

وزارة التعليم العالي و البحث العلمي

Université Badji Mokhtar –Annaba-

Badji Mokhtar –Annaba- University



جامعة باجي مختار- عنابة-

Année: 2013

Faculté des sciences de l'ingénieur
Département d'informatique

THÈSE

Pour obtenir le diplôme de
Docteur 3^{ème} cycle

La fouille de données audio

Filière : Informatique

Spécialité : Science et Technologie de l'Information et de la Communication

Préparée par

Faiz Maazouzi

Jury :

Président : Mohamed Tarek Khadir

Rapporteur : Halima Bahi-Abidet

Examineur : Labiba Souici-Meslati

Examineur : Mohamed Benmohamed

Examineur : Yamina Mohamed Benali - Hamel

Invité : Abdallah Benouareth

Prof U. ANNABA

Prof U. ANNABA

Prof U. ANNABA

Prof U. Mentouri Constantine

MCA U. ANNABA

Dr U. ANNABA

DEDICACE

Je dédie ce modeste travail :

A mes parents qui m'ont encouragé durant toutes mes études, faisant preuve de beaucoup de patience et de sacrifice. Je dédie ce mémoire en témoignage de ma reconnaissance, ainsi qu'à toute ma famille.

A ma chérie Imene

A mes frères et mes sœurs.

A tout mes amis

Enfin, je remercie toutes les personnes qui m'ont aidé de loin ou de proche dans mon cycle des études.

Remerciement

J'exprime d'abord mon profonds remerciements à Dieu qui ma donné la force, la volonté et le courage pour accomplir ce mémoire, pour que tous mes efforts soient fructueux et bénéfiques.

*Je voudrais remercier chaleureusement mon encadreur
Pr Halima Bahi pour sa disponibilité, sa gentillesse et pour ses suivies continus
durant la période de la réalisation de ce travail.*

*Je tiens à remercie ma familles pour ma avoir donné la possibilité de continuer
mes études et pour leurs soutient moral et leurs encouragement.*

*Je remercie tous ceux qui ont contribué de près ou de loin pour faciliter mon
travail.*



Abstract

Abstract

Online music diffusion and distribution is becoming increasingly important and commercial and personal databases are increasing in a considerable way. Nowadays, it's necessary to have tools that allow classifying and reaching these bases by carrying out analyzes through musical contents.

The work included in this thesis consists of knowing, in this time when it is a habit to say that the musical tastes of our young people are poor, whether the singers they prefer do not have really great vocal qualities. This work lies within the scope of audio mining including of the heterogeneous data. Work can be divided into two parts:

In the first part we treat the audio data to make a classification of the Algerian singers' voices. There exist two categories of the singing voices classification (type and quality), each category has several classes. Within this framework, we proposed three systems for the classification of the singing voices.

For the first system, we proposed to use a vector of 130 parameters with the GMM "Gaussian Mixture Models". In the second, we proposed an algorithm for the selection of the characteristics and the number of the Gaussians and in the third, we proposed to use a vector of characteristics which contains the descriptors of MPEG-7 + the descriptors Not-MPEG-7 and Standard T2 FGMMs "Type 2 Fuzzy Gaussian Mixture Models" for modeling and classification of singing voice. By using the results of the singing voices classification with the statistics of every singer, a new database has been created.

The second part of work consists of the data mining to answer our starting question "are our young people's musical tastes poor". In this part we proposed a new method of data excavation, based on the decision trees called: Multi Decision Tree (MDT). Finally, we used the MDT method to excavate our database.

Keywords: Data Base, Data Mining, Audio Mining, MDT, MPEG-7 Standard, Singing Voice Classification, GMMs, T2-FGMMs.



Résumé

Résumé

La diffusion et la distribution de la musique en ligne devenant de plus en plus importantes. Les bases de données commerciales et personnelles, augmentent de façon considérable. Il est devenu nécessaire de disposer d'outils permettant de classer et d'accéder à ces bases en effectuant des analyses par le contenu musical.

Le travail présenté dans notre thèse consiste à savoir, en ces temps où il est coutume de dire que les goûts de nos jeunes sont médiocres, si vraiment les chanteurs qu'ils préfèrent ne disposent pas de grandes qualités vocales, ce travail entre dans le cadre de fouille de données audio incluant des données hétérogènes. Le travail peut être divisé en deux parties.

Pour la première partie nous traitons les données audio pour faire une classification des voix des chanteurs algériens. Il existe deux catégories de classification des voix chantées (type et qualité), chaque catégorie dispose de plusieurs classes. Dans ce cadre nous avons proposé trois systèmes pour la classification des voix chantées.

Pour le première système nous avons proposé d'utiliser un vecteur de 130 paramètres avec les GMM « Gaussian Mixture Models », dans le deuxième système nous avons proposé un algorithme de sélection des caractéristiques et de nombre de gaussienne et dans le troisième système nous avons proposé d'utiliser un vecteur de caractéristiques qui contient les descripteurs de MPEG-7 + les descripteurs Non-MPEG-7 et T2 FGMMs « Type 2 Fuzzy Gaussian Mixture Models » pour la modélisation et la classification. En utilisant les résultats de la classification des voix chantées avec les statistiques sur chaque chanteur une nouvelle base de données a été créée.

La deuxième partie du travail consiste à fouiller les données pour répondre à notre question de départ « est ce que les goûts de nos jeunes sont médiocres ». Dans cette partie nous avons proposé une nouvelle méthode de fouille de données, basé sur les arbres de décision et appelée : Multi Decision Tree (MDT). Finalement, nous avons utilisé la méthode MDT pour fouiller notre base de données.

Mot clés : Fouille de données, Fouille de données audio, MDT, MPEG-7 Standard, Classification des voix chantées, GMMs, T2-FGMMs.



Liste des figures

Liste des figures :

Figure I.1: Schéma général du système.....	17
Figure II.1: Schéma général de la classification.....	21
Figure II.2: Schéma général d'un traitement acoustique.....	22
Figure II.3: Diagramme de séparation voix / musique.....	27
Figure II.4: Chevauchement avec une fenêtre triangulaire avec les différentes valeurs de M.....	29
Figure II.5: Différence entre le signal original (voix+ musique) et signal vocal (voix seulement).....	29
Figure II.6: Schéma explicatif de la méthode SVM. H est l'hyperplan séparateur. Les vecteurs supports sont grisés.....	32
Figure II.7: Neurone formel.....	33
Figure II.8: Structure d'un réseau de neurones.....	34
Figure II.9: Structure générale d'un system d'évaluation de la qualité vocale.....	35
Figure III.1: Schéma du système de modélisation et de classification basé sur le modèle de mélanges de gaussiennes.....	46
Figure III.2: Schéma de système de sélection des caractéristiques et du nombre de mélange de gaussiennes pour la classification des voix chantées.....	48
Figure III.3: Les résultats de la sélection des caractéristiques.....	56
Figure III.4 : (a) GPMF-UM « Gaussian primary membership function (MF) with uncertain mean » et (b) GPMF-Ustd « Gaussian primary membership function (MF) with uncertain std »	59

Figure IV.1: Illustration cartonne du processus de fouille de données.....	65
Figure IV.2: Les différents modèles de SVMs.....	69
Figure IV.3:Principe général de la combinaison de classifieurs.....	72
Figure IV.4: Schéma général de la classification basée MDT.....	75
Figure IV.5: Structure d'un MDT	76
Figure IV.6: Illustration du MDT - la base de données iris	77
Figure IV.7: Les variable (P_n , $PT_{n,n}$, PC_n) dans un MDT	80
Figure IV.8: Comparaison de la précision de C4.5, ADT, BDT, MDT et RFDT pour les bases de données ' iris, pima, glass, wine, ionosphere'	82
Figure IV.9: Comparaison des taux d'erreurs de C4.5, ADT, BDT, MDT et RFDT pour les bases de données ' iris, pima, glass, wine, ionosphere'	82
Figure V.1: Système proposé pour la fouille de données audio.....	85
Figure V.2: Les chanteurs et les albums les plus visités.....	87
Figure V.3: Résultats de séparation des voix chantées/ musique.....	88
Figure V.4 : Modèle obtenu après la modélisation avec MDT.....	91



Liste des tableaux

Liste des tableaux :

Tableau II.1: Récapitulatif des différentes études de la séparation voix chantées / musique.....	26
Tableau II.2: Fréquences en hertz des notes de la gamme tempérée.....	39
Tableau II.3 : Représente une comparaison des résultats des études de [Zwan et al., 2008] et [Prasert et al., 2008] pour l'évaluation de la qualité des voix chantées	41
Tableau II.4: Représente une comparaison des résultats d'étude de [Zwan et al., 2008] pour la classification en type des voix chantées.....	41
Tableau III.1: Les descripteurs non-MPEG-7.....	49
Tableau III.2 : Résultats de sélection ont utilisons le premier vecteur de caractéristiques pour l'évaluation de la qualité des voix chantées.....	51
Tableau III.3: Le nombre des échantillons utilise pour chaque catégorie.....	52
Tableau III.4 : Résultats de la sélection ont utilisons le deuxième vecteur de caractéristiques pour l'évaluation de la qualité des voix chantées.....	53
Tableau III.5 : Résultats de la sélection ont utilisons le deuxième vecteur de caractéristiques pour la classification des voix chantées en type.....	53
Tableau III.6 : Résultats de la sélection ont utilisons les deux vecteurs de caractéristiques pour la classification des voix chantées en type.....	53
Tableau III.7: Comparaison des résultats de les trois approches proposé et les études de [Prasert et al., 2008; Nakano et al., 2006 ;Zwan, 2008] pour la classification des qualités des voix chantées.....	60
Tableau III.8: Les nombres des échantillons utilisé pour chaque catégorie	60
Tableau III.9 : Comparaison des résultats des trois approches proposées et l'étude de [Zwan, 2008] pour la classification des voix chantées en type.....	61

Tableau IV.1: Informations sur les modèles construits à partir des bases (iris, Pima, glass, wine and ionospheres).....	78
Tableau IV.2: Exemple d'échantillons de base de données IRIS utilisé pour le teste	80
Tableau IV.3: Résultats de la classification.....	81
Tableau IV.4: Les ensembles de données utilisés, le nombre d'attributs, et d'apprentissage / test (tailles VC-10 représente 10 fois la validation croisée a été utilisée).....	81
Tableau V.1: Les informations de la base de données musicale	86
Tableau V.2: les différentes classes des voix chantées existantes en fonction de l'état de l'art....	87



Sommaire

SOMMAIRE

CHAPITRE I : INTRODUCTION GENERALE	15
I.1. Problématique de départ : Une enquête sociale?	16
I.2. Démarche générale.....	17
I.2.1. La classification des voix chantées dans la music commercial	17
I.2.1.1. La discrimination voix /musique	17
I.2.1.2. Classification des voix chantées	18
I.2.2. Collection des données	18
I.2.3. Application de technique de fouille de donnée.....	18
I.3. Organisation de la thèse	18
 CHAPITRE II : LA CLASSIFICATION DES VOIX CHANTEES DANS LA MUSIQUE COMMERCIALE	20
II.1. Introduction	21
II.2. Caractéristiques du signal vocal.....	22
II.2.1. L'analyse acoustique	22
II.2.2. Perception de la parole	22
II.2.2. Analyse et paramétrisation de la voix chantée	22
II.2.3.1. Le timbre	22
II.2.3.2. Coefficient d'harmonicité	23
II.2.3.3. Vibrato de la voix chantée	23
II.2.3.4. Les formants et le formant du chanteur	23
II.2.3.5. Autres paramètres	24
II.3 La séparation voix/musique	24
II.3.1. La méthode du masquage temps / fréquence	27
II.3.1.1. Etapes de l'algorithme de séparation	28
II.4 Les méthodes de classification et de modélisation de la voix chantée	30
II.4.1 Les Modèles de Markov Caches	30
II.4.2 Les Modèles de Mélanges de Gaussiennes	30
II.4.3 Les Machines à Vecteur de Support	31
II.4.4 Les k-Plus Proches Voisins	32
II.4.5 Les Réseaux de Neurones	32
II.4.6 La Quantification Vectorielle	34
II.5. L'évaluation de la qualité vocale	34
II.5.1. Le processus d'évaluation de la qualité vocale	34

II.5.2. Travaux antérieurs	35
II.6. La classification des voix chantées en type	36
II.6.1. Les méthodes de classification en type	37
II.6.1.1. Classification des voix selon la tessiture	37
II.6.1.1.1. Les types de voix selon la tessiture	38
II.6.1.2 Fréquence fondamentale	38
II.6.2. Les travaux de Zwan	40
II.6.3 Le travail de Prassert	40
II.6.4 Tableau récapitulatif	41
II.7. Conclusion	41
CHAPITRE III : APPROCHES PROPOSE POUR LA CLASSIFICATION DES VOIX CHANTEES	43
III.1.Introduction	44
III.2.Utilisation des Modèle de mélanges gaussiens pour la classification des voix chantées	44
III.2.1. L'extraction des caractéristiques.....	44
III.2.2. La classification (Modélisation et classification avec MMG « Modèle de Mélanges Gaussiens »)	45
III.2.2.1. Modélisation par GMMs.....	46
III.2.2.2. Critère de classification	47
III.3. La sélection des caractéristiques et de nombre de mélanges de gaussiennes pour la classification des voix chantées	48
III.3.1. Extraction des caractéristiques.....	48
III.3.1.1. MPEG-7 low-level descriptors	48
III.3.1.2 Non- MPEG-7 descripteurs	49
III.3.2. Modélisation et classification avec MMG « Modèle de Mélanges Gaussiens » .	50
III.3.3 La sélection des caractéristiques et de nombres de gaussien	50
III.4. Utilisation de T2 FGMMs (Type-2 Fuzzy Gaussian Mixture Models) pour la classification des voix chantées	54
III.4.1. Extraction des caractéristiques.....	54
III.4.1.1. MPEG-7 low-level descriptors	54
III.4.1.2 Non- MPEG-7 descripteurs	55
III.4.2. La sélection des caractéristiques	55
III.4.3. Modélisation et classification avec T2 FGMM « Type-2 Fuzzy Gaussian Mixture Models»	56
III.4.3.1. Modélisation avec T2 FGMMs	57
III.4.3.2. T2 FGMMs Classification	59
III.5. Expérimentations et résultats	59
III.6. Conclusion	61
CHAPITRE IV : LA FOUILLE DE DONNEES.....	63

IV.1.Introduction	64
IV.2. Le processus de fouille de données.....	64
IV.3. Les techniques de fouille de données	66
IV.3.1 Les algorithmes d'apprentissage.....	66
IV.3.1.1. Le Classifieur Bayésien	67
IV.3.1.1.1. Théorème de bayes	67
IV.3.1.1.2. Le classifieur bayésien naïf	68
IV.3.1.2. Machines à Vecteurs Supports	69
IV.3.1.3. Les arbres de décision	70
IV.3.1.3.1. La construction d'un arbre de décision	70
IV.3.1.3.2. Les algorithmes des arbres de décision	71
IV.4. Les ensembles de classifieurs.....	71
IV.4.1 Méthodes de Construction des ensembles de classifieurs	72
IV.4.2 Le Bagging.....	72
IV.4.3 Le Boosting.....	73
IV.4.4 L'apprentissage en cascade.....	73
IV.4.5 Les forêts aléatoires (Random Forest)	73
IV.5. La technique proposée (Multi Decision Tree ou multiple arbres de décision) ..	74
IV.5.1 Phase d'apprentissage	75
IV.5.2.1. Exemple illustratif d'apprentissage à l'aide de MDT	76
IV.5.2. La phase de décision.....	78
IV.5.2.1. Exemple illustrative de processus de décision de la méthode MDT.....	80
IV.6. Expérimentations et résultats.....	81
IV.7.Conclusion	83
CHAPITRE V : LA FOUILLE DE DONEES AUDIO.....	84
V.1.Introduction.....	85
V.2. L'établissement des données à collecter.....	86
Les données offertes par le site	86
V.3. La classification des voix chantées	87
V.3.1. La séparation de voix chantées / musique	88
V.3.2. Utilisation de T2 FGMMs pour la classification des voix chantées	89
V.3.2.1. L'extraction des caractéristiques	89
V.3.2.2. La modélisation et la classification avec T2FGMMs	89
V.3.3. Construction de la base de données	89
V.4. La fouille de données et résultats	89

Conclusion générale	94
Bibliographie	97



Chapitre I

Introduction générale

- 1. Introduction*
- 2. Contexte de l'étude*
- 3. Démarche générale*
- 4. Organisation de la thèse*

I.1. Problématique de départ : Une enquête sociale?

Les avancées technologiques et les prospérités économiques ont fait que la population mondiale se rencontre de plus en plus sur internet, et qu'elle en fait son espace de loisir de prédilection. Parmi les ressources qui sont très prisées sur internet, nous trouvons les chansons. En effet, la presque totalité des jeunes qui surfent sur internet écoutent de la musique. A cet effet, beaucoup de site au travers le monde offrent une grande palette de styles musicaux.

En Algérie (et presque dans tous les pays du monde), il est de coutume de dire que les jeunes n'ont pas la fibre artistique et que les goûts se dégradent. Ayant cette formidable ressource qui est le web, nous nous fixons comme objectif, de prospecter les goûts des auditeurs et fans des sites de musiques. Et ce pour confirmer ou infirmer ces allégations.

Pour cela, on considère un site de musique qui offre aux internautes un panel de chanteurs algériens. Les sites de ce type offrent de nombreuses statistiques sur les chanteurs les plus écoutés, les albums les plus prisés, ... A partir de ces données, nous tentons de répondre à notre question de départ, à savoir, note public a-t-il de bonnes ou de mauvaises préférences artistiques ?. Cette tâche relève de la fouille de données hétérogène, où l'aspect audio est omniprésent par le biais de la chanson.

I.2. Contexte de l'étude

Notre travail se situe au confluent de deux thématiques : la classification des voix des chanteurs et la fouille de données.

La classification des voix chantées est une discipline qui relève de la reconnaissance de formes et emprunte beaucoup des techniques de la reconnaissance de locuteur.

Dans cette tâche, il s'agit de catégoriser les chanteurs en se basant sur leur voix en des classes prédéfinies. Dans la littérature, il existe globalement, deux typologies. La première est bien établie et est basé sur le registre de la voix, tels que : soprano, ténor, ...etc. La seconde est plus subjective et concerne une évaluation de la voix en ce qu'elle est bonne ou moins bonne. Cette partie, nous est indispensable car c'est elle qui en plus des données fournies par les sites web qui permet de donner les caractéristiques des chanteurs préférés par le public.

D'autre part, nous aurons à appliquer une fouille sur les données obtenus. La fouille de données se présente comme un ensemble de techniques permettant à un utilisateur de découvrir des informations utiles et potentiellement inconnues à partir d'une masse de données [Vautier, 2008]. Plusieurs tentatives ont été proposées dans le but de généraliser toutes ces techniques. Le point commun de ces généralisations est que l'application de la fouille de données sur une masse de données consiste à enchaîner l'exécution d'opérateurs

dans le but d'aboutir à une généralisation appelée modèle des données que l'utilisateur interprète [Vautier, 2008]. À cet aspect quantitatif s'ajoutent des caractéristiques qualitatives et techniques : les données factuelles, textuelles et structurées ne sont plus les seules à être traitées ; viennent s'associer les images, les sons, etc.

I.3. Démarche générale

La conduite de notre projet induit un certain nombre d'étapes, qui se répartissent en deux volets assez importants :

Le premier volet inclut toutes les procédures que nous effectuons jusqu'à l'obtention d'un système de classification et d'évaluation qualitative de la voix.

Le deuxième volet consiste à regrouper les données issues du site web et du système de classification/évaluation et de les utiliser dans un contexte de fouille de données dont la réponse devrait nous permettre de connaître les caractéristiques de la voix préférée pour le public algérien (sur le web).

Cette démarche peut être détaillée en un certain nombre d'étapes que nous présentons dans ce qui suit :

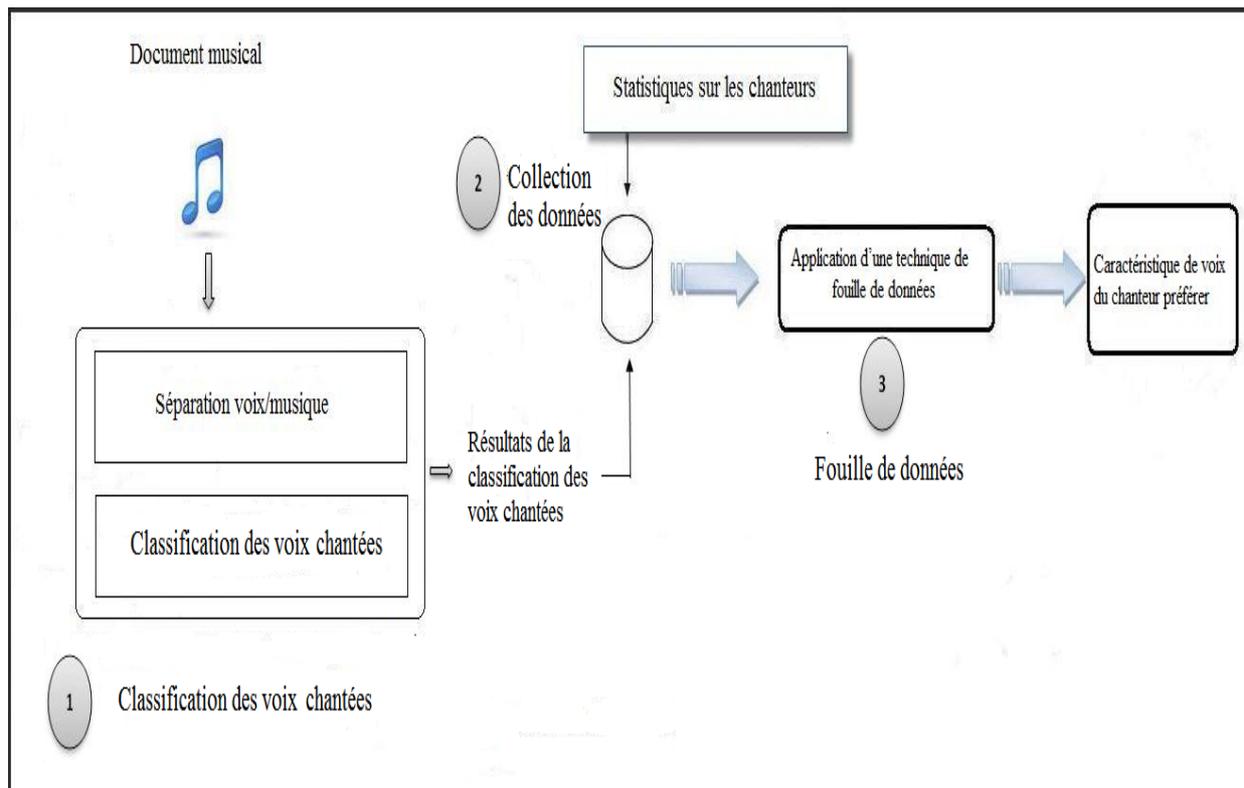


Figure I.1 : Schéma général du système

I.3.1. La classification des voix chantées dans la musique commerciale

Pour réaliser la catégorisation des voix des chanteurs, une première étape s'impose, il s'agit de séparer la voix du chanteur de la musique dans les enregistrements musicaux. Ensuite pourra intervenir l'étape de classification.

I.3.1.1 La discrimination voix /musique

Il existe de nombreux travaux qui se sont intéressés à cette problématique qui reste ouverte. Toutefois, ces dernières années certaines techniques ont donné des résultats intéressants. On retiendra la technique du masquage temps-fréquence pour réaliser le démixage. Dans cette technique, des coefficients de DFT (Discrete Fourier Transform) sont regroupés par pan adjacent, la différence de phase entre les canaux, la magnitude et l'ampleur de la variance avec un interactif en temps réel interface graphique. Les résultats produit par cette technique nous semble très satisfaisant au regard de notre application.

I.3.1.2 Classification des voix chantées

La classification des voix chantées représente une étape fondamentale pour notre système de fouille de données audio. Dans la littérature de classification des voix chantées, la plupart des chercheurs ont défini deux catégories de classification (Type et Qualité). Cette étape consiste à classer les voix des chanteurs en trois classes pour chaque catégorie.

Généralement, la classification des voix chantées se fait en trois étapes :

- 1) L'extraction du vecteur de caractéristiques,
- 2) la modélisation,
- 3) et la classification.

I.3.2. Collection des données

zikdalgerie.com est un site de musique algérien qui contient des milliers de chansons, ces chansons sont regroupées par albums où chaque album représente un chanteur. Ce site offre plusieurs statistiques ou données importantes dans notre étude.

I.3.3. Application de techniques de fouille de données

Une fois, nos données rassemblées et agencé, nous utilisons une technique de fouille de données très populaire que sont les arbres de décision. A cet effet, nous avons proposé un nouvel algorithme pour la construction des arbres de décision que nous avons appelé : MDT pour (Multi Décision Tree).

I.4. Organisation de la thèse

Ce mémoire est structuré autour de deux thèmes conducteurs : la caractérisation des voix chantées et la fouille de données. Ainsi, les chapitres 2 et 3 sont dédiés à la classification des voix chantées, et les chapitres 4 et 5 sont eux consacrés à la fouille de données.

Le chapitre 1 consiste en une introduction du travail principalement, au travers de sa problématique et son positionnement dans son contexte.

Dans le chapitre 2 nous introduisons d'abord le signal vocal et les différents paramètres qui peuvent le caractériser. On retiendra, en particulier, les paramètres issus de la norme MPEG-7 standard. Nous y décrivons aussi la méthode de masquage temps/fréquence, qui est la technique que nous avons utilisée pour la séparation entre la voix et la musique dans une chanson. Nous y présentons aussi les principales études qui ont traité de la classification des voix chantées.

Le chapitre 3 est dédié à la présentation de notre contribution dans cette thématique. Nous l'avons résumé en trois aspects. D'abord, nous présentons en premier, notre système de classification basé GMM (Gaussian Mixture Model). Ensuite, nous présentons nos expérimentations qui ont permis de dégager un ensemble de caractéristiques optimal. Finalement, nous présentons les détails de notre proposition qui est un système de classification basé sur les GMM flou (Fuzzy GMM). Ce chapitre comprend de nombreuses comparaisons qui situent nos systèmes par rapport à ceux de l'état de l'art.

Le chapitre 4 introduit le domaine de la fouille de données. On y présente les différentes techniques utilisées dans ce domaines. On retiendra, en particulier, les arbres de décision. Dans le contexte des arbres de décision, nous proposons un algorithme de construction qui offre une nouvelle méthode de Bagging. Cet algorithme, nous l'avons appelé : Multi Decision Tree (MDT).

Dans le chapitre 5, nous revenons sur notre projet de départ, en rassemblant les différentes composantes. A ce stade, nous construisons nos ensembles de données et procédons à une fouille de données en utilisant notre algorithme. Dans ce chapitre, on découvrira aussi la réponse à la question de départ.

Une conclusion et des perspectives clôturent cette thèse.



LA CLASSIFICATION DES VOIX CHANTEES DANS LA MUSIQUE COMMERCIALE

- 1. Introduction*
- 2. Caractéristiques du signal de la voix chantée*
- 3. La séparation voix/musique*
- 4. Les méthodes de classification et de modélisation de la voix chantée*
- 5. L'évaluation de la qualité vocale*
- 6. La classification en type*
- 7. Conclusion*

II.1. Introduction

Dans ce chapitre, il s'agit de faire le tour des travaux qui se sont intéressés à la classification des chanteurs selon leurs capacités vocales. Il faudrait rappeler que la première partie de ce travail est d'établir une caractérisation des chanteurs en se basant sur leur voix, pour pouvoir dire si le succès que rencontre un chanteur sur le web est dû (même en partie) à ses talents vocaux. La réalisation de cette tâche induit un certain nombre d'étapes. D'abord, il faut partir des fichiers audio disponibles sur le site, procéder à une séparation entre la voix et la musique. Ensuite, extraire les caractéristiques du signal vocal, et sur la base de cela, on tente de classer les chanteurs dans des catégories prédéfinies (Figure II.1).

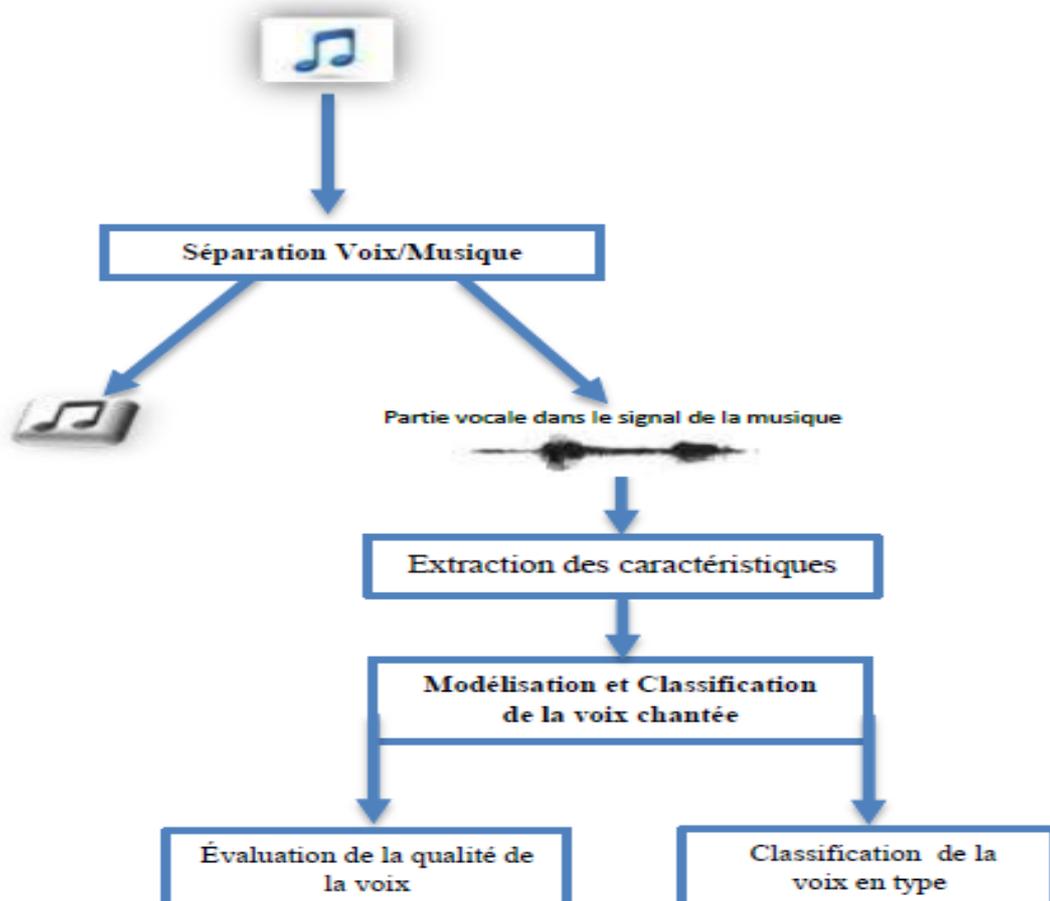


Figure II.1 : Schéma général de la classification

II.2. Caractéristiques du signal vocal

II.2.1 L'analyse acoustique

Le signal de parole, qui résulte de la conversion d'une onde acoustique en un signal électrique par un microphone, est un signal temporel unidimensionnel. Le système de paramétrisation utilise, en entrée, le signal de parole et retourne, en sortie, des vecteurs de paramètres à intervalle de temps régulier [Reynolds, 1994]. Ces vecteurs de paramètres sont calculés sur une fenêtre temporelle glissante dont la durée varie généralement entre 20 et 50 millisecondes. Il arrive souvent que les fenêtres temporelles utilisées pour extraire deux segments de signal consécutifs se recouvrent partiellement. La fréquence couramment utilisée pour l'extraction de ces paramètres est 100Hz. Ces caractéristiques ainsi que la dimension des vecteurs de paramètres acoustiques varient selon l'application et le type d'information extraite.



Figure II.2: Schéma général d'un traitement acoustique

II.2.2 Perception de la parole

Le signal de parole est un vecteur acoustique porteur d'informations d'une grande complexité, variabilité, redondance et influence de l'environnement. Les caractéristiques de ce signal sont appelées traits acoustiques. Chaque trait acoustique a une signification sur le plan perceptuel.

Le premier trait est **la fréquence fondamentale**, fréquence de vibration des cordes vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur (où les sons s'ordonnent du grave à l'aigu). Seuls les sons quasi-périodiques (voisés) engendrent une sensation de hauteur tonale bien définie.

Le deuxième trait est **le spectre fréquentiel** dont dépend principalement le timbre de la voix. Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix. Le timbre dépend de la corrélation entre la fréquence fondamentale et les harmoniques qui sont les multiples de cette fréquence.

Le dernier trait acoustique est **l'énergie** correspondant à l'intensité sonore. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés.

II.2.3 Analyse et paramétrisation de la voix chantée

II.2.3.1. Le timbre

Les paramètres les plus souvent utilisés sont ceux qui visent à décrire le timbre. Le timbre est en partie décrit par le spectre du signal. Les paramètres de timbre sont des vecteurs contenant

un nombre réduit de coefficients qui permettent de modéliser la forme de l'enveloppe spectrale. Les paramètres les plus souvent rencontrés dans la littérature sont:

Le vecteur de MFCC (Mel Frequency Cepstral Coefficients) : Les MFCC, calculés à partir de l'échelle des mels, peuvent avoir des variantes.

Les coefficients de prédiction linéaire (LPC) sont également utilisés pour décrire le timbre. À l'origine, ces coefficients sont utilisés pour localiser les formants dans la voix parlée. On trouve également des variantes de la LPC.

Il existe d'autres paramètres pour décrire le timbre : les Perceptual Linear Prediction coefficients (PLP), les warped LPC (WLPC), les Log Frequency Power Coefficient (LFPC).

II.2.3.2. Coefficient d'harmonicité

La voix chantée possède la particularité d'être fortement harmonique. Cette caractéristique, relatée à l'aide d'un coefficient d'harmonicité ou de voisement, a été utilisée à de nombreuses reprises. Chou définit le coefficient harmonique comme la moyenne harmonique du maximum de l'auto corrélation dans le domaine temporel et dans le domaine fréquentiel [Chou et Gu, 2001].

II.2.3.3. Vibrato de la voix chantée

Un ensemble de filtres est construit afin de détecter différents types de vibrato. D'autres recherches, dont les applications se trouvent dans la classification des sons, proposent des techniques d'extraction et d'estimation des paramètres du vibrato. Rossignol dans sa thèse [Rossignol et al., 1999] propose différentes méthodes pour détecter le vibrato ou extraire ses paramètres sur des sons monophoniques.

II.2.3.4. Les formants et le formant du chanteur

Les fréquences des formants sont déterminées par la forme du conduit bucco nasal. En général la localisation de trois formants permet d'identifier les voyelles de la voix parlée. Sundberg a montré dans [Sundberg et Rossing, 1990] que le spectre du chant lyrique est caractérisé par une importante énergie entre 2.5 et 3kHz. Ce pic du spectre, appelé formant du chanteur, permet à la voix de passer facilement au-dessus de l'accompagnement orchestral. Pour intégrer les formants et particulièrement le formant du chanteur aux paramètres acoustiques, les auteurs de [Lay et Wang, 2004] utilisent un ensemble de filtres en sous-bandes centrés sur les fréquences des formants.

Pour exploiter ces trois dernières caractéristiques, la sortie des filtres est analysée. Une approche similaire à celle utilisée pour le calcul des MFCC est employée pour obtenir un nombre réduit de coefficients : Une transformée en cosinus discret (DCT) est appliquée au logarithme de l'énergie de chaque sous bande du signal filtré puis les 13 premiers coefficients sont conservés.

II.2.3.5. Autres paramètres

Une autre façon plus générale pour déterminer les paramètres de la voix chantée est d'utiliser la description des signaux tels que les descripteurs de contenu audio comme la norme MPEG-7 standard. Bien que ces paramètres ne soient pas liés à la biomécanique de la voix chantante, ils peuvent être utiles dans le processus de reconnaissance. Le vecteur de caractéristiques de [Zwan et al., 2008] contient dans le groupe « d'autres paramètres » les paramètres suivants:

- ASE (Audio Spectrum Envelope) décrit le spectre de puissance à court terme de la forme d'onde. Les valeurs moyennes et les variances de chaque coefficient dans le temps. Ils sont désignés comme $ASE_1 \dots ASE_n$ et $ASE_{1var} \dots ASE_{nvar}$ respectivement.
- ASC (Audio Spectrum Centroid) décrit le centre de gravité du log fréquence spectre de puissance. La valeur moyenne et la variance sont désignées comme ASC et ASCvar respectivement.
- ASS (Audio Spectrum Spread) La valeur moyenne et la variance dans le temps, elles sont désignés comme l'ASS et ASS_{var} respectivement.
- SFM (Spectral Flatness Measure) calculée pour chaque bande de fréquence. Les valeurs moyennes et les écarts sont notés $SFM_1 \dots SFM_n$ et $SFM_{1var} \dots SFM_{nvar}$.
- Paramètres liés à des valeurs discrètes harmoniques: HSD (Harmonic Spectral Deviation), HSS (Harmonic Spectral Spread), HSV (Harmonic Spectral Variation).

Les paramètres présentés dans cette section sont inclus dans le groupe de paramètre de MPEG-7 descripteurs.

II.3. La séparation voix/musique

Les documents audio contiennent souvent plusieurs sources sonores (chanteurs, instruments, bruits) mélangées par enregistrement live ou en studio. La séparation de sources vise à reconstruire les signaux sources afin de les écouter individuellement. La séparation de sources dans un signal audio est un problème posé depuis longtemps et qui a fait l'objet de plusieurs travaux. Dans le domaine de la chanson, la séparation entre la voix du chanteur et de la musique est doublement difficile car les deux signaux sont souvent corrélés et harmonieux entre eux.

Il y a plusieurs techniques de séparation de voix/musique, ces méthodes reposent toutes sur des approches dites "aveugle" consistant à extraire du signal des descripteurs audio "génériques" utilisés ensuite pour apprendre les deux classes de segments "chanté" et "non chanté" à l'aide d'algorithmes statistiques.

On recense dans le tableau suivant pour chaque étude les descripteurs et les systèmes de classification qui ont fourni les meilleurs résultats. La base utilisée et la méthode de séparation sont également précisées. On note que les études concernant l'identification des chanteurs ne nécessitent pas une détection de la voix aussi complète que celles dont le but est la séparation automatique d'un morceau en deux parties : voix chantées et musique.

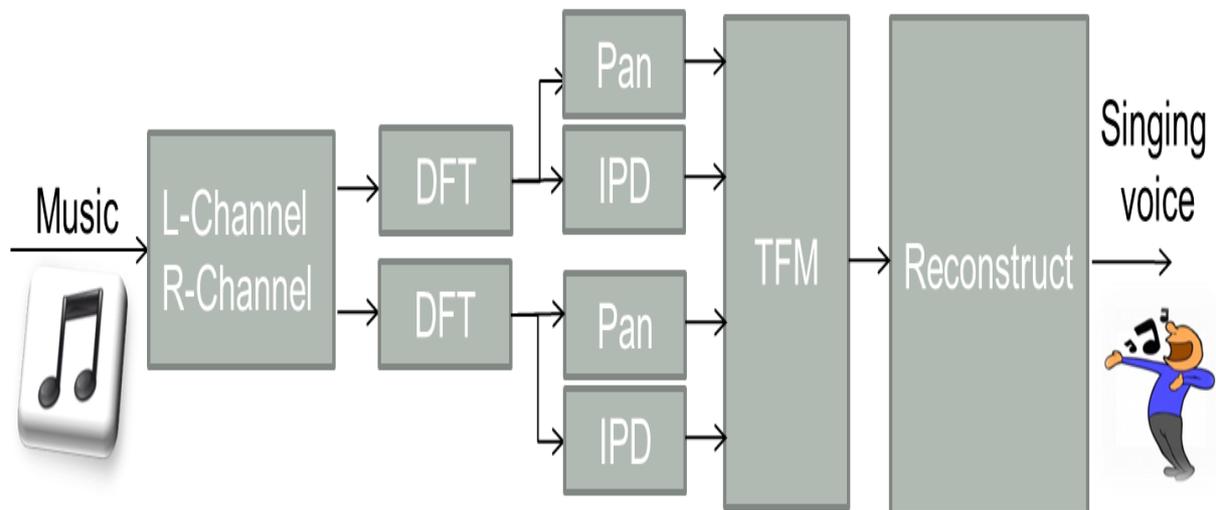
Étude	But de l'étude	Descripteurs	Classification	Base	Trames	Résultats
[Berenzweig et Ellis, 2001]	Détection par reconnaissance de phonèmes	PPF $\log(L_{mus})$ $\log(L_{voc})$ Entropie des PPF	HMM	245 segments de 15 sec	16ms	82.2%
[Chou et Gu, 2001]	Discrimination Parole musique amélioré par la détection de chant	MFCC	GMM Parole/ Musique Crossvalidation	26 minutes de parole, musique et chant		
[Kim et Whitman, 2002]	Identification du chanteur	Harmonicité	Seuil	20 chansons	100ms	PI : 79.3% I+V : 30.7% Moy : 54.9%
[Berenzweig et Ellis, 02]	Identification du chanteur	PLP	MLP (multi-layer Perceptron)	80 échantillons de 15sec	32ms	40% des trames vocales
[Zhang, 2003]	Identification du chanteur, instant de départ de la voix	Energie + ZRC + harmonicité + flux Spectral	GMM pour chaque chanteur	45 chansons (8 chanteurs)	16ms	
<i>suite page suivante</i>						

[New et Wang, 2004]	Détection automatique	LFPC	MM-HMM + Boostraping	20 chansons	1000ms	PI : 82% I+V : 86.6% Moy: 84.3%
[Li et Wang, 2006]	Extraction de la mélodie chantée	MFCC	GMM	10 chansons	86.65%	
[Fujihara et la., 2005]	Identification du chanteur	MFCC	64 mixtures vocal -GMM, 80 mixture non-vocal -GMM	242 chansons avec un seul chanteur, 22 avec deux chanteurs et 174 instrument ales		
[Ramona et Richard, 2008]	Détection automatique des segments vocaux	centroid + slope + width + asymetry + flux + MFCC + LPC + Zero Crossing rate + Estimation de F0 + algorithme IRMFSP pour trier les descripteurs	SVM + lissage par HMM	84 chansons	190ms	PI : 82% I+V : 84.4% Moy : 83.2%

Tableau II.1: Récapitulatif des différentes études de la séparation voix chantées / musique**II.3.1 La méthode du masquage temps / fréquence**

Les documents audio contiennent souvent plusieurs sources sonores (chanteurs, instruments, bruits) mélangées par enregistrement live ou en studio. La séparation de sources vise à reconstruire les signaux sources afin de les écouter individuellement.

Dans le cas des mélanges stéréo, l'étape de séparation peut s'effectuer par la méthode proposée par [Vinyes et al., 2006] « masquage temps-fréquence », qui consiste à utiliser une sélection des coefficients DFT. Les coefficients de DFT sont regroupés par Pan et l'IPD (Inter - Channel Phase-Difference).

**Figure II.3:** Diagramme de séparation voix / musique

Nous envisageons des chansons stéréo en entrée. Nous utilisons L variables de l'étiquette liée au canal gauche et L pour le canal droit R .

Original audio mixte (out_L, out_R) \longrightarrow Les signaux audio ($(S_i^R)[k], (S_i^L)[k]$)

Avec

Pan:

$$p_i = \arctan \left| \frac{DFT_p(S_i^R)[f]}{DFT_p(S_i^L)[f]} \right| .2\pi \quad (II.1)$$

IPD (Inter - Channel Phase-Difference):

$$|\text{Arg}(DFT_p(S_i^L)[f]) - \text{Arg}(DFT_p(S_i^R)[f])| = 0 \quad (\text{II.2})$$

II.3.1.1 Etapes de l'algorithme de séparation

1. La première étape consiste à découper le signal en deux canaux : out_L et out_R
2. Chaque cadre est fenêtré. Ensuite nous calculons la transformée de Fourier Discrète (DFT) des canaux gauche et droit, toutes les informations sont stockées dans les premiers $(N/2)+1$ coefficients. Nous ferons référence à leurs valeurs comme suit:

$$\begin{aligned} &DFT_0(out_L)[0] \cdots DFT_0(out_L)[N/2] \\ &DFT_0(out_R)[0] \cdots DFT_0(out_R)[N/2] \\ &\dots \\ &DFT_{P-1}(out_L)[0] \cdots DFT_{P-1}(out_L)[N/2] \\ &DFT_{P-1}(out_R)[0] \cdots DFT_{P-1}(out_R)[N/2] \end{aligned}$$

3. Ensuite, les cadres DFT de $(S_i^R)[k]$, $(S_i^L)[k]$ sont construits par maintien de certains coefficients de la DFT obtenus et fixant les autres à 0.

En d'autres termes, nous appliquons un masque binaire pour les coefficients de DFT. Soit $p \in 0 \dots P-1$,

$$DFT_p(S_i^L)[f] = \begin{cases} 0 \\ DFT_p(out^L[f]) \end{cases} \quad (\text{II.3})$$

$$DFT_p(S_i^R)[f] = \begin{cases} 0 \\ DFT_p(out^R[f]) \end{cases}$$

C'est l'étape où les différents paramètres peuvent être choisis pour générer des sons différents. Pour chaque cadre et chaque coefficient DFT, nous pouvons choisir de mettre à zéro ou à conserver sa valeur. En conséquence, une grande famille de solutions peut être générée: jusqu'à $2^{(N/2+1) \cdot P}$ différents sons.

4. La IDFT (Inverse Transformée de Fourier discrète) de ces cadres est effectuée après le remplissage de leurs coefficients de $(N/2)+2$ à $N-1$ avec des valeurs que la force de la symétrie hermitienne (le IDFT sortie doit être un signal réel). Ensuite, nous multiplions par l'inverse de la fenêtre de valeurs utilisées avant le calcul de la TFD.

5. Dans la dernière étape, nous ajoutons ces images pour obtenir $(S_i^R)[k]$, $(S_i^L)[k]$. Le recouvrement est effectué en utilisant une fenêtre triangulaire située autour du centre et complétée avec des zéros lorsque $M < N/2$. Deux exemples sont affichés dans les figures II.7 (a) et (b).

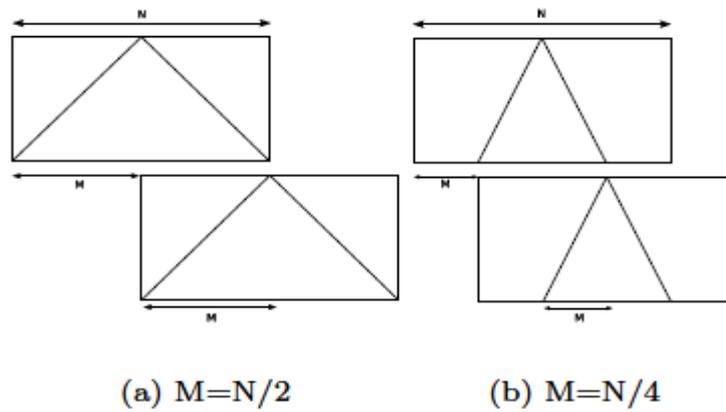


Figure II.4: Chevauchement avec une fenêtre triangulaire avec les différentes valeurs de M .

Dans la figure suivante, nous avons représenté le signal vocal obtenu après séparation voix/musique à partir du signal du haut.

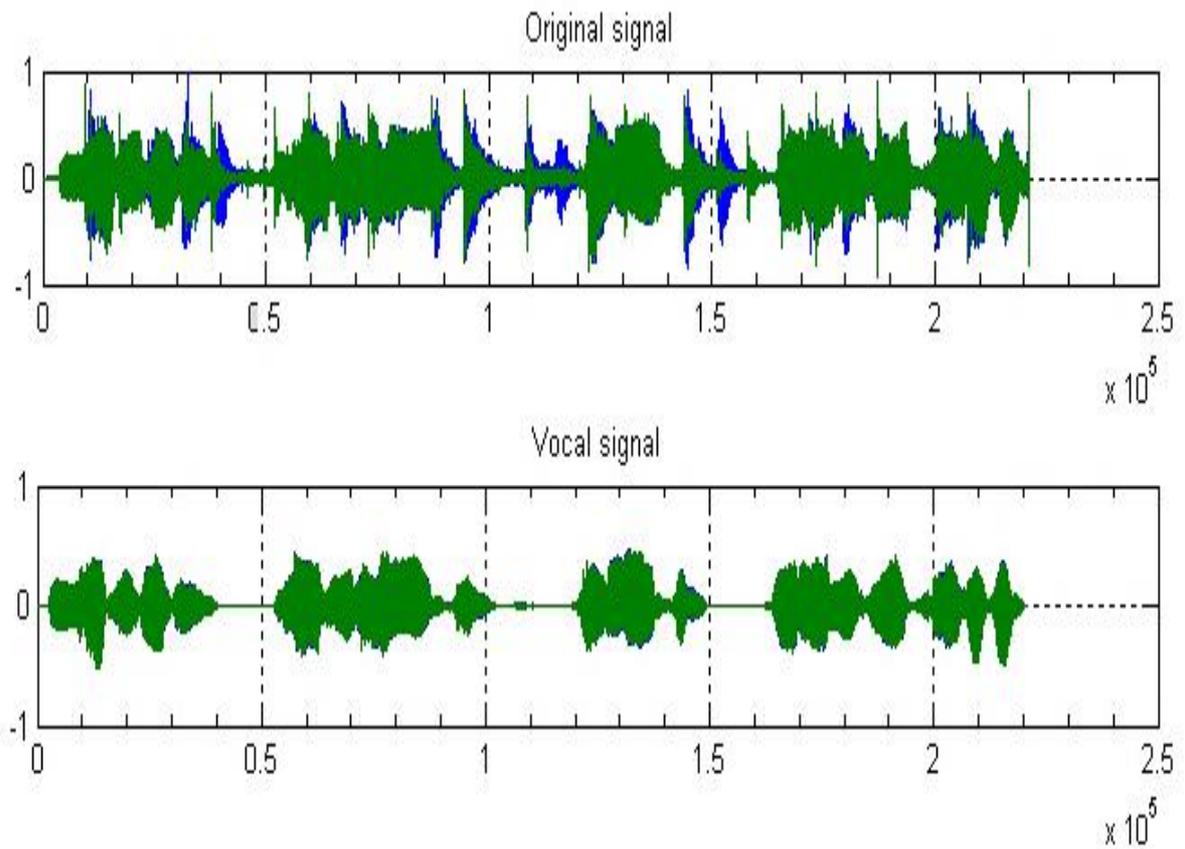


Figure II.5: Différence entre le signal original (voix+ musique) et signal vocal (voix seulement)

Cette technique est incluse dans le groupe des techniques pour obtenir des solutions au problème de séparation des sources ABS (Audio Blind Separation).

II.4 Les méthodes de classification et de modélisation de la voix chantée

Les études rapportées dans ce document relèvent principalement du domaine de la reconnaissance des formes. Les méthodes de classification utilisées sont des outils classiques dérivant des approches génératives (modélisation probabiliste essentiellement). Nous rappelons brièvement leurs principes.

II.4.1 Les Modèles de Markov Cachés

Les Modèles de Markov Cachés (Hidden Markov Models ou HMM en anglais) permettent de modéliser des enchaînements temporels au travers d'états, en tenant compte de leur durée. Cette méthode, développée dans les années 1965-1970 par Leonard E. Baum [Lee et Ellis, 2008] a été utilisée pour la reconnaissance de la parole ; elle est idéale pour modéliser chaque mot comme un enchaînement de sons de longueur variable. De la même façon, ils peuvent être utilisés pour modéliser n'importe quelle séquence temporelle qui respecte une certaine grammaire. En musique, ils sont utilisés par exemple pour modéliser les enchaînements d'accords [Mauch et al., 2007] ou de tonalités [Inoshita et Katto, 2009].

Les tonalités, tout comme les accords au sein d'une tonalité, s'enchaînent en respectant certaines règles. Pour construire les HMM, on peut ainsi, soit utiliser les nombreuses théories musicales, soit utiliser des corpora annotés manuellement. Dans ce cas, il faut estimer les probabilités de transitions entre états à l'aide des algorithmes classiques d'apprentissage des HMM, pour rendre compte de ces règles (pour les algorithmes d'apprentissage, voir [Rabiner 1989]).

À titre d'exemple, citons la reconnaissance d'accords : à partir d'une suite d'observations de type « chroma vector », l'étape de classification/reconnaissance se fait en cherchant, dans le HMM représentant le modèle, le chemin qui maximise la probabilité d'observation de cette suite.

II.4.2 Les Modèles de Mélanges de Gaussiennes

Les modèles de Mélanges de Gaussiennes (Gaussian Mixture Models ou GMM en anglais) sont des lois de probabilité uni- ou multidimensionnelles, très utilisées pour modéliser des répartitions inconnues ou susceptibles de présenter plusieurs modes dont le nombre est souvent inconnu. Ces mélanges sont couramment utilisés pour modéliser la voix d'un locuteur, ou la variabilité d'un son en parole. De fait, elles présentent l'avantage d'avoir été très étudiées ; leurs propriétés sont connues, les algorithmes d'estimation des paramètres sont approuvés.

Un modèle de mélange de Gaussiennes est la somme pondérée de N Gaussiennes uni-ou multidimensionnelles. Pour un nombre de Gaussiennes N fixé, la loi probabiliste de la distribution est définie de la manière suivante :

$$g(x, \mu_1, \Sigma_1, \mu_2, \Sigma_2, \dots, \mu_N, \Sigma_N) = \sum_{k=1}^N \pi_k f(x, \mu_k, \Sigma_k) \quad (\text{II.4})$$

Avec $f(x; \mu_k; \Sigma_k)$ la loi normale uni- (resp. multi-) dimensionnelle de moyenne (resp. vecteur de moyennes) μ_k et de variance (resp. matrice de covariance) Σ_k et π_k le poids de la $k^{\text{ième}}$ composante. Théoriquement, ces mélanges permettent d'approcher le nombre de distributions probabilistes, pourvu que le nombre de composantes soit suffisant.

Dans un problème à M classes, la distribution des paramètres pour chaque classe est modélisée par un GMM, le processus de décision se fait ensuite par la méthode du maximum de vraisemblance [Duda et al., 2001].

Les GMM sont utilisés beaucoup dans le domaine de traitement de signal, et particulièrement dans celui de la reconnaissance du locuteur et du chanteur.

II.4.3 Les Machines à Vecteur de Support

Les Machines à Vecteur de Support [Boser et al., 1992] (Support Vector Machine, ou SVM en anglais) sont des outils de classification discriminants développés pour les problèmes à deux classes. Dans le cas de données séparables linéairement dans un espace à N dimensions, les deux classes sont séparables par un hyperplan.

Ceci consiste à rechercher le meilleur hyperplan H, c'est-à-dire celui qui maximise sa distance aux frontières de chaque classe. Tout l'intérêt des SVM est que cet hyperplan optimal peut être caractérisé par les points de chaque classe qui en sont les plus proches.

Dans le cas où les données ne sont pas séparables dans l'espace de représentation, elles sont projetées dans un espace de dimension supérieure où elles sont séparables. Le problème est évidemment de trouver le bon espace image, au travers d'une fonction noyau adéquate qui correspond au produit scalaire dans ce nouvel espace.

Des méthodes ont été développées pour étendre cet outil aux problèmes à plusieurs classes :

-La méthode « un contre un » : N(N - 1) SVM bi-classe sont créés, pour apprendre toutes les séparations existant entre chaque couple de classes.

-La méthode « un contre tous » : N SVM bi-classe sont créés, pour apprendre les frontières de chacune des classes.

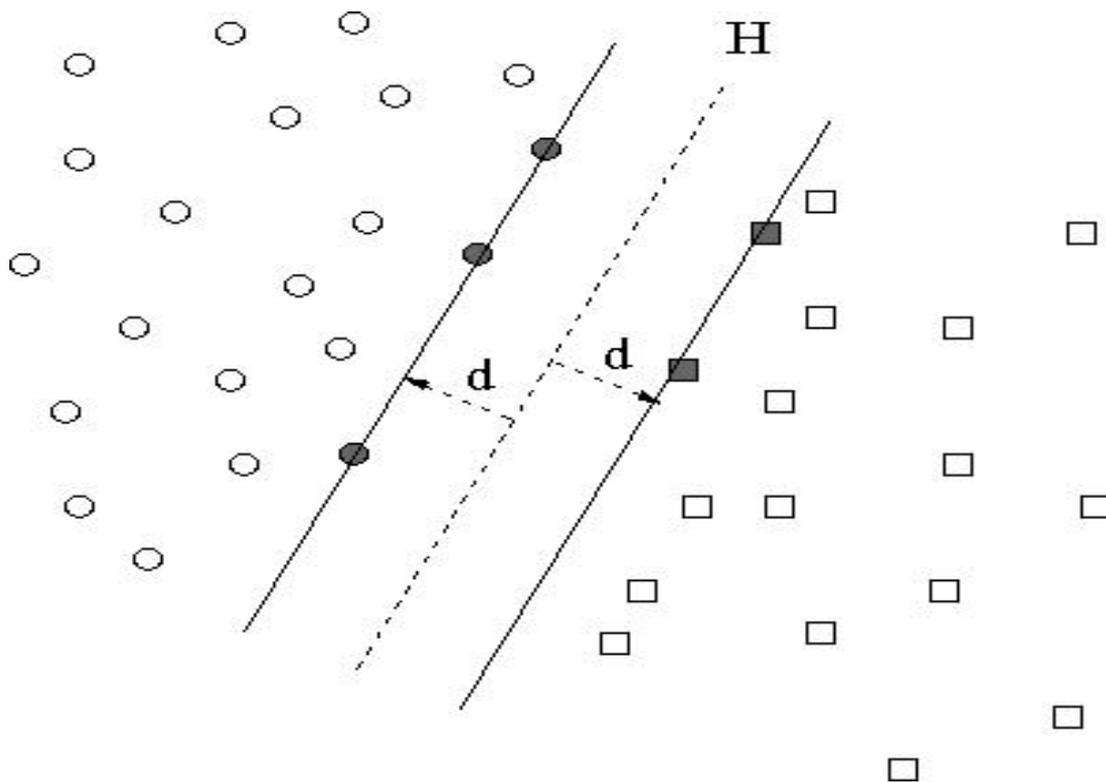


Figure II.6: Schéma explicatif de la méthode SVM. H est l'hyperplan séparateur. Les vecteurs supports sont grisés.

II.4.4 Les k-Plus Proches Voisins

Les k-Plus Proches Voisins (k-PPV, k-Nearest Neighbor ou k-NN en anglais) sont un algorithme de classification qui permet d'ignorer la distribution probabiliste des données. Il est basé sur une estimation locale de la densité de probabilité pour chaque classe.

A partir d'un ensemble d'apprentissage composé d'observations vectorielles, chacune étant associée à une classe, la décision se prend, pour chaque nouvelle observation, par vote majoritaire sur les classes des k observations de l'ensemble d'apprentissage les plus proches. La mise en œuvre de cet algorithme nécessite de régler la variable k.

Il est certain que l'avantage de cet algorithme est de prendre en compte un nombre quelconque de classes (contrairement à la technique SVM), et de n'avoir aucun à priori sur les lois probabilistes.

En revanche, le coût calculatoire est relativement élevé, même si des algorithmes ont été développés pour ne pas avoir à calculer la distance à tous les vecteurs de la base d'apprentissage, mais seulement à une partie d'entre eux, ce qui réduit considérablement le coût de calcul [Indyk. et Motwani, 1998].

II.4.5 Les Réseaux de Neurones

Les Réseaux de Neurones [**Kohonen, 1984**] (Artificial Neural Networks ou ANN en anglais) les plus employés en classification automatique sont les Perceptrons Multicouches (Multilayer Perceptron).

Le neurone formel (Figure II.7) est connu pour résoudre un problème de classification à deux classes. Le perceptron multicouche (Figure II.8) permet de réaliser une classification en k classes de données non séparables linéairement.

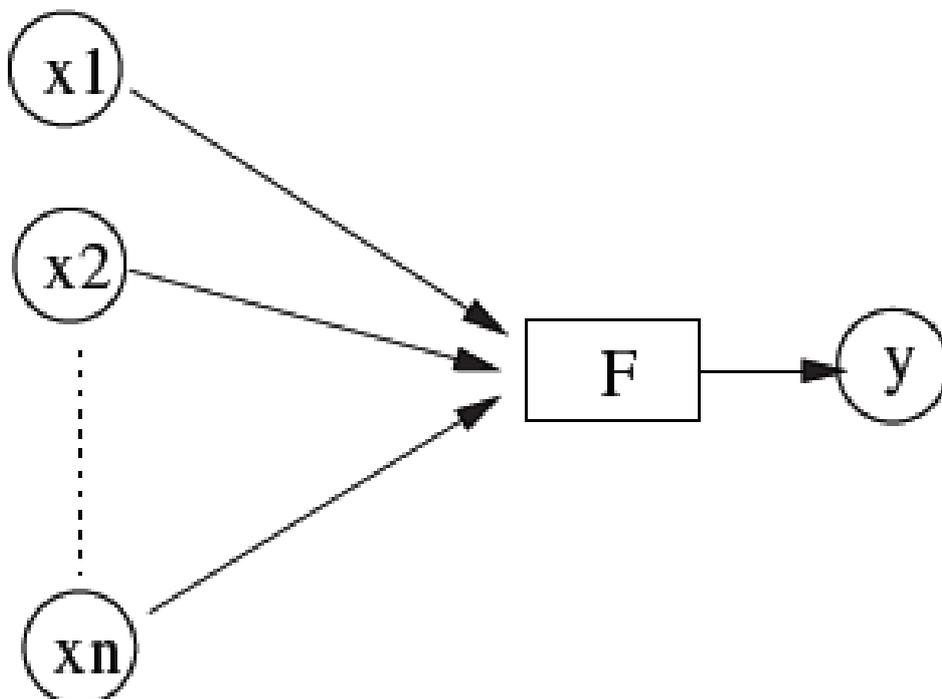


Figure II.7: Neurone formel.

Les sorties de chaque neurone de sortie peuvent être interprétées comme des valeurs prédictives, ou assimilées à des scores probabilistes. De ce fait, le perceptron est mis en œuvre pour fusionner des décisions issues de plusieurs classifieurs primaires.

Un autre intérêt du perceptron est de disposer d'algorithmes d'apprentissage efficaces (par exemple l'algorithme de retropropagation du gradient [**Krose et van der Smagt, 1996**]), mais un inconvénient réside dans la difficulté de définir sa topologie (nombre de couches, nombre de neurones par couche). [**Zwan et al., 2008**] ont utilisé les réseaux de neurone artificiel (RNA) dans leur système de décision pour classer les voix des chanteurs.

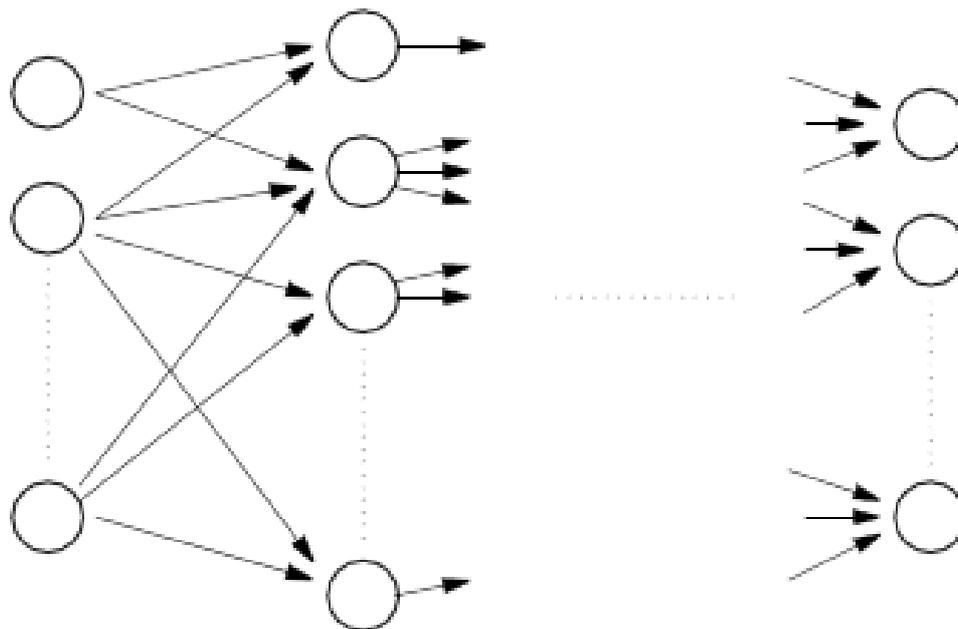


Figure II.8: Structure d'un réseau de neurones.

II.4.6 La Quantification Vectorielle

La quantification vectorielle (Vector Quantization en anglais) consiste à représenter tout vecteur acoustique x de dimension k par un autre vecteur y_i de même dimension mais appartenant à un ensemble E fini de L vecteurs. Les y_i sont appelés les vecteurs représentants ou codes vecteurs, E est appelé le dictionnaire ou catalogue des formes. La VQ détermine implicitement un partitionnement de l'espace acoustique en L régions appelées classes. Cette méthode élimine beaucoup d'informations et nécessite donc un temps de parole très long pour les phases d'apprentissage et de test.

II.5. L'évaluation de la qualité vocale

Dans cette étape, nous cherchons à évaluer les chanteurs en fonction de leur qualité vocale. C'est une évaluation subjective qui nécessite l'aide d'un spécialiste du domaine. Il est donc normal que l'on puisse y rencontrer plusieurs classes. A titre d'exemple, dans [Prasert et al., 2008], les auteurs utilisent trois termes pour évaluer les chanteurs: les chanteurs **amateurs**, les chanteurs **semi-professionnels** les chanteurs **professionnels**, tandis que [Prasert et al., 2008] utilisent deux termes : **bonne voix** et **mauvaise voix**.

II.5.1 Le processus d'évaluation de la qualité vocale

On distingue habituellement par l'évaluation de la qualité vocale un processus de décision consistant à utiliser les caractéristiques du signal de la voix pour déterminer la qualité de la

voix. Dans tous les travaux concernant l'évaluation de la qualité vocale des chanteurs, le schéma général pour la phase de reconnaissance est le suivant :

L'extraction des paramètres, la modélisation, l'apprentissage et enfin la classification.

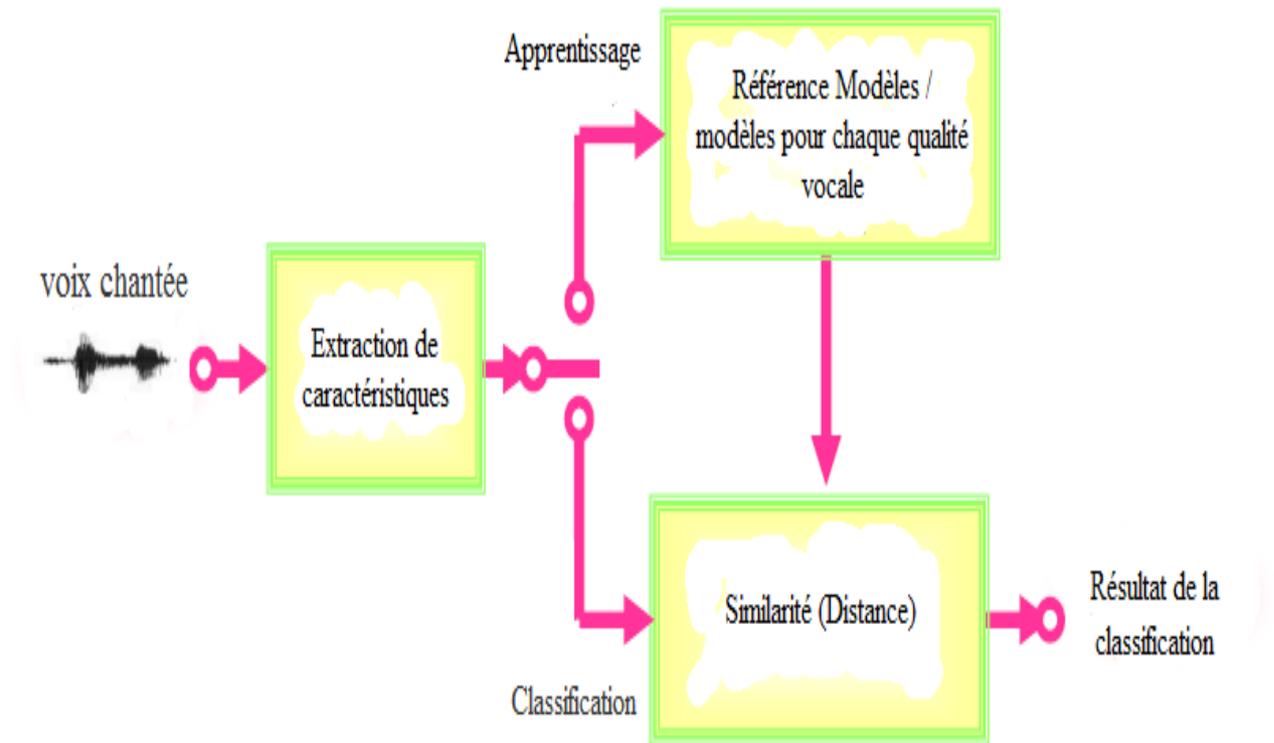


Figure II.9: Structure générale d'un système d'évaluation de la qualité vocale

II.5.2 Travaux antérieurs

Dans l'étude entreprise dans [Zwan et al., 2008] un vecteur de caractéristiques (FV) contenant 331 paramètres a été formé, les paramètres sont divisés dans les groupes suivants:

- Les paramètres de formants - 46 paramètres,
- Les paramètres de la glotte - 59 paramètres,
- Les paramètres de la hauteur (intonation et le vibrato) - 18 paramètres,
- D'autres paramètres (généraux) - 208 paramètres.

Le système de décision employé dans cette étude est : les réseaux de neurones artificiels (ANN) et les ensembles approximatifs (rough set-based (RS)), le système de décision a la capacité de classer automatiquement une voix chantée est discutée dans le cadre de la

comparaison de l'efficacité des systèmes ANN et RS dans la catégorie «la qualité de la voix» (classes: amateur, semi-professionnel, professionnel). En outre, [Zwan et al., 2008] utilisent des méthodes statistiques et approximatives pour juger les paramètres et appliquer différentes méthodes permettant de réduire la redondance dans le vecteur de caractéristiques. Finalement à partir des résultats obtenus dans ce travail on peut dire que l'utilisation des réseaux de neurones et les ensembles approximatifs a donné des bons résultats.

Les chercheurs dans [Prasert et al., 2008] ont utilisés une base de données qui se compose de voix de 10 vétérans chanteurs du Tokyo Institute of Technology. Tous les chanteurs ont chanté avec 2 styles de voix : «voix bonne» et «voix mauvaise».

Dans cette recherche, les chercheurs ont étudié deux types de vecteurs de caractéristiques, MFCC (Mel-Frequency Cepstre Coefficient) avec 11 dimensions et les vecteurs FBANK constitué des coefficients de log-mel filtres avec 12 dimensions.

Dans l'étape de construction du modèle et de décision, [Prasert et al., 2008] ont manuellement segmentés chaque phrase en phonèmes. Avec la méthode GMM, [Prasert et al., 2008] ont construit un modèles pour les deux voix celle qui chante bien et celle qui chante mal.

II.6. La classification des voix chantées en type

Pour chaque sexe, on distingue 3 catégories et de nombreuses sous-catégories. Les chanteurs et chanteuses sont classés en fonction du registre (étendue) de leur voix en plusieurs catégories.

- Les voix aiguës :

Soprano : - La plus aiguë des voix féminines. Le soprano voix haute et claire rappelle traditionnellement la féminité et la vertu. [Cross, 2008].

Sous-types de Soprano : Comme avec toutes les catégories de voix, des sopranos sont souvent divisés en différentes sous-catégories basées sur la tessiture, la couleur ou le timbre vocal, l'intensité de la voix. Ces sous-catégories incluent : Soprano de colorature, Soubrette, Soprano lyrique, Spinto, et Soprano dramatique. [Boldrey, 1994].

Ténor : - La plus aiguë des voix masculines. [Cross, 2008].

Sous-types de ténor : Il y a plusieurs sous-types des ténors comprenant Soprano ou soprano masculin, Haute-contre, et moderne.

- Les voix moyennes :

Mezzo-Soprano : - Également appelé mezzo, c'est la voix féminine de timbre moyen. [Cross, 2008].

Sous-types de Mezzo-soprano : Des Mezzo-sopranos sont souvent décomposés en deux catégories : Mezzo-soprano lyrique et Mezzo-soprano dramatique. [Boldrey, 1994].

Baryton : - Type de voix masculine dont la tessiture est moyennement grave. [Cross, 2008] .
Sous-types de baryton : Des barytons sont souvent divisés en différentes sous-catégories basées sur la tessiture, la couleur ou le timbre vocal, l'intensité de la voix.
Ces sous-catégories incluent : Baryton lyrique, Baryton de chant de bel (colorature), kavalierbariton, Baryton dramatique, Baryton de Verdi, baryton-noble, et Bariton/Baryton-Martin. [Boldrey, 1994] .

- **Les voix graves** :

Contralto : - La plus grave et la plus rare des voix féminines. [Richard, 2008].
Sous-types de contralto : Des contraltos sont souvent décomposés en deux catégories : Contralto lyrique et Contralto dramatique. [Boldrey, 1994].

Basse : - La plus grave des voix masculines, son étendue et son timbre rappellent le trombone ou le basson [Richard, 2008].
Sous-types bas : Des basses sont souvent divisées en différentes sous-catégories basées sur la tessiture, la couleur ou le timbre vocal, l'intensité de la voix. Ces sous-catégories incluent : Basso Profondo, Basso Buffo / Basse de chant de bel, Basso Cantante, Basse dramatique, et Basse-baryton. [Boldrey, 1994].

II.6.1 Les méthodes de classification selon le type

Dans la classification automatique des voix chantées par type on distingue deux approches de classification, la première approche est basée sur l'utilisation d'une classification statistique, tandis que, la deuxième approche (La méthode acoustique) c'est la plus simple et la plus utilisée dans ce genre de classification.

Il existe des travaux qui utilisent la première approche, parmi les projets qui se sont orientés dans ce contexte, Zwan dans [Zwan et al., 2008] utilisent également comme descripteur un vecteur de 331 paramètres avec les algorithmes de classification les réseaux de neurones artificiels (RNA) et rough set-based (RS).

La deuxième approche est une méthode simple et n'utilise pas un apprentissage, on appelle cette méthode la classification des voix selon la tessiture.

II.6.1.1. Classification des voix selon la tessiture

La tessiture

La tessiture représente la mesure de la hauteur de la voix chantée. Échelle de la voix qui indique la note la plus aiguë et la note la plus grave qu'une personne émet. C'est l'ensemble des sons émis sans efforts par une voix donnée.

Les voix peuvent être classées selon leur hauteur et leurs possibilités musicales à chanter des notes les plus graves aux plus aiguës.

Les voix féminines sont plus aiguës que les voix masculines à cause de la longueur des cordes vocales. On admet dans l'ensemble que les cordes vocales longues correspondent plutôt à des voix graves, et les cordes vocales courtes, aux voix aiguës. Les hommes ont des cordes vocales de 17 mm (voix de ténor) à 24 mm (voix de basse) et les femmes de 14 mm (voix de soprano) à 18 mm (voix de contralto).

II.6.1.1.1. Les types de voix selon la tessiture

D'après Cross dans [Cross, 2008] on classe ainsi les voix des plus aiguës aux plus graves :

- **Soprano** : - La tessiture habituelle de la soprano s'étend deux octaves au-dessus du do^3 avec parfois quelques notes aiguës supplémentaires.
- **Mezzo-Soprano** : - Sa tessiture s'étend le plus souvent du la^2 au la^4 .
- **Contralto** : - La voix de contralto s'étend le plus souvent du fa^2 au fa^4 .
- **Ténor** : - Sa voix s'étend le plus souvent du do^2 au do^4 .
- **Baryton** : - Type de voix masculine dont la tessiture est moyennement grave. L'étendue de sa tessiture se situe habituellement du la^1 au la^3 .
- **Basse** : - Sa tessiture s'étend habituellement du fa^1 au fa^3 .

II.6.1.1.2 Fréquence fondamentale

La classification basée sur de la fréquence fondamentale de la voix (F_0) de chanteur, chaque chanteur a une fréquence fondamentale ou en anglais le pitch (de 80 à 100 Hz chez l'homme, de 175 à 300 Hz pour la femme et de 200 à 600 Hz chez l'enfant). La fréquence fondamentale (F_0) est calculée par un estimateur d'auto corrélation [Quatieri, 2001].

La F_0 donne la hauteur de la voix du chanteur.

Chaque F_0 a une note, par exemple fréquence du $la^3 = 440$ Hz;

	Octave 0	Octave 1	Octave 2	Octave 3	Octave 4	Octave 5	Octave 6	Octave 7	Octave 8
do	32.70	65.41	130.8	261.6	523.3	1046.5	2093.0	4186.0	8372.0
do_# ré_b	34.65	69.30	138.6	277.2	554.4	1108.7	2217.5	4434.9	8869.8
ré	36.71	73.42	146.8	293.7	587.3	1174.7	2349.3	4698.6	9397.3
ré_# mi_b	38.89	77.78	155.6	311.1	622.3	1244.5	2489.0	4978.0	9956.1
mi	41.20	82.41	164.8	329.6	659.3	1318.5	2637.0	5274.0	10 548
fa	43.65	87.31	174.6	349.2	698.5	1396.9	2793.8	5587.7	11 175
fa_# sol_b	46.25	92.50	185.0	370.0	740.0	1480.0	2960.0	5919.9	11 840
sol	49.00	98.00	196.0	392.0	784.0	1568.0	3136.0	6271.9	12 544
sol_# la_b	51.91	103.8	207.7	415.3	830.6	1661.2	3322.4	6644.9	13 290
la	55.00	110.0	220.0	440.0	880.0	1760.0	3520.0	7040.0	14 080
la_# si_b	58.27	116.5	233.1	466.2	932.3	1864.7	3729.3	7458.6	14 917
si	61.74	123.5	246.9	493.9	987.8	1975.5	3951.1	7902.1	15 804

Tableau II.2: Fréquences en hertz des notes de la gamme tempérée

À partir de ces notes on peut donc classer les chanteurs selon la tessiture.

II.6.2. Les travaux de Zwan

Dans cette étude. Les systèmes formés avec des échantillons sonores, dont une grande partie 1700 échantillons ont été enregistrés à partir de 42 chanteurs dans un studio. Les chanteurs se composent de trois groupes: les amateurs (Gdańsk Université de chanteurs Chœur de la technologie), semi-professionnels (Gdańsk Académie de musique, étudiants de la Faculté Vocal), et les professionnels (Chanteurs qualifié, diplômé de la Faculté Vocal de l'Gdańsk Académie de Musique). Et le reste 1200 échantillons ont été extraits d'enregistrements professionnels sur CD. Pour chaque échantillon sonore, un vecteur de caractéristiques contenant 331 paramètres a été formé, les paramètres sont divisés dans les groupes suivants:

- Les paramètres de formants - 46 paramètres,
- Les paramètres de la glotte - 59 paramètres,
- Les paramètres de la hauteur (intonation et le vibrato) - 18 paramètres,
- D'autres paramètres (généraux) - 208 paramètres.

Les réseaux de neurones artificiels (ANN) et les ensembles approximatifs (rough set-based (RS)) ont été employés dans le but de déterminer la qualité de la voix chantée et le type de la voix chantée.

Dans cette étude le système de décision a la capacité de classer automatiquement une voix chantée est discutée dans le cadre de la comparaison de l'efficacité de la voix des systèmes ANN et RS dans deux catégories de reconnaissance: «Type de la voix » (classes: bas, baryton, ténor, alto, mezzo-soprano, soprano) et «Qualité de la voix» (classes: amateur, semi-professionnel, professionnel). En outre, [Zwan et al., 2008] utilisent des méthodes statistiques et approximatif pour juger les paramètres et appliquer différentes méthodes permettant de réduire la redondance dans le vecteur de caractéristiques, de nouveaux classificateurs ont été formés. Finalement à partir des résultats obtenus dans ce travail on peut dire que l'utilisation des réseaux de neurones et les ensembles approximatifs ont donné des bons résultats.

II.6.3 Le travail de Prassert

Dans cette étude la base de données se compose de voix de 10 chanteurs vétérans de TIT (Tokyo Institute of Technology). Tous les chanteurs ont chanté avec 2 styles de voix : «voix bonne» et «voix mauvaise».

Dans cette recherche, [Prasert et al., 2008] ont étudié deux types de vecteurs de caractéristiques, MFCC (Mel-Frequency Cepstre Coefficient) avec 11 dimensions et les vecteurs FBANK constitué des coefficients de log-mel filtres avec 12 dimensions.

Dans l'étape de construction du modèle et de décision, [Prasert et al., 2008] ont manuellement segmentés chaque phrase en phonèmes. Avec la méthode GMM, [Prasert et al., 2008] ont formé un modèles pour les deux voix qui chantent bien et qui chantent mal.

II.6.4 Tableau récapitulatif

Les deux tableaux suivants représentent une comparaison des différentes études existantes pour la classification des voix chantées (type et qualité).

Étude	Descripteurs	Classification	Base	rappel	précision
[Zwan et al., 2008]	Vecteur de 331 paramètres	Les réseaux de neurones artificiels (RNA)	1700 ECH 42 CH	93.6	93.4
[Nakano et al., 2006]	Pitch et Vibrato	Support Vector Machine (SVM)	600 ECH 12 CH	84.5	83.4
[Prasert et al., 2008]	MFCC	MMG	10 CH	93.7	97.8
[Prasert et al., 2008]	FBANK	GMM	10 CH	92.1	93.3

Tableau II.3 : Représente une comparaison des résultats des études de [Zwan et al., 2008] et [Prasert et al., 2008] pour l'évaluation de la qualité des voix chantées.

Étude	Descripteurs	Classification	Base	rappel	précision
[Zwan et al., 2008]	Vecteur de 331 paramètres	Les réseaux de neurones artificiels (RNA)	1700 ECH 42 CH	89.5	89.6
[Zwan et al., 2008]	Vecteur de 331 paramètres	rough set-based (RS)	1700 ECH 42 CH	64	67.6

Tableau II.4: Représente une comparaison des résultats d'étude de [Zwan et al., 2008] pour la classification en type des voix chantées.

II.7. Conclusion

Dans ce chapitre nous avons présenté la méthode de masquage temps/fréquence utilisée pour la séparation entre la voix et la musique dans les enregistrements. Nous avons également présentés les travaux de [Zwan, 2008] et [Prasert et al., 2008] qui nous ont inspirés, particulièrement par leur méthodologie de travail. Nous avons également présenté les GMMs

qui représentent l'outil le plus utilisé en reconnaissance de locuteur, et que nous avons utilisé et enrichi par la notion du flou pour la modélisation des voix chantées.

Le chapitre suivant sera destiné à replacer tous ces concepts et toutes ces techniques dans le contexte de l'application présente.

Approches proposées pour la classification des voix chantées

1. Introduction

2. Utilisation des Modèle de mélanges gaussiens pour la classification des voix chantées

3. La sélection des caractéristiques et le nombre de mélange de gaussiennes pour la classification des voix chantées

4. Utilisation de T2 FGMMs (Type-2 Fuzzy Gaussian Mixture Models) pour la classification des voix chantées

5. Expérimentations et résultats

6. Conclusion

III.1. Introduction

Lors de la mise en œuvre de notre système de caractérisation des chanteurs, nous avons adoptés progressivement trois choix conceptuels :

- 1- Utilisation des Modèle de mélanges gaussiens pour la classification des voix chantées [Maazouzi et Bahi , 2011].
- 2- La sélection des caractéristiques et le nombre de mélange gaussiennes pour la classification des voix chantées [Maazouzi et Bahi , 2012].
- 3- Utilisation de T2 FGMMs (Type-2 Fuzzy Gaussian Mixture Models) pour la classification des voix chantées [Maazouzi et Bahi, 2013].

Pour chaque système nous avons proposé différents vecteurs de caractéristiques.

III.2. Utilisation des Modèle de mélanges gaussiens pour la classification des voix chantées

Comme tout système de classification des voix chantées, le système proposé comprend deux étapes : l'extraction des caractéristiques et la classification.

III.2.1. L'extraction des caractéristiques

L'étape d'extraction des caractéristiques joue un rôle fondamental dans les systèmes de reconnaissance des formes, son rôle est d'assurer une représentation appropriée de l'information pour effectuer une classification optimale. Cependant, le problème de la représentation optimale du signal est maintenant loin d'être résolu.

L'objectif de l'extraction de caractéristiques en reconnaissance consiste à exprimer la forme primitive numérique connue ou codage symbolique. La difficulté ici est de trouver de bonnes caractéristiques. Certaines «bonnes» caractéristiques permettent de reconnaître facilement les différentes classes d'objets, nous disons alors qu'ils font de la discrimination.

Dans le domaine du traitement automatique de musique, il peut arriver que les mêmes méthodes de codage (en particulier, Mel-Frequency Cepstral Coefficients) sont utilisées à la fois pour la reconnaissance vocale, identification du locuteur ou de la reconnaissance chanteur.

La paramétrisation du premier vecteur de caractéristique (VCi) est fondée sur les coefficients cepstraux, le signal audio est échantillonné à 16 kHz et les coefficients MFCC (Mel-Frequency Cepstral Coefficients) sont calculés à partir d'un banc de 24 filtres Mel appliqué toutes les 10 ms sur des fenêtres de 30ms.

Une autre façon plus générale pour déterminer les paramètres de la voix chantée est d'utiliser la description des signaux tels que les descripteurs de contenu audio de la norme MPEG-7 standard. Le deuxième vecteur contient un certain nombre de descripteurs norme MPEG-7 et d'autres descripteurs. Les descripteurs sont divisés dans les groupes suivants:

- Energie: Audio Power (AP).
- Harmonique: Audio Fundamental Frequency (AFF)
- Spectral: Audio Spectrum Spread (ASS), Mel-Frequency Cepstral Coefficients (MFCC).
- Temporal: Log-Attaque Time (LAT), Temporal Centroid, (TP).
- Divers: Audio Spectrum Flatness (ASF).

A partir de ces descripteurs, un vecteur de caractéristiques (VCj) contenant 130 paramètres à été formé, les paramètres sont divisés dans les groupes suivants:

- Paramètre d'énergie: 4 paramètres.
- Paramètre d'harmonique: 4 paramètres.
- Paramètre spectral: 54 paramètres.
- Paramètre temporel: 36 paramètres.
- Divers paramètre: 32 paramètres.

III.2.2. La classification (Modélisation et classification avec MMG « Modèle de Mélanges Gaussiens »)

La classification des chanteurs est réalisée avec un système à base de MMG, Cette classification a comme but de classer les chanteurs sur la base de leur type de voix et leur qualité vocale.

Dans la classification en type nous avons utilisé 24 mélanges de distributions de Gauss pour modéliser les 6 classes, 12 pour les voix des chanteurs masculin et 12 pour les voix des chanteurs féminin [Maazouzi et Bahi , 2011].

Les types de voix masculin « Ténor, Brayton, basse ».

Les types de voix féminin « Soprano, Mezzo-soprano, Alto».

Dans L'évaluation de la qualité vocale nous avons utilisé 12 mélanges de distributions de Gauss pour modéliser les 6 classes, 6 pour les voix des chanteurs masculin et 6 pour les voix des chanteurs féminin [Maazouzi et Bahi , 2011].

Les trois qualités des voix masculin et féminin sont « Good », « Medium », « Bad ».

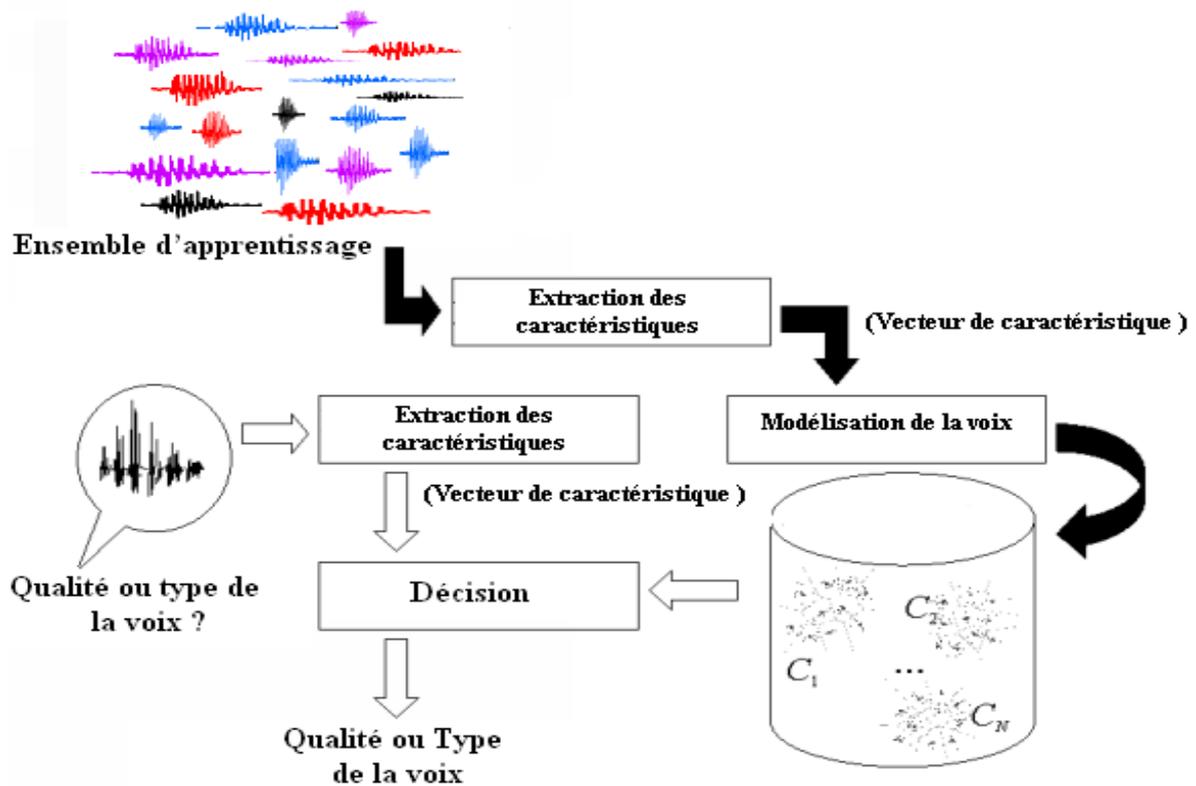


Figure III.1: Schéma du système de modélisation et classification basé sur le modèle de mélange de gaussiennes [Maazouzi et Bahi , 2011].

III.2.2.1. Modélisation par GMMs

Pour chaque classe un modèle de mélange de gaussiennes (GMM) est utilisé pour modéliser la distribution des données dans un descripteur de l'espace à D dimensions. Cet espace est obtenu par combinaison linéaire de plusieurs fonctions de densité de probabilité (PDF) aux dimensions D peut approcher la forme globale de la collection de descripteurs. La probabilité d'observer le vecteur λ Descripteur = $\{\mu, \Sigma\}$ étant donné le modèle multi-gaussien défini par ses paramètres x est alors :

$$p(x|\lambda) = \sum_{i=1}^M p_i g_i(x) \quad (\text{III.1})$$

Où

- x est un vecteur de descripteur à D dimensions,
- p_i les poids associés aux différentes gaussiennes et
- $g_i(x)$ une densité de probabilité définie par :

$$g_i(x) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} e^{-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)} \quad (\text{III.2})$$

Où μ représente la moyenne estimée et Σ la matrice de covariance. Les paramètres des densités de probabilité des gaussiennes sont estimés par l'algorithme EM (Expectation-

Maximization), dont la description peut être trouvée dans [McAulay et Quatieri, 1986]. EM est un algorithme itératif qui converge sur des paramètres qui sont localement optimale en fonction de la fonction de log-probabilité. Ces paramètres sont représentés collectivement par la notation: $\lambda = \{p_i, \mu_i, \Sigma_i\}$.

Pour l'évaluation de la qualité des voix chantées:

Le GMM de la classe "Good" (λ_g) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme bonne voix

Le GMM de la classe "Medium" (λ_m) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme moyen voix.

Le GMM de la classe "Bad" (λ_b) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme mauvais voix.

Pour la classification en type :

- Les types des voix male :

Le GMM de la classe " Ténor " (λ_{te}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme type de voix est Ténor

Le GMM de la classe " Baryton " (λ_{br}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme type de voix est Brayton

Le GMM de la classe " basse " (λ_{ba}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme type de voix est basse

- Les types de voix féminine :

Le GMM de la classe " Soprano " (λ_{so}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme type de voix est Soprano

Le GMM de la classe " Mezzo-soprano " (λ_{ms}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme moyen type de voix est Mezzo-soprano

Le GMM de la classe " Alto " (λ_{al}) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme type de voix est Alto.

III.2.2.2. Critère de classification

Le critère de classification retenu pour la classification est le critère de maximum a posteriori. Suivant ce critère, tout vecteur d'observation x est affecté à la classe qui maximise la probabilité d'observation sachant le modèle λ associé.

- Pour l'évaluation de la qualité de la voix:

$$\text{Classe}(x) = \{i \mid p(x \mid \lambda_i)\} = \{p \max(x \mid \lambda_g), p(x \mid \lambda_m), p(x \mid \lambda_b)\}$$

- Pour la classification de type pour les chanteurs:

$$\text{Classe}(x) = \{i \mid p(x \mid \lambda_i)\} = \{p \max(x \mid \lambda_{te}), p(x \mid \lambda_{br}), p(x \mid \lambda_{ba})\}$$

- Pour le classement de type pour les chanteuses:

$$\text{Classe}(x) = \{i \mid p(x \mid \lambda_i)\} = \{p \max(x \mid \lambda_{so}), p(x \mid \lambda_{ms}), p(x \mid \lambda_{al})\}$$

III.3. La sélection des caractéristiques et de nombre de mélange de gaussiennes pour la classification des voix chantées

Comme nous l'avons déjà souligné, l'ensemble de caractéristiques choisies pour modéliser une forme est très important pour les performances du classifieur. A cet effet, nous avons réalisé une étude pour sélectionner l'ensemble optimal pour cette modélisation.

Nous allons décrire l'ensemble des descripteurs de départ et la démarche poursuivie pour aboutir à cet ensemble. Nous avons, en outre, effectué quelques expérimentations pour déterminer le nombre de distribution Gaussienne à mélanger pour cette modélisation.

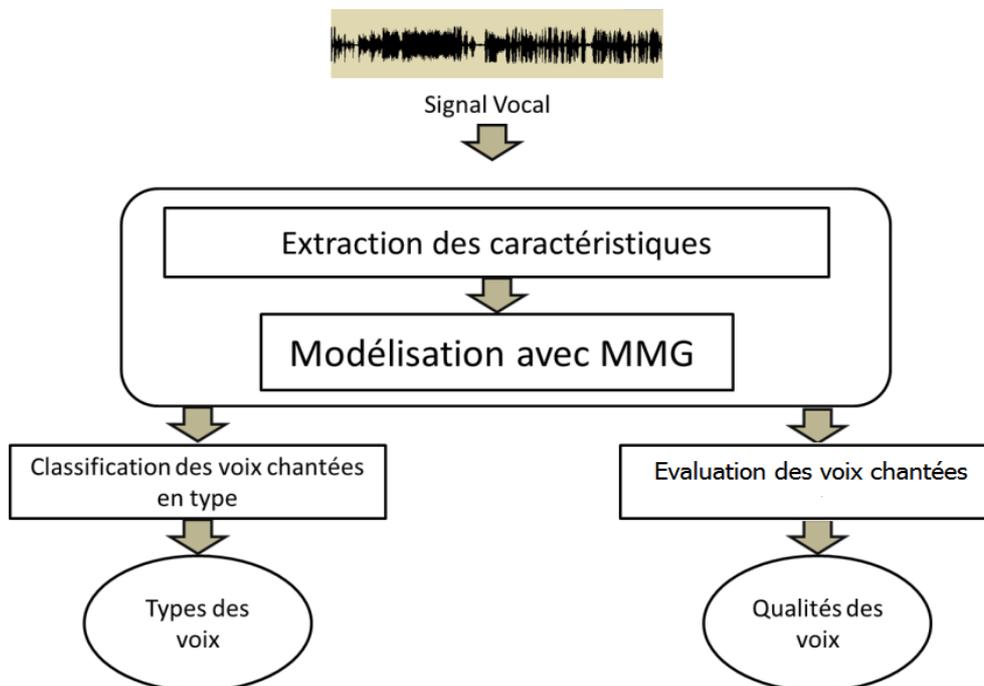


Figure III.2: Schéma du système de sélection des caractéristiques et du nombre de mélanges de gaussiennes pour la classification des voix chantées

III.3.1. Extraction des caractéristiques

Dans cette section, nous discutons les MPEG-7 LLDs « MPEG-7 low-level descriptors » et les non-MPEG-7 descripteurs qui sont pertinents pour cette étude.

III.3.1.1. MPEG-7 low-level descriptors

MPEG-7 LLDs représentent une collection de fonctionnalités audio de bas niveau qui décrivent les propriétés et les caractéristiques des sons tels que l'harmonicité, le formant et le timbre [Kim et al., 2005]. L'agrégation des descriptions de plusieurs cadres de la description d'un objet sur les médias n'est pas une partie normative de la norme MPEG-7. Les descripteurs de timbre et de spectre en charge d'une valeur de résumé unique pour tous les objets multimédias. Pour les autres MPEG-7 LLDs la norme propose des opérations

mathématiques, telles que minimale, maximale, moyenne et la variance pour un résumé [Mitrovic et al 2006]. Il existe 17 descripteurs temporelle et spectrale qui peuvent être utilisés dans plusieurs domaines et application [Lindsay et Herre, 2001]. Les descripteurs de MPEG-7 LLDs sont répartis dans les groupes suivants:

- Basic Descriptors (BD) : audio waveform (AWF), audio power (AP).
- Basic Spectral Descriptors (BSD): audio spectrum envelope (ASE), audio spectrum centroid (ASC), audio spectrum spread (ASS), audio spectrum flatness (ASF).
- Basic Signal Parameters (BSP): audio harmonicity (AH), audio fundamental frequency (AFF).
- Temporal Timbral Descriptors (TTD): log attack time (LAT) and temporal centroid (TC).
- Spectral Timbral Descriptors (STD): harmonic spectral centroid (HSC), harmonic spectral deviation (HSD), harmonic spectral spread (HSS), harmonic spectral variation (HSV) and spectral centroid (SC).
- Spectral Basis Representations (SBR): audio spectrum basis (ASB) and audio spectrum projection (ASP).

Les premiers vecteurs MPEG-7 LLDs (VCI) contiennent 121 paramètres. Ces paramètres sont répartis dans les groupes suivants [Maazouzi et Bahi, 2012]:

- 1) BD: 8 paramètres
- 2) BSD: 52 paramètres
- 3) BSP: 13 paramètres
- 4) TTD: 8 paramètres
- 5) STD: 28 paramètres
- 6) SBR : 12 paramètres

$VCI_{121} = [B_8, BS_{52}, BSIG_{13}, TT_8, ST_{28}, SBR_{12}]$

III.3.1.2. Non- MPEG-7 descripteurs

Le descripteur non-MPEG-7 contient un certain nombre de descripteurs utilisés dans plusieurs études du domaine de classification voix chantée. Les descripteurs non-MPEG-7 sont organisés par l'étude et répertoriés dans le Tableau III.1 :

Etude	Descripteurs	Abréviation
[Maazouzi et Bahi , 2011] [Nakano et al., 2006]	Mel-Frequency Cepstral Coefficients	MFCC
[Zwan, 2008]	Vibrato	VIB
[Zwan et al., 2007]	Warped Linear Predictive Coding	WLPC

Tableau III.1: Les descripteurs non-MPEG-7 [Maazouzi et Bahi , 2012].

Le deuxième vecteur non-MPEG-7 (VCII) contient 48 paramètres, les paramètres sont répartis dans les groupes suivants:

- 1) MFCC: 12 paramètres
- 2) VIB: 24 paramètres
- 3) PITCH: 1 paramètre
- 4) WLPC: 11 paramètres

$VCII_{48} = [MFCC_{12}, VIB_{24}, PITCH_1, WLPC_{11}]$.

III.3.2. Modélisation et classification avec MMG « Modèle de Mélanges Gaussiens »

Dans cette phase nous avons utilisé le même système de modélisation et de classification utilisé dans l'approche précédente [Maazouzi et Bahi, 2011] (voir sec III.1.2.).

III.3.3. La sélection des caractéristiques et de nombres de gaussien

La tâche de sélection de caractéristiques est utilisée pour réduire la taille, la suppression des données non pertinentes, ce qui augmente la précision de l'apprentissage et l'amélioration de l'intelligibilité. Il s'agit d'un processus de sélection d'un sous-ensemble de fonctionnalités pour sélectionner les caractéristiques, les variables et les mesures les plus intéressantes et pertinentes pour un système donné, de sorte que l'espace caractéristique est optimale.

Avant de poursuivre par la présentation de l'algorithme que nous proposons pour la sélection des paramètres acoustiques pour la classification des voix chantées, les deux algorithmes de recherche de solution les plus connus, sont : SFS (Sequential Forward Search) et SBS (Sequential Backward Search).

Dans cette section, nous avons introduit le système de sélection des caractéristiques et de nombre de mélanges gaussiens. Dans notre étude, nous avons proposé deux algorithmes pour la sélection. Le premier vecteur MPEG-7 LLDs (VCI) et l'autre vecteur les non-MPEG-7 descripteurs (VCII).

Les deux algorithmes sont basés sur l'algorithme SBS [Whitney, 1971]. L'algorithme SBS débute par la totalité de l'ensemble des paramètres et à chaque itération le pire paramètre est éliminé dans le sens de $J(x)$. Donc, pour l'algorithme proposé à chaque itération du processus de sélection, à partir de chaque groupe de descripteurs, le pire descripteur choisi est celui utilisée par la fonction *Find Worst (S)* est défini comme suit:

$$Find\ Worst(S): fw = \min\{J(S) - J(S - \{f\})\} \quad (III. 3)$$

Les deux algorithmes de sélection proposée dans cette étude [Maazouzi et Bahi, 2012] sont présentés ci-dessus:

Algorithme 1 : Pour le vecteur MPEG-7 LLDs (VCI)L'ensemble initial $VC \xleftarrow{\text{reçoit}} VCI$ S_j : Group des descripteurs NS_j : Nombres des descripteurs dans chaque groupe NGM : Nombres de mélange de gaussiennes**While** ($NS_j \neq 1$)For ($i=32$ to 128) $NGM = i$ For ($j=1$ to 5)Find Worst (S_j): fw From VC delete (fw) $j=j+1$

End

Accuracy calculate (VC , NGM) $i=i+32$

End

End

Vecteurs	Descripteurs	Groups des descripteurs	Précession		
			32 MG	64 MG	128 MG
VCI ₁₂₁	AWF ₄ , AP ₄ , ASE ₁₂ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ , AFF ₁ , LAT ₄ , TC ₄ , HSC ₈ , HSD ₄ , HSS ₄ , HSV ₈ , SC ₄ , ASB ₈ , ASP ₄	B ₈ , BS ₅₂ , BSIG ₁₃ , TT ₈ , ST ₂₈ SBR _{,12}	92,1%	93,4%	93%
VCI ₈₈	AP ₄ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₈ , HSS ₄ , HSV ₈ , SC ₄ , ASP ₄	B ₄ , BS ₄₀ , BSIG ₁₂ , TT ₄ , ST ₂₄ SBR _{,4}	92,1%	93,5%	93,3%
VCI ₈₀	AP ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ , TC ₄ , HSC ₈ , HSS ₄ , HSV ₈ , ASP ₄	B ₄ , BS ₃₆ , BSIG ₁₂ , TT ₄ , ST ₂₀ SBR _{,4}	90,4%	91,7%	90%
VCI ₅₂	AP ₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₈ , HSS ₄ , ASP ₄	B ₄ , BS ₁₂ , BSIG ₁₂ , TT ₄ , ST ₁₆ SBR _{,4}	87,3%	86,2%	85%
VCI ₄₈	AP ₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₁₂ , ASP ₄	B ₄ , BS ₁₂ , BSIG ₁₂ , TT ₄ , ST ₁₂ SBR _{,4}	81,3%	81,1%	79,2%

Tableau III.2 : Résultats de sélection ont utilisons le premier vecteur de caractéristiques pour l'évaluation de la qualité des voix chantées

Vecteurs	Descripteurs	Groupes des descripteurs	Précision		
			32 MG	64 MG	128 MG
VCI ₁₂₁	AWF ₄ , AP ₄ , ASE ₁₂ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ AFF ₁ , LAT ₄ , TC ₄ , HSC ₈ , HSD ₄ , HSS ₄ , HSV ₈ , SC ₄ , ASB ₈ , ASP ₄	B ₈ , BS ₅₂ , BSIG ₁₃ , TT ₈ , ST ₂₈ SBR ₁₂	91,1%	91,8%	91,5%
VCI ₈₈	AP ₄ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₈ , HSS ₄ , HSV ₈ , SC ₄ , ASP ₄	B ₄ , BS ₄₀ , BSIG ₁₂ , TT ₄ , ST ₂₄ SBR ₄	90,5%	91,7%	91,2%
VCI ₈₀	AP ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₈ , HSS ₄ , HSV ₈ , ASP ₄	B ₄ , BS ₃₆ , BSIG ₁₂ , TT ₄ , ST ₂₀ SBR ₄	89,1%	89,7%	89,3%
VCI ₅₂	AP ₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₈ , HSS ₄ , ASP ₄	B ₄ , BS ₁₂ , BSIG ₁₂ , ST ₄ , TT ₁₆ SBR ₄	85,2%	84,4%	83,5%
VCI ₄₈	AP ₄ , ASF ₁₂ , AH ₁₂ TC ₄ , HSC ₁₂ , ASP ₄	B ₄ , BS ₁₂ , BSIG ₁₂ , TT ₄ , ST ₁₂ SBR ₄	79,7%	78,1%	76,2%

Tableau III.3: Résultats de la sélection ont utilisons le premier vecteur de caractéristiques pour la classification des voix chantées en type

A partir des résultats obtenus dans les tableaux III.2 et III.3, nous avons choisi pour chaque catégorie de classification le système qui a donné les meilleurs résultats :

- VCI₈₈ avec 64 mélanges de gaussiens pour l'évaluation des qualités vocales des chanteurs
- VCI₁₂₁ avec 64 mélanges de gaussiens pour la classification des types des voix chantées.

Selon les résultats ont montré dans les tableaux III.2 et III.3, les descriptions (AP, ASC, ASS, ASF, AH, TC, HSC, HSS, HSV, ASP) est important dans le système de classification des voix chantées.

Algorithme 2 : Pour le vecteur Non-MPEG-7 descripteurs (VCI)

L'ensemble initial VC ← VCI
reçoit

D_j: Descripteurs

ND_j: Nombres des descripteurs dans chaque VC

NGM: Nombres de mélange de gaussien

For (i=32 to 128)

NGM = i

For (j= to ND_j)

j=j+1

Accuracy calculate (D_j, NGM)

End

Accuracy calculate (VC, NGM)

i=i+32

End

Vecteurs	Descripteurs	Précision		
		32 MG	64 MG	128 MG
VCII ₁₂	MFCC ₁₂	90,6%	90,5%	89,3%
VCII ₂₄	VIB ₂₄	90,5%	90,1%	90%
VCII ₁₁	WLPC ₁₁	87,2%	87,1%	86,9%
VCII ₄₈	MFCC ₁₂ , VIB ₂₄ , WLPC ₁₁	95,5%	95,9%	94,7%

Tableau III.4 : Résultats de la sélection ont utilisons le deuxième vecteur de caractéristiques pour l'évaluation de la qualité des voix chantées.

Vecteurs	Descripteurs	Précision		
		32 Gaussian Mixtures	64 Gaussian Mixtures	128 Gaussian Mixtures
FVII ₁₂	MFCC ₁₂	91,5%	91,1%	90,2%
FVII ₂₄	VIB ₂₄	92,4%	92,3%	88,9%
FVII ₁₁	WLPC ₁₁	86,8%	86,2%	80,4%
FVII ₄₈	MFCC ₁₂ , VIB ₂₄ , WLPC ₁₁	94,7%	94,5%	93,9%

Tableau III.5 : Résultats de la sélection ont utilisons le deuxième vecteur de caractéristiques pour la classification des voix chantées en type

Les résultats des tableaux III.4 et III.5 montrent que le système VCII₄₈ et 32 de mélanges gaussiens donnent les meilleurs résultats pour les deux catégories de classification, ce qui montre que chaque descripteur dans le VCII (MFCC, VIB et WLPC) est importante dans le système de classification voix chantée.

Dans la dernière étape de la sélection, nous voulons créer un nouveau vecteur contient les meilleurs descripteurs de (VCI + VCII), puis nous avons testé le nouveau vecteur (VCI + VCII). Les résultats montrent dans le tableau III.7

Catégories	Descripteurs	Précision		
		32 MG	64 MG	128 MG
Type	AP ₄ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ , TC ₄ , HSC ₈ , HSS ₄ , HSV ₈ , SC ₄ , ASP ₄ , MFCC ₁₂ , VIB ₂₄ , WLPC ₁₁	94%	94,7%	94,6%
Qualité	AWF ₄ , AP ₄ , ASE ₁₂ , ASC ₄ , ASS ₂₄ , ASF ₁₂ , AH ₁₂ , AFF ₁ , LAT ₄ , TC ₄ , HSC ₈ , HSD ₄ , HSS ₄ , HSV ₈ , SC ₄ , ASB ₈ , ASP ₄ , MFCC ₁₂ , VIB ₂₄ , WLPC ₁₁	96,5%	96,9	97%

Tableau III.6 : Résultats de la sélection ont utilisons les deux vecteurs de caractéristiques pour la classification des voix chantées en type

D'après les résultats du tableau III.7, nous avons observé que (VCI + VCII) donnent des résultats très bons. Enfin, nous avons obtenu deux systèmes, l'un pour la classification des types des voix et l'autre pour l'évaluation des voix chantées.

Le premier système: VC_T avec 64 mélanges gaussiens.

Le second système: VC_Q 128 avec les mélanges gaussiens.

- $VC_T = [AP_4, ASC_4, ASS_{24}, ASF_{12}, AH_{12}, TC_4, HSC_8, HSS_4, HSV_8, SC_4, ASP_4, MFCC_{12}, VIB_{24}, WLPC_{11}]$
- $VC_Q = [AWF_4, AP_4, ASE_{12}, ASC_4, ASS_{24}, ASF_{12}, AH_{12}, AFF_1, LAT_4, TC_4, HSC_8, HSD_4, HSS_4, HSV_8, SC_4, ASB_8, ASP_4, MFCC_{12}, VIB_{24}, WLPC_{11}]$

III.4. Utilisation de T2 FGMMs (Type-2 Fuzzy Gaussian Mixture Models) pour la classification des voix chantées

Pour cette approche nous avons utilisé T2 FGMMs, qui est une amélioration de GMM normale utilisée dans les approches 1 et 2 et présenté dans la section 2.1.

III.4.1. Extraction des caractéristiques

Dans ce travail, nous considérons deux vecteurs pour la classification des voix chantées (MPEG-7 LLDs et Non-MPEG-7).

III.4.1.1. MPEG-7 LLDs

Pour plus de détails sur le vecteur MPEG-7 LLDs « Voir la sec 2.1.1 ». La différence entre les deux approches est que les paramètres de Vecteur MPEG-7 LLDs (VC) sont énumérés d'une manière différente, voir ci-dessous [Maazouzi et Bahi, 2013]:

- 1) Paramètres 1, 2: mean Audio Power, variance Audio Power
- 2) Paramètres 3, 4: Audio WaveForm(minValue, maxValue)
- 3) Paramètres 5-38: mean ASE (34 frequency bands)
- 4) Paramètres 39-72: variances ASE (34 frequency bands)
- 5) Paramètres 73, 74: mean ASC, variance ASC
- 6) Paramètres 75, 76: mean ASS, variance ASS
- 7) Paramètres 77-100: mean ASF (24 frequency bands)
- 8) Paramètres 101-124: variances ASF (24 frequency bands)
- 9) Paramètres 125, 126: Harmonic Ratio (HR), UpperLimit of Harmonicity (ULH)
- 10) Paramètres 127, 128: mean AFF, variance AFF
- 11) Paramètres 129, 130: LAT, TC
- 12) Paramètres 131-140: Spectral Timbral Descriptors (HSC, HSD, HSS, HSV, SC)
- 13) Paramètres 141–198: Spectral Basis Representations (ASB, ASP)

III.4.1.2. Non-MPEG-7 descripteurs

Basé sur l'analyse des spectrogrammes de la voix chantée nous avons constaté que la voix chantée a deux descripteurs des caractéristiques majeures:

- Vibrato: De nombreux scientifiques soulignent que le vibrato joue un rôle très important dans la perception de la qualité de la voix [Zwan, 2006].
- Pitch: Le pitch [Li and Wang, 2006] représente la fréquence de vibration des cordes vocales. Cet élément est différent pour la voix d'un homme (entre 120Hz et 150Hz), la voix d'une femme (aux alentours de 250Hz) ou celle d'un enfant (entre 300Hz et 400Hz). Il existe plusieurs méthodes d'extraction automatique de pitch du son, la méthode autocorrélation est la plus utilisée [Kim and Whitman, 2002]. Pour cela la méthode utilisée dans notre système pour l'extraction automatique du pitch est le processus d'autocorrélation.

Le vecteur Non-MPEG-7 contient les descripteurs des caractéristiques majeures (pitch et vibrato) et d'autres descripteurs utilisés dans plusieurs études du domaine de classification des voix chantées. Les descripteurs Non-MPEG-7 sont énumérés dans le tableau III.1.

Sur les paramètres précédents, un ensemble appelé VCNM (Vecteur de Caractéristiques Non-MPEG-7) est formé, il comprend:

1. Paramètres 1-24: 24 first Mel-Frequency Cepstral Coefficients
2. Paramètres 25- 49: Vib
3. Paramètre 50: Pitch
4. Paramètres 51-64: WLPC

En utilisant les deux vecteurs (VCM, VCNM) présentés ci-dessus, le vecteur des caractéristiques (VC) est construit. Il contient 262 paramètres:

$$VC_{1-262} = [VCM_{1-198}, VCNM_{199-262}].$$

III.4.2. La sélection des caractéristiques pour le vecteur de 262 paramètres et T2 FGMMs

Dans ce travail, l'algorithme SBS [Whitney, 1971] est utilisé pour effectuer le processus de sélection des caractéristiques (pour plus de détail sur l'algorithme SBS voir la sec III.3.3.).

Les résultats de la sélection des caractéristiques sont indiqués dans la Figure III.4. Donc, à chaque itération nous avons affiché le taux d'erreur. Il peut être clairement vu que les taux d'erreurs de vecteur de caractéristiques utilisés pour la classification des voix chantées en type est proche du vecteur de caractéristiques utilisés pour l'évaluation de la qualité vocale.

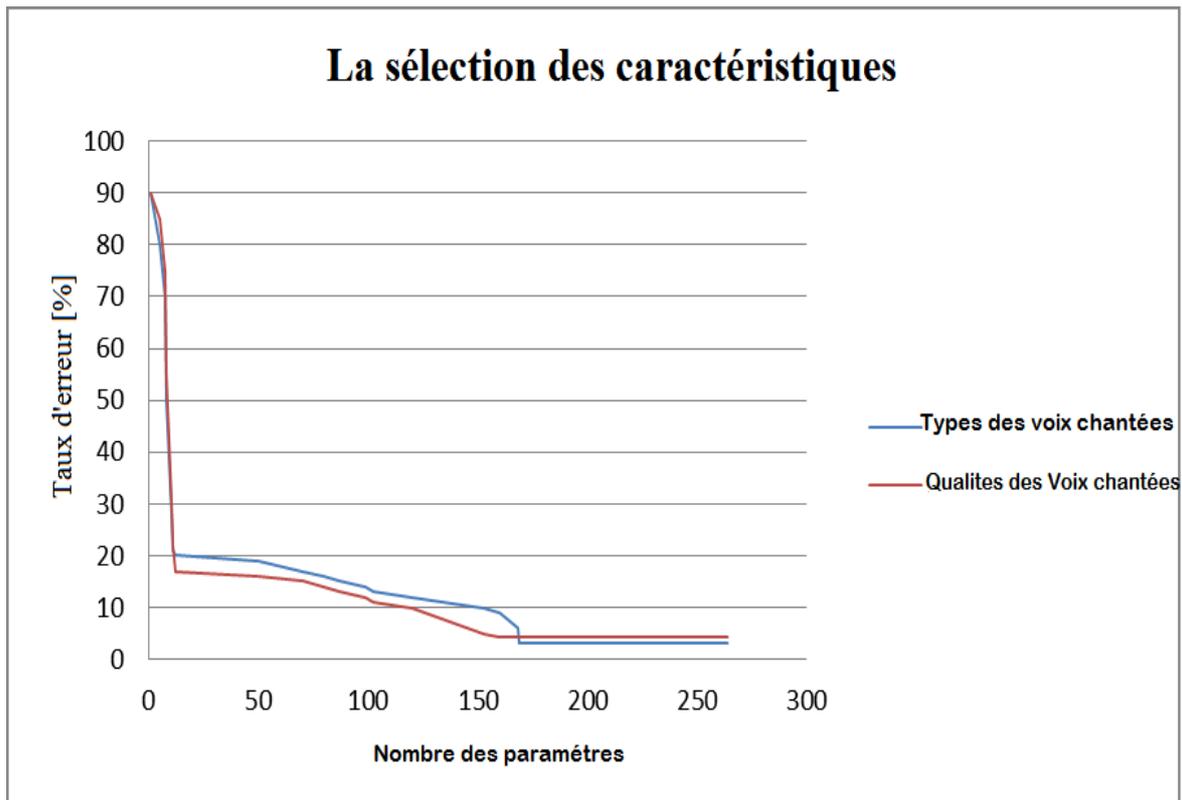


Figure III.3: Les résultats de la sélection des caractéristiques [Maazouzi et Bahi , 2013]

Après la phase de sélection à base de SBS, nous obtenons deux vecteurs de caractéristiques réduits: l'un pour la classification de type des voix chantées et un autre pour l'évaluation de la voix chantée.

- Vecteur de caractéristiques de type (VCT_{1-169}): [VCM_{1-121} , $VCNM_{123-169}$]
- Vecteur de caractéristiques de qualité (VCQ_{1-160}): [VCM_{1-124} , $VCNM_{125-160}$]

III.4.3. Modélisation et classification avec T2 FGMM « Type-2 Fuzzy Gaussian Mixture Models »

Les GMMs sont largement utilisés dans le domaine de la reconnaissance automatique du son, en séparation des sources [Todros and Tabrikian, 2007; Routtenberg and Tabrikian, 2009], identification du chanteur [Chang, 2009] et la classification de la voix chantées [Maazouzi and Bahi, 2011; Prasert et al., 2008]. T2 FGMMs méthode proposée par [Zeng et al. 2008], est une extension des GMMs selon le type de 2-ensembles flous (T2 FSS). Les T2 FGMMs utilisent un intervalle plutôt que d'un nombre précis réel pour exprimer les paramètres de GMM [Zeng et al. 2008]. Il suppose que la moyenne et la variance varient dans un intervalle avec la possibilité uniforme. Récemment, T2 FGMMs est appliqué avec succès dans le domaine de reconnaissance des formes [Zeng et al. 2008], ainsi que dans des autres domaines [Baf et al., 2008; Kai et al., 2010; Bouwmans and Baf, 2009].

III.4.3.1. Modélisation avec T2 FGMMs

La densité de mélange gaussien est une somme pondérée des densités des M composants [Maazouzi and Bahi, 2011] donné par l'équation (III.2).

T2 MFs (Type 2 Membership Functions) représentent tous les systèmes de logique floue nécessitant la spécification précise des MFs (Membership Functions) pour produire T2 FGMM-UM (T2 FGMM Uncertain Mean vector) ou T2 FGMM-UV (T2 FGMMs Uncertain Variance).

Les T2 MFs sont utilisées pour représenter la multi variée gaussienne avec UM ou UV, et remplacer les parties correspondantes de l'équation : (III.2).

Compte tenu d'une observation a d-dimension du vecteur x , le vecteur de moyen μ et la matrice de covariance diagonale $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$, la gaussienne multivariée UM ou UV:

$$g(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{\left[-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 \right] \dots \left[-\frac{1}{2} \left(\frac{x_d - \mu_d}{\sigma_d} \right)^2 \right]} \quad (\text{III.4})$$

$$\mu_1 \in [\underline{\mu}_1, \bar{\mu}_1], \dots, \mu_d \in [\underline{\mu}_d, \bar{\mu}_d],$$

Ou

$$g(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{\left[-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 \right] \dots \left[-\frac{1}{2} \left(\frac{x_d - \mu_d}{\sigma_d} \right)^2 \right]} \quad (\text{III.5})$$

Chaque composante exponentielle dans les équations. (III.4 et III.5) est GPMF-UM (Gaussian primary membership function (MF) with uncertain mean) ou l'écart type, comme indiqué à la Fig. III.4.

Dans GPMF- UM, le MF supérieure est donnée par:

$$\begin{aligned} \bar{h}(x) &= f(x, \underline{\mu}, \sigma) & \text{if } x < \underline{\mu} \\ \bar{h}(x) &= 1 & \text{if } \underline{\mu} \leq x \leq \bar{\mu} \\ \bar{h}(x) &= f(x, \bar{\mu}, \sigma) & \text{if } x > \bar{\mu} \end{aligned} \quad (\text{III.6})$$

Ou

$$f(x, \underline{\mu}, \sigma) = e^{\left[-\frac{1}{2} \left(\frac{x - \underline{\mu}}{\sigma} \right)^2 \right]} \text{ ET } f(x, \bar{\mu}, \sigma) = e^{\left[-\frac{1}{2} \left(\frac{x - \bar{\mu}}{\sigma} \right)^2 \right]} \quad (\text{III.7})$$

La MF inférieure est:

$$\underline{h}(x) = f\left(x, \underline{\mu}, \sigma\right) \quad \text{if } x \leq \frac{\mu + \bar{\mu}}{2} \quad (\text{III.8})$$

$$\underline{h}(x) = f\left(x, \bar{\mu}, \sigma\right) \quad \text{if } x > \frac{\mu + \bar{\mu}}{2}$$

Dans GPMF-Ustd, le MF supérieure : $\bar{h}(x) = f(x, \mu, \bar{\sigma})$ est et le MF inférieur est: $\underline{h}(x) = f(x, \mu, \bar{\sigma})$. Les facteurs k_m et k_v contrôlent les intervalles dans lesquels les paramètres varient dans les équations. (III.9 et III.10).

Parce que une gaussienne unidimensionnelle a 99,7% de sa masse de probabilité dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$ Ce, contrainte $K_m \in [0, 3], K_v \in [0.3, 1]$

$$\underline{\mu}1 = \mu - K_m \sigma, \bar{\mu}1 = \mu + K_m \sigma, K_m \in [0, 3], \quad (\text{III.9})$$

$$\underline{\sigma}1 = K_v \sigma, \bar{\sigma}1 = \frac{1}{K_v \sigma}, K_v \in [0.3, 1], \quad (\text{III.10})$$

Pour déterminer les meilleurs paramètres du modèle des densités de probabilité gaussiennes de chaque catégorie, nous utilisons la technique bien connue Expectation Maximization (EM) [Maazouzi et Bahi, 2013].

Pour l'évaluation des voix chantées, chaque classe est représentée par un T2 FGMM et est référée à son modèle λ .

Pour l'évaluation des voix des males chanteurs : Le T2 FGMM de la classe "Professionnelle" ($(\lambda_{\text{Professionnelle -Male}})$) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme Professionnelle male chanteur.

Le T2 FGMM de la classe "Professionnelle" ($(\lambda_{\text{semi-professionnelle -Male}})$) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme Semi Professionnelle male chanteur.

Le T2 FGMM de la classe "Professionnelle" ($(\lambda_{\text{non-professionnelle -Male}})$) : l'apprentissage ce fait sur les descripteurs extraits des sections annotées comme Non Professionnelle male chanteur.

Pour l'évaluation des femelle voix chantées et de classification des voix chantées en type:

Qualités des femelle voix chantées: $\lambda_{\text{Professionnelle -Female}}, \lambda_{\text{Semi- Professionnelle -Female}}, \lambda_{\text{Non- Professionnelle -Female}}$

Type des voix male : $\lambda_{\text{Tenor}}, \lambda_{\text{Baritone}}, \lambda_{\text{Bass}}$

Type des voix femelle: $\lambda_{\text{Soprano}}, \lambda_{\text{Mezzo-Soprano}}$ et $\lambda_{\text{Contralto}}$

Ajouter les facteurs k_m ou k_v pour produire T2 FGMM-UMs or T2 FGMM-UVs

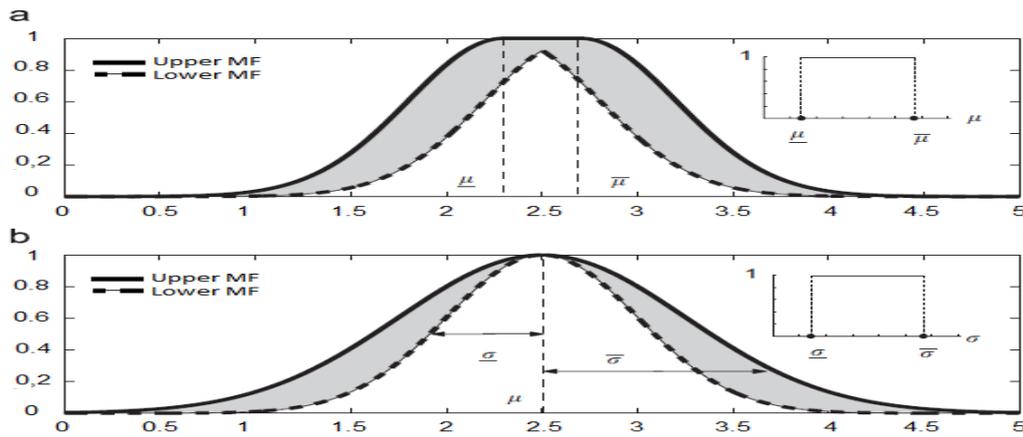


Figure III.4 : (a) GPMF-UM « Gaussian primary membership function (MF) with uncertain mean » et (b) GPMF-Ustd « Gaussian primary membership function (MF) with uncertain std »
[Maazouzi et Bahi , 2013]

III.4.2.2. T2 FGMMs Classification

Dans la phase de la classification des voix chantées en utilisent les T2 FGMMs, le critère de classification est le critère maximum a posteriori. Donc comme on a dite dans la (sec III.2.2.2.), selon le critère de maximum a posteriori : « tout vecteur d'observation x est affecté à la classe qui maximise la probabilité d'observation sachant le modèle λ associé ».

➤ Pour l'évaluation de la qualité des voix male:

$$\text{Class}(x) = \{i \mid \underline{h}(x \mid \lambda_i)\} = [\max \{\underline{h}(x \mid \lambda_{\text{Professional-Male}}), \underline{h}(x \mid \lambda_{\text{Semi-Professional-Male}}), \underline{h}(x \mid \lambda_{\text{Non-Professional-Male}})\} \text{ et } \max \{\bar{h}(x \mid \lambda_{\text{Professional-Male}}), \bar{h}(x \mid \lambda_{\text{Semi-Professional-Male}}), \bar{h}(x \mid \lambda_{\text{Non-Professional-Male}})\}]$$

➤ Pour l'évaluation de la qualité des voix femelle:

$$\text{Class}(x) = \{i \mid \underline{h}(x \mid \lambda_i)\} = [\max \{\underline{h}(x \mid \lambda_{\text{Professional-Female}}), \underline{h}(x \mid \lambda_{\text{Semi-Professional-Female}}), \underline{h}(x \mid \lambda_{\text{Non-Professional-Female}})\} \text{ et } \max \{\bar{h}(x \mid \lambda_{\text{Professional-Female}}), \bar{h}(x \mid \lambda_{\text{Semi-Professional-Female}}), \bar{h}(x \mid \lambda_{\text{Non-Professional-Female}})\}]$$

➤ Pour la classification des mâles type:

$$\text{Class}(x) = \{i \mid \underline{h}(x \mid \lambda_i)\} = [\max \{\underline{h}(x \mid \lambda_{\text{Tenor}}), \underline{h}(x \mid \lambda_{\text{Baritone}}), \underline{h}(x \mid \lambda_{\text{Bass}})\} \text{ et } \max \{\bar{h}(x \mid \lambda_{\text{Tenor}}), \bar{h}(x \mid \lambda_{\text{Baritone}}), \bar{h}(x \mid \lambda_{\text{Bass}})\}]$$

➤ Pour le classement à des femelles type:

$$\text{Class}(x) = \{i \mid \underline{h}(x \mid \lambda_i)\} = [\max \{\underline{h}(x \mid \lambda_{\text{Soprano}}), \underline{h}(x \mid \lambda_{\text{Mezzo-Soprano}}), \underline{h}(x \mid \lambda_{\text{Contralto}})\} \text{ et } \max \{\bar{h}(x \mid \lambda_{\text{Soprano}}), \bar{h}(x \mid \lambda_{\text{Mezzo-Soprano}}), \bar{h}(x \mid \lambda_{\text{Contralto}})\}]$$

III.5. Expérimentations et résultats

Nos premiers résultats sont encourageants, lors d'une expérimentation des trois systèmes, nous avons pris le même nombre des échantillons pour chaque système (Voir Tableau III.7) de différentes catégories (Type et Qualité). Dans la littérature nous distinguons trois types des

voix male (Ténor, Baryton, Basse) et trois types des voix femelle (Soprano, Mezzo-Soprano, Contralto). Pour la catégorie qualités nous avons utilisée dans la première et la deuxième approche : Bonne, Moyen, Mauvais, et pour la troisième approche : Professional, Semi-Professional, Non-Professional.

Genre	Voix Type	Nombres des échantillons	Voix Qualité	Nombres des échantillons
Male	Soprano	220 échantillons	Professional	290 échantillons
	Mezzo-Soprano	150 échantillons	Semi-Professional	220 échantillons
	Contralto	240 échantillons	Non- Professional	100 échantillons
Femelle	Ténor	230 échantillons	Professional	270 échantillons
	Barytone	190 échantillons	Semi-Professional	200 échantillons
	Bass	170 échantillons	Non- Professional	130 échantillons

Tableau III.7: Les nombres des échantillons utilisé pour chaque catégorie [Maazouzi et Bahi, 2013]

Dans les tableaux suivants nous avons présenté une comparaison des résultats des trois approches proposées et les études de [Prasert et al., 2008; Nakano et al., 2006; Zwan, 2008] pour la classification des voix chantées.

Etude	Descripteurs	Classifieur	Base	Rappel	Précision
[Maazouzi et Bahi, 2011]	MFCC 12 dimensions	Gaussian Mixture Model (GMM 12 mixtures)	1200 échantillons 100 chanteurs	90.5%	90.1%
[Maazouzi et Bahi, 2011]	Vecteur de 130 paramètres	Gaussian Mixture Model (GMM 12 mixtures)	1200 échantillons 100 chanteurs	95.5%	96.1%
[Prasert et al., 2008]	FBANK 12 dimensions	Gaussian Mixture Model (GMM)	10 chanteurs	92.1%	93.3%
[Nakano et al., 2006]	Pitch et Vibrato	Support Vector Machine (SVM)	600 échantillons 12 chanteurs	84.5%	83.4%
[Zwan, 2008]	Vecteur de 331 paramètres	The Artificial Neural Networks (ANN)	1700 échantillons 42 chanteurs	93.6%	93.4%
[Zwan, 2008]	Vecteur de 331 paramètres	Rough Set-based (RS)	1700 échantillons 42 chanteurs	90.2%	95.6%
[Maazouzi et Bahi, 2012]	Vecteur de 168 paramètres	Gaussian Mixture Model (GMM 128 mixtures)	1200 échantillons 100 chanteurs	96.2%	96.5%
[Maazouzi et Bahi, 2013]	Vecteur de 160 paramètres	Type-2 Fuzzy Gaussian Mixture Models (T2 FGMM-Ums, $km = 1.5$)	1200 échantillons 100 chanteurs	97%	97%

Tableau III.8: Comparaison des résultats des trois approches proposées et les études de [Prasert et al., 2008; Nakano et al., 2006 ;Zwan, 2008] pour la classification des qualités des voix chantées

Etudes	Descripteurs	Classifieurs	Base	rappel	précision
[Maazouzi et Bahi, 2011]	MFCC12dimension s	Gaussian Mixture Model (GMM 12 mixtures)	1200 échantillons 100 chanteurs	88.9%	90.1%
[Maazouzi et Bahi, 2011]	Vecteur de 130 paramètres	Gaussian Mixture Model (GMM 12 mixtures)	1200 échantillons 100 chanteurs	95.5%	93.2%
[Zwan, 2008]	Vecteur de 331 paramètres	The Artificial Neural Networks (ANN)	1700 échantillons 42 chanteurs	89.5%	89.6%
[Zwan, 2008]	Vecteur de 331 paramètres	Rough Set-based (RS)	1700 échantillons 42 chanteurs	64%	67.6%
[Maazouzi et Bahi, 2012]	Vecteur de 135 paramètres	Gaussian Mixture Model (GMM 64 mixtures)	1200 échantillons 100 chanteurs	95.9%	96%
[Maazouzi et Bahi, 2013]	Vecteur de 169 paramètres	Type Fuzzy Gaussian Mixture Models (T2 FGMM-UVs, $k_V=0.7$)	1200 échantillons 100 chanteurs	96.85%	96.83%

Tableau III.9: Comparaison des résultats des trois approches proposées et l'étude de [Zwan, 2008] pour la classification des voix chantées en type

Les deux tableaux présentent une comparaison des résultats des trois systèmes proposées avec les travaux [Prasert et al., 2008; Nakano et al., 2006 ;Zwan, 2008], on remarque qu'il existe toujours une différence entre les systèmes, soit dans le vecteur de caractéristiques ou dans les systèmes de modélisation et de classification. On constate qu'il n'existe pas de différence significative entre ces résultats.

A partir des résultats du tableau III.8 et III.9

- Si on compare les résultats obtenus avec le vecteur MPEG-7 et les vecteurs Non MPEG-7, Le vecteur de caractéristique (MPEG 7) donne les meilleurs résultats, par contre le vecteur de caractéristiques qui combine entre les descripteurs MPEG-7 et les descripteurs Non MPEG-7 donne des résultats meilleurs que le vecteur MPEG-7
- Les GMMs sont plus performant que les autres systèmes de modélisation (RNA, SVM...) et de classification pour la classification des voix chantées.
- Les T2FGMMs présentent les meilleurs résultats pour la classification des voix chantées pour les deux catégories.

III.6. Conclusion

Dans ce chapitre, nous avons présenté nos expérimentations dans le cadre de la catégorisation des voix chantées pour les deux catégories (type et qualité). Ceci a nécessité de nombreux calibrages et des tests pour définir l'ensemble de caractéristiques qui soit optimal. Nous avons aussi apporté une contribution qui consiste en l'utilisation des T2 Fuzzy GMM pour la classification des voix chantées.

Au final, nous obtenons un système qui étant donné l'enregistrement d'une chanson, nous pouvons obtenir les caractéristiques vocales du chanteur. Ceci nous permet de compléter les statistiques fournies par les sites musicaux.

La fouille de données

- 1. Introduction*
- 2. Le processus de fouille de données*
- 3. Les techniques de fouille de données*
- 4. Les ensembles des classifieurs*
- 5. La technique proposée (Multi Decision Tree ou multiple arbres de décision)*
- 6. Expérimentations et résultats*
- 7. Conclusion*

IV.1. Introduction

La fouille de données est un composant du domaine passionnant de l'apprentissage automatique et le calcul adaptatif. L'exploration de données est souvent définie comme l'extraction automatique des informations nouvelles et intéressantes de grands ensembles de données. L'exploration de données, telle que nous la connaissons, a ses racines dans les statistiques, la théorie des probabilités, les réseaux de neurones, et les systèmes des experts de l'intelligence artificielle.

La fouille de données n'est pas une discipline limitée, mais elle nécessite une combinaison de plusieurs disciplines et techniques. L'exploration de données se distingue des statistiques traditionnelles dans le sens que maintenant nous traitons des ensembles de données potentiellement très importants qui peuvent aller de giga-octets, téraoctets, à penta bytes. Pendant un certain temps, un problème a été considéré comme un problème de data mining que si les données pouvaient pas être stockées dans la mémoire de travail d'un ordinateur tout-à-temps. D'autres définitions du data mining ont insisté pendant un certain temps que les données doivent provenir de différentes bases de données. Ainsi, des problèmes intéressants et extrêmement difficiles, comme la découverte de gènes et le repliement des protéines en bio-informatique, ne seraient pas admissibles en tant que légitimes problèmes d'extraction de données en vertu de ces définitions restrictives. La fouille de données est différente des méthodes plus traditionnelles en ce sens que pour les grandes quantités de données, de nombreux algorithmes classiques, tels que l'algorithme K-means pour la classification, ne s'adaptent pas bien. En général, on peut résumer la fouille de données typique par la présence de :

- A. Un ensemble de données qui peut être très important
- B. Un problème qui est généralement difficile et n'est souvent pas bien défini
- C. L'existence de données manquantes ou défectueuses
- D. L'existence de redondances dans les données domaines, mais les champs redondants n'ont pas tous la même qualité.

Dans ce chapitre nous allons introduire le domaine de la fouille de données (en anglais, data mining) et les techniques qui existent dans ce domaine. Après un bref état de l'art sur la fouille de données nous allons présenter une nouvelle approche de fouille de données basées sur les arbres de décision.

IV.2. Le processus de fouille de données

De nombreuses applications de fouille de données peuvent être représentées dans un modèle de dessin animé que nous appellerons le processus d'exploration de données standard. Ce

processus implique la collecte de données, le nettoyage des données, le prétraitement des données et la transformation d'un sous-ensemble de données dans un fichier plat, en un ou plusieurs modèles qui peuvent être des modèles prédictifs, des graphes ou des visualisations de données qui conduisent à la formulation de règles et enfin reconstituer la plus grande image. Ce processus est décrit dans la figure suivante :

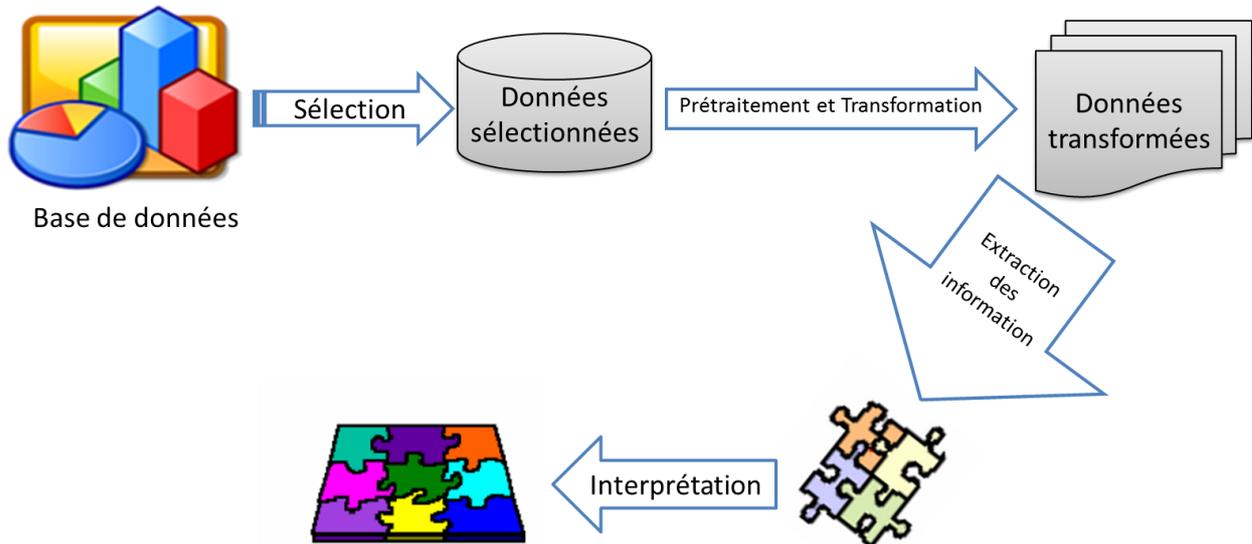


Figure IV.1: Illustration cartonne du processus de fouille de données

La figure précédente montre que le processus de fouille de données se passe en quatre étapes :

- 1) La sélection de données
- 2) Prétraitement et transformation
- 3) L'extraction des informations
- 4) L'interprétation

Nous en déduisons que la fouille de données est l'étape de l'Extraction de Connaissances à partir de Données qui vise à extraire des régularités (ou des irrégularités) de l'ensemble de données préparées. Les tâches primaires du data mining sont:

- ❖ **Caractérisation et Comparaison** : La description de données fournit une vue générale d'une classe de données ou une distinction d'un ensemble de classes comparatives. Elle permet de décrire les ensembles de données à différents niveaux de granularité et selon différents angles. La caractérisation fournit un regroupement «summarization» de la collection de données, cependant la comparaison de classe ou de concept (aussi connue comme la discrimination) fournit des descriptions comparant des collections de données.
- ❖ **Classification et Prédiction**: Il s'agit de trouver des modèles ou fonctions qui décrivent et distinguent les classes de données ou concepts. Le modèle trouvé est basé sur l'analyse d'un ensemble de données « training data ». Il s'agit simplement de trouver un modèle en utilisant les objets pour lesquels les classes sont connues dans un

but de prévoir les classes des objets pour lesquels leurs classes sont inconnues. Le modèle trouvé peut être représenté sous différentes formes comme règle de classification «IF-THEN-ELSE», arbre de décision, formule mathématique ou réseaux de neurones. Dans le cas où les valeurs à prévoir sont des données numériques la tâche est appelée prédiction.

- ❖ **Le regroupement «Clustering en anglais»:** Le regroupement est le processus qui conduit à générer des classes à partir d'analyse des objets de base. Les objets sont groupés en se basant sur la maximisation de similarités intra classes et la minimisation de similarité extra classes. Le regroupement peut faciliter la taxonomie (hiérarchies de classes). A l'encontre de la classification, le clustering est une tâche descriptive.
- ❖ **Analyse d'associations:** Elle consiste à trouver des modèles décrivant des dépendances significatives entre les variables. L'analyse d'association est largement utilisée dans le «market basket» et l'analyse des transactions de données [Raorane et al., 2012]. La découverte des relations peuvent être représentées sous la forme de règles d'association ou des ensembles de motifs fréquents.
- ❖ **Analyse d'évolution et de déviation:** L'analyse d'évolution des données décrit et modélise les régularités des objets pour lesquels le comportement change dans le temps (évolution générale) ainsi que leurs déviations (la différence entre la valeur mesurée et sa correspondance de référence telle que la valeur précédente ou la valeur normative). Un exemple de cette analyse est la prédiction des tendances boursières, ou encore l'étude des changements climatiques.

IV.3. Les techniques de fouille de données

Plusieurs techniques de fouille de données ont été proposées dans les dernières années. Ces techniques sont appliquées dans les différents sous domaines de la fouille de données tel que : la fouille de données textuelle, la fouille de données audio, la fouille de données image et la fouille de données vidéo.

Globalement on peut dire qu'il existe deux types de techniques de fouille de données : les algorithmes d'apprentissage automatique, et les ensembles des classifieurs où nous utilisons plusieurs classifieurs.

IV.3.1 Les algorithmes d'apprentissage

Avant de citer les différents algorithmes d'apprentissage automatique il faut répondre à la question : à quoi ça sert l'apprentissage automatique?

L'apprentissage automatique (Machine Learning en anglais) correspond au domaine se consacrant au développement d'algorithmes permettant à une machine d'apprendre à partir d'un ensemble de données [Larochelle, 2008]. Bien que la motivation originale de ce domaine était de permettre la mise sur pied de systèmes manifestant une intelligence artificielle, les algorithmes issus de ce domaine sont maintenant répandus dans bien d'autres domaines, tels que l'intelligence artificielle, les statistiques, les sciences cognitives, la théorie des probabilités, et l'optimisation du signal et de l'information... Généralement, il existe deux

types d'apprentissage automatique (Apprentissage supervisé et Apprentissage non supervisé), dans ce cadre plusieurs algorithmes sont proposés tels que : Classifieur Bayésien, RB (Rules – Based Algorithm), Algorithme Génétique, Réseaux de Neurones Artificiels (RNA), Machines à Vecteurs Supports (MVS), K Plus Proches Voisins (K Nearest Neighbors), Arbre de décision.

IV.3.1.1. Le Classifieur Bayésien :

La classification bayésienne représente une méthode d'apprentissage supervisé, aussi bien qu'une méthode de classification statistique. En supposant un modèle sous-jacent probabiliste et il nous permet de saisir l'incertitude sur le modèle d'une manière raisonnée, en déterminant les probabilités des résultats. Il peut résoudre les problèmes de diagnostics et de prédiction.

Ce classifieur est basé sur le théorème de Bayes et l'hypothèse de maximum a posteriori. L'Hypothèse de Bayes naïve d'indépendance conditionnelle est souvent faite pour réduire le coût de calcul.

IV.3.1.1.1. Théorème de Bayes

Soit $X = \{x_1, x_2, \dots, x_n\}$ un échantillon, dont les composantes représentent les valeurs faites sur un ensemble de n attributs. En termes bayésiens, X est considéré comme «Preuves». Soit H une hypothèse, comme celle que la donnée X appartient à une classe spécifique C . Pour les problèmes de classification, notre objectif est de déterminer $P(H | X)$, la probabilité que l'hypothèse H est titulaire compte tenu de la «preuve», (c'est à dire les données X observées de l'échantillon). En d'autres termes, nous cherchons la probabilité que X l'échantillon appartient à la classe C , étant donné que nous connaissons la description de l'attribut de X .

$P(H | X)$ est la probabilité à posteriori de H conditionnée par X . Par exemple, supposons que nos échantillons de données ont pour attributs: l'âge et le revenu, et que X l'échantillon est un client de 30 ans avec un revenu de 500,000 DA. Supposons que H est l'hypothèse que notre client va acheter un ordinateur. Alors $P(H | X)$ est la probabilité que X clients vont acheter un ordinateur sachant l'âge du client et son revenu.

En revanche, $P(H)$ est la probabilité à priori de H . Pour notre exemple, il s'agit de la probabilité qu'un client achète un ordinateur, indépendamment de l'âge, le revenu, ou d'autres informations. La probabilité à posteriori $P(H | X)$ est basée sur plusieurs informations (sur le client) de la probabilité à priori $P(H)$, qui est indépendante de X .

De même, $P(X | H)$ est la probabilité à posteriori de X conditionnée par H . Autrement dit, c'est la probabilité qu'un client X , a 30 ans et gagne 500,000 DA, sachant que le client va acheter un microordinateur. $P(X)$ est la probabilité à priori de X . Dans notre exemple, c'est la probabilité qu'une personne de notre ensemble de clients est de 30 ans et gagne 500,000 DA.

Selon le théorème de Bayes, la probabilité que l'on veut calculer $P(H | X)$ peut être exprimée en termes de probabilités $P(H)$, $P(X | H)$, et $P(X)$ telle que :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}, \quad (\text{IV.1})$$

IV.3.1.1.2. Le classifieur bayésien naïf

Le classificateur bayésien fonctionne comme suit :

Soit T un ensemble d'échantillons de l'apprentissage, chacun avec son étiquette de classe. Il existe k classes, C_1, C_2, \dots, C_k . Chaque échantillon est représenté par un vecteur à n dimensions, $X = \{x_1, x_2, \dots, x_n\}$, représentant n valeurs de mesure des n attributs A_1, A_2, \dots, A_n , respectivement.

Étant donné un échantillon X, le classificateur prédit que X appartient à la classe ayant la plus haute probabilité a posteriori, conditionnés par X. D'où X appartient à la classe C_i si et seulement si

$$P(C_i|X) > P(C_j|X) \text{ pour } 1 \leq j \leq m, j \neq i \quad (\text{IV.2})$$

Ainsi, nous trouvons la classe qui maximise $P(C_i|X)$. La classe C_i pour lequel $P(C_j|X)$ est maximisée est appelé l'hypothèse maximum a posteriori. Par le théorème de Bayes

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (\text{IV.3})$$

Comme $P(X)$ est la même pour toutes les classes, que $P(X|C_i)P(C_i)$ ne doit être maximisée. Si la classe des probabilités a priori, $P(C_i)$, ne sont pas connus, il est communément admis que les classes sont équiprobables, c'est-à-dire $P(C_1) = P(C_2) = \dots = P(C_k)$, il s'agit donc de maximiser $P(X|C_i)$. Sinon, nous maximisons $P(X|C_i)P(C_i)$. Notez que la classe des probabilités a priori peut être estimée par $P(C_i) = \text{freq}(C_i, T) / |T|$.

Compte tenu de jeux de données avec de nombreux attributs, il serait coûteux en calcul de calculer $P(X|C_i)$. Afin de réduire calcul dans l'évaluation de $P(X|C_i)P(C_i)$, l'hypothèse de Bayes naïve d'indépendance conditionnelle est faite. Cela suppose que les valeurs des attributs soient conditionnellement indépendantes les uns des autres, compte tenu de l'étiquette de classe de l'échantillon. Mathématiquement, cela signifie que

$$P(X|C_i) \approx \prod_{k=1}^n P(x_k|C_i) \quad (\text{IV.4})$$

Les probabilités $P(X_1|C_i), P(X_2|C_i), \dots, P(X_n|C_i)$ peuvent facilement être estimée à partir de l'ensemble d'apprentissage. Rappelons que x_k ici se réfère à la valeur de l'attribut A_k pour l'échantillon X.

- ❖ Si A_k est catégorique, alors $P(X_k|C_i)$ est le nombre d'échantillons de la classe C_i en T ayant le x_k valeur de l'attribut A_k , divisé par $\text{freq}(C_i, T)$, le nombre d'échantillons de la classe C_i en T.

- ❖ Si A_k est continu-évaluées, alors nous supposent généralement que les valeurs ont une distribution gaussienne avec une moyenne μ et l'écart type σ défini par

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2} \quad (\text{IV.5})$$

De sorte que

$$P(X_k|C_i) = g(X_k, \mu_{C_i}, \sigma_{C_i}) \quad (\text{IV.6})$$

Nous avons besoin de calculer μ_{C_i} et σ_{C_i} qui sont la moyenne et l'écart type des valeurs de l'attribut A_k pour les échantillons de formation de la classe C_i .

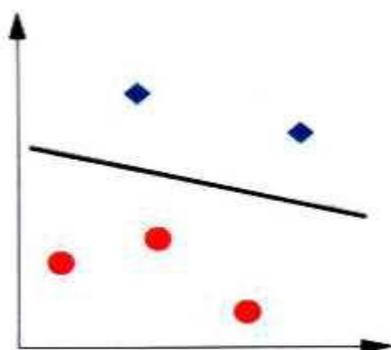
Afin de prédire l'étiquette de la classe de X , $P(X|C_i)P(C_i)$ est évaluée pour chaque classe C_i . Le classificateur prédit que l'étiquette de classe de X est C_i si et seulement si elle est la classe qui maximise $P(X|C_i)P(C_i)$.

IV.3.1.2. Machines à Vecteurs Supports

La Machines à Vecteurs Supports est une nouvelle méthode d'apprentissage par rapport aux autres méthodes d'apprentissage. La Machines à Vecteurs Supports (Support Vector Machine or SVM en anglais) a été introduite en 1992 par Vapnik et ses collègues [Boser et al., 1992]. Dans sa forme originale, SVM est un algorithme d'apprentissage pour la classification linéaire. C'est seulement plus tard qu'il a été utilisé pour la régression, l'analyse en composantes principales, la détection de nouveautés et aussi pour les cas non-linéaires.

Parmi les modèles des SVMs, on distingue les cas linéairement séparables et les cas non-linéairement séparables. Dans les premiers cas, il est facile de trouver le classifieur linéaire.

Cas linéairement séparable



Cas non linéairement séparable

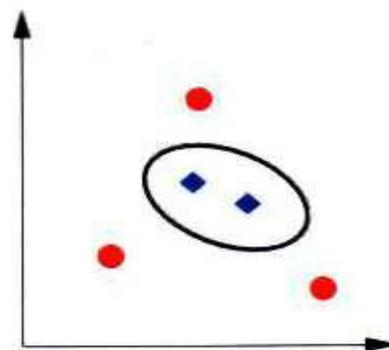


Figure IV.2: les types des modelés de SVMs

Dans la plupart des problèmes réels, il n'y a pas de séparation linéaire possible entre les données. Pour les cas non linéairement séparables l'idée est de changer l'espace des données afin de surmonter l'inconvénient des cas non-linéairement séparables. La transformation non-linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel

espace, espace de représentation. Intuitivement, plus la dimension de l'espace de représentation est grand, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée.

IV.3.1.3. Les arbres de décision

Les arbres de décision sont les méthodes les plus populaires d'apprentissage et ils sont très utilisés dans le domaine de la fouille de données. Les arbres de décision sont une représentation commode de fonctions de classifications, moins puissante que les règles de production mais plus facile à utiliser. Un arbre de décision permet de classer un objet à l'aide de questions : chaque nœud de l'arbre représente une question, chaque lien est une réponse à la question, et chaque feuille est une classe. En règle générale, la construction de l'arbre de décision est un processus récursif [Taniar et al., 2008].

Un arbre de décision est représenté sous forme d'une arborescence où: chaque nœud interne teste un attribut, chaque branche correspond à une valeur d'attribut et chaque nœud feuille attribue un classement. Alors, un arbre de décision peut être utilisé pour classer un exemple en partant de la racine de l'arbre et en se déplaçant à travers elle jusqu'à un nœud de feuille, qui fournit la classification de l'instance.

IV.3.1.3.1. La construction d'un arbre de décision

Généralement, la construction d'un arbre de décision optimal est un problème NP-complet, optimal au sens où il minimise le nombre d'erreurs de classification. Aussi, il ne faut pas avoir l'espoir de construire l'arbre de décision optimal pour un jeu d'exemples donnée [Taniar et al., 2008]. On va se contenter d'en construire un qui soit correct. Plusieurs algorithmes ont été proposés, notamment CART dans les années 1980 par [Breiman, 1994]. On décrit ici l'algorithme ID3 de R. Quinlan proposé en 1986 qui a été raffiné par la suite (C4.5 puis C5) (cf. [Quinlan, 1986; Quinlan, 1993]). On constate expérimentalement que ces algorithmes sont très performants : ils construisent rapidement des arbres de décision qui prédisent avec une assez grande fiabilité la classe de nouvelles données.

On distingue deux type d'arbres de décision, les arbres de décision binaires (ou toutes les questions ont deux réponses) et les arbres de décision non binaires.

Algorithme de construction de l'arbre de décision binaire

Procédure Construit_arbre(X)

debut

 Si tous les points de X sont dans la même classe alors

 Créer une feuille de cette classe

 sinon

 choisir le meilleur *sélecteur* (la meilleure question)

 pour créer un noeud. Séparer X suivant ce sélecteur en X_d et X_g .

 Construit_arbre(X_d)

 Construit_arbre(X_g)

 finsi

Fin

Pour les arbres de décision non binaires il existe plusieurs algorithmes tels que (C4.5, C5, CHAID...).

IV.3.1.3.2. Les algorithmes des arbres de décision

ID3 (arbre de décision inductif) proposé par [Quinlan, 1986]. Il est basé sur l'algorithme de Hunt (l'un des premiers algorithmes) [Hunt, 1962]. Comme d'autres algorithmes d'arbres de décision, l'arbre est construit en deux phases; la croissance des arbres et des arbres PRUNIN [Anyanwu et Shiva, 2009]. IDE3 utilise la mesure du gain d'information dans le choix de l'attribut de fractionnement. Il n'accepte que les attributs catégoriques dans la construction d'un modèle d'arbre [Quinlan, 1986, 1987].

C4.5 est le successeur de l'ID3. Proposé par [Quinlan, 1993]. Cet algorithme est implémenté dans plusieurs outils de data mining comme [Weka, 2012]. Il est basé sur l'algorithme de Hunt lui aussi comme ID3, il est mis en œuvre en série. L'élagage de C4.5 ce fait en remplaçant le nœud interne avec un nœud de feuille en réduisant ainsi le taux d'erreur [Podgorelec et al., 2002]. Contrairement à ID3, C4.5 accepte les attributs continus et catégoriels dans la construction de l'arbre de décision. Il a une méthode d'élagage des arbres améliorée qui réduit le bruit dû aux erreurs de classification ou des détails trop nombreux dans l'ensemble de données d'apprentissage.

CART (classification and regression trees) a été proposé par [Breiman et al., 1984]. La construction d'un arbre de décision par CART est basé sur le partage binaire des attributs [Podgorelec et al., 2002]. Il utilise la mesure de l'index Gini pour la sélection de l'attribut de fractionnement. L'élagage est effectuée dans CART en utilisant une partie de l'ensemble de données d'apprentissage [Breiman et al., 1984]. CART utilise les attributs numériques et catégoriels pour la construction de l'arbre de décision et possède des caractéristiques intégrées qui traitent les attributs manquants [Quinlan, 1993]. CART est unique par rapport à d'autres algorithmes, il est également utilisé pour l'analyse de régression à l'aide des arbres de régression [Quinlan, 1993]. La fonction de l'analyse de régression est utilisée dans la prévision d'une variable dépendante (résultat) étant donné un ensemble de variables prédictives sur une période de temps donnée [Podgorelec et al., 2002].

C5 est une version plus récente d'algorithmes ID3 et C4.5. Il était mis en œuvre dans les logiciels commerciaux, et il n'est pas possible d'avoir les détails.

IV.4. Les ensembles des classifieurs

Les méthodes d'apprentissage apprennent à identifier les sous-espaces de l'espace d'entrée pour lesquels la solution est identique. Lorsqu'un nouveau cas est soumis au système, celui-ci identifie le sous-espace correspondant et retourne la réponse associée.

D'autres familles d'algorithmes distribuent la tâche entre plusieurs méthodes d'apprentissage et combinent les solutions partielles pour obtenir la solution générale. L'apprentissage consiste alors à déterminer les solutions partielles et à trouver une manière efficace pour les combiner. Lorsque la combinaison des réponses des experts (systèmes d'apprentissage) ne

dépend pas de l'entrée, on parle de combinaison statique. Les méthodes de boosting et bagging, très étudiées actuellement, sont l'archétype de cette approche. Lorsque la combinaison dépend de l'entrée, on parle alors de structures dynamiques, dont les mélanges d'experts et les mélanges hiérarchiques d'experts sont les méthodes les plus représentatives.

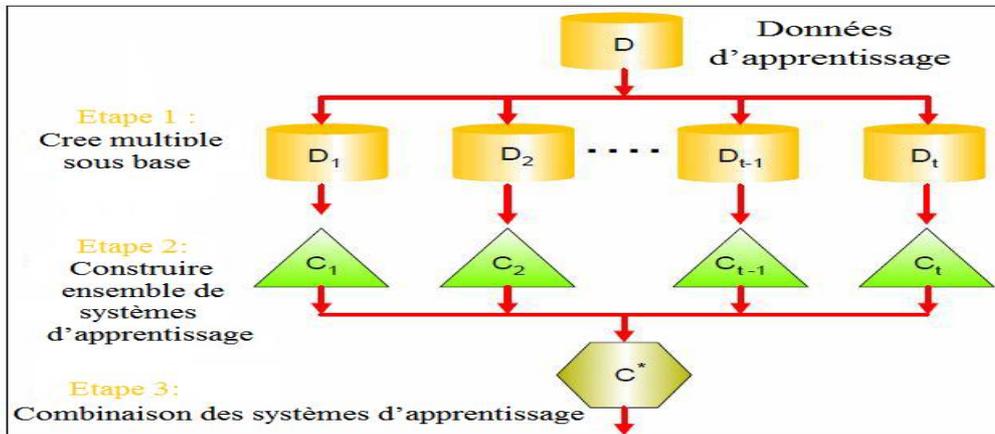


Figure IV.3: Idée générale [Bouguessa et Wang, 2011]

Les recherches en apprentissage artificiel initiées au début des années 90 montrent qu'il est possible d'atteindre une décision aussi précise que souhaitée par une combinaison judicieuse d'experts imparfaits mais correctement entraînés.

IV.4.1 Méthodes de Construction des ensembles de classifieurs

De nombreuses approches ont été proposées pour construire des ensembles d'apprentissage telles que :

- Utiliser les connaissances a priori du problème et/ou du système d'apprentissage : cette méthode consiste à faire varier le type de classifieur utilisé, l'architecture, les paramètres pour créer des classifieurs complémentaires
- Injecter de l'aléatoire : par exemple l'algorithme de rétro-propagation du gradient dans un MLP est souvent exécuté plusieurs fois en utilisant différentes (aléatoires) conditions initiales (les poids synaptiques)
- Manipuler les données d'apprentissage : consiste à entraîner N classifieurs sur N bases différentes (cross-validated committees), bagging, boosting
- Manipuler les caractéristiques d'entrée : bruite les caractéristiques d'entrée, Random subspace method
- Manipuler les caractéristiques de sortie : cette méthode consiste à partitionner l'ensemble des classes de différentes façons (ECOC)

IV.4.2 Le Bagging

C'est une méthode proposée par [Breiman, 1996] pour construire des ensembles de classifieurs, chaque classifieur étant entraîné sur une réplique différente de la base d'apprentissage. Le bagging (de bootstrap aggregation) est une méthode qui, comme le boosting, combine des hypothèses pour obtenir une hypothèse finale. Cependant la méthode est plus simple et généralement moins performante. L'idée de base est d'entraîner un

algorithme d'apprentissage élémentaire sur plusieurs bases d'apprentissage obtenues par tirage avec remise de A' (avec $A' < A$) exemples d'apprentissage dans l'échantillon d'apprentissage Z . Pour chaque tirage b (pour *bag*), une hypothèse Hyp_b est obtenue. L'hypothèse finale est simplement la moyenne des hypothèses obtenues sur S tirages au total :

$$T(x) = \frac{1}{S} \sum_{b=1}^S Hyp_b(x) \quad (IV.7)$$

L'une des justifications de cette méthode est que si les hypothèses Hyp_b calculées pour chaque tirage b ont une variance importante (donc sont sensibles à l'échantillon des A' exemples d'apprentissage), alors leur moyenne T aura une variance réduite.

IV.4.3 Le Boosting

La méthode du Boosting a été inspirée de l'algorithme d'apprentissage Hedge (β) [**Freund et Schapire., 1995**]. Le principe du Boosting est de réutiliser un classifieur plusieurs fois en attribuant aux instances d'apprentissage, à chaque fois, une pondération différente et puis combiner les résultats afin de trouver un seul classifieur « fort/très précis » [**Meir et Ratsch., 2003**].

AdaBoost (Adaptive Boosting) est une méthode algorithmique connue dans le domaine de l'apprentissage. Cette méthode a été initialement conçue par Robert Schapire et Yoav Freund en 1995 [**Freund et Schapire., 1995**]. Cette approche est la dérivée la plus pratiquée de la méthode du Boosting qui vise à stimuler la performance de l'algorithme d'apprentissage.

Le Boosting, et particulièrement l'algorithme AdaBoost, a été employé avec succès avec de nombreux algorithmes d'apprentissage « faibles ». Il est utilisé souvent par exemple avec des arbres de décision et dans les systèmes d'apprentissage des règles. Le Boosting a été testé sur des domaines d'application variés. En général, l'utilisation du Boosting a pour résultat d'améliorer souvent sensiblement les performances en apprentissage.

IV.4.4 L'apprentissage en cascade

Dans l'apprentissage en cascade, les classificateurs élémentaires ne sont pas placés en parallèle comme dans le vote ou le boosting, mais en série. Autrement dit, l'organisation est la suivante : le premier classificateur élémentaire est lancé sur la donnée. Si le résultat produit est considéré comme fiable, il est conservé. Sinon, c'est le second classificateur élémentaire qui traite la donnée. Si son résultat est fiable, on le conserve, sinon on passe la main au classificateur élémentaire suivant, et ainsi de suite.

IV.4.5 Les forêts aléatoires (Random Forest)

Les forêts aléatoires ou Random Forest (RF) en anglais, ont été inventées par Breiman ([**Breiman, 2001**]). Elles sont en général plus efficaces que les simples arbres de décision mais possède l'inconvénient d'être plus difficilement interprétables.

La méthode RF [**Montillo, 2009**] est un exemple de méta-apprentissage : un algorithme est utilisé pour traiter les représentations de concepts issus de multiples machines d'apprentissage.

Concernant le vocabulaire des forêts aléatoires, il existe une ambiguïté dans la littérature. Leo Breiman, dans son article de 2001 définit les forêts aléatoires comme une famille de méthodes. Or, dans le même article, il présente un cas particulier de forêts aléatoires, appelées

RF, qu'il a implémentées [Breiman et Cutler, 2005]. Par suite, ce sont ces RF qui ont été quasi-systématiquement utilisées dans de très nombreuses applications réelles. Et pour cause, le programme est accessible à tous, est facile d'utilisation et la méthode atteint des performances exceptionnelles. Finalement, la dénomination "forêts aléatoires" désigne maintenant très souvent les RF. On trouve également le terme de "forêts aléatoires de Leo Breiman" pour désigner les RF.

RF est certainement l'une des méthodes les plus efficaces pour l'apprentissage d'arbres de décision [Do et al., 2009]. Cette méthode consiste à créer une famille d'arbres non élagués (où à chaque nœud la séparation des individus qui est réalisée à partir d'un sous-ensemble des attributs prédictifs choisis aléatoirement dans l'ensemble des attributs) à partir d'échantillon bootstrap (échantillon construit par tirage aléatoire avec remise à partir de l'ensemble d'apprentissage). Les performances d'une forêt d'arbres dépendent de la qualité des individus la composant et de leur indépendance. Les RFs sont fondées sur des arbres non élagués afin de réduire l'erreur de biais. En outre, le processus aléatoire permet d'assurer une faible corrélation entre les arbres [Do et al., 2009].

L'algorithme RF se fait en trois principales étapes :

- 1) Construire un arbre dépendant d'un vecteur aléatoire
- 2) Collecter les prédictions de chaque arbre
- 3) Décider à la majorité

IV.5. La technique proposée (Multi Decision Tree ou multiple arbres de décision)

L'objectif de multi-classifieur basé sur les méthodes d'apprentissage automatique est de construire un ensemble de modèles de prévision et de combiner leurs résultats dans une seule prédiction [Maazouzi et Bahi, 2012a].

Dans ce contexte, plusieurs contribution de méthodes d'apprentissage sont proposées [Gal-Or et al, 2010; Qiu et al, 2008]. L'objectif de notre recherche est d'améliorer la classification et le processus de prédiction avec des arbres de décision. Dans ce contexte, au cours des dernières années, plusieurs méthodes ont été développées, comme bagging des arbres de décision [Loh, 2008], Les forêts aléatoires des arbres de décision [Loh et Shih, 1997] et Boosting des arbres de décision ou BDT (Boosting Decision Tree) [Roe et al., 2006].

Les deux algorithmes les plus reconnus dans le domaine de fouille de données sont : CART [Breiman et al., 1984] et C5 (la version la plus récente de ID3 et C4.5 [Breiman et al., 1984]). Ces algorithmes sont utilisés pour leurs performances [Maazouzi et Bahi , 2012].

Dans le travail de [Maazouzi et Bahi, 2012a] nous avons proposé une nouvelle méthode appelée MDT (Multi Decision Tree) pour la fouille de données, cette méthode est basée sur les arbres de décision. Dans cette section nous avons présenté MDT algorithme. Le processus de MDT modélisation et classification est affiché dans la figure suivante :

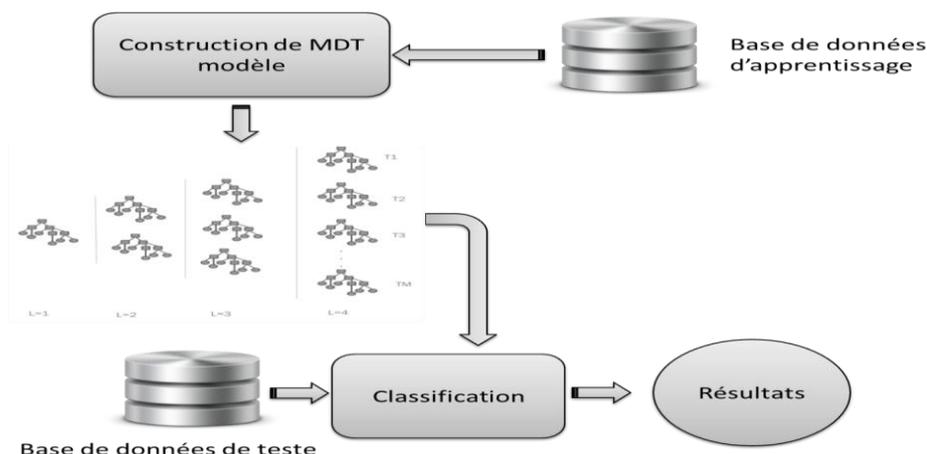


Figure III.4: Schéma général de MDT classification

Dans ce qui suit, nous présentons les deux étapes de Multi decision tree (MDT): la phase d'apprentissage pour déterminer les caractéristiques des classes et la phase de reconnaissance (classification).

IV.5.2.1. Phase d'apprentissage

Notre modèle consiste organiser des arbres de décision combinées sous la forme d'un réseau. Où, l'arbre de décision représente une représentation graphique d'un processus de classification.

Comme dans la méthode Bagging d'arbres de décision et Forêt d'arbres décisionnels, notre méthode est une agrégation de plusieurs arbres de décision comme classifieurs de base. L'objectif de l'utilisation de notre méthode est d'améliorer le processus d'apprentissage et de classification on utilisant les arbres de décision.

Dans notre système nous avons partitionné les données en sous bases, et pour chaque sous base nous construisons un arbre différent. Donc, pour chaque arbre une classe peut dépendre d'un attribut différent. La construction de modèle se fait en trois étapes [Maazouzi et Bahi, 2012a]:

- 1) La première étape est de choisir le nombre de couches n .
- 2) Dans la deuxième étape nous avons construit des sous base (SB) pour chaque couche tel que le nombre des attributs dans chaque couche est :

$$C_i SB_j = B / NuC \quad (\text{III.8})$$

Où

B : La Base complète

NuC : Numéro de la couche.

- 3) La dernière étape consiste à appliquer l'algorithme d'arbre de décision (C4.5) [Adrian et al., 1996] pour chaque SB.

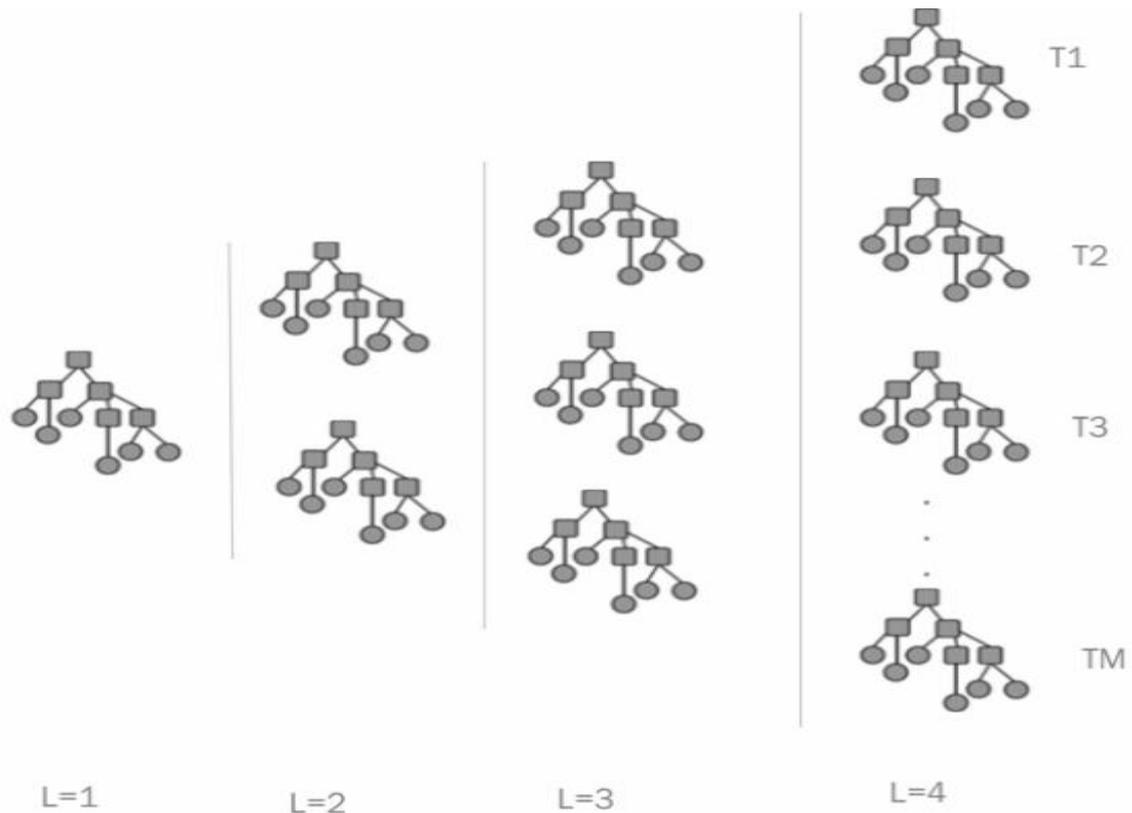


Figure IV.5: Structure du MDT [Maazouzi et Bahi, 2012a]

Dans la phase de construction du modèle le nombre de couche dépend toujours de la taille de base de données, si la taille de l'arbre est grande le nombre de couches est grand et vice-versa.

Remarque : dans notre étude nous avons appliqué l'algorithme des arbres de décisions (C4.5) pour construire chaque arbre dans le modèle.

IV.5.1.1. Exemple illustratif d'apprentissage à l'aide de MDT

Dans cette section, nous appliquons la méthode d'apprentissage proposée (MDT) avec l'ensemble de données IRIS (MLR, 2012), qui est une référence dans le domaine de fouille de données. La base de données contient 150 échantillons, tous les échantillons sont utilisés pour l'apprentissage. Pour cela, nous avons choisi $n=2$ (deux couches).

La deuxième étape consiste à calculer C_iSB_j

$$C_1SB_1 = \frac{150}{1} = 150$$

$$C_2SB_1 = \frac{150}{2} = 75$$

$$C_2SB_2 = \frac{150}{2} = 75$$

Enfin, nous avons appliqué l'algorithme C4.5 pour chaque ensemble de données C_iSB_j . La figure 4 montre le modèle obtenu après la phase d'apprentissage de MDT.

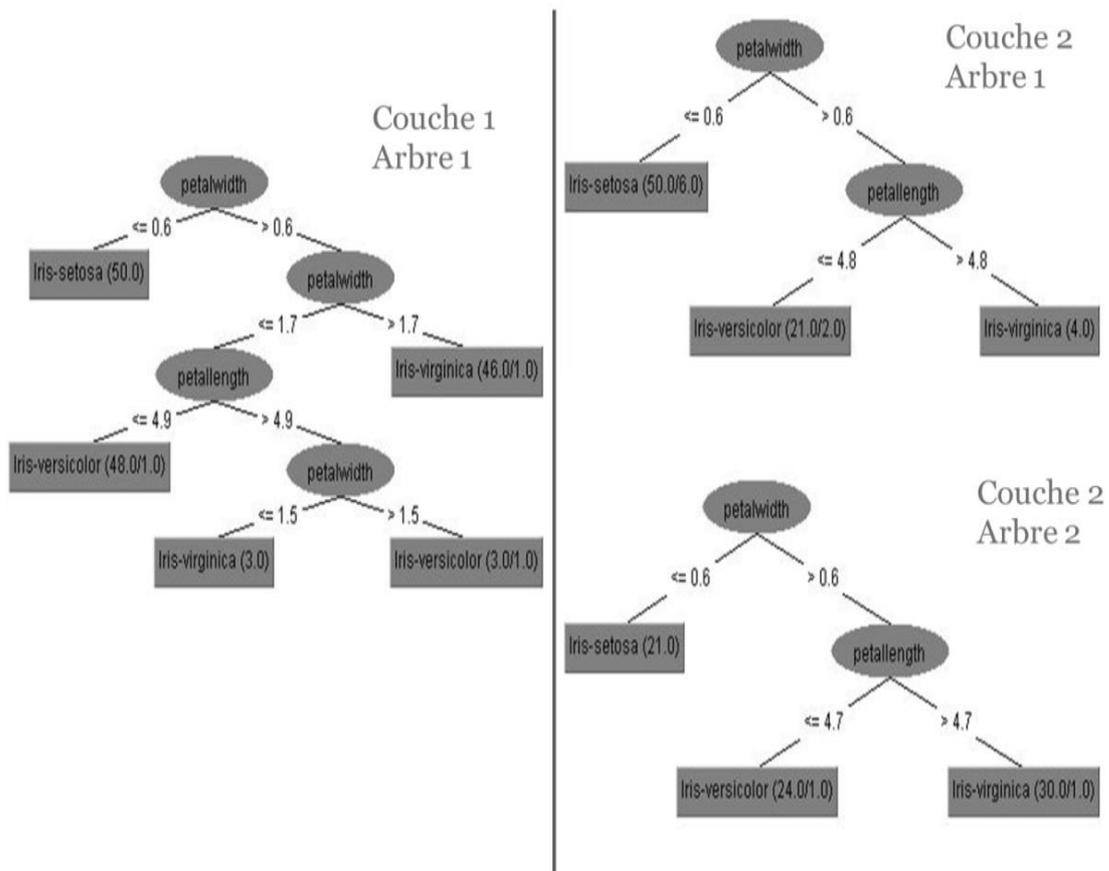


Figure IV.6: Illustration Du MDT - la base de données iris [Maazouzi et Bahi, 2012a]

Selon la figure IV.6, la taille d'un arbre dans la couche n est plus grande que la taille de celles de la couche n + 1 (couche n > n+1). Mais cela ne fonctionne pas toujours.

A titre d'illustration, les informations sur les modèles construits à partir des ensembles de données: iris, Pima, glass, wine and ionospheres [MLR, 2012] sont présentés dans le tableau IV.1

Base	Nombre des échantillons	Nombre de couches	la taille SB de chaque couche	Nombre d'arbres	Nombres de classes
Iris	150	2	C1SB1 :150 C2SB1 :75 C2SB2 :75	3	3
Pima	306	4	C1SB1 :306 C2SB1 :153 C2SB2 :153 C3SB1 :102 C3SB2 :102 C3SB3 :102 C4SB1 :76 C4SB2 :76 C4SB3 :77 C4SB4 :77	10	2
Glass	214	3	C1SB1 :214 C2SB1 :107 C2SB2 :107 C3SB1 :71 C3SB2 :71 C3SB3 :72	6	7
wine	178	2	C1SB1 :178 C2SB1 :89 C2SB2 :79	3	3
Iono- sphere	351	4	C1SB1 :351 C2SB1 :175 C2SB2 :176 C3SB1 :117 C3SB2 :117 C3SB3 :117 C4SB1 :87 C4SB2 :88 C4SB3 :88 C4SB4 :88	10	2

Tableau IV.1: Informations sur les modèles construits à partir des bases (iris, Pima, glass, wine and ionospheres) [Maazouzi et Bahi, 2012a].

IV.5.2. La phase de décision

Dans cette section, nous présentons le processus de décision dans la méthode MDT. Soit Pr un problème de M classes tel que CM représente la classe M [Maazouzi et Bahi, 2012a].

Le but du processus de décision est de calculer le pourcentage P_C qu'une donnée de test appartient à la classe C_M .

La première étape consiste à calculer x , qui est une variable utilisée pour calculer le poids (P_i) d'une couche de l'équation (IV.9).

$$x^{-1} = \sum_{k=1}^n \frac{1}{k} \quad (\text{IV.9})$$

n représente le nombre de couche dans le modèle.

La deuxième étape du processus de décision consiste à calculer le poids (P_i) pour chaque couche

$$P_i = \frac{1}{i} * x \quad (\text{IV.10})$$

Tel que :

$$\sum_{k=1}^n P_i = 1 \text{ et } P_i > P_{i-1} \quad (\text{IV.11})$$

Dans l'étape suivante nous avons calculé la probabilité (PC_i) de la classe C_M pour chaque couche i ($i = 1 \dots n$)

$$PC_i = \frac{\sum_{j=1}^i PT_{i,j}}{i} \quad (\text{IV.12})$$

Où

$$PT_{i,j} = \begin{cases} 1 & \text{si le résultat de classification} = C_M \\ 0 & \text{si le résultat de classification} \neq C_M \end{cases}$$

La dernière étape est de calculer P_C tel que :

$$P_C = \frac{\sum_{i=1}^n PC_i * P_i}{n} * 100 \quad (\text{IV.13})$$

La figure suivante montre la position des variables utilisées dans le processus de décision.

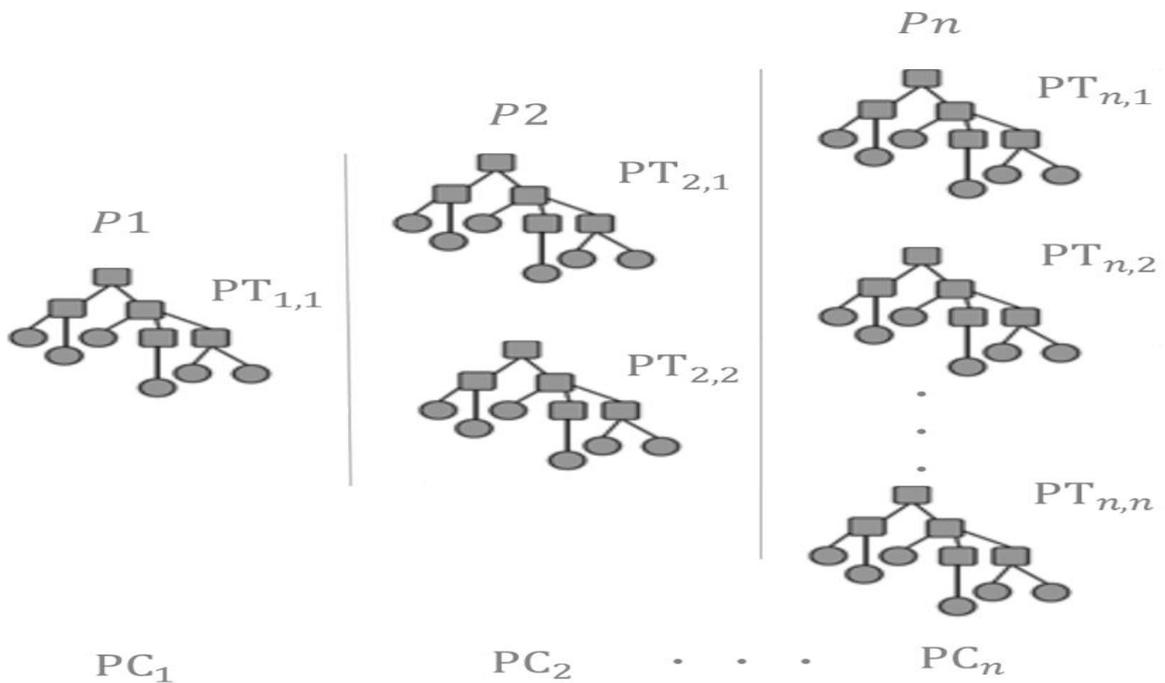


Figure IV.7: les variable ($P_n, PT_{n,n}, PC_n$) dans MDT[Maazouzi et Bahi, 2012a]

Finalement, nous avons obtenu un pourcentage pour chaque classe. Ainsi, pour donner un exemple à classer, il appartient à la classe avec le pourcentage le plus élevé.

Résultat de la classification = $\max (P_{C1}, P_{C2} \dots P_{Cm})$.

Dans le cas de $P_C = P_{C+1}$ le résultat de la classification est le max de PCi .

IV.5.2.1. Exemple illustratif du processus de décision de la méthode MDT

Dans cette section, nous avons utilisé le modèle de base de données IRIS obtenu dans la section IV.5.1.1 (voir figure IV.6). La première étape de la phase de reconnaissance MDT consiste à calculer le poids (P_i) pour chaque couche.

$P1 = 2 / 3$ et $P2 = 1 / 3$

Les classes définies dans la base de données iris sont (Iris virginica-Iris-versicolour, et Iris-setosa). Nous avons essayé de classer l'exemple montré dans le tableau IV.2.

Statue	Sepal length	Sepal width	Petal length	Petal width	Classe
Test	7.2	3.0	5.8	1.6	Iris-virginica

Tableau IV.2: Exemple d'échantillons de base de données IRIS utilisé pour le test [Maazouzi et Bahi, 2012a]

Dans la prochaine étape, nous calculons les probabilités ($P_n, PT_{n,n}, PC_n$) de chaque classe dans notre modèle. Les résultats sont présentés dans le Tableau IV.3.

Class	Probabilité d'arbre			Probabilité de couche		Probabilité de classe
	$PT_{1,1}$	$PT_{2,1}$	$PT_{2,2}$	PC_1	PC_2	P_c
Iris-virginica	0	1	1	0	2	67%
Iris-versicolor	1	0	0	1	0	67%
Iris-setosa	0	0	0	0	0	0

Tableau IV.3: Résultats de la classification [Maazouzi et Bahi, 2012a]

A partir des résultats affichés dans le tableau, la probabilité d'Iris-virginica est égale à la probabilité d'Iris-versicolor (67%). Dans ce cas, nous devons trouver le maximum (PC_1 , PC_2) pour déterminer le résultat de la classification. Le résultat de la classification de l'exemple montré dans le tableau x est Iris virginica-car $PC_1 < PC_2$.

Dans la prochaine section, la méthode décrite ci-dessus (MDT) a été appliquée à certains ensembles de données prises à partir du Machine Learning Repository [MLR, 2012] afin de comparer les performances du système avec d'autres générateurs connus d'arbre de décision.

IV.6. Expérimentations et résultats

Dans cette expérience, nous présentons une étude comparative de la technique de classification MDT avec C4.5 algorithme d'arbre de décision, l'algorithme ADT (AdaBoost Decision Tree), BDT (Bagging Decision Tree) et de RFDT (Random Forest Decision Tree). Nous avons testé notre méthode avec des bases de données : iris, Pima, glass, wine and Ionosphere (Voir Tableau IV.4.).

Chaque ensemble de données a été répartis au hasard en deux groupes, l'ensemble d'apprentissage, qui comprend 100% des données d'apprentissage et les données de test, qui comprend 10 fois- validation croisée [Maazouzi et Bahi, 2012a].

Base de données	Nombre d'attributs	Taille de l'ensemble d'apprentissage	Taille de l'ensemble de test
Iris	4	150	VC-10
Pima	8	768	VC-10
Glass	10	214	VC-10
Wine	13	178	VC-10
Ionosphere	34	351	VC-10

Tableau IV.4: Les ensembles de données utilisés, le nombre d'attributs, et d'apprentissage / test (taille VC-10 représente 10 fois la validation croisée a été utilisée) [Maazouzi et Bahi, 2012a]

Nous comparons les résultats du système MDT avec les résultats des algorithmes C4.5 (mis en œuvre dans WEKA que J48 [Weka ; 2012]) et le résultat de ADT, BDT et RFDT, pour ce faire, nous calculons la précision et le taux d'erreur des deux techniques. Les résultats sont présentés dans la Figure IV.8 et Figure IV.9.

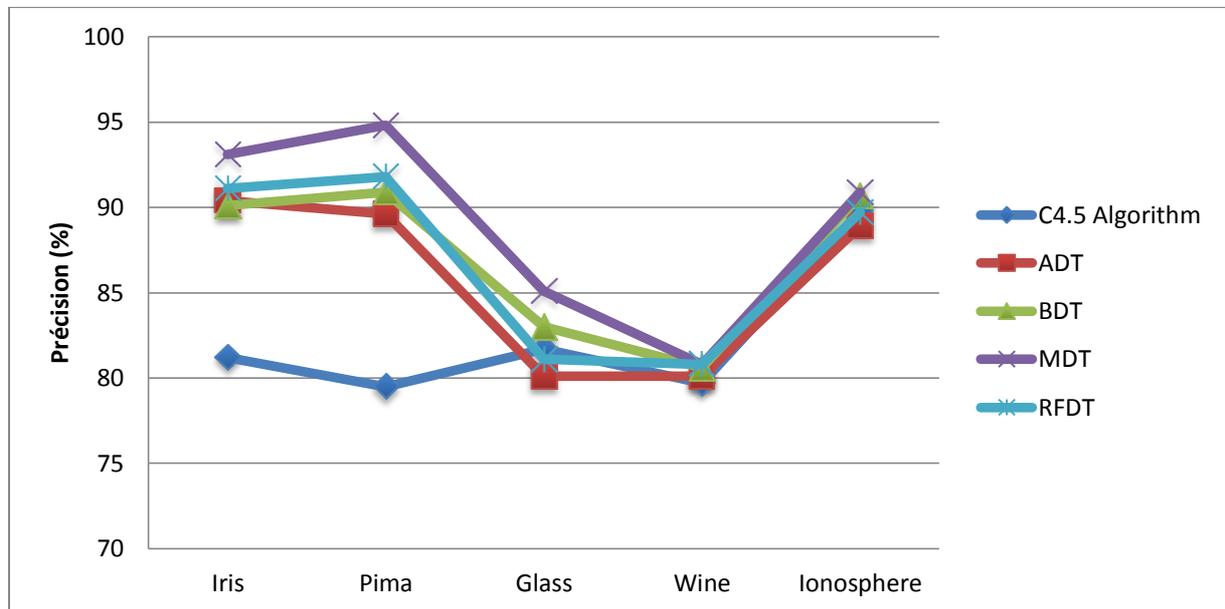


Figure IV.8: Comparaison de la précision de C4.5, ADT, BDT, MDT et RFDT pour les bases de données 'iris, pima, glass, wine, ionosphere' [Maazouzi et Bahi, 2012a].

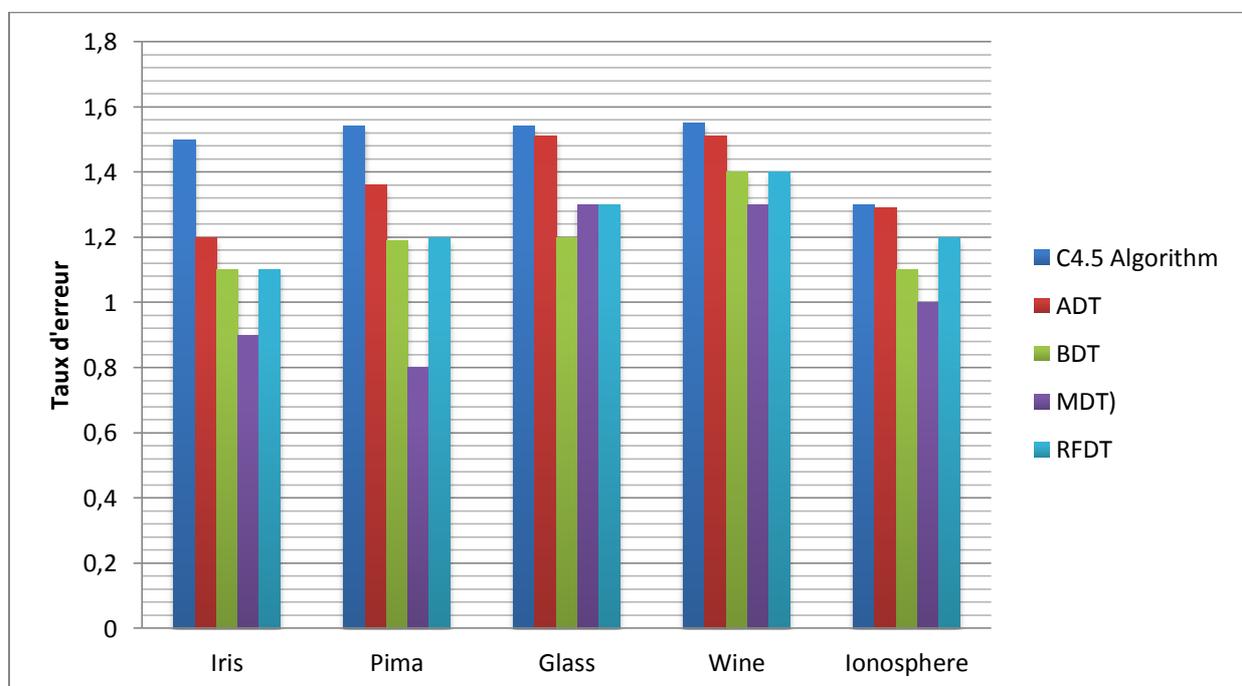


Figure IV.9: Comparaison des taux d'erreurs de C4.5, ADT, BDT, MDT et RFDT pour les bases de données 'iris, pima, glass, wine, ionosphere' [Maazouzi et Bahi, 2012a].

- La précision de la MDT est élevée (Fig. IV.8) pour la plupart des bases de données utilisées pour les tests, ce qui est hautement nécessaire.
- La précision la plus élevée est obtenue en utilisant MDT avec Pima base de données (94,8%).
- On peut observer à partir de la figure IV.9 que le taux d'erreur de la MDT est la plus basse comparé avec C4.5 algorithme d'arbre de décision, ADT, BDT et RFDT, ce qui est le plus souhaitable.
- Le taux d'erreur minimum est obtenu en utilisant MDT avec Pima base de données (0,8).

Selon les expérimentations et l'analyse des résultats présentés dans ce document, on constate que la technique MDT est préférable de classer les ensembles de données avec une meilleure précision. Cela peut s'expliquer par le fait que l'agrégation des classifieurs est de meilleurs résultats que d'un seul classifieur. D'autre part, le fait de partager les exemples d'apprentissage, implique que, pour un exemple donné, il va rencontrer des arbres de plus en plus précis que nous avançons en couches.

IV.6. Conclusion

Dans les dernières années plusieurs techniques de fouille de données sont proposées pour améliorer les systèmes de classification et de prédiction. Les arbres de décision sont souvent utilisés dans le domaine de l'exploration de données, dans ce papier, nous présentons une technique de fouille de données basé sur des arbres de décision appelé: Multi Decision Tree (MDT). Dans la méthode MDT, les résultats de la classification de chaque instance sont en pourcentage pour chaque classe.

Nous avons comparé notre approche avec d'autres approches basées sur les arbres de décision (C4.5 algorithme d'arbre de décision, ADT, BDT et RFDT), les résultats montrent qu'il est possible d'améliorer le système de classification en utilisant la méthode MDT. L'induction de l'approche MDT, qui construit des arbres de décision différentes pour différents résultats de classification, peut réduire la cardinalité de règle, et d'améliorer la concision et la précision à la fois de la règle de classification sur un inducteur ou un seul arbre de décision traditionnel.

***Application de la Fouille de données sur
un site de chansons***

1. Introduction

2. L'établissement des données à collecter

3. La sélection des caractéristiques et le nombre de mélange de gaussiennes pour la classification des voix chantées

4. La classification des voix chantées

5. La fouille de données

6. Conclusion

V.1. Introduction

Le domaine de la fouille de donnée audio est très riche et multidisciplinaire. Dans ce contexte, nous nous proposons d'établir une enquête sociale dont le but est de prospecter les goûts artistiques des auditeurs algériens. Pour cela, nous considérons un site de musique visité par les auditeurs algériens, et nous essayons de « fouiller » les données dont il regorge pour voir si vraiment l'auditeur algérien accorde de l'importance aux capacités vocales du chanteur qu'il écoute le plus, pour cela plusieurs étapes doivent être dégagées et bien étudiées.

Le système de fouille de données est présenté dans la figure V.1

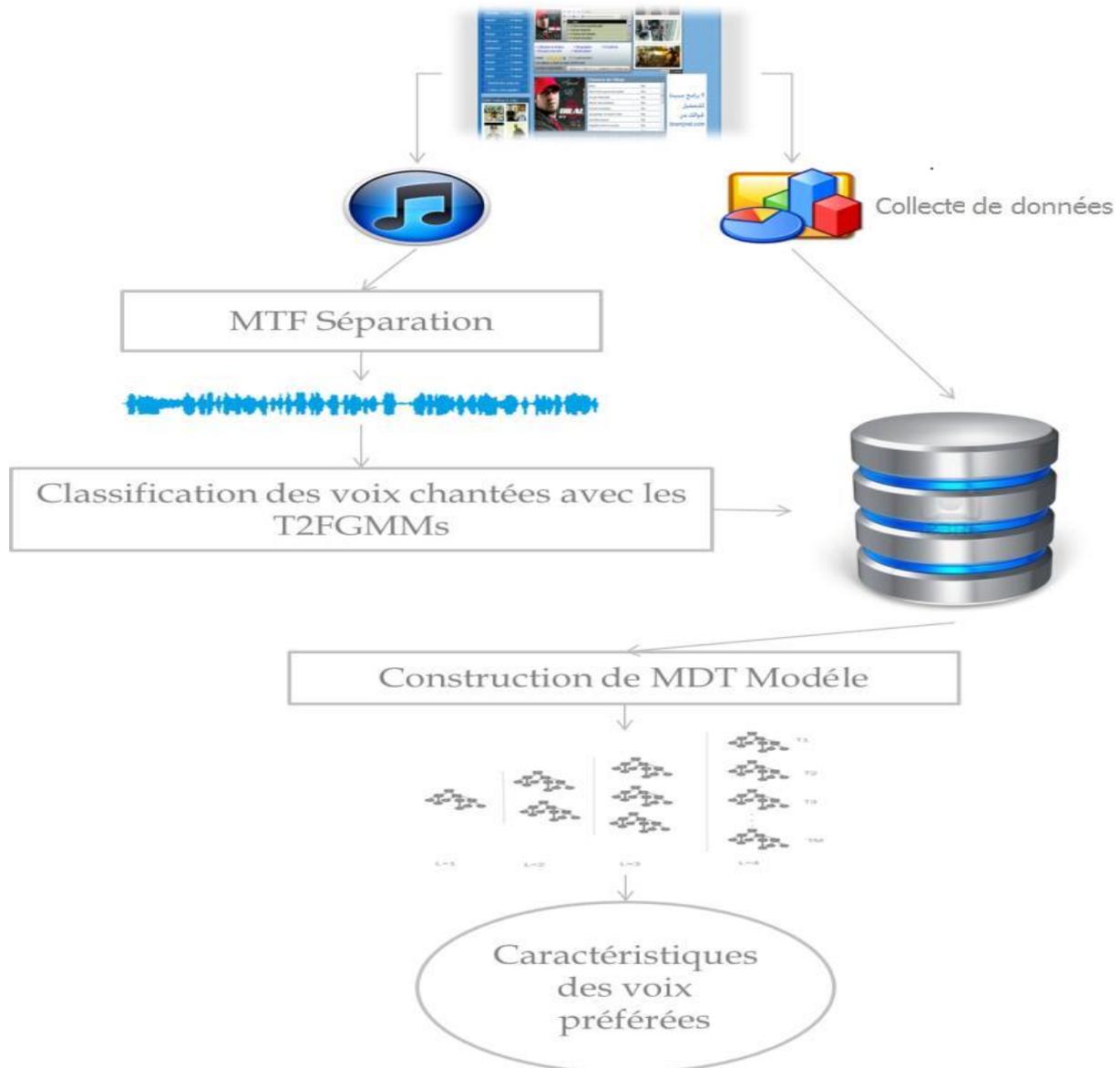


Figure V.1: Système proposé pour la fouille de données audio

V.2. L'établissement des données à collecter

Aujourd'hui la plupart des gens utilisent les sites internet pour écouter la musique, à partir d'un site qui contient des milliers des musiques algériennes dans tous les genres (rai, staifi, chabi,...ect). Nous considérons le site [zikdalgire, 2012] qui est un site de musique algérienne qui contient des milliers de chansons, ces chansons sont regroupées par albums, chaque album représente un chanteur. Ce site offre plusieurs statistiques ou données importantes pour notre étude.

V.2.1. Les données offertes par le site

Il y a deux types des données pertinentes que le site offre :

Type 1 : Le premier type c'est les fichiers musicaux, ou les fichiers audio (musicaux) sont des fichiers de type mp3 converti au type wav, ces fichiers représentent le corpus utilisé dans la prochaine section (la classification des voix chantées).

Dans cette étude, nous considérons dans une partie, la base de données qui comprend les voix de 140 chanteurs algériens extraites du site web [zikdalgire, 2012]. L'ensemble de données musicales (DM) contient 1600 échantillons mélanges (voix + musique). Tous les échantillons dans notre base de données sont au format WAV.

- Canal: 2 (stéréo)
- Fréquence d'échantillonnage: 44.1 kHz
- Débit: 128 kbps
- Duration: 15 secs

Genre	Type	Nombres des échantillons	Qualité	Nombres des échantillons
Mâle	Soprano	220 échantillons	Professionnel	290 échantillons
	Mezzo-Soprano	150 échantillons	Semi-Professionnel	220 échantillons
	Contralto	240 échantillons	Non- Professionnel	100 échantillons
Femelle	Ténor	230 échantillons	Professionnel	270 échantillons
	Barytone	190 échantillons	Semi- Professionnel	200 échantillons
	Basse	170 échantillons	Non- Professionnel	130 échantillons

Tableau V.1: Les informations de la base de données musicale

Type 2 : Le deuxième type c'est les statistiques sur les chanteurs tel que : les données concernant les albums et les données concernant les chanteurs. Pour notre étude nous utilisons les données concernant les chanteurs tel que : le nom de chanteur, le genre, le nombre de

visites du chanteurs, le classement de chaque chanteur, et quelle sont les chanteurs et les albums les plus visités dans une journée, un mois ou une semaine ? Voir la Figure V.2.



Figure V.2: Les chanteurs et les albums les plus visités

V.3. La classification des voix chantées

Le but de cette étape (ou sous système) est de faire une classification des chanteurs selon leur capacité vocale. Dans ce contexte, nous avons traité deux catégories de classification: La première porte sur les compétences d'un chanteur (de qualité), et le second avec le genre de la voix chantée (Type).

Le tableau suivant présente les différentes classes existantes en fonction de l'état de l'art.

Etudes	Catégories	Classes
[Maazouzi et Bahi, 2012 ; Zwan, 2008]	Type	Soprano, Mezzo-Soprano, Alto, Ténor, Barytone, Bass
	Qualité	Amateur, Semi-Professionnel, Professionnelle
[Nakano et al., 2006]	Qualité	Bonne, Mauvais
[Prasert et al., 2008]	Qualité	Bonne, pauvres
[Nakano et al., 2006 ; Maazouzi et Bah, 2012b]	Type	Soprano, Mezzo-Soprano, Alto, Ténor, Barytone, Basse
	Qualité	Bonne, Moyen, Mauvais

Tableau V.2: les différentes classes des voix chantées existantes en fonction de l'état de l'art

Avant de faire une classification des voix des chanteurs algériens une étape obligatoire consiste à extraire les voix des chanteurs à partir des enregistrements musicaux.

V.3.1. La séparation de voix chantées / musique

Dans cette étape nous nous intéressons à l'extraction de voix chantée dans des enregistrements musicaux, c'est-à-dire à la séparation de la voix par rapport à l'accompagnement musical. Le problème de l'extraction de voix chantée dans des enregistrements musicaux monophoniques, c'est-à-dire la séparation voix / musique avec un seul capteur, est étudié. Les approches utilisées sont basées sur les modèles statistiques a priori de deux sources. Une étude comparative des différents modèles et estimateurs est effectuée [Ozerov et al., 2007], ainsi qu'une étude d'impact de l'hétérogénéité entre les données d'apprentissage et les données à séparer.

Il y a plusieurs projet de séparation de voix/musique, ces méthodes reposent toutes sur des approches dites "aveugle" consistant à extraire du signal des descripteurs audio "génériques" utilisés ensuite pour apprendre les deux classes de segments "chanté" et "non chanté" à l'aide d'algorithmes statistiques, parmi les projet qui se sont orienter dans ce contexte, [Li et Wang, 2006] utilisent également comme descripteur les MFCC avec l'algorithme de classification MMG .

Pour le cas des mélanges stéréo (notre cas), nous avons proposé d'utiliser la méthode masquage tems fréquence [Vinyes et al., 2006] (pour plus de détaille voir la section II.4.1).

Les résultats de cette technique prouvent une mise en œuvre et peut être utilisée aujourd'hui pour demixer les pistes des chansons commerciales.

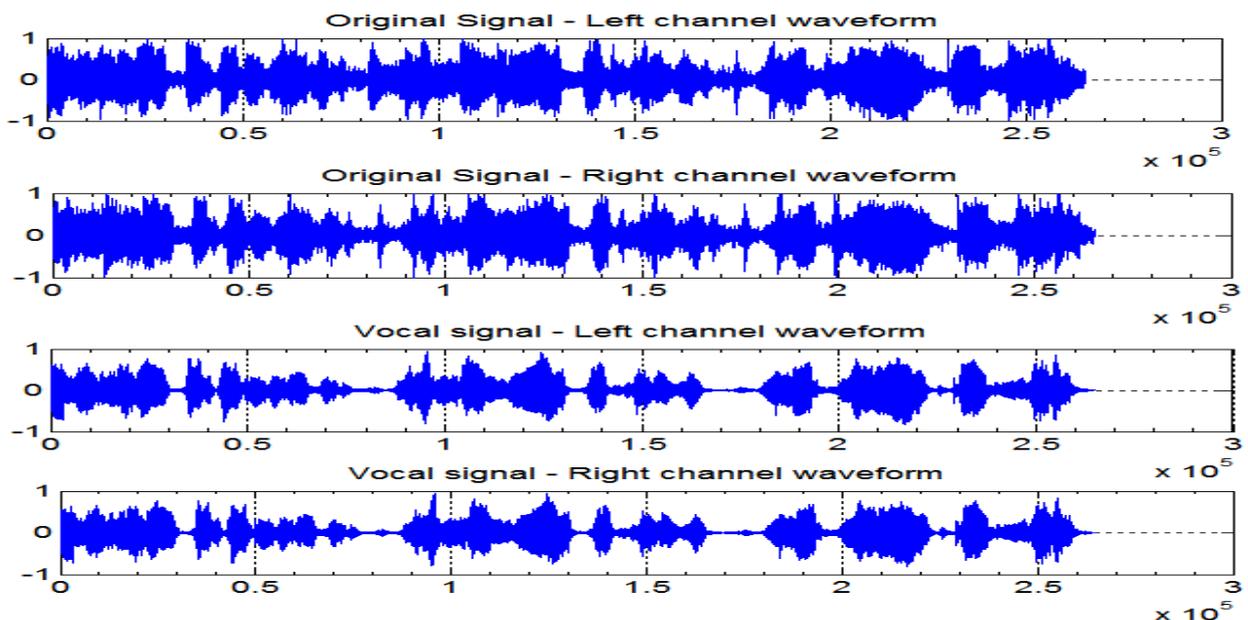


Figure V.3: Résultats de séparation des voix chantées/ musique [Maazouzi et Bahi, 2013]

V.3.2. Utilisation de T2 FGMMs pour la classification des voix chantées

La conduite de ce travail induit un certain nombre d'étapes, en particulier celles relatives à la caractérisation des chanteurs. Il s'agit alors de classer les chanteurs selon leur registre mais aussi d'évaluer leur qualité de la voix. Comme tous les systèmes de classification supervisée, notre système de classification se fait dans deux étapes: l'extraction des caractéristiques, la modélisation et la classification.

V.3.2.1. L'extraction des caractéristiques

deux vecteurs de caractéristiques sont utilisés dans cette étape le premier vecteur pour la classification des voix en type (*VCT*) et le deuxième vecteur pour l'évaluation de la qualité vocale des chanteurs (*VCQ*).

VCT_{1-169} : [VCM_{1-121} , $VCNM_{123-169}$]

VCQ_{1-160} : [VCM_{1-124} , $VCNM_{125-160}$]

Les deux vecteurs sont obtenus par une phase de sélection (voir section III.3.1).

V.3.2.2. La modélisation et la classification avec T2FGMMs

Après les résultats obtenus dans le tableau II. Nous avons proposé d'utiliser la méthode T2FGMMs méthode proposée par [Zeng et al. 2008] (voir section III.4.3) pour la classification des voix chantées. Pour chaque catégorie de classification nous avons utilisé un modèle différent :

- 1) la classification des voix en type nous avons utilisé Type 2 Fuzzy Gaussian Mixture Models Uncertain Mean vector (T2 FGMM-Ums) avec $k_m=1,5$
- 2) l'évaluation de qualité vocale nous avons utilisé Type 2 Fuzzy Gaussian Mixture Models Uncertain Mean vector Uncertain Variance (T2 FGMM-UVs) avec $k_v=0,7$.

Enfin, la phase de classification se fait en utilisant le critère de maximum à posteriori.

V.3.3. Construction de la base de données

Cette étape représente une fusion de base de données, c'est à dire, nous avons créé une base de données qui contient les statistiques sur chaque chanteur tel que : le nom de chanteur, le classement de chanteur par rapport de top 5, le genre, le genre de musique et la qualité et le type vocale de chaque chanteur et chanteuse. Nous avons utilisé une base de 140 chanteurs et chanteuses.

V.4. La fouille de données et résultats

La fouille de données audio est généralement utilisée dans les logiciels de reconnaissance vocale et l'analyse musicale. Cette technologie offre à l'utilisateur la possibilité de rechercher par la parole ou audio de musique qui a été analysée pour des caractéristiques spécifiques.

Pour notre système, à partir d'une base d'apprentissage qui contient des statistiques sur 140 chanteurs avec la qualité vocale de chaque chanteur, nous avons appliqué l'algorithme MDT (Multi Decision Tree) proposé dans [Maazouzi et Bahi, 2012a] pour déterminer les goûts du public algérien visitant de site de musique. La classification se baser sur les chanteurs du top 5, il y a 3 classe possibles pour chaque chanteur : « 1 -Très préférer, 2-Préférer, 3-Non préférer ». Voir Figure V.4.

La base de données contient 140 échantillons, tous les échantillons sont utilisés pour l'apprentissage. Pour cela, nous avons choisi $n=2$ (deux couches).

Ensuite, on va calculer C_iSB_j

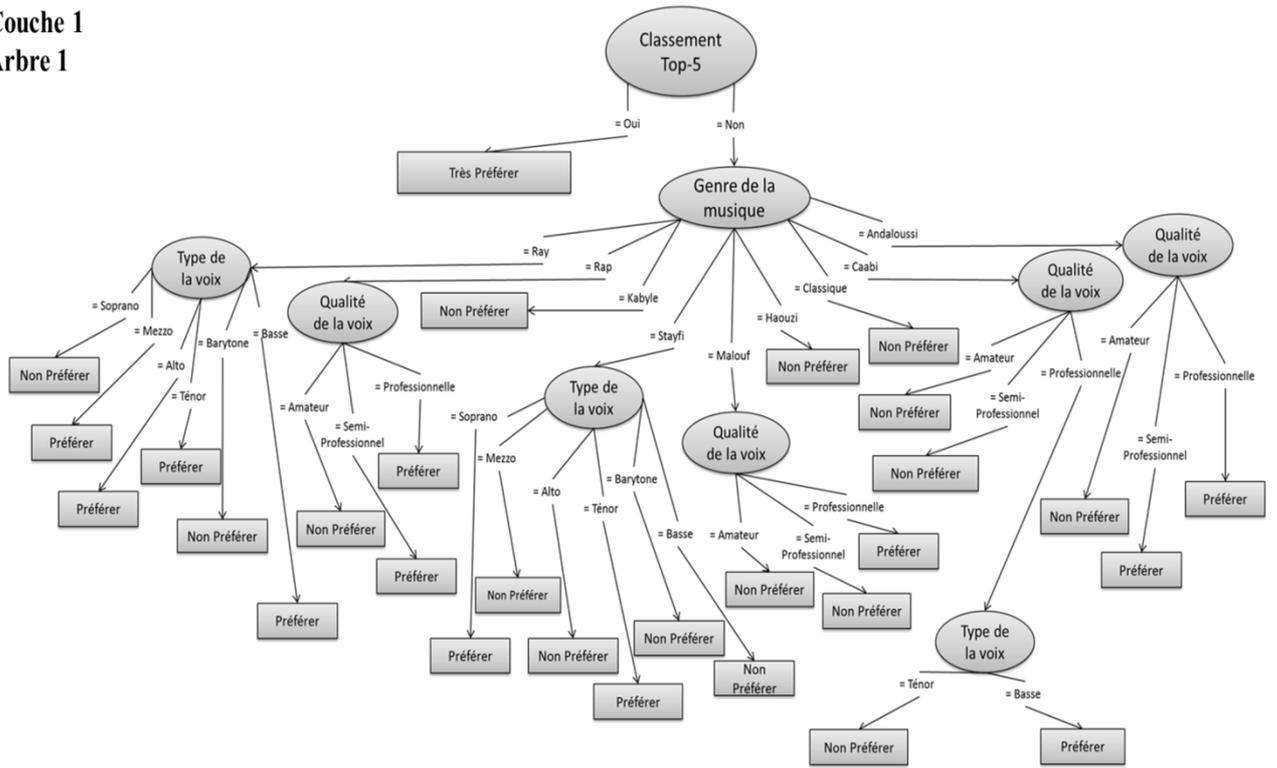
$$C_1SB_1 = \frac{140}{1} = 140$$

$$C_2SB_1 = \frac{140}{2} = 70$$

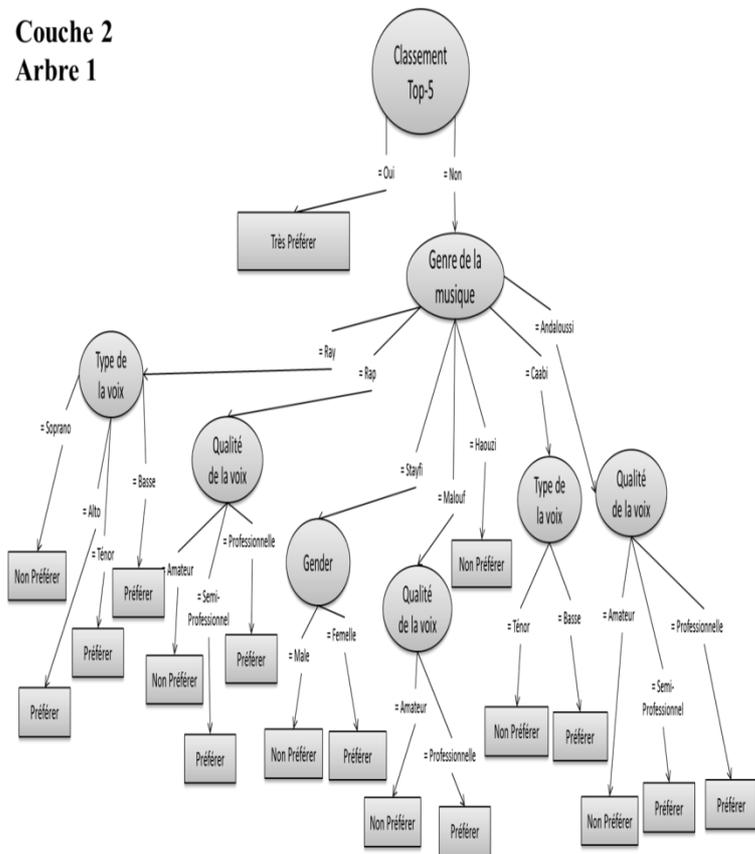
$$C_2SB_2 = \frac{140}{2} = 70$$

La dernière étape consiste à appliquer l'algorithme C4.5 pour chaque ensemble de données C_iSB_j . La figure montre le modèle obtenu après la phase d'apprentissage de MDT.

Couche 1
Arbre 1



Couche 2
Arbre 1



Couche 2
Arbre 2

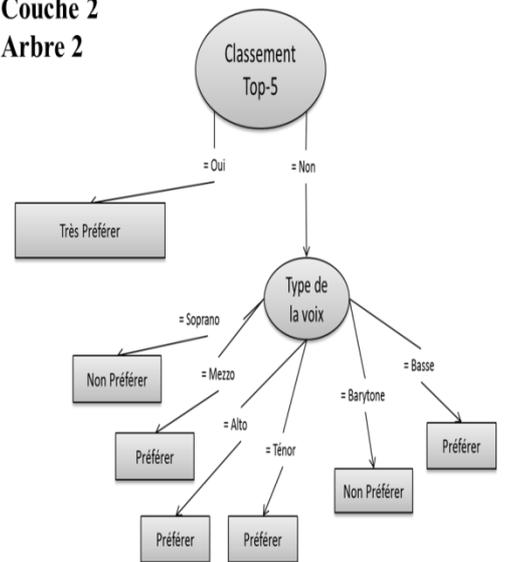


Figure V.4 : Modèle obtenu après la modélisation avec MDT

A partir de l'arbre construit, nous pouvons faire les déductions suivantes concernant les préférences des auditeurs algériens :

- Automatiquement les chanteurs du top 5 sont classés dans la classe « Très préféré », et on remarque que le genre de tous les chanteurs du top 5 est le Rai, mais il y a une chanteuse parmi les tops 5 qui a la qualité vocal alto, la classification des autres chanteurs se fait sur la base de leurs genres.
- Les chanteurs de rap qui ont la qualité de la voix « professionnelle » sont des chanteurs « préféré » et les chanteurs qui ont de qualité de la voix « Amateur » sont des chanteurs « non préféré », on remarque que les chanteuse de genre de musique staifi sont préférés, mais les chanteurs ne sont pas préférés pour les auditeurs.
- La plupart des auditeurs algériens écoutent le rai. Il y a des chanteurs qui ont le de la voix ténor mais ils ne sont « pas préféré » pour le public, malgré qu'il existe des chanteurs avec une qualité vocale basse préférés pour le public.
- Les chanteurs du Chaabbi qui ont qualité de voix « professionnelle » et type de voix « bass » sont des chanteurs préférés pour le public algérien. Il existe aussi des genres de musique qui ont pas préférés pour le publique tel que : Kabyle, Haouzi et Classique.
- Finalement, on peut dire qu'il n'y a pas de relation entre la qualité, le type de la voix des chanteurs algériens et le gout de publique algérien par ce que il existe des chanteurs qui ont de faible qualité vocal mais ils sont très préférés pour le public.



Conclusion générale

Conclusion et perspectives

1. Bilan

Le travail que nous avons mené dans cette thèse entre dans le cadre des travaux de fouille de données audio. Le but est d'effectuer une étude qui nous permettra de sortir les caractéristiques des voix préférées pour le public algérien qui visitent un site de musique. Notre contribution dans cette thèse revêt deux aspects (classification des voix chantées et fouille de données), il s'agit d'abord d'un processus de classification des voix chantées. Dans ce cadre, l'étude bibliographique que nous avons effectuée, nous a permis de constater que le processus de classification des voix chantées se fait en trois étapes :

- L'extraction des caractéristiques
- La sélection des caractéristiques
- La modélisation et la classification

La principale difficulté que nous soulignons par rapport aux autres travaux dans ce domaine, et que nous avons utilisé des fichiers audio mix (musique et voix chantées). De ce fait, avant de faire une classification des voix chantées, nous avons séparé les deux sources « musique et voix chantées », pour extraire les voix des chanteurs. Pour prendre en charge cette particularité dans notre étude, nous avons utilisé la méthode de Masquage Temps Fréquence (MTF) pour séparer la voix du chanteur et la musique.

L'objectif de cette classification est de classer les chanteurs sur la base de leurs qualités vocales en trois classes « Professionnel, Semi-Professionnel et Non- Professionnel » pour chaque genre et selon leurs types des voix (voir sec II.6.1.1.). Dans cette phase (Classification des voix chantées) nous avons construits quatre systèmes, dans chaque système nous avons utilisé un vecteur de caractéristique différent et une méthode d'apprentissage automatique différente (méthode de modélisation et de classification).

Dans le premier système, nous avons utilisé un vecteur MFCC de 12 paramètres avec les GMMs pour la modélisation et la classification [Maazouzi et Bahi, 2011]. Ce système a donné de bons résultats dans le domaine de reconnaissance de la parole.

Pour le deuxième système nous avons utilisé un vecteur de caractéristiques de 130 paramètres obtenus après une phase de sélection des caractéristiques. Ce vecteur contient les descripteurs de la norme MPEG-7 Standard [Maazouzi et Bahi, 2011]. La technique de modélisation et de classification utilisée dans ce système est les GMMs. Après une comparaison des résultats du premier système et du deuxième système on a conclu que l'utilisation des descripteurs MPEG-7 a donné les meilleurs résultats pour les deux catégories de classification des voix chantées.

Ensuite, pour améliorer les résultats de la classification, nous avons proposé un algorithme de sélection des caractéristiques et de nombre de nombre de gaussienne [Maazouzi et Bahi, 2012].

Dans ce système deux groupes de descripteurs sont proposés, le premier groupe contient les Non- MPEG-7 Descripteurs. Ce sont les descripteurs les plus utilisés dans le domaine de musique et qui n'existent pas dans la norme MPEG-7 Standard tel que : Le vibrato, MFCC et WLPC. Le deuxième groupe des descripteurs contient les descripteurs MPEG-7 LLDs [Maazouzi et Bahi, 2012]. L'algorithme de sélection des caractéristiques proposé dans ce système est basé sur la méthode SBS. Nous avons appliqué cet algorithme sur le vecteur qui contient les Non-MPEG-7 descripteurs et le vecteur qui contient les MPEG-7 LLDs descripteurs. Finalement, Les résultats montrent des améliorations substantielles par rapport aux deux premiers systèmes.

Les deux groupes de descripteurs utilisés dans le système précédent sont regroupé pour construire un vecteur de caractéristiques qui contient 262 paramètres dans ce système [Maazouzi et Bahi, 2013]. La nouvelle idée dans ce système consiste à utiliser la technique T2 FGMMs pour la modélisation et la classification, l'étude de [Zeng et al., 2008] a prouvé que cette technique est meilleure que les GMMs dans le domaine de reconnaissance des formes. Donc dans le quatrième système nous avons utilisé deux vecteurs de caractéristiques (un pour la classification des voix chantées en types et le deuxième pour l'évaluation de la qualité des voix chantées) avec la technique T2 FGMMs [Maazouzi et Bahi, 2013]. La comparaison des résultats des quatre systèmes proposés avec d'autres systèmes de classification des voix chantées, montre que le quatrième système est le meilleur système en termes de performance. C'est celui que nous allons utiliser dans le système de fouille de données audio.

Le deuxième aspect dans notre contribution rentre dans le cadre de la fouille de données, qui est un domaine de recherche très large et plusieurs techniques de fouille de données sont proposées dans les dernières années. Après une analyse bibliographique du domaine FDD (Fouille de données), on peut distinguer deux types de techniques de fouille de données : les algorithmes d'apprentissage automatique et les ensembles des classifieurs. Dans le cadre de ce dernier type, nous avons proposé une nouvelle technique de fouille de données MDT (Multi Decision Tree) [Maazouzi et Bahi, 2012a]. Cette méthode se base sur les arbres de décision et s'inspire des réseaux des Réseaux de Neurones Artificiel (RNA). Où le modèle contient des arbres de décisions combinées sous la forme d'un réseau de plusieurs couches. Ensuite, cette méthode est utilisée pour explorer notre nouvelle base de données. L'analyse de modèle obtenu (voir Fig. V.4) montre qu'il n'existe pas une relation entre la qualité, le type des voix et le goût du public algérien.

Globalement notre contribution consiste en :

- La séparation des voix chantées,
- La classification des voix chantées,
- La fouille de données.

2. Perspectives de la classification des voix chantées dans la musique commerciale

Les résultats des systèmes proposés pour la classification des voix chantées sont alors assez bons mais ils ne sont pas satisfaits pour certain raison, parmi les raisons : la séparation de la voix du chanteur et la musique. Une perspective du travail, serait donc de tester d'autres méthodes ou de proposer une nouvelle méthode pour la séparation des voix chantées basé sur les descripteurs de vibrato avec les descripteurs de la norme MPEG-7 LLDs.

Les nouveaux travaux des reconnaissances des formes ont prouvé qu'un system basé sur un ensemble de classifieurs, peut donner des résultats meilleurs qu'un seul classifieur. Comme perspective, on propose un système basé sur un ensemble des classifieurs pour la classification des voix chantées. L'ensemble des classifieurs utilisé dans le système seraient les classifieurs existants dans la littérature de la classification des voix chantées. C'est à dire les : RNA, F2 FGMMs et SVM.

3. Perspective en fouille de données

La comparaison de la méthode proposée (MDT) pour la fouille de données avec les autres méthodes montre qu'il existe des cas où notre système ne donne pas des bons résultats. Donc nous avons réfléchi à faire des modifications au cours de la phase de modélisation tels que : le calcul de nombres des couches et l'utilisation des différents algorithmes pour la construction des arbres de décision.



Bibliographie

Bibliographie

- [**Anyanwu et Shiva, 2009**] M.N. Anyanwu, et S.Shiva, Comparative analysis of serial decision tree classification algorithms', International Journal of Computer Science and Security, Vol. 3, No. 3, pp.154-271, 2009.
- [**Baf et al., 2008**] E.F. Baf, T. Bouwmans, et B. Vachon, Type-2 fuzzy mixture of Gaussians model: application to background modeling, Proceedings of the International Symposium of visual computer, Las Vegas, USA, pp.72-781, 2008.
- [**Berenzweig et al., 2002**] A. Berenzweig, D.P.W. Ellis, et S. Lawrence, Using voice segments to improve artist classification of music, AES 22nd International Conference, 2002.
- [**Boldrey, 1994**] R. Boldrey, Guide des rôles et des arias d'opéra. Caldwell Publishing Company. ISBN 978-1877761645.
- [**Bougoussa et Wang, 2011**] M. Bougoussa et S. Wang. Apprentissage par ensemble, 2011.
- [**Bouwmans et Baf, 2009**] T. Bouwmans, et E.F. Baf, Modeling of dynamic backgrounds by Type-2 fuzzy gaussian mixture models, MASAUM Journal of Basics and Applied Sciences, Vol. 1, pp.265-277, 2009.
- [**Boser et al., 1992**] B. E., Boser, I. M., Guyon et V. N.Vapnik, A training algorithm for optimal margin classifiers, Proceedings of the 5th annual ACM workshop on Computational Learning Theory, ACM Press, pp. 144-152, 1992.
- [**Breiman et al., 1984**] L. Breiman, J. Friedman, L. Olshen, et J. Stone, Classification and Regression trees. Wadsworth Statistics/Probability Series, CRC press Boca Raton, Florida, USA, 1984.
- [**Breiman, 1994**] L. Breiman, Bagging predictors, Journal Machine Learning, Vol. 24, No. 2, pp.123-140, 1994.
- [**Breiman et Cutler, 2005**] L., Breiman et A.,Cutler, Random Forests, Berkeley. Available from <http://www.stat.berkeley.edu/users/breiman/RandomForests/>, 2005.
- [**Berenzweig et Ellis, 2001**] AL Berenzweig et DPW Ellis. Locating singing voice segments within music signals. IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, pages 119-122, 2001.
- [**Chang, 2009**] P. Chang, Pitch oriented automatic singer identification in pop music, , IEEE International Conference semantic computing ICSC '09, p.161-166, 2009.

- [Chou et Gu, 2001]** W. Chou et L. Gu. Robust singing detection in speech/music discriminator design. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processin, 2001.
- [Cross, 2008]** Richard Cross. Classification vocal. CSV 2008, 2008.
- [Do et al., 2009]** T., Do, S., Lallich, N., Pham et P., Lenca, Un nouvel algorithme de forêts aléatoires d'arbres obliques particulièrement adapté à la classification de données en grandes dimensions, Extraction et gestion des connaissances, pp.79-90, 2009.
- [Dodane, 2008]** C. Dodane, cour de production de la parole, Université Montpellier 3, Laboratoire Dipralang (EA 739), 2008
- [Duda et al., 2001]** R. Duda, P. Hart, et D. Stork. Pattern Classi_cation. Willey-Interscience, 2001.
- [Freund et Schapire., 1995]** Y. Freund et R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, In Second European Conference on Computational Learning Theory, March 1995.
- [Fujihara et la., 2005]** H. Fujihara, T. Kitahara, M. Goto, K. Komatani, T. Ogata, et H.G. Okuno, Singer identification based on accompaniment sound reduction and reliable frame selection. Proc. ISMIR, pages 329-336, 2005.
- [Gal-Or et al, 2010]** M.Gal-Or, J.H. May, et W.E. Spangler, When to choose an ensemble classifier model for data mining, IJBIDM, Vol. 5, No. 3, pp.297-318, 2010.
- [Hunt, 1962]** E.B.Hunt, Concept Learning: An Information Processing Problem, Wiley, 1962.
- [Indyk. et Motwani, 1998]** P. Indyk et R. Motwani. Approximate Nearest Neighbors : Towards Removing the Curse of Dimensionality. In 30th Symposium on Theory of Computing, 1998.
- [Inoshita et Katto, 2009]** T. Inoshita et J. Katto. Key Estimation Using Circle of Fifths. In Proc. of the 15th International Multimedia Modeling Conference (MMM'09), pp.287-297, 2009.
- [Kai et al., 2010]** X. Kai, W. Fangfang, et Q. Kun, An image segmentation method based on Type-2 fuzzy gaussian mixture models, Proceedings of International Conference on Image Analysis and Signal Processing, Zhejiang, Chine, pp.363-366, 2010.
- [Kim et Whitman, 2002]** Y.E. Kim, et B. Whitman, Singer identification in popular music recordings using voice coding features, In Proceedings of the 3rd International Conference on Music Information Retrieval(ISMIR), pp.164-169, 2002.
- [Kim et al., 2005]** H. G Kim, N. Moreau et T. Sikora, MPEG-7 audio and beyond. West Sussex: Wiley, ISBN: 047009334X, 2005.

- [Kohonen, 1984]** T. Kohonen. Self-Organization and Associative Memory. Springer-Verlag, 1984.
- [Krose et van der Smagt, 1996]** B. Krose et P. van der Smagt. An Introduction to Neural Networks. The University of Amsterdam, 8 edition, 1996.
- [Larochelle, 2008]** H. Larochelle, Étude de techniques d'apprentissage non-supervisé pour l'amélioration de l'entraînement supervisé de modèles connexionnistes, 2008. (Thèse)
- [Lay et Wang, 2004]** T. L. Nwe et Y. Wang. Automatic detection of vocal segments in popular songs. In ISMIR, 2004.
- [Lee et Ellis, 2008]** K. Lee et D. Ellis. Detecting Music in Ambient Audio by Long-Window Autocorrelation, In International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.9-12, 2008.
- [Li and Wang, 2006]** Y. Li et D. Wang, Singing voice separation from monaural recordings. In ISMIR, pages 176-179, 2006.
- [Li et Wang, 2006]** Y.Li, et D.L. Wang, Separation of singing voice from music accompaniment for monaural recordings, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, pp.1475-1487, 2006.
- [Lindsay et Herre, 2001]** A. Lindsay et J. Herre, MPEG-7 and MPEG-7 Audio - An Overview. J. Audio Eng. Society, Vol. 49 No. (7/8), pp.589-594, 2001.
- [Loh, 2008]** W.Y.Loh, Classification and regression tree methods, Encyclopedia of Statistics in Quality and Reliability, pp.315-323, 2008.
- [Loh et Shih, 1997]** W.Y. Loh, et Y.S. Shih, Split selection methods for classification trees, Statistica Sinica, Vol. 7, pp.815-840, 1997.
- [Maazouzi et Bahi, 2011]** F. Maazouzi et H. Bahi. Use of Gaussian Mixture Models and Vector Quantization for Singing Voice Classification in Commercial Music Productions. Proceedings of ISPS 2011, pp: 130-135, 2011.
- [Maazouzi et Bahi, 2012]** F. Maazouzi et H. Bahi. Features and Number of Gaussian Mixtures Selection for Singing Voice Classification in Commercial Music Production, The Mediterranean Journal of Computers and Networks. Vol. 8, No.1, 2012.
- [Maazouzi et Bahi , 2012a]** F. Maazouzi et H. Bahi, Using multi decision tree technique to improving decision tree classifier, International Journal of Business Intelligence and Data Mining, Vol.7 No. 4, pp. 274-287, 2012.
- [Maazouzi et Bahi, 2013]** F. Maazouzi et H. Bahi. Type-2 Fuzzy Gaussian Mixture Models for Singing Voice Classification in Commercial Music Production, International Journal of Signal and Imaging Systems, Vol.6 No.2, pp.111-118, 2013.

- [**Mauch et al., 2007**] M. Mauch, S. Dixon, C. Harte, M. Casey, et B. Fields. Discovering Chord Idioms Through Beatles and Real Book Songs. In Proc. of the 8th International Conference on Music Information Retrieval (ISMIR'07), 2007.
- [**McAulay et Quatieri, 1986**] R.McAulay et T. Quatieri. Speech analysis /synthesis based on a sinusoidal representation, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.34, No.4, pp.744-754, 1986.
- [**Meir et Ratsch., 2003**] R. Meir et G. Ratsch. An introduction to Boosting and Leveraging, 2003.
- [**Mitrovic et al., 2006**] D. Mitrovic, M. Zeppelzauer et H. Eidenberger, Analysis of the Data Quality of Audio Descriptors of environmental Sounds. Proc. the Fourth Special Workshop of Multimedia Semantics. Chania, Greece, pp.70-79, 2006.
- [**MLR, 2012**] Machine Learning Repository, [online] <http://archive.ics.uci.edu/ml/datasets.html>, (accessed May 2012).
- [**Montillo, 2009**] A.A. Montillo, Random forests, Lecture in Statistical Foundations of Data Analysis, 2009.
- [**New et Wang, 2004**] T. L. N. et Y. Wang. Automatic detection of vocal segments in popular songs. In ISMIR, 2004.
- [**Ozerov et al., 2007**] A. Ozerov, R. Gribonval, P. Philippe, et F. Bimbot, Choix et adaptation des modèles pour la séparation de voix chantée à partir d'un seul microphone, Traitement du signal, vol. 24, no. 3, pp. 211-224, 2007.
- [**Podgorelec et al., 2002**] V.Podgorelec, P.Kokol, B. Stiglic, et I. Rozman, Decision trees: an overview and their use in medicine', Journal of Medical Systems, Vol. 26, No. 5, pp.445-463, Kluwer Academic/Plenum Press, 2002.
- [**Prasert et al., 2008**] P. Prasert, K. Iwano et S. Furui, An Automatic Singing Voice Evaluation Method for Voice Training Systems. No. 2-5-12 pp. 911-912, 2008.
- [**Qiu et al, 2008**] D. Qiu, Y. Wang, et B. Bi, Identify cross-selling opportunities via hybrid classifier, IJDWM, Vol. 4, No. 2, pp.55-62, 2008.
- [**Quatieri, 2001**][6] T. F. Quatieri, Discrete Time Speech Signal Processing: Principles and Practice", Prentice Hall PTR, 2001.
- [**Quinlan, 1986**] J.R. Quinlan, Induction of decision trees, Machine Learning, Vol. 1, No. 1, pp.81-106, 1986.
- [**Quinlan, 1993**] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, Vol. 16, pp.236-240, 1993.

- [Raorane et al., 2012]** A.A. Raorane, R.V. Kulkarni et B.D. Jitkar, Association Rule Extracting Knowledge Using Market Basket Analysis, Research Journal of Recent Sciences, Vol. 1, pp. 19-27, 2012.
- [Ramona et al., 2008]** M. Ramona, B. Richard et G. David, Vocal detection in music with support vector machine, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 1885- 1888, March 2008.
- [Roe et al., 2006]** B.P. Roe, H-J. Yang, et J. Zhub, Boosted decision tree – a powerful event classifier, Statistical Problems in Particle Physics, Astrophysics and Cosmology, pp.139-143, 2006.
- [Routtenberg et Tabrikian, 2009]** T. Routtenberg, et J. Tabrikian, MIMO-AR system identification and blind source separation for GMM-distributed sources, IEEE Transactions Signal Processing, Vol. 57, pp.1717–1730, 2009.
- [Ros et al., 1999]** S. Rossignol, P. Depalle, J. Soumagne, X. Rodet, et J.L. Collette. Vibrato: detection, estimation, extraction, modification. Proceedings of Digital Audio Effects Workshop, 1999.
- [Serra et al., 2005]** M.-H. Serra, M. Fingerhut, M. David-Calvet, A. Bonardi, Guillaume Dimanche. La voix et l'ordinateur [Ressource électronique] / IRCAM. – Paris : IRCAM, Centre Pompidou, [s.d.].Accès : World Wide Web. – URL : <http://catalogue.ircam.fr/sites/Voix/index.html> .
- [Sundberg et Rossing, 1990]** J. Sundberg et T.D. Rossing, The science of singing voice, The Journal of the Acoustical Society of America, Vol. 87, pp.462, 1990.
- [Taniar et al., 2008]** D., Taniar, C.H.C Leung, W., Rahayu, et S. Goel, High Performance Parallel Database Processing and Grid Databases, John Wiley & Sons, 2008.
- [Todros et Tabrikian, 2007]** K. Todros et J. Tabrikian, Blind separation of independent sources using gaussian mixture model, IEEE Transactions Signal Processing, Vol. 55, pp.3645–3658, 2007.
- [Vautier, 2008]** A.Vautier, Fouille de données sans information a priori sur la structure de la connaissance, 2008. (Thèse)
- [Weka, 2012]** Weka, Data Mining with Open Source Machine Learning Software, [online] <http://www.cs.waikato.ac.nz/ml/weka/> (accessed 24 January 2012), 2012.
- [Whitney, 1971]** A.W. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Computers, Vol. 20, No. 9, pp. 1100– 1103, 1971.
- [Zeng et al. 2008]** J.Zeng, L. Xieb, et Z.–Q. Liuc, Type-2 fuzzy gaussian mixture models, Pattern Recognition J., Vol. 41, pp.3636–3643, 2008.
- [Zhang., 2003]** T. Zhang, Automatic singer identification. ICME'03. Proceedings of International Conference on Multimedia and Expo, 2003.

[zikdalgeire, 2012] www.zikdalgeire.com, dernière visite, 25 Dec. 2012.

[Zwan, 2006] P. Zwan, Expert system for automatic classification and quality assessment of singing voices, Proceedings of Audio Engineering Society 121th Convention. San Francisco, CA, USA, 2006.

[Zwan et al., 2007] P. Zwan, P. Szczuko, B. Kostek et A. Czyzewski, Automatic singing voice recognition employing neural networks and rough sets, Rough Sets and Intelligent Systems Paradigms, Vol. 4585, pp.793–802, 2007.

[Zwan et al., 2008] P. Zwan, P. Szczuko, B. Kostek et A. Czyzewski, Automatic Singing Voice Recognition Employing Neural Networks and Rough Sets. T. Rough Sets 9, PP. 455-473, 2008.

