

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

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

Faculté des Sciences et de la technologie

Département d'électronique.

Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

FILIERE : Télécommunications

Spécialité : Systèmes des Télécommunications

Par

➤ **Benaldjia Abdelmoumene**

➤ **Boussouar Nihad**

Intitulé

***Extraction des paramètres statistiques pour la Reconnaissance acoustique
des émotions***

Soutenu le : 03/07/2023

Devant le Jury composé de :

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>DR : Bekkouche Tewfik</i>	<i>MCA</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>DR : Hacine gharbi Abdenour</i>	<i>MCA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Pr : Messali Zoubeida</i>	<i>Professeur</i>	<i>Examineur</i>	<i>Univ-BBA</i>
<i>DR : Ghezzali Fatah</i>	<i>DR</i>	<i>Co-Encadreur</i>	<i>Univ-BBA</i>

Année Universitaire 2022/2023

Remerciement

Tout d'abord, nous tenons à remercier 'ALLAH', le clément et le miséricordieux de nous avoir donné la force, la sante, la volonté, et le courage de mener à bien ce mémoire. Ce travail ne seras pas aussi riche et n'aurais pas pu avoir le jour sans l'aide et l'encadrement de **DR. HACINE GHARBI ABDENOUR**, en le remercie pour la qualité de son encadrement, pour sa patience, son aide pratique et sa disponibilité durant notre préparation de ce mémoire, Nos remerciements s'adressent à **DR. GHEZZALI FATAH** pour son aide pratique et son soutien moral et ses encouragements. Nous tenons aussi à exprimer nos remerciements aux membres du jury qui ont accepté d'évaluer ce modeste travail Nos remerciements aussi au département de l'électronique de l'université de Mohamed el Bachir el Ibrahim BBA enseignants et étudiants. Enfin, nous remercions nos collègues pour les motivations et les encouragements qu'ils nous ont donnés.

Dédicace

*Tout d'abord, je tiens à remercier **ALLAH***

De m'avoir donné la force et le courage de mener

À bien ce modeste travail.

Je tiens à dédier cet humble travail à :

*A ma tendre mère **OUANOUGHE- R** et mon très cher père **MALEK***

*A ma chère sœur que j'aime du fond du cœur **LINA***

*Et ma sœur **NESRINE***

*A mes frères: **ABDELDJALILE ; ABDELHAI***

A mes meilleurs amis :

HICHAME ; ABDELOUAHABE ; SIHEM ; KHAOULA

Abdelmoumene

Dédicace

Je voudrai dédier ce travail à 'ALLAH' tout puissant.

*À mon grand-père **Mohammed Tayeb** qui m'a élevé et m'a toujours motive a rechercher la connaissance, que dieu ait pitié de lui et le laisse entrer dans son vaste paradis.*

*À ma grand mère **Oum Noune** que dieu te donne une longue et joyeuse vie.*

À celui qui m'a enseigné à être tendre sans attendre, à celui à qui je suis fier de porter son nom, celui qui m'a appris et guidé jusqu'ici, celui qui travaille dur pour mon bonheur, je prie Dieu de vous accorder une longue vie et une bonne santé en reconnaissance des sacrifices que vous avez consentis pour nous , Mon cher père affectueux.

À mon ange dans ma vie, celui qui incarne l'amour et la compassion, celui qui se dévoue sans relâche, à la lumière qui éclaire mon chemin d'études, pour chaque goutte de sueur perlant sur son front, à ma chère maman bien-aimée.

À mes frères Akram et Wassime, a ma sœur Imane, pour leurs amour profond et leur soutien inconditionnel.

À tout mes amis spatialement

Aya Malek, Mebarka, khawla, Hadjer, zahwa.

Et à tous ceux qui j'aime, et qui m'aiment merci pour votre soutien.

Nihad

Résumé

Plusieurs travaux de recherche ont proposé un système de reconnaissance acoustique des émotions (RAE) basé sur le classificateur KNN combiné avec l'extraction des coefficients MFCC ainsi que la stratégie de règle de vote. Cependant, l'algorithme KNN est gourmand en temps de calcul et d'espace mémoire. Pour remédier ce problème, nous proposons de réduire le nombre de vecteurs de paramètres MFCC en un nombre limité de vecteurs de paramètres statistiques pour minimiser cette complexité. Notre contribution consiste à chercher la bonne configuration qui donne un taux élevé entre la précision et la complexité du système. Les résultats obtenus montrent la pertinence des paramètres statistiques des coefficients MFCC avec un taux de classification maximal égal à 82.17%.

Mots clés : Reconnaissance acoustique des émotions, coefficients MFCC, paramètres statistiques, classificateur KNN, stratégie de la règle de vote.

Abstract

Several research works have proposed acoustic emotion recognition (EAR) system based on the KNN classifier combined with the extraction of MFCC coefficients and the voting rule strategy. However, the KNN algorithm is computationally and memory intensive. To address this problem, we propose reducing the number of MFCC features vectors to a limited number of statistical features vectors to minimize this complexity. Our contribution lies in seeking the optimal configuration to achieve the best trade-off between system accuracy and complexity. The obtained results demonstrate the relevance of statistical parameters of MFCC coefficients, with a classification rate of 82.17%.

Keywords: Acoustic emotion recognition, MFCC coefficients, statistical parameters, KNN classifier, voting rule strategy.

ملخص

اقترحت العديد من الابحاث نظاما للتعرف على المشاعر الصوتية بالاعتماد على مصنف KNN و استخراج ميزات MFCC بالإضافة إلى استراتيجية قاعدة التصويت. ومع ذلك، يستهلك مصنف KNN الكثير من وقت للحساب و مساحة الذاكرة. لحل هذه المشكلة، نقترح نظام لتقليل عدد اشعة معاملات MFCC إلى عدد محدود من متجهات الاشعة الإحصائية لتقليل هذه التعقيد. تتمثل مساهمتنا في البحث عن الإعدادات المناسبة التي تسمح بالحصول على أفضل توازن بين الدقة والتعقيد في النظام. تُظهر النتائج التي تم الحصول عليها فعالية الأشعة الإحصائية لميزات MFCC بمعدل تصنيف يبلغ 82.17%.

الكلمات المفتاحية: التعرف على المشاعر، ميزات MFCC، الاشعة الإحصائية، مصنف KNN، استراتيجية قاعدة التصويت.

Table des matières

Introduction générale concept de base	1
Chapitre I : Etat de l'art sur la reconnaissance acoustique des émotions	
I.1 Introduction	3
I.2 Définitions.....	3
I.2.1 Emotions.....	3
I.2.2 Reconnaissance des émotions.....	3
I.2.3 Reconnaissance acoustique d'émotion	4
I.3 Classification des émotions.....	4
I.3.1 Émotions primaires.....	4
I.3.2 Émotions secondaires	6
I.4 Types d'émotions	6
I.4.1 Émotions positives.....	7
I.4.2 Émotions négatives.....	7
I.5 Architecture standard d'un système RAE	7
I.6 Méthodes d'extraction des paramètres.....	9
I.7 Méthodes de classification	9
I.7.1 Classificateur KNN	9
I.7.1.1 Algorithme de base du classificateur KNN	9
I.8 Récapitulatif sur les systèmes RAE	11
I.9 Conclusion.....	13
Chapitre II : Système de reconnaissance acoustique des émotions basé sur l'extraction des paramètres statistiques	
II.1 Introduction.....	16
II.2 Coefficients cepstraux sur l'échelle Mel (MFCC)	16
II.2.1 Etapes d'extraction des paramètres acoustiques MFCC	17
II.2.1 Prétraitement du signal.....	18
II.2.2 Fenêtrage	18
II.2.3 Analyse de Fourier à court terme	19

II.2.4 Bacs de filtres en échelle Mel	19
II.2.5 Calcul des énergies.....	19
II.2.6 Calcul du logarithme des énergies	19
II.2.7 Transformation en domaine fréquentiel	19
II.2.8. Vecteurs de paramètres MFCC	19
II.3 Calcul des paramètres (caractéristiques) dynamiques des MFCC.....	20
II.4 Paramètres statistiques.....	21
II.4.1 Définition de paramètre statistique	21
II.4.2 Etat de l'art sur les paramètres statistiques	21
II.4.3 Extraction des paramètres statistiques	22
II.4.3.1 Moyenne	23
II.4.3.2 écart-type	23
II.4.3.3 L'asymétrie	23
II.4.3.4 L'aplatissement	23
II.5 Travaux en relation	24
II.6 Système RAE proposé	24
II.7 Conclusion	25

Chapitre III :Implémentation et résultats

III.1 Introduction	29
III.2 Présentation de la base de données EMO-DB.....	30
III. 3 Implémentation du système RAE proposé	31
III.3.1 Phase d'apprentissage	31
III.3.1.1 Extraction des paramètres MFCC.....	32
III.3.1.2 Extraction des paramètres statistiques	33
III.3.1.3 Enregistrement des vecteurs de références.....	34
III.3.2 Phase de test.....	34
III.3.3 Classificateur KNN combiné avec la stratégie de règle de vote.....	35
III.3.4 Mesure de performances du système RAE proposé	35
III.4 Expériences et résultats	36
III.4.1 combinaison optimale des paramètres statistiques	36
III.4.2 Nombre optimal des vecteurs statistiques.....	37
III .4.3 Configuration optimale du classificateur KNN.....	40

III.4.4 Combinaison optimale des paramètres MFCC	41
III.4.5 L'effet d'utilisation des paramètres statistiques sur la complexité du système	42
III.5 Conclusion	43
Conclusion générale	44

Bibliographie

Liste des figures

Figure.I.1: Architecture d'un système RAE	8
Figure.II.1: Exemple de conversion des hertz en mel	17
Figure.II.2: Etapes de calcul d'un vecteur caractéristique de type MFCC	18
Figure.II.3: Calcul des dérivés premières et secondes de coefficients MFCC	43
Figure.II.4: Schéma du système RAE proposé.....	24
Figure.II.5: Système RAE basé sur l'extraction des paramètres statistiques.....	25
Figure.III.1: Schéma synoptique du système RAE proposé.....	31
Figure.III.2: Etape d'extraction des paramètres statistiques	34
Figure.III.3: Taux de classification TCV en fonction de nombre des vecteurs instantanés NVIS.....	40
Figure.III.4: Taux de classification TCS en fonction de nombre des vecteurs instantanés NVIS.....	40

Liste des tableaux

Tableau.I.1: Principales categories des emotions de base	5
Tableau. I.2: Mesure de similarite	10
Tableau.I.3: Les differentes methodes d'extraction de parametres	10
Tableau.II.1: Tableau de l'evolution des techniques d'extraction des parametres.....	22
Tableau III 1: Repartition de la base de donnees Emo-DB	31
Tableau III 2:Fichier de configuration (analysis.conf).....	32
Tableau III 3: Meilleure combinaison des parametres statistiques.....	36
Tableau III 4: Taux de classification TCV et tcs en fonction du nombre des vecteurs instantanes NVIS.....	37
Tableau III 5: Configuration optimale du classificateur knn.....	41
Tableau III 6: Taux de classification pour differentes combinaisons de parametres MFCC	42
Tableau III 7: Comparaison des performances des systemes RAE	43

Liste des abréviations

ADI: Average Demand Interval.
ANN: Artificial Neural Network.
DCT: Discrète Cosine Transform.
ECG : L'électrocardiogramme.
EEG: Electroencephalogram.
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.
KNN: k-Nearest Neighbors.
LDA : Linear Discriminant Analysis.
LPC: Linear Predictive Coding.
LPCC: Linear Prediction Cestrum Coefficients.
MFCC: Mel-Frequency Cepstral coefficients.
NLP: Natural Language Processing.
NVIS : Nombre de vecteurs instantanés par ce séquence.
PLP: Personal Learning Plan.
RL : Linear Regression.
RNN: Recurrent Neural Networks.
SVM: Support vector machine.
TCS: Taux de classification des signaux.
TCV: Taux de classification des vecteurs.

.



Introduction générale



Introduction générale

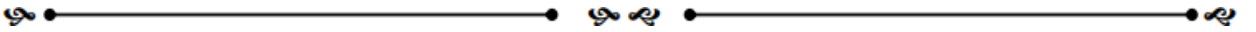
Introduction générale

Les émotions sont des réactions complexes qui sont déclenchées par des stimuli externes ou internes. Elles peuvent être positives, telles que la joie, l'amour et l'excitation, ou négatives, comme la colère, la peur et la tristesse. Les émotions nous permettent de communiquer avec les autres, de transmettre des sentiments et de comprendre des opinions [1].

La reconnaissance automatique des émotions (RAE) est utilisée dans plusieurs domaines, tels que la psychologie, marketing, sécurité ...etc. Les émotions sont souvent considérées comme des réponses subjectives à des événements ou des stimulus qui activent nos systèmes cognitifs, comportementaux et physiologiques [2]. Les systèmes RAE peuvent se baser sur plusieurs modalités, telles que l'image faciale, la parole, le signal EEG (électro-encéphalogramme), le signal ECG (électro-cardio-gramme)...etc. Dans notre travail, on s'intéresse à la reconnaissance acoustique des émotions à partir du signal parole.

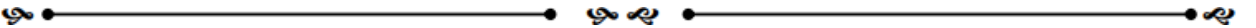
La conception d'un système RAE se base sur une phase d'apprentissage pour apprendre les différents modèles de classes et une phase de test pour identifier les signaux d'entrées et évaluer les performances du système RAE. Ces deux phases exigent une étape d'extraction de paramètres. Cette dernière consiste à convertir chaque signal d'entrée en une séquence de vecteurs de paramètres acoustiques tels que les coefficients MFCC (Mel-Frequency Cepstral coefficients), PLP et LPCC. Le fonctionnement du système RAE se déroule en une étape d'extraction et une étape de classification en se basant sur un des classificateurs standard tels que KNN, GMM, ANN, SVM, HMM, ADI, etc. Dans [3], les auteurs ont implémenté un système RAE basé sur le classificateur KNN et l'extraction des paramètres MFCC combinés avec la stratégie de règle de vote. Cependant, le classificateur KNN exige plus d'espace mémoire et temps de calcul. Notre objectif est de minimiser cette complexité en réduisant le nombre de vecteurs de paramètres en un nombre limité de vecteurs de paramètres statistiques. Notre contribution consiste à chercher la bonne configuration permettant d'obtenir le meilleur compromis entre la précision et la complexité du système.

Ce manuscrit est organisé en trois chapitres. Le premier chapitre est consacré à l'état de l'art sur la reconnaissance acoustique des émotions. Le deuxième chapitre est dédié à l'extraction des paramètres statistiques des coefficients MFCC. Le troisième chapitre présente les résultats du système proposé. En fin, nous terminons le manuscrit par une conclusion générale et des perspectives.



Chapitre I

Etat de l'art sur la reconnaissance acoustique des
émotions



I.1 Introduction

Avec la complexité permanente de la vie sociale, les émotions jouent un rôle crucial dans la définition et la régulation de nos relations avec les autres. Nos réponses émotionnelles à d'autres personnes ont une certaine incidence sur la façon dont ces autres personnes réagissent à nous et dans une certaine mesure [4]. La plupart des systèmes de reconnaissance automatique des émotions se basent sur l'analyse des expressions faciales ou l'analyse du signal parole en tenant compte plusieurs catégories émotionnelles, telles que le bonheur, la tristesse, colère, la surprise, la peur et le dégoût.

L'objectif de notre travail est de concevoir un système automatique de reconnaissance des émotions (RAE) à partir du signal parole. Avant de pouvoir concevoir ce système, il est nécessaire de comprendre le phénomène émotionnel ainsi que le fonctionnement d'un système RAE. Ce chapitre vise à fournir des généralités sur les émotions, leurs différents types et le fonctionnement d'un système RAE.

I.2 Définitions

I. 2.1 Emotions

L'émotion est un état de motivation constitué d'impulsions, de désirs ou d'aversion ; elles impliquent des changements de motivation. Ces émotions poussent la personne à modifier sa relation avec un objet, l'état du monde, son propre état, ou à entretenir une relation. Un processus émotionnel comprend des composantes de réaction, de réflexion, d'apprentissage et de mémorisation [5].

I.2.2 Reconnaissance des émotions

Le processus d'identification des émotions humaines est connu sous le nom de reconnaissance des émotions. La majorité des travaux à ce jour se sont concentrés sur l'automatisation de la reconnaissance des émotions faciales à partir de la vidéo, des expressions parlées à partir de l'audio, des expressions écrites à partir du texte et des mesures physiologiques effectuées par des appareils spéciaux. La voix, la posture, les gestes, la

direction des yeux et les expressions faciales sont tous des moyens par lesquels les gens peuvent transmettre leurs émotions [6].

I.2.3 Reconnaissance acoustique d'émotion

La reconnaissance acoustique d'émotions est une technique qui permet d'analyser les caractéristiques acoustiques d'un signal audio pour déterminer l'émotion. Cette technique est souvent utilisée en traitement automatique du langage naturel, en particulier pour l'analyse de la parole.

Les émotions peuvent être détectées à partir des différents indices acoustiques tels que la hauteur, la durée, le débit, le volume, le rythme, le timbre et l'intensité de la voix. Par exemple, un ton de voix haut et un débit rapide peuvent indiquer de la colère, tandis qu'un ton plus bas et une cadence plus lente peuvent indiquer de la tristesse ou de la dépression.

Pour reconnaître les émotions à partir de signaux audio, on peut utiliser des techniques de traitement automatique du langage naturel (NLP : Natural Language Processing) telles que les réseaux de neurones, les arbres de décision, les méthodes de classification bayésienne, etc. Ces techniques utilisent des algorithmes sophistiqués pour extraire des caractéristiques acoustiques du signal audio, puis les classer en fonction des émotions qu'elles représentent [7].

La reconnaissance acoustique d'émotion peut être utilisée dans de nombreux domaines, tels que la reconnaissance de la parole, l'analyse de sentiments et la surveillance de la santé mentale. Par exemple, dans le domaine de la santé mentale, la reconnaissance acoustique d'émotion peut aider à détecter les signes précurseurs de la dépression ou de l'anxiété chez les patients en analysant leur voix lorsqu'ils parlent [7].

I.3 Classification des émotions

Les émotions peuvent être divisées en deux classes, à savoir les émotions primaires et les émotions secondaires.

I.3.1 Émotions primaires

Les émotions primaires sont considérées comme des émotions de base, innées et universelles. Elles sont présentes chez tous les individus, quel que soit leur contexte culturel.

Les émotions sont classées en 7 catégories comme suit : la tristesse, le bonheur, la surprise, le dégoût, la peur, la neutralité et la colère [8]. Le tableau I.1 regroupe les principales catégories des émotions de base:

Tableau I.1:Principales catégories des émotions de base

Référence	Emotions
[9]	Colère, dégoût, joie, tristesse, peur, surprise, neutralité.
[10]	Acceptation, colère, anticipation, dégoût, peur, joie, tristesse, surprise
[4]	Intérêt, joie, désir, chagrin, émerveillement
[11]	Colère, intérêt, mépris, dégoût, détresse, peur, joie, honte, surprise
[12]	Courage, colère, aversion, désir, désespoir, tristesse, amour, espoir, abattement, haine, peur
[13]	Souffrance, plaisir
[14]	Colère, peur, joie, tristesse, dégoût

- **Joie** : est une émotion positive est liée à des expériences agréables ou à la satisfaction de besoins [15].
- **Tristesse** : est une émotion associée à la perte, la séparation ou la détresse émotionnelle [15].
- **Peur** : est une émotion de survie se déclenche face à une menace ou un danger potentiel [15].
- **colère** : est une émotion liée à la frustration, l'injustice ou la violation de nos limites [15].
- **dégoût** : est une émotion négative est ressentie face à des stimuli répugnants ou repoussants [15].
- **surprise** : est une réaction émotionnelle face à un événement inattendu ou surprenant [15].
- **Neutralité** : est un état non émotionnel, comprenant le calme, la relaxation, la satisfaction, le contentement, la somnolence etc. [16].

Ces émotions primaires sont souvent considérées comme les blocs de construction fondamentaux des expériences émotionnelles.

I.3.2 Émotions secondaires

Les émotions secondaires, également appelées émotions sociales ou complexes, sont considérées comme des combinaisons des émotions primaires. Elles peuvent être influencées par des facteurs tels que la culture, l'éducation, les normes sociales et les expériences individuelles [17]. Les émotions secondaires peuvent inclure:

- **amour** : est une émotion complexe qui peut découler de la connexion émotionnelle, de l'attachement ou de l'affection envers quelqu'un [18].
- **culpabilité** : est une émotion qui survient lorsque nous nous sentons responsables ou blâmables pour nos actions ou leurs conséquences [18].
- **jalousie** : est une émotion qui se manifeste lorsque nous ressentons une menace pour une relation ou une possession que nous considérons comme importante [18].
- **honte** : est une émotion liée à la perception d'une violation des normes sociales ou à une atteinte à l'estime de soi [18].
- **envie** : est une émotion ressentie lorsque nous désirons ce que possède quelqu'un d'autre [18].

Les émotions secondaires sont souvent plus complexes et peuvent varier considérablement d'une personne à l'autre en fonction de leur histoire personnelle et de leur environnement culturel [18].

I.4 Types d'émotions

Effectivement, dans de nombreuses approches de la psychologie des émotions, on distingue généralement deux grands types d'émotions : les émotions positives et les émotions négatives. Ces catégories permettent de classer les différentes expressions émotionnelles en fonction de leur valence émotionnelle, c'est-à-dire de leur caractère agréable ou désagréable.

I.4.1 Émotions positives

Les émotions positives sont associées à des expressions agréables, plaisantes ou gratifiantes. Elles ont tendance à générer des sensations de bien-être, de satisfaction ou de bonheur, Ces émotions positives jouent un rôle important dans notre bien-être émotionnel et contribuent à renforcer nos relations sociales, notre motivation et notre résilience face aux difficultés [19].

I.4.2 Émotions négatives

Les émotions négatives sont associées à des expressions désagréables, pénibles ou stressantes. Elles peuvent résulter de situations perçues comme menaçantes, frustrantes ou tristes, Ces émotions négatives peuvent jouer un rôle adaptatif en nous aidant à réagir aux situations stressantes, à nous protéger ou à nous mobiliser pour résoudre les problèmes.

Il est important de noter que les émotions positives et négatives ne sont pas mutuellement exclusives [19].

I .5 Architecture standard d'un système RAE

Le système de reconnaissance acoustique d'émotions est un système qui utilise des techniques de traitement de signal ainsi que des algorithmes d'apprentissage automatique pour identifier l'état émotionnel d'une personne à partir de sa voix.

L'architecture standard d'un système RAE est illustrée sur la figure (I.1). Cette figure montre les étapes nécessaires de fonctionnement d'un système RAE. Cette architecture peut être variée selon les applications et les algorithmes utilisés dans chaque étape du système. Les étapes seront décrites dans les sous sections suivantes.

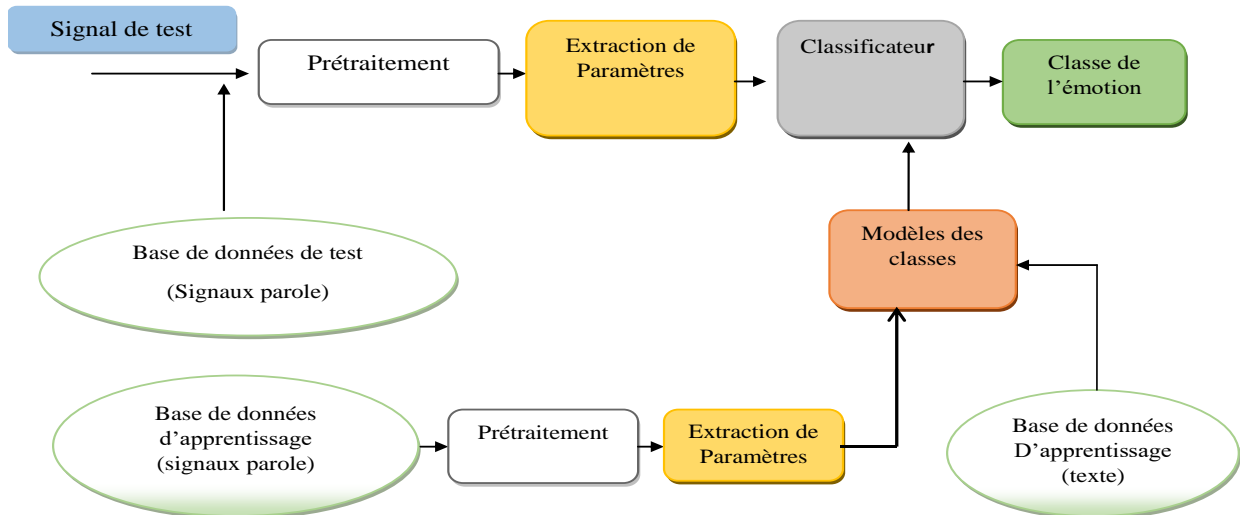


Figure I 1: Architecture d'un système RAE [20]

- **Acquisition du signal**

Le signal sonore est capté à l'aide d'un microphone ou d'un autre dispositif d'enregistrement audio [21].

- **Prétraitement du signal**

Le signal brut est ensuite filtré, amplifié et éventuellement échantillonné pour obtenir un signal numérique de bonne qualité [21].

- **Extraction de paramètres**

Les paramètres acoustiques discriminants sont extraits du signal audio numérique à l'aide d'algorithmes de traitement du signal tels que la transformée de Fourier ou la transformée en ondelettes [21],..., etc.

- **Modélisation**

Les paramètres extraits sont ensuite utilisés pour créer un modèle acoustique de chaque classe d'émotion qui peut être utilisé pour reconnaître des sons similaires en classes d'émotions [21].

- **Reconnaissance**

L'étape de reconnaissance consiste à appliquer des algorithmes de classification tels que KNN, ANN et GMM, sur la séquence des vecteurs de paramètres MFCC correspond au signal d'entrée en utilisant les modèles de classes préalablement obtenus dans la phase d'apprentissage [21].

I.6 Méthodes d'extraction des paramètres

L'étape d'extraction de paramètres consiste à convertir chaque signal en une séquence de vecteurs de paramètres caractérisant les différentes classes en utilisant des méthodes de traitement du signal parole, à savoir : l'analyse de Fourier, L'analyse par ondelettes, l'analyse cepstrale. Plusieurs types de paramètres sont utilisés dans les systèmes RAE, tels que : les paramètres MFCC, PLP et LPCC. Les paramètres couramment utilisés sont les coefficients MFCC [20].

I.7 Méthodes de classification

Il existe plusieurs méthodes de classification, telles que la régression linéaire (RL), les machines vectorielles de support (SVM), K les plus proches voisins (KNN), les modèles de chaîne de Markov cachées (HMM), le modèle de mélange de Gaussiennes (GMM), les réseaux neuronaux artificiels(ANN) et les réseau de neurones récurrents. (RNN) [22].

Le classificateur KNN est généralement très utilisé dans les systèmes de reconnaissance de formes, vue de sa simplicité et son efficacité [[20] [23] [3]].

I.7.1 Classificateur KNN

L'algorithme K plus proche voisin (KNN pour K-Nearest Neighbors en anglais) est une méthode d'apprentissage automatique supervisé utilisé pour la classification et la régression. L'idée fondamentale derrière le classificateur KNN est de trouver les K échantillons d'entraînement les plus proches de l'échantillon de test à classer, en fonction d'une mesure de similarité, telle que la distance euclidienne. La classe majoritaire parmi celles des K échantillons est ensuite attribuée à l'échantillon à classer [24].

Le classificateur KNN est relativement simple à mettre en œuvre et ne nécessite pas d'apprentissage préalable. Cependant, il peut être sensible à la dimensionnalité des données et peut devenir coûteux en termes de temps de calcul pour de grands ensembles d'entraînement. De plus, il n'apprend pas explicitement de modèle et ne fournit pas d'interprétation des caractéristiques. Néanmoins, il reste un algorithme populaire et efficace pour de nombreux problèmes de classification [24]. Le tableau I.2 représente quelques exemples de distances pour ce dernier.

Tableau I. 2: Mesure de similarité

Distances	Formules
Euclidienne	$De(X, Y) = \sqrt{\sum_{i=1}^D (x_i - y_i)^2}$
Cityblock	$D_{\text{cityblok}}(X, Y) = \sum_{i=1}^d (x_i - y_i)$
Cosinus	$D_{\text{cos}}(X, Y) = 1 - \frac{\sum_{i=1}^d x_i y_i}{\sqrt{\sum_{i=1}^d x_i^2} \sqrt{\sum_{i=1}^d y_i^2}}$
Corrélation	$D_{\text{corr}}(X, Y) = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^d (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^d (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^d (y_i - \bar{y})^2}} \right)$
Hamming	$D_{\text{ham}}(X, Y) = \sum_{i=1}^L 1_{x_i \neq y_i}$

I.7.2 Algorithme de base du classificateur KNN

L'algorithme de base du classificateur KNN est donné comme suit [25] :

Début d'algorithme :

- Introduire les entrées ;

- vecteurs de paramètres d'apprentissage dans une matrice MA,
- vecteurs de paramètres de test dans une matrice MT,
- étiquettes de classes des vecteurs d'apprentissage dans un vecteur CA,

- étiquettes de classes des vecteurs de test dans un vecteur CT,
- la valeur de K,
- la distance utilisée (choisie).

- Pour chaque vecteur VTi de (MT):

1. calculer les distances entre chaque vecteur de test VTi et tous les vecteurs de la base d'apprentissage MA,
2. Chercher les K vecteurs de la base d'apprentissage MA les plus proche au VTi en termes de distance choisie.
3. Attribuer au vecteur VTi la classe la plus dominante ou la plus votée,
4. Refaire l'étape 1, 2 et 3 pour tous les vecteurs de la base de test MT.

Fin de l'algorithme.

I.8 Récapitulatif sur les systèmes RAE

Le tableau I.3 illustre un récapitulatif sur les systèmes RAE en présentant les méthodes d'extraction, les classificateurs, les bases de données utilisées et le taux de classification.

Tableau I. 3: les différentes méthodes d'extraction de paramètres

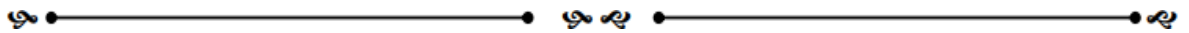
Référence	Année	La base des données	Extraction des paramètres	Méthode de classification	Résultat CR%
[26]	2014	Emo-DB	Energie + MFCC + ZCR + Probabilité de vocalisation + Fréquence fondamentale	FCBF + MVT	84.19
[27]	2017	Spanish	MFCC + MS	RNN	90.05
[28]	2018	Emo-DB	GCZCMT	SVM	84.45
[29]	2009	Emo-DB	Prosodique + sous-bande + MFCC + LPC	SFS algorithme + SVM	83.00
[30]	2018	eNTERFACE'05	Energie + Pitch + formants, intensité +	SVM	90.00

Chapitre I : Etat de l'art sur la reconnaissance acoustique des émotions

			ZCR + texte		
[31]	2018	CASIA	Prosodique + les caractéristiques de qualité + MFCC	analyse de corrélation + Fisher + arbre de décision ELM	89.60
[32]	2019	IEMOCAP	IS10 + MFCCs + eGemaps+SoundNet + VGGish	SVM	64.00
[33]	2015	Emo-DB	Caractéristiques prosodiques + caractéristiques paralinguistiques	SVM	94.90
[34]	2018	CASIA	MFCC	GA-BEL + PCA + LDA	90.28
[35]	2017	CASIA	MFCC, pitch, énergie à court terme + ZCR	SVM + DBN	95.80
[36]	2015	Emo-DB	eGeMAPS	Système d'empilage CSS +SVM	82.45
[37]	2019	Emo-DB	MFCCs	ensemble de SVMs	92.45
[38]	2016	RAVDESS	CWT, les coefficients prosodiques	SVM	60.10
[39]	2019	Spanish	SMFCC, ECC, MFF, MS et EFCC	RNN	91.16
[40]	2020	RAVDESS/SAVEE	Prosodique + spectrale	RDF	99.55
[41]	2019	Emo-DB	MFCC, PLP, pitch, Energie, LPCC	GMM	86.05

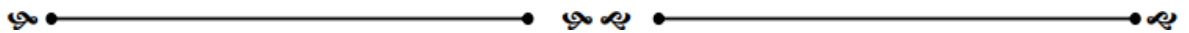
I.9 Conclusion

Dans ce chapitre, nous avons présenté des généralités sur la reconnaissance acoustique des émotions. Egaleme nt, nous avons décrit brièvement les différentes phases et étapes de fonctionnement d'un système RAE. Plus particulièrement, nous avons présenté le classificateur KNN.



Chapitre II

Systeme de reconnaissance acoustique des émotions
basé sur l'extraction des paramètres statistiques



II.1 Introduction

Le fonctionnement d'un système de reconnaissance acoustique des émotions (RAE) se déroule en deux phases principales : une phase d'apprentissage et une phase de test. Ces deux phases nécessitent une étape d'extraction de paramètres qui consiste à convertir chaque signal en une séquence de vecteurs de paramètres. Ensuite, cette séquence est classifiée par un des classificateurs standards pour déterminer la classe de l'état émotionnel correspondante au signal vocal d'entrée. Les paramètres extraits fournissent des informations précieuses sur les caractéristiques acoustiques de la voix, telles que la fréquence, l'intensité et la durée des sons émis [42].

Dans ce chapitre, nous présenterons les méthodes couramment utilisées pour l'extraction des paramètres pertinents du signal vocal dans le cadre de la reconnaissance automatique des émotions (RAE), notamment la méthode basée sur le descripteur MFCC.

II.2 Coefficients cepstreaux en l'échelle Mel (MFCC)

Les coefficients MFCC (Mel-Frequency Cepstral Coefficients) sont une représentation numérique des caractéristiques spectrales d'un signal audio. Ils sont largement utilisés dans le domaine du traitement automatique de la parole et de la reconnaissance vocale. Les coefficients MFCC sont calculés à partir d'une série d'étapes. Tout d'abord, le signal audio est divisé en petits segments temporelles. Ensuite, pour chaque segment, la transformation de Fourier est appliquée pour obtenir le spectre de puissance. Cependant, plutôt que d'utiliser une échelle linéaire des fréquences, les fréquences sont converties en une échelle non linéaire, laquelle approxime la façon dont l'oreille humaine perçoit les différentes fréquences. Cette échelle est appelée échelle de Mel. Ensuite, le logarithme du spectre de puissance est pris et des filtres triangulaires sont appliqués pour capturer les bandes de fréquences importantes. Les coefficients MFCC sont finalement obtenus en effectuant une transformation en cosinus discrète (DCT) sur les valeurs du logarithme des puissances spectrales filtrées. Les coefficients MFCC sont utilisés pour représenter les caractéristiques distinctives du signal audio et sont

particulièrement efficaces pour la reconnaissance de la parole et la classification des locuteurs. [43] . La fréquence Mel-échelle est définie par :

$$B(f) = 2595 \log_{10} (1 + f/700) \quad [43]$$

Où f est la fréquence en Hz, (B) est la fréquence en échelle Mel de f . la figure II.1 présente un exemple de conversion d'hertz en Mel.

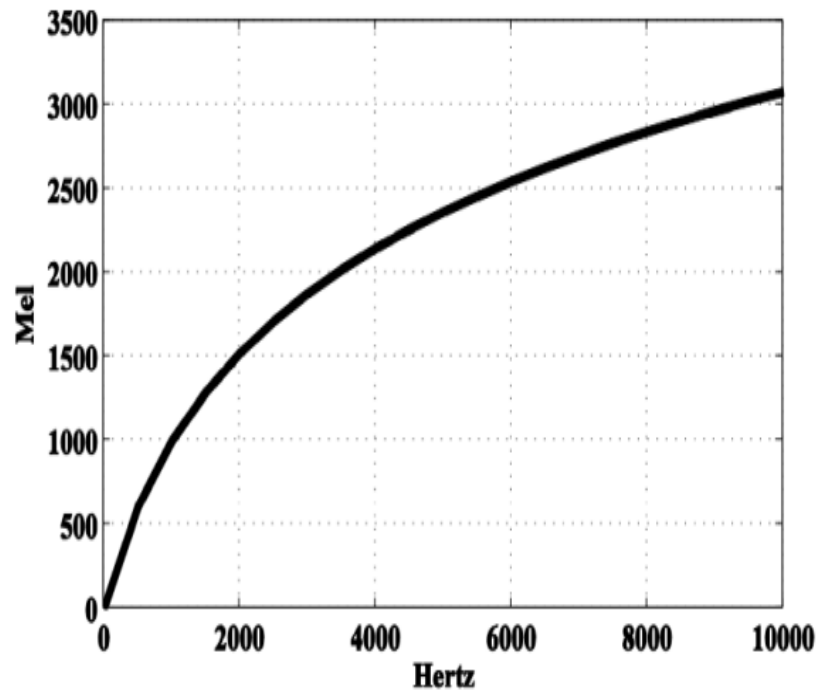


Figure II. 1:Exemple de conversion des hertz en Mel [44]

II.2.1 Etapes d'extraction des paramètres acoustiques MFCC

L'extraction des paramètres acoustiques MFCC se déroule généralement en plusieurs étapes telles qu'illustrées sur la Figure II.2

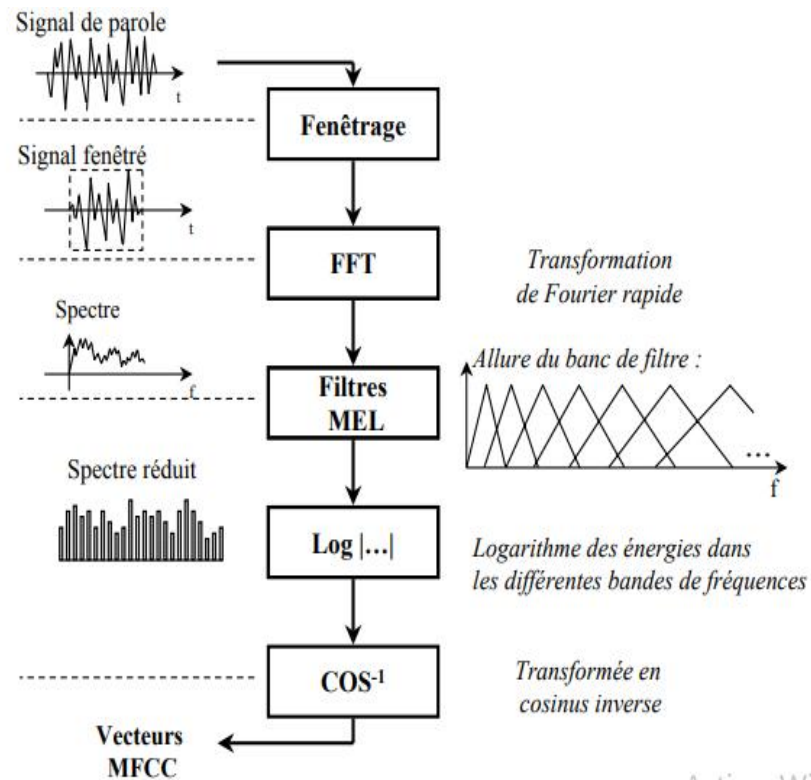


Figure II. 2: Étapes de calcul d'un vecteur caractéristique de type MFCC [45]

II.2.1.1 Prétraitement du signal

Tout d'abord, le signal audio brut est prétraité pour éliminer les bruits indésirables et les artefacts. Cela peut impliquer des étapes telles que le filtrage, la normalisation de l'amplitude, la suppression des silences, etc [46].

II.2.1.2 Fenêtrage

Le signal audio est divisé en petites trames (ou fenêtres) de taille fixe, généralement de 20 à 40 millisecondes. Une fenêtre typiquement utilisée est la fenêtre de Hamming, qui atténue les effets des bords lors du découpage du signal [46].

II.2.1.3 Analyse de Fourier à court terme

Pour chaque fenêtre, la transformée de Fourier discrète est calculée pour obtenir une représentation fréquentielle du signal [47].

II.2.1.4 Bancs de filtres en échelle Mel

Les filtres de bancs de filtres Mel sont utilisés pour extraire les caractéristiques spectrales du signal. Ces filtres sont espacés de manière non linéaire sur l'échelle de Mel, qui est une échelle de fréquence perceptuelle. Le nombre de filtres dépend généralement de l'application, mais 20 à 40 filtres sont couramment utilisés [47].

II.2.1.5 Calcul des énergies

Pour chaque filtre Mel, l'énergie spectrale est calculée en sommant les amplitudes des coefficients de la transformée de Fourier à court terme qui sont couverts par chaque filtre. [47]

II.2.1.6 Calcul du logarithme des énergies

Les énergies sont ensuite converties en échelle logarithmique. Cela est fait pour approximer la sensibilité de l'oreille humaine aux variations de fréquence. [48]

II.2.1.7 Transformation en domaine fréquentiel

La transformation en cosinus discrète (DCT) est appliquée sur les énergies logarithmiques pour obtenir les coefficients cepstraux. La DCT est similaire à la transformée de Fourier, mais elle produit des coefficients réels et compacte l'information en se concentrant sur les parties basses fréquences [49].

II.2.1.8 Vecteurs de paramètres MFCC

Les premiers coefficients cepstraux obtenus après la DCT sont les MFCC. Selon l'application, un certain nombre de coefficients MFCC est sélectionné (par exemple, les 12 premiers) et utilisé comme des descripteurs pour représenter le signal audio [50].

Ces étapes représentent un flux de travail de base pour l'extraction des paramètres acoustiques MFCC. Il existe également des variantes et des améliorations possibles, comme l'inclusion des coefficients différentiels (dynamiques) de premier et deuxième ordre, appelés respectivement delta (Δ) et delta delta ($\Delta\Delta$).

II.3 Calcul des paramètres (caractéristiques) dynamiques des MFCC

Les paramètres dynamiques des MFCC sont calculés pour capturer les variations temporelles des caractéristiques acoustiques d'un signal vocal. Pour calculer ces paramètres, on utilise généralement les premières et les deuxièmes dérivées des MFCC. Les premières dérivées (Δ) représentent la vitesse de variation des MFCC, tandis que les deuxièmes dérivées ($\Delta\Delta$) reflètent l'accélération des changements. Les paramètres dynamiques des MFCC peuvent être utilisés pour améliorer la robustesse de la reconnaissance de la parole en capturant les transitions rapides et les variations d'intensité dans le signal vocal. Ces informations supplémentaires peuvent aider à mieux représenter les caractéristiques temporelles de la parole et à améliorer les performances des systèmes de reconnaissance automatique de la parole. Soit $C_k(t)$ le coefficient cepstral d'indice k de la trame t , alors le coefficient différentiel de premier ordre $\Delta C_k(t)$ peut être calculé sur T trames en utilisant la formule donnée par [51] :

$$\Delta C_k(t) = \frac{\sum_{i=1}^T i (C_k(t+i) - C_k(t-i))}{2 \cdot \sum_{i=1}^T i^2}$$

De la même manière, les coefficients différentiels du second ordre $\Delta\Delta$ sont calculés à partir des coefficients du premier ordre.

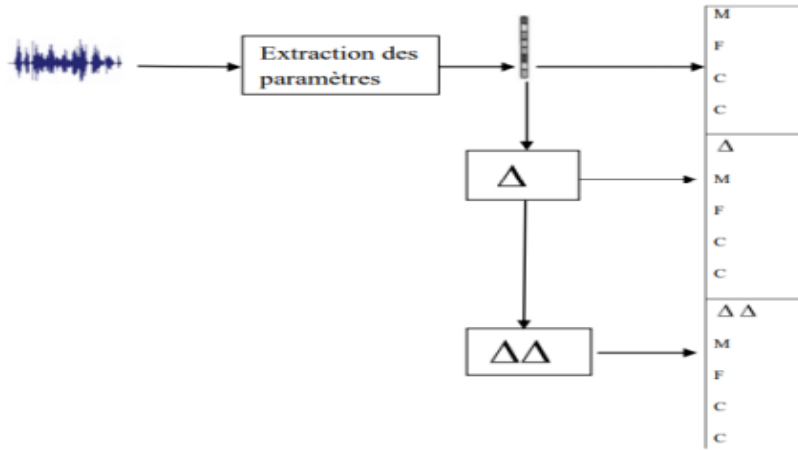


Figure II. 3: Calcul des dérivées premières et secondes de coefficients MFCC [44]

Dans ce travail, on s'intéresse à appliquer cette méthode d'analyse acoustiques pour extraire les coefficients MFCC, l'énergie et leurs paramètres dynamiques en utilisant la commande HCopy de librairie de HTK [50]. Ces coefficients seront extraits pour la tâche de reconnaissance acoustique des émotions.

II.4 Paramètres statistiques

II.4.1 Définition de paramètre statistique

Les paramètres statistiques sont des mesures numériques utilisées pour résumer et décrire les caractéristiques d'un ensemble de données. Ils fournissent des informations quantitatives sur la tendance centrale, la dispersion et la forme de la distribution des valeurs, parmi les paramètres statistique il y a la moyenne, l'écart-type, les moments statistiques, la variance, les enveloppes temporelles...etc [52].

II.4.2 Etat de l'art sur l'extraction des paramètres statistiques

L'extraction des paramètres statistiques dans le domaine de la reconnaissance acoustique des émotions est un domaine de recherche relativement récent qui s'est développé au cours des dernières décennies. Le tableau II.1 présente un historique de cette évolution :

Chapitre II : Système RAE basé sur l'extraction des paramètres statistiques

Tableau II. 1: Tableau de l'évolution des techniques d'extraction des paramètres

Année	Référence	L'historique
2000	[53]	Les recherches se sont intensifiées, avec un accent particulier sur l'utilisation de techniques d'apprentissage automatique pour la classification des émotions. Les paramètres statistiques, tels que la moyenne, l'écart-type et les quantiles, ont été utilisés pour résumer les caractéristiques acoustiques extraites des enregistrements vocaux.
2010	[54]	L'exploration des paramètres statistiques s'est étendue pour inclure des mesures plus avancées, telles que l'entropie spectrale, les coefficients cepstraux, les modèles de Markov cachés et les réseaux de neurones profonds. Ces avancées ont permis d'améliorer la précision de la reconnaissance des émotions à partir des signaux acoustiques.
2020	[55]	Les recherches dans le domaine de la reconnaissance acoustique des émotions ont continué à se développer, en explorant de nouvelles techniques et en intégrant des approches multimodales. Les paramètres statistiques ont été combinés avec d'autres informations, comme les expressions faciales et les gestes, pour améliorer la précision et la compréhension des émotions.

Au fur et à mesure que le domaine de la reconnaissance acoustique des émotions progresse, de nouvelles techniques d'extraction de paramètres statistiques et d'apprentissage automatique continuent d'émerger, permettant une meilleure compréhension et une meilleure classification des émotions à partir des signaux acoustiques. Cette évolution contribue à des applications pratiques telles que la détection d'émotions dans les systèmes de dialogue homme-machine, la surveillance de la santé mentale et l'analyse des réactions des utilisateurs dans le domaine du marketing et de la publicité [56].

II.4.3 Extraction de paramètres statistiques

L'extraction des paramètres statistiques pour la reconnaissance acoustique des émotions est une méthode couramment utilisée pour analyser les signaux sonores et identifier les émotions exprimées par une personne. Les paramètres statistiques extraits à partir de ces signaux fournissent des informations quantitatives qui aident à caractériser les différentes émotions, parmi les paramètres statistiques utilisés, on retrouve :

II.4.3.1 Moyenne

Elle représente la valeur centrale d'un ensemble de données V représentant les valeurs d'un coefficient MFCC, calculée en sommant toutes les valeurs d'indice r et en les divisant par le nombre total des valeurs R [57]

$$\mu = \frac{1}{R} \sum_{r=1}^R V(r)$$

II.4.3.2 Écart-type

Il mesure la dispersion des valeurs autour de la moyenne. Un écart-type plus élevé indique une plus grande variabilité des données. [57]

$$\delta = \sqrt{\frac{1}{R} \sum_{r=1}^R (V(r) - \mu)^2} \quad [23]$$

II.4.3.4 L'asymétrie

L'asymétrie est un concept qui décrit une situation où les éléments d'un système ne sont pas équilibrés ou équivalents, créant ainsi une disparité ou une différence notable entre eux. [57]

$$SK = \frac{\frac{1}{R} \sum_{r=1}^R (V(r) - \mu)^3}{(\delta)^3}$$

II.4.3.5 L'aplatissement [23]

L'aplatissement statistique est un phénomène qui se produit lorsque la distribution des données devient plus uniforme, avec une réduction des écarts entre les valeurs extrêmes, ce qui peut indiquer une diminution de la variabilité dans un ensemble de données [57].

$$Ku = \frac{\frac{1}{R} \sum_{r=1}^R (V(r) - \mu)^4}{(\delta)^4}$$

Ces paramètres statistiques permettent de comprendre et de résumer les propriétés essentielles d'un ensemble de données, facilitant ainsi leur analyse et leur interprétation. Dans notre travail nous recherchons la meilleure combinaison des paramètres statistiques pour obtenir de meilleures performances.

II.5 Travaux en relation

Dans [3] les auteurs ont proposé un système RAE basé sur l'extraction des paramètres MFCC et le classificateur KNN combiné avec la stratégie de règle de vote. Cette combinaison est proposée dans le domaine d'identification des appareils électrique par [23]. Le système proposé dans [3] est illustré sur Figure II.4:

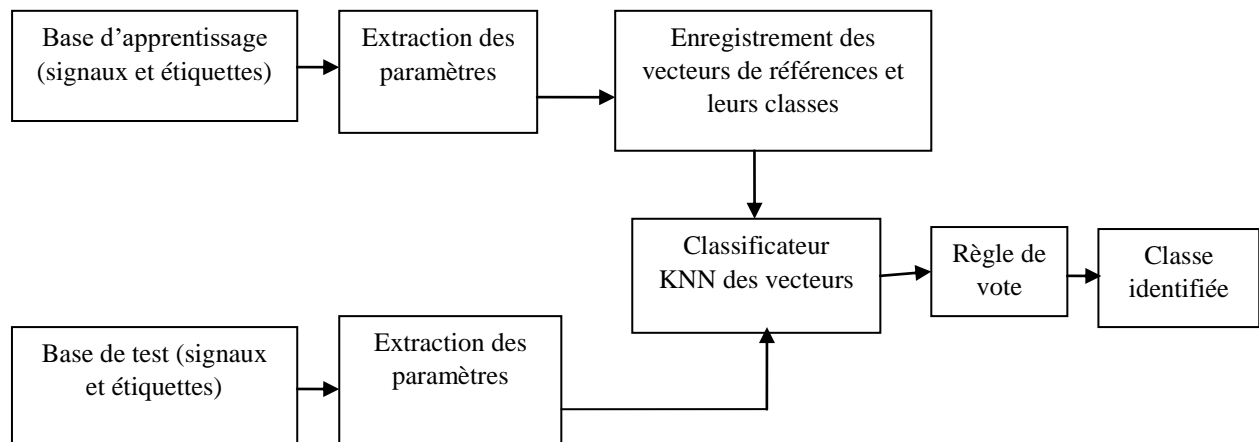


Figure II. 4: Schéma du système RAE proposé dans [3]

II.6 Système RAE proposé

Dans ce travail, nous proposons un système RAE basé sur l'extraction des paramètres statistiques des vecteurs de coefficients MFCC et le classificateur KNN combiné avec la stratégie de la règle de vote. Ces paramètres statistiques sont : la moyenne, l'écart-type, l'asymétrie et l'aplatissement. Les blocs avec des caractères en vert mettent en évidence les nouveautés ajoutées par rapport au système proposé par [3] (voir figure II.5).

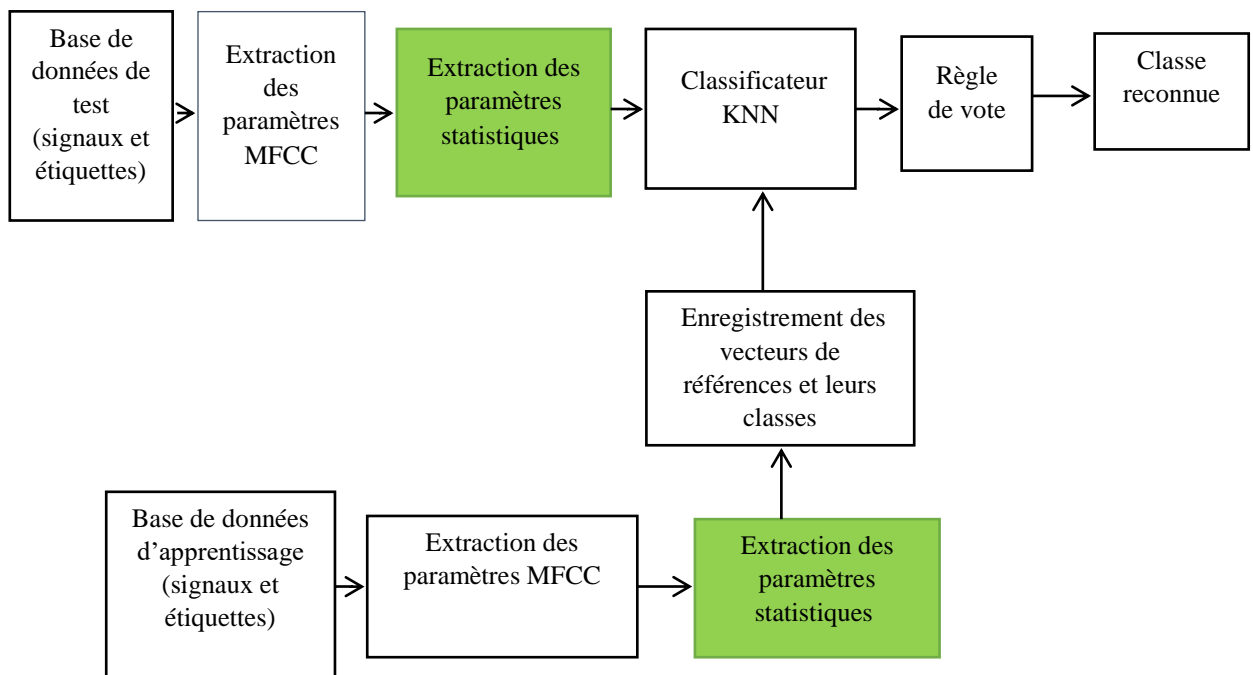
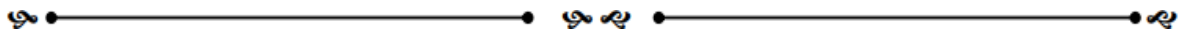


Figure II. 5: système RAE basé sur l'extraction des paramètres statistiques

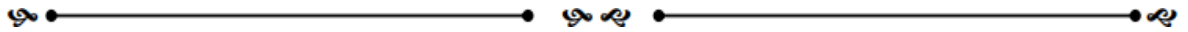
II.7 Conclusion

Dans ce chapitre nous avons présenté des généralités sur les techniques d'extraction de paramètres statistiques pour un système de reconnaissance acoustique des émotions. Puis, nous avons décrit brièvement les étapes d'extraction des paramètres MFCC, ainsi que leurs paramètres statistiques. Finalement, nous avons présenté le système RAE proposé.



Chapitre III

Implémentation et résultats



III.1 Introduction

La reconnaissance acoustique des émotions est une technique utilisée pour déterminer l'état émotionnel d'une personne à partir de son signal vocal. Elle est basée sur l'idée que les émotions humaines sont souvent exprimées par des changements dans l'intonation, le rythme, l'intensité et d'autres caractéristiques acoustiques de la parole. La conception d'un tel système est basée généralement sur une phase d'apprentissage et une phase de test. Chacune de ces phases, nécessite une étape d'extraction des paramètres permettant de convertir chaque signal en une séquence de vecteurs de paramètres. Dans notre travail on propose l'extraction des paramètres statistiques des coefficients MFCC appliquée à la reconnaissance acoustique des émotions basée sur le classificateur KNN combiné avec la stratégie de la règle de vote.

Les questions essentielles pour trouver la bonne configuration de notre système proposé sont les suivantes :

1. Comment implémenter un système RAE basé sur l'extraction des paramètres statistiques combiné avec le classificateur KNN et la stratégie de règle de vote ?
2. Quelle est la meilleure combinaison des paramètres statistiques des coefficients MFCC?
3. Quel est l'effet de la règle de vote sur les performances du système RAE ?
4. Quelle est la configuration optimale du classificateur KNN permettant d'obtenir les meilleures performances
5. Quelle est la combinaison optimale des paramètres statistiques MFCC?

Diverses expériences sont menées afin de répondre à ces questions. La réponse sur la première question consiste à présenter les étapes nécessaires pour l'implémentation de notre système RAE sous l'environnement de programmation Matlab. Plus particulièrement, l'étape d'extraction s'effectue en utilisant la boîte à outils HTK. Le système proposé est évalué sur la base de données de références EMO-DB (Berlin Data base of Emotional Speech), qui sera brièvement décrite dans la section suivante.

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

La base de données EMO-DB est une des bases des signaux de parole très utilisée dans les systèmes RAE [26] [28] [29] [33] [36] [37] [41]. Elle est riche en vocabulaire, du nombre de locuteurs, de la qualité des signaux et des différentes durées des phrases. Cette base de données est composée de 10 phrases de textes différents. Un ensemble, appelé A, est constitué de 5 phrases courtes, tandis qu'un autre ensemble nommé B constitué de 5 phrases longues. Ces phrases sont prononcées en allemand par 10 acteurs, comprenant 5 hommes et 5 femmes. Les acteurs simulent sept états émotionnels primaires : 'colère', 'dégoût', 'peur', 'joie', 'neutre', 'tristesse' et 'surprise'.

La collection complète des signaux comprend 800 énoncés au total, dont 535 énoncés ont été validés par des auditeurs (20 auditeurs au total). Par conséquent, la base de données utilisée dans notre étude est constituée de 535 signaux représentant sept classes d'émotions. Le tableau III.1 présente la répartition des enregistrements de phrases parmi les 7 états émotionnels, ainsi que le nombre d'occurrences utilisées pour les phases d'apprentissage et de test. Les enregistrements ont été initialement effectués à une fréquence d'échantillonnage de 48 kHz, puis sous-échantillonnés à 16 kHz. Dans notre étude, l'ensemble A, composé de 277 énoncés (représentant 51,78 % de la base totale), est utilisé comme base de données d'apprentissage, tandis que l'ensemble B, composé de 258 énoncés (représentant 48,22 % de la base totale), est utilisé comme base de données de test. Par conséquent, notre système RAE développé fonctionne de manière indépendante du texte.

Tableau III 1: Répartition de la base de données EMO-DB

Classes	Classes							
	colère	surprise	dégoût	peur	joie	neutre	tristesse	Totale
Base de données								
Base d'apprentissage	65	41	25	35	38	41	32	277
Base de test	62	40	21	34	33	38	30	258
Totale	127	81	46	69	71	79	62	535

III. 3 Implémentation du système RAE proposé

Les étapes de conception de ce système sont illustrées sur la Figure III.1.

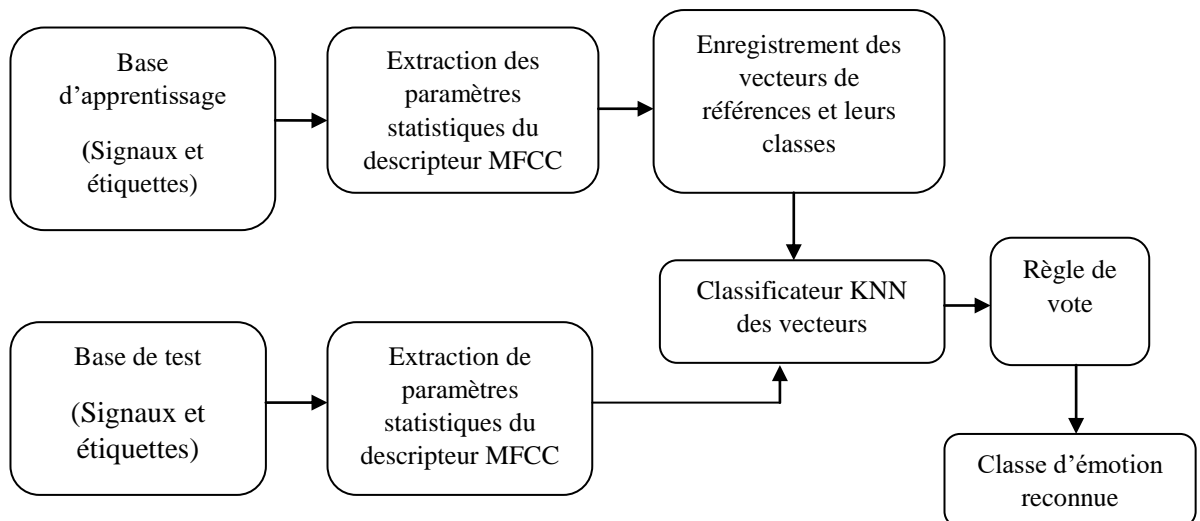


Figure III. 1: Schéma synoptique du système RAE proposé

III.3.1 Phase d'apprentissage

L'extraction des paramètres consiste à convertir chaque signal parole en une séquence de vecteurs de paramètres. Cette étape est effectuée en suivant plusieurs étapes à savoir l'étape de prétraitement, l'étape de fenêtrage et l'étape d'analyse des signaux utilisant la boîte à outils HTK afin d'extraire les coefficients MFCC pour chaque fenêtre d'analyse. La Figure III.1 illustre le schéma synoptique de l'étape d'extraction des paramètres proposée.

III.3.1.1 Extraction des paramètres MFCC

L'étape d'extraction des paramètres MFCC consiste à découper premièrement le signal parole en une séquence de fenêtres d'analyse de 30 ms chevauchées de 20 ms (66.67%). Ensuite, chaque séquence est convertie en un vecteur constitué de 39 paramètres du descripteur MFCC. Ces paramètres sont ordonnés comme suit : 12 coefficients MFCC statiques, log énergie (E), 12 paramètres dynamiques Δ des coefficients MFCC, paramètre dynamique de l'énergie ΔE , 12 paramètres dynamiques $\Delta\Delta$ des coefficients MFCC ainsi que du paramètre dynamique $\Delta\Delta E$. Les paramètres dynamiques Δ et $\Delta\Delta$ sont respectivement notés D et A. Chaque vecteur est composé de 39 paramètres, lesquels sont obtenus en utilisant l'outil Hcopy de la bibliothèque HTK [58] pour extraire les paramètres à partir du fichier audio au format .wav qui représente le signal d'entrée. Pour convertir un signal d'entrée en une séquence de vecteurs contenant 39 paramètres MFCC, on peut utiliser la commande suivante :

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

Le tableau III-2 représente le fichier de configuration de l'outil HCopy tel que :

- **analysis.conf**: est utilisé pour configurer l'extraction des paramètres acoustiques MFCC. Il s'agit d'un fichier texte qui contrôle le processus d'extraction. Un exemple de configuration est présenté dans le tableau III-2.
- **fichier_son.wav** : est un fichier audio au format wav qui représente le signal d'entrée.
- **fichier_param.mfc** : est un fichier écrit en format HTK, dans lequel la séquence des vecteurs de paramètres est enregistrée.

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

SOURCEFORMAT	=	WAV
SOURCEKIND	=	WAVEFORM
HNET: TRACE	=	1
TARGETKIND	=	MFCC_E_D_A
# 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

Avec :

- **WINDOWSIZE:** la durée d'une période de temps (exemple: 30ms) mesurée en utilisant l'unité de HTK (10⁻⁷).
- **TARGETRATE:** la durée de décalage des fenêtres d'analyse (exemple: 10 ms).
- **TARGETKIND:** L'identifiant des coefficients à utiliser. MFCC_E_D_A signifie que l'ensemble des paramètres est constitué des coefficients MFCC statiques, l'énergie ainsi que leurs paramètres dynamiques Δ (D) et $\Delta\Delta$ (A).
- **NUMCEPS:** Le nombre de coefficients MFCC statiques.
- **USEHAMMING:** Le choix de la fonction de Hamming pour le fenêtrage
- **PREEMCOEF:** Le coefficient de préaccentuation.
- **NUMCHANS:** Nombre de canaux du banc de filtres.
- **CEPLIFTER:** La longueur du liftering cepstral.
- **SOURCERATE:** représente La période d'échantillonnage.

III.3.1.2 Extraction des paramètres statistiques

Cette étape consiste à estimer les paramètres statistiques (moyenne, écarte type, l'asymétrie et l'aplatissement) sur un ensemble de vecteurs instantanés, le nombre optimal de ces vecteurs sera discuté dans la partie expérimentale. L'objectif de cette étape comme déjà évoqué est de minimiser le nombre de vecteurs de paramètres représentant le signal vocal. Il est à noter que cette étape réduise le nombre de vecteurs, mais en contrepartie augmente légèrement la taille de ces vecteurs (nombre de paramètres), comme illustré dans la figure III.2.

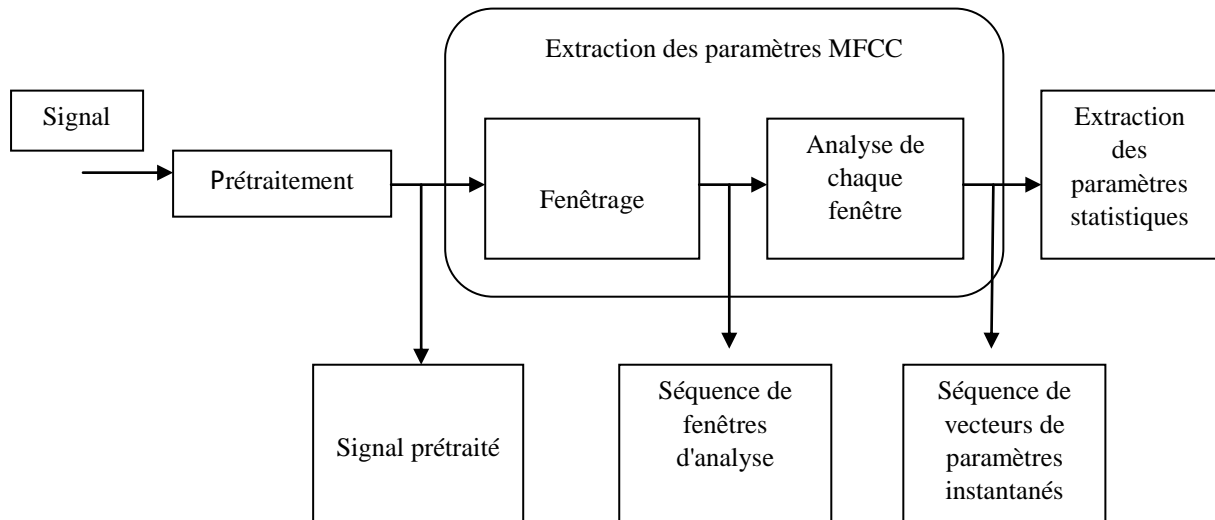


Figure III. 2: Etape d'extraction des paramètres statistiques [59]

III.3.1.3 Enregistrement des vecteurs de références

Dans cette étape, les vecteurs extraits sont enregistrés dans une matrice appelée matrice de paramètres d'apprentissage. Chaque ligne de cette matrice représente un vecteur de paramètres qui correspond à une fenêtre d'analyse. Les classes des vecteurs sont également enregistrées dans un tableau (de type ligne ou colonne) appelé tableau de classes. La matrice de vecteurs et le tableau de classes d'apprentissage seront utilisés dans l'étape de classification KNN.

III.3.2 Phase de test

Cette phase a pour objectif de reconnaître la classe d'émotion de chaque signal et d'évaluer les performances de notre système RAE en termes de taux de classification. La reconnaissance de la classe d'émotion du signal repose sur deux étapes : l'extraction des paramètres statistiques et la classification de la séquence de vecteurs de paramètres statistiques qui représente le signal de test. Cette étape est réalisée de la même manière que celle de la phase d'apprentissage.

III.3.3 Classificateur KNN combiné avec la stratégie de la règle de vote

La classification d'une séquence de vecteurs de paramètres implique plusieurs étapes. Tout d'abord, chaque vecteur de paramètres statistiques est classifié en utilisant le classificateur KNN. Ensuite, la stratégie de règle de vote est appliquée sur la séquence d'indices de classes pour déterminer la classe la plus votée, représentant ainsi la classe d'émotion du signal de test.

Plus précisément, le classificateur KNN compare chaque vecteur de la séquence de vecteurs de paramètres de test avec tous les vecteurs de paramètres présents dans la matrice de paramètres d'apprentissage. Il détermine ensuite l'indice de classe du vecteur ayant la distance minimale par rapport au vecteur de test. Cette étape permet d'associer une classe au vecteur de test en se basant sur les vecteurs les plus similaires présents dans la matrice d'apprentissage.

III.3.4 Mesure de performances du système RAE proposé

Notre travail consiste à évaluer les performances du classificateur KNN en considérant initialement le taux de classification des vecteurs de paramètres statistiques de tous les signaux de la base de données de test, sans prendre en compte la classification des signaux eux-mêmes

$$TCV = \frac{N_{cv}}{N_{vt}} * 100 \quad \text{III.1}$$

Avec :

- N_{vt} est le nombre des vecteurs de paramètres de tous les signaux de la base de test,
- N_{vc} est le nombre de vecteurs de paramètres correctement classifiés.

Sans tenir compte de l'appartenance des vecteurs aux signaux, le taux TCV offre une évaluation globale de l'efficacité du classificateur KNN.

En réalité, le système RAE a pour objectif de classifier chaque signal afin d'identifier sa classe émotionnelle. Par conséquent, nous proposons ensuite d'évaluer les performances du système RAE en tenant compte le taux de classification des signaux de test après l'application de la stratégie de la règle de vote :

$$TCS = \frac{N_{sc}}{N_{st}} * 100 \quad \text{III.2}$$

Avec :

- N_{ST} : est le nombre total des signaux de la base de test.
- N_{sc} : est le nombre de signaux correctement reconnus.

III.4 Expériences et résultats

Cette section est consacrée à l'explication des diverses expériences menées dans notre étude, visant à répondre aux questions soulevées dans l'introduction de ce chapitre. Nous procéderons également à discuter et interpréter les résultats obtenus pour chaque expérience.

III.4.1 combinaison optimale des paramètres statistiques

L'objectif de cette expérience est de chercher la meilleure combinaison des quatre types de paramètres statistiques : moyenne, écarte type, l'asymétrie et l'aplatissement, en termes de taux de classification. Dans cette expérience, chaque signal vocal est converti en un seul vecteur de 156 paramètres statistiques estimés sur toute séquence des vecteurs de 39 paramètres MFCC. Par conséquent, la classification KNN a été effectuée sans l'application de la règle de vote

Tableau III. 3:Meilleure combinaison des paramètres statistiques

Combinaison des paramètres statistiques	Nombres de paramètres	CR%
μ	39	65.89
Δ	39	48.44
SK	39	33.33
Ku	39	36.82
μ, δ	78	68.99
μ, SK	78	65.89
μ, Ku	78	67.82
δ, SK	78	52.32
δ, Ku	78	56.20
SK, Ku	78	39.92
μ, SK, Ku	117	68.21
μ, δ, Ku	117	74.41
δ, SK, Ku	117	55.81

μ, δ, SK	117	69.76
μ, δ, SK, Ku	156	74.41

A partir du tableau III. 2 on remarque que la meilleure combinaison des types de paramètres statistiques est celle constituée de la moyenne et de l'écart type et l'aplatissement formant un vecteur de paramètres de 78 paramètres. Cette combinaison permet d'obtenir les taux de classification les plus élevés avec un CR de 74.41% Ce résultat est obtenu aussi avec la combinaison des quatre type paramètres statistiques (moyenne, écarte type, l'asymétrie et l'aplatissement) avec un vecteur de paramètres de 156 paramètres. La combinaison de trois mesures statistiques (μ, σ, Ku) sera choisie dans les expériences suivantes.

III.4.2 Nombre optimal des vecteurs statistiques

Comme déjà évoqué, notre système RAE est basé sur l'utilisation des paramètres statistiques pour minimiser la complexité du système en termes du temps de calcul et de l'espace mémoire. Cependant, le processus de classification des émotions se déroule de manière permanente, ce qui exige de limiter la durée de l'estimation des paramètres statistiques en divisant chaque séquence de vecteurs de paramètres MFCC instantanés en sous séquences dont chacune est convertie en un vecteur de paramètres statistiques. L'objectif de cette expérience est de déterminer le nombre de vecteurs instantanés NVIS optimal de chaque sous séquence [59]. Le tableau III.4, présente le taux de classification pour tous les cas possibles des valeurs de NVIS allant de 2 à 50. Les résultats sont présentés en termes du taux de classification des vecteurs TCV et du taux de classification des signaux TCS obtenu par l'application de la méthode de règle de vote.

Tableau III. 4: Taux de classification TCV et TCS en fonction du nombre des vecteurs instantanés NVIS

NVIS	TCV	TCS
2	40.37	82.17
3	45.39	79.06
4	50.79	79.45
5	55.35	75.96

Chapitre III : Implémentation et résultats

6	59.04	73.64
7	61.81	73.64
8	64.09	75.19
9	66.00	73.64
10	67.90	76.35
11	69.32	72.48
12	70.83	70.93
13	71.42	74.03
14	72.64	73.64
15	73.81	72.48
16	74.16	70.54
17	74.58	72.82
18	75.29	74.03
19	75.73	72.09
20	76.18	74.41
21	76.65	74.41
22	76.87	74.80
23	77.46	71.70
24	77.67	74.80
25	77.94	72.86
26	77.94	72.48
27	78.21	72.48
28	78.42	70.54
29	78.76	70.58
30	78.68	69.76
31	79.04	70.93
32	79.28	71.31
33	79.64	73.64
34	79.63	73.64
35	79.65	72.86
36	79.75	72.09
37	79.86	76.74
38	80.00	73.25
39	80.16	74.03

40	80.35	74.41
41	80.41	74.03
42	80.32	71.70
43	80.50	75.96
44	80.81	74.41
45	81.00	73.25
46	81.04	73.25
47	81.11	74.80
48	81.26	75.19
49	81.19	74.41
50	81.17	73.25

A partir de ces résultats, nous pouvons conclure que le nombre optimal de vecteurs instantanés NVIS utilisés pour l'estimation de chaque vecteur de paramètres statistiques est égal à deux avec un taux TCS égal à 82.17 %. On peut remarquer que l'augmentation du nombre de vecteurs NVIS améliore le taux TCV, mais dégrade le taux TCS. Cette dégradation peut être justifiée par le nombre réduit des vecteurs de paramètres statistiques utilisés dans l'algorithme de la stratégie de règle de vote. Les figures III. 3 et III. 4 illustrent respectivement les taux de classification TCV et TCS pour les différentes valeurs du nombre de vecteurs instantanés NVIS.

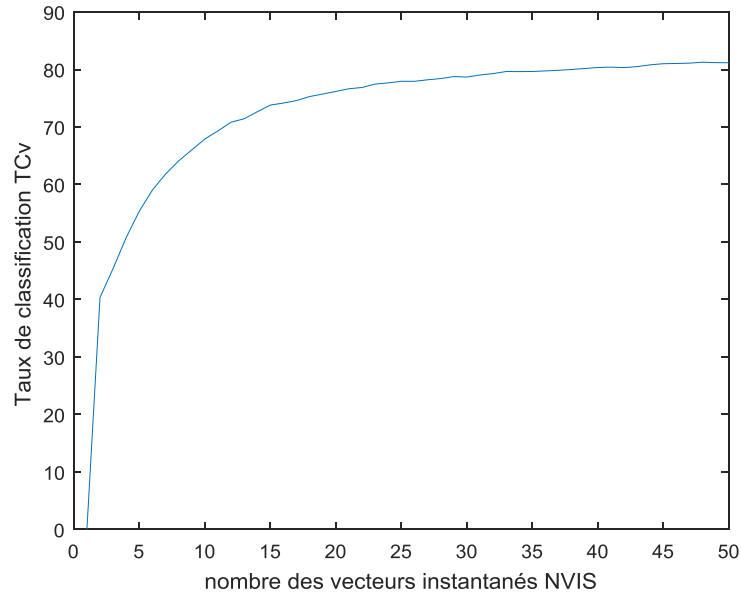


Figure III.3: Taux de classification TCv en fonction de nombre des vecteurs instantanés NVIS

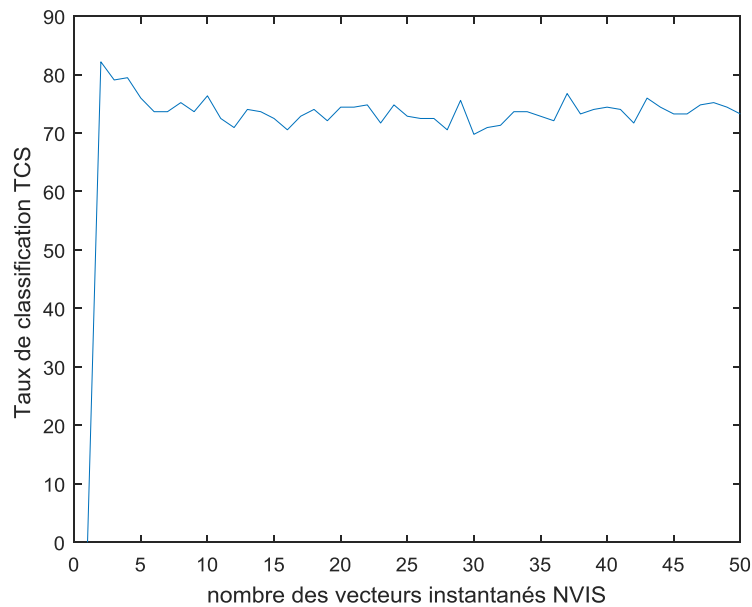


Figure III.4: Taux de classification TCS en fonction de nombre des vecteurs instantanés NVIS.

III.4.3 Configuration optimale du classificateur KNN

L'objectif de cette expérience consiste à chercher la configuration optimale du classificateur KNN en déterminant à la fois la valeur optimale K des vecteurs les plus proches voisins et la

distance optimale en termes de taux de classification des signaux TCS. Les résultats des taux de classification TCS pour les différentes valeurs de k (allant de 1 à 20) et pour les différentes distances (Euclidienne, Cityblock, Cosine, Corrélation) sont présentés dans le tableau III.5.

Tableau III. 5: configuration optimale du classificateur KNN

Distance \ K	Euclidean	Cityblock	Cosine	Correlation
1	82.17	81.39	79.45	80.62
2	82.17	81.39	79.45	80.62
3	79.84	80.62	77.90	79.45
4	79.45	79.06	78.29	78.68
5	79.84	79.84	77.51	77.51
6	78.68	79.84	77.13	77.13
7	77.51	77.90	77.13	77.51
8	77.51	77.90	75.96	76.74
9	76.35	78.29	75.96	75.19
10	78.29	79.06	75.58	75.96
11	77.90	77.51	74.80	75.85
12	75.96	77.90	72.86	75.96
13	75.19	76.74	73.64	74.80
14	75.19	77.51	73.25	74.41
15	76.35	77.13	72.48	71.70
16	76.35	76.74	72.86	72.09
17	74.80	76.74	71.70	72.86
18	75.58	76.35	72.09	71.70
19	75.19	76.74	71.31	71.31
20	74.80	75.58	71.31	71.31

A partir du tableau III.5, on peut remarquer les points suivants :

1. Le taux de classification TCS atteint des valeurs élevées pour les deux premières valeurs de k (k=1,2), quelle que soit la distance utilisée.
2. Le taux de classification maximal de 82.17% est atteint en utilisant la distance euclidienne avec des valeurs de k égales à 1 et 2.

III.4.4 Combinaison optimale des paramètres MFCC

L'objectif de cette expérience est d'identifier la meilleure combinaison de paramètres du descripteur MFCC en se basant sur les types de paramètres suivants :

- Paramètres MFCC statiques.

- Paramètre énergétique E.
- Paramètre dynamique D (vitesse).
- Paramètre dynamique A (accélération).

Les différentes combinaisons considérées sont données comme suit : MFCC avec 36 coefficients, MFCC_E avec 39 coefficients, MFCC_D avec 72 coefficients, MFCC_ED avec 78 coefficients, MFCC_EDA avec 117 coefficients. Le tableau III.6 présente le taux de classification du système RAE basé sur classificateur KNN combiné avec la règle de vote.

Tableau III. 6:Taux de classification pour différentes combinaisons de paramètres MFCC

Types de descripteur	EDA	MFCC	MFCC_E	MFCC_D	MFCC_ED	MFCC_EDA
Nombre de Coefficients	9	36	39	72	78	117
TCS	27.90	79.45	77.90	81.00	81.39	82.17

A partir du tableau III.6, on peut donner les points suivants :

1. L'utilisation de tous les paramètres (MFCC_EDA) permet d'obtenir le meilleur taux de classification de 82.17%. Cela indique que l'inclusion de tous les paramètres conduit à une meilleure capacité de classification et à de meilleures performances globales du système.
2. Les paramètres énergétiques seuls donnent un taux de classification relativement faible de 27.90%. Cela suggère que ces paramètres ne contiennent pas suffisamment d'informations discriminantes pour une classification précise des données.

III.4.5 L'effet d'utilisation des paramètres statistiques sur la complexité du système

L'objectif de cette expérience est de comparer les performances de notre système RAE avec celle proposé dans [3], en termes de complexité (temps de calcul est espace mémoire). Le tableau III.7 nous présente les résultats des deux systèmes RAE. Cette expérience est effectuée en utilisant un PC portable I3 de 4 Gb de RAM.

Tableau III. 7:comparaison des performances des systèmes RAE

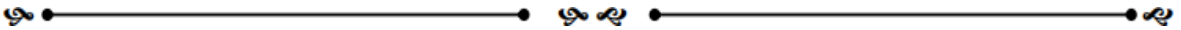
Système RAE proposé dans [3],			Notre système RAE proposé		
TCS%	Temps de calcul [seconde]	Nombre de vecteurs d'apprentissages	TCS%	Temps de calcul [seconde]	Nombre de vecteurs d'apprentissages
76.36	242.31	60093	82.17	140.96	29973

Les résultats obtenus nous montrent clairement l'efficacité de l'utilisation des paramètres statistiques, en termes de complexité et de précision. En effet, cette utilisation a permis de réduire le temps de calcul et de l'espace mémoire, expliquée par la taille de la matrice d'apprentissage. De plus, elle a amélioré le taux de classification d'environ 6%.

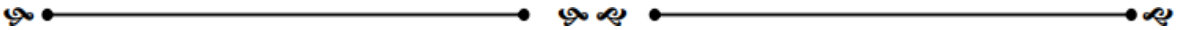
III.5 Conclusion

Dans ce chapitre, nous avons présenté un système RAE. Ce système est basé sur le classificateur KNN combiné avec la stratégie de la règle de vote et l'extraction des paramètres statistiques du descripteur des paramètres MFCC. Nous avons réalisé plusieurs expériences afin de déterminer les configurations optimales pour ce système. Les résultats obtenus nous ont permis de conclure ce qui suit :

- La meilleure combinaison des paramètres statistiques des coefficients MFCC correspond au choix de la moyenne statistique, l'écart-type et l'aplatissement de ces coefficients, qui permettent d'obtenir un taux de classification maximal de 82.17%.
- L'application de la stratégie de règle de vote permet d'améliorer les résultats de classification TCS.
- L'utilisation de la totalité d'espace de paramètres MFCC-EDA donne les meilleures performances du système.
- Les paramètres statistiques ont un rôle efficace dans la réduction du temps de calcul et de l'espace de stockage, ainsi que l'amélioration du taux de classification.



Conclusion générale



Conclusion générale

Notre travail entre dans le cadre des systèmes de reconnaissance acoustique des émotions. Plusieurs travaux de recherches ont utilisé l'algorithme KNN pour classifier le signal parole en une classe d'émotions parmi les 7 émotions standards: Colère, dégoût, joie, tristesse, peur, surprise, neutralité. Cependant, ce classificateur souffre d'une complexité en termes d'espace mémoire et temps de calcul. Notre travail consiste à minimiser cette complexité en réduisant chaque séquence de vecteurs de paramètres MFCC en un nombre réduit de vecteurs de paramètres statistiques. Plus particulièrement, on a implémenté un système RAE basé sur l'extraction des paramètres statistiques des coefficients MFCC et le classificateur KNN combiné avec la stratégie de règle de vote. Ce système est implémenté en utilisant le langage de programmation Matlab ainsi que la boîte à outils HTK.

Les résultats obtenus nous ont montré que la combinaison de paramètres statistiques (moyenne μ , écart type δ et aplatissement ku) donnent le meilleur taux de classification de 74.41%, ainsi que le choix de 2 vecteurs de paramètres instantanés utilisés pour l'estimation de chaque vecteur de paramètres statistiques améliore le taux de classification TCS à 82.17%. Cette technique d'extraction permet de réduire l'espace mémoire nécessaire en extrayant les paramètres les plus pertinents des données d'émotion, tout en préservant les informations essentielles pour la classification. De plus, elle permet d'accélérer le processus de comparaison en utilisant des représentations statistiques des données plutôt que les vecteurs bruts complets.

Finalement, plusieurs perspectives sont envisagées à savoir :

- L'évaluation du système sous une autre base de données plus large qu'EMO-DB.
- L'utilisation d'autres types de classificateurs, tels que GMM, LDA, SVM, ANN ...etc.
- L'utilisation d'autres descripteurs de paramètres.

Bibliographie

- [1] F. Mehran, «Les émotions positives,» *Psychologie positive et personnalité*, 2010.
- [2] T. g. Nathan c.hall, *Emotion ,moivation,and self-regulation*, university of konstanz, 2013.
- [3] S. A. ZITOUNI SABER, «Reconnaissance acoustique des émotions basée sur le classificateur KNN,» 2022.
- [4] N. H. F. Anna Tcherkassof, «Les émotions : une conception relationnelle,» vol. 114, 2014.
- [5] L. V. Laurence Devillers, «Revue d'intelligence artificielle,» vol. 20, n° %14-5, 2006.
- [6] Pitpitt, «Reconnaissance d'émotions,» 24/05/2023. [En ligne]. Available: https://datafranca.org/wiki/Reconnaissance_d%27%C3%A9motions.
- [7] G. R. Chloé clavel, «Livre Clavelrichard - Telecom Paris,» 25/05/2023. [En ligne]. Available: https://perso.telecom-paristech.fr/grichard/Publications/livre_ClavelRichard.pdf.
- [8] G. RICHARD, «Curriculum vitae - Telecom Paris,» 15/05/2023. [En ligne]. Available: https://perso.telecom-paristech.fr/grichard/Publications/CV_2020_english_short.pdf.
- [9] P. F. W. V. & E. P. Ekman, «Emotion in the human face: Guidelines for research and an integration of findings. Pergamon Press.,» 1972.
- [10] R. Plutchik, «The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist,,» vol. 89, n° %14, 2001.
- [11] Tomkis, «L'émergence des émotions dans les sciences psychologiques,» 2016.
- [12] B.Arlond, «Arnold's theory of emotion in historical perspective,» November 200.
- [13] M. K. K. E. M. a. S. N. Grimm, «Analyse et détection des émotions verbales dans les interactions orales,» 2007.
- [14] Darwin, *Darwin et expression des émotions*, 1993.
- [15] «Quelles sont les émotions de base ?,» 08/2022. [En ligne]. Available: <https://amelioretasante.com/quelles-sont-les-emotions-de-base/>.

-
- [16] G. L.-P. C. Jia Rong, «Acoustic feature selection for automatic emotion recognition from speech,» vol. 45, n° 13, 2009.
- [17] «Les émotions secondaires et leur manifestation chez les enfants,» 12 Décembre 2018. [En ligne]. Available: <https://etrepants.com/les-emotions-secondaires-et-leur-manifestation-chez-les-enfants/>.
- [18] D. Goleman, «Emotional intelligence,» 1999.
- [19] R. Thibodeau, «Emotion et coryances,» 2002.
- [20] A. Hacine-Gharbi et P. Ravier, «On the optimal number estimation of selected features using joint histogram based mutual information for speech emotion recognition,» vol. 31, n° 11, 2019.
- [21] B. V. Bjorn Schuller, «Acoustic emotion recognition: A benchmark comparison of performances,» 2009.
- [22] K. Leila, «Analyse acoustique de la voix pour la détection des émotions du locuteur ...,» 18/05/2023. [En ligne]. Available: <https://theses.hal.science/tel-02925116>. [Accès le 28/08/2020].
- [23] A. H.-G. ., R. ., FATEH GHAZALI, «Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method,» vol. 27, 1/1/2019.
- [24] T. & H. P. Cover, « Nearest neighbor pattern classification. IEEE Transactions on Information Theory,» 1967.
- [25] V. K. J. R. Q. J. G. Q. Y. H. M. G. J. M. A. N. X. Wu, «Top 10 algorithms in data mining,» vol. 14, n° 11, 2008.
- [26] A. K. M. R. ., A. S. Md. Kamruzzaman, « Emotion Recognition from Human,» May 2014..
- [27] S. Y. M. M. K. R. e. M. M. A. K. Leila, «A Review on Speech Emotion Recognition: Case of Pedagogical Interaction in Classroom,» 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP),» 2017.
- [28] S. Y. e. Z. Xue-Ying, ««Characteristics of human auditory model based on compensation of glottal features in speech emotion recognition,» Future Generation Computer Systems, pp.,» 2018.
- [29] H. A. e. G. Polat, ««Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection,» Expert Systems with Applications,» vol. 36, 2009. .
- [30] K. S. e. P. N. J. Bhaskar, ««Hybrid approach for emotion classification of audio conversation based on text and speech mining,» Procedia Computer Science,,» vol. 36, 17 January 2018..

-
- [31] Q. X. M. W. W. C. Y. M. e. J. M. Z. Liu, «Speech emotion recognition based on feature selection and extreme learning machine decision tree,» *Neurocomputing,*» vol. 273, 17 January 2018.
- [32] Z. W. J. J. X. H. e. C. L. W. Jiang, ««Speech emotion recognition with heterogeneous feature unification of deep neural network,» *Sensors,*» vol. 19, 18 June 2019.
- [33] J. C. M. M. ., C. T. J. Alonso, « «New approach in quantification of emotional intensity from the speech signal: Emotional temperature,» *Expert Systems With Applications,*» vol. 42, 30 December 2015.
- [34] Q. X. M. W. W. C. Y. M. J. M. Z. Liu, ««Speech emotion recognition based on an improved brain emotion learning model,» *Neurocomputing,*» vol. 309, 2 October 2018.
- [35] L. C. D. Z. J. Z. ., W. Z. L. Zhu, ««Emotion recognition from chinese speech for smart affective services using a combination of SVM and DBN,» *Sensors,*» vol. 17, 24 July 2017. .
- [36] B. S. A. A. J. L. e. N. G.-V. A. Álvarez, ««Classifier subset selection for the stacked generalization method applied to emotion recognition in speech,» *Sensors,*» vol. 16, 25 Décembre 2015.
- [37] P. C. e. R. S. A. Bhavan, ««Bagged support vector machines for emotion recognition from speech,» *Knowledge-Based Systems,*» vol. 184, 15 November 2019.
- [38] P. S. e. P. Sircar, ««Continuous wavelet transform based speech emotion recognition,» *In Proceedings of the 2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS),»* December 2016.
- [39] Y. S. K. R. M. M. M. e. C. C. L. Kerkeni, ««Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO,» *Speech Communication,*» 2019.
- [40] Z. K. e. O. Oludayo, ««Ensemble Learning of Hybrid Acoustic Features for Speech Emotion Recognition,» *MDPI,*» 22 March 2020.
- [41] A. H.-g. e. R. Philippe, ««On the optimal number estimation of selected features using joint histogram based mutual information for speech emotion recognition,» *Journal of King Saud University,*» 23 July 2019. .
- [42] Schuller, Björn;, «Communication vocale,» vol. 53, n° 19-10, 2011.
- [43] «Wikipedia,» [En ligne]. Available: https://en.wikipedia.org/wiki/Mel-frequency_cepstrum.
- [44] N. ASBAI, «Identification et Authentification de Locuteurs, par les Techniques de Fusion des,» 2015.
- [45]]. R. C. e. F. MOHAMADI, «Reconnaissance Acoustique des émotions basée sur les modèles,»

2018.

- [46] steven davis , paul mermelstein, «Readings in Speech Recognition,» 1990.
- [47] Patricia Thaine , Gerald Penn, «Extracting mel-frequency and bark-frequency cepstral coefficients from encrypted signals,» 2019.
- [48] B. L.Rabiner, «Hidden markov models for speech recognition — strengths and limitations,» 1992.
- [49] D. Ronnie Smith, «Spoken Natural Language Dialog Systems,» 1995.
- [50] R.Swingle, Results of analyses of Tank 23H and 24h Saltstone WAC samples HTK-531-HTK-528, 2003.
- [51] P. S.Davis, «Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,» vol. 28, n° 14, 1980.
- [52] Les Définitions, 28/05/2023. [En ligne]. Available: <https://lesdefinitions.fr/parametre..>
- [53] B. W. (. Schuller, «Emotion recognition :Emotion recognition: A pattern analysis approach. Springer.,» 2018.
- [54] F. S. K. R. S. B. W. S. J. A. E. B. C. .. & W. M. Eyben, «The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing,» vol. 7.
- [55] M. G. D. J. B. & S. B. Soleymani, «acoustic emotion recognition: A review of the most recent databases and approaches. IEEE Transactions on Affective Computing,» vol. 7.
- [56] A. B. Björn Schuller, «Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge,» vol. 53, n° 19-10, 2011.
- [57] R. L. B. George Casella, Statistical Inference.
- [58] S. Young, The HTK book (HTK version 3.4)., Cambridge, 2006.
- [59] D. G. Fatah, « THESE DE DOCTORAT : Extraction et sélection des paramètres pertinents appliquées à l'identification des appareils électriques,» 2022.