- **Transformé de Fourier rapide (Fast fourier transform :FFT):**

On passe du domaine temporel vers le domaine spectral où le signal parole (résultat de convolution de la source par le conduit vocale) devient un simple produit. La FFT est algorithme rapide qui est définie pour un signal x_l de E échantillons par [41] :

$$X_n = \sum_{K=0}^{E-1} X_k e^{\frac{-2j*\pi*K*n}{E}} \quad \text{(II.3)}$$

$pour \quad 0 \leq K \geq E - 1$.

Le résultat obtenu est considéré comme étant le spectre du signal.

- **Application de l'échelle de Mel :**

Le spectre présente plusieurs fluctuations, et afin de réduire la taille des vecteurs cepstraux on ne s'intéresse seulement à l'enveloppe du spectre. Pour réaliser ce lissage nous multiplions le spectre par un blanc de filtre tenant compte de la réponse acoustique que l'oreille humaine.

Un banc du filtres est une série de filtres à bande passante repartie d'une façon équidistante dans l'échelle de Mel.

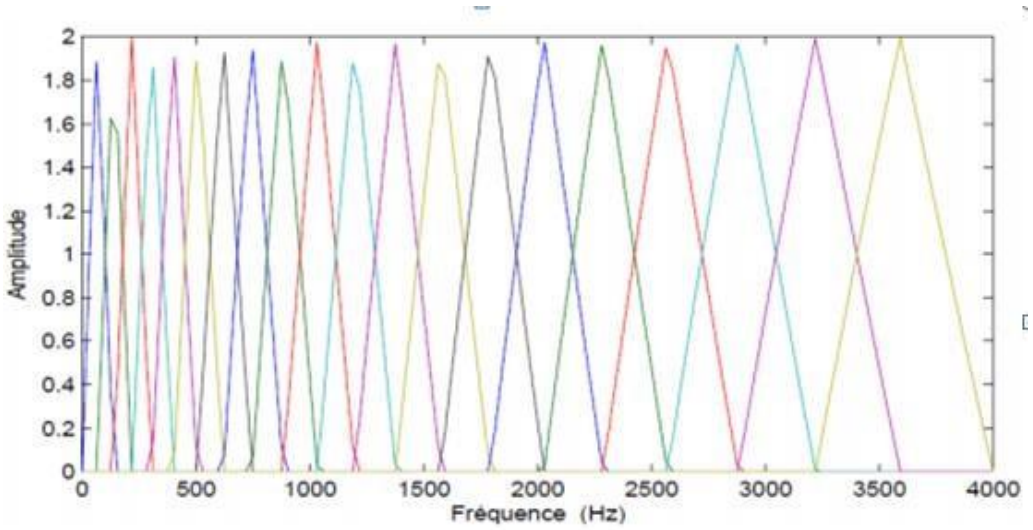


Figure II.4 : Banc de filtres Triangulaires équidistance en échelle Mel.

- **Logarithme :**

Nous prenons le logarithme de cette enveloppe et nous multiplions chaque coefficient par 20 afin d'obtenir l'enveloppe spectrale en dB.

- **Transformé en cosinus discrète (Discrete Cosine Transform : DCT) :**

Les coefficients cepstraux sont obtenus par une transformée en cosinus discrète à partir des logarithmes des énergies issues du banc de filtres. L'expression de ces coefficients :

$$C_n = \sum_{k=1}^{L} S_k * \cos[n(k-\frac{1}{2})\frac{\pi}{k}] \quad \text{(II.4)}$$

Avec : k=1, 2, 3, …, L ;

S_k : Les coefficients spectraux calculés précédemment ;

L : le nombre de coefficients nombre de coefficients spectraux calculés précédemment.

<u>Le codage prédictif linéaire (Linear prédiction coefficients : LPC) :</u>

Cette méthode a pour objectif une représentation directe du signal vocal sous la forme d'un nombre limité de paramètres. Le principe de cette méthode est fondé sur l'hypothèse selon laquelle un échantillon du signal de parole x (nTs), où « Ts » est la période d'échantillonnage, peut être prédit approximativement par une somme pondérée linéairement de « p » échantillons le précédant immédiatement, p est appelé l'ordre de prédiction. [25]

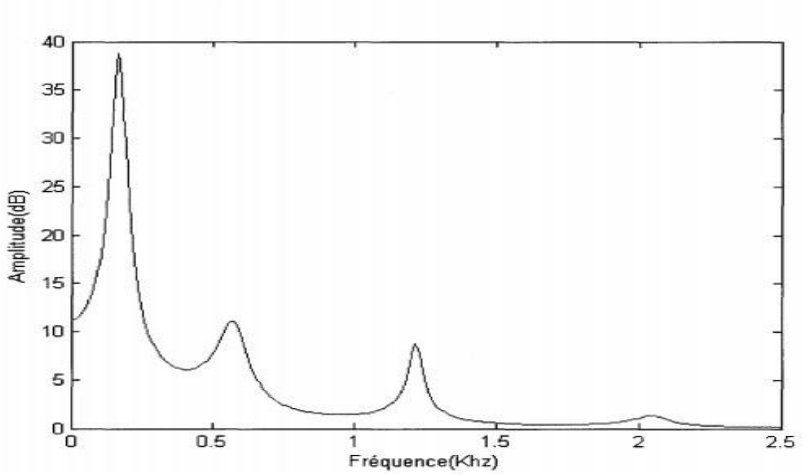


Figure II.5 : Exemple d'un spectre LPC.

On distingue plusieurs modèles de prédiction, les modèles auto régressifs **AR** est le plus utilisé.

- **Modèle AR :**

Comme vu précédemment, on peut représenter le mécanisme de production de la parole par le système suivant :

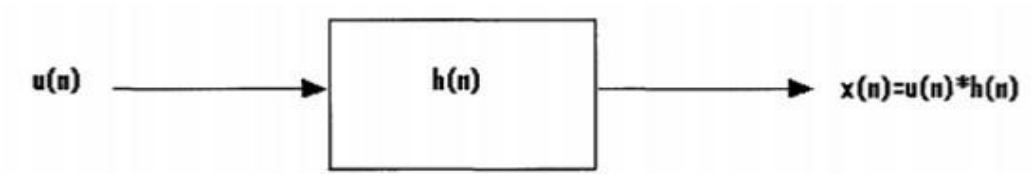


Figure II.6 : Modèle simplifié de la production de la parole.

Où h(n) est la réponse impulsionnelle du filtre et u(n) un signal d'excitation.

Nous savons que pour un signal voisé l'excitation est un train périodique d'impulsions d'amplitude unité :

$$u(n) = \sum_k \delta\,(n - kP) \quad \text{(II.5)}$$

Avec P la période du fondamental.

Une approximation nous donne généralement La transformée en Z de H(n) par [42] :

$$TZ\big(h(n)\big) = \frac{\sigma}{A(z)} = \frac{\sigma}{1-\sum_{i=1}^{p} a(i)Z^{-i}} \quad \text{(II.6)}$$

Donc on peut exprimer X(Z) :

$$X(z) = U(z) * \frac{\sigma}{A(z)} \quad \text{(II.7)}$$

Ce qui nous donne dans le domaine temporel :

$$X(n) = \sum_{i=1}^{p} a(i)x(n-1) + \sigma u(n) \quad \text{(II.8)}$$

Ainsi on peut prévoir l'échantillon x(n) à partir d'une combinaison linéaire des « p » échantillons qui le précèdent. Les coefficients « a(i) » du filtre sont appelés les coefficients de prédiction.

- **L'estimation des coefficients de prédiction :**

Après avoir établi le modèle, reste le problème de déterminer ses paramètres optimaux, c'est à dire les coefficients du filtre tous pôles pour lesquels l'erreur de prédiction est minimale, avec la seule information à priori le signal de sortie, le signal de l'entrée lui étant inconnu. Supposons qu'on a une estimation (ou prédiction) d'un échantillon $\tilde{x}$(n) à partir des « p » échantillons qui le précèdent. L'erreur de prédiction est définie par :

$$e(n) = x(n) - \tilde{x}(n) \quad \text{(II.9)}$$

Ainsi on peut définir l'énergie résiduelle de prédiction par [42] :

$$E_n = \sum_m e^2(m) = \sum_m (x_n(m) - \tilde{x}(m))^2 \quad \text{(II.10)}$$

La minimisation de cette erreur est à la base de la détermination des coefficients de prédiction « a(i) ». Ainsi les coefficients a_k optimaux seront tirés de l'équation suivante :

$$\frac{\delta E_n}{\delta a_n} = 0 \quad \text{(II.11)}$$

Un système d'équations en découle appelée équations de « Yule Walker » dont la résolution a été approchée par plusieurs méthodes comme : « la méthode de l'autocorrélation » et « la méthode de la covariance » [42].

<u>Les Coefficients cepstraux de prédiction linéaire (Linear Prediction Cepstral Coefficients LPCC) :</u>

Par définition Le cepstre est le résultat de la transformée de Fourier inverse appliquée au logarithme de la transformée de Fourier du signal de parole. Son but est de simplifier l'opération consolative qui se passe entre le signal de la source et la fonction de transfert du conduit vocal. À partir des coefficients prédis par la méthode LPC, on peut calculer les paramètres cepstraux.

soient a_k les coefficients calculés avec la méthode LPC tel que $k = 1,2, \ldots, p$,alors on peut calculer les coefficients LPCC suivant les équations suivantes [43] :

$$C_0 = \log_e p \quad \text{(II.12)}$$

pour $1 \leq m \leq p$:

$$C_m = a_m + \sum_{m}^{m-1} \frac{k}{m} * C_k * a_{m-k} \quad \text{(II.13)}$$

Pour $m > k$:

$$C_m = \sum_{k=m-p}^{m-1} \frac{k}{m} * C_k * a_{m-k} \quad \text{(II.14)}$$

Sachant que C_m sont les coefficients LPCC que nous cherchons.

II.3) Modélisation :

Une fois que le vecteur de paramètres a extrait le système RAL effectue la création d'un modèle du locuteur qui sera plus tard comparé à d'autres modèles. Cette opération consiste à classer les séquences de vecteurs acoustiques extraits à partir du signal de parole en une ou plusieurs classes. Chaque classe est rattachée à un locuteur. Dans ce qui suit nous allons voir les techniques de modélisation les plus connus et utilisées de nos jours.

II.3.1) le modèle de mélange de Gaussiennes (Gaussian Mixture Model GMM):

Le modèle de mélange de Gaussiennes est un modèle statistique où la distribution des données est un mélange de plusieurs lois gaussienne [44].

La densité de probabilité du GMM est comme suite :

$$P(x\backslash\lambda) = \sum_{m=1}^{M} w_n N(x\backslash\mu_m, \sum_m) \quad \text{(II.15)}$$

Où :

$$N(x\backslash\mu_m, \sum_m = \frac{1}{\sqrt{(2\pi)^D |\sum_m|}} \exp(-1/2(x-\mu_m)^T \sum_m^{-1}(x-\mu_m)) \quad \text{(II.16)}$$

Sachant que :

μ_m : Vecteur de moyennes.

$\sum_m$: Matrice de covariance.

W_m : vecteur de pondération $\sum_{m=1}^{M} w_n = 1$

Ainsi on peut définir les paramètres de chaque locuteur par son modèle $\lambda = \{\mu_m, \sum_m, w_m\}_{m=1}^{M}$.

L'apprentissage d'un modèle GMM consiste à estimer l'ensemble de données d'apprentissage λ en fonction des vecteurs $X= \{x_1, x_2, \ldots, x_T\}$issus de l'étape précédente. Il existe plusieurs techniques pour estimer les paramètres d'un GMM [45], l'estimation a maximum de vraisemblance (MLE-Maximum Likelihood Estimation) est une méthode simple qui permet d'aboutir la plupart du temps à un bon estimateur. L'algorithme EM (Expectation Maximization) est très utilisé de nos jours pour déterminer les paramètres du modèle de chaque locuteur réalisant le maximum de vraisemblance [45].

II.3.2) La quantification vectorielle :

La quantification vectorielle (Vector Quantization : VQ) repose sur un partitionnement de l'espace acoustique en sous-espaces. Chaque sous-espace est associé à leur vecteur centroïde (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 centroïdes, appelé dictionnaire de quantification (codebook).

Ainsi l'espace acoustique d'un locuteur donné « X » est réparti en un ensemble de « M » sous-espaces représentés par leur vecteurs centroïdes « C », ces vecteurs centroïdes forment un dictionnaire (de taille M) qui modélise ce locuteur.

La rapidité et les performances de cette technique dépendent fortement de la taille du dictionnaire plus la taille du dictionnaire augmente, meilleures sont les performances sinon, le processus devient plus lent.

II.3.3) Modèle de Markov-caché (Hidden Markov Modeling HMM):

Un modèle de Markov-caché (HMM) est un modèle stochastique. Un modèle stochastique est une suite d'expériences dont le résultat dépend du hasard. En certains temps ce système peut être dans l'un des états d'une collection finie d'états possibles. Ainsi un modèle stochastique est un phénomène temporel où le hasard intervient, c'est-à-dire une variable aléatoire X(t) qui évolue en fonction du temps [48].

Un processus stochastique est markovien si son évolution est entièrement déterminée par une probabilité initiale et une probabilité de transition entre états, ce qui veut dire que son évolution ne dépend pas de son passé mais uniquement de son état présent. L'Etat courant du système contient toute l'information pour prédire son état futur. Les modèles de Markov-caché modélisent des phénomènes dont on suppose qu'ils sont composés au premier niveau d'un processus aléatoire de transition entre deux états inobservables (les états cachés) et à un second niveau d'un état qui génère des valeurs observables [46].

L'apprentissage d'un modèle HMM consiste à estimer les paramètres optimaux de l'HMM de chaque locuteur.Il faut donc calculer pour chaque modèle :

- ✓ Les probabilités initiales
- ✓ Les probabilités de transition
- ✓ Les probabilités d'émission définies par : les vecteurs de moyennes, les matrices de covariance, les poids de pondération.

Différentes approches d'apprentissages ont été proposés. L'approche la plus utilisée s'appuie sur le maximum de vraisemblance (MLE) estimé par l'algorithme de Baum-Welch [47].

II.4) Comparaison et décision :

Ce module entre en jeu dans la phase test où l'on considère de nouveaux échantillons de la parole. Le système doit nous fournir l'information qui appartiennent ces échantillons dans le cas de l'identification ou l'informer que la personne est acceptée ou rejeté dans le cas de la vérification. Cependant, avant la comparaison nous devons reprendre toute la procédure d'extraction des paramètres vu précédemment. Une fois les paramètres extraits, ils seront introduits dans le comparateur.

La compression dépend de la méthode de modélisation utilisée :

II.4 .1) En quantification vectorielle :

Lors de la phase de reconnaissance [48], l'erreur de quantification moyenne est calculée avec chaque dictionnaire de locuteur calculé dans la phase d'apprentissage avec la formule suivante.

$$D(X,C)=\frac{1}{T}\sum_{t=1}^{T}\min_{M\geq m\geq 1} d(x_t, c_m) \quad \text{(II.17)}$$

Où « $d\,(x_t\,, c_m)$» est une mesure de distance au sens d'une certaine métrique liée à la parametrisation.et « T » le nombre de donnés de test du locuteur.

La décision est prise selon le faite que plus l'erreur est faible plus vraisemblablement la séquence de parole a été dite par le locuteur.

II.4 .2) Dans le modèle de mélange gaussien :

Jusqu'ici la base de données contient un groupe de « R » locuteurs représentés par des GMM : λ_1, λ_2... λ_R . La comparaison consiste à trouver la probabilité a postériori, qu'on appellera score, la plus grande à partir d'une séquence observée X= {x_1, x_2,…,x_L}[44]; c'est-à-dire :

$$s = \arg\left(\max_{1\geq r\geq R} p\left(\frac{\lambda_r}{X}\right)\right) \quad \text{(II.18)}$$

D'après la loi de Bayes on aura :

$$s = \arg\max_{1\geq r\geq R} \frac{p(X/\lambda_r}{p(X)} p(\lambda_r) \quad \text{(II.19)}$$

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

$$p(\lambda_r) = \frac{1}{R} \quad \text{(II.20)}$$

Et en constatant que les termes $p(\lambda_r)\ et\ p(X)$ sont constants par rapport à la variation de « r » et que le calcul de maximums est un calcul de dérivée on peut simplifier l'équation avec :

$$s = \arg\max_{1\geq r\geq R} p(X/\lambda_r) \quad \text{(II.25)}$$

Vu que ce calcul est un produit on peut, pour simplifier en utilisant la propriété d'indépendance, inclure le logarithme :

$$s_{log} = \arg\max_{1\geq r\geq R} \sum_{l=1}^{L} \log(x_l/\lambda_r) \quad \text{(II.21)}$$

Où $p(x_l/\lambda_r)$ est donnée par l'équation (II.15)

Sachant que les modèles GMM sont un cas particulier des HMM. Les mêmes étapes de comparaison sont suivies dans cette dernière.

II.5) Evaluation des performances du système :

Dans le cas de l'identification du locuteur, le calcul de la performance se fait de la façon suivante :

$$\text{performance} = \frac{\text{nombre de locuteurs identifiés justement}}{\text{nombre totale des locuteurs testés}} \times 100 \ldots\ldots\ldots\ldots\ldots \quad \text{(II.22)}$$

II.6) Comparaison entre systèmes :

Pour pouvoir comparer l'efficacité entre les systèmes nous avons introduit le calcul de taux d'amélioration/dégradation du système qui se calcule ainsi :

$$taux = \frac{p-pb}{pb} \times 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(II.23)}$$

P : performance du système après la tentative d'amélioration.

Pb : performance du système de base sans aucune tentative d'amélioration.

Si le résultat est positif on dit que le système s'est amélioré, s'il est négatif on dit que le système s'est dégradé.

II.7) Compensation de la variabilité du canal :

Dans cette section nous allons présenter les techniques que nous jugeons pertinents pour l'amélioration du système de base en termes de variabilité du canal. Ce sont des techniques d'intervention sur les différents blocs du système.

II.7.1) Intervention sur le bloc parameterisation (prétraitement) :

II.7.1.1) préaccentuation :

Dans le signal parole les hautes fréquences sont mal représentées par rapport aux basses fréquences, pour compenser le niveau faible des aigus on utilise généralement un filtre passe haut dit de préaccentuation [23]. Ce filtre a souvent pour fonction de transfert :

$$\boldsymbol{H(z) = 1 - az^{-1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(II.24)}$$

Où « a » est une valeur proche de 1. Dans la littérature nous trouvons que la valeur typique est : a=0.95 [49].

II.7.1.2) Suppression des zones des silences :

Dans tout signal de la parole, il existe toujours des régions où il n'y a pas d'activité vocale ; ces régions souvent bruitées, ont très peu d'énergie et ne contienne aucune information sur le locuteur. Les supprimer nous permet de ne conserver que les parties essentielles, ce qui va réduire la taille de donnée tout en conservant la caractéristique de la voix.

Un dicteur d'activité vocale sera sollicité pour effectuer cette fonction. Cette opération est très difficile à mener à cause de la présence de bruit qui change les caractéristiques du signal de la parole. Une des techniques les plus utilisées est l'algorithme de Rabiner et Sambur [50] qui applique les étapes suivantes :

Découper le signal de parole en plusieurs trames non chevauchantes ;

Calculer l'amplitude moyenne est le taux de passage par zéros (TPZ) de chaque trame suivant les équations données par (II.25 et II.26) ;

Si l'amplitude moyenne d'une trame (Am) est supérieure à un seuil maximal (SMX), la trame est considérée comme une trame de signal de parole.

Si l'amplitude moyenne d'une trame (Am) est inférieure à SMX et supérieur à un seuil minimal (SMN), et l'amplitude moyenne de la trame précédente est supérieur à SMX, la trame est considérée comme une trame de signal de parole.

Si l'amplitude moyenne d'une trame (Am) est inférieure à SMX et supérieur à SMN, l'amplitude moyenne de la trame précédente est inférieure à SMX et le taux de passage par zéros de cette trame est supérieur à un seuil SPZ la trame est considérée comme une trame de signal de parole

Dans les autres cas, la trame est considérée comme une zone de silence. Avec SMX et SMN sont les seuils maximal et minimal respectivement pour l'amplitude moyenne et SPZ est le seuil pour le taux de passage par zéros. Ces différents seuils sont donnés par les équations (II.26 et II.30).

Avec :

- ✓ AM : Amplitude moyenne ; TPZ : taux de passage par zéros.
- ✓ $\mathrm{Am} = \frac{1}{\mathrm{T}}\sum_{t=1}^{N}|\mathrm{X(t)}|$ ……...……..…….…… (II.25)
- ✓ $TPZ = \sum_{i=0}^{l}|sgn[X(t+i+1)] - sgn[X(t+1)]|$……...…...….. (II.26)
- ✓ $\overline{TPZ}$: Est la moyenne de TPZ pendant le silence, ¾TPZ est l'écart type de TPZ, IFS est un facteur choisi suivant l'expérience.
- ✓ $SPZ = \min(IFS, \overline{TPZ}, 2\sigma_{TPZ})$…….……... . (II.27)
- ✓ E1 = 0.03× (EMX-EMN) +EMN ……...……..…...…..…. (II.28)

Avec EMX et EMN sont l'énergie maximale et minimale des trames pendant le silence. Ils sont estimés à partir des premiers 75 msec du signal de parole.

- ✓ E2 = 4×EMN ……..……...…….…… (II.29)
- ✓ SMN = min (E1, E2) ……...……...……... (II.30)
- ✓ SMX = 5×SMN ……...…….……..……. (II.31)

La (figure II.1) montre un signal de parole après l'application de l'algorithme de détection des zones de silence.

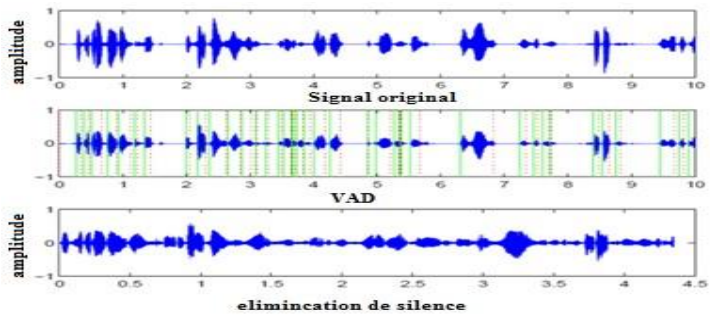


Figure II.7 : Signal de parole après le passage par un éliminateur de silence.

II.7.2) Intervention sur le bloc de modélisation (GMM-UBM):

GMM-UBM est une méthode identique à la méthode GMM. En méthode GMM l'estimation des paramètres des modèles se fait avec l'algorithme EM. Cet algorithme démarre avec des données (Vecteur de moyennes, Matrice de covariance, un poids) initiales aléatoires. Un modèle GMM du monde ou UBM (Universal Background Model) est ainsi appris par l'algorithme EM sur un certain nombre du locuteur qui n'existent pas dans la base de données à tester. Puis les paramètre résultants (Vecteur de moyennes, Matrice de covariance, un poids) sont utilisés comme paramètre initiaux dans l'algorithme EM lors de l'apprentissage des locuteurs concernés par le système [51].

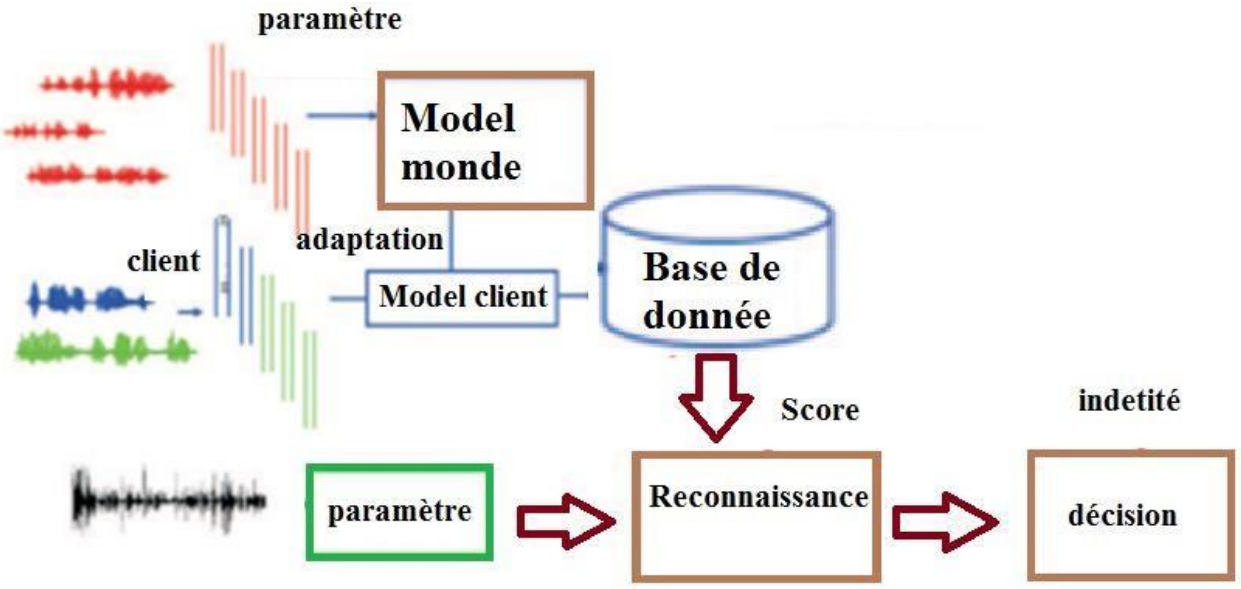


Figure II.8 : Structure générale d'un système RAL à base GMM-UBM.

II.8) Conclusion :

Dans ce chapitre nous avons vu la structure détaillée du système de reconnaissance du locuteur. Puis nous avons présentés les principales techniques utilisées dans les différents blocs de ce système. Après, nous avons fait une brève définition des paramètres à calculer pour pouvoir comparer entre l'efficacité de différents systèmes. Enfin, Nous avons présenté les techniques que nous avons jugés la possibilité d'être efficaces pour compenser la variabilité du canal

Chapitre III

Compensation de la variabilité du canal

III.1) Introduction

Dans ce chapitre nous allons essayer d'améliorer le système de reconnaissance du locuteur de base en nous intéressant à la variabilité du canal.

En premier temps nous allons parler du contexte expérimental dans lequel nous présenterons le système sur lequel nous travaillons, la base de données d'enregistrements utilisée dans nos expériences, le logiciel qui sera mis en œuvre dans ce travaille.

Puis nous allons effectuer un certain nombre d'expériences en présentant leurs résultats afin de choisir les paramètres optimaux avec les quels notre système donne de meilleurs résultats. Tout en mettant en avant l'effet de la variabilité du canal sur les résultats du système.

Enfin, sur la base de ces paramètres nous allons tester des techniques pour essayer de compenser cette variabilité en présentant les résultats des expériences tout en les analysant.

III.2) Le contexte expérimental :

III .2.1) Le système de base :

Nous allons travailler sur un système de base qui sera indépendant du texte et qui remplira la tâche d'identification du locuteur. Les techniques utilisées dans chaque bloque sont :

- Paramétrisation : MFCC
- Modélisation : GMM

Sa composition est présentée dans la (figure III.1) :

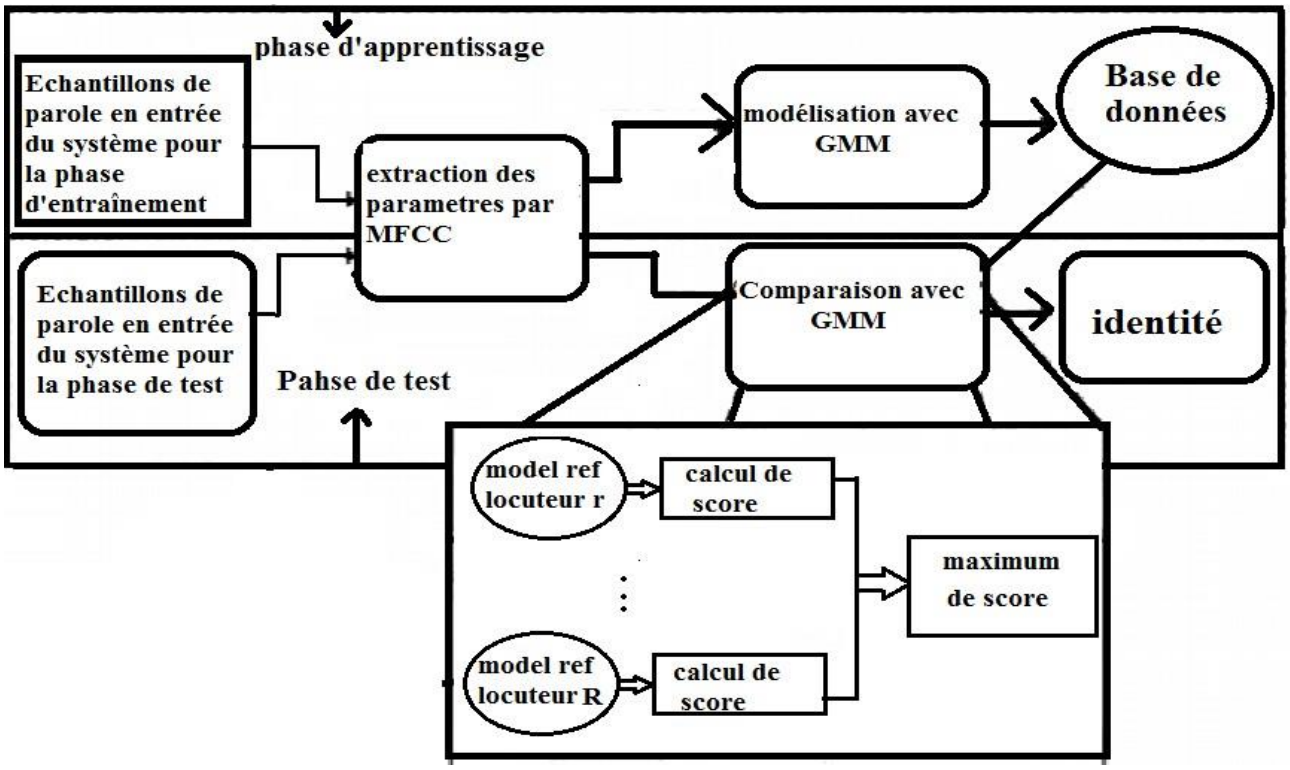


Figure III.1 : Structure du système de base.

III .2.2) La base de données :

La base de données est constitué d'enregistrement de 39 personnes avec 3 différents canaux d'enregistrement (téléphone mobile, téléphone fixe, microphone). Effectués par le centre de développement des technologies avancées « CDTA » situé à Alger.

Nous avons divisé chaque enregistrement en deux parties :

- Une partie de 50 secondes pour la phase d'apprentissage.
- Une partie de 30 secondes pour la phase de test.

III.2.3) Le logiciel utilisé :

MATLAB est une abréviation de MAtrix LABoratory. Écrit à l'origine, en Fortran, par C. Moler, était destiné à faciliter l'accès au logiciel matriciel développé dans les projets LINPACK et EISPACK. La version actuelle, écrite en C par the Math Works Inc., existe en version professionnelle et en version étudiant. Sa disponibilité est assurée sur plusieurs plateformes : Sun, Bull, HP, IBM, compatibles PC (DOS, Unix ou Windows), Macintoch, iMac et plusieurs machines parallèles.

Aujourd'hui, MATLAB est un environnement puissant, complet et facile à utiliser destiner au calcul scientifique. Il apporte aux ingénieurs, chercheurs et à tout scientifique un système interactif intégrant calcul numérique et visualisation. C'est un environnement performant, ouvert et programmable qui permet de remarquables gains de productivité et de créativité [52].

III.3) Fixation des paramètres optimaux.

Le nombre de coefficients MFCC extraits par les enregistrements, le nombre de GMM dans le modèle de chaque locuteur, le nombre d'itérations lors de chaque estimation des paramètres des GMM (vu que le calcul de ces paramètres se base sur une technique itérative) ; tous des paramètres auxquels la performance du système est sensible. Sachant qu'il n'existe pas de théorie qui nous permet de déterminer les paramètres d'une façon que le système soit optimal, chaque base de données peut avoir ses paramètres optimaux qu'on peut déterminer par la seule voie qui est expérimentale.

Dans ce cadre nous avons effectué une série d'expériences faisant varier ces paramètres d'un côté et faisant varier le canal d'un autre coté, en calculant à chaque fois la performance, suivant la logique ci-dessous :

A) Nous avons des enregistrements effectués avec 3 différents canaux de sorte que chaque canal enregistre la voix de 39 mêmes personnes. Nous avons fait varier le canal pour chacune des deux phases de reconnaissance comme suite :

N.B : dans ce qui suit nous allons utiliser « M » pour désigner « le téléphone mobile », « T » pour désigner le « téléphone fixe » et « µ » pour désigner « le microphone ».

	Canal variable						Même canal		
Apprentissage	M	M	µ	µ	T	T	T	µ	M
Test	T	µ	T	M	µ	M	T	µ	M

Tableau III.1 : Tableau explicatif des différentes variations du canal.

Ainsi nous avons 9 expériences à faire.

B) Pour chaque expérience nous allons faire la variation des paramètres :

- Nombre de coefficients MFCC :de 12 vers 20 avec un pas de 2.
- Nombre de GMM : des puissances de 2 allant de 2 vers 64.
- Nombre d'itération pour calculer le paramètre de chaque GMM :de 2 vers 20 avec un pas de 2.

Pour chaque situation on fait la modélisation des 39 enregistrements, puis nous injectons les enregistrements de test dans le système et nous calculons la performance.

Comme le calcul des paramètres GMM est une estimation avec des condition initiales aléatoires (fonctionnement de l'algorithme EM), ceci nous conduit vers des résultats variables (on peut avoir des différents résultats pour le même calcule de performance). Dans ce contexte nous avons refait la simulation un certain nombre de fois puis nous prenons la moyenne. A chaque fois la variation diminue, avec un nombre de 30 expériences nous avons réussi à stabiliser les résultats.

Ceci dit, la réalisation de ces expériences prendrait beaucoup de temps pour un simple PC (Personal Computeur). Nous avons sollicité les services de l'université de Médéa et de Babzouar pour avoir accès à leurs centres de calcules intensifs. Nous nous sommes retrouvés avec des refus pour cause de panne dans les deux cas. Comme issus nous avons mobilisé un laboratoire dans notre faculté et quelques PC personnels pour une somme de 21 PC qui ont travaillé toute une semaine sans arrêt pour aboutir aux résultats obtenus.

Après l'analyse des résultats nous avons extraits les paramètres pour lesquels la performance se retrouve au maximum pour chaque expérience vers le tableau III.2 :

	Canal variable						Même canal		
apprentissage	M	M	µ	µ	T	T	T	µ	M
Test	T	µ	T	M	µ	M	T	µ	M
MFCC	20	18	16	14	12	16	12	16	16
GMM	2	2	2	4	2	2	64	2	2
Itération	4	10	4	16	2	10	2	4	20
Performance	6.64	11.15	4.02	3.68	4.55	16.73	76.32	69.23	83.46

Tableau III.2 : Paramètres optimaux pour chaque expérience accompagnée de la performance atteinte.

On peut constater l'énorme dégradation de la performance en comparant celles des expériences avec un même canal et celles des expériences avec un canal variable. Effectivement, la variabilité du canal présente un inconvénient non négligeable.

III.4) Compensation de la variabilité du canal :

Dans cette partie nous avons appliqué sur notre système les techniques cités dans la section compensation de la variabilité du canal dans le chapitre II. Nous nous sommes intéressées aux expériences qui concernent le cas de canal variable en utilisant les paramètres optimaux que nous avons extraits précédemment pour le calcul de performance à chaque expérience.

Nous avons divisé la tentative d'amélioration en deux parties. L'intervention sur le bloc parameterisation et l'intervention sur le bloc modélisation. Nous avons appliqué sur chaque cas de canal variable l'intervention.

III.4.1) Intervention sur le bloc parameterisation :

Dans cette intervention nous avons appliqué les techniques de prétraitement « application du filtre de préaccentuation, élimination du silence »et l'association d'un paramètres prosodique « l'énergie »

III.4.1.1) Le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe :

Les performances sont représentées en fonction de l'intervention dans le Tableau III.3 :

N.B : nous utiliserons « I » pour désigner « intervention », « P » pour performance en (%) et « T » pour le taux d'amélioration ou de dégradation en (%).

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	6.64	0
	Application du filtre de préaccentuation	5.90	-11.14
	Application de l'élimination du silence	7.86	18.37
	Application du filtre de préaccentuation et de l'élimination du silence	8.46	27.41
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	5.77	-13.10
	Application du filtre de préaccentuation	5.21	-21.54
	Application de l'élimination du silence	8.72	31.33
	Application du filtre de préaccentuation et de l'élimination du silence	8.97	35.09

Tableau III.3 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe avec l'intervention sur le bloc parameterisation. On constate que dans ce cas l'amélioration a bien été réalisé même si avec un faible taux.

L'application de la préaccentuation seulement, avec et sans ajout de l'énergie, ou l'ajout de l'énergie seulement présentent une dégradation de la performance. Les autres interventions présentent une amélioration avec des taux variables.la meilleure intervention qui a su améliorer la performance avec le taux le plus élevé est l'application du filtre de préaccentuation avec l'élimination du silence tout en ajoutant l'énergie aux coefficients cepstraux.

III.4.1.2) Le cas d'apprentissage avec téléphone mobile et test avec microphone :

Les performances sont représentées en fonction de l'intervention dans le Tableau III.4:

Dans ce cas toutes les interventions présentent une dégradation de la performance sauf pour deux cas. Effectivement l'application du filtre de préaccentuation a présenté une très faible amélioration. Cependant l'ajout de l'énergie aux coefficients cepstraux au système basique présente une amélioration.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	11.15	0
	Application du filtre de préaccentuation	10.09	-9.50
	Application de l'élimination du silence	8.89	-20.26
	Application du filtre de préaccentuation et de l'élimination du silence	6.84	-38.65
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	14.94	33.99
	Application du filtre de préaccentuation	11.28	1.17
	Application de l'élimination du silence	7.18	-35.61
	Application du filtre de préaccentuation et de l'élimination du silence	7.61	-31.75

Tableau III.4 : performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec microphone avec l'intervention sur le bloc parameterisation.

III.4.1.3) Le cas d'apprentissage avec microphone et test avec téléphone fixe :

Les performances sont représentées en fonction de l'intervention dans le Tableau III.5 :

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	4.02	0
	Application du filtre de préaccentuation	2.91	-27.61
	Application de l'élimination du silence	4.87	21.14
	Application du filtre de préaccentuation et de l'élimination du silence	4.70	16.91
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	4.23	5.22
	Application du filtre de préaccentuation	3.33	-17.16
	Application de l'élimination du silence	5.13	27.61
	Application du filtre de préaccentuation et de l'élimination du silence	5.47	36.07

Tableau III.5 : performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone fixe avec l'intervention sur le bloc paramétrisation.

Dans ce cas l'amélioration se retrouve à un taux très faible pour l'application de l'élimination du silence et l'application de l'élimination du silence avec filtrage de préaccentuation tout deux sans ajout de l'énergie au coefficient cepstraux. L'ajout de de l'énergie seulement présente un taux faible aussi. Cependant, l'application de l'élimination du silence avec filtrage de préaccentuation tous deux, avec ajout de l'énergie au coefficient cepstraux, présente un meilleur taux d'amélioration. Les autres interventions présentent une dégradation de la performance.

III.4.1.4) Le cas d'apprentissage avec microphone et test avec téléphone mobile :

Les performances sont représentées en fonction de l'intervention dans le Tableau III.6 :

Dans ce cas toutes les interventions ont présenté une amélioration avec des taux variables. Le taux le plus faible a été obtenu en ajoutant l'énergie au coefficients cepstraux. On constate justement que l'ajout de l'énergie dans tous les cas dégrade significativement le taux de l'amélioration vu que les

meilleurs taux sont enregistrés dans le cas sans ajout de l'energie.la meilleure performance a été obtenu par l'application de l'élimination du silence sans ajout de l'énergie aux coefficients cepstraux.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	3.68	0
	Application du filtre de préaccentuation	7.43	101.90
	Application de l'élimination du silence	7.36	100
	Application du filtre de préaccentuation et de l'élimination du silence	5.64	53.26
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	3.85	4.62
	Application du filtre de préaccentuation	4.36	18.48
	Application de l'élimination du silence	5.47	48.64
	Application du filtre de préaccentuation et de l'élimination du silence	4.19	13.86

Tableau III.6 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone mobile avec l'intervention sur le bloc parameterisation.

III.4.1.5) Le cas d'apprentissage avec téléphone fixe et test avec microphone :

Les performances sont représentées en fonction de l'intervention dans le tableau III.7 :

Dans ce cas toutes les interventions ont présenté une dégradation de la performance sauf pour deux cas où une légère amélioration a été observée. Effectivement l'ajout de l'énergie et l'application du filtrage de préaccentuation avec énergie ont obtenu un résultat positif. La meilleure amélioration pour ce cas a été obtenu avec la préaccentuation en associant l'énergie aux coefficients cepstraux.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	4.55	0
	Application du filtre de préaccentuation	4.27	-6.15
	Application de l'élimination du silence	4.10	-9.89
	Application du filtre de préaccentuation et de l'élimination du silence	2.91	-36.04
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	4.62	1.54
	Application du filtre de préaccentuation	5.21	14.50
	Application de l'élimination du silence	2.73	-40
	Application du filtre de préaccentuation et de l'élimination du silence	2.99	-34.29

TableauIII.7 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc parameterisation.

III.4.1.6) Le cas d'apprentissage avec téléphone fixe et test avec microphone :

Les performances sont représentées en fonction de l'intervention dans le tableau III.8 :

Dans ce cas aucune intervention n'a pu améliorer la performance, au contraire on enregistre des dégradations avec des taux variables.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base sans intervention	16.73	0
	Application du filtre de préaccentuation	13.76	-17.75
	Application de l'élimination du silence	11.97	-28.45
	Application du filtre de préaccentuation et de l'élimination du silence	10.34	-38.19
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	15.21	-9.08
	Application du filtre de préaccentuation	13.76	-17.75
	Application de l'élimination du silence	10.26	-38.67
	Application du filtre de préaccentuation et de l'élimination du silence	11.03	-34.07

Tableau III.8 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc parameterisation.

III.4.2) Intervention sur le bloc modélisation :

Dans cette section nous allons ajouter la méthode UBM à la technique de modélisation. Pour former l'UBM nous avons procédé ainsi :

- Nous savons que notre base de donnée contient l'enregistrement de 39 voix avec 3 canaux différents.
- Nous avons pris 9 enregistrements de chaque cas d'enregistrement, nous avons extrait 2 minutes de chaque enregistrement et nous avons mélange tous les extraits pour nous retrouver avec un mélange de 54 minutes.
- Nous avons fait l'apprentissage de l'UBM avec ce mélange, ainsi pour être objectifs il nous reste 30 enregistrement pour chaque canal afin de faire l'apprentissage et l'identification du locuteur.

En terme de comparaison de performances et calcul du taux d'amélioration il est plus convenable d'avoir le même nombre de locuteurs dans les expériences du système de base et les expériences du système avec GMM-UBM. Ainsi nous avons supprimé les enregistrements utilisés dans l'apprentissage de l'UBM puis nous avons refait les à partir de ça pour le système de base, les résultats sont sur le tableau III.9 :

Apprentissage	M	M	µ	µ	T	T
Test	T	µ	T	M	µ	M
MFCC	20	18	16	14	12	16
GMM	2	2	2	4	2	2
Itération	4	10	4	16	2	10
Performance	6.67	13.33	3.33	6.67	6.67	16.67

Tableau III.9 : Performance du système avec les paramètres optimaux pour 30 personnes.

Donc comme prévu nous avons appliqué la méthode GMM sur le système de base puis nous avons appliqués les différentes techniques de prétraitement et l'association du paramètre prosodique (l'énergie) sur chaque cas de reconnaissance et identification du locuteur. Dans ce qui suit nous trouverons les résultats des expériences.

III.4.2.1) Le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe :

Les performances sont représentées en fonction de l'intervention dans le tableauIII.9 :

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	3.33	-50.07
	Application du filtre de préaccentuation	6.67	0
	Application de l'élimination du silence	6.67	0
	Application du filtre de préaccentuation et de l'élimination du silence	6.67	0
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	3.33	-50.07
	Application du filtre de préaccentuation	6.67	0
	Application de l'élimination du silence	3.33	-50.07
	Application du filtre de préaccentuation et de l'élimination du silence	10	49.92

Tableau III.10 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec téléphone fixe avec l'intervention sur le bloc modélisation.

Dans ce cas le remplacement de GMM par GMM-UBM dans le système de, avec ou sans énergie, et application de l'élimination du silence ont présenté une dégradation. Le reste des interventions n'ont eu aucun effet sur la performance sauf pour le cas d'Application du filtre de préaccentuation et de l'élimination du silence qui a présenté une amélioration avec un taux considérable.

III.4.2.2) Le cas d'apprentissage avec téléphone mobile et test avec microphone :

Les performances sont représentées en fonction de l'intervention dans le tableauIII.10 :

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	16.67	25.00
	Application du filtre de préaccentuation	16.67	25.00
	Application de l'élimination du silence	13.33	0
	Application du filtre de préaccentuation et de l'élimination du silence	20	50.04
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	23.33	75.01
	Application du filtre de préaccentuation	23.33	75.01
	Application de l'élimination du silence	20	50.07
	Application du filtre de préaccentuation et de l'élimination du silence	10	-24.98

Tableau III.11 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone mobile et test avec microphone avec l'intervention sur le bloc modélisation.

Dans ce cas, l'application de l'élimination du silence sans ajout de l'énergie au coefficients cepstraux n'a eu aucun effet sur la performance. Le reste des interventions ont tous présentés une amélioration avec des taux différents sauf pour l'application du filtre de préaccentuation et de l'élimination du silence avec énergie qui a présenté une certaine degradation.la meilleure performance a été obtenu avec deux interventions différente : ajout de l'énergie au système de base avec GMM et l'application du filtre de préaccentuation avec énergie. Pour choisir dans ce cas le facteur temps de calcul de la performance nous aidera à distinguer entre les deux. Cependant la meilleure intervention est ajout de l'énergie au système de base avec GMM.

III.4.2.3) Le cas d'apprentissage avec microphone et test avec téléphone fixe :

Les performances sont représentées en fonction de l'intervention dans le tableauIII.12 :

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	6.67	100.30
	Application du filtre de préaccentuation	3.33	0
	Application de l'élimination du silence	6.67	100.30
	Application du filtre de préaccentuation et de l'élimination du silence	3.33	0
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	3.33	0
	Application du filtre de préaccentuation	3.33	0
	Application de l'élimination du silence	3.33	0
	Application du filtre de préaccentuation et de l'élimination du silence	6.67	100.30

TableauIII.12 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone fixe avec l'intervention sur le bloc modélisation.

Dans ce cas, le remplacement de GMM par GMM-UBM dans le système, application de l'élimination du silence sans énergie et Application du filtre de préaccentuation et de l'élimination du silence avec énergie ont présentés une amélioration avec un taux identique. Le reste des interventions n'a eu aucun effet sur la performance. Dans le cas de taux identique on choisit toujours la meilleure intervention en considérant le facteur temps. Ainsi on enregistre que la meilleure performance a été obtenu avec le remplacement de GMM par GMM-UBM dans le système.

III.4.2.4) Le cas d'apprentissage avec microphone et test avec téléphone mobile :

Les performances sont représentées en fonction de l'intervention dans le tableau III.13 :

Dans ce cas, l'Application du filtre de préaccentuation et de l'élimination du silence sans ajout de l'énergie n'a eu aucun effet sur la performance. Une nette dégradation a été observé avec les intervention qui consiste à ajouter l'énergie aux coefficients cepstraux. Cependant, les reste des interventions présente une amélioration avec des taux différent. On peut enregistrer que la meilleure

performance a été obtenue avec l'application du filtre de préaccentuation sans ajout de l'énergie aux coefficient cepstraux.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	10	49.93
	Application du filtre de préaccentuation	16.67	149.93
	Application de l'élimination du silence	13.33	99.85
	Application du filtre de préaccentuation et de l'élimination du silence	6.67	0
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	3.33	-50.08
	Application du filtre de préaccentuation	3.33	-50.08
	Application de l'élimination du silence	3.33	-50.08
	Application du filtre de préaccentuation et de l'élimination du silence	3.33	-50.08

Tableau III.13 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec microphone et test avec téléphone mobile avec l'intervention sur le bloc modélisation.

III.4.2.5) Le cas d'apprentissage avec téléphone fixe et test avec microphone :

Les performances sont représentées en fonction de l'intervention dans le tableauIII.14 :

Dans ce cas, l'ajout de l'énergie aux coefficients cepstraux, l'application de l'élimination du silence sans énergie et le remplacement de GMM par GMM-UBM dans le système présente une amélioration avec un taux identique. L'application du filtre de préaccentuation sans énergie et application de l'élimination du silence avec énergie présentent une dégradation. Le reste des interventions n'ont eu aucun effet sur la performance. Cependant, pour choisir l'intervention la plus performante nous considérons toujours le temps de calcul de performance. Cependant, on enregistre que l'ajout de l'énergie aux coefficients cepstraux est la meilleure amélioration qu'on peut avoir dans ce cas.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	10.	49.93
	Application du filtre de préaccentuation	6.67	0
	Application de l'élimination du silence	10	49.93
	Application du filtre de préaccentuation et de l'élimination du silence	3.33	-50.08
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	10	49.93
	Application du filtre de préaccentuation	3.33	-50.08
	Application de l'élimination du silence	6.67	0
	Application du filtre de préaccentuation et de l'élimination du silence	3.33	-50.08

Tableau III.14 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec microphone avec l'intervention sur le bloc modélisation.

III.4.2.6) Le cas d'apprentissage avec téléphone fixe et test avec téléphone mobile :

Les performances sont représentées en fonction de l'intervention dans le tableau III.15 :

Dans ce cas, le e remplacement de GMM par GMM-UBM et l'application du filtre de préaccentuation et de l'élimination du silence avec énergie n'a eu aucun effet sur la performance. Cependant le reste des interventions ont présentés une dégradation de performance sauf pour le cas de l'ajout de l'énergie aux coefficients cepstraux.

	I	P	T
Sans ajout de l'énergie aux coefficients cepstraux	Système de base	16.67	0
	Application du filtre de préaccentuation	13.33	-20.03
	Application de l'élimination du silence	6.67	-59.99
	Application du filtre de préaccentuation et de l'élimination du silence	13.33	-20.03
Avec ajout de l'énergie aux coefficients cepstraux	Système de base	20	19.98
	Application du filtre de préaccentuation	13.33	-20.03
	Application de l'élimination du silence	6.67	-59.99
	Application du filtre de préaccentuation et de l'élimination du silence	16.67	0

TableauIII.15 : Performance et taux d'amélioration/dégradation du système pour le cas d'apprentissage avec téléphone fixe et test avec téléphone mobile avec l'intervention sur le bloc modélisation.

III.5) Comparaison du taux d'amélioration entre le cas de modélisation GMM et le cas modélisation GMM-UBM :

Ici nous allons prendre le meilleur taux d'amélioration, avec GMM et GMM-UBM, pour chaque cas d'apprentissage et d'identification du locuteur et les comparer, tableau III.16 résume ces taux :

Apprentissage	M	M	µ	µ	T	T
Test	T	M	T	M	µ	M
Taux avec GMM	35.09	33.99	21.14	101.90	14.50	0
Taux avec GMM-UBM	49.92	75.01	100.30	149.93	49.93	19.98

Tableau III.16 : Comparaison entre le taux d'amélioration obtenu avec GMM et celui obtenu avec GMM-UBM.

On observe que le taux d'amélioration avec la méthode GMM-UBM est plus important que celui de la GMM.

III.6) Conclusion :

- La variabilité du canal dégrade considérablement les performances du système de reconnaissance du locuteur.
- Lors de l'intervention sur le bloc de paramétrisation une amélioration du système a été observée. Cette amélioration dépend du couple de canaux utilisés lors de l'apprentissage et de test du locuteur. C'est à dire que chaque couple détermine la technique que nous devons utiliser pour améliorer les performances.
- L'utilisation de la technique de modélisation GMM-UBM nous a encore, permît d'obtenir de meilleurs résultats qu'avec GMM simple.

Conclusion Générale

Dans ce mémoire nous nous sommes intéressés a l'étude du système de reconnaissance du locuteur. Nous avons commencé par voir de façon globale et générale les différents concepts qui constituent le traitement de la parole. Ensuite Nous avons vu détaillé les différentes étapes qui se trouvent au cœur de ce système. Enfin nous avons essayé pratiquement les différentes techniques que nous avons proposés pour compenser la variabilité du canal en dressant un protocole expérimental.

L'analyse de tous les résultats obtenus par ces expériences et tirés les conclusions suivantes :

- La variabilité du canal dégrade considérablement les performances du système de reconnaissance du locuteur.
- Lors de l'intervention sur le bloc de paramétrisation une amélioration du système a été observée. Cette amélioration dépend du couple de canaux utilisés lors de l'apprentissage et de test du locuteur. C'est à dire que chaque couple détermine la technique que nous devons utiliser pour améliorer les performances.
- L'utilisation de la technique de modélisation GMM-UBM nous a, encore, permît d'obtenir de meilleurs résultats qu'avec GMM simple.

Comme perspective nous proposons d'essayer de nouvelles techniques qui donneront des résultats qui ne dépendent pas du couple de canaux utilisés lors de l'apprentissage et test du locuteur.

Résumé :

La reconnaissance du locuteur permet d'identifier ou de vérifier l'identité d'un individu, qu'il proclame, à travers sa voix. Cette tache constitue l'une des nombreuses applications que regroupe la science du traitement de la parole. La variabilité du canal (support avec lequel les enregistrements sont effectués) représente un problème qui est au cœur des recherches les plus récentes. Ce travail regroupe dans sa partie théorique, très riche en références, une suite logique d'informations qui ont pour but, en premier lieu, d'expliquer ce qu'est le traitement de la parole et la situation qu'occupe la reconnaissance du locuteur dans cette science, puis, de détailler le système de reconnaissance en mettant en avant différentes techniques utilisées dans ses constituants. Enfin, de présenter des solutions que nous jugeons probables pour compenser cette variabilité du canal comme l'ajout de l'énergie aux au paramètres extraits du signal de parole, des techniques de prétraitement (élimination du silence et la préaccentuation) et l'utilisation, en modélisation, de la méthode du modèle de mélange de Gaussiennes avec un modèle du monde ou UBM (Universal Background Model). Dans sa partie pratique, il est constitué d'expériences rigoureuses qui mettent en avant l'effet de cette variabilité. L'analyse des résultats obtenus nous a permet de constater une amélioration des performances par la réduction des taux d'erreurs.

Abstract:

Speaker recognition consists of person identification or verification through his voice. This task is one of many applications of speech processing science. The channel variability (the medium with which the recordings are made) represents one of the most considered problems in recent researches. This work gathers in its theoretical part, very rich in references, a logical suite of information. Which aims, in the first place, to explain what's the speech processing and the situation occupied by the speaker recognition system. Then, to detail the recognition system by highlighting various techniques used in its components. Finally, to present a solution that we consider likely to channel compensation for this variability such as the addition of energy to the extracted parameters from the speech signal, pretreatment techniques (silence elimination and preaccentuation) and the use, in modeling stage, of the Gaussian Mixture Models (GMM) method with a world model or UBM (Universal Background Model). In its practical part, it consists of rigorous experiments that highlight the effect of this variability. Analysis of the obtained results showed that performance improved by reducing error rates

bibliographies

[1] A. Jain, R. Bolle, S. Pankanti, « Biometrics: Personal Identification in Networked Society », Kluwer, New York, 1998.

[2] S. Liu, M. Silverman, « A Practical Guide to Biometric Security Technology », IEEE Computer Society, IT Pro-Security, 2001.u

[3] W. B. Kheder, « Reconnaissance du locuteur en milieux difficiles. Informatique et langage ». Université d'Avignon, 2017.

[4]http://www.donboscotournai.be/ExpoDB/ExposPrecedentes/Expo/Ondes/fichiers%20son/Traite-parole.pdf. Consulté le 14 aout 2019.

[5] http://www.claudegabriel.be/Cine%20acoustique%209.pdf. Consulté le 14 aout 2019.

[6] J.J. Ohala, « The origin of the sound patterns in vocal tract constraints Speech », SpringerVerlag, New-York, 1983.

[7] X. Huang, A. Acero, &H.-W. Hon, « Spoken language processing: a guide ta theory, algorithm, and system development », Prentice Hall PTR, New Jersey, 2001.

[8] J. P.J. Campbell, « Speaker recognition: A tutorial », Proceedings of the IEEE, 1997.

[9] J. Trémolière, « La synthèse de la parole », Electronique Applications N°62, octobre 1988.

[10] www.r.battault.free.fr/probatoire. Consulté le 17 octobre 2019.

[11] R. C. Rose, « course ECSE 570, Automatic Speech Recognition », McGill University, Winter 2005.

[12] Hachette Multimedia, encyclopédie ,2004.

[13] Jean Hennebert, « Traitement de la parole », Université de Fribourg, Suisse, 2005.

[14] Fletcher « auditory patterns », Reviews of Modern Physics, 1940.

[15] Calliope, nom collectif représentant les 36 auteurs de cet ouvrage, « La parole et son traitement automatique », Editions Masson, 1989.

[16] L. R. Rabiner, & B. H. Juang, « Fundamentals of speech recognition », PTR Prentice Hall, New Jersey, 1993.

[17] J. L. Flanagan, « Speech analysis; synthesis and perception », Springer-Verlag, New York,1972.

[18] Stevens, Volkmann, « the relation of pitch to frequency », Journal of psychology,1940.

[19] Carter Paul, « Structured Variation in British English Liquids: the role of resonance », University of York,2002.

[20] Othman Lachhab. « Reconnaissance Statistique de la Parole Continue pour Voix Laryngée et Alaryngée », Université Mohammed V de Rabat (Maroc), 2017.

[21] B. F. Gring, « Language of the world », Summer Instep of Linguistics, Novembre 2000.

[22] M. A. Zissman et K. M. Berkling, « Automatic Language Identificatio », Speech Communication, 2001.

[23] S. Furui, « Digital speech processing, synthesis, and recognition », Marcel Dekker, New York, 2001.

[24] H. Hollien, « Phonetician as expert witness. Ethics and responsibilities ». Annals of the New York Academy of Sciences,1990.

[25] R. W. Rabiner, L. R., & Schafer, « Digital processing of speech signals », Englewood Cliffs, Prentice-Hall, New Jersey, 1978.

[26] B. Corona, J. Torry, M. Hind, R. Vincent, «Effects of respiration on heart sounds using timefrequency analysis», IEEE Trans.,2001.

[27] O'Shaughnessy, « Speech Communications: Human & Machine», IEEE Press, 1999;

[28] B. S. Atal, « Automatic recognition of speakers from their vices », Proceedings of the IEEE, 1976.

[29] G. R. Doddington, « Speaker recognition-identifying people by their voices ». Proceedings of the IEEE, 1985.

[30] O'Shaughnessy, « Speaker recognition », IEEE ASSP Magazine, 1986.

[31] S. Furui, « An overview of speaker recognition technology», Paper presented at the Proceedings of Workshop on Automatic Speaker Recognition, Identification and Verification, Switzerland, 1994.

[32] G. chalet, F. Bimbot, « Assessment of speaker verification systems», EAGLES Handbook, 1995.

[33] J. M. Naik, « Speaker verification: A tutorial », IEEE Communications Magazine, 1990.

[34] A. E. Rosenberg, « Automatic speaker verification: a review», Proceedings of the IEEE, 1976.

[35] F. Bimbot, A. Paoloni, G. Chalet, « Assessment Methodology for Speaker Identification and Verification Systems. »Technical report - Task 2500 - Report 19, SAM-A ESPRIT Project 6819. 1993.

[36] F. Bimbot, & L. Mathan, « Second-order statistical measures for text-independent speaker identification », Paper presented at the Proceedings of Workshop on Automatic Speaker Recognition, Identification and Verification, Switzerland, 1994.

[37]http://repository.usthb.dz/bitstream/handle/123456789/583/resum%E9.pdf;jsessionid=C4216EDF206F519B5A16970AC5B19EDE?sequence=1 , consulté le 5 septembre 2019

[38] A. Amrouche « Reconnaissance automatique de la parole par les modèles connexionnistes », USTHB. 1989.

[39] http://www.speech.kth.se/wavesurfer consulté le 5 septembre 2019

[40] E. Wong and S. Sridharan, « Comparison of linear prediction cepstrum coefficients and Melfrequency cepstrum coefficients for language identification», presented at Intelligent Multimedia, 2001.

[41] C. Gargour, V. Ramachandran, « Traitement numérique des signaux », Montréal École de technologie supérieure, 2001.

[42] R. Boite et M. Kunt, « Traitement de la parole », Presses Polytechniques Romandes, Lausanne, 1987.

[43] L. Rabiner , B.Juang, « Fundamentals of speech recognition », Prentice Hall, New Jersey, 1993.

[44] D. Reynolds, R. Rose, « Robust text-independent speaker identification using Gaussian mixture speaker models », IEEE Transactions on Speech and Audio Processing, 1995.

[45] A. P. Dempster, N. M. Laird, and D. B. Rubin, « Maximum likelihood from incomplete data via the EM algorithm », Journal of the Royal Statistical Society, 1977.

[46] L.R. Rabiner, « A tutorial on hidden Markov models and selected applications in speech recognition », Proceedings of the IEEE, 1989.

[47] P. Brown, P. de Souza, R. Mercer, L., «Maximum mutual information estimation of hidden markov model parameters for speech recognition», IEEE International Conference on Acoustics, Speech, and Signal Processing, 1986.

[48] R. Hagen, P. Hedelin, « Robust vectorquantization in spectral coding », Proc. IEEE Int.Conf. Acoust., Speech, Signal Processing, 1993.

[49] B. R. Wildermoth, K. Paliwal, « GMM based speaker recognition on readily available databases»,2003.

[50] L. R. Rabiner & M. R. Sambur, « An algorithm for determining the endpoints of isolated utterances », The Bell System Technical Journal, 1975.

[51] W. Campbell, D. Sturim, et D. Reynolds, « Support Vector Machines Using GMM Supervectors for Speaker Verification », IEEE Signal Processing Letters ,2006.

[52] https://www.iro.umontreal.ca/~mignotte/IFT2425/Matlab.pdf, consulté le 25/09/2019.