

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

FILIERE : Télécommunications

Spécialité : Systèmes des Télécommunications

Par

➤ **ZITOUNI SABER**

➤ **SAIDANI ADEL**

Intitulé

***Reconnaissance acoustique des émotions basée sur le
classificateur KNN***

Devant le Jury composé de :

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>Mme. MESSALI .Zoubaida</i>	<i>Professeur</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mr. HACINE GHARBI Abdenour</i>	<i>MCA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Mme. HACINI .Latifa</i>	<i>MCA</i>	<i>Examineur</i>	<i>Univ-BBA</i>
<i>Mr. GHAZALI Fateh</i>	<i>DR</i>	<i>Co-encadreur</i>	<i>Univ-BBA</i>

Année Universitaire 2021/2022

Remerciement

*On remercie tout d'abord le dieu qui nous a donné la sante et la volonté
d'entamer et de terminer 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 exceptionnel, pour sa patience, sa rigueur, son aide
pratique et sa disponibilité durant notre préparation de ce mémoire*

*Nos remerciements s'adressent à **DR. GHAZALI FATEH** pour son aide pratique
et son soutien moral et ses encouragements. Nos remerciements s'adressent
également à tous nos professeurs pour leur générosité et la grande patience
dont il sut faire preuve malgré leurs charges académiques et professionnelles.*

*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.*

Résumé

Ce travail a pour objectif de proposer et d'implémenter un système de reconnaissance automatique des émotions (RAE) performant. Deux systèmes RAE proposés sont basés sur le classificateur KNN (k-Nearest Neighbors) et la stratégie de la règle de vote. Ces deux systèmes utilisent deux méthodes d'extraction différentes. Le premier système se base sur l'extraction des paramètres MFCC (Mel-Frequency Cepstral coefficients), alors que le deuxième système utilise l'analyse en ondelettes pour extraire des paramètres énergétiques appelés DWE (Discret Wavelet Energy), calculés à des niveaux de décomposition dyadique. Chacun des deux descripteurs MFCC et DWE inclut également ses paramètres dynamiques Δ et $\Delta\Delta$. Les performances du système sont évaluées en termes du taux de classification, sous la base de données EMO-DB. Les résultats obtenus montrent la pertinence des paramètres MFCC avec un taux de classification de signaux égal à 76.74%, alors que les paramètres DWE présentent un faible taux de classification de 56.58%. De plus, la combinaison de ces deux descripteurs améliore légèrement le taux de classification à 79.04 %, mais avec une augmentation du temps d'exécution.

Mots clés : Reconnaissance acoustique des émotions, Transformée en ondelettes discrète (DWT), coefficients MFCC, classificateur KNN, stratégie de la règle de vote.

Abstract

The objective of this work is to propose and implement an powerful automatic emotion recognition system (AER). Two AER systems proposed are based on the KNN (k-Nearest Neighbors) classifier combined with voting rule strategy. These two systems use two different features extraction methods. The first system based on the extraction of MFCC (Mel-Frequency Cepstral coefficients) features, while the second system uses wavelet analysis to extract energy features called DWE (Discrete Wavelet Energy), calculated at dyadic decomposition levels. The two descriptors MFCC and DWE include their dynamic features Δ and $\Delta\Delta$. The performance of the system is evaluated in terms of the classification rate, under the EMO-DB database. The obtained results show the relevance of the MFCC features with a classification rate equal to 76.74%, while the DWE features gave a low classification rate of 56.58%.

Moreover, the combination of these two descriptors improves the classification rate to 79.04%, but with an increase in execution time.

Keywords : Acoustic emotion recognition, Discrete Wavelet Transform (DWT), MFCC coefficients, KNN classifier, Voting rule strategy.

ملخص

يهدف هذا العمل إلى اقتراح وتنفيذ نظام التعرف الآلي على المشاعر (RAE). تم اقتراح نظامين (RAE) يعتمدان على مصنف KNN (k-Nearest Neighbours) مع قاعدة التصويت. هذان النظامان يستعملان طريقتين مختلفتين لاستخراج الميزات، يعتمد النظام الأول على استخراج الميزات MFCC ، أما النظام الثاني يستعمل طريقة تحويل الموجات لاستخراج الميزات الطاقوية تدعى (DWE). كل من الواصفين MFCC و DWE يحتوي كذلك على ميزات ديناميكية Δ و $\Delta\Delta$. مردود هذان النظامان يتم تقييمها عن طريق نسبة التصنيف باستعمال قاعدة بيانات (EMO-DB).

النتائج المحصل عليها تثبت أهمية المعاملات MFCC بمعدل تصنيف يساوي 76.74%، بينما ميزات DWE تعطي معدل تمييز منخفض نسبياً يقدر بـ 56.58%. علاوة على ذلك، يؤدي الجمع بين هذين الواصفين إلى تحسين معدل التصنيف بشكل طفيف إلى 79.04%، ولكن مع زيادة وقت التنفيذ.

الكلمات المفتاحية: التعرف الآلي على المشاعر، تحويل الموجات (DWT)، الميزات MFCC ، المصنف KNN، استراتيجية قاعدة التصويت.

Table des matières

Remerciement	ii
Résumé.....	iii
Abstract	iii
Introduction générale	1
Chapitre I: Généralités sur la reconnaissance automatique des émotions	
I.1 Introduction	4
I.2 Généralités sur les émotions.....	4
I.2.1 Définition des émotions	4
I.3 Classification des émotions.....	5
I.3.1 Emotions primaires	5
I.3.2 Émotions secondaire	7
I.4 Types d'émotions	7
I.4.1 Émotions positives	7
I.4.2 Émotions négatives	8
I.5 Types de corpus des émotions.....	8
I.5.1 Émotions naturelles.....	8
I.5.2 Émotions simulées	8
I.5.3 Émotions induites.....	8
I.6.1 Reconnaissance des émotions à partir d'images fixes ou des vidéos (analyse d'expressions faciales).....	9
I.6.2 Reconnaissance des émotions à partir des signaux physiologiques.....	10
I.6.3 Système de reconnaissance des émotions à partir de signal singal parole	11
• Prétraitement.....	12
• Bloc de fenêtrage	12
• Etape d'extraction des paramètres.....	12
• Classificateur	12
I.7 Classificateur KNN	13
a). Distance Euclidienne	13
b). Distance de Cityblock	13
c). Distance Cosinus	13

Table des matières

d). Distance corrélation	14
I.8 Comparaison entre les différentes modalités utilisées pour la reconnaissance des émotions	14
I.9 Domaines d'applications de la reconnaissance des émotions	15
I.10 Conclusion.....	15

Chapitre II : Extraction des paramètres acoustiques

II.1 Introduction.....	17
II.2 Etat d'Art sur les techniques d'extraction de paramètres pour la reconnaissance automatique des émotions	17
II.3 Extraction de paramètres	17
II.4 Extraction des paramètres acoustique basée sur la transformée en ondelettes	19
II.5 Coefficients cepstraux en l'échelle Mel (MFCC)	21
II.5.1 Étapes de calcul du vecteur des paramètres de type MFCC.....	22
a. segmentation en trame (Frame blocking)	23
b. fenêtrage	23
c. Calcul de la transformée de Fourier rapide (Fast Fourier Transform, FFT) :.....	23
d. Filtrage en l'échelle Mel :.....	23
e. Calcul du cepstre sur l'échelle Mel :	24
II.5.2 Calcul des caractéristiques dynamiques des MFCC	24

Chapitre III: Système de reconnaissance acoustique d'émotions basé sur le classificateur KNN combiné avec la règle de vote

III.1 Introduction	27
III.2 Description de la base de données	28
III.3 Implémentation du système RAE proposé	29
III.3.1 Phase d'apprentissage.....	30
III.3.1.1. Extraction des paramètres.....	30
III.3.1.2. Enregistrement des vecteurs de références.....	33
III.3.2 Phase de test	33
III.3.3 Classification des signaux de test en appliquant l'algorithme de classification KNN combiné avec la règle de vote.....	33
III.3.4 Mesure de performances du système RAE proposé	33
III.4 Expériences et résultats	34
III.4.1. Expériences effectuées sur le système basé sur l'extraction des paramètres MFCC ...	34

Table des matières

III.4.1.1. Configuration optimale du classificateur KNN du premier système	34
III.4.1.2. Combinaison optimale des paramètres MFCC.....	37
III.5.1.1. Choix de la fonction mère optimale de la famille Daubechies DB	38
III.5.1.2. Configuration optimale du classificateur KNN du deuxième système	39
III.5.1.3. Combinaison optimale des paramètres DWE.....	41
III.6. Effet de l'application de la stratégie de règle de vote	42
III.7. Combinaison des descripteurs MFCC et DWE.....	43
III.8. Conclusion.....	43
Conclusion générale	47
Référence bibliographie	

Liste des figures

Figure I.1: Architecture du système de reconnaissance automatique des émotions.....	11
Figure II.1: Etape d'extraction de paramètres.	19
Figure II.2: Exemples d'ondelettes mères $\psi(t)$	20
Figure II.3: Processus d'extraction de caractéristiques DWE.....	21
Figure II.4 : Exemple de conversion des hertz en Mel	22
Figure II.5: Étapes de calcul d'un vecteur caractéristique de type MFCC	22
Figure II.6: Calcul des dérivées premières et secondes de coefficients MFCC	25
Figure III.1: Schéma synoptique du système RAE proposé.....	30
Figure III.2: Taux de reconnaissance TCS du premier système RAE avec différentes valeurs de k et différentes distances du classificateur KNN.....	36
Figure III.3: Taux de classification pour différentes fonctions mère DB.....	39
Figure III.4: Taux de classification TCS pour les différentes valeurs de K ainsi pour les différentes distances pour le deuxième système.....	41

Liste des tableaux

Tableau I.1: Principales catégories des émotions de base	5
Tableau I.2: Comparaison entre les corpus émotionnelles	9
Tableau I.3 : Expressions faciales correspondant à chaque émotion.....	10
Tableau I.4 : Comparaison entre les systèmes de reconnaissance des émotions.....	14
Tableau II.1: Comparaison des méthodes apparentées	17
Tableau III.1: Répartition des phrases de la base 'EMO-DB	29
Tableau III.2: Fichier de configuration (analysis.conf)	31
Tableau III.3: Taux de classification TCS du premier système RAE pour différentes valeurs de k et différentes distances.....	35
Tableau III.4: Taux de classification pour différentes combinaisons de paramètres MFCC ..	37
Tableau III.5: Taux de classification pour différentes fonctions mère DB	38
Tableau III.6: Configuration optimale du classificateur KNN (avec les paramètres DWE) ...	40
Tableau III.7: Taux de classification pour différentes combinaisons de paramètres DWE ...	42
Tableau III.8: Taux de classification TCV et TCS	42
Tableau III.9: Taux de classification TCS	43

Liste des abréviations

- ECG** : L'électrocardiogramme.
- EMG** : Electromyogramme.
- EEG** : Electroencéphalogramme.
- KNN** : k-Nearest Neighbors.
- SVM** : Support Vector Machine.
- ANN** : Artificial Neural Network.
- HMM** : Hidden Markov Model.
- PLP** : La prédiction linéaire perceptive.
- MFCC** : Mel-Frequency Cepstral coefficients.
- LPC** : Linear Predictive Coding.
- LPCC** : Linear Prediction Cepstrum Coefficients.
- LDA** : Linear Discriminant Analysis.
- DWT** : Discrete Wavelet Transform.
- GMM** : Gaussian Mixtures Model.
- CNN** : Convolution Neural Network.
- MLP** : Multilayer Perceptron.
- RNN** : Recurrent Neural Networks.
- RBM** : Restricted Boltzmann Machine.
- DWE** : Discrete Wavelet Energy.
- DCT** : Discrete Cosine Transform.
- FFT**: *Fast Fourier Transform.*
- DFT** : Discret Fourier Transforme.
- HTK** : Hidden markov models ToolKit.
- EMO DB** : Berlin Data base of Emotional Speech



Introduction générale



Introduction générale

L'émotion joue un rôle important dans la communication humaine. Elle aide les êtres humains à comprendre les opinions des autres en transmettant des sentiments. Cela ouvre un nouveau champ de recherche appelé reconnaissance automatique des émotions (RAE), dont les objectifs fondamentaux sont de comprendre et d'identifier les émotions.

Actuellement, la reconnaissance automatique des émotions trouve son intérêt dans plusieurs domaines d'application tels que : la surveillance audio/vidéo, sécurité dans les aéroports, l'apprentissage basé sur le Web, les applications commerciales, les études cliniques, le divertissement, la banque, les centres d'appels, jeux informatiques et diagnostic psychiatrique. Bien que la perception humaine donne des bons résultats de reconnaissance des émotions à partir du signal vocal, la conception d'un système RAE performant reste un défi pour ces différentes applications. Les performances des systèmes RAE sont liées aux modalités utilisées dans ces différentes applications, telles que : les expressions faciales, le signal parole, les signaux physiologiques, etc. Néanmoins, le signal parole est une modalité moins coûteuse et moins contraignante.

Généralement, un système RAE consiste à reconnaître automatiquement l'état émotionnel à partir d'un signal d'entrée en utilisant les étapes de reconnaissance de formes telles que l'extraction de paramètres et la classification. La conception d'un tel système se base sur une phase d'apprentissage permettant de modéliser les différentes classes de données ainsi qu'une phase de test permettant de classifier des signaux de test et d'évaluer les performances du système sous une base de données de références. Ces deux phases exigent une étape d'extraction de paramètres permettant de convertir chaque signal en une séquence de vecteurs de paramètres. L'objectif de ces deux phases est de concevoir un système RAE performant. Le fonctionnement du système RAE conçu se déroule en une étape d'extraction de paramètres et une étape de classification utilisant un des classificateurs standard tels que KNN, SVM, ADI, ANN, etc. Ainsi, la conception d'un système performant consiste à choisir le meilleur classificateur et la bonne méthode d'extraction en termes de précision et de complexité. Couramment, le classificateur KNN est le classificateur le plus utilisé vu de sa simplicité et son efficacité.

Dans notre travail on s'intéresse à reconnaître l'état émotionnelle à partir du signal parole. Plus particulièrement, notre travail consiste à concevoir un système RAE basé sur l'extraction des paramètres acoustiques combinée avec le classificateur KNN et la stratégie de règle de vote. La proposition de cette combinaison est justifiée par l'application de cette stratégie de classification (KNN avec la règle de vote) dans plusieurs travaux de recherches récents tels que : l'identification des appareils électriques [1], le diagnostic de la maladie de parkinson [2], et la classification des signaux PCG (phono cardiogramme) [3]. L'extraction de paramètres se base sur deux types de descripteurs : le descripteur DWE issue de l'analyse par ondelettes et le descripteur MFCC. Ainsi, ces deux descripteurs conduisent à implémenter deux systèmes RAE. L'évaluation de ces systèmes s'effectue sous la base de données (EMO-DB) qui considère sept états émotionnels (colère, dégoût, peur, joie, tristesse, neutralité et surprise). L'objectif principal de notre travail est de chercher la bonne configuration permettant d'obtenir les meilleures performances.

Le manuscrit de ce travail est réparti en 3 chapitres :

Le premier chapitre présente quelques notions sur les émotions, les différentes modalités de système de reconnaissance des émotions. Nous présenterons l'architecture d'un système de reconnaissance d'émotions utilisant la modalité du signal parole et nous décrirons ses différentes étapes de son fonctionnement ainsi que les principales méthodes de reconnaissances d'émotions. Nous décrivons également la technique de classification KNN.

Le deuxième chapitre donne un état de l'art sur les méthodes d'extraction de paramètres et décrit la méthode d'extraction des paramètres basée sur l'analyse par ondelettes ainsi que la méthode d'extraction des coefficients MFCC.

Le chapitre trois décrit le système RAE proposé et présente les étapes d'implémentation du système ainsi que les différentes expériences et résultats obtenus.

Finalement, nous terminons notre manuscrit par une conclusion générale.



Chapitre I

*Généralités sur la reconnaissance automatique des
émotions*



I.1 Introduction

Le signal vocal est un signal riche d'informations telles que le texte parlé, la langue, l'identité du locuteur, son genre, son âge et ses émotions. Les émotions permettent d'assurer une meilleure compréhension du message véhiculé entre des personnes, ainsi de s'adapter à une situation donnée. Les émotions jouent un rôle important dans la prise de décision, ils influencent également sur le comportement et la formation de la personnalité. L'objectif de notre travail est de proposer un système de reconnaissance automatique des émotions (RAE) à partir du signal parole. Pour accomplir cette tâche de reconnaissance, il est nécessaire de comprendre le phénomène émotionnel avant de pouvoir concevoir ce système. Ce chapitre consiste à donner des généralités sur les émotions, leurs types de classification, et également les différentes modalités des systèmes de reconnaissance des émotions.

I.2 Généralité sur les émotions

I.2.1 Définition des émotions

L'émotion selon Kleinginna et al [4] " est une interaction complexe entre des facteurs objectifs et subjectifs, présentée par un système neural/hormonal qui donne naissance à des expériences affectives telles que le sentiment d'excitation, de plaisir/déplaisir. Elle génère un processus cognitif comme les effets perceptuels émotionnels, le jugement et le processus de labellisation. Elle active un vaste ensemble de régulation pour les conditions d'excitation et entraîne un comportement souvent, mais pas toujours, expressif, dirigé vers un but et adaptatif [4].

La majorité des scientifiques s'accordent qu'une émotion présente par plusieurs composantes :

- **La composante comportementale expressive** : une expression (faciale, vocale ou encore posturale) permet l'interaction avec l'environnement et transmet les émotions
- **La composante physiologique** : une réaction corporelle (changement du rythme cardiaque, changement du corps musculaire, changement de la respiration)
- **La composante cognitive** : une évaluation cognitive déterminer la nature et l'intensité de la réaction.

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 Emotions primaires

Les Emotions primaires appelées souvent émotions de base sont définies par un ensemble des émotions universelles considérées comme naturelles qui change selon l'événement. Paul Ekman [5], affirme que les émotions sont classées en 7 catégories comme suit : la tristesse, le bonheur, la surprise, le dégoût, la peur, la neutralité et la colère. Le tableau I.1 regroupe les principales catégories des émotions de base

Tableau I.1: Principales catégories des émotions de base [6].

Auteurs	Emotions
EKMAN ET AL	Colère, dégoût, joie, tristesse, peur, surprise, neutralité.
PLUTCHIK	Acceptation, colère, anticipation, dégoût, peur, joie, tristesse, surprise
FRIDJA	Intérêt, joie, désir, chagrin, émerveillement
TOMKINS	Colère, intérêt, mépris, dégoût, détresse, peur, joie, honte, surprise
Arnold	Courage, colère, aversion, désir, désespoir, tristesse, amour, espoir, abattement, haine, peur
Mower	Souffrance, plaisir
Darwin	Colère, peur, joie, tristesse, dégoût

I.3.1.1. Bonheur : Est une émotion qui signifie un sentiment de satisfaction ou de plaisir. Parmi ses manifestations physiologiques nous citons [7] :

- Sourires ou expressions faciales liés à la félicité ;
- Posture du corps détendue, lâche et expressive ;
- Ton de voix vif, rapide et aigu.

I.3.1.2. Tristesse : La tristesse apparaît quand une personne souffre. L'origine de la tristesse est typiquement une perte de quelque chose. Cette émotion calme et non impulsive, est souvent accompagnée de larmes. Pendant l'émotion, les muscles du visage perdent la tension qui peut entraîner les caractéristiques physiologiques typiques suivantes [7] :

- Faible niveau d'énergie ;
- Besoin de se retirer des autres ;
- bouleversement émotionnel ;
- Sentiment de vide ou d'inutilité.

I.3.1.3. Dégoût : Le dégoût est une émotion négative généralement évoquée par l'odorat, le goût ou la vision. Contrairement aux autres émotions, les objets évoquant le dégoût ne sont pas universels, mais culturels ou personnels, par ex. aliments. La réponse physiologique extrême est le vomissement [8]. Cette émotion se distingue par les expressions suivantes [7] :

- Une réaction automatique pour rester à l'écart de l'objet qui génère du ressentiment ;
- Nausées ou vomissements ;
- Rincer le visage.

I.3.1.4. Colère : La colère est une réaction émotionnelle forte et peut également être une émotion dangereuse car elle pourrait provoquer la violence. La colère a un impact important sur tout le corps [8]. Cette émotion se distingue par les expressions suivantes :

- L'augmentation de la pression sanguine ;
- le visage rouge ;
- la tension dans les muscles généralement reflétée.

I.3.1.5. Surprise : La surprise est une émotion soudaine. Cela vient sans réfléchir et ne dure que peu de temps. On le développe dans des situations inattendues ou imprévues, cela n'implique pas toujours le côté positif. En fait, il y a trois manifestations possibles de cela : neutre, agréable et désagréable. Cette distinction est très importante, car la plupart des gens l'associent à la joie ou au bonheur. Cependant, un événement inattendu qui provoque de l'insatisfaction ou de la peur est aussi une expression de surprise. Parmi ses caractéristiques physiologiques externes, on souligne [7] :

- Sourcils levés ;
- Ouverture complète des paupières ;
- Mâchoire légèrement relâchée ;
- Réaction gestuelle qui accompagne la situation.

I.3.1.6. Peur : La peur est une émotion résultante d'un sentiment intense souvent désagréable, causée par un danger, présent ou futur réel ou imaginaire. La peur est l'une des principales émotions qui découlent naturellement de la menace [7]. Parmi les caractéristiques de l'émotion de peur :

- une augmentation du rythme cardiaque ;
- la raideur des muscles ;
- le besoin de se protéger.

En général, les peurs irrationnelles sont classées comme des phobies [7].

I.3.1.6. Neutralité : La neutralité est un état non émotionnel, comprenant le calme, la relaxation, la satisfaction, le contentement, la somnolence etc [9].

I.3.2 Émotion secondaire

Les émotions secondaires dites également complexes sont issues des émotions primaires et résultent d'un mélange de ces dernières : la colère peut par exemple donner l'agressivité, la haine à partir de la peur, la culpabilité, l'angoisse... Selon Ekman [10], elles sont nommées parfois des émotions mixtes. Contrairement aux émotions primaires, les émotions secondaires ne sont pas innées. Parmi ces émotions secondaires, on trouve aussi la jalousie, la culpabilité, l'embarras et l'envie [10].

I.4 Types d'émotions

Généralement, deux types d'émotions sont envisagés :

- Emotion positive ;
- Emotion négative.

I.4.1 Émotions positives

Les émotions positives appelées aussi émotions saines, affectent positivement le bien-être de la personne par exemple : la joie, l'amour.

I.4.2 Émotions négative

Les émotions négatives sont des émotions désagréables (La peur ou la tristesse) qui affecte négativement le bien-être des personnes.

I.5 Types de corpus des émotions

Nous distinguons essentiellement trois types de corpus émotionnels utilisés dans le domaine de la détection automatique des émotions : les émotions naturelles, simulées et induites.

I.5.1 Émotions naturelles

Les corpus naturels ou réalistes sont obtenus à partir des enregistrements spontanés et naturels. Ce corpus de données est caractérisé par une très haute validité écologique. L'inconvénient est que ces données sont très limitées en nombre de locuteurs, de courtes durées, souvent de Mauvaise qualité, en plus sont très difficiles à collecter et à étiqueter en classes d'émotions [11].

I.5.2 Émotions simulées

Une émotion simulée est une émotion produite par un professionnel ou un semi-professionnel sur la base d'un type d'émotion et/ou d'un scénario typique. La majorité des données utilisées dans les études menées sur la reconnaissance automatique des émotions (RAE) sont basées sur ce type des émotions

I.5.3 Émotions induites

Les émotions induites étaient à l'origine utilisées dans le domaine de la psychologie. Les émotions de cette catégorie sont générées expérimentalement dans un laboratoire à l'aide des techniques d'induction. Un nombre impressionnant de techniques d'induction ont été élaborées par des psychologues et parmi celles-ci, on peut citer la visualisation d'images animées ou de films, écoute de la musique, imagination...etc. Le tableau I.2 présente une comparaison entre les différents corpus émotionnels.

Tableaux I.2 : Comparaison entre les corpus émotionnelles [12].

Corpus	Réaliste	Qualité d'enregistrement	Reconnaissance par l'oreille
Naturel	Plus réaliste	Mauvaise qualité	N'est pas garantie
Acté	Hors contexte	Très bon qualité	plus garantie
Induit	Moins réaliste	bon qualité	Moins garantie

I.6 Reconnaissance automatique des émotions

La RAE est un domaine de recherche particulièrement dynamique qui couvre un large champ d'applications. Parmi les applications on trouve par exemple : le marketing, la sécurité, les systèmes embarqués pour contrôler l'état du conducteur de voiture, les systèmes de traduction automatique. Un système de RAE est utilisé aussi pour les applications nécessitant une interaction homme-machine, telles que les films Web, l'e-learning et les didacticiels. Il peut également être utilisé comme un outil de diagnostic pour les thérapeutes [11].

Il y a trois modalités des systèmes de reconnaissance d'émotions [13] :

- La reconnaissance des émotions à partir d'images fixes ou de vidéos (analyse d'expressions faciales) ;
- La reconnaissance des émotions à partir des signaux physiologiques ;
- La reconnaissance des émotions à partir de signal acoustique.

Dans notre travail, nous intéressons à la reconnaissance d'émotions par l'analyse de signal acoustique.

I.6.1 Reconnaissance des émotions à partir d'images fixes ou des vidéos (analyse d'expressions faciales)

Notre visage est riche d'informations (telles que l'identité individuelle, les expressions de communication verbale et non verbale, les émotions) à travers la direction du regard et les expressions faciales. La reconnaissance des émotions à partir des expressions faciales est la technique la plus simple et la plus ancienne. Elle a une longue histoire car on pense qu'elles laissent entrevoir les aspects importants des émotions. De nombreuses approches ont donc été

introduites pour classifier les émotions à partir des expressions faciales. L'idée principale derrière toutes ces approches réside dans le fait que les signes émotionnels peuvent être découverts à partir du positionnement spatial de points et de régions spécifiques du visage [14]. Le tableau ci-dessous donne les expressions faciales des émotions primaires.

Tableau I. 2 : Expressions faciales correspondant à chaque émotion [15].

EMOTION	EXPRESSION FACIAL
Joie	<ul style="list-style-type: none"> • Le coin des lèvres est élargi • Joux relevées
Tristesse	<ul style="list-style-type: none"> • Triangulation de la peau sous le sourcil • Paupière supérieure tirée vers l'intérieur
Colère	<ul style="list-style-type: none"> • Regard fixe • Narines dilatées • Lèvres pressées fermement
Peur	<ul style="list-style-type: none"> • Sourcils levés • Bouche ouverte
Dégoût	<ul style="list-style-type: none"> • Lèvres relevées • Nez plissé
Surprise	<ul style="list-style-type: none"> • Sourcils hauts et courbés • Blanc des yeux plus visible

I.6.2 Reconnaissance des émotions à partir les signaux physiologiques

Bien que les signaux physiologiques ne soient pas directement liés aux émotions, les informations qu'ils contiennent peuvent être utilisées pour la reconnaissance des émotions. Ils peuvent être classés en deux grandes catégories, en fonction de leur point d'origine [16]. La première catégorie de signaux provient du système nerveux périphérique, comme l'électrocardiogramme (ECG) et la conductivité de la peau (EMG), tandis que la seconde provient du système nerveux central, comme les signaux du cerveau (EEG).

I.6.3 Système de reconnaissance des émotions à partir de signal acoustique de la parole

Ce type de reconnaissance se base sur une analyse qui tient en compte le mécanisme de génération du signal vocal, en extrayant des caractéristiques (paramètres, attributs,...) contenant des informations émotionnelles de la voix du locuteur, et en utilisant des méthodes de reconnaissance de formes appropriées pour identifier les états émotionnels. Généralement, un système de reconnaissance acoustique des émotions se base sur deux étapes principales : extraction des paramètres et classification. Les différentes étapes constituant un système de reconnaissance acoustique des émotions sont illustrées sur la figure I-1

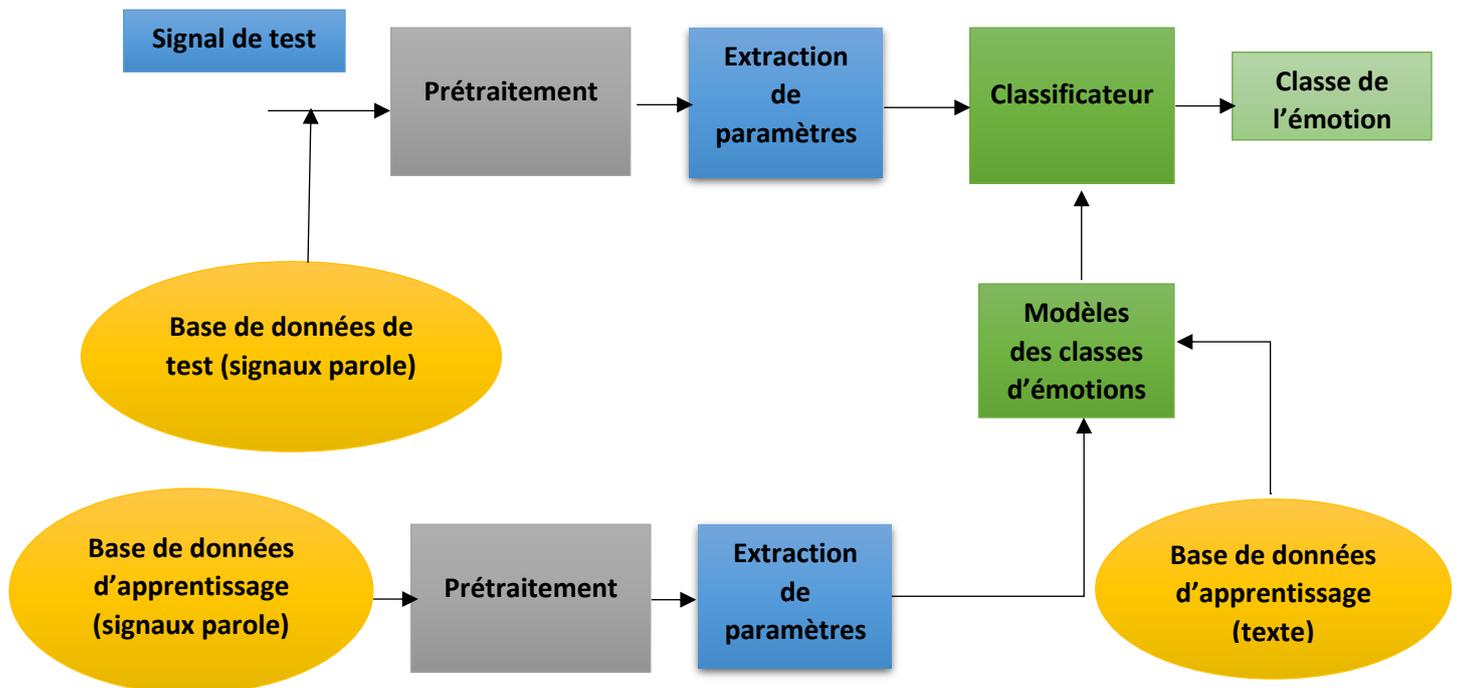


Figure I.1 : Architecture d'un système RAE [17].

La conception d'un système de reconnaissance automatique des émotions se déroule en deux phases, une phase d'apprentissage et une phase de test. La phase d'apprentissage a pour objectif d'apprendre les modèles classes ou de références. La phase de test consiste à évaluer les performances du système RAE en utilisant des mesures de performances telles que le taux de classification. Cette phase exige une étape de reconnaissance qui consiste à utiliser les références ou les modèles entraînés durant la phase d'apprentissage, pour classifier chaque

signal de la base de test en utilisant des algorithmes de classification standards tels que (KNN, SVM, ANN et HMM....).

- **Prétraitement**

Une étape de prétraitement de signal a pour objectif d'avoir un signal convenable pour la tâche de classification. Cette étape a pour rôle de détecter l'intervalle de fonctionnement en extrayant le début et la fin à partir des fichiers texte de la base de données.

- **Bloc de fenêtrage**

Il est difficile de traiter un signal non stationnaire comme le signal de la parole sans le segmenter en trames. Une analyse à court terme montre que le signal vocal est quasi stationnaire sur des tranches temporelles de durées de 10 à 30 ms. Cette analyse est effectuée à l'aide des fenêtres telles que [18] : fenêtre Hamming et fenêtre Hanning.

- **Etape d'extraction des paramètres**

La phase d'apprentissage et la phase de test exigent chacune une étape d'extraction des paramètres. Généralement, cette étape découpe chaque signal en une séquence de fenêtres d'analyse chevauchées. Ensuite, elle convertit chaque fenêtre d'analyse en un vecteur composé d'un ensemble de coefficients en appliquant des différentes méthodes d'extraction de paramètres telles que : la prédiction linéaire perceptive (PLP) [19], l'extraction des coefficients cepstraux en échelle fréquentielle Mel (MFCC) [19], l'extraction des coefficients cepstraux basée la prédiction linéaire (LPCC) [20], l'extraction des coefficients RASTAPLP [21], l'analyse discriminante linéaire (LDA) [22], la transformée en ondelettes discrète (DWT) [22].

Dans notre travail on s'intéresse à l'extraction des paramètres du signal acoustique par l'application de la transformée en ondelettes discrète (DWT) et à l'extraction des coefficients MFCC.

- **Classificateur**

Le classificateur est utilisé pour identifier la classe émotionnelle. Dans cette étape, l'état émotionnel peut être identifié par plusieurs classificateurs. Les classificateurs les plus utilisés sont : les k-plus proche voisins(KNN), les modèles de Markov cachés (HMM), les réseaux de neurones artificiels (ANN : artificiel neural network), modèles de mélanges de gaussiennes

(GMM), Support Vector Machine (SVM), Various Neural Network model like Convolution Neural Network (CNN), Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Restricted Boltzmann Machine (RBM) [23].

Dans notre travail on s'intéresse à utiliser le classificateur KNN vue de sa simplicité. Ce classificateur est combiné avec la règle de vote pour identifier la classe émotionnelle correspondant au signal d'entrée du système.

I.7 Classificateur KNN

L'algorithme des KNN appartient à la famille des algorithmes d'apprentissage automatique. La méthode K-NN [24] se base sur une idée fondatrice simple, qui consiste en 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 consiste en un calcul de distances entre ces instances. Puis, l'instance à classer est assignée la classe majoritaire parmi les classes des k instances les plus proches. Les distances utilisées dans la méthode KNN sont : Euclidienne, Minkowski, Hamming, Cityblock, Cosinus, Corrélation, et Gaussienne.

Dans notre travail on utilise les distances suivantes : Euclidienne, Cityblock, Cosinus, Corrélation.

a). Distance Euclidienne

La distance Euclidienne entre les deux vecteurs X et Y est définie par :

$$d_E(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad \text{I.1}$$

b). Distance de Cityblock

Cette distance est définie comme suit :

$$d_E(X, Y) = \sum_{i=1}^m |x_i - y_i| \quad \text{I.2}$$

c). Distance Cosinus

Cette distance est définie comme suit :

$$d_{\cos}(X, Y) = 1 - \frac{\sum_{i=1}^m X_i Y_i}{\sqrt{\sum_{i=1}^m X_i^2} \sqrt{\sum_{i=1}^m Y_i^2}} \quad \text{I.3}$$

d). Distance corrélation

Cette distance est définie comme suit :

$$d_{\text{cor}}(X, Y) = 1 - \frac{\sum_{i=1}^m (X_i - \bar{Y}_i)}{\sqrt{\sum_{i=1}^m (X_i - Y_i)^2}} \frac{(X_i - Y_i)}{\sqrt{\sum_{i=1}^m (X_i - Y_i)^2}} \quad \text{I. 4}$$

I.8 Comparaison entre les différentes modalités utilisées pour reconnaissance des émotions

Dans cette section, nous présenterons une comparaison entre les différents systèmes de reconnaissance des émotions en termes de précision. Le tableau I.4 présente les avantages et les inconvénients de chaque système.

Tableaux I 3 : Comparaison entre les systèmes de reconnaissance des émotions [15].

Système de reconnaissance	Précision	Avantages	Inconvénient
Expression facial	90 %	Le Calcul facile est la technique la plus utilisée.	Dans certaines situations, les gens agissent également dans le seul but de tromper les autres.
Signal acoustique	70 %	La configuration expérimentale requise est moindre	La parole ne peut pas toujours être associée à l'état émotionnel
ECG	67 %	Le signal ne peut être masqué car il n'est pas sous le contrôle de l'utilisateur.	Il est difficile d'étiqueter les émotions de moindre intensité
EMG	-	L'acquisition du signal est relativement facile	Les mesures de la valence et de l'excitation sont obtenues à partir de deux signaux différents.
EEG	98 %	Les résultats sont plus précis car les signaux du cerveau sont utilisés directement.	Le calcul est complexe et coûteux

I.9 Domaines d'applications de la reconnaissance des émotions

La reconnaissance automatique d'émotions trouve son intérêt dans plusieurs domaines d'applications tels que [25] [26] [27] [28] [29]:

- **Marketing** : des applications pour mesurer la satisfaction des clients, prévoir les produits qui intéressent.
- **Sécurité** : détection du stress, Surveillance de sécurité.
- **Interaction homme-machine** : Robot d'escorte, voiture intelligente.
- **Education** : apprentissage à distance, programmes éducatifs.
- **Commerce** : Relations avec les consommateurs.
- **Médecine** : Le diagnostic psychiatrique, Autisme, La musicothérapie pour la dépression
- **Surveillance** : Pilote -Car - détecter l'état du pilote et de l'alerte.
- **Divertissement** : Joueur de musique : Reconnaître l'humeur et l'émotion de l'utilisateur et de ses besoins.
- **Loi** : Dans la détection du mensonge.

I.10 Conclusion

Dans ce chapitre, nous avons introduit brièvement des généralités sur les émotions, ensuite nous avons décrit les différentes étapes d'un système de reconnaissance automatique des émotions à partir du signal acoustique. Ainsi, nous avons donné un état de l'art sur les différentes sur modalités des systèmes de reconnaissance des émotions, puis nous avons terminé ce chapitre par l'importance d'utilisation de système de reconnaissance des émotions dans plusieurs domaines d'applications.



Chapitre II

Extraction des paramètres acoustiques



II.1 Introduction

Le fonctionnement d'un système RAE se déroule en deux phases principales, une phase d'apprentissage et une phase de test. Ces deux phases exigent une étape d'extraction de paramètres qui permet de convertir chaque signal en une séquence de vecteurs de paramètres. Cette séquence sera classifiée par un des classificateurs standards pour reconnaître la classe de l'état émotionnel correspondant au signal parole d'entrée.

Dans ce chapitre, nous allons présenter les méthodes les plus utilisées pour l'extraction des paramètres pertinents du signal parole pour la reconnaissance automatique des émotions (RAE), à savoir : Méthode basée sur l'analyse par ondelettes (DWE), Méthode basée sur le descripteur MFCC.

II.2 Etat de l'Art sur les techniques d'extraction de paramètre pour la reconnaissance automatique des émotions

La recherche sur la reconnaissance des émotions de la parole est passionnante depuis longtemps et plusieurs articles de recherche ont présenté différentes manières pour le développement des systèmes de reconnaissance des émotions humaines. Un système RAE se base sur des techniques d'extraction de paramètres ainsi que sur des algorithmes de classification. Le tableau II.1 montre le résultat en termes de taux de reconnaissance pour les différentes méthodes d'extraction de paramètres, de classification, ainsi que de la base de données utilisée

II.3 Extraction de paramètres

Le signal acoustique des émotions contient un grand nombre de paramètres qui reflètent les caractéristiques émotionnelles. L'extraction de ces caractéristiques est une étape importante dans le système RAE pour convertir chaque signal d'émotion en une séquence de vecteurs de caractéristiques obtenus en divisant le signal en fenêtres chevauchantes et calculer à partir de chaque fenêtre un ensemble de caractéristiques qui constitue le vecteur de caractéristiques. Cette étape s'effectue en plusieurs étapes secondaires à savoir : le prétraitement, le fenêtrage, et l'analyse à court terme.

Tableaux II .1: Comparaison des méthodes apparentées

Reference	La base des donnes	Extraction des paramètres	Méthode de classification	Résultat CR%
[30]	Emo-DB	Energie + MFCC + ZCR + Probabilité de vocalisation + Fréquence fondamentale	FCBF + MVT	84.19
[31]	Spanish	MFCC + MS	RNN	90.05
[32]	Emo-DB	GCZCMT	SVM	84.45
[33]	Emo-DB	Prosodique + sous-bande + MFCC + LPC	SFS algorithme + SVM	83.00
[34]	eNTERFACE'05	Energie + Pitch + formants, intensité + ZCR + texte	SVM	90.00
[35]	CASIA	Prosodique + les caractéristiques de qualité + MFCC	analyse de corrélation + Fisher + arbre de décision ELM	89.60
[36]	IEMOCAP	IS10 + MFCCs + eGemaps+SoundNet + VGGish	SVM	64.00
[37]	Emo-DB	Caractéristiques prosodiques + caractéristiques paralinguistiques	SVM	94.90
[38]	CASIA	MFCC	GA-BEL + PCA + LDA	90.28
[39]	CASIA	MFCC, pitch, énergie à court terme + ZCR	SVM + DBN	95.80
[40]	Emo-DB	eGeMAPS	Système d'empilage CSS +SVM	82.45
[41]	Emo-DB	MFCCs	ensemble ensaché de SVMs	92.45
[42]	RAVDESS	CWT, les coefficients prosodiques	SVM	60.10
[43]	Spanish	SMFCC, ECC, MFF, MS et EFCC	RNN	91.16
[44]	RAVDESS/SAVEE	Prosodique + spectrale	RDF	99.55
[45]	Emo-DB	MFCC, PLP, pitch, Energie, LPCC	GMM	86.05

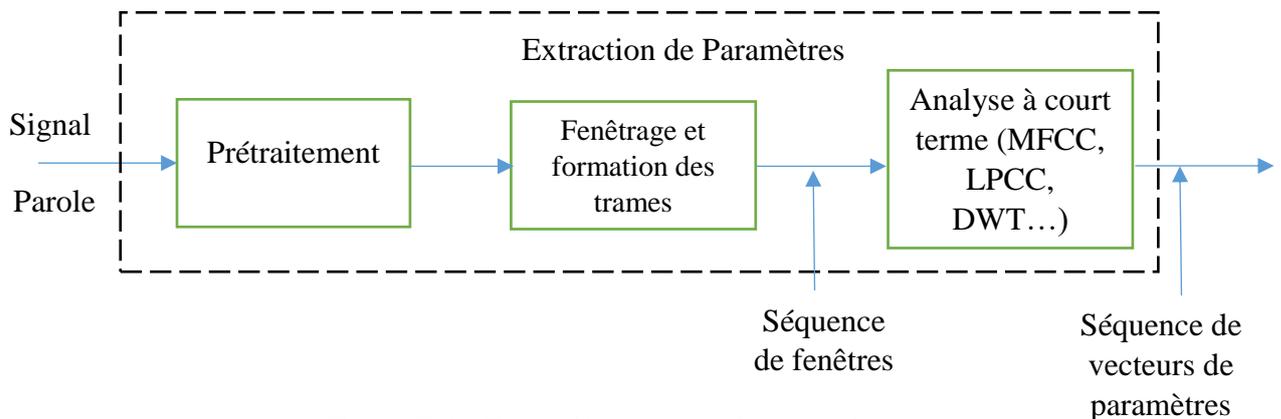


Figure II.1 : Etape d'extraction de paramètres.

II.4 Extraction des paramètres acoustique basée sur la transformée en ondelettes

La transformation en ondelettes est utilisée en traitement de signal afin de répondre à la problématique de la résolution statique de la transformée de Fourier à court terme. Bien que le concept fut proposé par Alfred Haar en 1910, c'est Jean Morlet qui a été le premier à l'avoir utilisé et lui a donné son nom [46]. La transformée en ondelettes constitue une méthode d'analyse qui découpe les données et les signaux en composantes fréquentielles suivant une résolution adaptée. Elle peut être continue ou discrète comme dans le cas de la transformée de Fourier, et il en existe plusieurs variétés

Le principe est le suivant : une famille de fonctions élémentaires est construite à partir d'une fonction de base $\psi(t)$ qu'on appelle ondelette-mère ou analysante. Cette famille est générée par l'expression suivante [47] :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right); \quad a > 0, b \in \mathbb{R} \quad \text{II.1}$$

Les deux paramètres a et b de cette expression sont respectivement les paramètres d'échelle et de translation.

Il existe plusieurs familles d'ondelettes (voir Fig. II.2).

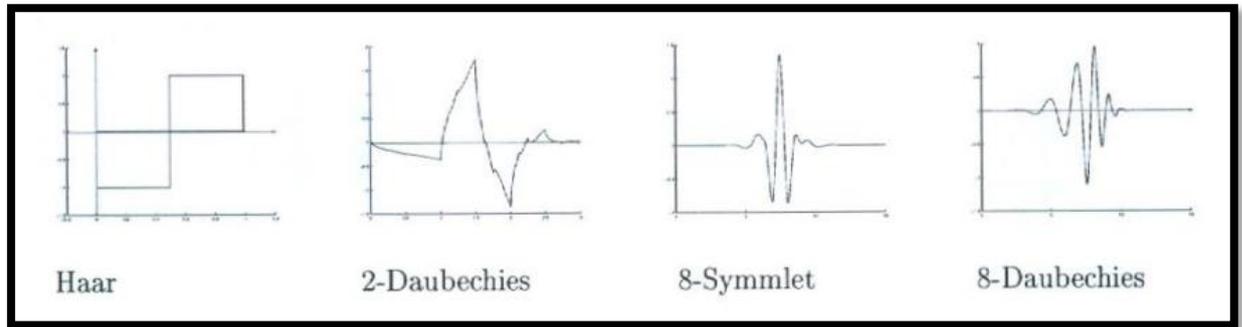


Figure II.2 : Exemples d'ondelettes mères $\psi(t)$ [48].

La transformation en ondelettes d'un signal $x(t)$ est donnée par l'équation II.2 [49].

$$X(a, b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}(t)dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right) dt \quad \text{II.2}$$

La fonction, ou l'ondelette ψ joue le rôle de la fenêtre d'analyse dans la transformée de Fourier, mais leurs caractéristiques sont totalement différentes. En effet, contrairement aux fenêtres d'analyse, les ondelettes sont de nature oscillante et leur intégrale est nulle. Afin qu'elles soient performantes en termes de résolution en temps et fréquence, elles convergent vers 0 assez rapidement. En pratique, l'ondelette doit en même temps osciller puis s'amortir rapidement.

La transformée en ondelettes est souvent utilisée pour le traitement du signal et aussi pour l'analyse de signal dans les systèmes de reconnaissance de formes afin d'extraire un ensemble de paramètres de chaque fenêtre d'analyse (signal acoustique, image ...). La durée de chaque fenêtre d'analyse est généralement égale à 30 ms. Le succès de cette technique repose essentiellement sur la flexibilité offerte par l'ondelette mère dont le choix dépend des propriétés des données ainsi que de l'application visée [50].

En outre, la transformation en ondelettes est particulièrement adaptée à l'analyse de données provenant du monde réel en raison de sa propriété d'échelle qui est conforme à de nombreux comportements physiques : les événements de longue durée doivent être analysés avec des formes d'onde longue et lente, tandis que les événements de courte durée ou impulsifs doivent être analysés avec des formes d'onde courtes et rapide. La transformée en ondelettes est souvent utilisée pour le calcul des énergies normalisées à partir des coefficients d'ondelettes DWE (Discrete Wavelet Energy). Ce calcul permet d'avoir une bonne réduction de la dimensionnalité.

L'ensemble de N coefficients d'ondelettes (N étant le nombre d'échantillons) est réduit à un ensemble de $p + 1$ (p est le niveau maximal de la décomposition dyadique) valeurs d'énergie qui sont définies par [51] [52] comme suit :

$$\begin{cases} E(d_j) = \frac{1}{E_T} \sum_{n=0}^{n=N_j-1} |d_j[n]|^2 \\ E(a_p) = \frac{1}{E_T} \sum_{n=0}^{n=N_p-1} |a_p[n]|^2 \end{cases} ; j = 1 \dots p \quad \text{II.3}$$

L'entier N_j est le nombre de coefficients d'ondelettes qui sont calculés à chaque échelle j , $N_j = \frac{N}{2^j}$. $a_p[n]$ sont les coefficients d'approximation et $d_j[n]$ sont les coefficients de détail au niveau j . Les valeurs d'énergies sont normalisées par rapport à l'énergie totale (E_T).

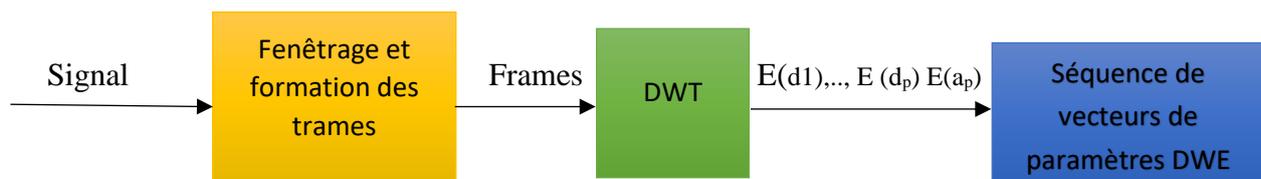


Figure II.3 : Processus d'extraction de paramètres DWE [50].

Dans notre système, on propose d'appliquer les coefficients DWE au système de reconnaissance acoustique des émotions basé sur le classificateur KNN combiné avec la méthode des règles de vote.

II.5 Coefficients cepstraux sur l'échelle Mel (MFCC)

Les coefficients cepstraux en échelle Mel (MFCC pour Mel-Frequency Cepstral Coefficients en anglais) [53], sont des coefficients très souvent utilisés en reconnaissance automatique de la parole et du locuteur. L'extraction des coefficients MFCC utilise une échelle fréquentielle non-linéaire ou échelle Mel.

La fréquence Mel-échelle est définie par [54] :

$$B(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad \text{II.4}$$

Où f est la fréquence en Hz, B est la fréquence en échelle Mel.

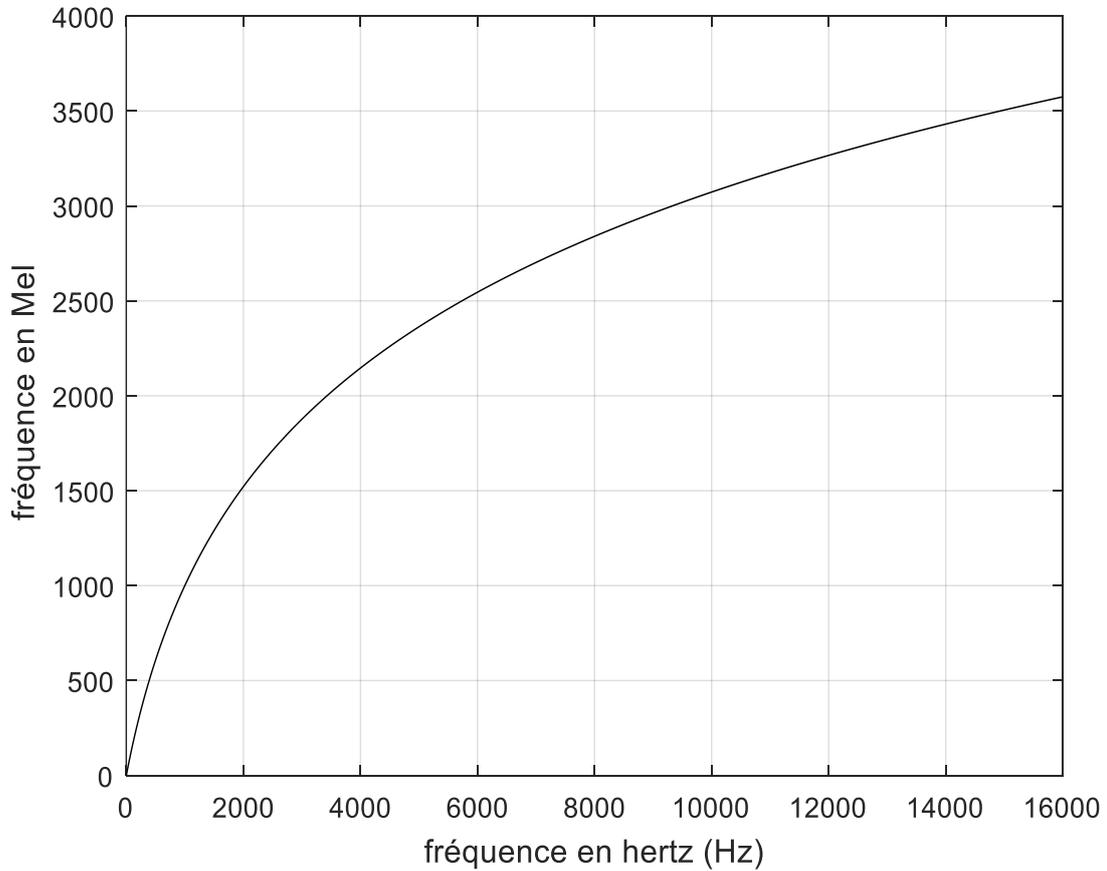


Figure II.4 : Exemple de conversion des hertz en Mel

II.5.1 Étapes de calcul du vecteur caractéristique de type MFCC

Dans ce qui suit, nous décrivons chacune des étapes nécessaires pour l'obtention d'un vecteur de coefficients MFCC, tel qu'illustré par la Figure II.5

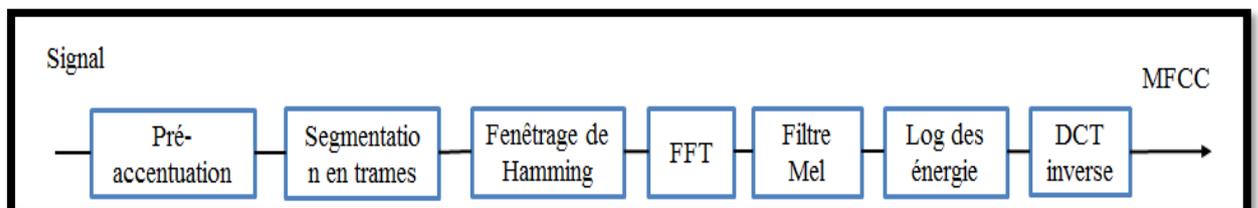


Figure II.5 : Étapes de calcul d'un vecteur caractéristique de type MFCC [55].

a. Segmentation en trame (Frame blocking) :

Le signal acoustique continu est segmenté en trames de N échantillons, avec un pas d'avancement de M trames ($M < N$), c'est-à-dire que deux trames consécutives se chevauchent sur $N - M$ échantillons. Comme prétraitement, il est d'usage de procéder à la préaccentuation du signal en appliquant l'équation de différence du premier ordre aux échantillons $x(n)$ avec $0 \leq n \leq N - 1$ [55].

$$x'(n) = x(n) - kx(n - 1), 0 \leq n \leq N - 1 \quad \text{II.5}$$

K : représente un coefficient de préaccentuation qui peut prendre une valeur dans l'étendue $0 \leq K \leq 1$

b. Fenêtrage :

Si nous définissons $w(n)$ comme fenêtre où $0 < n < N - 1$ et N représente le nombre d'échantillons dans chacune des trames, alors le résultat du fenêtrage est le signal x_a , donné par la formule (II.6) [55].

$$x_e(n) = x(n)w(n), 0 \leq n \leq N - 1 \quad \text{II.6}$$

La fenêtre la plus utilisée est la fenêtre de Hamming dont sa fonction est donnée par la formule suivante :

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N - 1 \quad \text{II.7}$$

c. Calcul de la transformée de Fourier rapide (Fast Fourier Transform, FFT) :

Au cours de cette étape chacune des trames, de N valeurs, est convertie du domaine temporel au domaine fréquentiel. La FFT est un algorithme rapide pour le calcul de la transformée de Fourier discret (DFT) et est définie comme suit, Les valeurs obtenues sont appelées le spectre [56].

$$X[k] = \sum_{n=0}^{N-1} x_a[n] e^{-j \frac{2\pi nk}{N}}, 0 \leq k \leq N - 1 \quad \text{II.8}$$

En général, les valeurs $X[k]$ sont des nombres complexes et nous utilisons que leurs valeurs absolues (énergie de la fréquence) [56].

d. Filtrage sur l'échelle Mel :

Le spectre d'amplitude est pondéré par un banc de M filtres triangulaires espacés selon l'échelle Mel. Dans l'échelle de mesure Mel, la correspondance est approximativement linéaire

sur les fréquences au-dessous de 1kHz et logarithmique sur les fréquences supérieures à celle-ci. Cette relation est donnée par la formule II.4 [56] [11] :

Logarithme de l'énergie de chaque filtre est calculé selon l'équation [55] [56] :

$$S[m] = \ln \left[\sum_{i=0}^{N-1} X_a[k] H_m[k] \right], \quad 0 < m \leq M \quad \text{II.9}$$

e. Calcul du cepstre en échelle Mel

Le cepstre sur l'échelle de fréquence Mel est obtenu par le calcul de la transformée en cosinus discrète du logarithme de la sortie des M filtres (reconversion du log-Mel-spectre vers le domaine temporel) [56].

$$c[n] = \sum S[m] \cos(\pi n(m - 1/2)/M), \quad 0 \leq n < M \quad \text{II.10}$$

Le premier coefficient, $c[0]$, représente l'énergie moyenne dans la trame de la parole $c[1]$ reflète la balance d'énergie entre les basses et hautes fréquences ; pour $i > 1$, $c[i]$ représente des détails spectraux de plus en plus fins [11].

II.5.2 Calcul des caractéristiques dynamiques des MFCC

La dynamique du signal parole joue un rôle important dans la perception humaine. Cette dynamique peut être expliquée par les changements temporels dans le cepstre (c). Ces changements peuvent être mesurés par les dérivées des coefficients des MFCC statiques. Considérons, $C_k(t)$ 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 [55] :

$$\Delta C_k(t) = \frac{\sum_{i=1}^{i=T} i (C_k(t+i) - C_k(t-i))}{2 \sum_{i=1}^T i^2} \quad \text{II.11}$$

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

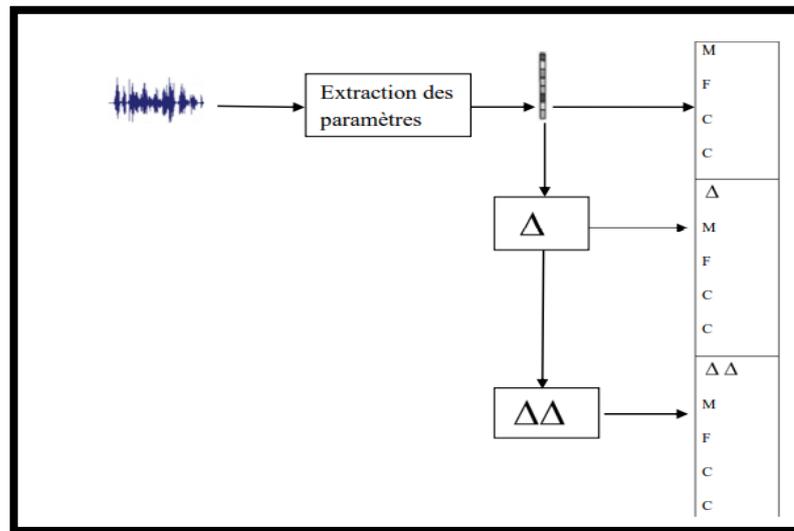


Figure II.6 : Calcul des dérivées premières et secondes de coefficients MFCC [54].

Les MFCC donnent des meilleurs résultats dans les systèmes de RAE, ce qui montre plus généralement l'intérêt d'un prétraitement par banc de filtres, d'une échelle fréquentielle non linéaire, et de la représentation cepstrale [45]. Dans ce travail, on s'intéresse à appliquer cette méthode d'analyse acoustiques pour d'extraire les coefficients MFCC, l'énergie et leurs paramètres dynamiques en utilisant la commande HCopy de librairie de HTK [57]. Ces coefficients seront extraits pour la tâche de reconnaissance acoustique des émotions.

II.6 Conclusion

Dans ce chapitre nous avons donné l'état de l'art sur les techniques d'extraction de paramètres pour la reconnaissance automatique des émotions à partir du signal parole. Puis, nous avons cité les étapes d'extraction de paramètres d'un système de RAE. Ensuite, nous avons détaillé l'extraction de paramètres basée sur l'analyse par ondelette DWE. Finalement, nous avons décrit les différentes étapes de calcul d'un vecteur de caractéristiques de type MFCC ainsi que le calcul de leurs paramètres dynamiques des MFCC (Δ et $\Delta\Delta$).



Chapitre III

*Systeme de reconnaissance acoustique d'émotions
basé sur le classificateur KNN combiné avec la
règle de vote*



III.1 Introduction

La reconnaissance acoustique d'émotions a pour objectif d'identifier l'état émotionnel d'une personne à partir du signal vocal. La conception d'un tel système se base sur une phase d'apprentissage permettant de modéliser les différentes classes d'émotions et une phase de test permettant de reconnaître les classes de signaux et d'évaluer les performances du système de reconnaissance. Ces deux phases exigent ainsi une base de données d'apprentissage et une base de données de test. De plus chacune, des phases exige une étape d'extraction de paramètres permettant de convertir chaque signal en une séquence de vecteurs de paramètres. Dans notre travail on propose la conception d'un système de reconnaissance acoustique d'émotion basé sur le classificateur KNN combiné avec la stratégie de la règle de vote. Deux méthodes d'extraction de paramètres sont utilisées. La première méthode convertit chaque fenêtre d'analyse du signal en un ensemble de coefficients MFCC concaténés avec l'énergie et leurs paramètres dynamiques Δ et $\Delta\Delta$. La deuxième méthode convertit chaque fenêtre en un ensemble de paramètres énergétiques calculés sur chaque niveau de décomposition dyadique de la transformée discrète par ondelettes, concaténés avec leurs paramètres dynamiques Δ et $\Delta\Delta$. Les deux types de paramètres extraits par les deux méthodes sont appelés respectivement descripteur MFCC et descripteur DWE (Discret Wavelet Energy).

Ainsi, deux systèmes RAE sont proposés. Le premier se base sur l'extraction des paramètres MFCC combinée avec le classificateur KNN et la règle de vote. Le deuxième se base sur l'extraction des paramètres DWE combinée avec le classificateur KNN et la règle de vote.

Les questions les plus importantes pour chercher la bonne configuration des deux systèmes proposés sont comme suit :

- 1- Comment implémenter un système de reconnaissance acoustique des émotions basé sur la classificateur KNN combiné avec la stratégie de la règle de vote ?
- 2- Quelle est la meilleure combinaison des paramètres MFCC ?
- 3- Quelle est la meilleure fonction mère d'ondelette et le meilleur niveau de décomposition par ondelettes ?
- 4- Quelle est la configuration optimale du classificateur KNN permettant d'obtenir les meilleures performances du système RAE pour chaque système proposé ?

- 5- Est-ce que l'utilisation de la règle de vote contribue à l'amélioration des performances du système reconnaissance acoustique des émotions ?
- 6- Quel est le meilleur système RAE parmi les deux systèmes proposés.

Différentes expériences sont effectuées pour répondre à ces questions, La réponse sur la première question consiste à décrire les différentes étapes permettant d'implémenter un système sous l'environnement de programmation Matlab. Le premier système se base sur l'extraction de paramètres MFCC extraits en utilisant la boîte à outils HTK. Le deuxième système se base sur l'extraction de paramètres DWE extraits en utilisant des fonctions en Matlab permettant de calculer les coefficients de détails et d'approximations et calculer l'énergie dans chaque niveau de décomposition. De plus les paramètres dynamiques de ce dernier descripteur sont extraits en utilisant la boîte à outils HTK. Les performances des systèmes proposés sont évaluées en utilisant la base de données de références EMO-dB (Berlin Data base of Emotional Speech) qui sera décrite brièvement dans la section suivante.

III.2 Description de la base de données

Plusieurs travaux de recherches ont utilisé des bases de données contenant des signaux parole pour évaluer les performances des systèmes RAE. La base de données EMO-DB est couramment utilisée pour ce type d'évaluation [30] [32] [33] [37] [41] [45], vue de sa richesse en vocabulaire, nombre de locuteurs, qualité de signaux, différentes durées des phrases. Ainsi, dans notre travail, nous avons utilisé cette base de données.

La base de données est composée de 10 phrases de textes différents (5 phrases courtes constituent un ensemble A et 5 phrases longues constituent un ensemble B) prononcées en langue allemand par 10 acteurs (5 hommes, 5 femmes) qui simulent sept états émotionnels primaires ('*anger*', '*disgust*', '*fear*', '*happiness*', '*neutral*', '*sadness*' and '*surprise*') qui peuvent être traduites respectivement par : ('colère', 'dégoût', 'peur', 'joie', 'neutre', 'tristesse' et 'surprise').

La collection totale des signaux se compose de 800 énoncés dont 535 énoncés sont considérés par une validation humaine (20 auditeurs). Ainsi, la base de données considérée dans notre travail est constituée de 535 signaux représentant sept classes d'émotions.

Le tableau III.1 illustre la répartition des enregistrements de phrases entre les 7 états émotionnels ainsi que le nombre d'occurrences utilisées respectivement pour les phases de test

et d'apprentissage. Les enregistrements ont été effectués premièrement avec une fréquence d'échantillonnage de 48 kHz et sous-échantillonnés ensuite à 16 kHz.

Dans notre travail, l'ensemble A composé de 277 énoncés (51.78% de la base totale) est pris comme base de données d'apprentissage, et l'ensemble B composé de 258 énoncés (48.22% de la base totale) est pris comme base de données de test. Ainsi, notre système RAE conçu fonctionne en mode indépendant du texte.

Tableaux III.1 : Répartition des phrases d'EMO-DB

Classe \ Base	Anger	Boredom	Disgust	Fear	Hapines	Neutral	Sadness	Totale
APP	65	41	25	35	38	41	32	277
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 différentes étapes de conception de ce système sont données comme suit :

- Préparation de données d'apprentissage et de test.
- Extraction des paramètres durant la phase d'apprentissage et la phase de test.
- Enregistrement des vecteurs de références et leurs classes d'émotions durant la phase d'apprentissage.
- Classification des signaux de test en appliquant l'algorithme de classification KNN combinée avec la règle de vote.
- Evaluation des performances du système en termes de taux de classification.

Le schéma synoptique d'implémentation de notre système RAE est illustré sur la figure III.1.

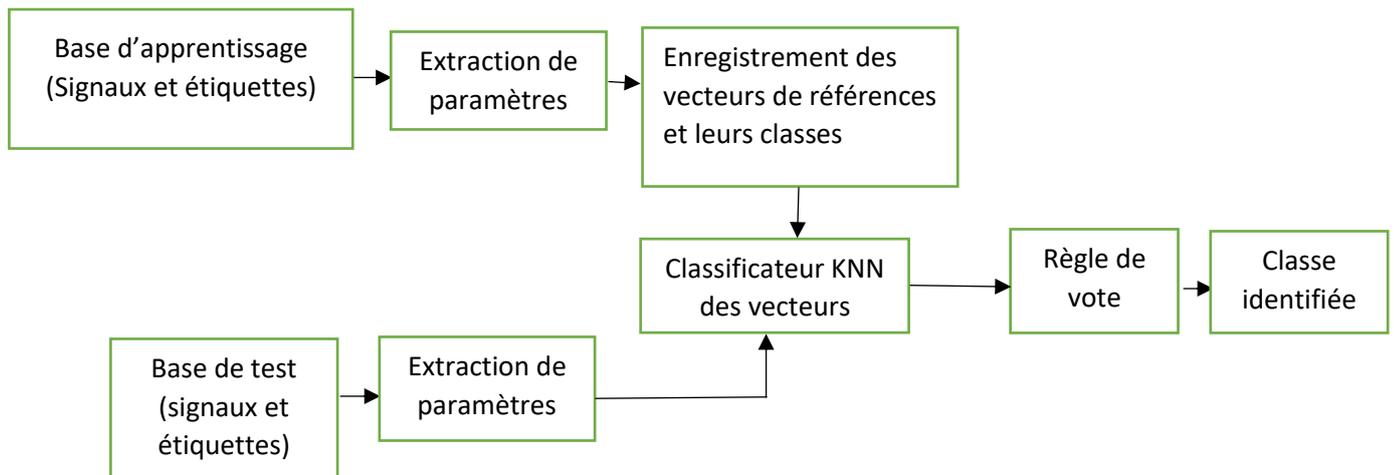


Figure III .1 : Schéma synoptique du système RAE proposé

Les différentes étapes des phases d'apprentissage et de test seront décrites dans les deux sections suivantes.

III.3.1 Phase d'apprentissage

Cette phase consiste premièrement à convertir chaque signal parole d'apprentissage en une séquence de vecteurs de paramètres (MFCC ou DWE) et d'enregistrer ensuite toutes les séquences des vecteurs dans une matrice appelée matrice de paramètres d'apprentissage dont chaque ligne est un vecteur de paramètres représentant une ligne d'analyse, et d'enregistrer également les classes des vecteurs 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.3.1.1. Extraction des paramètres

Cette étape consiste à découper premièrement le signal parole en une séquence de fenêtres d'analyse de 30 ms chevauchées de 20 ms (66.67%) et d'extraire ensuite pour chaque fenêtre d'analyse un ensemble de paramètres. Les deux descripteurs considérés dans notre travail sont les paramètres MFCC et les paramètres DWE.

Dans le premier descripteur, chaque vecteur est constitué et 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 $\Delta\Delta$ seront notés respectivement par D et A. Ainsi, chaque vecteur est constitué de 39 paramètres. Ces paramètres sont extraits en appliquant l'outil Hcopy de librairie HTK [57] sur le fichier de type **wav** représentant le signal d'entrée. La commande suivante permet de convertir un signal d'entrée en une séquence de vecteurs de 39 paramètres MFCC :

Le fichier de configuration de l'outil HCopy est représenté dans le tableau III-2. La commande qui permet l'extraction de ces paramètres est la suivante :

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

Tel que :

- **analysis.conf**: est un fichier de type texte permettant la configuration du processus d'extraction des paramètres acoustiques MFCC. Un exemple d'un fichier de configuration est illustré dans le tableau III-2.
- **fichier_son.wav** : est un fichier son de type **wav** représentant 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:** La période d'échantillonnage.

Dans le deuxième descripteur DWE, chaque vecteur de paramètres est constitué de l'ensemble des énergies normalisées calculées chacune à partir des coefficients de détails d'un niveau de décomposition dyadique de la transformée DWT, ainsi que de l'énergie obtenue par les coefficients d'approximation. Ainsi, le nombre de paramètres dépend du niveau de décomposition dyadique N_d . Chaque vecteur est constitué de (N_d+1) paramètres DWE statiques concaténés avec leurs paramètres dynamiques Δ et $\Delta\Delta$. L'implémentation de l'extraction des paramètres du descripteur DWE est effectuée en utilisant la commande Matlab Wavedec. Alors que les paramètres dynamiques sont obtenus en utilisant l'outil Hcopy de librairie HTK.

III.3.1.2. Enregistrement des vecteurs de références

Cette étape consiste à concaténer les séquences des vecteurs de paramètres dans une matrice appelée Matrice de paramètres d'apprentissage, dont chaque ligne est un vecteur de paramètres. De plus, chaque indice de classe d'un vecteur de paramètres de cette matrice est enregistré dans un tableau (ligne ou colonne) appelé tableau de classes d'apprentissage.

III.3.2 Phase de test

Cette phase consiste à reconnaître la classe d'émotion de chaque signal et d'évaluer les performances de notre système RAE en termes du taux de classification. La reconnaissance de la classe d'émotion du signal se base sur une étape d'extraction de paramètres et une étape de classification de la séquence de vecteurs de paramètres représentant le signal de test.

L'étape d'extraction de paramètres durant cette phase s'effectue de la même manière que celle de l'apprentissage. Dans la section suivante, on décrira la classification de la séquence de paramètres représentant le signal de test considéré.

III.3.3 Classification des signaux de test en appliquant l'algorithme de classification KNN combiné avec la règle de vote

La classification d'une séquence de vecteurs de paramètres consiste premièrement à classer chaque vecteur de paramètre en utilisant le classificateur KNN, ensuite appliquer la stratégie de la règle de vote sur la séquence d'indices de classes pour obtenir la classe la plus votée représentant la classe d'émotion du signal de test. Plus particulièrement le classificateur KNN compare chaque vecteur de la séquence de vecteurs de paramètres de test avec tous les vecteurs de paramètres de la matrice de paramètres d'apprentissage et fait sortir ensuite l'indice de classe du vecteur ayant la distance minimale avec le vecteur de test.

III.3.4 Mesure de performances du système RAE proposé

Dans notre travail, on propose d'évaluer premièrement les performances du classificateur KNN en considérant le taux de classification des vecteurs de paramètres de tous les signaux de la base de données sans tenir en compte la classification des signaux :

$$TCV = \frac{Nv_C}{Nv_T} \cdot 100 \quad \text{III.1}$$

Avec :

- 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 TCV permet d'évaluer globalement l'efficacité du classificateur KNN sans tenir compte l'appartenance des vecteurs aux signaux.

Réellement, l'objectif est de classifier chaque signal pour identifier sa classe émotionnelle. Ainsi, on propose deuxièmement d'évaluer les performances du système RAE en considérant le taux de classification des signaux de test après l'application de la stratégie de la règle de vote :

$$TCS = \frac{Ns_C}{Ns_T} \cdot 100 \quad \text{III.2}$$

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 décrire les différentes expériences effectuées dans notre travail dans le but de répondre sur les différentes questions posées dans l'introduction de ce chapitre. Nous allons également discuter et interpréter les différents résultats obtenus pour chaque expérience. Dans les sections suivantes, nous allons décrire les expériences effectuées sur le premier système, ensuite, nous décrire celles du deuxième système.

III.4.1. Expériences effectuées sur le système basé sur l'extraction des paramètres MFCC

III.4.1.1. Configuration optimale du classificateur KNN pour les paramètres MFCC

Le but de cette expérience est de chercher la configuration optimale du classificateur KNN en choisissant le nombre de vecteurs les plus proches voisins K ainsi que la Distance optimale

en termes du taux de classification des signaux TCS. Le tableau III.3 présente les taux de classification TCS pour les différentes valeurs de k variant de 1 à 20 et pour les différentes distances (Euclidean, cityblock, cosine, correlation).

Tableau III.3 : Taux de classification TCS du premier système RAE pour différentes valeurs de k et différentes distances.

Distance \ K	Euclidean	Cityblock	Cosine	Correlation
1	76,36	75,97	75,58	75,58
2	76,36	75,97	75,58	75,58
3	76,36	76,74	74,42	73,64
4	76,36	74,03	74,42	75,58
5	76,74	72,87	72,87	72,87
6	75,97	72,48	74,03	72,87
7	74,81	71,31	72,48	72,48
8	73,64	72,09	70,54	72,48
9	73,64	71,71	71,32	70,93
10	71,32	70,15	70,54	71,70
11	72,87	70,54	69,77	71,70
12	71,31	69,77	70,93	70,54
13	72,48	69,38	70,54	70,54
14	70,15	67,44	70,15	70,93
15	68,99	67,83	69,77	69,77
16	68,99	67,83	68,60	70,15

17	68,99	68,60	68,60	68,99
18	68,60	67,44	68,60	69,38
19	68,60	67,83	68,21	68,99
20	67,83	67,83	66,67	68,60

Ces résultats sont illustrés sur la figure III.2 :

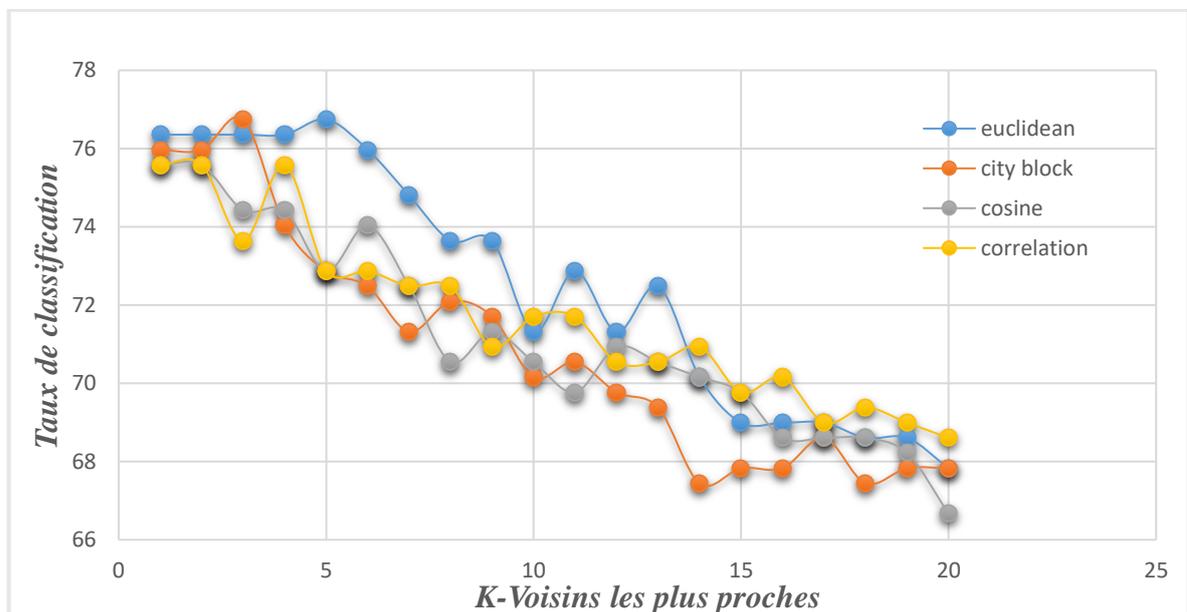


Figure III.2 : Taux de reconnaissance TCS du premier système RAE avec différentes valeurs de k et différentes distances du classificateur KNN.

A partir de la Figure III.2 et du tableau III.3 on remarque que :

- Le taux de classification TCS atteint les grandes valeurs pour les cinq premières valeurs de k pour toutes les distances. Ensuite, le taux commence à chuter au fur et à mesure que k augmente.
- Le taux de classification maximal de 76.74% est obtenu en choisissant les distances euclidienne et cityblock avec des valeurs de k égaux respectivement à 5 et 3.

III.4.1.2. Combinaison optimale des paramètres MFCC

Cette expérience a pour objectif de chercher la meilleure combinaison des types de paramètres du descripteur MFCC, en considérant les types de paramètres suivants :

- paramètres MFCC statiques,
- paramètre énergétique E,
- paramètre dynamique D (vitesse).
- Paramètre dynamique A (accélération).

Ainsi, les différentes combinaisons considérées sont données comme suit : MFCC avec 12 coefficients, MFCC_E avec 13 coefficients, MFCC_ED avec 26 coefficients, MFCC_D avec 24 coefficients, MFCC_EDA avec 39 coefficients. Le tableau III.4 donne le TCS du système de reconnaissance acoustique des émotions basé sur classificateur KNN combinée avec la règle de vote.

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

Types de descripteurs	EDA	MFCC	MFCC_E	MFCC_D	MFCC_ED	MFCC_EDA
Nombre de Coefficients	3	12	13	24	26	39
TCS	33.33	72	74.41	75.58	75.58	76,74

A partir de ce tableau, on peut donner les points suivants :

- L'utilisation de la totalité des paramètres (MFCC_EDA) présente le meilleur taux de classification de 76.74%
- Les paramètres énergétiques donnent un taux de classification TCS relativement faible de 33.33%.
- La combinaison MFCC_E représente le meilleur compromis entre le taux de classification et la dimension des vecteurs.

III.5.1. EXPERIENCES DU SYSTEME RAE BASE SUR L'EXTRACTION DES PARAMETRES DWE

III.5.1.1. Choix de la fonction mère optimale de la famille Daubechies DB

L'objectif de cette expérience est de chercher la fonction mère optimale de la famille d'ondelettes Daubechies (DB) en prenant le niveau maximal de décomposition N_{dm} obtenu par l'application de la fonction Matlab « **wmaxlev** ». Plus particulièrement, chaque vecteur de paramètres est constitué de : $(N_{dm}+1)$ paramètres DWE statiques, L'énergie E totale, leurs paramètres dynamiques de type D, leurs paramètres dynamiques de type A. Ainsi, chaque vecteur est constitué de $(N_{dm}+2)*3$ paramètres.

Le tableau III.5 présente le Taux de classification pour les différentes fonctions mère DBn (n varie de 1 à 10). Cette expérience est effectuée en prenant la distance Euclidienne avec k égal à 1.

Tableau III.5 : Taux de classification pour différentes fonctions mère DB

DBn	niveau de décomposition maximal	TCS (%)
DB1	8	56.58
DB2	7	54,65
DB3	6	53,87
DB4	6	54,65
DB5	5	54,65
DB6	5	54,65
DB7	5	53,10
DB8	5	52.71
DB9	4	50
DB10	4	50

La figure III.3 illustre le taux de classification en fonction de l'ordre n de la fonction mère DB

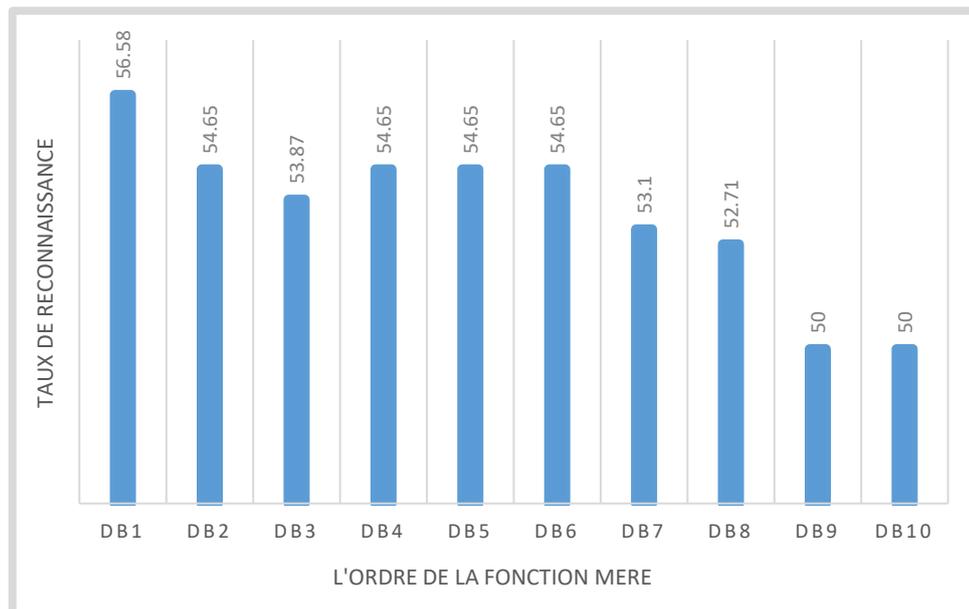


Figure III.3 : Taux de classification pour différentes fonctions mère DB.

D'après le tableau III.5 et la Figure III.3 on remarque que :

- la valeur TCS la plus élevée de 56.58% est obtenue avec l'ondelette mère Daubechies d'ordre 1 avec un niveau de décomposition égal à 8.
- la valeur TCS la plus faible de 50% est obtenue avec l'ondelette mère Daubechies d'ordres 9 et 10 avec un niveau de décomposition à 4.

III.5.1.2. Configuration optimale du classificateur KNN

L'objectif de cette expérience est de chercher la bonne configuration du classificateur KNN à savoir : la valeur de k vecteurs plus proche voisins ainsi que la distance optimale correspondante.

Il est à noter que cette expérience est effectuée en utilisant la configuration optimale de l'expérience précédente (DB1 avec niveau 8).

Le tableau III.6 présente les résultats obtenus en terme du taux de classification (TCS) pour les différentes de K (k=1 à 20), ainsi que la distance optimale correspondante.

Tableau III.6 : Configuration optimale du classificateur KNN (avec les paramètres DWE)

Distance \ K	Euclidean	Cityblock	Cosine	Correlation
1	56.58	56.20	51.55	52.33
2	56.58	56.20	51.55	52.33
3	54.26	54.65	51.55	50.00
4	51.55	54.26	50.77	50.78
5	52.71	53.48	50.00	50.39
6	52.32	52.71	50.77	50.78
7	52.32	51.94	50.00	50.39
8	52.32	50.78	50.00	49.22
9	52.32	51.16	50.00	48.06
10	51.16	50.78	50.00	49.61
11	49.61	51.94	51.16	50.78
12	50.78	50.78	49.22	50.39
13	50.39	52.32	48.84	50.00
14	49.22	51.94	48.84	50.00
15	49.22	51.94	49.61	49.61
16	50.39	51.16	48.84	48.45
17	49.61	51.55	47.67	48.84
18	50.00	50.78	48.84	49.61
19	50.39	51.55	48.84	48.45
20	50.00	50.78	48.45	48.84

La figure III.4 donne une vision globale de ces résultats.

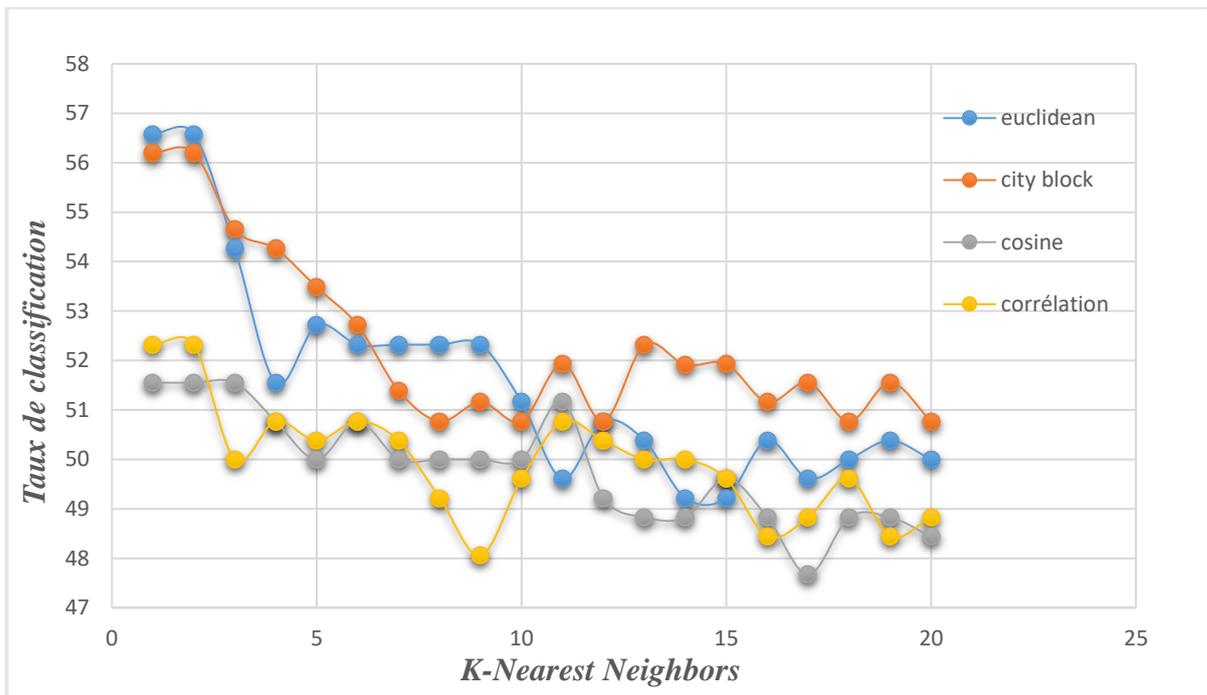


Figure III.4 : Taux de classification TCS pour les différentes valeurs de K ainsi pour les différentes distances pour l'ondelette (DB1_8).

D'après la figure III.4 et les résultats tableau III.6 on remarque que :

- Le taux de classification TCS atteint ses grandes valeurs pour les 5 premières valeurs de k. Ensuite, il se décroît pour k supérieur à 5.
- Le taux de classification TCS maximal de 56.58% est obtenu en choisissant la distance Euclidienne avec k=1.

III.5.1.3. Combinaison optimale des paramètres DWE

Cette expérience se déroule en suivant la même stratégie utilisée dans la section III.5.1.2. Pour choisir la meilleure combinaison des types de paramètres du descripteur DWE en choisissant la fonction DB1 avec le niveau de décomposition égal à 8. Les types de paramètres considérés sont comme suit :

- Paramètres DWE statiques,
- Paramètre énergétique E,
- Paramètre dynamique D (vitesse).

- Paramètre dynamique A (accélération).

Ainsi, les différentes combinaisons considérées sont données comme suit : DWE avec 9 coefficients, DWE_E avec 10 coefficients, DWE_ED avec 20 coefficients, DWE_D avec 18 coefficients, DWE_EDA avec 30 coefficients. Le tableau III.7 présente le TCS du système RAE basé sur classificateur KNN combinée avec la règle de vote.

Tableau III.7 : Taux de classification pour différentes combinaisons de paramètres DWE.

Types de descripteurs	EDA	DWE	DWE_E	DWE_D	DWE_ED	DWE_EDA
Nombre de Coefficients	3	9	10	18	20	30
TCS	33.33	50	50	53.10	52.32	56.58

A partir de ce tableau, on peut donner les points suivants :

- L'utilisation de la totalité des paramètres (DWE_EDA) donne le meilleur taux de classification de 56.58%
- Les paramètres énergétiques donnent un taux de classification TCS relativement faible de 33.33%.

III.6. Effet de l'application de la stratégie de règle de vote

Cette expérience a pour objectif de montrer l'efficacité de la stratégie de la règle de vote. Le tableau III.8 présente le taux de classification des vecteurs TCV ainsi que le taux de classification des signaux TCS pour les différents descripteurs (MFCC_EDA, DWE_EDA) en prenant les configurations optimales du classificateur KNN.

Tableau III.8 : Taux de classification TCV et TCS

Descripteurs	Nombre de paramètres	Taux TCV	Taux TCS
DWE_EDA	30	32.99	56.58
MFCC_EDA	39	38.76	76.74

A partir de ce tableau, on peut remarquer que l'application de la stratégie de règle de vote améliore les résultats de classification pour les deux types de descripteurs

III.7. Combinaison des descripteurs MFCC et DWE

Cette expérience consiste à extraire des vecteurs de paramètres en concaténant les paramètres du descripteur MFCC et ceux du descripteur DWE. La configuration par défaut du classificateur KNN est choisie. Cette expérience est effectuée en utilisant un PC portable I5 de 4 Gb de RAM.

Le tableau III.9 illustre le taux de classification TCS ainsi que le temps d'exécution de la phase de test avant et après la concaténation.

Tableau III.9 : Taux de classification TCS

Descripteurs	Nombre de paramètres	Taux TCS (%)	Temps d'exécution (S)
DWE_EDA	30	56.58	117.89
MFCC_EDA	39	76,36	131.34
MFCC_EDA & DWE_EDA	69	79.07	258.16

A partir de ce tableau, on peut remarquer les points suivants :

- La concaténation des deux descripteurs améliore le taux de classification TCS à 79.07% mais augmente considérablement le temps d'exécution.
- Le système basé sur l'extraction des paramètres MFCC donne le meilleur compromis entre le taux TCS et le temps d'exécution.

III.8. Conclusion

Dans ce chapitre nous avons décrit premièrement les deux systèmes RAE proposés qui se basent sur le classificateur KNN et la stratégie de la règle de vote. Plus particulièrement, le premier système se base sur le descripteur de paramètres MFCC, alors que le deuxième se base sur le descripteur DWE extrait à partir de là l'analyse par ondelettes. Plusieurs expériences ont été effectuées pour répondre sur des questions posées sur les bonnes configurations des deux systèmes proposées. A partir des résultats obtenus, nous concluons les points suivants :

- Le premier système donne des meilleures performances en choisissant le descripteur MFCC_EDA et le classificateur KNN avec l'utilisation de la distance Euclidienne et un nombre de vecteurs les plus proches voisins égal à 5.
- Le deuxième système donne son meilleur taux TCS en choisissant le descripteur DWE_EDA et le classificateur KNN avec l'utilisation de la distance Euclidienne et un nombre de vecteurs les plus proches voisins égal à 1. La meilleure fonction mère de la famille des ondelettes « Daubechies », conduisant aux bonnes performances est la fonction Db1.
- La combinaison du descripteur MFCC_EDA avec celui de DWE améliore légèrement le taux de classification, mais en revanche elle augmente la dimension des vecteurs et le temps d'exécution.

Conclusion générale

Conclusion générale

La reconnaissance acoustique des émotions (RAE) consiste à reconnaître l'état émotionnel d'une personne à partir du signal parole. Le fonctionnement d'un système RAE se base sur une phase d'apprentissage permettant de modéliser les différentes classes d'émotions et une phase de test permettant de classifier chaque signal de test et d'évaluer les performances du système RAE. Chacune des phases exige une étape d'extraction de paramètres permettant de convertir chaque signal en une séquence de vecteurs de paramètres.

Notre étude consiste à concevoir et d'implémenter deux systèmes RAE basé sur le classificateur KNN combinée à la stratégie de la règle de vote. Les deux systèmes se diffèrent dans l'étape d'extraction de paramètres. Plus particulièrement, le premier système se base sur l'extraction de paramètres MFCC, alors que le deuxième système se base sur l'extraction de paramètres appelés DWE calculés en utilisant l'analyse par ondelettes. Ainsi, les différents types de paramètres étudiés dans notre travail sont les coefficients MFCC, DWE, l'énergie et leurs paramètres dynamiques Δ et $\Delta\Delta$. Ces deux systèmes sont implémentés en utilisant le langage de programmation MATLAB ainsi que la boîte à outils HTK.

Dans le but d'évaluer les performances de ces systèmes, nous avons utilisé la base de données allemande de référence EMO-DB. Cette base considère 7 émotions (*'anger', 'disgust', 'fear', 'happiness', 'neutral', 'sadness' and 'surprise'*).

Plusieurs expériences sont effectuées pour répondre aux différentes questions posées sur la meilleure configuration de chaque système, à savoir : l'ordre optimal d'onde mère 'DB' et son niveau maximal de décomposition, la configuration optimale du classificateur KNN, l'effet de l'application de la stratégie de la règle de vote sur les performances des deux systèmes, la combinaison des paramètres MFCC avec les paramètres DWE, l'effet de l'ajout des paramètres dynamiques Δ et $\Delta\Delta$.

Les résultats obtenus montrent la pertinence des paramètres MFCC par rapport à ceux de DWE. Les résultats ont montré l'efficacité de la nouvelle architecture du système RAE basée sur le classificateur KNN combiné avec la stratégie de la règle de vote. Cette architecture fonctionne convenablement en utilisant le descripteur de paramètres MFCC dans l'étape d'extraction de paramètres. Ces résultats confirment également la pertinence des paramètres MFCC selon l'état de

L'art. De plus, la concaténation des deux descripteurs MFCC et DWE ont amélioré légèrement les résultats de reconnaissance.

Finalement, pour l'objectif d'améliorer les performances des systèmes RAE proposés, 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.
- L'utilisation d'autres descripteurs de paramètres.
- L'utilisation d'autres familles d'ondelettes.

- [1] F. Ghazali, A. Hacine-gharbi, P. Ravier et T. Mohamadi, «Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method », *turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, No.14, pp. 2980-2997, 2019.
- [2] H. Bengacemi, A. Hacine-gharbi, P. Ravier, Karim Abed-Meraim, O. Buttelli. « Surface EMG signal classification for Parkinson's disease using WCC descriptors and ANN classifier. » 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Feb 2021, Online Streaming, France. (hal-03348690).
- [3] R. Touahria, A. Hacine-gharbi et P. Ravier, « Discrete Wavelet based Features for PCG Signal Classification using Hidden Markov Models » 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Feb 2021, Online Streaming, Austria. pp.334-340.
- [4] P. Kleinginna et M. A. Kleinginna A, «categorized list of emotion definitions, with suggestions for a consensual definition» *Motivation and Emotion*, vol. 5, No.14, pp. 345-379, 1981.
- [5] Site web : <https://yestherapyhelps.com/the-8-types-of-emotions-classification-and-description-11658>.
- [6] M. Tamazirt et D. Naghraoui « Benchmark pour la reconnaissance automatique des émotions », Mémoire Master, Spécialité : automatique , École supérieure en sciences appliquées ,Tlemcen, 2020.
- [7] Site web : <https://amelioresetasante.com/quelles-sont-les-emotions-de-base/date:10/03/2022>.
- [8] K. Lekdioui, « Reconnaissance d'états émotionnels par analyse visuelle du visage et apprentissage machine » ,thèse de doctorat, Université Bourgogne Franche-Comté, Kénitra, 2019.
- [9] R. Jia, L. Gang et C. Yi-Ping Phoebe, « Acoustic feature selection for automatic emotion recognition from speech » *Information Processing And Man agement*, vol. 45, No 113, pp. 315–328, 2009.
- [10] Site web : <https://etrepants.com/les-emotions-secondaires-et-leur-manifestation-chez-les-enfants/date:10/03/2022>.

- [11] Y. attabi, « Reconnaissance automatique des émotions spontanées à partir du signal de parole », thèse de doctorat, école de technologie supérieure université du québec, 30 novembre 2015.
- [12] L. Kerkeni, « Analyse acoustique de la voix pour la détection des émotions du locuteur », Université de Sousse, Tunisie, 2020..
- [13] H. Hamdi, « Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques : Application à la simulation d'entretiens d'embauch », thèse de doctorat, Université d'Angers, 2012.
- [14] A. Ralph, « Recognizing Emotion from Facial Expressions: Psychological and Neurological Mechanisms » Behavioral and Cognitive Neuroscience Reviews, p. 21–62, 1 March 2002.
- [15] P. Gayathri et S. Matilda, « Emotion Recognition : A Survey » International Journal of Advanced Computer Research, Vol. 3, January 2015.
- [16] C. Guillaume, R. Cyril, B. Mireille et P. Thierry, « Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty » IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 41, No 16, 2011.
- [17] 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 - Computer and Information Sciences, Vol 33. No 9, pp. 1074–1083, 2019.
- [18] F. J. Harris, « On the use of windows for harmonic analysis with the discrete Fourier transform » Proceedings of the IEEE, Vol. 66, No 11, January 1978.
- [19] D. Namrata, « Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition » IJARET, Vol. 1, No. 6 , pp. 01 – 05, July 2013.
- [20] P. Yixiong, S. Peipei et S. Liping, « Speech Emotion Using Support Vector Machine » International Journal of Smart Home Vol. 6, No. 2, April, 2012
- [21] P. S. R. D. Suman K. Saksamudre, « A Review on Different Approaches for Speech Recognition » International Journal of Computer Applications, Vol. 115, No 122, April 2015.

- [22] G. Mayur R, D. Kinnal et B. Ninad, « Classification Techniques for Speech Recognition » International Journal of Emerging Technology and Advanced Engineering, Vol. 5, No 12, pp. 58-62, February 2015..
- [23] K. Leila, S. Youssef, M. Mohamed, R. Kosai et M. Mohamed Ali, « A Review on Speech Emotion Recognition: Case of Pedagogical Interaction in Classroom » International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) , 2017.
- [24] N. S et Altman, « An introduction to kernel and nearest-neighbor nonparametric regression » The American Statistician, Vol. 46, No 13, pp. 175-185, Aug,1992.
- [25] B. Schuller, R. Gerhard et L. Manfred, « Reconnaissance des émotions de la parole combinant des caractéristiques acoustiques et des informations linguistiques dans une architecture hybride de réseau de croyance-machine à vecteurs de support », 1. I - 577. 10.1109/ICASSP.2004.1326051. 2004.
- [26] J. Daniel, G. S. Richard, S. Stephen, S. Marilyn et W. D.Mitchell, « Acoustical properties of speech as indicators of depression and suicidal risk. » IEEE Transactions on Biomedical Engineering, Vol. 47, No.7,pp. 829–837.
- [27] J. H. Hansen et A. C. Douglas. « ICARUS : Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments » Speech Communication, Vol.16 No.4,pp 391–422. 1995.
- [28] J.Ma , H. Jin, L.Tianruo Yang et J. Tsai « Ubiquitous intelligence and computing » Third International Conference , UIC 2006, Wuhan, Chine, 3–6 septembre 2006.
- [29] M. El Ayadi, S. K. Mohamed et K. Fakhri. « Survey on speech emotion recognition: Features, classification schemes, and databases» Pattern Recognition, Vol 44 No 3, pp. 572–587, 2011.
- [30] S. Md. Kamruzzaman, A. Kazi Md. Rokibul et Arifuzzaman, « Emotion Recognition from Human Speech: Emphasizing on Relevant Feature Selection and Majority Voting Technique » International Conference on Informatics, Electronics & Vision (ICIEV) , Bangladesh, May 2014.
- [31] K. Leila, S. Youssef, M. Mohamed, K. Raoof et M. Mohamed Ali, « A Review on Speech Emotion Recognition: Case of Pedagogical Interaction in Classroom » International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 1-7, 2017.

- [32] S. Ying et Z. Xue-Ying, « Characteristics of human auditory model based on compensation of glottal features in speech emotion recognition » *Future Generation Computer Systems*, pp. 291-296, 2018.
- [33] H. Altun et G. Polat, « Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection » *Expert Systems with Applications*, Vol. 36, No 14, pp. 8197-8203, 2009.
- [34] J. Bhaskar, K. Sruthi et P. Nedungadi, « Hybrid approach for emotion classification of audio conversation based on text and speech mining » *Procedia Computer Science*, Vol. 46, pp. 635-643, 2015.
- [35] Z. Liu, M. Wu, W. Cao, J. Mao, J. Xu et G. Tan, « Speech emotion recognition based on feature selection and extreme learning machine decision tree », *Neurocomputing*, Vol. 273, pp. 271- 280, 17 January 2018.
- [36] W. Jiang, Z. Wang, J. Jin, X. Han et C. Li, « Speech emotion recognition with heterogeneous feature unification of deep neural network, » *Sensors (Basel, Switzerland)*, Vol. 19, No. 112, p. 2730, 18 June 2019.
- [37] J. Alonso, J. Cabrera, M. Medina et C. Travieso, « New approach in quantification of emotional intensity from the speech signal: Emotional temperature » *Expert Systems With Applications*, Vol. 42, No 124, pp. 9554-9564, 30 December 2015.
- [38] Z. Liu, Q. Xie, M. Wu, W. Cao, Y. Mei et J. Mao, « Speech emotion recognition based on an improved brain emotion learning model » *Neurocomputing*, Vol. 309, pp. 145-156, 2 October 2018.
- [39] L. Zhu, L. Chen, D. Zhao, J. Zhou et W. Zhang, « Emotion recognition from chinese speech for smart affective services using a combination of SVM and DBN » *Sensors*, Vol. 17, No 7, 24 July 2017.
- [40] A. Álvarez, B. Sierra, A. Arruti, J. Lópezgil et N. Garay-Vitoria, « Classifier subset selection for the stacked generalization method applied to emotion recognition in speech » *Sensors*, Vol. 16, No 121, 25 Décembre 2015.

- [41] A. Bhavan, P. Chauhan et R. Shah, «Bagged support vector machines for emotion recognition from speech » *Knowledge-Based Systems*, Vol. 184, 15 November 2019.
- [42] P. Shegokar et P. Sircar, « Continuous wavelet transform based speech emotion recognition » In *Proceedings of the 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, Australia, pp 1-8, December 2016.
- [43] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. Mahjoub et C. Cleder, « Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO » *Speech Communication*, Vol. 114, pp. 22-35, 2019.
- [44] Z. Kudakwashe et O. Oludayo, « Ensemble Learning of Hybrid Acoustic Features for Speech Emotion Recognition » *MDPI, Algorithms*, Vol 13, No 3, 22 March 2020.
- [45] 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 9, pp. 1074-1083, 2021.
- [46] F. Truchetet, « Ondelettes pour le signal numérique » *Hermes Science Publications*, Paris, 1998.
- [47] A. N. Akansu, W. A. Serdijn et I. W. Selesnick, « Emerging applications of wavelets: A review » *Physical Communication* , Vol. 3, No. 111, pp. 1-18, Mars 2010.
- [48] C. Catherine, « Some wavelet applications to signal and image processing », *Thèse de doctorat en mathématique*, Université de Liège 2003.
- [49] C. Gasquet et P. Witomski, « *Analyse de Fourier et Applications*», Dunod, collection Sciences sup, 1990.
- [50] H. Bengacemi, A. Hacine-gharbi, P. Ravier, A.-M. Karim et B. Olivier, « Surface EMG signal segmentation based on HMM modelling: Application on Parkinson's disease » *ENP Engineering Science Journal*, Vol. 1, No.1, pp. 63-74, July, 2021.
- [51] F. Ghazali, A. Hacine-gharbi et P. Ravier, « Statistical features extraction based on the discrete wavelet transform for electrical appliances identification » *Proceedings of the 1st International Conference on Intelligent Systems and Pattern Recognition, ISPR'20*, pp. 22-26. Virtual Event, Tunisia. October 16–18, 2020,

- [52] A. Hacine-gharbi et P. Ravier, « Wavelet Cepstral Coefficients for Electrical Appliances Identification using Hidden Markov Models » ICPRAM, Vol. 1, pp. 541-549, Portugal, 2018.
- [53] Y. Yujin, Z. Peihua et Z. Qun, « Research of speaker recognition based on combination of LPCC and MFCC, » In Intelligent Computing and Intelligent Systems (ICIS), Vol. 3, pp. 765-767, October 2010.
- [54] O'Shaghnessy, Douglas. Speech Communication: Human and Machine. Reading, MA: Addison-Wesley Publishing Company, 1987.
- [55] R. Charif et F. Mohamadi « Reconnaissance Acoustique des émotions basée sur les modèles de Markov cachés HMM » Mémoire Master, Spécialité: télécommunication, université de Bordj Bou Arreridj, 2018.
- [56] K. Belghitri et F. Didi, « Système sécurisé à base vocale » Mémoire master en Informatique, université de tlemcen, 2015.
- [57] S. Young, « The HTK book (HTK version 3.4) », Cambridge University Engineering Department, 2006.