الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج
-البويرة-

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Akli Mohand Oulhadj - Bouira -
Tasdawit Akli Muḥend Ulḥağ - Tubirett -

جامعة البويرة

كلية العلوم والعلوم التطبيقية
Faculté des Sciences et des Sciences Appliquées

المرجع:...../م م/ 2021
Référence : ...../MM/2021

Mémoire de Master

Présenté au

Département : Génie Électrique
Domaine : Sciences et Technologies
Filière : Télécommunications
Spécialité : Systèmes des Télécommunications

Réalisé par :

OULMI Manel

Thème

Reconnaissance et identification du genre par empreinte acoustique

Soutenu le: **...../..../2022**

Devant la commission composée de :

Mr :	BENGHENIA Hadj AEK	MAA	Univ. Bouira	Président
	REZKI Mohamed	M.C.A	Univ. Bouira	Rapporteur
	KASMI Reda	M.C.A	Univ. Bouira	Examinateur

Année Universitaire : 2021-2022

Dédicace

De part ces quelque lignes écourtées,

Je transmets un message bien édité

A ma mère qui a tant sacrifié

Pour qu'un jour elle me voie émancipée ;

A mon père toujours occupé

Mais de par ses conseils il ne m'a jamais privée

Au reste de la famille bien adorée ;

Mes frères qui ont su m'épauler ;

A la mémoire de mon grand-père duquel j'ai appris que le savoir la science doivent

être vénérés !

Je termine par un être cher

Mehdi à qui je voue respect et fidélité

Car de bonheur d'abnégation et de bienveillance il m'a toujours comblée

Manel

Remerciements

S'il est coutume en de telle circontance d'adresser un remerciment au-delà d'une simple tradition, je tiens à remercier particulièrement mon promoteur Mr Rezki pour ses énormes qualités d'homme de sciences, cela suscite respect et admiration.Merci pour votre simplicité, votre disponibilité ainsi que votre souci constant à l'aboutissement de ce projet.

J'adresse mes sincères remerciments à tous les professeurs, intervenants et toutes personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté à me recontrer et répondre à mes questions durant mes recherches.

Je remercie Mr Kasmi et que toute ma gratitude lui soit réservée pour avoir accepté d'évaluer mon travail.

Mes sincères remerciments à Mr Benghenia qui m'a fait l'honneur d'accepter la présidence du jury.

Je remercie mes très chers parents, qui ont toujours été là pour moi, « vous avez tout sacrifié pour vos enfants n'apargnant ni santé ni effort. Vous m'aviez donné un magnifique modèle de labeur et de persévérance. Je suis redavable d'une éducation dont je suis fier ».

Je remercie mes frères Nabil et Mehdi pour leur encouragement.

Une pensée distinctive s'adresse à Mehdi qui n'a ménagé aucun effort pour m'aider et m'apporter du réconfort.

Je remercie très spécialement Redha, Noureddine Anis qui ont toujours été là pour moi.

Je tiens à remercier Bahdja, Assala et Anfel, pour leur amitié, leur soutien inconditionnel et leur encouragement.

Enfin je remercie tous mes Ami(e)s que j'aime tant Lahna, Melissa, Rosa, Amira, Hibatellah, Tarek, Walid, Mohammed, Yanis, Abdelmalek, pour leur sincére amitié et confiance et à qui je dois ma reconnaissance et mon attachement.

A tous ces intervenant, je présente mes remerciments, mon respect et ma gratitute.

Résumé

La reconnaissance vocale a de nombreuses applications, notamment l'interaction homme-machine, le tri des appels téléphoniques en fonction du genre, la catégorisation des vidéos avec étiquetage, etc. Actuellement, l'apprentissage automatique est une tendance populaire qui a été largement utilisée dans divers domaines et applications, en exploitant le développement récent des technologies numériques et l'avantage des capacités de stockage des médias électroniques.

Récemment, la recherche s'est concentrée sur la combinaison de techniques d'apprentissage d'ensemble afin de construire des classificateurs plus précis.

Dans cette étude, nous nous concentrons sur la reconnaissance du genre par la voix en utilisant des algorithmes (Covariance, Energie, Mfcc) pris indépendamment, puis un nouvel algorithme d'ensemble sera fait pour démontrer son efficacité en termes de précision.

Mots clés : algorithme, apprentissage d’ensemble, covariance, énergie, mfcc, reconnaissance du genre.

Table des matières

Chapitre 01 : La Reconnaissance Automatique du Locuteur

Chapitre 02 : L'analyse du signal vocal

Chapitre 03 : Application à la reconnaissance du genre du locuteur

Liste des figures

Liste des tableaux

Liste des Acronymes et Symboles

➢ *Acronymes*

ADN	Acide Désoxyribo Nucléique
COV	Covariance
DCT	Discrete Cosine Transform (transformée de Fourier discrète)
DSP	Digital Signal Processor
DWT	Discrete Wavelet Transform
ENG	Energie
FFT	Fast Fourier Transform (transformée de Fourier rapide)
FPGA	Field Programmable Gate Arrays
IAL	Identification Automatique du Locuteur
LPCC	Linear Predictive Cepstral Coefficients (coefficient cepstraux prédictifs linéaires)
MFCC	Mel-Frequency Cepstral Coefficients (coefficient cepstraux de fréquence Mel)
PIN	Personal Identification Number (numéro d'identification personnel)
RAL	Reconnaissance Automatique du Locuteur
RAP	Reconnaissance Automatique de la Parole
TF	Transformée de Fourier
TFCT	Transformée de Fourier à Court Terme
TFD	Transformée de Fourier Discrète
TO	Transformée en Ondelette
TPZ	Taux de Passage par Zéro
VAL	Vérification Automatique du Locuteur
ZC	Zero Crossing (passage à zéro)

➢ *Symboles*

f_0 *Fréquence du fondamental*

f_e *Fréquence d'échantillonnage*

T_e *Période d'échantillonnage*

F_{max} *Fréquence maximale*

ϕ_{xx} *Autocorrélation*

INTRODUCTION

Introduction générale

La parole constitue l'un des moyens les plus populaires et les plus importants pour les humains decommuniquer, d'exprimer leurs émotions, leurs états cognitifs et leurs intentions les uns aux autres.

La parole est produite par les humains à l'aide d'un mécanisme biologique naturel dans lequel les poumons évacuent l'air et le convertissent en parole en passant par les cordes vocales et les organes chaque élément jouant un rôle très important pour un échange parfait de l'information. Au niveau de la production du son, la juxtaposition de tous ces éléments entraîne la possibilité de créer des sons très variés, que ce soit en termes de composantes fréquentielles, d'intensité ou de hauteur. Au niveau de la perception, l'objectif semble plus simple : décoder les informations contenues dans l'onde sonore. Le dispositif associé à cette tâche est pourtant bien plus alambiqué, notamment en raison de la nécessité de transformer la vibration de l'air en information compréhensible par le cerveau.[01]

Si l'être humain a le privilège de comprendre un message vocal d'une autre personne quelconque quel que soit l'environnement, la syntaxe, le vocabulaire utilisé, la machine sera-elle un jour capable de faire autant ? Une solution robuste et efficace sera-elle trouvée et proposée pour satisfaire ces contraintes ? Le moins qu'on puisse dire c'est que malgré son importance actuellement, seules des solutions partielles sont proposées dans le langage machine, qui peuvent faire des taches déjà prédéfinies et déjà préenregistrées, mais est incapable de faire ce dont l'homme est capable.[02]

La sécurité est depuis tout temps la préoccupation majeure des individus. Que ce soit pour les biens, les personnes ou les données, les techniques de sécurité ont connu une véritable avancée. Passant de ce que l'on possède (clef, badge, etc.) ou de ce que l'on sait (mot de passe, etc.), la tendance actuelle de sécurité se base sur ce qu'on est. Cette approche, nommée Biométrie, a apporté simplicité et confort aux utilisateurs.

Dans ce travail, nous nous intéressons à la biométrie vocale et plus exactement à l'Identification Automatique du Locuteur (IAL) en mode indépendant du texte. Il s'agit de reconnaître le genre d'une personne (homme ou femme) à partir de sa voix.
Ce mémoire sera donc essentiellement dédié à la conception d'un système d'Identification Automatique du Locuteur et à la Paramétrisation de ce dernier.
Ainsi, ce travail s'articule autour de trois chapitres.

Dans le premier chapitre, nous introduisons le domaine de la sécurité en exposant la biométrie avec toutes ses techniques, ses avantages et inconvénients tout en mettant l'accent sur la biométrie vocale. Une description anatomique détaillée du Locuteur est aussi passée en revue.

Le deuxième chapitre traite les outils de base pour le traitement et la paramétrisation du signal vocal, et à ce propos on a cité quelques systèmes et méthodes d'analyse.

Le troisième chapitre nous décrirons notre système ainsi que les algorithmes utilisés pour l'élaborer, et enfin nous conclurons par les résultats expérimentaux obtenus.

Partie bibliographique

CHAPITRE 01 :

La Reconnaissance Automatique du Locuteur

1.1. Introduction

La parole est une manifestation du langage humain qui est, lui-même, un mécanisme de communication d'information entre les êtres humains, sa particularité tient sans doute à la complexité des fonctions que le cerveau met en œuvre pour la produire ou la comprendre, et ceci d'une manière pratiquement instantanée. Ces informations servent en particulier à déterminer l'identité du Locuteur

Les systèmes de Reconnaissance Automatique du Locuteur (RAL) s'intéressent précisément à ces caractéristiques particulières du signal de parole. C'est un terme générique qui regroupe les problèmes relatifs à l'identification ou à la vérification du Locuteur sur la base de l'information contenue dans le signal acoustique : il est question de reconnaître une personne à partir de sa voix. [03]

Ce chapitre est une introduction au domaine de RAL ; Nous allons décrire le Locuteur sous ses deux facettes anatomiques et acoustiques afin de justifier l'utilisation de la voix dans le domaine de reconnaissance, nous présenterons aussi les différentes tâches liées à ce domaine ainsi que leur application.

1.2. La voix

La voix est un instrument paradoxal. Il est à la fois banal et précieux, fragile et puissant. [04] La voix de chaque personne dépend des caractéristiques, à la fois anatomiques et comportementales. Avant de parler de la reconnaissance automatique du Locuteur, il est important de le décrire anatomiquement pour comprendre le processus d'émission de la voix et connaître les paramètres qui différencient un Locuteur d'un autre.

1.3. Description de l'appareil phonatoire

Ce que l'on appelle l'appareil vocal humain est constitué des poumons, de la trachée, du larynx, du pharynx, et enfin des cavités nasales et orales. Connectés les uns aux autres, ils forment une sorte de tube représentant l'appareil de production de la parole. [05]

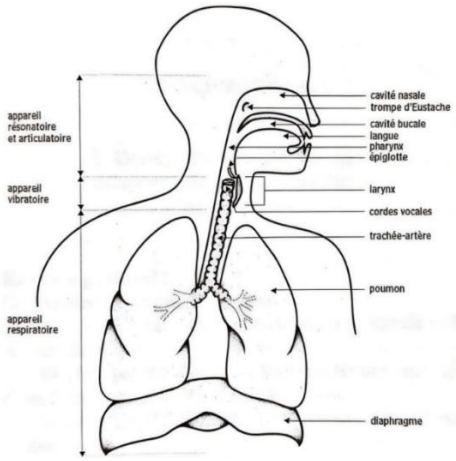

Figure 1.1: Schéma de l'appareil phonatoire humain. [06]

1.3.1. Les poumons et la trachée

Les poumons et la trachée constituent la soufflerie qui fournit la source d'énergie nécessaire à l'ensemble de l'appareil phonatoire. La parole est essentiellement produite lors de l'expiration de l'air, qui est à l'origine de la formation d'une surpression en dessous du larynx. [07]

1.3.2. Le larynx

Son rôle est la production des sons. C'est un ensemble de muscles et de cartilages mobiles, les cartilages les plus importants et les plus connus sont les cordes vocales. Ces plis vocaux sont un des organes musculaires de la phonation, constitué de replis fermes et souples à la fois produisant ainsi des variations de pressions dans l'air qui sont perçues comme du son par l'oreille humaine.[07]

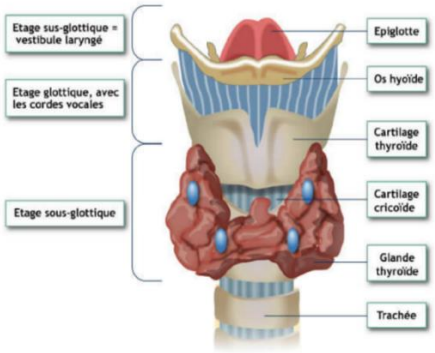

Figure 1.2: Vue postérieure du larynx. [09]

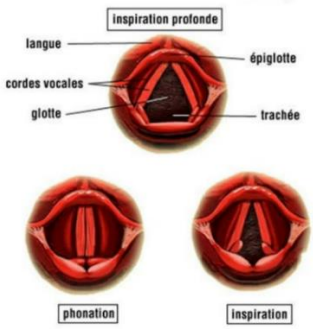

Figure 1.3: Vue supérieure du larynx. [10]

1.3.3. Le conduit vocal

C'est tout ce qui se situe entre les cordes vocales et les lèvres. Il est composé de plusieurs cavités reliées entre elles [08], on retrouve :

- La cavité pharyngale.
- La cavité nasale.
- La cavité buccale.
- La cavité labiale.

1.4. Production de la parole

Le processus de production de la parole est un mécanisme très complexe qui repose sur l'interaction entre le système nerveux et le système physiologique. Il existe de nombreux organes et muscles impliqués dans la production des sons du langage naturel. La fonction de l'appareil vocal humain est basée sur l'interaction entre trois organes majeurs : les poumons, le larynx et les cavités supra-glottiques.

Les deux premières catégories fournissent ce qui est nécessaire pour produire n'importe quel son, qu'il s'agisse de musique ou de parole : les sources d'air et les sources de bruit. La troisième classe contient des organes qui peuvent modifier le son produit par le travail combiné des deux premières classes.

Lors de la production de la parole les poumons fournissent une source d'air, pendant la phase d'inspiration, l'action combinée du diaphragme contracté et abaissé et des muscles intercostaux crée un vide dans les poumons, qui est rempli par la perméation de l'air. Lors de l'expiration, le diaphragme se détend, ce qui expulse l'air des poumons. Quand l'air est évacué des poumons, il traverse le larynx qui par le biais des cordes vocales va générer une vibration, c'est ici, sous l'influence des vibrations, que la pression de l'air se transforme en quelque sorte en une série de pulsations quasi-périodiques ou aléatoires. Une première distinction très importante s'impose donc à l'égard des différents types de sons produits : les sons engendrés par une vibration pseudopériodique des cordes vocales sont dits *voisés*, alors que les sons provenant d'un simple flux d'air à travers la glotte sont dits *non voisés*. [07] Une fois traversée l'espace entouré par ces cordes vocales (glotte), la colonne d'air atteint les cavités supra-glottiques. [11]

1.4.1. La fréquence du fondamental (le pitch)

La vitesse d'ouverture et de fermeture des cordes vocales au cours du processus de phonation, émet une vibration variable appelée fréquence fondamentale f_0 dont la valeur est étroitement liée à la taille de l'appareil vocal de la personne. C'est elle qui véhicule une grande partie de l'information prosodique (variation de ton, de durée, d'intensité) peut faire ressortir bien des caractéristiques du locuteur, comme son genre, ses émotions, etc

Il s'agit d'une fréquence quasi-stationnaire pour un signal de type voisé [12], elle varie de :

- De 110 à 165 Hz pour une voix masculine.
- De 220 à 330Hz pour une voix féminine.
- De 200 à 600Hz pour une voix d'enfant.

1.4.2. Les formants

Un formant est le large maximum spectral résultant d'une résonance acoustique du tractus vocal humain, les formants sont numérotés dans leur ordre d'apparition depuis les fréquences basses jusqu'aux fréquences hautes. Du point de vue perceptif, seuls les trois premiers formants jouent un rôle essentiel pour caractériser le spectre vocal. [13]

1.4.3. Le phonème

Un phonème est la plus petite unité présentée dans la parole qui susceptible par sa présence de changer la signification d'un mot [14]. Selon la théorie des traits distinctifs, chaque phonème se caractérise par un ensemble de traits qui le différencie des autres phonèmes, ces traits reflètent des propriétés de nature acoustico-auditive ou articulatoire. [15]

Le nombre de phonèmes est toujours très limité, en générale inférieur à 50.

1.5. Description Physique du Signal Vocal

En plus du message linguistique servant à la communication entre individus, le signal de parole véhicule des informations caractéristiques de la personne qui l'a émis comme le timbre de sa voix, sa façon de parler, son état émotionnel ou pathologique, etc.

Ces informations caractéristiques du Locuteur peuvent être classées en deux catégories distinctes :

- Les informations de nature statique telles que les paramètres spectraux caractérisant les conduits vocal et nasal, la moyenne et les variations de la fréquence fondamentale.
- Les informations de nature dynamique reflétant les phénomènes de coarticulation, les trajectoires formantiques ainsi que les informations temporelles (vitesse d'élocution, distribution des pauses).

Nous parlerons ici des caractéristiques statiques du signal vocal. Ce dernier peut être défini par quatre (4) paramètres principaux [16] [17] [18]

a. **Intensité :** L'intensité d'un son correspond à l'amplitude de la vibration acoustique ; elle caractérise le volume sonore qui nous permet de distinguer un son fort d'un son faible. L'intensité vocale varie surtout en fonction de la pression sous glottique.
b. **Timbre :** Le timbre permet de différencier deux sons de même hauteur et de même amplitude. Il est constitué d'un ensemble de fréquences appelé spectre. La richesse du spectre permettra de dire qu'un son est riche, brillant, profond, etc. Le timbre est fonction des trois critères suivants : des conditions d'accolement des cordes vocales, de leur épaisseur et enfin des caractéristiques anatomiques des cavités de résonance (pharynx, bouche et cavités nasales).
c. **Hauteur :** La hauteur dépend de la fréquence de la variation de pression acoustique correspondant au son. Elle est fonction de la périodicité du mouvement des lèvres glottiques, c'est-à-dire en pratique, du nombre d'ouvertures glottiques par seconde. La hauteur dépend aussi de la taille du larynx : plus les cordes vocales sont longues, plus la voix est grave.
d. **Fréquence :** Elle représente le nombre de vibrations de l'air en une seconde.

1.6. L'appareil auditif

L'émetteur d'information (l'appareil phonatoire), serait inutile si l'information produite ne pouvait pas être captée et analysée par un récepteur. L'oreille, organe récepteur de l'information sonore. [20]

Le système auditif est en lui-même une merveille du traitement acoustique et cognitif avec un détecteur qui est l'oreille et un centre de traitement qui est le cerveau. [19]

1.6.1. Anatomie de l'oreille humaine

L'oreille est divisée en trois parties distinctes

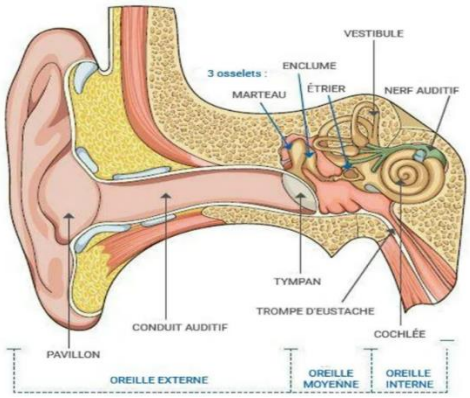

Figure 1.4 : Composition de l'oreille humaine. [21]

1.6.1.1. L'oreille externe

Recueille les vibrations sonores et les oriente vers le tympan [22], elle est constituée du :

- Pavillon dont la fonction est de détecter les ondes sonores et de les diriger vers le canal auditif.
- Canal auditif responsable de la transmission du son à la deuxième partie de l'oreille.
- Tympan, membrane située à l'extrémité du conduit auditif qui vibre sous l'influence des ondes sonores.

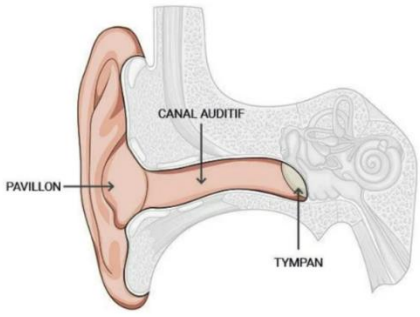

Figure 1.5 : L'oreille externe. [21]

1.6.1.2. L'oreille moyenne

S'étend du tympan aux fenêtres ovales et rondes L'oreille moyenne comporte une série d'osselets (marteau, enclume, étriers) maintenus ensemble par des articulations son rôle est de détecter

un son aérien (vibration aérienne) et de le convertir en vibration solidienne (le bruit passe par la structure de l'oreille). [22]

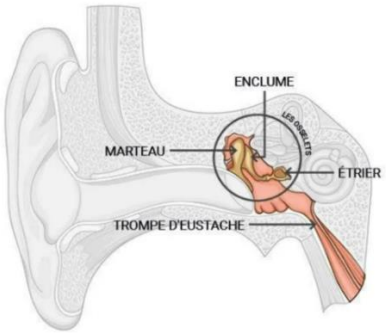

Figure 1.6 : L'oreille moyenne. [21]

1.6.1.3. L'oreille interne

Elle communique avec l'oreille moyenne par la fenêtre ovale. Elle est divisée en plusieurs parties [22]

- Le vestibule et les canaux semi-circulaires, qui interviennent dans l'équilibre et la posture.
- La cochlée, se compose de liquide et ses parois sont recouvertes de cellules ciliées. On en compte environ 15 000, et joue un rôle décisif dans la transmission du son au cerveau. Les cellules ciliées baignent dans la périlymphe qui permet la conversion d'une vibration en message nerveux. Son codage contient la fréquence, l'intensité ainsi que la composition du son et sa source d'émission.

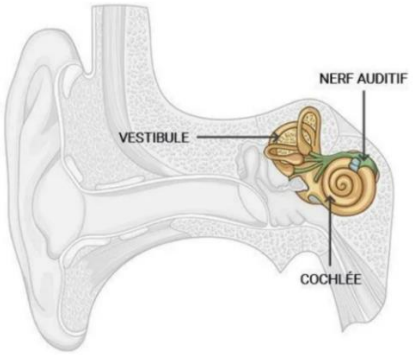

Figure 1.7 : L'oreille interne. [21]

1.6.2. Fonctionnement du système auditif

Le pavillon reçoit le son et le guide à travers le conduit auditif externe jusqu'au tympan.[21] Celui-ci vibre et ces vibrations sont émises à la fenêtre ovale par les osselets qui amplifient les informations sonores. Ces vibrations sont transmises de l'oreille interne à la cochlée.[23]

La cochlée convertit le son en signaux électriques qui sont envoyés au cerveau par le nerf cochléaire. A l'intérieur de la cochlée, les cellules ciliées sont suspendues dans un liquide (endolymphe). Avec les vibrations, le fluide bouge et déplace les cils de la cellule. Le mouvement des cils libère des messagers (neurotransmetteurs) qui transmettent un message nerveux par le nerf cochléaire : on parle donc d'information sonore.[24]

1.7. Modélisation du mécanisme de la production de la parole

Nous pouvons retenir que chaque parole est le résultat d'actions successives des poumons, larynx et du conduit vocal, tous agissant de façon indépendante les uns des autres permettant à l'être humain de produire beaucoup de sons différents. Cette relative indépendance des sources sonores et de leurs transformations est la base de la théorie acoustique de la production de la parole. Cette théorie prend en compte le terme source et un filtre linéaire qui transforme le signal source en modifiant son enveloppe spectrale. Il est ainsi possible de résumer le cheminement du son, depuis son origine jusqu'à sa sortie du corps humain et assimiler chaque organe producteur de parole à un élément plus « mécanique » [25], comme indiqué sur la figure 1.8

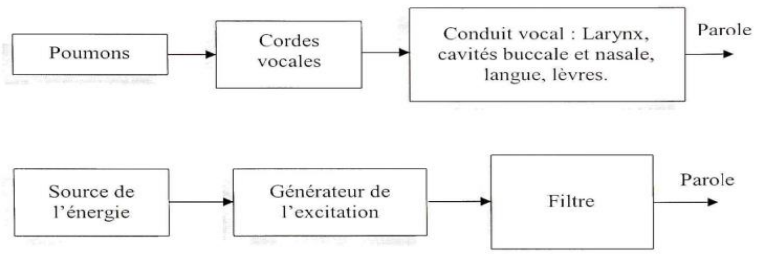

Figure 1.8 : Schématisation du processus de production de la parole.

Par conséquent, une source sonore peut être modélisée par une série d'impulsions périodiques, pour les sons voisés, ou un bruit blanc, pour les sons non voisés, excitant un filtre dit *tous-pôles*, dont les éléments représentent les caractéristiques du conduit vocal.

1.8. La biométrie

La biométrie comme son étymologie l'indique est composée des deux mots : vie et mesure. C'est une technique qui vise à établir l'identité d'un être vivant (personne) en se basant sur ses différentes caractéristiques physiques et comportementales, qui rendent la personne unique et la distinguent des autres individus dans une population très large.

1.8.1. Les différentes techniques d'authentification biométrique

On compte plus d'une dizaine de technologies biométriques classées en trois catégories : les premières reposent sur l'analyse morphologique ; les secondes sur l'analyse comportementale et les troisièmes enfin sur l'analyse de traces biologiques.

1.8.1.1. La biométrie morphologique

La biométrie morphologique est basée sur l'identification de traits physiques particuliers. Elle regroupe notamment, mais pas exclusivement, la reconnaissance des empreintes digitales, de la forme de la main, du visage, de la rétine et de l'iris de l'œil.

1.8.1.2. La biométrie comportementale

La biométrie comportementale est basée sur l'analyse de certains comportements d'une personne, comme le tracé de sa signature, l'empreinte de sa voix, sa démarche, sa façon de taper sur un clavier, etc.

1.8.1.3. La biométrie biologique

La biométrie biologique est basée sur l'analyse des traces biologiques d'une personne, comme l'ADN, le sang, la salive, la thermographie faciale, les odeurs, etc.

1.8.2. Architecture d'un système biométrique

L'architecture d'un système biométrique se compose de six modules (voir Figure 1.9).

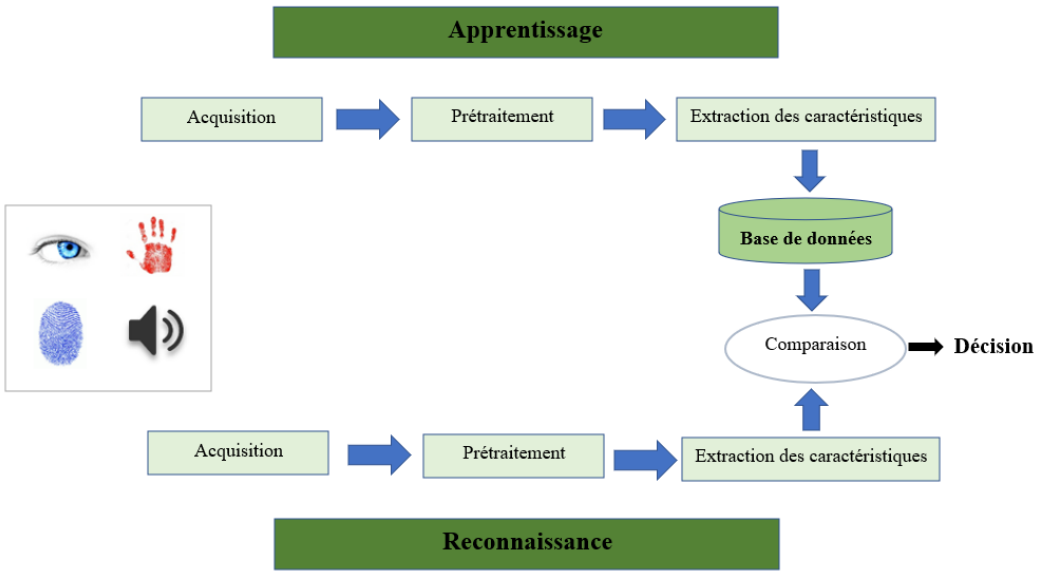

Figure 1.9: Architecture d'un système biométrique.

1.8.3. Biométrie vocale

La biométrie vocale est un domaine scientifique et technologique qui vise à développer des applications permettant de vérifier l'identité d'une personne seulement grâce à sa voix. Si la reconnaissance de la parole sert à déchiffrer « ce qui est dit » dans un enregistrement sonore, la reconnaissance du locuteur (ou biométrie vocale) cherche à savoir « qui l'a dit ».

1.8.3.1. Son Fonctionnement

- Lors de son inscription, l'individu fournit un échantillon de sa voix (échantillon d'inscription).

- Puis, chaque fois qu'il désire s'identifier, l'individu doit fournir un échantillon test (soit un mot de passe précis ou simplement un échantillon de sa voix sans contenu spécifique).
- Le système compare l'échantillon test à un modèle entraîné à partir d'information provenant d'échantillons vocaux de nombreux autres locuteurs. La machine ne compare donc pas directement l'échantillon test avec l'échantillon d'inscription de la personne qui tente de s'authentifier. Le système cherche plutôt à mesurer à quel point la voix du test ressemble plus au modèle de voix lié à l'utilisateur tentant de s'identifier qu'au modèle de référence du système.
- De plus, les développeurs doivent déterminer le niveau de sévérité de leur système de biométrie vocale. C'est-à-dire qu'ils doivent choisir la « note de passage », le score de similarité à partir duquel la voix test sera considérée comme celle de l'individu inscrit. Ce niveau de sévérité dépend souvent du coût d'une erreur : il sera évidemment plus élevé pour autoriser des transactions bancaires que pour avoir accès à votre compte de bibliothèque !
- Enfin, selon le score de similarité obtenu et le niveau de sévérité choisi, le logiciel fournit une réponse sous forme binaire : il accepte ou il rejette l'identification.

L'approche automatique sous tous ses aspects sera détaillée dans la suite de ce chapitre.

1.9. Reconnaissance automatique du Locuteur

1.9.1. Généralités

La Reconnaissance Automatique du Locuteur (RAL) s'inscrit dans le cadre général du traitement automatique de la parole ; son objectif principal est de déterminer l'identité d'une personne à l'aide de sa voix [26]. Elle tire son essence de la variabilité interlocuteur. Cette variabilité est due principalement aux différences morphologiques de l'appareil vocal d'un individu à l'autre. On peut classifier les systèmes de RAL suivant leur dépendance au texte. On distingue les systèmes dépendants du texte des systèmes indépendants du texte. Dans le mode dépendant du texte [27], la reconnaissance est réalisée à l'aide d'un message connu a priori par le système que le Locuteur doit prononcer (mot de passe, code PIN, phrase, ...). Ce message peut être choisi par le Locuteur ou imposé par le système. Le mode indépendant du texte contrairement au premier n'impose aucune contrainte, sur le message à prononcer, au Locuteur

Les systèmes de RAL sont sensibles à certains facteurs qui peuvent altérer leur performance ; ces facteurs peuvent être intrinsèques ou extrinsèques au Locuteur. On peut citer :

- L'état pathologique du Locuteur (maladie, émotion, …)
- Vieillissement.
- Facteurs socioculturels.

- Locuteurs non coopératifs.
- Conditions de prise de son.
- Bruit ambiant, …

1.9.2. Différentes tâches en RAL

Un nombre de tâches centrées sur l'identité des locuteurs ont été étudiées dans la dernière trentaine d'années. Ces tâches répondent à des besoins applicatifs et permettent de tirer parti de l'identité du locuteur de différentes manières dépendant de l'application ciblée. Les deux tâches pionnières des systèmes de Reconnaissance Automatiques du Locuteur (RAL) sont l'Identification Automatique du Locuteur (IAL) et la Vérification Automatique du Locuteur (VAL). [27] [28] [29] Récemment, des besoins spécifiques ont stimulé l'apparition de nouvelles tâches comme l'indexation du Locuteur qui consiste à indiquer à quel moment chaque Locuteur intervenant dans une conversation a pris la parole.

Dans cette section, nous allons décrire les principales tâches de la RAL qui sont l'IAL et la VAL

1.9.2.1.Vérification (authentification) Automatique du Locuteur (VAL)

Consiste à accepter ou refuser l'identité proclamée par un locuteur, en se basant sur un modèle qui lui est associé. Elle traite le scénario où le système prend en entrée un énoncé de test ainsi qu'une identité proclamée (figure 1.9). La tâche consiste alors à prendre une décision binaire qui va confirmer ou infirmer le fait que l'enregistrement de test est effectivement prononcé par le locuteur proclamé.

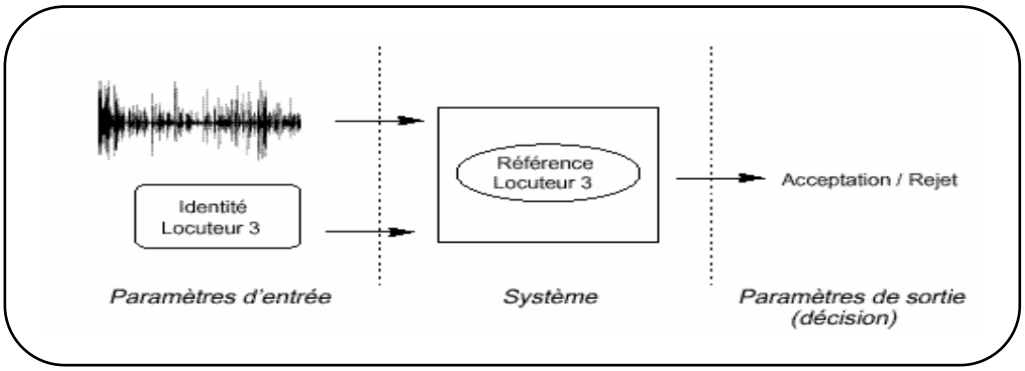

Figure 1.10 : Principe de base de la tâche de Vérification Automatique du Locuteur.[30]

1.9.2.2. Identification Automatique du Locuteur (IAL)

Consiste à déterminer, parmi une population de locuteurs connus, la personne ayant prononcé un message donné. D'un point de vue schématique (voir figure 1.10), une séquence de parole est donnée en entrée du système d'IAL. Pour chaque locuteur connu du système, la séquence de parole est comparée à une référence caractéristique du locuteur : identité du locuteur dont la référence est la

plus proche de la séquence de parole est donnée en sortie du système d'IAL. Deux modes sont proposés en IAL :

- L'identification en ensemble fermé pour lequel on suppose que la séquence de parole est effectivement prononcée par un locuteur connu du système ;
- L'identification en ensemble ouvert pour lequel le locuteur peut ne pas être connu.

En mode « ensemble ouvert », le système d'IAL doit décider de la fiabilité de son jugement en acceptant ou rejetant l'identité qu'il a trouvée. De par son principe - déterminer une identité parmi les identités potentielles - les performances des systèmes d'IAL se dégradent généralement au fur et à mesure que la population de locuteurs augmente.

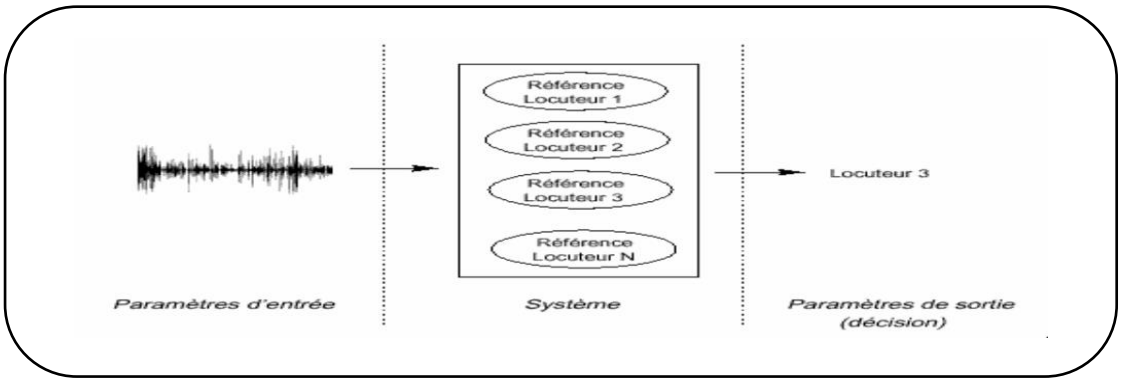

Figure 1.11 : Principe de base de la tâche de d'Identification Automatique du Locuteur. [30]

1.10. Structure d'un système de RAL

Les systèmes de vérification du locuteur sont composés de trois modules comme le montre le schéma (voir figure 1.12)

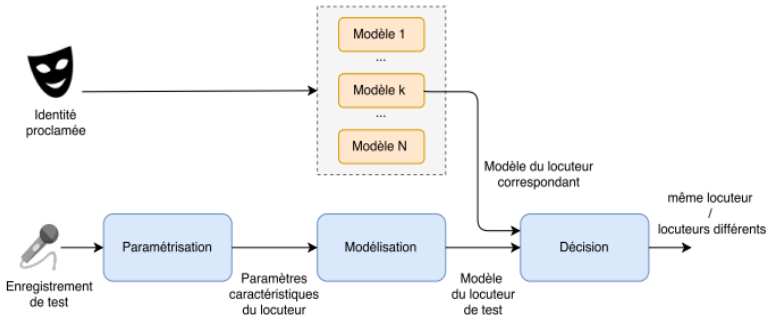

Figure 1.12 : Structure d'un système de vérification du locuteur.[32]

1.10.1. Paramétrisation

Cette étape vise à capturer des paramètres caractéristiques de la parole d'une personne donnée. Suite à de nombreux travaux de recherche [31] [32], il s'est avéré que les paramètres basés sur la représentation spectrale de la parole sont les plus pertinents pour la RAL. Ces paramètres sont corrélés

à la forme du conduit vocal et sont les plus utilisés dans les systèmes de RAL modernes. Cependant, les paramètres prosodiques qui décrivent le style de parole du locuteur sont aussi utilisés en pratique.

1.10.2. Modélisation

Les paramètres acoustiques extraits d'un enregistrement donné sont utilisés pour construire un modèle qui résume l'information acoustique correspondante.

1.10.3. Décision

La phase de décision désigne l'identité du locuteur reconnu. Dans le cas de la vérification, cette décision est binaire et consiste à confirmer ou infirmer la correspondance de la session de test à une identité proclamée. Vu qu'il est impossible d'avoir une similarité de 100% entre le signal du locuteur de test et celui des locuteurs clients, les modèles sont conçus de telle sorte qu'une telle comparaison fournisse un score (une valeur scalaire) indiquant si les deux énoncés correspondent au même locuteur. Si ce score est supérieur (inférieur) à un seuil prédéfini, le système accepte (rejette) le locuteur de test.

1.11. Domaine d'applications des technologies de RAL

Un intérêt croissant est accordé aux technologies basées sur l'identification vocale dans les domaines public et industriel. En effet, la RAL intervient de nos jours dans un grand nombre d'applications

- **Applications sur site**

La personne doit être physiquement présentée en un lieu précis.

- Serrures vocales (pour des locaux, un compte informatique …)
- Interactivité matérielle (retrait d'argent à un guichet automatique…).

- **Applications liées aux télécommunications**

La vérification s'opère à distance

- Accès à des services pour des abonnés, ou des données confidentielles.
- Transaction à distance.

- **Applications commerciales**

- Associer un même mot de passe pour une petite population de locuteur (membre d'une famille, d'une société).
- Protection de matériel contre le vol.

➢ **Applications judiciaires**

- Recherche de suspects, d'éléments de preuve, de preuve.
- Les juges, les avocats, les enquêteurs de police ou de la gendarmerie souhaitent utiliser des procédés de reconnaissance vocale pour mener une enquête ou confirmer un verdict de culpabilité ou d'innocence.

➢ **Applications stratégiques**

- Ecoutes téléphoniques.
- Protection de la démocratie.
- Intrusion dans la vie privée.

1.12. Conclusion

L'être humain reste une merveille technologique méconnue. Nous sommes capables, sans aucuns efforts, instantanément et de façon robuste, d'isoler un flux de parole d'un individu donné, à partir d'un paysage sonore complexe et bruité.

Au cours de ce chapitre nous avons étudié en revue le mécanisme de la production de la parole, ses principes auditifs et les caractéristiques générales du signal vocal. Ensuite nous avons introduit le principe de la Reconnaissance Automatique du Locuteur et les différentes tâches liées à ce domaine et aussi les différentes étapes d'un système de RAL ont été présentées.

Il existe plusieurs modèles pour le traitement du signal vocal, le modèle le plus pratique et le plus utilisé est l'équivalent électrique du mécanisme de la production de la parole.

Nous étudierons dans le prochain chapitre les différents outils de traitements de la parole, nous présenterons aussi quelques méthodes de paramétrisation du signal vocal.

CHAPITRE 02 :

L’analyse du signal vocal

2.1 Introduction

Au cours du premier chapitre, les signaux vocaux ont été décrits comme des mouvements d'air qui transportent à la fois des informations verbales et émotionnelles, produites par un ensemble de muscles et d'organes. À partir de ce point et pour les besoins de notre étude, il sera traité comme une forme d'onde qui contient un ensemble complet d'informations sur le message qu'il véhicule. L'objet de ce chapitre est de comprendre la base du traitement de la parole et de s'appuyer sur les caractéristiques spécifiques du signal de parole pour déterminer la méthode d'extraction des paramètres la plus adaptée.

2.2 Les signaux de parole

Le signal de parole est l'entité fondamentale sur laquelle les chercheurs travaillent lors de la conception des systèmes de reconnaissance de parole. Le signal vocal prend la forme donnée dans la figure 2.1, et est générer par les fluctuations de la pression de l'air produites par l'appareil phonatoire humain. Cette figure montre ces fluctuations en fonction du temps. Un signal de parole est donc non-stationnaire, c'est-à-dire que ses propriétés statistiques (moyenne, écart-type...) varient en fonction du temps. Les chiffres 1 , 2, 3 et 4 représentent leur prononciation par le locuteur (voir Figure2.1).

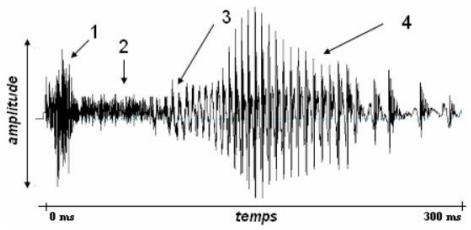

Figure 2.1 : Représentation d'un signal sonore.

2.3 Le prétraitement du signal de la parole

Le signal vocal est de nature analogique (continue en temps), en revanche, les systèmes de traitement sont des systèmes numériques, d'où la nécessité de cette phase.

La numérisation consiste à convertir un signal analogique contenant un nombre infini d'amplitudes en un signal numérique contenant un nombre fini de valeurs.

Le passage de l'analogique au numérique repose sur trois étapes : l'échantillonnage, la quantification, et le codage.

Ainsi, une fois le signal numérisé, la première action que doit subir le signal est la préaccentuation afin de relever les hautes fréquences puis segmenter en trame, chaque trame est constituée d'un nombre de N d'échantillons de parole, ces dernier (trames) sont des tranches allant de 10 à 30ms (d'où l'appellation d'analyse à court-terme). Ces ensuite sur ces trames que de parole que peuvent s'appliquer les opérations destinées à préparer le signal d'entrée pour la phase d'extraction des informations. Nous décrirons brièvement ces étapes.

2.3.1 La numérisation

2.3.1.1 L'échantillonnage

L'échantillonnage consiste à transformer un signal à temps continu ***x(t)*** en un signal à temps discret ***x(nTe)***

Le signal analogique est découpé en échantillons. Le nombre d'échantillons par seconde représente la fréquence d'échantillonnage f_e(Hz), qui est elle-même l'inverse de la période d'échantillonnage T_e. Le choix de la fréquence d'échantillonnage doit être adéquate (prélever assez de valeurs pour ne pas perdre l'information contenue) c'est-à-dire respecter le ***théorème de Shannon*** [34].

$$f_e \geq 2f_{max} \qquad (2.1)$$

f_e **:** la fréquence d'échantillonnage.

f_{max} **:** la fréquence maximale du signal vocal.

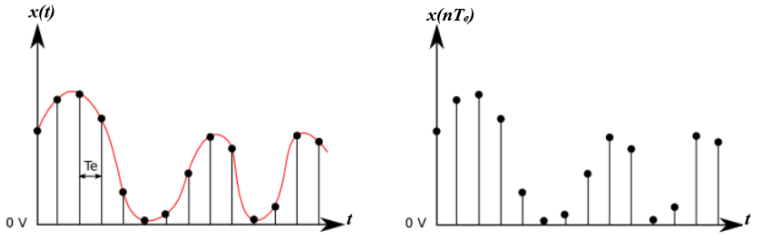

Figure 2.2 : Signal échantillonné. [33]

2.3.1.2 La quantification

La quantification d'un signal implique à placer les amplitudes des échantillons sur une échelle de valeurs à intervalles fixes. Chaque impulsion correspond ainsi à un nombre binaire unique. Une quantification à *n* bits permet d'utiliser 2^n valeurs différentes. La quantification a pour effet d'arrondir l'amplitude de chaque échantillon à l'une de ces 2^n valeurs. Ainsi on peut dire que le nombre de bits de quantification détermine donc *la précision en amplitude* ou *la dynamique* de la conversion, alors que la fréquence d'échantillonnage détermine *la précision temporelle* de la conversion. [34]

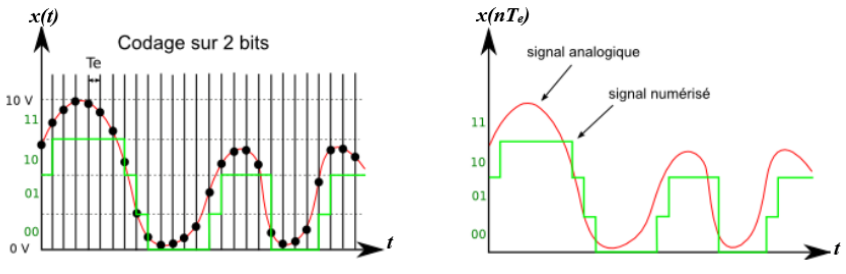

Figure 2.3 : Signal quantifié. [33]

2.3.1.3 Le codage

On appelle codage la transformation des différentes valeurs quantifiées en langage binaire qui permet le traitement du signal sur machine. [34]

2.3.2 Préaccentuation

Afin d'égaliser les sons aigus qui ont toujours une énergie plus faible que les sons graves, on réalise une opération de préaccentuation, qui consiste à favoriser la transmission des fréquences élevées, et à atténuer les fréquences basses (l'oreille humaine est plus sensible à la région du spectre autour des 1kHz)

Ainsi, un filtre de préaccentuation va amplifier la région centrale du spectre en faisant passer le signal vocal à travers un filtre passe-haut appelé préaccentuation, dont la fonction de transfert est [35]

$$\boldsymbol{H(z) = 1 - \propto z^{-1}} \qquad \boldsymbol{avec\ 0.9 \leq \propto \leq 1} \qquad (2.2)$$

Le signal préaccentué est donc donné par la formule

$$\boldsymbol{y(n) = s(n) - \propto s(n-1)} \qquad (2.3)$$

Où ***s(n)*** correspond au n^{ième} échantillon du signal d'entrée.

2.3.3 Fenêtrage

Pour traiter le signal vocal et s'assurer de sa stationnarité, il faut limiter le nombre d'échantillons. Pour cela, on effectue une opération de fenêtrage dont le but est d'adoucir les discontinuités en réduisant le poids des échantillons situés aux extrémités, par rapport à ceux situés au centre [35]. Pour ne pas perdre les informations, les fenêtres sont assez larges et se recouvrent entre elles.

Donc, si nous définissons $w(n)$ fenêtre où $0 \leq n \leq N-1$, où N représente le nombre d'échantillon dans chacune des trames, alors le résultat du fenêtrage est le signal $\boldsymbol{x(n)}$, donné par la formule :

$$\boldsymbol{x(n) = s(n)w(n)} \qquad (2.4)$$

La fenêtre la plus utilisée dans le domaine de la reconnaissance vocale est la fenêtre de Hamming, que l'on peut réaliser à l'aide de la formule [36]

$$w(n) = \begin{cases} \boldsymbol{0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)} & \boldsymbol{avec\ 0 \leq n \leq N-1} \\ \boldsymbol{0} & \boldsymbol{ailleurs} \end{cases} \qquad (2.5)$$

2.4 L'analyse du signal de parole

2.4.1 L'analyse temporelle

2.4.1.1 Energie, amplitude et puissance

L'un des outils permettant de représenter fidèlement la variation de l'amplitude du signal vocal $\boldsymbol{x(n)}$ dans le temps est l'énergie, selon le type et l'amplitude du son produit, l'énergie peut être très faible pendant le silence et élevée pendant l'activité vocale. Généralement, elle est définie par [10]

$$E_n = \sum_{n=0}^{N-1} \boldsymbol{x^2(n)} \qquad (2.6)$$

$\boldsymbol{x(n)}$: signal de '*n*' échantillons.

$\boldsymbol{N}$: nombre d'échantillon par fenêtre.

Le temps de calcul de l'énergie à court terme avec une élévation carrée pour chaque échantillon est très élevé. En pratique, il est préférable d'utiliser une autre forme, l'amplitude moyenne, qui se définit comme suit

$$M_n = \sum_{n=0}^{N-1} \boldsymbol{x^2(n)} \qquad (2.7)$$

La puissance à court terme d'un segment de parole de longueur *N* est définie par

$$P_n = \frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{x^2(n)} \qquad (2.8)$$

Il faut noter que l'énergie court terme et la puissance court terme fournissent à un facteur près $(\frac{1}{N})$ la même information.

2.4.1.2 Taux de passage par zéro (TPZ)

Un autre outil très pratique pour le traitement de la parole est le taux de passage par zéro. Pour un signal échantillonné, le passage par zéro se produit lorsque deux échantillons successifs sont de signe opposé, il est particulièrement utile pour distinguer une zone voisée d'une zone non voisée [35]. Le taux de passage par zéro à court terme est estimé par la formule

$$ZC = \frac{1}{2N} \sum_{n=0}^{N-1} \left[\boldsymbol{sgn}\big(\boldsymbol{x(n)}\big) - \boldsymbol{sgn}\big(\boldsymbol{x(n-1)}\big) \right] \qquad (2.9)$$

Puisqu'une zone non voisée voit son énergie distribuée principalement dans les hautes fréquences, les valeurs de ZC seront nettement plus élevées pour un son non voisé que pour un son voisé.

2.4.1.3 Autocorrélation

L'autocorrélation d'un signal est la corrélation de ce signal avec lui-même, et son estimation est donnée par la formule suivante [37]

$$\phi_{xx}(k) = \frac{1}{N-k} \sum_{n=0}^{N-1-k} [\, \boldsymbol{x(n)} \times \boldsymbol{x(n+k)} \,] \qquad (2.10)$$

L'idée d'utiliser la fonction d'autocorrélation est de déterminer à quel point deux échantillons successifs d'un signal sont similaires. Parmi ses autres applications, on peut s'y référer pour estimer la fréquence de la fréquence fondamentale (ou pitch). En effet, ce dernier apparaît comme un pic dans la fonction d'autocorrélation, qui doit être suffisamment isolé pour pouvoir l'évaluer. [38]

2.4.1.4 Covariance

Dans cette méthode on fixe dès le départ une portion du signal [0, N-1], sur laquelle l'énergie résiduelle est évaluée.

Dans ce cas on n'applique pas de fenêtrage au signal, alors contrairement à la méthode de l'autocorrélation où la covariance peut être réduite à une simple autocorrélation, la fonction de covariance est utilisée directement dans le système d'équations pour trouver les coefficients a_i. [35]

$$\phi_{xx}(i,k) = \sum_{n=-i}^{N-i-1} x(n) \times x(m+1-k); \qquad 1 \leq i \leq p \; et \; 0 \leq k \leq p \qquad (2.11)$$

Ainsi le système d'équations à résoudre aura la forme matricielle suivante

$$\begin{bmatrix} \phi_{xx}(1,1) & \phi_{xx}(1,2) \dots & \phi_{xx}(1,p) \\ \phi_{xx}(2,1) & \phi_{xx}(2,2) \dots & \phi_{xx}(2,p) \\ \vdots & \vdots & \vdots \\ \phi_{xx}(p,1) & \phi_{xx}(p,2) & \phi_{xx}(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} \phi_{xx}(1,0) \\ \phi_{xx}(2,0) \\ \vdots \\ \phi_{xx}(p,0) \end{bmatrix} \qquad (2.12)$$

2.4.2 L'analyse fréquentielle

2.4.2.1 Analyse par Spectrogramme

Il est souvent intéressant de représenter l'évolution temporelle d'un signal, sous la forme d'un spectrogramme. Le spectrogramme est une représentation tridimensionnelle, où le temps est représenté sur l'axe des abscisses, la fréquence sur l'axe des ordonnées et l'énergie apparaît sous la forme de niveau de gris pour un temps et une fréquence donnée.

Il existe deux types de spectrogrammes, les spectrogrammes à bandes larges et les spectrogrammes à bandes étroites. Les premiers sont obtenus avec des fenêtres de courte durée. Ils mettent en évidence l'enveloppe spectrale (formants) du signal, où les périodes voisés apparaissent sous forme de bandes verticales plus sombres (voir figure 2.4).

Les spectrogrammes à bande étroite sont obtenus avec des fenêtres de 30 à 40 ms, et ils offrent une bonne résolution au niveau de la fréquence, où les harmoniques du signal dans les zones voisées apparaissent comme des bandes de fréquence horizontales (voir figure 2.5). [39]

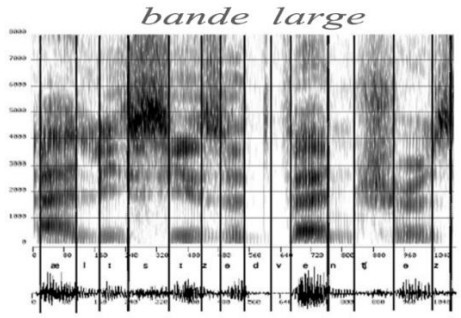

Figure 2.4 : Spectrogramme à bande large.

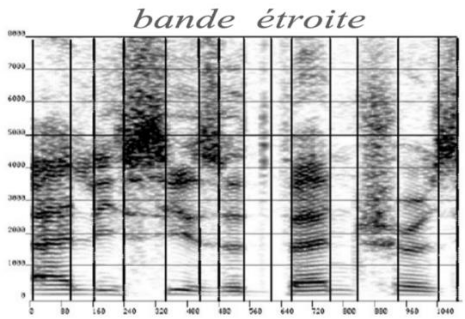

Figure 2.5 : Spectrogramme à bande étroite.

2.4.2.2 Analyse par Transformée de Fourier à court terme (TFCT)

La transformée de Fourier (TF) a suscité beaucoup d'intérêt et reste un outil essentiel dans le traitement du signal. Or, si cet opérateur peut connaître les différentes fréquences contenues dans le signal (spectre de fréquence) néanmoins, il ne peut connaître à quel instant ces fréquences sont émises. En d'autres termes, FT donne des informations globales plutôt que locales sur les composantes de fréquence. [40]

Cette perte de localité est préjudiciable pour l'analyse de signaux non stationnaires ou quasi stationnaires, notamment pour les signaux de parole où la connaissance temporelle de la distribution spectrale est cruciale pour y remédier, c'est-à-dire aux problèmes par l'inadéquation de la TF dans le cas des signaux non stationnaire on définit la TFCT

La TF est définie par

$$X(f) = \int_{-\infty}^{+\infty} \boldsymbol{x(t)\, e^{-j2\pi ft}\, dt} \qquad (2.13)$$

La TF discrète

$$X(k) = \sum_{n=0}^{N-1} \boldsymbol{x(n)\, e^{-j\frac{2\pi}{N}k.n}} \qquad (2.14)$$

$f = \frac{k}{N}$ $o\grave{u}\ k = 0, 1, \dots, N-1$ $\quad et \quad$ $n = 0, 1, \dots, N-1$

$x(n)$: *le signal échantillionné*

N: *le nombre de point de la suite temporel* $x(n)$

Afin de réduire le temps de calcul de la TFD, on applique la transformée de Fourier rapide (FFT).

- **La transformée de Fourier Rapide (FFT)**

La transformée de Fourier rapide est simplement une TFD calculée à partir d'un algorithme qui réduit le nombre d'opérations, notamment le nombre de multiplications à effectuer. Cependant, il convient de noter que réduire le nombre d'opérations arithmétiques à effectuer n'équivaut pas à réduire le temps d'exécution, et cela dépend de l'architecture du processeur effectuant le traitement.

L'algorithme exige que N soit une puissance de 2. Le principe de l'algorithme consiste à décomposer le calcul de la TFD d'ordre N = 2^l en l étapes successives, mais celle-ci ne peut déduire à quel instant se trouve l'amplitude maximale de $x(t)$. [41]

Pour pallier cette lacune, une transformée de Fourier à Court Terme (TFCT) est mise en œuvre. Il s'agit alors d'appliquer une fenêtre d'observation glissante au signal (en général de Hamming), qui sera transformée dans le domaine temporel de sorte que l'hypothèse de stationnarité soit localement vérifiée, et qu'au lieu d'appliquer la TF au signal global $x(t)$, elle sera appliquée à chacune des tranches ainsi découpées. Ainsi un spectre « local » est obtenu. L'ensemble des spectres « locaux » montre alors l'évolution du spectre dans le temps.

L'axe horizontal porte le nombre de fenêtre (temps) et l'axe vertical porte la fréquence. En effet on présume que le signal est à stationnaire (à l'arrêt) sur toute la longueur de chaque fenêtre. Ces fenêtres permettent d'améliorer de manière significative le spectre.

Pour un signal $x(t)$, la TFCT est définie par

$$TFCT(\tau, f) = \int_{-\infty}^{+\infty} \boldsymbol{x(t) e^{-j2\pi f t}\, w(t-\tau)\ dt} \qquad (2.15)$$

τ : représente le paramètre de translation de la fenêtre d'analyse.

$w(t-\tau)$: représente une fonction de fenêtrage centrée en τ.

Pour obtenir la représentation spectrale autour de τ , il suffit de déplacer par translation la fenêtre w et d'effectuer une transformation de Fourier sur le signal ainsi fenêtré.

La TFCT est la transformée d'un produit donc c'est la convolution des transformées.

$$TFCT = X(\omega) * w(\omega) \qquad (2.16)$$
$$o\grave{u}\ \omega = 2\pi f\ \ (pulsation)$$

2.4.2.3 Analyse cepstrale

Le spectre donné par la FFT contient des informations sur la source et le conduit vocal, mais leur intermodulation fait qu'il est difficile de mesurer la $\boldsymbol{f_0}$ (fréquence fondamentale) et des fréquences des formants qui caractérisent respectivement la source et le conduit. Le lissage cepstral est une méthode qui vise à séparer la contribution du conduit et de la source d'excitation par déconvolution. Pour cela nous supposons que le signal vocal $x(n)$ est produit par un signal excitateur $g(n)$ (source glottique) traversant un système linéaire passif de réponse impulsionnelle $b(n)$ (conduit). Avec ces suppositions, nous pouvons écrire

$$\hat{x}(n) = g(n) \otimes b(n) \qquad (2.17)$$

Afin de déconvoluer plus facilement $\boldsymbol{x(n)}$ il suffit de transposer le problème par homomorphisme dans un espace ou l'opérateur ⊗ (convolution) correspond à l'opérateur + (addition). En pratique, cette transposition homomorphe se fait par les étapes suivantes

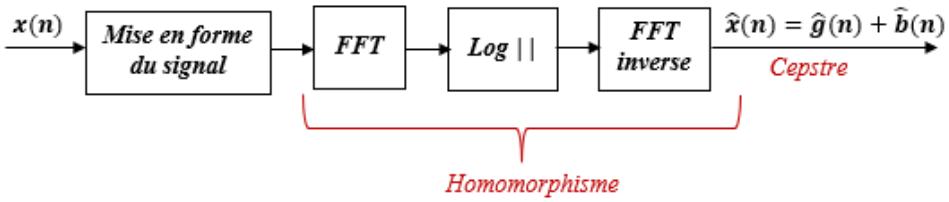

Figure 2.6 : Analyse homomorphique de la parole

Où $\hat{\boldsymbol{x}}(\boldsymbol{n})$ sont les coefficients cepstraux approchés, prenant des valeurs dans un domaine pseudo-temporel réel appelé domaine quéfrentiel. La structure de la parole et les hypothèses sur la source d'excitation et le conduit vocal permettent de dire :

$\hat{\boldsymbol{g}}(\boldsymbol{n})$ Se réduit théoriquement à une séquence d'impulsions de période n_0 (n_0 correspond à la fréquence fondamentale $\boldsymbol{f_0}$).

$\hat{\boldsymbol{b}}(\boldsymbol{n})$ Décroît rapidement avec n (en $1/n$) et devient négligeable lorsque $n>n_0$.

Dans ces conditions, on peut dire que la contribution du conduit est localisée aux basses fréquences ($n < n_0$) et que la séquence d'impulsions reflète la contribution de la source.[43]

Il existe deux principaux types de coefficients cepstraux [42] :

- Les coefficients MFCCs.
- Les coefficients LPCCs.

Plusieurs travaux ont été publiés afin de comparer les diverses techniques de paramétrisation, le défi de ces travaux consistait à cibler les meilleurs paramètres qui représentent efficacement les propriétés caractéristiques de chaque locuteur. Les résultats les plus probants ont été obtenus en utilisant la méthode MFCC. [44]

2.4.2.3.1 Les coefficients cepstraux MFCCs

Les coefficients cepstraux MFCC (*Mel-Frequency Cepstral coefficients*) ont été intensivement utilisés comme vecteur de traits caractéristiques dans les systèmes de reconnaissance de la parole et du locuteur, ces coefficients, sont basés sur une échelle appelée échelle de *Mel*. Cette échelle redistribue les fréquences selon une échelle non linéaire (linéaire jusqu'à 1000Hz et logarithmique au-delà de 1000Hz) qui consiste en la définition de bandes critiques de perception (à l'aide d'un banc de filtres). Celle-ci correspond à la répartition fréquentielle du mécanisme d'audition. [45]

Le passage de l'échelle fréquentielle à l'échelle de Mel est régi par l'équation suivante [45]

$$Mel\ (f) = \mathbf{2595}.\boldsymbol{log_{10}}\left(\mathbf{1} + \frac{f}{\mathbf{700}}\right) \qquad (2.18)$$

On utilise un banc de filtres triangulaire, positionnés uniformément sur l'échelle Mel, c'est à dire non uniformément dans le domaine fréquentiel.

Le schéma suivant permet de récapituler le processus pour déterminer les coefficients MFCC

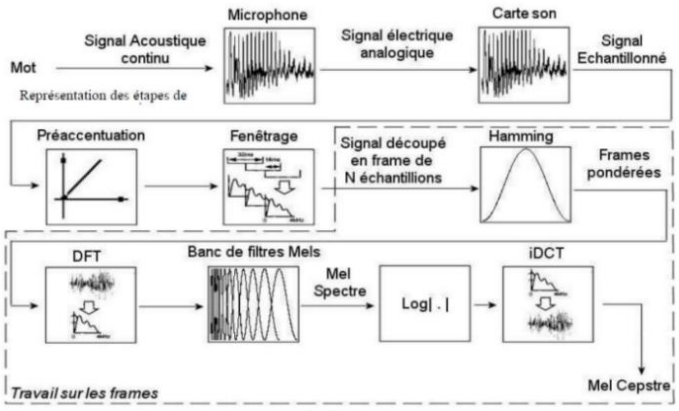

Figure 2.7 : Processus pour l'obtention des coefficients MFCCs

- ✓ ***Prétraitement*** : on effectue une préaccentuation du signal au début du traitement, puis on utilise une fenêtre de type Hamming pour décomposer le signal en un ensemble de segments d'échantillons.

- ✓ ***FFT*** : En pratique, la transformée en Z est remplacée par une transformée discrète de Fourier (ou FFT) qui a les mêmes propriétés linéaires que la transformée en Z.

- ✓ ***Banc de filtres*** : Une suite de filtres triangulaires appliqués selon l'échelle Mel pour à la fois lisser le spectre et réduire les informations à traiter.

- ✓ ***Log()***: le logarithme est appliquer pour transformer la multiplication en addition

- ✓ ***DCT(Discret Cosine Transform)*** : A la fin du processus, on revient dans l'espace temporel par FFT inverse. Puisque nous ne traitons que la partie réelle du signal, la DCT peut facilement faire la transformée inverse.

Si on pose '*K'* le nombre de filtres et '*L'* le nombre de coefficients qu'on veut avoir, les coefficients MFCCs seront

$$\hat{c}(n) = \sum_{\boldsymbol{k=1}}^{\boldsymbol{k}} (\boldsymbol{log}\ \widehat{\boldsymbol{E}}_{\boldsymbol{k}})\ \boldsymbol{cos}\left[\boldsymbol{n}\left(\boldsymbol{k} - \frac{\mathbf{1}}{\mathbf{2}}\right)\frac{\boldsymbol{\pi}}{\boldsymbol{k}}\right] \qquad (2.19) \quad ;\ n = 1,2,\ldots.,L$$

$avec\ \hat{E}_k\ L'énergie$ à $la\ sortie\ des\ filtres,\quad k = 1,2,\ldots,K$

On note que le coefficient $\hat{\boldsymbol{c}}_{\boldsymbol{0}}$ a été écarté, cela en raison du fait qu'il représente l'énergie moyenne dans la trame de la parole et ne contribue pas de manière significative dans les applications *RAP*.

2.4.3 L'analyse temps-fréquence

2.4.3.1 L'analyse par les ondelettes (transformée en ondelettes)

Contrairement à la transformée de Fourier, la transformée en ondelettes (*TO*) n'est pas limitée aux techniques d'analyse fréquentielle. En appliquant une transformée en ondelettes à un signal, son comportement peut être observé dans le domaine fréquentiel mais aussi temporel. Cette analyse temps-fréquence la place dans le groupe d'analyse des méthodes multi-échelles telles que la transformée de Fourier à fenêtre glissante et la transformée en cosinus.

Le principe de base consiste à convoluer le signal analysé avec une fonction appelée ondelette (wavelet).

Une ondelette Ψ est une fonction de moyenne nulle [46]

$$\int_{-\infty}^{+\infty} \boldsymbol{\Psi}(\boldsymbol{t})\, \boldsymbol{dt} = \mathbf{0} \qquad (2.20)$$

Qui peut être dilatée par un paramètre d'échelle ***'u'*** et translatée de ***'s'***

$$\Psi_{u,s}(t) = \frac{\mathbf{1}}{\sqrt{\boldsymbol{s}}}\boldsymbol{\Psi}\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{\boldsymbol{s}}\right) \qquad (2.21)\quad avec\quad s>0\quad et\quad u\in\mathbb{R}$$

L'ondelette Ψ, appelée ondelette mère, produit une base orthonormée de fonctions appelées ondelettes filles ou plus simplement ondelettes. La transformée en ondelettes d'un signal x(t) à une échelle ***'u'*** et une position ***'s'*** est obtenue en corrélant le signal avec l'ondelette

$$WT_x(u,s) = \int_{-\infty}^{+\infty} \boldsymbol{x}(\boldsymbol{t})\frac{\mathbf{1}}{\sqrt{\boldsymbol{s}}}\boldsymbol{\Psi}^*\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{\boldsymbol{s}}\right)\boldsymbol{dt} \qquad (2.22)$$

- **La transformée en ondelette discrète**

La transformée discrète en ondelettes (*DWT - Discrete Wavelet Transform*) est une application numérique de TO. Son utilisation est populaire car elle peut être facilement implémenté sur des circuits numériques (FPGA, DSP). *DWT* utilise une technique de fenêtrage qui consiste à traiter les signaux un par un. Le principe général de la *DWT* est de décomposer un signal en plusieurs sous-signaux. Un moyen efficace de mettre en œuvre DWT à l'aide de bancs de filtres. [47] [48]

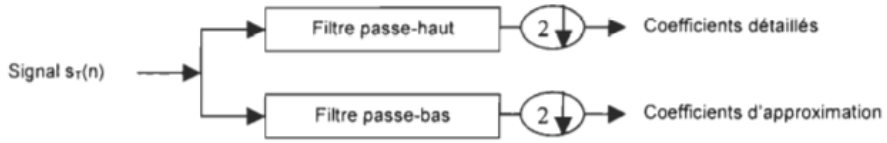

Figure 2.8 : Principe de décomposition en ondelettes du signal sT (n)

Le signal de base $s_T(n)$ (le signal à traiter) traverse dans un filtre passe-haut et parallèlement dans un filtre passe-bas. Ce signal discret est dyadique c'est-à-dire qu'il se compose de 2^k échantillons, où k est un entier. Il convient de noter que, les coefficients des filtres passe-haut et passe-bas sont les mêmes. C'est juste que leur ordre est inversé, Autrement dit, le premier coefficient du filtre passe-bas correspond au dernier coefficient du filtre passe-haut, et ainsi de suite. À la sortie de chaque filtre, une opération de sous-échantillonnage par 2 est effectué. On obtient les coefficients A et D, appelés respectivement coefficients d'approximation et coefficients des détails. Les équations correspondantes sont comme suit

$$a(k) = \sum_{n=1}^{N} \boldsymbol{g(n-2k)\, s_T(n)} \qquad (2.23)$$

$$d(k) = \sum_{n=1}^{N} \boldsymbol{h(n-2k)\, s_T(n)} \qquad (2.24)$$

g(n) et h(n) correspond au filtre passe-bas et passe-haut, respectivement.

En réalisant l'opération présentée à la figure 16, on passe du niveau j au niveau *j+1*.Le signal de base est généralement au niveau *j=0.*

Avec le facteur de dilatation $s = s_0^j$ et le paramètre de translation $u = k\, s_0^j$ discrétisés, la *DWT* est donnée par 1'équation suivante

$$DWT_x(j,k) = \int_{-\infty}^{+\infty} \boldsymbol{x(t)}\, \frac{\mathbf{1}}{\sqrt{\boldsymbol{a_0^j}}}\, \boldsymbol{\Psi^*\left(a_0^{-j} t - k\right) dt} \qquad (2.25)$$

2.5 Conclusion

Ce chapitre nous a permis d'introduire les éléments de base de la mise en œuvre d'un système de reconnaissance vocale qui consiste à extraire les informations contenues dans les ondes sonores. Divers outils pour le traitement du signal vocal ont été étudiés ; Les signaux de parole ne peuvent être considérés comme quasi-stationnaires que pour des intervalles de temps de durée limitée.

On s'intéressera au prochain chapitre à la reconnaissance automatique du locuteur coté expérimental.

Partie expérimentale

CHAPITRE 03 :

Application à la reconnaissance du genre du locuteur

3.1. Introduction

Dans les chapitres précédents nous avons présenté les aspects généraux de la parole (production et audition), les différents outils nécessaires pour son traitement et sa paramétrisation, ainsi qu'un aperçu sur les principales approches de reconnaissance qu'on retrouve dans les différents systèmes RAL. Nous allons maintenant aborder l'aspect pratique par une implémentation de la méthode covariance, MFCC et énergie, ensuite, faire une fusion globale des méthodes prises deux à deux afin d'évaluer la validité de notre système par une analyse des résultats obtenus.

3.2. Description et acquisition de la base de données

Nous disposons d'une base de données TIMIT composée de cent (100) locuteurs dont cinquante (50) femmes et cinquante (50) hommes.

Nous avons choisi d'évaluer nos systèmes de reconnaissance automatique du locuteur avec la base de données acoustiques TIMIT [49] pour plusieurs raisons.

- Parce que c'est une base de référence communément utilisée par les chercheurs pour comparer leurs résultats.
- Parce qu'elle est fournie avec une segmentation phonétique manuelle, qui simplifie l'apprentissage des modèles phonétiques d'un système RAP continue.

TIMIT est un corpus de parole dédié à la reconnaissance de la parole continue indépendante du locuteur. Dans cette base de données, 630 locuteurs américains répartis sur 8 dialectes régionaux ont participé à la procédure d'enregistrement sonores des phrases.

L'enregistrement sonore des phrases c'est déroulé dans de bonnes conditions (le signal sonore est échantillonné à 16KHz avec 16 bits de codage pour chaque échantillon), les fichiers sons sont en format Wave (*'.wav'*).

3.3. Le langage de programmation

Le langage utilisé est Matlab (Matrix Laboratory) qui est un logiciel interactif, développé par Math Works, destiné notamment au traitement numérique des données. Ce logiciel possède un langage de programmation puissant et simple à utiliser, précis, robuste et rapide.

3.4. Fonctionnement du système de reconnaissance du locuteur

La reconnaissance du locuteur se déroule en deux phases

a) La phase d'apprentissage.
b) La phase de reconnaissance.

3.5. Répartition de la base de données

Notre base de données est composée de cent (100) fichiers audios dont 50 femmes et 50 hommes.

- Les fichiers audios femmes sont numérotés de 1 à 50.
- Les fichiers audios hommes sont numérotés de 51 à 100.

Répartition naïve

On considère une base de données qu'on divise en 2/3 apprentissage et 1/3 test, on aura donc [50]

$$\boldsymbol{apprentissage} = \mathbf{100} \times \frac{\mathbf{2}}{\mathbf{3}} \approx \mathbf{60}\ fichiers\ audio$$

$$\boldsymbol{test} = \mathbf{100} \times \frac{\mathbf{1}}{\mathbf{3}} \approx \mathbf{40}\ fichiers\ audio$$

3.5.1. Répartition 01

Le tableau 3.1 représente la première répartition de cent (100) fichiers audios, l'apprentissage compte les soixante (60) premiers fichiers et le test les quarante (40) derniers.

Fichiers audio	Femmes	Hommes
Apprentissage	1 à 30	51 à 80
Test	31 à 50	81 à 100

Tableau 3.1 : Première répartition des fichiers audios.

3.5.2. Répartition 02

Le tableau 3.2 représente la deuxième répartition de cent (100) fichiers audios, le test compte les quarante (40) premiers fichiers et l'apprentissage les soixante (60) derniers.

Fichiers audio	Femmes	Hommes
Apprentissage	21 à 50	71 à 100
Test	1 à 20	51 à 70

Tableau 3.2 : Deuxième répartition des fichiers audios.

3.5.3. Répartition 03

Le tableau 3.3 représente la troisième répartition de cent (100) fichiers audios, le test compte les quarante (40) fichiers du milieu et l'apprentissage les soixante (60) restants.

Fichiers audio	Femmes	Hommes
Apprentissage	1 à 30	71 à 100
Test	31 à 50	51 à 70

Tableau 3.3 : Troisième répartition des fichiers audios.

3.6. Méthodologie proposée pour l'extraction des paramètres désirés

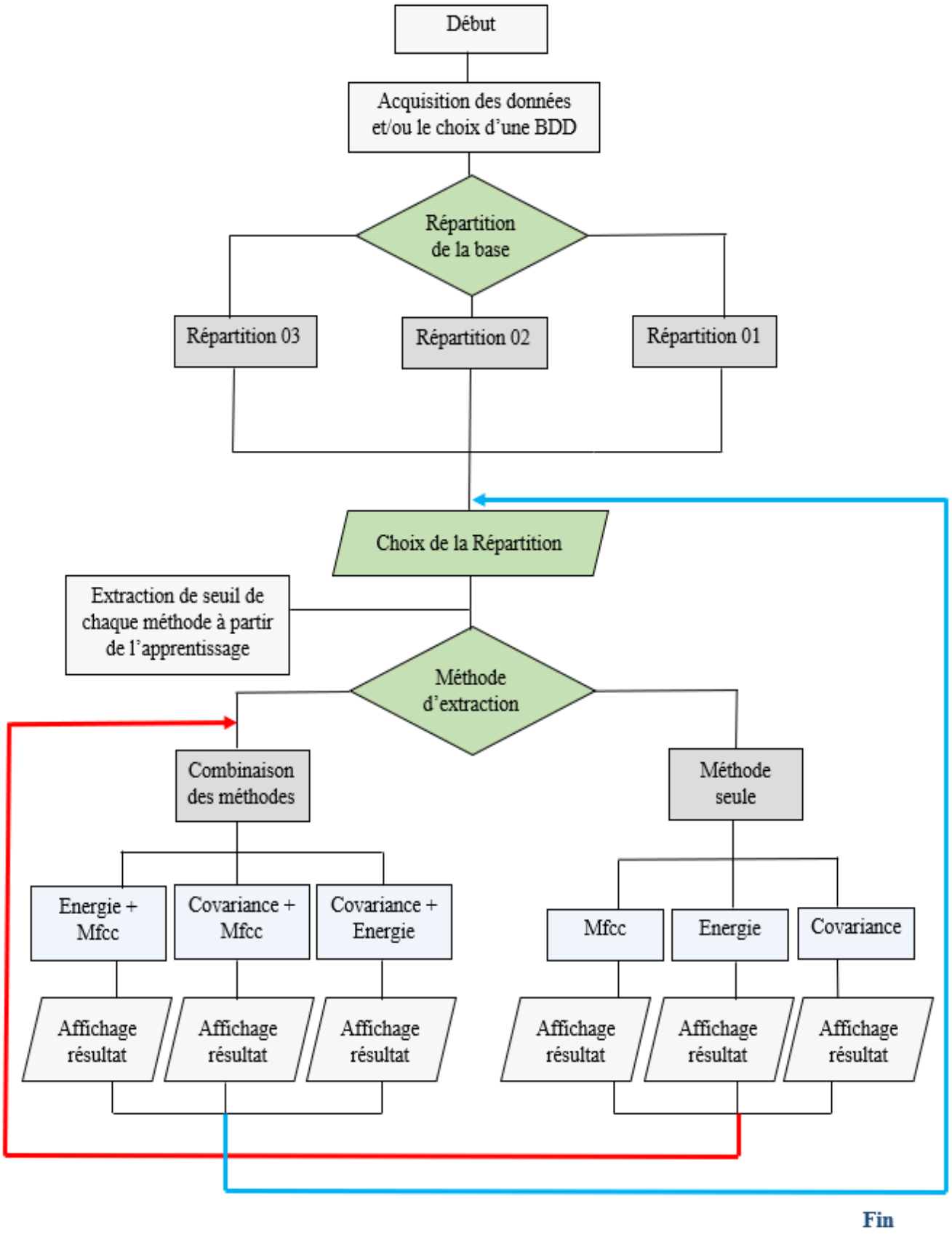

Figure 3.1 : Organigramme représentant la méthodologie d'extraction des paramètres.

3.7. Résultats et discussions

Dans notre travail nous avons utilisée deux approches des méthodes utilisés (covariance, mfcc, énergie) pour comparer les résultats et laquelle des deux est meilleur. Nous commençons d'abord par les méthodes seule ensuite les méthodes combinées.

3.7.1. Une seule méthode

Dans cette partie-là nous allons tester chaque méthodes seules et voir le meilleur taux de précision de celle-ci pour chaque répartition ensuite comparer les méthodes pour chaque répartition.

3.7.1.1.Résultats des répartitions de chaque méthode

Pour obtenir les résultats on passe par une phase d'apprentissage pour tirer le seuil, ensuite, le seuil obtenu sera utilisé dans la phase test.

Un exemple des programmes du système de reconnaissance des méthodes utilisées, en occurrence celui de la méthode « énergie » est présenté ci-dessous :

```
clear;clc;


mfcc_tot=[];
energy_tot=[];
for i=1:40

    i
    fichier = strcat('C:\Users\USER\Desktop\test','\','s (',int2str(i),').wav');
    [x,fs]=audioread(fichier);

    Fd=0.025;
    P=3*log(fs);
    N=Fd*fs;
    inc=N/2;
    S = melcepst(x, fs,'e',12,P,N,inc);
    energy=S(:,13);
    energy=mean(energy);
    mfcc_coeff=S(:,1:12);
    mfcc_coeff=mean(mfcc_coeff);
    mfcc_coeff=mean(mfcc_coeff);
    fichier1 = strcat('C:\Users\USER\Desktop\test','\','s(',int2str(i),')_MFCC');
    fichier2 = strcat('C:\Users\USER\Desktop\test','\','s(',int2str(i),')_energy');
    save(fichier1, 'S');
    save(fichier2, 'energy');

 %seuil de l'energie
seuil_energy=mean(energy);
energy_tot=[energy_tot,seuil_energy];
seuil_energy=mean(energy_tot);

%seuil du mfcc
seuil_mfcc=mean(mfcc_coeff);
mfcc_tot=[mfcc_tot,seuil_mfcc];
seuil_mfcc=mean(mfcc_tot);

%
%  si la valeur est supérieure à 1 pour cette observation particulière
if energy<-0.6365
    disp('La voix de femme est reconnue')

else
    disp('La voix d homme est reconnue')

end
end
```

Figure 3.2 : Programme du système de reconnaissance pour la méthode énergie

3.7.1.1.1. Répartition 01

a) COVARIANCE

Pour obtenir les résultats dans le tableau 3.4, nous avons utilisé un seuil de détection $f_{max} = 165$ extrait de la référence [12]

- Si $f_{max} > 165$, le locuteur est une femme.
- Si $f_{max} < 165$, le locuteur est un homme.

Le tableau 3.4 représente le taux de précision total (Femmes + Hommes) de **87.5%**

Méthode	**Taux de précision**		
COVARIANCE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{\mathbf{35}}{\mathbf{40}} \times \mathbf{100} = \mathbf{87.5\%}$

Tableau 3.4 : Taux de précision des deux genres (femmes + hommes) avec la méthode de covariance.

Le taux de précision pour les femmes est de **90%** et celui des hommes de **85%** et d'après ces résultats, nous remarquons que la méthode de covariance donne un meilleur résultat pour les femmes que pour les hommes.

b) ENERGIE

Pour obtenir les résultats dans le tableau 3.5, nous avons utilisé un seuil de détection $seuil_{energy} = -0.6365$ obtenu de la phase d'apprentissage

- Si $energy < seuil_{energy}$, le locuteur est une femme.
- Si $energy > seuil_{energy}$, le locuteur est un homme.

Le tableau 3.5 représente le taux de précision total (Femmes + Hommes) de **67.5%**

Méthode	**Taux de précision**		
ENERGIE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{11}{20} \times 100 = \mathbf{55\%}$	$\frac{27}{40} \times 100 = \mathbf{67.5\%}$

Tableau 3.5 : Taux de précision des deux genres (femmes + hommes) avec la méthode d'énergie.

Le taux de précision pour les femmes est de **80%** et celui des hommes de **55%** et d'après ces résultats, nous remarquons que la méthode d'énergie donne un meilleur résultat pour les femmes que pour les hommes.

c) MFCC

Pour obtenir les résultats dans le tableau 3.6 nous avons utilisé un seuil de détection $seuil_{mfcc} = -0.3847$ obtenu de la phase d'apprentissage (Le nombre des coefficients MFCC est de 12 coefficients).

- Si $mfcc_{coeff} < seuil_{mfcc}$, le locuteur est une femme.
- Si $mfcc_{coeff} > seuil_{mfcc}$, le locuteur est un homme.

Le tableau 3.6 représente le taux de précision total (Femmes + Hommes) de **97.5%**

Méthode	**Taux de précision**		
MFCC	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\mathbf{\frac{39}{40} \times 100 = 97.5\%}$

Tableau 3.6 : Taux de précision des deux genres (femmes + hommes) avec la méthode de Mfcc.

Le taux de précision pour les femmes est de **95%** et celui des hommes de **100%** et d'après ces résultats, nous remarquons que la méthode mfcc donne un meilleur résultat pour les hommes que pour les femmes.

Interprétation : nous avons déduit dans la première répartition que le meilleur taux de précision des trois méthodes, est celui de la méthode mfcc.

3.7.1.1.2. Répartition 02

a) COVARIANCE

Pour obtenir les résultats dans le tableau 3.7, nous avons utilisé un seuil de détection $f_{max} = 165$ extrait de la référence [12]

- Si $f_{max} > 165$, le locuteur est une femme.
- Si $f_{max} < 165$, le locuteur est un homme.

Le tableau 3.7 représente le taux de précision total (Femmes + Hommes) de **95%**

Méthode	**Taux de précision**		
COVARIANCE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\mathbf{\frac{38}{40} \times 100 = 95\%}$

Tableau 3.7 : Taux de précision des deux genres (femmes + hommes) avec la méthode de covariance.

Le taux de précision pour les femmes est de **90%** et celui des hommes de **100%** et d'après ces résultats, nous remarquons que la méthode de covariance donne un meilleur résultat pour les hommes que pour les femmes.

b) ENERGIE

Pour obtenir les résultats dans le tableau 3.8, nous avons utilisé un seuil de détection $seuil_{energy} = -0.7439$ obtenu de la phase d'apprentissage

- Si $energy < seuil_{energy}$, le locuteur est une femme.
- Si $energy > seuil_{energy}$, le locuteur est un homme.

Le tableau 3.8 représente le taux de précision total (Femmes + Hommes) de **62.5%**

Méthode	**Taux de précision**		
ENERGIE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{9}{20} \times 100 = \mathbf{45\%}$	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{\mathbf{25}}{\mathbf{40}} \times \mathbf{100} = \mathbf{62.5\%}$

Tableau 3.8 : Taux de précision des deux genres (femmes + hommes) avec la méthode d'énergie.

Le taux de précision pour les femmes est de **45%** et celui des hommes de **80%** et d'après ces résultats, nous remarquons que la méthode énergie donne un meilleur résultat pour les hommes que pour les femmes.

c) MFCC

Pour obtenir les résultats dans le tableau 3.9 nous avons utilisé un seuil de détection $seuil_{mfcc} = -0.3775$ obtenu de la phase d'apprentissage (Le nombre des coefficients MFCC est de 12 coefficients).

- Si $mfcc_{coeff} < seuil_{mfcc}$, le locuteur est une femme.
- Si $mfcc_{coeff} > seuil_{mfcc}$, le locuteur est un homme.

Le tableau 3.9 représente le taux de précision total (Femmes + Hommes) de **100%**

Méthode	**Taux de précision**		
MFCC	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$

Tableau 3.9 : Taux de précision des deux genres (femmes + hommes) avec la méthode de Mfcc.

Dans ce cas la méthodes mfcc donne un taux de précision de 100% pour les deux genres (Femmes + Hommes).

Interprétation: nous avons déduit dans la deuxième répartition que le meilleur taux de précision des trois méthodes, est celui de la méthode mfcc.

3.7.1.1.3. Répartition 03

a) COVARIANCE

Pour obtenir les résultats dans le tableau 3.10, nous avons utilisé un seuil de détection $f_{max} = 165$ extrait de la référence [12]

- Si $f_{max} > 165$, le locuteur est une femme.
- Si $f_{max} < 165$, le locuteur est un homme.

Le tableau 3.10 représente le taux de précision total (Femmes + Hommes) de **95%**

Méthode	**Taux de précision**		
COVARIANCE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\mathbf{\frac{38}{40} \times 100 = 95\%}$

Tableau 3.10: Taux de précision des deux genres (femmes + hommes) avec la méthode de covariance

Le taux de précision pour les femmes est de **90%** et celui des hommes de **100%** et d'après ces résultats, nous remarquons que la méthode covariance donne un meilleur résultat pour les hommes que pour les femmes.

b) ENERGIE

Pour obtenir les résultats dans le tableau 3.11, nous avons utilisé un seuil de détection $seuil_{energy} = -0.7349$ obtenu de la phase d'apprentissage

- Si $energy < seuil_{energy}$, le locuteur est une femme.
- Si $energy > seuil_{energy}$, le locuteur est un homme.

Le tableau 3.11 représente le taux de précision total (Femmes + Hommes) de **72.5%**

Méthode	**Taux de précision**		
ENERGIE	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{13}{20} \times 100 = \mathbf{65\%}$	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\mathbf{\frac{29}{40} \times 100 = 72.5\%}$

Tableau 3.11 : Taux de précision des deux genres (femmes + hommes) avec la méthode d'énergie.

Le taux de précision pour les femmes est de **65%** et celui des hommes de **80%** et d'après ces résultats, nous remarquons que la méthode énergie donne un meilleur résultat pour les hommes que pour les femmes.

c) MFCC

Pour obtenir les résultats dans le tableau 3.12 nous avons utilisé un seuil de détection $seuil_{mfcc} = -0.3895$ obtenu de la phase d'apprentissage (Le nombre des coefficients mfcc est de 12 coefficients).

- Si $mfcc_{coeff} < seuil_{mfcc}$, le locuteur est une femme.
- Si $mfcc_{coeff} > seuil_{mfcc}$, le locuteur est un homme.

Le tableau 3.12 représente le taux de précision total (Femmes + Hommes) de **97.5%**

Méthode	**Taux de précision**		
MFCC	**Femmes**	**Hommes**	**Total**
Pourcentage	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\mathbf{\frac{39}{40} \times 100 = 97.5\%}$

Tableau 3.12 : Taux de précision des deux genres (femmes + hommes) avec la méthode de Mfcc.

Le taux de précision pour les femmes est de **95%** et celui des hommes de **100%** et d'après ces résultats, nous remarquons que la méthode énergie donne un meilleur résultat pour les hommes que pour les femmes.

Interprétation: nous avons déduit dans la deuxième répartition que le meilleur taux de précision des trois méthodes, est celui de la méthode mfcc.

3.7.1.2. Comparaison des méthodes pour chaque répartition

L'histogramme de la figure 3.3 représente la comparaison du taux de précision total des trois méthodes (covariance, énergie, mfcc) pour chaque répartition.

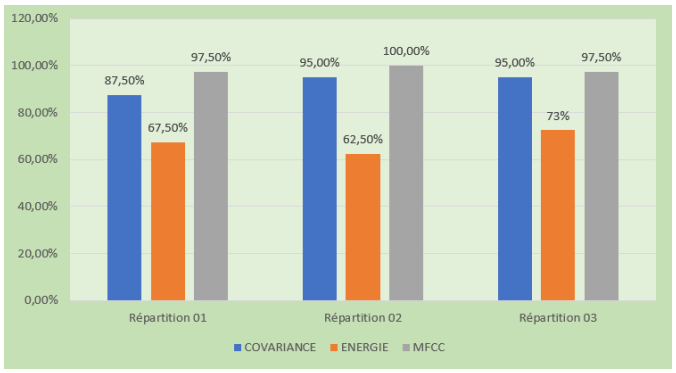

Figure 3.3 : Histogramme comparatif des méthodes pour chaque répartition.

Nous avons constaté que dans ce cas la méthode qui donne de meilleur résultat pour les trois répartitions est celle de **mfcc** avec un taux de précision de **100 %** dans la **répartition 02.**

3.7.1.3. Comparaison des répartitions pour chaque méthode

L'histogramme de la figure 3.4 représente la comparaison des trois répartitions pour chaque taux de précision de chaque méthode

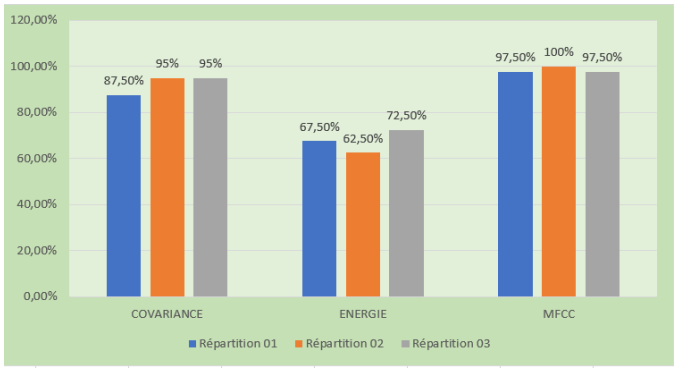

Figure 3.4 : Histogramme comparatif des répartitions pour chaque méthode.

Nous avons constaté que dans ce cas la répartition qui donne le meilleur résultat est la **répartition 02** avec un taux de précision de **100%** dans la méthode **mfcc**.

3.7.2. Combinaison de deux méthodes

Dans cette partie, nous avons utilisé une combinaison de deux méthodes, plusieurs pourcentages ont été attribué à chacune d'elle pour un meilleur taux de précision.

Les seuils de détection utilisés ont été calculés dans la première partie (une seule méthode), chaque répartition a ses méthodes et chaque méthode son propre seuil de détection.

Un score a été utilisé pour détecter le genre du locuteur (homme ou femme).

3.7.2.1. Résultats des répartitions de chaque méthode

Le programme du système de reconnaissance de chaque méthode combinée est présenté ci-dessous Les mêmes programmes utilisés dans les premières approches sont utilisés dans celle-ci sauf qu'ils sont combinée et le seuil de détection change. On utilise un score pour chaque méthode pour améliorer le taux de précision.

```
86      %Donner des scores et trouver un nombre pour une observation particulière
87
88 -    score=(0.6*(fmax/fundamental_freq_level))+(0.4*(energy/energy_level))
89 -    seuilmarks=mean(marks);
90
91      % si la valeur est supérieure à 1 pour cette observation particulière
92 -    if score>1
93 -        disp('La voix de femme est reconnue')
94
95 -    else
96 -        disp('La voix d homme est reconnue')
97
98 -    end
99 -    end
```

Figure 3.5 : Programme du système de reconnaissance pour la méthode covariance+énergie (seuil de détection)

```
86      %Donner des scores et trouver un nombre pour une observation particulière
87
88 -    score=(0.6*(fmax/fundamental_freq_level))+(0.4*(mfcc_coeff/mfcc_level))
89 -    seuilmarks=mean(marks);
90
91      %  si la valeur est supérieure à 1 pour cette observation particulière
92 -    if score>1
93 -        disp('La voix de femme est reconnue')
94
95 -    else
96 -        disp('La voix d homme est reconnue')
97
98 -    end
99 -    end
```

Figure 3.6 : Programme du système de reconnaissance pour la méthode covariance+mfcc (seuil de détection)

```
%Donner des scores et trouver un nombre pour une observation particulière

score=(0.3*(mfcc_coeff/mfcc_level))+(0.7*(energy/energy_level))
seuilmarks=mean(marks);

%  si la valeur est supérieure à 1 pour cette observation particulière
if score>1
    disp('La voix de femme est reconnue')

else
    disp('La voix d homme est reconnue')

end
end
```

Figure 3.7 : Programme du système de reconnaissance pour la méthode mfcc+énergie (seuil de détection)

3.7.2.1.1. Répartition 01

Les seuils de détection de chaque méthode utilisée dans cette répartition sont :

$$\boldsymbol{Seuil}_{energy} = -\,0.6365 \quad ; \quad \boldsymbol{Seuil}_{mfcc} = -\,0.3847 \quad ; \quad \boldsymbol{f}_{max} = 165$$

a) COVARIANCE + ENERGIE

Le tableau 3.13 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **82.5%**

Score		Taux de Précision (%)		
COVARIANCE	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{31}{40} \times 100 = \mathbf{77.5\%}$
0.3	0.7	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{14}{20} \times 100 = \mathbf{70\%}$	$\frac{30}{40} \times 100 = \mathbf{75\%}$
0.7	0.3	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{33}{40} \times 100 = \mathbf{82.5\%}$
0.4	0.6	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{30}{40} \times 100 = \mathbf{75\%}$
0.6	0.4	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{32}{40} \times 100 = \mathbf{80\%}$

Tableau 3.13 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Energie

Cette combinaison donne de meilleurs résultats pour les femmes que pour les hommes.

Le meilleur taux de précision est celui des scores 0.7 (cov) et 0.3 (eng).

Le taux de précision pour les femmes est de **90%** et celui des hommes de **75%**

b) COVARIANCE + MFCC

Le tableau 3.14 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **95%**

Score		Taux de Précision (%)		
COVARIANCE	**MFCC**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{38}}{\mathbf{40}} \times \mathbf{100 = 95\%}$
0.3	0.7	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{37}}{\mathbf{40}} \times \mathbf{100 = 92.5\%}$
0.7	0.3	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{\mathbf{37}}{\mathbf{40}} \times \mathbf{100 = 92.5\%}$
0.4	0.6	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{19}{20} \times 100 = 95\%$	$\frac{\mathbf{38}}{\mathbf{40}} \times \mathbf{100 = 95\%}$
0.6	0.4	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{18}{20} \times 100 = 90\%$	$\frac{\mathbf{37}}{\mathbf{40}} \times \mathbf{100 = 92.5\%}$

Tableau 3.14 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Mfcc

Cette combinaison donne de meilleurs résultats pour les femmes que pour les hommes.

Les meilleurs taux de précision sont ceux des scores 0.5 (cov) , 0.5 (mfcc) et 0.4 (cov) , 0.6 (mfcc).

Le taux de précision pour les femmes est de **95%** et celui des hommes de **95%**.

c) MFCC + ENERGIE

Le tableau 3.15 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **97.5%**

Score		Taux de Précision (%)		
MFCC	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{20}{20} * 100 = \mathbf{100\%}$	$\frac{17}{20} * 100 = \mathbf{85\%}$	$\frac{\mathbf{37}}{\mathbf{40}} \times \mathbf{100 = 92.5\%}$
0.3	0.7	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{\mathbf{33}}{\mathbf{40}} \times \mathbf{100 = 82.5\%}$
0.7	0.3	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100 = 97.5\%}$
0.4	0.6	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{\mathbf{34}}{\mathbf{40}} \times \mathbf{100 = 85\%}$
0.6	0.4	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{\mathbf{38}}{\mathbf{40}} \times \mathbf{100 = 95\%}$

Tableau 3.15 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Mfcc et Energie

Cette combinaison donne de meilleurs résultats pour les femmes que pour les hommes.

Le meilleur taux de précision est celui des scores 0.7 (mfcc) et 0.3 (eng).

Le taux de précision pour les femmes est de **100%** et celui des hommes de **95%**.

3.7.2.1.2 Répartition 02

Les seuils de détection de chaque méthode utilisée dans cette répartition sont :

$$Seuil_{energy} = -0.7439 \quad ; \quad Seuil_{mfcc} = -0.3775 \quad ; \quad f_{max} = 165$$

a) COVARIANCE + ENERGIE

Le tableau 3.16 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **95%**

Score		Taux de Précision (%)		
COVARIANCE	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{14}{20} \times 100 = \mathbf{70\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{32}{40} \times 100 = \mathbf{80\%}$
0.3	0.7	$\frac{12}{20} \times 100 = \mathbf{60\%}$	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{29}{40} \times 100 = \mathbf{72.5\%}$
0.7	0.3	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{38}{40} \times 100 = \mathbf{95\%}$
0.4	0.6	$\frac{14}{20} \times 100 = \mathbf{70\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{32}{40} \times 100 = \mathbf{80\%}$
0.6	0.4	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{36}{40} \times 100 = \mathbf{90\%}$

Tableau 3.16 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Energie

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Le meilleur taux de précision est celui des scores 0.7 (cov) et 0.3 (eng).

Le taux de précision pour les femmes est de **90%** et celui des hommes de **100%**.

b) COVARIANCE + MFCC

Le tableau 3.17 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **100%.**

Score		Taux de Précision (%)		
COVARIANCE	**MFCC**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{39}{40} \times 100 = \mathbf{97.5\%}$
0.3	0.7	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{40}{40} \times 100 = \mathbf{100\%}$
0.7	0.3	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{39}{40} \times 100 = \mathbf{97.5\%}$
0.4	0.6	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{40}{40} \times 100 = \mathbf{100\%}$
0.6	0.4	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{39}{40} \times 100 = \mathbf{97.5\%}$

Tableau 3.17 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Mfcc

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Les meilleurs taux de précision sont ceux des scores 0.3 (cov) , 0.7 (mfcc) et 0.4 (cov) , 0.6 (mfcc)

Le taux de précision pour les femmes est de **100%** et celui des hommes de **100%**.

c) MFCC + ENERGIE

Le tableau 3.18 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **100%.**

Score		Taux de Précision (%)		
MFCC	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$
0.3	0.7	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{34}}{\mathbf{40}} \times \mathbf{100} = \mathbf{85\%}$
0.7	0.3	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$
0.4	0.6	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{37}}{\mathbf{40}} \times \mathbf{100} = \mathbf{92.5\%}$
0.6	0.4	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$

Tableau 3.18 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Mfcc et Energie

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Les meilleurs taux de précision sont ceux des scores 0.5 (mfcc) , 0.5 (eng) et 0.7 (mfcc) , 0.3 (eng) et 0.6 (mfcc) , 0.4 (eng)

Le taux de précision pour les femmes est de **100%** et celui des hommes de **100%**.

3.7.2.1.3 Répartition 03

Les seuils de détection de chaque méthode utilisée dans cette répartition sont :

$$\boldsymbol{Seuil_{energy}} = -\,0.7349\,; \quad \boldsymbol{Seuil_{mfcc}} = -\,0.3895\,; \quad \boldsymbol{f_{max}} = 165$$

a) COVARIANCE + ENERGIE

Le tableau 3.19 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **95%**

Score		Taux de Précision (%)		
COVARIANCE	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{\mathbf{34}}{\mathbf{40}} \times \mathbf{100} = \mathbf{85\%}$
0.3	0.7	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{\mathbf{32}}{\mathbf{40}} \times \mathbf{100} = \mathbf{80\%}$
0.7	0.3	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{38}}{\mathbf{40}} \times \mathbf{100} = \mathbf{95\%}$
0.4	0.6	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{\mathbf{33}}{\mathbf{20}} \times \mathbf{100} = \mathbf{82.5\%}$
0.6	0.4	$\frac{16}{20} \times 100 = \mathbf{80\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{36}}{\mathbf{40}} \times \mathbf{100} = \mathbf{90\%}$

Tableau 3.19 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Energie

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Les meilleurs taux de précision sont ceux des scores 0.7 (cov) , 0.3 (eng) .

Le taux de précision pour les femmes est de **90%** et celui des hommes de **100%**.

b) COVARIANCE + MFCC

Le tableau 3.20 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **97.5%**

Score		Taux de Précision (%)		
COVARIANCE	**MFCC**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100} = \mathbf{97.5\%}$
0.3	0.7	$\frac{18}{20} \times 100 = \mathbf{90\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{38}}{\mathbf{40}} \times \mathbf{100} = \mathbf{95\%}$
0.7	0.3	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100} = \mathbf{97.2\%}$
0.4	0.6	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100} = \mathbf{97.2\%}$
0.6	0.4	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100} = \mathbf{97.2\%}$

Tableau 3.20 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Mfcc

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Les meilleurs taux de précision sont ceux des scores 0.5 (cov) , 0.5 (mfcc) et 0.7 (cov) , 0.3 (mfcc) et 0.6 (cov) , 0.4 (mfcc) et 0.6 (cov) , 0.4 (mfcc).

Le taux de précision pour les femmes est de **95%** et celui des hommes de **100%**.

c) MFCC + ENERGIE

Le tableau 3.21 représente le taux de précision total des différents scores (Femmes + Hommes), le meilleur taux de précisions est de **100%.**

Score		Taux de Précision (%)		
MFCC	**ENERGIE**	**Femmes**	**Hommes**	**Total**
0.5	0.5	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{39}}{\mathbf{40}} \times \mathbf{100} = \mathbf{97.5\%}$
0.3	0.7	$\frac{15}{20} \times 100 = \mathbf{75\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{34}}{\mathbf{40}} \times \mathbf{100} = \mathbf{85\%}$
0.7	0.3	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$
0.4	0.6	$\frac{17}{20} \times 100 = \mathbf{85\%}$	$\frac{19}{20} \times 100 = \mathbf{95\%}$	$\frac{\mathbf{36}}{\mathbf{40}} \times \mathbf{100} = \mathbf{90\%}$
0.6	0.4	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{20}{20} \times 100 = \mathbf{100\%}$	$\frac{\mathbf{40}}{\mathbf{40}} \times \mathbf{100} = \mathbf{100\%}$

Tableau 3.21 : Taux de précision des deux genre (femmes + hommes) avec la méthode de Covariance et Mfcc

Cette combinaison donne de meilleurs résultats pour les hommes que pour les femmes.

Les meilleurs taux de précision sont ceux des scores 0.7 (mfcc) , 0.3 (eng) et 0.6 (mfcc) , 0.4 (eng) .

Le taux de précision pour les femmes est de **100%** et celui des hommes de **100%**.

3.7.2.2. Comparaison des méthodes pour chaque répartition

3.7.2.2.1. Score 0.5 , 0.5

L'histogramme de la figure 3.8 représente la comparaison du taux de précision total des trois méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) pour chaque répartition avec un score de 0.5 , 0.5.

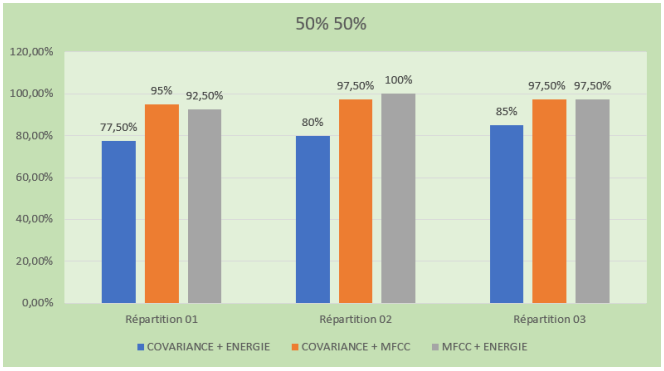

Figure 3.8 : Histogramme comparatif des méthodes pour chaque répartition, score (0.5 , 0.5)

Nous avons constaté que dans ce cas la combinaison qui donne de meilleur résultat pour les trois répartitions est celle de **mfcc + energie** avec un taux de précision de **100 %** dans la **répartition 02**.

3.7.2.2.2. Score 0.3 , 0.7

L'histogramme de la figure 3.9 représente la comparaison du taux de précision total des trois méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) pour chaque répartition avec un score de 0.3 , 0.7.

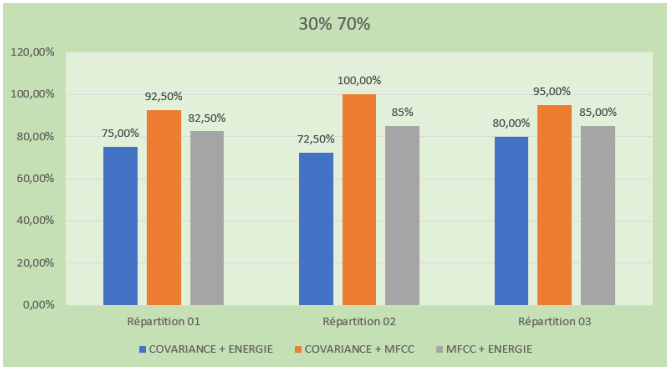

Figure 3.9 : Histogramme comparatif des méthodes pour chaque répartition, score (0.3 , 0.7)

Nous avons constaté que dans ce cas la combinaison qui donne de meilleur résultat pour les trois répartitions est celle de **mfcc + energie** avec un taux de précision de **100 %** dans la **répartition 02**

3.7.2.2.3. Score 0.7 , 0.3

L'histogramme de la figure 3.10 représente la comparaison du taux de précision total des trois méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) pour chaque répartition avec un score de 0.7 , 0.3.

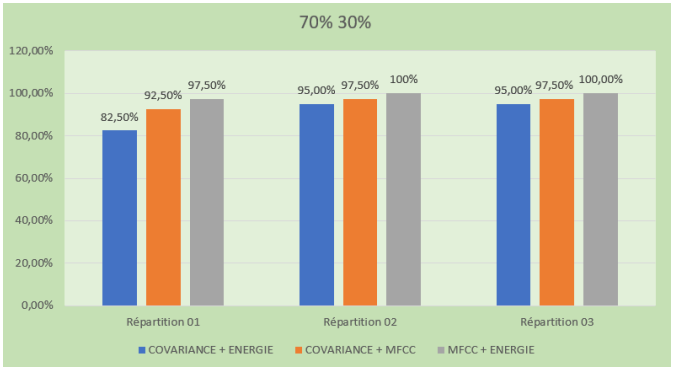

Figure 3.10 : Histogramme comparatif des méthodes pour chaque répartition, score (0.7 , 0.3)

Nous avons constaté que dans ce cas la combinaison qui donne de meilleur résultat pour les trois répartitions est celle de **mfcc + energie** avec un taux de précision de **100 %** dans la **répartition 02** et la **répartition 03.**

3.7.2.2.4. Score 0.4 , 0.6

L'histogramme de la figure 3.11 représente la comparaison du taux de précision total des trois méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) pour chaque répartition avec un score de 0.4 , 0.6.

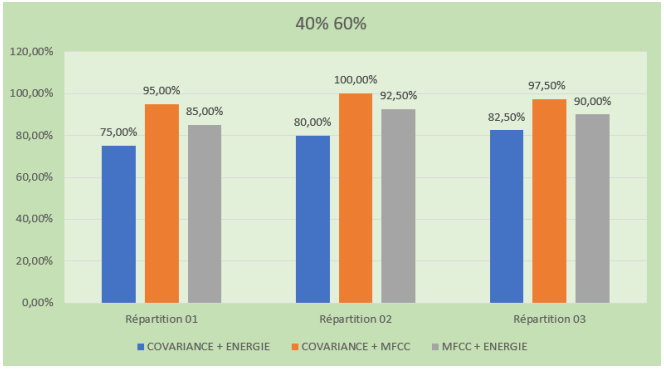

Figure 3.11 : Histogramme comparatif des méthodes pour chaque répartition, score (0.4 , 0.6).

Nous avons constaté que dans ce cas la combinaison qui donne de meilleur résultat pour les trois répartitions est celle de **covariance + mfcc** avec un taux de précision de **100 %** dans la **répartition 02**.

3.7.2.2.5. Score 0.6 , 0.4

L'histogramme de la figure 3.12 représente la comparaison du taux de précision total des trois méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) pour chaque répartition avec un score de 0.6 , 0.4.

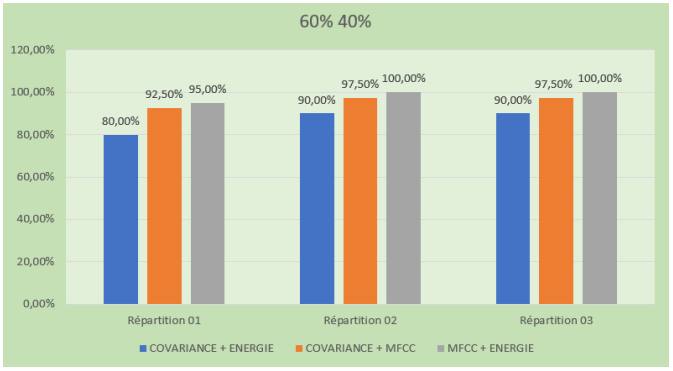

Figure 3.12 : Histogramme comparatif des méthodes pour chaque répartition, score (0.6 , 0.4)

Nous avons constaté que dans ce cas la combinaison qui donne de meilleur résultat pour les trois répartitions est celle de **covariance + mfcc** avec un taux de précision de **100 %** dans la **répartition 02**.

3.7.2.3. Comparaison des répartitions pour chaque méthode

3.7.2.3.1. Score 0.5 , 0.5

L'histogramme de la figure 3.13 représente la comparaison des trois répartitions pour chaque taux de précision des méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) avec un score de 0.5 , 0.5.

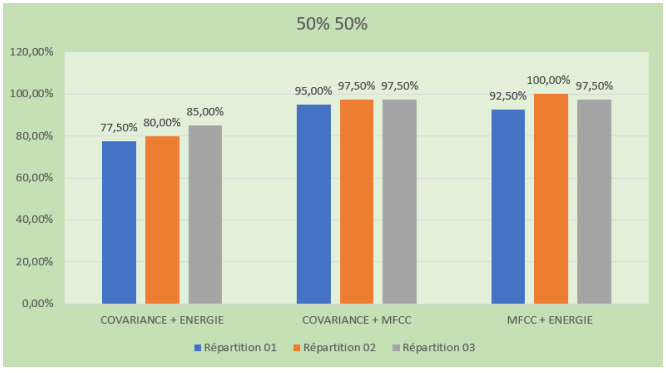

Figure 3.13 : Histogramme comparatif des répartitions pour chaque méthode, score (0.5 , 0.5).

Nous avons constaté que dans ce cas la répartition qui donne de meilleur résultat parmi les trois est la **répartition 02** avec un taux de précision de **100%** dans la méthode mfcc + énegie.

3.7.2.3.2. Score 0.3 , 0.7

L'histogramme de la figure 3.14 représente la comparaison des trois répartitions pour chaque taux de précision des méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) avec un score de 0.3 , 0.7.

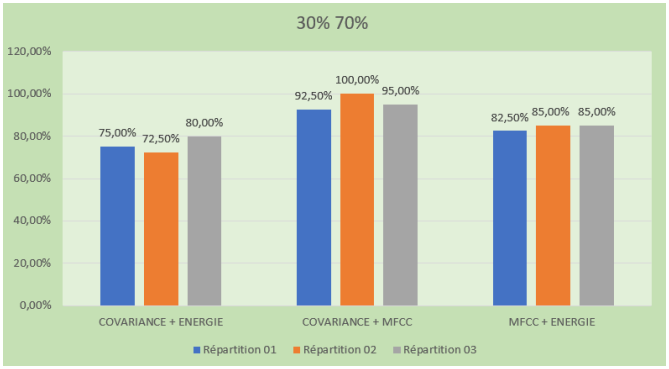

Figure 3.14 : Histogramme comparatif des répartitions pour chaque méthode, score (0.3 , 0.7)

Nous avons constaté que dans ce cas la répartition qui donne de meilleur résultat parmi les trois est la **répartition 02** avec un taux de précision de **100%** dans la méthode covariance + mfcc.

3.7.2.3.3. Score 0.7 , 0.3

L'histogramme de la figure 3.15 représente la comparaison des trois répartitions pour chaque taux de précision des méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) avec un score de 0.7 , 0.3.

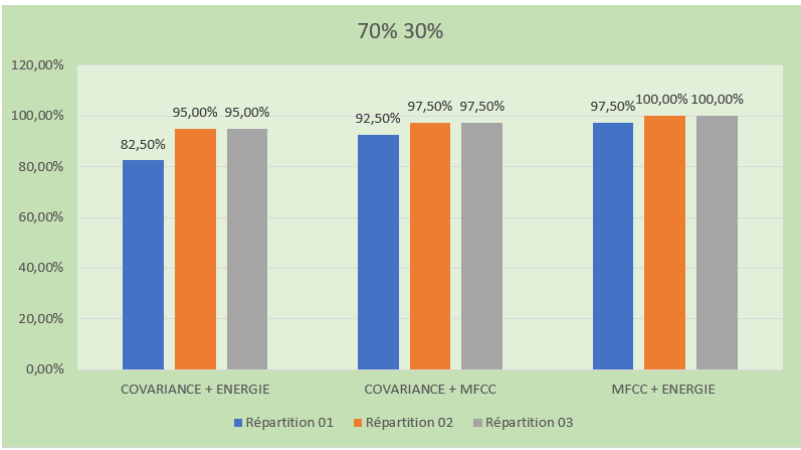

Figure 3.15 : Histogramme comparatif des répartitions pour chaque méthode, score (0.7 , 0.3)

Nous avons constaté que dans ce cas la répartition qui donne de meilleur résultat parmi les trois est la **répartition 02** et la **répartition 03** avec un taux de précision de **100%** dans la méthode **mfcc + énergie**.

3.7.2.3.4. Score 0.4 , 0.6

L'histogramme de la figure 3.16 représente la comparaison des trois répartitions pour chaque taux de précision des méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) avec un score de 0.4 , 0.6.

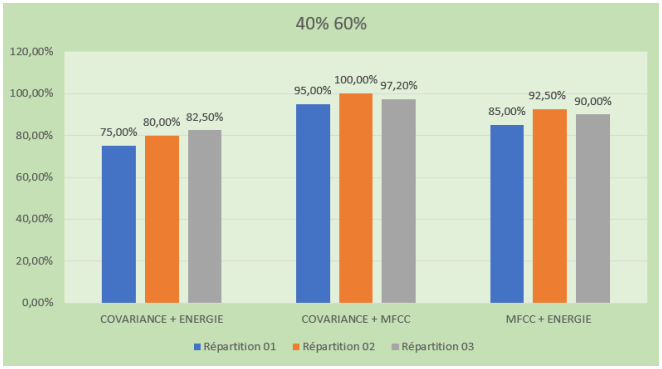

Figure 3.16 : Histogramme comparatif des répartitions pour chaque méthode, score (0.4 , 0.6)

Nous avons constaté que dans ce cas la répartition qui donne de meilleur résultat parmi les trois est la **répartition 02** avec un taux de précision de **100%** dans la méthode **covariance + mfcc**.

3.7.2.3.5. Score 0.6 , 0.4

L'histogramme de la figure 3.17 représente la comparaison des trois répartitions pour chaque taux de précision des méthodes (covariance + énergie ; covariance + mfcc ; mfcc + énergie) avec un score de 0.6 , 0.4.

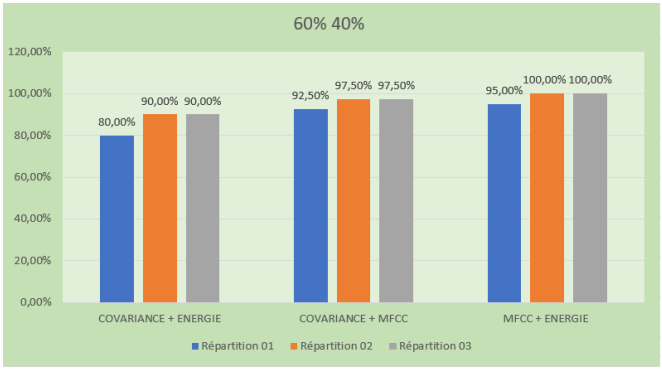

Figure 3.17 : Histogramme comparatif des répartitions pour chaque méthode, score (0.6 , 0.4)

Nous avons constaté que dans ce cas la répartition qui donne de meilleur résultat parmi les trois est la **répartition 02** avec un taux de précision de **100%** dans la méthode **covariance + mfcc**.

3.8. Conclusion

Dans ce chapitre, une méthodologie a été élaborée pour l'identification automatique du genre du locuteur. Cette méthodologie se résume à classer la base de données en trois répartitions et chacune de ces dernières dispose de deux parties, l'une utilise des méthodes seules et l'autre la combinaison de deux méthodes.

Les résultats obtenus montrent l'efficacité de la combinaison des méthodes pour un meilleur taux de précision.

La méthode avec le taux de précision le plus élevé est la combinaison de la méthode de Mfcc et énergie avec une précision allons jusqu'à 100% à condition que le score du Mfcc soit plus grand que celui de l'énergie.

Le taux de précision varie selon les répartitions et nous avons trouvé que la répartition 02 est la meilleure.

Conclusion générale

Cette thèse s'inscrit dans le domaine du traitement de la parole, nous nous sommes intéressés à la reconnaissance automatique du genre du locuteur et pour cela nous avons donné un aperçu du domaine de la sécurité en exposant la biométrie tout en mettant l'accent sur la biométrie vocale et le système de production de la parole est aussi passé en revue, ses caractéristiques, son acquisition et son analyse par les différentes méthodes temporelles et fréquentielles.

Des tests de reconnaissance ont été effectués sur la base de données TIMIT, cette dernière a été classée en trois répartitions et chacune d'elle dispose de deux parties, l'une utilise des méthodes seules et l'autre deux méthodes combinées.

Dans la première partie, nous avons choisi trois méthodes

- Covariance
- Energie
- Mfcc

Dans la seconde, nous avons combiné deux méthodes des trois citées ci-dessus

- Covariance + Energie
- Covariance + Mfcc
- Mfcc + Energie

A partir des résultats obtenus dans les deux parties, nous avons constaté une meilleure performance pour les méthodes combinées.
Les résultats de reconnaissance acquis, montrent clairement que la combinaison de la méthode mfcc avec une autre, est de loin meilleure.

Ce projet nous ouvre une porte vers de nouvelles technologies et vers de nouvelles méthodes, le monde de la biométrie est bien vaste et ne cesse de s'étendre, et quel que soit le niveau d'efficacité atteint par notre application il y'aura toujours un moyen de l'améliorer.

Références bibliographiques

[01] Fumi, Sadaoki. 1989. Digital Speech Processing, Synthesis. and Récognition. 1ère éd. Marcel Dekker lue, 390p.

[02] M. F. Clemente Giorio, Kinect in Motion - Audio and Visual Tracking by Example, Packt Publishing, 2013.

[03] Teva MERLIN .Amiral, une plateforme générique pour la reconnaissance automatique du locuteur de l'authentification à l'indexation. (thèse) Académie d'aix-Marseille université d'Avignon et des Pays de Vaucluse. 18 novembre 2004.

[04] « La voix humaine » www.musimem.com (consulté le 29/05/2022)

[05] S. Nefti, « Segmentation automatique de parole en phones », Doctorat, Université de Rennes 1, Rennes, France, 2004

[06] M.Pelletier, «La voix, support de la relation en psychomotricité», Diplôme d'état en psychomotricité, Institut de formation en psychomotricité, Université Bordeaux, France, 2020

[07] H.Knoerr, « Le mécanisme phonatoire », Cours de l'université d'Ottawa, Canada, 2011

[08] A.Ghio, « Modélisation du conduit vocal », Laboratoire Parole et Langage, Université de Provence, France, 2017

[09] D.Charline, « Cancer du larynx », https://www.sante-sur-le net.com/maladies/cancer/cancer-du-larynx/ (consulté le 18/02/2022)

[10] « Pathologies des cordes vocales-dysphonie », https://orl.nc/pathologies-du-cou/pathologie-des-cordes-vocales/ (consulté le 18/02/2022)

[11] N. Sturmel, « Analyse de la qualité vocale appliquée à la parole expressive », Ecole Doctorale, Université Paris-Sud 11, Paris, France, 2011.

[12] M.VIENNOT, « A propos d'une analyse objective de la voix de 40 sujets présentant des troubles musculo-squelettiques », UHP Nancy - Certificat de Capacité d'Orthophonie 2010

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

[14]Damien Vincent. Thèse« Analyse et contrôle du signal glottique en synthèse de la parole» l'École Nationale Supérieure des Télécommunications de Bretagne 2007.

[15] S. Nefti, « Segmentation automatique de parole en phones », Doctorat, Université de Rennes 1, Rennes, France, 2004.

[16] Zwicker, E., Feldtkeller, R., Psychoacoustique, CENT/ENST, collection technique et scientifique des télécommunications, Mason Paris, 1981.

[17] Reynolds, D. A., Experimental evaluation of features for robust speaker identification, IEEE transactions Speech Audio Processing, volume 2, pages 639-643, 1994.

[18] Homayounpour, M. M., Chollet, G., Performance comparison of some relevant spectral representations for speaker verification, Workshop on Automatic Speaker Recognition, Identification, Verification, pages 27-30, Martigny (Suisse), Avril 1994.

[19] http://outilsrecherche.over-blog.com/pages/Notes_111_Le_Systeme_Auditif_Humain-3080878.html (consulté le 09/03/2022)

[20] L. Buniet, « Traitement automatique de la parole en milieu bruité », Doctorat, Université Henri Poincaré - Nancy 1, France, 1997.

[21] « Le fonctionnement de l'oreille humaine », https://www.cotral.fr/blog/prevention-risques-auditifs/le-fonctionnement-de-l-oreille-humaine.html (consulté le 09/03/2022)

[22] Laboratoire Unisson « Le fonctionnement de votre oreille », https://www.laboratoires-unisson.com/fonctionnement-systeme-auditif.html (consulté le 09/03/2022)

[23] https://www.uvex-safety.com/blog/fr/comment-fonctionne-systeme-auditif/ (consulté le 09/03/2022)

[24] https://www.ideal-audition.fr/audition (consulté le 09/03/2022)

[25] Flanagan, J.L. 1972. Speech Analysis, Synthesis, and Pei-ception. 2"™ éd.. New York : Springer-Verlag.

[26] Charlet, D., Authentification vocale par téléphone en mode dépendant du texte, Thèse de l'Ecole Nationale Supérieure des Télécommunications, 1997.

[27] Atal, B. S., Automatic recognition of speakers from their voices, IEEE transactions, volume 64 (4), pages 460-475, 1976.

[28] Doddington, G. R., Speaker recognition Identifying people by their voices, IEEE transactions, volume. 73(11), pages 1651-1664, 1985.

[29] Rosenberg, A. E., Soong, F. K., Recent research in automatic speaker recognition, Advances in speech signal processing, 1991.

[30] JOUSSE, Vincent. Identification nommée du locuteur: exploitation conjointe du signal sonore et de sa transcription. 2011. Thèse de doctorat. Université du Maine.

[31] Sambur, 1975. Selection of acoustic features for speaker identification. IEEE Transactions on Acoustics, Speech, and Signal Processing 23(2), 176–182.

[32] J. Bonastre & H. Meloni, 1992. A study of spectral variability for speaker characterisation. 19èmes Journées d'Etudes sur la Parole 555.

[33] http://culturesciencesphysique.ens-lyon.fr/prepublication/db/csphysique/data/principe-numerisation.xml (consulté le 16/03/2022)

[34] M. Didiche, «Modélisation neuro-prédictive pour La classification phonétique», Thèse Doctorat, Université Mohamed Khider, Biskra, Algérie, 2014.

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

[36] C.-S. Gargour, Traitement numérique des signaux, École de technologie supérieure, 2001.

[37] B. Bouseksou « Reconnaissance automatique de la parole par la méthode globale ». Thèse de magister option électronique acoustique et physiologique de la Parole. Institut de linguistique et de phonétique, Alger,1983.

[38] T. Parsons, Voice and Speech Processing, McGraw-Hill, 1986.

[39] T. Dutoit, Introduction au traitement automatique de la parole, Faculté polytechnique de Mons, Belgique, 2000.

[40] M. Joseph « Analyse, synthèse et codage de la parole tome 1 et 2 ». Lavoisier-2002

[41] J. Dumas, « L'analyse temps – fréquence », (Document réalisé par: Groupe MVI technologies), limonest Lyon, Version février 2001.

[42] E. Wong and S. Sridharan, "Comparison oflinear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification," presented at Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of2001 International Symposium on, 2001.

[43] Calliope, « *La parole et son traitement automatique* », édition Masson, 1999.

[44] G. Mahmoud, La Parametrisation Mfcc En Vue D'Une Reconnaissance Robuste de Parole, 2015.

[45] K. Dash, A Novel Bpnn Approach for Speaker Identification Using Mfcc, 2012

[46] S.Mallat, "*A theory for multiresolution signal decomposition: the wavelet representation"*, IEEE Transactions on Pattern Analysis and machine Intelligence, 11(7), 1989, pp.674-693.

[47] T.AL ANI. '' Introduction aux ondelettes, Deuxième partie : Quelques concepts Généraux de la théorie des ondelettes. '' Département Informatique ESIEE-Paris.

[48] A.AMRANE, '' Détection de l'onde P de L'électrocardiogramme Par des algorithmes basés sur la transformée en Ondelette et modèle Markov caché '' Thèse magistère en électronique Université de Skikda.

[49] Garofolo, John S., et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

[50] http://www.xavierdupre.fr/app/mlstatpy/helpsphinx/notebooks/split_train_test.html (consulté le 26/06/2022).

Résumé :

La reconnaissance vocale a de nombreuses applications, notamment l'interaction homme-machine, le tri des appels téléphoniques en fonction du genre, la catégorisation des vidéos avec étiquetage, etc. Actuellement, l'apprentissage automatique est une tendance populaire qui a été largement utilisée dans divers domaines et applications, en exploitant le développement récent des technologies numériques et l'avantage des capacités de stockage des médias électroniques.

Récemment, la recherche s'est concentrée sur la combinaison de techniques d'apprentissage d'ensemble afin de construire des classificateurs plus précis.

Dans cette étude, nous nous concentrons sur la reconnaissance du genre par la voix en utilisant des algorithmes (Covariance, Energie, Mfcc) pris indépendamment, puis un nouvel algorithme d'ensemble sera fait pour démontrer son efficacité en termes de précision.

Mots clés : algorithme, apprentissage d'ensemble, covariance, énergie, mfcc, reconnaissance du genre.

Abstract :

Speech recognition has various applications including human to machine interaction, sorting of telephone calls by gender categorization, video categorization with tagging and so on. Currently, machine learning is a popular trend which has been widely utilized in various fields and applications, exploiting the recent development in digital technologies and the advantage of storage capabilities from electronic media.

Recently, research focuses on the combination of ensemble learning techniques to build more accurate classifiers.

In this study, we focus on voice-based gender recognition using independently taken algorithms (covariance, energy, Mfcc), then a new ensemble algorithm will be made to demonstrate its effectiveness in terms of accuracy.

Keywords: algorithm, ensemble learning, covariance, energy, mfcc, gender recognition.

ملخص :

تتضمن ميزة التعرف على الصوت العديد من التطبيقات، بما في ذلك التفاعل بين الإنسان والآلة، وفرز المكالمات الهاتفية حسب الجنس، وتصنيف مقاطع الفيديو مع وضع العلامات، وما إلى ذلك. وفي الوقت الحالي، يعد تعلم الآلة اتجاها شائعا تم استخدامه على نطاق واسع في مختلف المجالات والتطبيقات، مستغلا التطور الحديث للتقنيات الرقمية وميزة إمكانيات تخزين الوسائط الإلكترونية.

مؤخرا، ركزت الأبحاث على الجمع بين تقنيات التعلم لإنشاء مزيد من المصنفين الأكثر دقة.

في هذه الدراسة، نركز على التعرف على نوع الجنس عن طريق الصوت باستخدام الخوارزميات (التباين المشترك، الطاقة، م.س.ل.م) التي يتم أخذها بشكل مستقل، ثم سيتم إنشاء خوارزمية مجموعة جديدة لتوضيح فعاليتها من حيث الدقة.

الكلمات المفتاحية: الخوارزمية، التعلم العام، التباين المشترك، الطاقة ، م.س.ل.م ، الاعتراف بالفوارق بين الجنسين.