REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj

Faculté *des Sciences et de la technologie*

Département d'Electronique

Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

FILIÈRE : **Télécommunications**

Spécialité : Système des Télécommunications

Par

- **BENBOUZID YACINE**
- **BENKECHIDA LILIA**

Intitulé

Reconnaissance acoustique du genre humain basée sur les modèles GMM

Soutenu le : 26-06-2024

Devant le Jury composé de :

Nom & Prénom	*Grade*	*Qualité*	*Etablissement*
Mme. MEGUELLATI SABRINA	*MAA*	*Président*	*Univ-BBA*
Mr. HACINE GHARBI ABDENOUR	*MCA*	*Encadreur*	*Univ-BBA*
Mme. MESSALI ZOUBEIDA	*PROF*	*Examinateur*	*Univ-BBA*

Année Universitaire 2023/2024

Remerciement

Tout d'abord, nous tenons à remercier "ALLAH", le Tout-Puissant pour nous avoir donné la santé, la volonté, le courage et la patience pour mener à terme notre formation et pouvoir réaliser ce travail.

Nos remerciements s'adressent particulièrement au ***Dr. HACINE GHARBI ABDENOUR*** *pour son encadrement de qualité, sa motivation professionnelle, ses conseils et ses critiques constructives, ses corrections, sa gentillesse et sa patience, pour le temps qu'il nous a consacré pour la réalisation de notre travail.*

Nous remercions également les membres du jury pour leurs présences, leurs lectures attentives de ce mémoire, ainsi que pour les remarques qu'ils nous adresseront lors de cette soutenance afin d'améliorer notre travail.

Enfin, nous adressons nos plus sincères remerciements à nos familles, nos amis proches et nos collègues, qui nous ont accompagnés, aidés, soutenus et encouragés tout au long de la réalisation de ce mémoire.

Dédicace

« Louange à Allah, le seul et unique »

«اهدي تخرجي الى ذلك الرجل العظيم جدي رحمه الله لطالما تمنيت ان يقر عينه برؤيتي في يوم كهذا فرحتي ينقصها وجودك ونجاحي ينقصه فخرك بي »

Á mes chers parents aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être.

Je vous remercie pour tous le soutien et l'amour que vous me portez depuis mon enfance, Tout ce que j'espère, c'est que vous soyez fiers de moi aujourd'hui.

Mama,reposez votrecœur.Votre rêve et devenu réalité.Enfinvotre petite fille est devenue diplômée comme que tu as tant imaginée et désirée.

À ma grand sœur IMANE et son mari MAHDI pour ses soutiens moral et leurs conseils.et à mes chers frères, Abdelkader et Oussama et Chère tante MAROUA.

Á mes petit anges RINAD, FARIDE, LINA ET AHMED.

Á Mon âme sœur CHANEZ qui a été ma confidente mon pilier dans les moments difficiles. sa présence à mes côtes a rendu ce parcours plus doux.Et ma cousine manel pour sa amour.

Je dédie entièrement ce travail pour mon binôme BENBOUZID YACINE, vous avez toujours été là pour m'écouter, échanger des idées et apporter votre expertise. Votre sérieux et votre gentillesse font de vous une paire exceptionnelle. Et je suis fier du travail que nous avons accompli ensemble.

BENKECHIDA LILIA

Dédicace

« Louange à Allah, le seul et unique »

اهدي تخرجي الى اعز انسان والاقرب لقلبي "سارة" اختي الغالية رحمك الله، حلمك تحقق ويناديك يا اجمل خريجة.

ابتسمي ف لم انساك يوما في مشواري دراسي، كنتي ومازلتي سندي الاول، شكرا لكي على دعمي رحمك الله.

الى ملاكي وقرة عيني، وقوتي بعد الله، داعمتي الاولى والابدية" امي "اهديك هذا الانجاز الذي لولا تضحياتك لما كان له وجود ممتن لان الله قد أكرمني بك يا خير سند.

الى" أبي" الى من دعمني بلا حدود، واعطاني بلا مقابل، الى من احمل اسمه بكل فخر وعزة وشرف شكرا لك ودمت ابا و اخ وبارك الله فيك.

إلى أختي الكبيرة" منيرة "شكرا لكي على دعمك الدائم ونصائحك لي في أصعب الظروف شكرا.

الى أخواتي" حنان وايمان"وإلى ملائكتي الصغار «جواد و دينا "أشكركم على كل الدعم والحب الذي قدمتماه لي لقد كنتم أسراري وسندي في الأوقات الصعبة.

أهدي هذا العمل الى الخال الوحيد رحمه الله "عمار" و الى كل من علمني كلمة في مشواري الدراسي

أهدي هذا العمل بالكامل إلى زميلتي" بن كشيدة ليليا" التي كانتَ دائماً موجودةً للاستماع إليّ وتبادل الأفكار والمساهمة بخبرتها .شكرا لجديتك ولطفك شكرا لكي كثيرا كثيرا. وأنا فخور بالعمل الذي أنجزناه معا.

" يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ"

اللّهم كمّا أنعمّتَ فَزِدْ و كما زدتّ فَبارك و كما باركتَ فَتممّ و كما أتممت فثبت.

BENBOUZID YACINE

Résumé

La reconnaissance Automatique du genre humain est une tâche très importante utilisée dans plusieurs domaines d'application tels que : la sécurité, la surveillance, le marketing et la santé. Cette tâche consiste à reconnaître le genre d'une personne à partir de différents types de modalités telles que : la parole, le visage, le signal vidéo, l'écriture, ...etc. Plus particulièrement, un système de reconnaissance acoustique du genre (RAG) a pour objectif de classifier un signal vocal (parole) en classes de genre (H : masculin, F : féminin).

La conception d'un système RAG se base sur une phase d'apprentissage permettant de modéliser des différentes classes de genre en utilisant une base de données de signaux d'apprentissage, et une phase de test permettant de classifier des signaux appartenant à une base de données de test pour évaluer les performances du système. Plus particulièrement, l'algorithme de classification KNN est couramment utilisé pour cette tâche, vue de sa simplicité et sa facilité d'implémentation, néanmoins il exige plus d'espace mémoire et temps de calcul.

Notre travail consiste à concevoir un système de reconnaissance de genre basée sur l'algorithme de classification GMM (Gaussian Mixture Models) appliqué sur des vecteurs de paramètres MFCC et combiné avec la stratégie de règle de vote. Les résultats de différentes expériences menées nous ont montré que le classificateur GMM est plus performant par rapport au classificateur KNN de points de vue précision (taux de classification) et complexité (espace mémoire et temps de calcul).Plus particulièrement, le classificateur GMM atteint un taux de classification de 100% en utilisant la base de données EMO-DB (Berlin Data base of Emotional speech) en mode indépendant du texte et dépendant du locuteur, alors que le classificateur KNN atteint un taux de classification de 99.61% avec un temps de calcul plus long par rapport au classificateur GMM.

Mots clés : Reconnaissance acoustique du genre, coefficients MFCC, classificateur GMM, classificateur KNN, stratégie de la règle de vote.

Abstract

Automatic recognition of the human gender is a very important task used in several application areas such as: security, surveillance, biometric, marketing and health. This task consists in recognizing the gender of a person from different types of modalities such as:

writing, ...etc. In particular, a gender acoustic recognition system (RAG) aims to classify a vocal signal into gender classes (M: male, F: female).

The design of a RAG system is based on a learning phase allowing to model different gender classes using a database of learning signals, and a test phase to classify signals from a test database to assess system performance. In particular, the KNN classification algorithm is commonly used for this task, given its simple implementation, however it requires more memory space and computation time.

Our work consists in designing a gender recognition system based on the GMM (Gaussian Mixture Models) classification algorithm applied on MFCC features vectors and combined with the voting rule strategy. The results of different experiments have shown us that the GMM classifier is more efficient compared to the KNN classifier in terms of accuracy (classification rate) and complexity (memory space and calculation time). In particular, the GMM classifier achieves a 100% classification rate using the EMO-DB (Berlin Database of Emotional speech) database in text-independent and speaker-dependent mode, whereas the KNN classifier achieves a classification rate of 99.61% with a longer calculation time compared to the GMM classifier.

Keywords: Acoustic gender recognition, MFCC coefficients, GMM classifier, KNN classifier, voting rule strategy.

ملخص:

يعتبر التعرف الالي بالجنس البشري مهمة ذات أهمية للغاية تستخدم في العديد من مجالات التطبيق مثل: الأمن، المراقبة، البيومتري، التسويق والصحة. تتمثل هذه المهمة في التعرف على جنس الشخص انطلاقا من أنواع مختلفة من الطرائق مثل: الكلام، الوجه وإشارة الفيديو والكتابة... إلخ. وبشكل أكثر تحديدًا، يهدف نظام التعرف الصوتي على نوع الجنس إلى تصنيف إشارة صوتية إلى فئات من الجنس (M: ذكر، F: أنثى).

يعتمد تصميم نظام RAG على مرحلة تعلم تسمح بنمذجة الفئات المختلفة من الجنس البشري باستخدام قاعدة بيانات لإشارات التعلم، ومرحلة اختبار لتصنيف الإشارات من قاعدة بيانات الاختبار لتقييم أداء النظام. على وجه الخصوص، يتم استخدام خوارزمية تصنيف KNN بشكل شائع لهذه المهمة، نظرًا لتنفيذها البسيط، ولكنها تتطلب مساحة ذاكرة ووقت حساب أكبر.

يتمثل عملنا في تصميم نظام التعرف الالي بالجنس البشري بناءً على خوارزمية تصنيف GMM المطبقة على ميزات MFCC والمدمجة مع استراتيجية قاعدة التصويت. أظهرت لنا نتائج التجارب المختلفة أن مصنف GMM أكثر كفاءة مقارنة بمصنف KNN من حيث الدقة (معدل التصنيف) والتعقيد (مساحة الذاكرة ووقت الحساب). على وجه الخصوص،

يحقق مصنف GMM معدل تصنيف 100٪ باستخدام قاعدة بيانات EMO-DB (قاعدة بيانات برلين للكلام العاطفي) في وضع مستقل عن النص ومتعلق بالمتحدث، في حين أن مصنف KNN يحقق معدل تصنيف 99.61٪ مع وقت حساب أطول مقارنة بمصنف GMM.

الكلمات المفتاحية: التعرف الصوتي بنوع الجنس البشري، ميزات MFCC، مصنف GMM، مصنف KNN، استراتيجية قاعدة التصويت.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

ANN : Artificial Neural Network.

CNN: Convolution Neural Network.

KNN: k-Nearest Neighbors.

EMO DB: Berlin Data base of Emotional Speech.

FFT: Fast Fourier Transform

GMM: Gaussian Mixtures Model

HMM: Hidden Markov Model.

HTK: Hidden markov models ToolKit.

LPC: Linear Predictive Coding.

LPCC: Linear Prediction Cepstrum Coefficients.

MFCC: Mel-Frequency Cepstral Coefficients.

PLP: Perceptual Linear Prediction.

SVM: Support vector machine.

TCS : Taux de classification des signaux.

TCV : Taux de classification des vecteurs.

I. Introduction générale

Le signal parole est un signal acoustique apportant plusieurs informations telles que la langue, le texte parlé, l'identité du locuteur et son genre, son âge et ses émotions. Plus particulièrement, la Reconnaissance Acoustique du Genre (RAG) humain est une tâche qui trouve ses applications dans plusieurs domaines tels que : la sécurité, la surveillance, la biométrie, le marketing et la santé. L'objectif de la RAG consiste à reconnaître la classe du genre humain (H : Homme ou F : Femme) à partir du signal vocal (parole) en utilisant généralement des outils de la reconnaissance de formes et du traitement du signal.

La conception d'un système RAG se base généralement sur une phase d'apprentissage permettant de modéliser les classes H et F à partir des signaux d'apprentissage, et une phase de test permettant de classifier des signaux de test pour évaluer les performances du système. Chacune des phases exige une étape d'extraction de paramètres permettant de convertir généralement chaque signal en une séquence de vecteurs de paramètres extraits chacun sur une fenêtre d'analyse. Les paramètres acoustiques couramment utilisés pour la reconnaissance vocale sont les coefficients cepstraux MFCC qui jouent un rôle prédominant en raison de leur capacité à modéliser efficacement les propriétés perceptuelles du signal vocal humain. Dans la phase de test, chaque séquence de vecteurs de paramètres correspondant à un signal de test sera classifiée en utilisant un des algorithmes de classification tels que : KNN, ANN, SVM, GMM. Le classificateur KNN est couramment utilisé vue de sa simplicité d'implémentation [1]. Cependant, il exige plus d'espace mémoire dans la phase d'apprentissage et plus de temps de calcul dans l'étape de classification ou de reconnaissance.

L'objectif de notre travail est d'appliquer l'algorithme de classification GMM pour améliorer la complexité des systèmes RAG en termes d'espace mémoire et temps de calcul, par rapport au classificateur KNN. Notre travail consiste à concevoir un système RAG basé sur l'algorithme de classification GMM appliqué sur des séquences de vecteurs de paramètres MFCC, et combiné avec la stratégie de la règle de vote. Une étude comparative sera menée entre les performances du système basé sur le GMM et celles du système basé sur le KNN.

Le manuscrit est structuré en 3 chapitres. Le premier chapitre présente quelques généralités sur la RAG en décrivant brièvement l'architecture d'un système RAG et la technique de classification KNN. Le deuxième chapitre décrit la technique de classification GMM et la méthode d'extraction de paramètres MFCC. Le troisième chapitre décrit le système RAG proposé et présente les étapes d'implémentation du système ainsi que les différentes

expériences et résultats obtenus. Finalement, nous terminons notre travail par une conclusion générale.

Chapitre I

Généralités sur la reconnaissance automatique du genre

I.1 Introduction

Au cours des dernières années, l'évolution du langage naturel a connu des progrès significatifs : le langage parlé a été le premier à évoluer, suivie par le langage écrit. La parole a souvent été considérée comme la clé de l'accès universel à l'information, puisqu'elle est le moyen naturel d'interaction qui ne nécessite pas d'être alphabétisé. Le traitement du signal parole peut être classé généralement en codage de la parole et de l'audio, synthèse texte-parole, reconnaissance de la parole, reconnaissance du locuteur, amélioration de la parole et compréhension du langage parlé. Avec la préoccupation moderne de la sécurité dans le monde entier, la reconnaissance du genre a connu une grande attention de la part des chercheurs sur la parole. La reconnaissance acoustique du genre est utilisée pour reconnaître le genre d'une personne à partir de sa voix. La reconnaissance du genre peut être classée en deux catégories : l'identification et la vérification du genre [2].

La classification du signal parole par genre est utilisée dans les systèmes modernes de recherche d'informations multimédias pour diverses applications : reconnaissance vocale, Identification de locuteur, interaction intelligente entre l'homme et l'ordinateur, biométrie, robots sociaux, indexation de contenus audio ou vidéo, etc. Les informations relatives au genre sont utilisées pour normaliser les caractéristiques de la parole, afin de réduire le taux d'erreur de reconnaissance du locuteur. En général, il est important de déterminer le genre du locuteur [2].

I.2 Définition du genre

Dans un cadre social, le genre représente la construction socioculturelle des rôles masculins et féminins, contrairement au sexe qui est une caractéristique biologique (masculin/féminin). De plus, le genre est influencé par la société, la culture et l'époque. L'identité de genre désigne le sentiment d'être homme, femme [1].

La catégorisation du genre consiste à identifier le genre d'une personne, comme un homme ou une femme, en se basant sur ses attributs biométriques. En général, on utilise des images faciales pour extraire des caractéristiques, puis on applique un classificateur aux caractéristiques extraites afin de reconnaître le genre. C'est un sujet actif de recherche dans les domaines de la vision par ordinateur [1].

La reconnaissance du genre est principalement un problème de classification à deux niveaux. Malgré l'utilisation d'autres caractéristiques biométriques telles que la démarche et le visage

pour classer le genre, les méthodes basées sur la voix demeurent les plus courantes pour la discrimination de genre [3].

I.3 Reconnaissance automatique du genre

Le développement de RAG remonte au moins au début des années 1990. Des études précédentes ont examiné les compétences techniques de RAG et ses utilisations, telles que le marketing, la génétique, l'interaction homme-machine, la surveillance et la sécurité. Plusieurs raisons expliquent l'utilisation de RAG : elle a pour objectif d'améliorer l'expérience de l'utilisateur en fournissant à un système numérique plus d'informations sur l'utilisateur, afin qu'il puisse mieux s'adapter à lui.

Le système de reconnaissance du genre (RAG) est un ensemble de méthodes informatiques utilisées pour identifier le genre d'une personne en extrayant et en analysant les caractéristiques d'images, de vidéos et/ou de sons [4].

La reconnaissance automatique du genre (RAG) (ou classification du genre) fait référence aux méthodes algorithmiques, comme les technologies de reconnaissance faciale automatique et de reconnaissance corporelle, qui extraient des caractéristiques d'images, de vidéos ou de sons d'un ou de plusieurs individus afin de déterminer leur classe de genre. Dans RAG, on utilise souvent des algorithmes de vision par ordinateur et/ou des modules de reconnaissance vocale. On extrait souvent des caractéristiques des données visuelles et/ou audio d'une personne et on les extrait. Plusieurs méthodes ont été étudiées par les chercheurs afin de mettre en pratique RAG et d'améliorer sa précision, telles que la reconnaissance vocale en utilisant la fréquence fondamentale et les coefficients MFCC [4].

I.4 Application de la Reconnaissance automatique du genre

Le développement et les progrès de la technologie de reconnaissance de genre ont conduit à de nombreuses utilisations potentielles dans un large champ d'application, car les techniques de classification de genre peuvent améliorer considérablement les capacités de perception et d'interaction de l'ordinateur. Par exemple, la classification par genre est capable d'améliorer l'intelligence d'un système de surveillance, d'analyser les demandes des clients en matière de gestion du magasin et de permettre aux robots de percevoir le genre, etc. Pour être concrètes, les applications de la classification automatique du genre peuvent être classées dans les domaines suivants [5] :

- **Marketing ciblé :** Les entreprises pourraient utiliser la RAG pour diffuser des publicités plus pertinentes aux clients en fonction de leur genre.
- **Sécurité et surveillance :** La RAG pourrait être utilisée pour identifier les individus dans les foules ou pour contrôler l'accès aux zones restreintes.
- **Assistant personnel :** Les assistants personnels dotés de la RAG pourraient s'adapter aux préférences individuelles des utilisateurs.
- **Santé :** La RAG pourrait être utilisée pour développer de nouvelles technologies de diagnostic ou de traitement, notamment pour les maladies qui affectent différemment les hommes et les femmes.

I.5 Reconnaissance acoustique du genre

La reconnaissance du genre, un domaine fascinant de la reconnaissance vocale, se base sur les différences entre les voix masculines et féminines.

En analysant les caractéristiques acoustiques de la parole, les systèmes de reconnaissance du genre peuvent identifier et différencier les voix avec une précision remarquable. Plusieurs méthodes d'analyse acoustique du signal parole ont été utilisées pour extraire des paramètres acoustiques (caractéristiques) discriminants les classes du genre F et H. Plusieurs travaux de recherches ont utilisé différents descripteurs de paramètres tels que : la fréquence fondamentale (pitch), le taux de passage par zéro, les coefficients MFCC, les coefficients PLP, les coefficients LPCC, l'entropie d'énergie [1] [6] [7].

I.5.1 Signal parole (vocal)

L'ensemble des organes (appareil vocale) agit comme un filtre, considéré comme linéaire, dont la réponse impulsionnelle comporte des fréquences de résonance caractérisées par des pics, appelés formants, dans le spectre du signal de sortie. Le signal résultant est globalement non stationnaire mais peut être considéré comme stationnaire sur de très courtes périodes, de l'ordre de 30ms (signal pseudo-stationnaire). Sur un segment de parole de cette longueur la voix est habituellement et schématiquement séparée en deux classes distinctes :

- ***Voisée*** lorsqu'il y a vibration des cordes vocales, le signal est alors quasi périodique,
- ***Non voisée*** dans le cas d'un simple soufflement, le signal est alors considéré comme aléatoire.

Dans le premier cas, la source d'excitation est modélisée par un train d'impulsions périodique, de fréquence dite de voisement F_0, qui correspond à la fréquence de vibration des cordes vocales, la fréquence fondamentale ou pitch :

$$U(n) = \sum_k \delta(n - kT) \qquad \text{(I.1)}$$

Où T représente la période de la fréquence fondamentale (en anglais, « pitch »). Dans le second cas, la source est modélisée par un bruit blanc. Cette représentation binaire de la production de la parole a été introduite par [8]. Elle est reprise sur la figure I.1

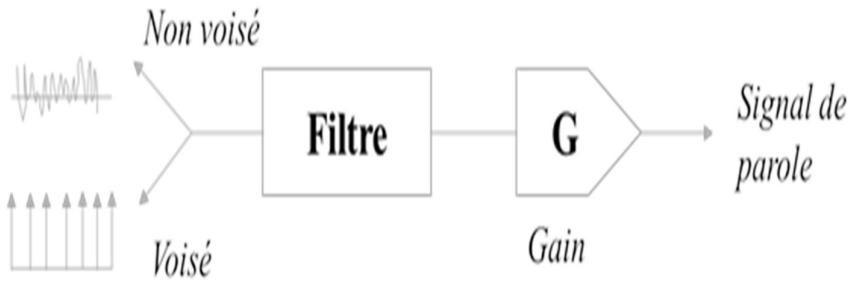


Figure I.1: Modèle de production de la parole [7]

Le volume dépend de la pression acoustique crée par la source d'air (le nombre de particules d'air déplacées). Plus elle est importante plus le volume est élevé. La hauteur tonale est définie par les vibrations de l'objet créant le son. Plus la fréquence est élevée, plus la longueur d'onde est petite et plus le son perçu est aigu. En doublant la fréquence d'une note, on obtient la même à l'octave supérieure. Et donc, en divisant la fréquence par deux, on passe à l'octave inférieure. Ce n'est qu'au-delà de 20 vibrations par seconde que l'oreille perçoit un son. Les infrasons, de fréquence inférieure à cette limite de 20 Hz sont inaudibles, de même que les ultrasons, de fréquence supérieure à 20000 Hz, soit un peu plus de 10 octaves [9].

I.5.2 Analyses du signal parole sur la RAG

Analyse et synthèse sont deux activités duales, l'analyse fournit une description du signal acoustique, alors que la synthèse utilise pour le reproduire. L'Analyse acoustique est une partie importante dans le traitement que subit le signal sonore pour pouvoir réaliser un système de haute qualité de synthèse, de compréhension, ou de reconnaissance de la parole. Cette opération consiste à tirer à partir du signal vocal un ensemble de paramètres pertinents, discriminants et robustes susceptibles de le représenter. Il y a Plusieurs techniques d'analyse sont utilisées parmi lesquelles on peut prend l'analyse par le spectrogramme [10].

I.5.3 Méthodes d'analyses du signal parole

La parole, élément fondamental du langage humain, peut être analysée en profondeur pour en extraire des informations précieuses. Cette analyse permet se décline en trois étapes principales :

1.Forme spectrale du signal : Cette approche se focalise sur les caractéristiques fréquentielles du signal vocal. Elle permet de dresser une cartographie précise de la répartition des fréquences dans le spectre de la parole.

2.Fréquence fondamentale (Le ton de la voix) : L'analyse de la fréquence fondamentale (F0) s'intéresse aux variations de la hauteur tonale dans la parole. Elle s'avère essentielle pour comprendre l'intonation, les mélodies vocales et les émotions véhiculées par le langage parlé.

3.Durée et paramètres associés (Le rythme de la parole) : cette catégorie se base sur l'analyse temporelle des différents segments de la parole, tels que les voyelles, les consonnes et les pauses. Elle s'intéresse également aux paramètres temporels comme la durée et le débit, qui contribuent au rythme et à la fluidité de la parole [11].

I.5.4 Fonctionnement d'un système RAG

Le schéma de fonctionnement d'un système de reconnaissance de genre est illustré sur la figure I.2.

Le fonctionnement d'un système RAG se déroule en une phase d'apprentissage et une phase de test. La phase d'apprentissage permet de modéliser les différentes classes du genre (F et H), alors que la phase de test permet de classifier chaque signal de test et d'évaluer les performances du système en utilisant des métriques d'évaluation telles que le taux de classification. Chacune des phases exige une étape d'extraction de paramètres permettant de convertir chaque signal d'entrée en une séquence de vecteur de paramètres (caractéristiques). Dans la phase d'apprentissage, les séquences de vecteurs de paramètres extraits des signaux d'apprentissage seront utilisées pour l'entrainement des modèles de classes. Dans la phase de test, chaque séquence de vecteur correspondant à un signal de test sera appliquée à l'entrée d'un classificateur pour reconnaître sa classe du genre. Les algorithmes de classification couramment utilisés dans les systèmes RAG sont : KNN, ANN, SVM, GMM et CNN [12].

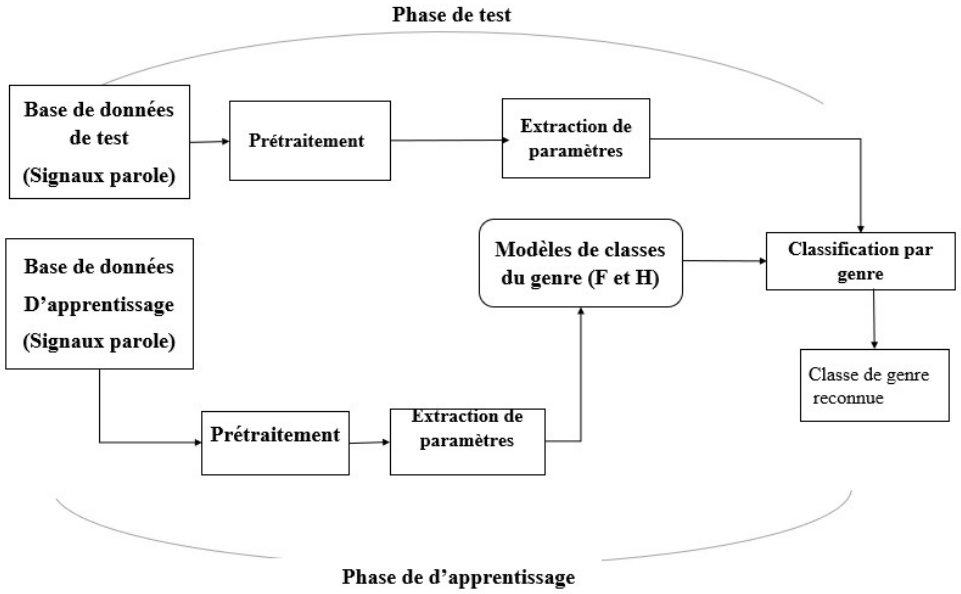


Figure I.2: Architecture d'un système RAG [13].

I.6 Méthodes de classification

Il existe plusieurs méthodes de classification qui peuvent être utilisées dans les systèmes de reconnaissance du genre, chacune ayant ses propres avantages et limites. Dans notre travail, on s'intéresse à utiliser les classificateur KNN et GMM.

I.6.1 Classificateur (KNN)

Le KNN est un algorithme de la famille des algorithmes d'apprentissage automatique. Cet algorithme repose sur une idée fondamentale simple, qui implique une comparaison directe entre le vecteur caractéristique de l'instance (nouvelle observation) à classer et les vecteurs des instances de la base d'apprentissage. La comparaison implique de déterminer les distances entre ces instances. Ensuite, on attribue à l'instance à classer la classe la plus importante parmi les classes des k instances les plus proches. On utilise différentes distances dans la méthode KNN: Euclidienne, Hamming, City block, Cosinus, Corrélation et Gaussienne [14].

Dans notre travail on utilise les distances suivantes : Euclidienne, City block, Cosinus et Corrélation. Chacune de ces distances est exprimée en fonction d'un vecteur $x(x_1, \dots, x_m)$ et d'un vecteur $y(y_1, \dots, y_m)$.

La distance Euclidienne entre les deux vecteurs *x* et *y* est définie par :

$$d_E(x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \qquad \text{(I.2)}$$

La distance City block est définie comme suit :

$$d_C(x,y) = \sum_{i=1}^{m} |x_i - y_i| \qquad \text{(I. 3)}$$

La distance Cosine est définie comme suit :

$$d_{\cos}(x,y) = 1 - \frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} x_i^2}} \frac{1}{\sqrt{\sum_{i=1}^{m} y_i^2}} \qquad \text{(I. 4)}$$

La distance corrélation est définie comme suit :

$$d_{\text{cor}}(x,y) = 1 - \frac{\sum_{i=1}^{m} (x_i - y_i)}{\sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}} \frac{(x_i - y_i)}{\sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}} \qquad \text{(I. 5)}$$

I.7 Etat de l'art sur les systèmes RAG

Le tableau I.1 illustre quelques travaux de recherches sur la reconnaissance du genre en détaillant brièvement le classificateur utilisé, la méthode d'extraction, la base de données, l'année de publication ainsi que le résultat obtenu.

Tableau I.1: Etat de l'art sur des systèmes RAG

Référence	Année	Classificateur utilisé	Extraction des paramètres	Base des données	Résultat (%)
[15]	2024	MLP	MFCC et ZCR et RMSF	CREMA-D et EMO-DB	98
[16]	2021	SVM	MFCC et PCA	RAVDESS	98.88
[17]	2015	GMM	MFCC	EMO-DB	80
[18]	2020	SVM et KNN	MFCC	TIMIT et RAVDESS et BGC	96.8
[19]	2018	SVM	MFCC	TIMIT	96.45
[20]	2020	Machine learning algorithm (J48)	MFCC et VQ	CMU_ARCTIC	100

I.8 Conclusion

Dans ce chapitre, nous avons donné quelques généralités sur la reconnaissance automatique du genre et les domaines d'application et les modalités utilisées. Nous avons décrit le fonctionnement d'un système reconnaissance acoustique du genre à partir d'analyse du signal parole, ensuite nous avons donné brièvement les méthodes d'analyse ainsi que les méthodes de classification, puis nous avons terminé ce chapitre par un état de l'art sur les systèmes RAG.

Chapitre II

Reconnaissance acoustique de genre basée sur les modèles GMM

II.1 Introduction

Le fonctionnement d'un système RAG se déroule généralement en une phase d'apprentissage et une phase de test. La conception d'un tel système doit tenir en compte le compromis entre la précision, la rapidité, la simplicité d'implémentation, le coût d'espace mémoire. Ce compromis dépend généralement de l'étape d'extraction de paramètres et de l'étape de classification.

Dans [1] les auteurs ont proposé l'application du classificateur KNN vue de sa simplicité d'implémentation, combiné avec l'extraction des paramètres MFCC pour la tâche de reconnaissance du genre. Cependant, ce classificateur exige un large espace mémoire et un grand temps de calcul.

Dans [21] les auteurs ont utilisé le classificateur GMM combiné avec l'extraction des paramètres MFCC. Ce dernier système exige moins d'espace mémoire et temps de calcul. Plus particulièrement, le classificateur GMM a montré son efficacité dans plusieurs tâches de reconnaissance vocale [7].

Ainsi, dans notre travail, on s'intéresse à la conception d'un système RAG basé sur le classificateur GMM combiné avec l'extraction des paramètres MFCC. Le système RAG proposé sera décrit dans le chapitre 3 et ses performances seront comparées avec celles du système basé sur le classificateur KNN.

Dans ce chapitre, nous allons présenter la méthode de classification GMM ainsi que la méthode d'extraction de paramètres MFCC.

II.2 Fonctionnement d'un système RAG basé sur le classificateur GMM

Un système RAG basé sur le classificateur GMM combiné avec la méthode d'extraction des paramètres MFCC est illustré sur la figure II.1. La conception d'un tel système se base sur une phase d'apprentissage et une phase de test. La phase d'apprentissage permet premièrement de convertir chaque signal d'apprentissage en une séquence de vecteurs de paramètres MFCC, en suite modéliser chaque classe de genre par un modèle GMM en utilisant les séquences de vecteurs extraits correspondants à cette classe. La phase de test permet premièrement de convertir chaque signal de test en une séquence de vecteurs MFCC, ensuite appliquer l'algorithme de classification GMM sur cette séquence pour reconnaître la classe du genre correspondant à ce signal. Puis, des

mesures de métriques tel le taux de classification et le temps de calcul seront effectuées pour évaluer la qualité des performances.

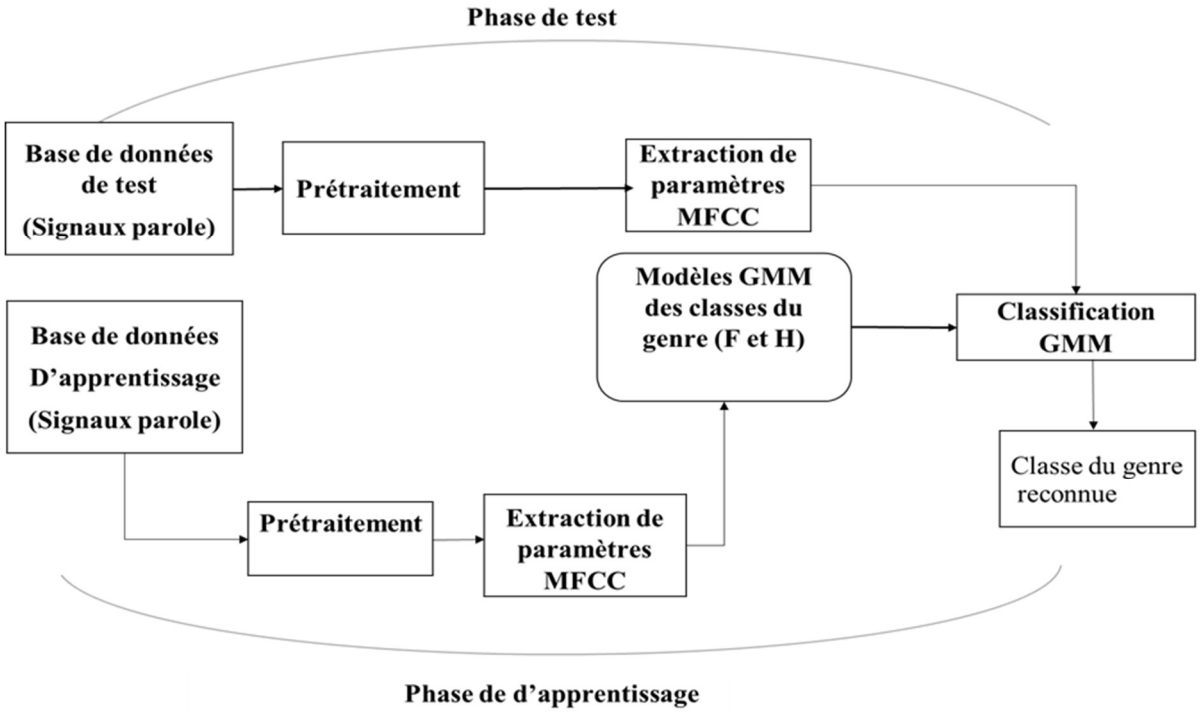


Figure II.1: Schéma d'un système RAG basé sur le classificateur GMM combiné avec la méthode d'extraction des parameters MFCC [13].

II.3 Classification GMM

Le fonctionnement d'un classificateur GMM se déroule en une étape d'apprentissage permettant l'entrainement des paramètres des modèles GMM représentant les classes considérées en utilisant des observations (échantillons) d'entrainement, et une étape de classification permettant de classifier des observations de test.

Un modèle de mélange gaussien (GMM) est une fonction de densité de probabilité paramétrique représentée par une somme pondérée de densités de composantes gaussiennes. Cette modélisation se base sur l'algorithme de regroupement (clustering) EM (Expectation-Maximization) qui consiste à regrouper l'ensemble de données d'une classe en groupes dont chacun est représenté par une fonction de densité de probabilité (fdp) gaussienne [17].

Généralement, les données sont des vecteurs composés chacun de *d* paramètres. Ces données peuvent être considérées comme des variables aléatoires multidimensionnelles *x* de dimension *d* gérées par des pdf de mélange gaussien décrites comme suit :

$$f(x) = \sum_{i=1}^{k} \frac{\alpha_i}{\sqrt{(2\pi)^d \left|\sum_{\mathrm{i}}^{n}\right|}} \exp\left(-\frac{1}{2}(x-\mu_i)'.\sum_{\mathrm{i}}^{-1}.(x-\mu_i)\right) \qquad \text{(II. 1)}$$

k : est le nombre de composantes gaussiennes.

$\mu_i$et$\sum_{\mathrm{i}}$sont respectivement le vecteur de moyennes et la matrice de covariance de la i$^{\text{ième}}$ gaussienne, $\alpha_i$est le poids de la $i^{\text{ième}}$ gaussienne vérifiant la condition : $\sum_{i=1}^{k} \alpha_i$=1.

$|\sum_{\mathrm{i}}|$ est le déterminant de la matrice$\sum_{\mathrm{i}}$.

Pratiquement, cette modélisation est effectuée en langage de programmation Matlab en utilisant la fonction gmdistribution.fit pour la modélisation et la fonction pdf pour la classification.

II.4 Coefficients cepstreaux en échelle Mel (MFCC)

Les coefficients MFCC sont calculés en appliquant la transformée cosinus inverse sur le logarithme des énergies spectrales estimées sur des bandes fréquentielles distribuées en échelle de MEL d'une fenêtre d'analyse. La transformée cosinus inverse permet d'obtenir des coefficients MFCC décolérés.

L'énergie spectrale est calculée en appliquant un ensemble de filtres espacés de manière uniforme sur une échelle fréquentielle modifiée, connue sous le nom d'échelle Mel [22]. La conversion de la fréquence en Mel-échelle est définie par :

$$B(f) = 2595 * \log10(1 + f/700) \quad [22] \qquad \text{(II. 2)}$$

Où f est la fréquence en Hz, (B) est la fréquence en échelle Mel de f. La figure II.2 présente un exemple de conversion d'Hertz en Mel.

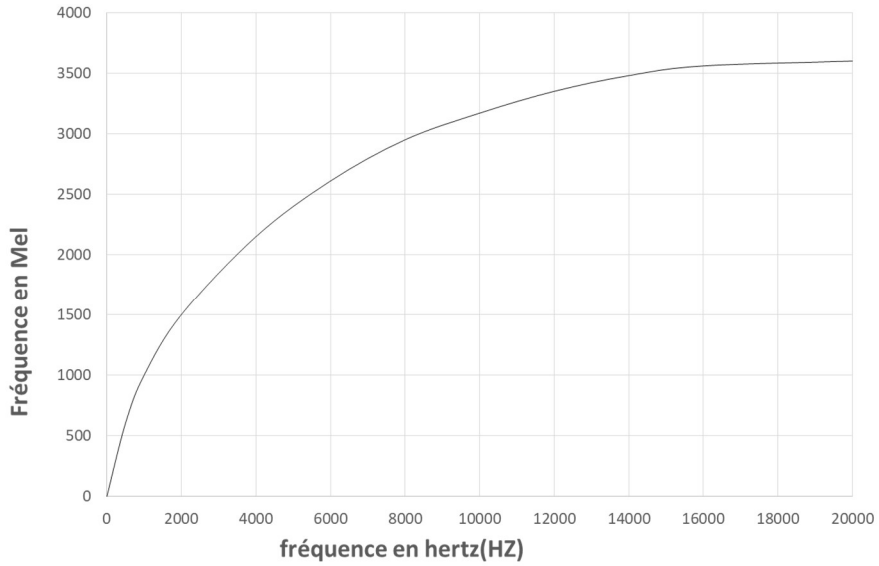
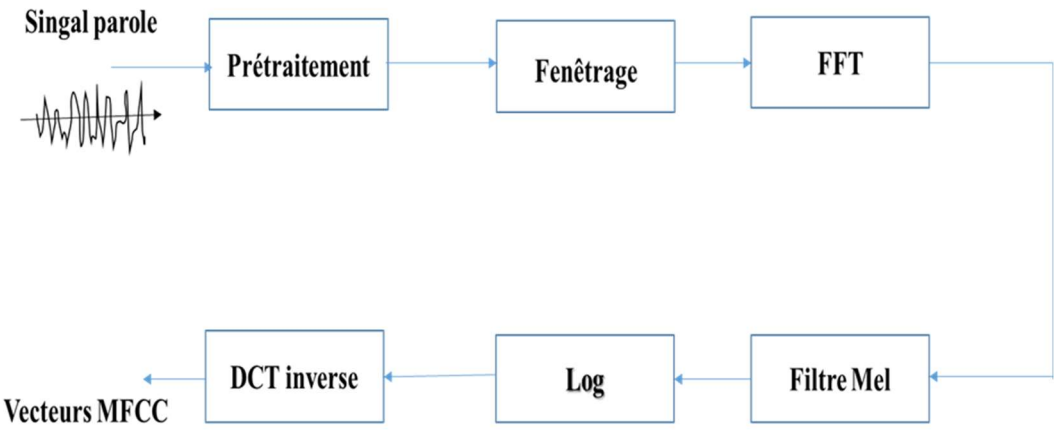


Figure II.2: Example de conversion des hertz en Mel [23].

II.5 Calcul du vecteur de paramètres MFCC

L'extraction des paramètres acoustiques MFCC se déroule généralement en plusieurs étapes telles qu'illustrées sur la Figure [24].



FigureII.3: Etapes de calcul d'un vecteur de paramètres MFCC [24]

II.5.1 Prétraitement du signal

L'objectif du prétraitement est d'éliminer les bruits indésirables et de réduire les déformations de signaux dans le but d'extraire des paramètres significatifs [25].

II.5.2 Fenêtrage

Le fenêtrage de la trame est utilisé pour minimiser les discontinuités du signal au début et à la fin de chaque trame. Le signal sonore est découpé en courts segments de durée fixe, appelés fenêtres ou trames, d'une taille habituelle comprise entre 20 et 40 millisecondes.

Afin de minimiser les artefacts sonores dus à cette découpe, on utilise couramment une fenêtre particulière appelée fenêtre de Hamming [25].

II.5.3 Calcul de la transformée de Fourier rapide (Fast Fourier Transform, FFT)

La transformée de Fourier rapide (FFT) est un algorithme rapide mathématique utilisé pour calculer la transformée de Fourier discrète (TFD) d'un signal. La TFD est une transformation qui décompose un signal en ses composantes de fréquence. Et est définie comme suit, Les valeurs obtenues sont appelées le spectre [23].

$$X[k] = \sum_{N=0}^{N-1} x_0[n] e^{-\frac{j2\pi nk}{N}}, \qquad 0 \leq K \leq N \qquad \text{(II. 3)}$$

Où k est l'indice de la composante fréquentielle,

N est le nombre d'échantillons,

X la transformée de Fourrier du signal x.

En général, les valeurs X[k] sont des nombres complexes et nous utilisons que leurs valeurs absolues (énergie de la fréquence) [23].

II.5.4 Filtrage sur l'échelle Mel

Les filtres Mel sont des filtres répartis logarithmiquement en fréquence, et dont la largeur et l'amplitude varient afin de conserver une énergie constante. Le logarithme de l'énergie de chaque filtre est calculé selon l'équation suivante [23]:

$$S[m] = ln\left[\sum_{k=0}^{N-1} |X[k]|\, H_m[k]\right], \qquad 0 < m \leq M \qquad \text{(II. 4)}$$

Où $H_m[k]$ est la fonction de transfert du filtre Mel

II.5.5 Calcul du cepstre sur l'échelle Mel

Le cepstre sur l'échelle de fréquence Mel est obtenu par le calcul de la transformée en cosinus discrète inverse du logarithme de la sortie des M filtres (reconversion du log-Mel-spectre vers le domaine temporel) [23].

$$c[n] = \sum_{m=0}^{N-1} S[m] \cos\left(\frac{\pi n\left(m-\frac{1}{2}\right)}{M}\right), \qquad 0 \leq n < M \qquad \text{(II. 5)}$$

L'ensemble des coefficients c[n] constitue le vecteur de paramètres MFCC.

II.6 Calcul des paramètres dynamiques des MFCC

Les paramètres dynamiques des MFCC sont des valeurs complémentaires aux coefficients MFCC statiques, qui permettent de capturer les variations temporelles (dynamique) des caractéristiques acoustiques d'un signal vocal. Ces variations représentent les changements rapides et les modulations d'intensité dans la voix humaine.

Le calcul des paramètres dynamiques des MFCC s'effectue généralement en utilisant les premières et deuxièmes dérivées des coefficients MFCC statiques. Les premières dérivées (notées Δ) représentent la vitesse de changement des MFCC, tandis que les deuxièmes dérivées (notées ΔΔ) reflètent l'accélération de ces changements.

L'intégration de ces paramètres dynamiques dans les systèmes de reconnaissance de la parole permet d'améliorer leur robustesse face à la variabilité de la voix humaine et aux bruits environnants. En effet, les paramètres dynamiques capturent des informations temporelles supplémentaires qui aident à mieux représenter les caractéristiques temporelles de la parole [7]. La figure II.4 montre le calcul des paramètres dynamiques Δ et ΔΔ.

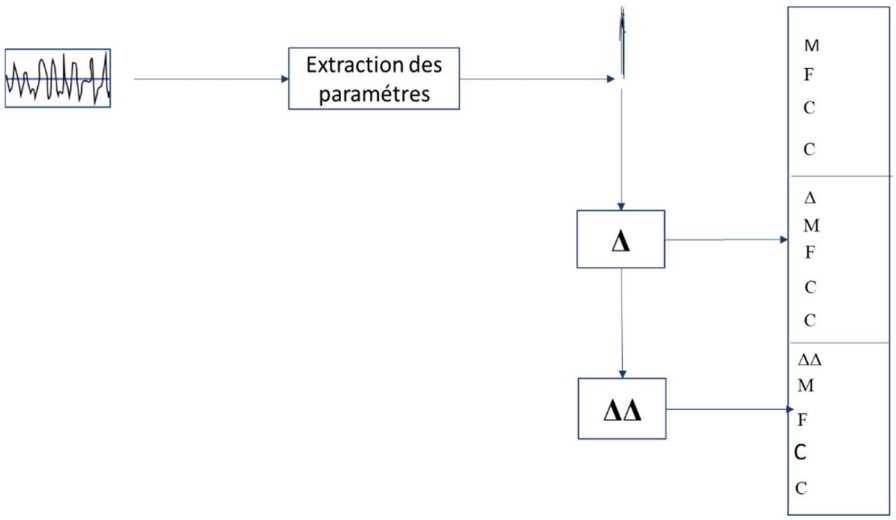


Figure II.4: Calcul des derives première et seconde des coefficients MFCC [14].

II.7 Conclusion

Dans ce chapitre nous avons présenté brièvement les étapes de fonctionnement d'un système RAG basé sur le classificateur GMM combiné avec la méthode d'extraction des paramètres MFCC. Plus particulièrement, nous avons détaillé le principe de fonctionnement du classificateur GMM ainsi que la méthode d'extraction des paramètres MFCC.

Chapitre III

Système de reconnaissance acoustique du genre basée sur le classificateur GMM combiné avec la règle de vote

III.1 Introduction

L'objectif de notre travail consiste à concevoir et implémenter un système RAG performant en termes de précision et de complexité, exigeant moins d'espace mémoire et temps de calcul. Un tel système exige une phase d'apprentissage pour modéliser les deux classes du genre (H et F), et une phase de test permettant de reconnaissance des signaux de test pour évaluer les performances du système en utilisant des algorithmes de classification tels que : KNN, SVM, GMM, ANN ou CNN. Ces deux phases utilisant une base de données d'apprentissage et une base de test exige chacune une étape d'extraction de paramètres permettant de convertir chaque signal en une séquence de vecteurs de paramètres en utilisant des techniques d'extraction de paramètres tels que les coefficients MFCC. Dans la phase de test, le classificateur utilise les modèles entrainés durant la phase d'apprentissage pour classifier les vecteurs de paramètres extrait de chaque signal en classe de genre.

Dans notre travail, nous proposons la conception et l'implémentation sous Matlab d'un système de reconnaissance acoustique du genre basé sur l'algorithme de classification GMM. Cet algorithme est appliqué sur les vecteurs de paramètres MFCC et combiné avec la stratégie de la règle de vote pour décider sur la classe du signal de test. De plus, nous proposons de comparer les performances de ce système avec celles du système basé sur l'algorithme de classification KNN.

Différentes questions peuvent se poser pour déterminer les bonnes configurations des deux systèmes RAG basés sur les algorithmes GMM et KNN.

1- Comment implémenter un système de reconnaissance acoustique du genre basé sur les classificateurs GMM et KNN combinés avec la stratégie de la règle de vote ?
2- Quel est le meilleur type distance et le nombre optimal k des vecteurs les plus proches voisins du classificateur KNN permettant d'obtenir les meilleures performances du système RAG ?
3- Quel est le nombre optimal des composantes gaussiennes des modèles GMM des classes du genre H et F ?
4- Est-ce que l'utilisation de la règle de vote contribue à l'amélioration des performances du système reconnaissance acoustique basé sur le classificateur GMM ?

5- Le classificateur GMM est-il efficace par rapport au classificateur KNN ?

La première étude s'est concentrée sur la description des différentes étapes nécessaires à la mise en œuvre d'un système RAG dans l'environnement de programmation Matlab. Le système utilise la boîte à outils HTK pour extraire les paramètres MFCC (Mel-Frequency Cepstral Coefficients) [26]. Les performances des systèmes à implémenter sont évaluées en utilisant la base de données de référence EMO-DB (Berlin Data base of Emotional Speech), qui sera brièvement présentée dans la section suivante.

III.2 Description de la base de données EMO-DB

La base de données EMO-DB est couramment utilisée comme une base de référence dans plusieurs travaux de recherches [16] [27] [7]. Plus particulièrement, cette base de données de signaux parole est utilisée pour évaluer les performances des systèmes de reconnaissance d'émotions, de genre et de l'âge [16]. Cette base de données se distingue par sa richesse en vocabulaire, la qualité des signaux audio, la variété des longueurs de phrases et sa diversité de locuteurs avec différents âges et genre. Cette base de données comprend 10 phrases provenant de différents textes. Elle est composée de deux groupes. Le premier groupe, nommé A, est constitué de 5 phrases courtes et un autre groupe, nommé B, est constitué de 5 phrases longues. Ces phrases en allemand sont prononcées par 10 acteurs, 5 hommes et 5 femmes. Sept états émotionnels primaires sont simulés par les acteurs : colère, dégoût, peur, joie, neutre, tristesse et surprise. Cette base est constituée réellement de 535 signaux qui représentent sept types d'émotions. Dans un premier temps, les enregistrements ont été réalisés à 48 kHz, puis sous-échantillonnés à 16 kHz. Dans notre étude, l'ensemble A (277 énoncés, 51,78 % de la base totale) est utilisé comme base de données d'apprentissage, alors que l'ensemble B (258 énoncés, 48,22 % de la base totale) est utilisé comme base de données de test. En conséquence, notre système RAG fonctionne en mode indépendant du texte et dépendent du locuteur. Le tableau III.1.a montré la répartition de ces bases en locuteurs et en classes de genre.

Dans notre travail, nous nous intéressons également à évaluer les performances du système RAG proposé fonctionnant en mode indépendant du locuteur et dépendant du texte. Dans ce mode, les

Locuteurs contribuant dans la phase d'apprentissage ne contribuent pas dans la phase de test. Cette répartition est illustrée dans le tableau III.1.b.

Tableau III.1.a : Répartition des locuteurs de la base EMO-DB

Classe	Féminin					Total	Masculin					Total
Locuteur	F_1	F_2	F_3	F_4	F_5	5	H_1	H_2	H_3	H_4	H_5	5
TEST	29	20	30	30	35	144	24	17	29	17	27	114
APP	29	23	31	39	36	158	25	21	26	18	29	119

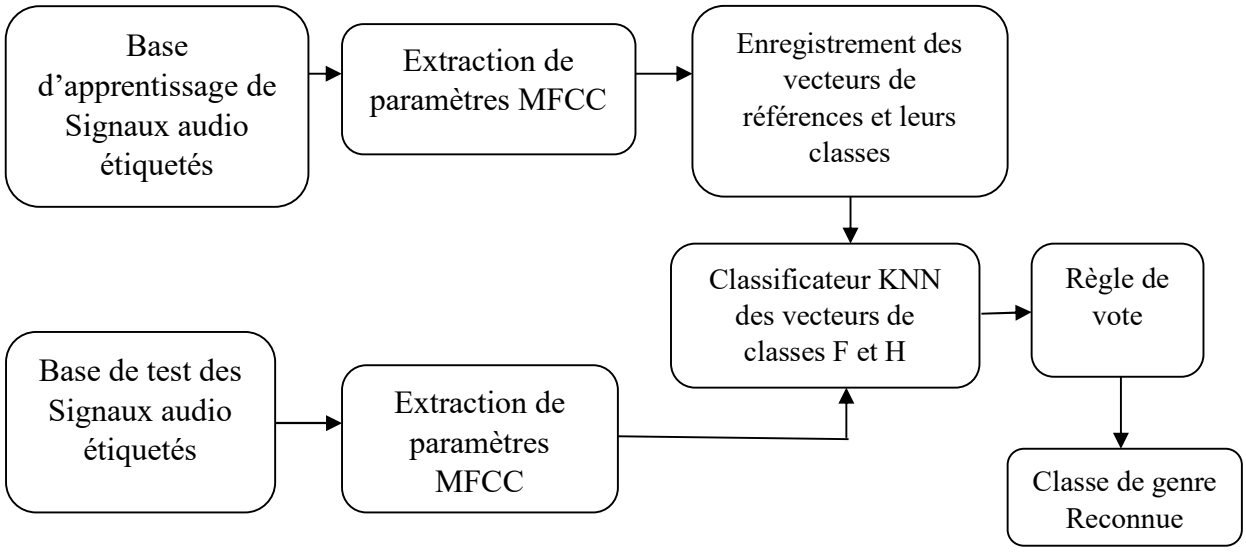
Tableau III.1.b: Répartition de la base de données EMO-DB en base d'apprentissage et base de test en mode indépendant du locuteur et mode dépendant du texte

Classe	APP					Total	TEST					Total
Locuteur	F_1	F_2	H_1	H_2	H_3	5	F_3	F_4	F_5	H_4	H_5	5
Nombre	58	43	49	38	55	243	61	69	71	35	56	292

III.3 Description de l'implémentation des étapes de fonctionnement du système RAG proposé

Dans notre travail, nous avons proposé d'appliquer les algorithmes KNN et GMM pour la classification des signaux parole en classes de genre H (Masculin) et F (Féminin) basée sur l'extraction des paramètres MFCC, combinée avec la règle de vote. Ainsi, le premier système se base sur le classificateur KNN, alors que le deuxième système se base sur le classificateur GMM. Le fonctionnement de chaque système se déroule en une phase d'apprentissage et une phase de test. Pour chaque phase, le signal d'entrée est converti en une séquence de vecteurs de paramètres MFCC en utilisant la boite à outils HTK. Durant la phase d'apprentissage, chaque classe est représentée soit par des vecteurs de références dans le premier système ou soit représentée par un modèle GMM dans le cas du deuxième système, en utilisant la base de données d'apprentissage. La phase de test consiste à classifier chaque vecteur de paramètres MFCC d'un signal de test en

utilisant le classificateur KNN ou GMM et d'appliquer ensuite la règle de vote sur la séquence d'indices de classes obtenue pour reconnaitre la classe du signal. Les deux systèmes RAG basés sur les classificateurs KNN et GMM sont illustrés respectivement sur les figures (III.1) et (III.2).



FigureIII.1: Schéma du système RAG propose, basé sur le classificateur KNN

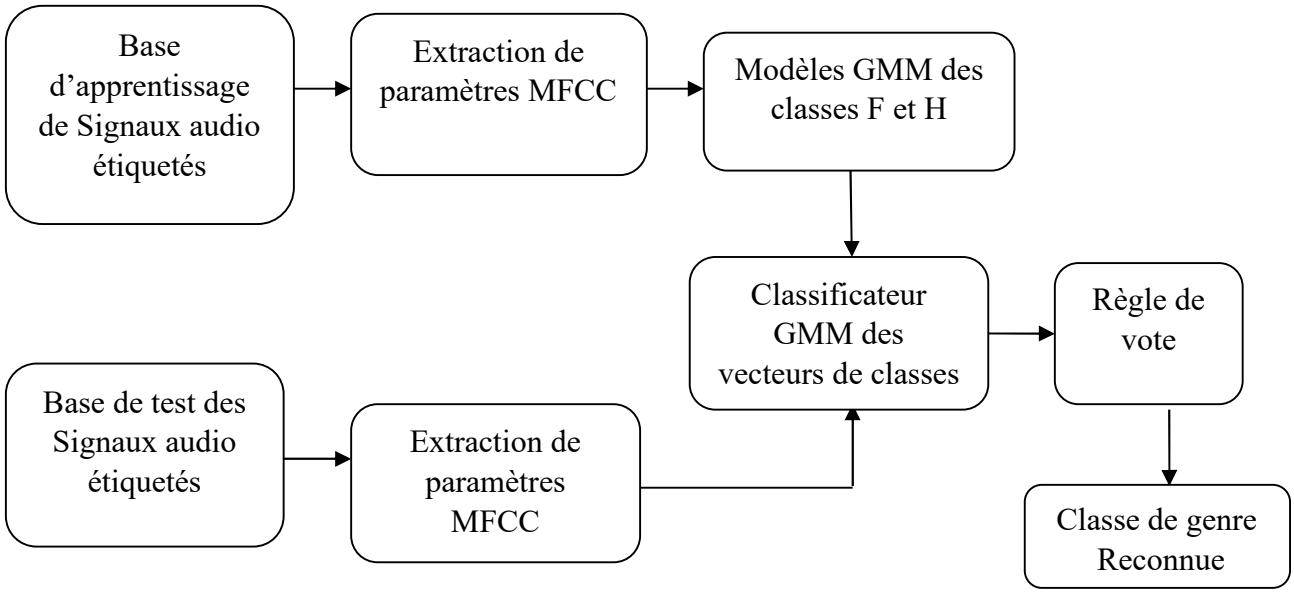


Figure III.2: Schéma proposé du système RAG basé sur GMM

Les différentes étapes d'implémentation des deux systèmes sont données comme suit :

- Préparation des données : Répartition de la base de données EMO-DB en une base de données d'apprentissage constituée de 277 signaux et une base de données de test constituée de 258 signaux.
- Extraction des paramètres : Extraire des paramètres MFCC des données d'apprentissage et de test.
- Classification des signaux : Appliquer les algorithmes KNN et GMM sur l'ensemble de vecteurs de paramètres de test.
- Application de la stratégie de la règle de vote sur chaque séquence d'indices de classes pour identifier la classe du signal correspondant.
- Évaluation des performances : Évaluer les performances du système en termes de taux de classification et temps de calcul.

III.3.1 Extraction des paramètres

L'étape d'extraction de paramètres consiste à convertir chaque signal d'entrée en une séquence de vecteurs de paramètres discriminants. Premièrement, chaque signal d'entrée est découpé en une séquence de fenêtres d'analyse de 30 ms chevauchées de 20ms. Ensuite, chaque fenêtre est convertie en un vecteur de paramètres en utilisant une technique d'extraction de paramètres.

Dans notre travail, nous nous sommes intéressés la méthode d'extraction des paramètres MFCCs dont le vecteur de paramètres est constitué de 12 coefficients MFCC statiques, de paramètre du logarithme de l'énergie totale (E) ainsi que de leurs paramètres dynamiques Δ et ΔΔ. Ainsi, chaque vecteur est composé de 39 paramètres. Cette étape est implémentée en utilisant l'outil Hcopy de la boite à outils HTK [26]:

HCopy –C analysis.conf fichier_son.wav fichier_param.mfc

Où :

- **Analysis.conf**: est un fichier de type texte permettant la configuration de l'étape d'extraction des paramètres acoustiques MFCC. Dans notre travail, les différents paramètres de configuration sont illustrés dans le tableau III.2.
- **fichier_son.wav** : est un fichier son de format wav, représentant le signal d'entrée.
- **fichier_param.mfc** : est un fichier de sortie dans lequel la séquences de vecteurs représentant les signaux d'entrée sont enregistrées sous format HTK.

Tableau III.2 : Fichier de configuration (analysis.conf).

SOURCEFORMAT	=	WAV
SOURCEKIND	=	WAVFORM
HNET : TRACE	=	1
TARGETKIND	=	MFCC_E_DA
# Unit	=	0.1 micro-second
SOURCERATE	=	625
SAVECOMPRESSED	=	F
SAVEWITHCRC	=	F
WINDOWSIZE	=	300000.0
TARGETRATE	=	100000.0
NORMALISE	=	F
NUMCEPS	=	12
USEHAMMING	=	T

PREEMCOEF	=	0.97
NUMCHANS	=	26
CEPLIFTER	=	22

III.3.2 Classification des vecteurs

Dans le premier système RAG, chaque signal de test est converti en une séquence de vecteurs de paramètres MFCC. Ensuite, chaque vecteur de test est comparé avec tous les vecteurs de références enregistrés préalablement dans la phase d'apprentissage en utilisant l'algorithme KNN. Cette comparaison a pour objectif de reconnaître la classe du vecteur de test en cherchant la classe la plus votée de l'ensemble des k vecteurs les plus proches. Ainsi, le classificateur KNN fait correspondre à chaque séquence de vecteurs représentant le signal de test, une séquence d'indices de classes.

Dans le deuxième système RAG, chaque vecteur de test est classifié en utilisant le classificateur GMM dont ses modèles de classes F et H sont entraînés dans la phase d'apprentissage.

III.3.3 Application de la stratégie de la règle de vote

Les deux classificateurs KNN et GMM ont permis d'associer à chaque signal de test une séquence d'indices de classes en classifiant chaque vecteur en un indice de classe F ou H. Cependant, l'objectif du système RAG est de reconnaître la classe du signal. Une solution permettant d'obtenir cette classe, consiste à appliquer la stratégie de la règle de vote en cherchant la classe la plus votée à partir de la séquence d'indices de classes obtenue précédemment. Plus particulièrement, cette stratégie combinée avec différents classificateurs a été utilisée dans plusieurs domaines d'applications tels que : l'identification des appareils électriques domestiques [28], la classification des signaux parole en classes injonction et non-injonction [29], la classification des signaux électromyogrammes (EMG) pour l'aide au diagnostic de la maladie de Parkinson [30].

III.3.4 Évaluation des performances du modèle RAG

L'évaluation des performances d'un système RAG peut être effectuée en utilisant le taux de classification comme une mesure de précision, et le temps de calcul et l'espace mémoire comme des mesures de complexité.

Dans notre travail, nous évaluons premièrement le taux de classification TCV de l'ensemble des vecteurs de paramètres des signaux de test, ensuite nous évaluons le taux de classification des signaux de test. Le taux TCV peut être considéré comme une mesure globale des performances du système RAG sans tenir en compte l'appartenance des vecteurs à leurs propres signaux.

La relation suivante définit le taux de classification TCV :

$$\mathrm{TCV} = \frac{\mathrm{Nv_C}}{\mathrm{Nv_T}}.100 \qquad \text{(III. 1)}$$

Avec :

- $\mathrm{Nv_T}$ est le nombre des vecteurs de paramètres de tous les signaux de la base de test.
- Nv_C est le nombre de vecteurs de paramètres correctement classifiés.

Le taux TCS permet d'évaluer globalement les performances du classificateur GMM ou KNN en tenant la classification des signaux test, en appliquant la stratégie de la règle de vote.

$$\mathrm{TCS} = \frac{\mathrm{Ns_c}}{\mathrm{Ns_t}}.100 \qquad \text{(III. 2)}$$

Avec :

- Ns_t : est le nombre total des signaux de la base de test.
- Ns_c: est le nombre de signaux correctement reconnus.

III.4 Expériences et résultats

Dans cette section, nous allons présenter les différentes expériences menées dans but de répondre aux différentes questions posées au début de ce chapitre. Les premières expériences sont effectuées en mode dépendant du locuteur et indépendant du texte.

III.4.1 Système RAG basé sur le classificateur KNN en mode dépendant du locuteur

Dans cette section, nous cherchons la meilleure configuration du classificateur KNN permettant d'obtenir les meilleurs taux TCV et TCS en mode dépendant du locuteur et indépendant du texte.

III.4.1.1 Choix du nombre des vecteurs les plus proches voisins K et le type de distance

L'objectif de cette expérience est de déterminer la configuration optimale du classificateur KNN en sélectionnant le nombre optimal des vecteurs les plus proches voisins K ainsi que le choix optimal de la distance, permettant d'obtenir le taux de classification TCS maximal. Dans cette expérience, la valeur de K varie de 1 à 50, alors que les différents types de distances utilisées sont : Euclidean, City block, Cosine et Correlation.

Le tableau III.3 présente les 25 premières valeurs des taux de classification TCV et TCS, pour différents types de distances. Les figures 3 et 4 illustrent respectivement les taux TCV et TCS en fonction de K qui varie de 1 à 50.

Tableau III. 3 : Taux de classification TCS du système RAG pour différentes valeurs de k et différents types de distances

Distance / K	Euclidean	Cityblock	Cosine	Correlation
1	98.83	98.44	99.61	99.61
2	98.83	98.44	99.61	99.61
3	98.83	98.44	98.83	98.83

4	98.83	98.83	98.44	98.83
5	98.83	97.66	98.05	98.05
6	98.83	97.27	98.05	98.44
7	97.66	97.27	98.05	98.44
8	98.44	96.88	98.44	98.44
9	97.66	96.88	98.05	98.05
10	97.27	96.88	98.05	98.05
11	97.27	96.49	97.66	97.27
12	97.66	96.88	97.66	97.66
13	97.27	96.49	97.23	97.27
14	97.27	96.88	97.66	97.27
15	97.27	96.49	97.66	96.88
16	97.66	96.49	97.66	96.88
17	97.27	96.49	97.27	96.10
18	97.27	96.88	97.27	97.27
19	97.88	96.49	97.27	96.49
20	97.27	96.49	97.27	96.49
21	96.88	96.49	96.49	96.10
22	96.88	96.49	96.49	96.49
23	96.88	96.49	96.88	96.10
24	97.27	96.49	96.49	96.49
25	96.88	96.49	96.49	96.10

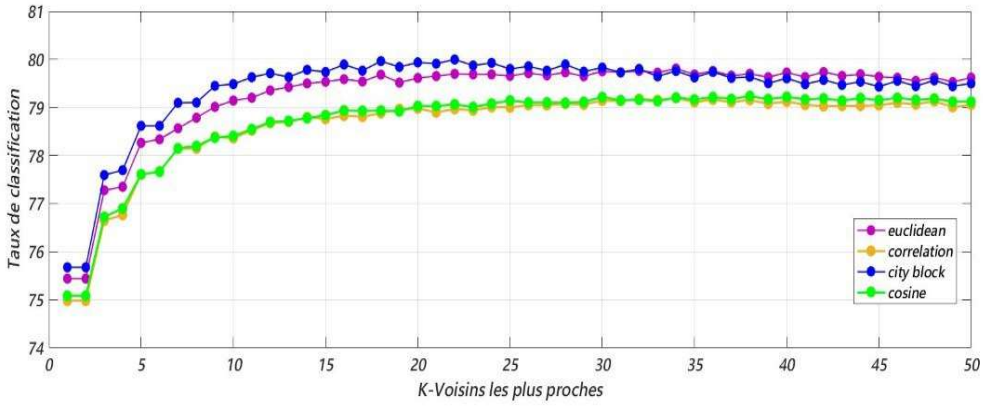


Figure III.3: Taux de reconnaissance TCV système RAG avec différentes valeurs de k du classificateur KNN

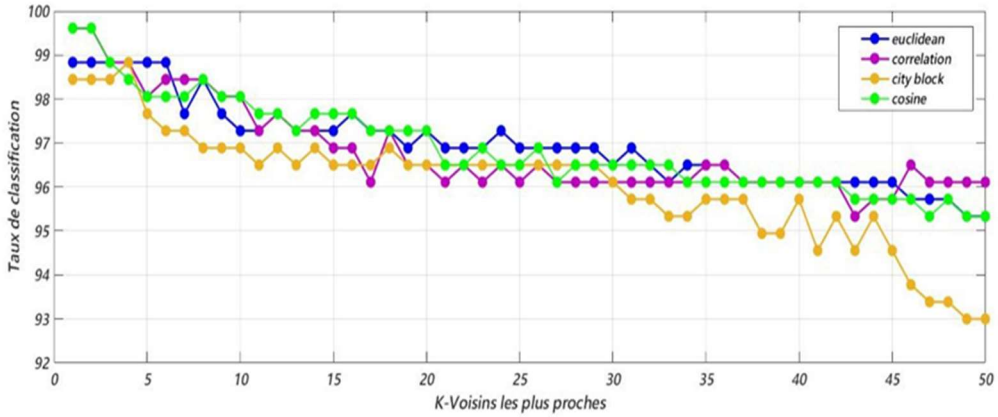


Figure III.4: Taux de reconnaissance TCS du système RAG avec différentes valeurs de k pour classificateur KNN

A partir des résultats illustrés sur le tableau III.3 ainsi que les figures III.3 et III.4, nous pouvons remarquer les points suivants :

- Les valeurs des taux TCV présentent une grande marge de variation indiquant l'importance du choix de la valeur de K et le type de distance.
- La courbe de TCV atteint approximativement un plateau au K égal à 10.

• La courbe de TCS présente une décroissance légère avec une variation égale approximativement à 2.

• Le meilleur taux TCS du système basé sur le classificateur KNN est égal à 99.61% en choisissant la distance 'correlation' ou 'cosine' avec K égal à 1.

• Les valeurs de taux TCS sont plus grandes que celles du taux TCV quelques soit K. Ceci montre l'efficacité de la stratégie de la règle de vote.

III.4.1.2 Meilleure combinaison des paramètres MFCC

Le tableau III.3 présente le TCS du système RAG basé sur classificateur KNN combiné avec la règle de vote.

Tableau III.4 : Taux de classification TCS pour différentes combinaisons de paramètres

Combinaison de paramètres	MFCC	MFCC_E	MFCC_D	MFCC_ED	MFCC_EDA
Nombre de paramètres	12	13	24	26	39
TCS	88.02	89.23	90.54	91.32	92.18

A partir de ce tableau, nous pouvons donner les remarques suivantes :

- L'utilisation de la totalité des paramètres (MFCC_EDA) présente le meilleur taux de classification TCS de 92.18%.
- L'ajout des paramètres énergétiques améliore le taux de classification TCS.
- L'ajout des paramètres dynamiques (D) et (A) améliore également le taux TCS.
- La combinaison MFCC_E représente le meilleur compromis entre le taux de classification et la dimension des vecteurs.

III.4.2 Système RAG basé sur le classificateur GMM en mode dépendant du locuteur

Cette section a pour objectif de chercher la meilleure configuration du classificateur GMM permettant d'obtenir les meilleurs taux TCV et TCS du système RAG en mode dépendant du locuteur et indépendant du texte. Plus particulièrement, cette configuration dépend du nombre de composantes gaussiennes des modèles GMM représentant les classes de genre H et F.

Les figures III.5 et III.6 présentent respectivement les taux TCV et TCS en fonction de nombre de composantes gaussiennes Ng.

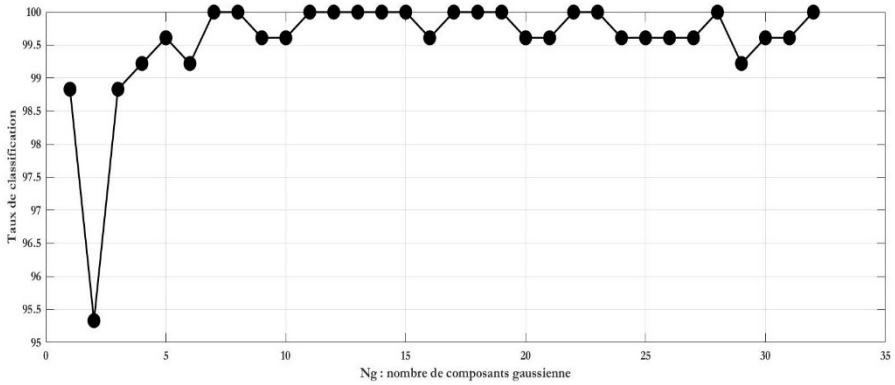


Figure III.5: Taux TCS pour classificateur GMM

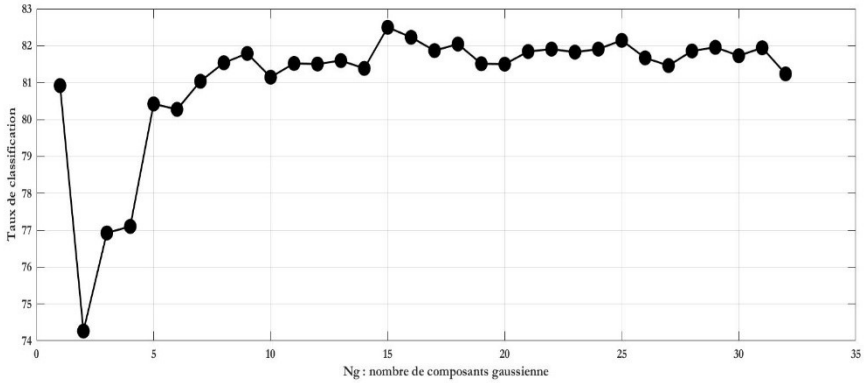


Figure III.6: Taux TCV pour classificateur GMM

A partir des figures III .5 et figure III.6, nous pouvons donner les remarques suivantes :

- La courbe du taux TCV augmente et atteint approximativement un plateau au Ng égal à 9.
- La courbe du taux TCS augmente et atteint sa valeur maximale de 100% en prenant un nombre de composantes gaussiennes Ng égal à 7. Ensuite, cette courbe présente un plateau avec des fluctuations légères.
- Les valeurs de taux TCS sont plus grandes par rapport à celles du taux TCV quelques soit le nombre de composantes gaussiennes Ng. Ce résultat montre l'importance de la stratégie de la règle de vote dans le système RAG proposé.

III.4.3 Etude comparative entre les performances des classificateurs KNN et GMM en mode dépendant du locuteur

Cette section a pour objectif de comparer les performances du système RAG en mode dépendant du locuteur, basé sur le classificateur KNN et celles du système basé sur le classificateur GMM, en termes de taux de classification TCS et de temps de calcul de la phase de test.

Dans cette expérience, nous utilisons les meilleures configurations des classificateurs KNN et GMM, obtenues précédemment. Le tableau III.5 illustre le taux TCS et le temps d'exécution durant la phase d'apprentissage et la phase de test des deux classificateurs. Cependant, le temps d'exécution de la phase d'apprentissage du classificateur KNN n'est pas présenté, puisque cette phase exige seulement l'enregistrement des vecteurs de paramètres et n'a pas besoin d'une modélisation.

D'après les résultats du tableau III.5, nous pouvons remarquer que le système RAG basé sur le classificateur GMM présente les meilleures performances en termes de taux TCS et de temps

d'exécution durant la phase de test. Cependant, ce système exige plus de temps d'exécution pour la modélisation des modèles GMM des classes H et F, durant la phase d'apprentissage.

Tableau III.5: Performances des classificateurs KNN et GMM en mode dépendant du locuteur

Classificateur	Temps d'exécution (s)		TCS (%)	TCV (%)
	Apprentissage	Test		
KNN (Corrélation)	-	90.87	99.61	79.19
GMM	35.95	1.37	100	82.49

III.4.4 Performances du système RAG en mode indépendant du locuteur

Cette section a pour objectif d'étudier l'influence de fonctionnement du système RAG proposé en mode indépendant du locuteur, sur les performances du système. Les performances sont évaluées en utilisant la répartition de la base de données EMO-DB illustrée sur le tableau III.1.b.

Les performances des systèmes RAG basés sur les classificateurs KNN et GMM fonctionnant en mode indépendant du locuteur avec l'utilisation de leurs configurations précédentes, sont illustrées sur le tableau III.6.

Tableau III.6: Performances des classificateurs KNN et GMM en mode indépendant du locuteur

Classificateur	Temps d'exécution (s)		TCS (%)	TCV (%)
	Apprentissage	Test		
KNN (Corrélation)	-	90.21	92.80	67.13
GMM	31.90	0.80	97.26	75.43

Les résultats du tableau III.6 nous montre clairement le l'efficacité du classificateur GMM en termes des taux de classification TCV et TCS et également le temps d'exécution de la phase de

test. Ces résultats montrent également la détérioration des performances en mode indépendant du locuteur par rapport au mode dépendant.

III.5 Conclusion

Dans ce chapitre, nous avons présenté les deux systèmes RAG proposés basés sur les classificateurs KNN et GMM. Nous avons décrit les différentes étapes d'implémentation des deux systèmes, à savoir : extraction de paramètres MFCC, classification des vecteurs de paramètres basée sur les algorithmes KNN et GMM, ainsi que la prise de décision basée sur l'application de la règle de vote. De plus, nous avons présenté les expériences permettant de nous répondre sur les différentes questions posées, ainsi que la discussion sur les différents résultats obtenus. Les résultats nous ont montré que les deux systèmes RAG basés sur les classificateurs KNN et GMM présenté de bonne précision. Cependant, le système basé sur le classificateur GMM montre un bon compromis entre la précision et la complexité en termes de temps de calcul et d'espace mémoire.

IV. Conclusion générale

La reconnaissance automatique du genre consiste à reconnaître la classe du genre (F ou H) en utilisant différentes modalités telles que la parole, le visage, le signal vidéo, l'écriture, la biométrie. Cette reconnaissance trouve ces applications dans plusieurs domaines tels que : le marketing, la surveillance, la biométrie, la parole, le visage, l'écriture, ...etc.

Dans notre travail, nous avons conçu et implémenter deux systèmes RAG basé l'extraction de paramètres MFCC. Le premier système s'est basé sur le classificateur KNN combiné avec la stratégie de la règle de vote. Ce classificateur est couramment utilisé dans la reconnaissance de forme, vue de sa simplicité et sa mise en œuvre. Néanmoins, il exige plus d'espace mémoire et temps d'exécution durant la phase de test.

Le deuxième système s'est basé sur le classificateur GMM combiné avec la stratégie de la règle de vote. Ainsi, plusieurs expériences sont menées pour répondre sur les différentes questions posées au début du chapitre. Les résultats nous ont montré clairement que le classificateur GMM fonctionnant en mode dépendant du locuteur est plus performant par rapport au classificateur KNN en termes de précision et rapidité. Les résultats ont montré également la pertinence des paramètres énergétiques et les paramètres dynamiques pour la tâche RAG. De plus, les résultats en mode indépendant du locuteur ont montré l'influence de ce mode sur les performances du système. Ainsi, le classificateur GMM présente une robustesse supérieure dans la classification du genre.

Comme perspectives à notre travail nous proposons de :

- ✓ Utiliser et tester d'autres paramètres acoustiques tels que : LPCC (Linear Predective Cepstral Coefficients), PLP (Perceptual Linear Predctive) ;
- ✓ Utiliser d'autres classificateurs tels que SVM, ANN et CNN.
- ✓ Utiliser une base de données plus large.

Références

[1] E. Yücesoy, "Gender Recognition from Speech Signal Using CNN, KNN, SVM and RF," *Procedia Computer Science,* vol. 235, pp. 2251-2257, 2024.

[2] T. Jayasankar, K. Vinothkuma and V. Arputha , "Automatic Gender Identification in Speech Recognition by Genetic Algorithm," *Applied Mathematics & Information Sciences,* vol. 11, no. 3, pp. 907-913, 2017.

[3] R. a. Brunelli, "Automatic person recognition by acoustic and geometric features," *Machine Vis. Apps,* vol. 8, p. 317–325, September 1995.

[4] F. Hamidi and M. Branham , *Gender Recognition or Gender Reductionism? The Social Implications of Automatic Gender Recognition System,* p. 13, 2018.

[5] "Y. Zhuang, Feng Lin and Wenyao Xu, "Human Gender Classification: A Review',," pp. 275-300, 2017.

[6] Levitan, Taniya Mishra and Srinivas Bangalore;, "Automatic Identification of Gender from Speech Sarah Ita Srinivas Bangalore".

[7] A. Hacine-gharbi et P. Ravier, "« On the optimal number estimation of selected features using joint histogram based mutual information for speech emotion recognition »," *Journal of King Saud University,* vol. 33, no. 1074-1083, p. 9, 2021.

[8] G. Fant and s-Gavenhage, "Acoustic theory of speech production," Mouton, 1960.

[9] Abdelhak and Souadkia, "Reconnaissance automatique de la parole arabe: Approche évolutionnist," Guelma, 2010.

[10] H. Tebbi, *ranscription Orthographique Phonétique vue de la synthèse de la parole à partir du texte en l'Arabe StandardMémoire de Magister Spécialité : Ingénierie des systèmes et des connaissances,* USD-Blida, : Ecole Doctorale Informatique, Présenté à l'Université de Guelma, 2010., Juin 2007.

[11] L. Yves, "analyse spectrale de la parole," 2009.

[12] P. Nedungadi, J. Bhaskar and K. Sruthi , "« Hybrid approach for emotion classification of audio conversation based on text and speech mining »Procedia Computer Science,," 2015.

[13] S. Chaudhari and Kagalkar,R, "Methodology for gender identification, classification and recognition of human age," *nternational Journal of Computer Applications, 975, 8887.,* p. 10, 2015.

[14] s. a. zitouni saber, *Reconnaissance acoustique des émotions basée sur le classificateur KNN; Mémoire de Master,,* Bordj Bou Arreridj: Université de Mohamed El-Bachir El-Ibrahimi - , 2022.

[15] R. Aron, Indra Sigicharla, l, Mohanapr and Mohanapr, *SEGAA: A Unified Approach to Predicting Age, Gender, and Emotion in Speech,* Chennai, 2024.

[16] A. Shamim Banu, A. Venkatachalam and V. Kavitha, "A Novel Gender Recognition in Emotional Environment," *International Journal of Computer Applications,* vol. 115, 2021.

[17] V. A. S. B. Kavitha and S. Venkatachalam, "A Novel Gender Recognition in Emotional Environment," vol. 115, p. 15, 2015.

[18] Mohammad Amaz Uddin, Refat Khan and Md Sayem Hossain, Gender Recognition from Human Voice using Multi-Layer Architecture, Chittagong, Bangladesh,: Dept. of Computer Science and Engineering BGC Trust University , 2020.

[19] S. Chaudhary and D. Kumar Sharma, Gender Identification based on Voice Signal Characteristics, Greater Noida, India: Davendra Kumar Sharma Electronics & Communication Engineering Meerut Institute of Engineering & Technology Meerut, 2018.

[20] M. Sahib Shareef, S. Yaqeen Mezaal and Thulfiqar Abd , Gender voice classification with huge accuracy rate, vol. Vol. 18, TELKOMNIKA Telecommunication, Computing, Electronics and Control, 2020, pp. pp. 2612-2617.

[21] . Y. Ergün and V. N. Vasif , "Gender identification of a speaker using MFCC and GMM," in *8th International Conference on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey, 2013.

[22] A.Gruson, "Rapport de Stage Représentations vectorielles de distributions de probabilités pour la similarité musicale," 2009.

[23] Y.Attabi, "Mémoire Reconnaissance automatique des émotions à partir du signal acoustique ,École de technologie supérieure," 2008.

[24] M. JEDRA, *Conception et réalisation d'un système d'authentification par biométrie vocale,* Université Ibn Tofail, 2012.

[25] N.ZERARI, "INTÉGRATION D'UN MODULE DE RECONNAISSANCE DE LA PAROLE AU NIVEAU D'UN SYSTÈME AUDIOVISUEL – APPLICATION TÉLÉVISEUR," 2021.

[26] S. Young, The HTK book (HTK version 3.4) », Cambridge University Engineering Department,, Cambridge University Engineering Department, 2006.

[27] Munmun Biswas, Refat Khan Pathan, Mohammad Amaz Uddin and Md Sayem Hossain, "Gender Recognition from Human Voice using Multi-Layer Architecture.," in *Dept. of Computer Science and Engineering*, Chittagong, Bangladesh, 2020.

[28] F. Ghazali, A. Hacine gharbi and P. Ravier, "Selection of statistical Wavelet features using wrapper approach for Electrical Appliances identification based on KNN classifier combined with voting rules method," *International Journal of Computational System engineering,* 2022.

[29] A. bougrine, P. Ravier, A. Hacine gharbi and H. Ouachour, "LSTM Network based on Prosodic Features for the Classification of Injunction in French Oral Utterances," in *In Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2022)*, 2022.

[30] H. Bengacemi, A. Hacine-Gharbi, P. Ravier and K. a. B. O. Abed-Meraim, "Surface EMG Signal Classification for Parkinson's Disease using WCC Descriptor and ANN Classifier.," in *In Proceedings of the 10th International Conference on Pattern Recognition Aplications and Methods (ICPRAM ,*, 2021.