

UNIVERSITÉ DU QUÉBEC

CLASSIFICATION DES SONS RESPIRATOIRES EN VUE D'UNE DÉTECTION
AUTOMATIQUE DES SIBILANTS

MÉMOIRE
PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À RIMOUSKI
COMME EXIGENCE PARTIELLE
DU PROGRAMME DE MAÎTRISE EN INGÉNIERIE

PAR
CHARLES PELLETIER

NOVEMBRE, 2006



Mise en garde/Advice

Afin de rendre accessible au plus grand nombre le résultat des travaux de recherche menés par ses étudiants gradués et dans l'esprit des règles qui régissent le dépôt et la diffusion des mémoires et thèses produits dans cette Institution, **l'Université du Québec à Chicoutimi (UQAC)** est fière de rendre accessible une version complète et gratuite de cette œuvre.

Motivated by a desire to make the results of its graduate students' research accessible to all, and in accordance with the rules governing the acceptance and diffusion of dissertations and theses in this Institution, the **Université du Québec à Chicoutimi (UQAC)** is proud to make a complete version of this work available at no cost to the reader.

L'auteur conserve néanmoins la propriété du droit d'auteur qui protège ce mémoire ou cette thèse. Ni le mémoire ou la thèse ni des extraits substantiels de ceux-ci ne peuvent être imprimés ou autrement reproduits sans son autorisation.

The author retains ownership of the copyright of this dissertation or thesis. Neither the dissertation or thesis, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

REMERCIEMENTS

Au terme de ce travail, je tiens à remercier mon directeur de recherche, Monsieur Mohammed Bahoura, professeur au département de Mathématiques, d'Informatique et de Génie (DMIG) de l'Université du Québec à Rimouski (UQAR), pour m'avoir donné l'opportunité d'effectuer cette recherche. Le grand intérêt qu'il a démontré, sa disponibilité ainsi que son évaluation des différentes versions de ce manuscrit m'ont permis de mener ce projet à terme.

J'adresse mes vifs remerciements aux professeurs Jean-François Méthot (UQAR) et Jean Rouat (Université de Sherbrooke) qui m'ont fait l'honneur de participer au jury d'évaluation de ce mémoire. Je tiens aussi à remercier mes anciens collègues messieurs Stéphane Boucher et Jean-François Garneau qui m'ont permis de me lancer dans cette étude, en appuyant de prime abord ma candidature.

Je désire exprimer ma reconnaissance et remercier toutes les autres personnes qui de près ou de loin m'ont aidé dans la réalisation de mes travaux et la rédaction de mon mémoire. Je remercie les étudiants gradués : Richard Lepage, Jean-Louis Rétiveau et Xiaoguang Lu pour les discussions fructueuses qui m'ont permis de mieux saisir le sujet de la reconnaissance de formes. Mes remerciements vont à Monsieur Edward Babulak, professeur invité à l'UQAR, pour les conseils judicieux tant qu'à la remise de ce mémoire.

Je remercie également le professeur Yvan Simard de l'Institut Maurice Lamontagne (IML) pour sa confiance en m'offrant un contrat de recherche. Je remercie Messieurs Éric Hudier et Émilien Caron pour les nombreuses discussions stimulantes.

Enfin, je remercie ma femme, Suzanne Tremblay, pour m'avoir soutenu pendant mon retour aux études.

Cette recherche a été rendu possible grâce au support financier du Conseil de Recherches en Sciences Naturelles et en Génie (CRSNG).

RÉSUMÉ

L'asthme est une maladie inflammatoire chronique des voies respiratoires. Il se manifeste par des crises de dyspnée sifflantes durant lesquels les sons émis sont appelés sibilants. Ces crises sont souvent nocturnes et réversibles, spontanément ou sous l'effet du traitement. Les sibilants caractérisant l'asthme se superposent aux sons respiratoires normaux et apparaissent le plus souvent à l'expiration.

Notre projet de recherche consiste à réaliser un détecteur automatique des sibilants. Cette problématique sera traitée comme un processus de reconnaissance de formes. Le processus de reconnaissance de formes se compose de deux modules indissociables : 1) l'extraction des caractéristiques des signaux, et 2) la modélisation permettant la discrimination des classes. Des étapes de prétraitement ou de post traitement des données peuvent s'ajouter pour perfectionner le système. Les performances du système de reconnaissance dépendent des combinaisons des types de modules adoptés dans ce système.

Pour ce projet, nous avons testé plusieurs combinaisons de techniques d'extraction des caractéristiques et de modélisation. Nous avons aussi analysé l'effet du prétraitement et du post traitement des données. Les résultats ont été comparés à l'aide de l'aire sous la courbe (*AUC - Area Under Curve*) de tracés de type (*ROC - Receiver Operating Characteristic*). Le meilleur résultat a été obtenu en utilisant, comme caractéristiques, 16 coefficients de type (*MFCC - Mel Frequency Cepstrum Coefficients*) extraits d'un signal échantillonné à 6000 Hz sur des segments du signal sonore de 1024 échantillons. Ces paramètres ont été modélisés par la technique (*GMM - Gaussian Mixture Model*) à 8 regroupements en utilisant des matrices de covariance de type diagonale. Aucun prétraitement, effectué sur les paramètres, ne s'est avéré efficace. Par contre, une technique de post traitement, effectuée sur les mesures obtenues lors de l'étape de classification, s'est avérée bénéfique.

TABLE DES MATIÈRES

RÉSUMÉ	iii
LISTE DES FIGURES	vii
LISTE DES TABLEAUX	x
LISTE DES ABRÉVIATIONS	xi
INTRODUCTION	1
CHAPITRE I RECONNAISSANCE DE FORMES POUR CLASSIFIER LES SONS RES- PIRATOIRES	3
1.1 Nomenclature et classification des sons respiratoires	3
1.1.1 Sons respiratoires normaux	3
1.1.2 Sons respiratoires adventices	4
1.2 Emplacement de la prise de son	6
1.3 Base de données utilisée	6
1.4 Reconnaissance de formes	6
1.5 Problématique	8
1.6 Objectifs	8
1.7 Méthode	8
1.7.1 Étude des différentes combinaisons des techniques d'extraction des caractéristiques et de modélisation	10
1.7.2 Optimisation des paramètres des techniques d'extraction des caractéristiques et de modélisation	11
1.7.3 Étude de l'effet des méthodes de prétraitement des données	11
1.8 Hypothèses	12
CHAPITRE II EXTRACTION DES CARACTÉRISTIQUES	13
2.1 Principe de l'extraction des caractéristiques d'un signal	13
2.2 Analyse de Fourier	15

2.2.1	Transformée de Fourier à court terme (<i>STFT</i>)	15
2.3	Analyse Cepstrale	20
2.3.1	Coefficients Cepstraux	20
2.3.2	Coefficients cepstraux à l'échelle de Mel (<i>MFCC</i>)	23
2.4	Modélisation linéaire prédictive (<i>LPC</i>)	26
2.4.1	Coefficients de la prédition linéaire	26
2.5	Analyse par ondelettes	28
2.5.1	Transformée en ondelettes continue (<i>CWT</i>)	28
2.5.2	Transformée en ondelettes discrète (<i>DWT</i>)	29
2.5.3	Transformée en paquets d'ondelettes (<i>WPT</i>)	32
2.6	Prétraitement	40
2.6.1	Prétraitement sur chaque segment du signal	41
2.6.2	Prétraitement sur le vecteur des caractéristiques	42
CHAPITRE III		
MODÉLISATION ET CLASSIFICATION DES SIGNAUX		46
3.1	Principe d'un système de classification	46
3.2	Nomenclature	47
3.3	Modélisation	48
3.4	Quantification vectorielle (<i>VQ</i>)	49
3.4.1	Phase d'apprentissage	51
3.4.2	Phase de classification	52
3.4.3	Algorithme <i>LBG</i>	54
3.5	Modélisation par mélange de gaussiennes (<i>GMM</i>)	55
3.5.1	Principe de la modélisation par mélange de gaussiennes	55
3.5.2	Phase d'apprentissage	57
3.5.3	Phase de classification	58
3.5.4	Algorithme EM (<i>Expectation Maximisation</i>)	59
3.6	Mélange de distributions Student (<i>SMM</i>)	63
3.6.1	Principe	63
3.6.2	Algorithme	64
3.7	Perceptron multi-couche (<i>MLP</i>)	66

3.7.1	Principe	66
3.7.2	Phase d'apprentissage	66
3.7.3	Phase de classification	68
3.8	Comparaison des discriminants pour différentes techniques de modélisation	69
3.9	Contrainte d'un discriminant complexe	70
3.10	Post traitement	70
CHAPITRE IV		
MÉTHODOLOGIE ET RÉSULTATS		72
4.1	Préparation de la base de données	72
4.2	Protocole d'expérimentation	73
4.3	Critères d'évaluation du système de reconnaissance	73
4.4	Résultats préliminaires	79
4.5	Non-reproductibilité des résultats du réseau <i>MLP</i>	80
4.6	Effets de différents paramètres	82
4.6.1	Effets du prétraitement	82
4.6.2	Effets du recouvrement	83
4.6.3	Effets de la largeur de la fenêtre d'analyse	84
4.6.4	Effets du nombre d'itérations	85
4.6.5	Effets du type d'initialisation et du type de matrice de covariance sur la classification par la méthode <i>GMM</i>	87
4.6.6	Effets de la dimension et du nombre de regroupements	87
4.7	Résultats optimaux pour chaque type de modélisation	90
4.7.1	Effets du post-traitement	90
CONCLUSION		94
RÉFÉRENCES		96
BIBLIOGRAPHIE		100

LISTE DES FIGURES

1.1	Reconnaissance de formes : Apprentissage et test	7
1.2	Section du spectrogramme contenant un sibilant monophonique	9
1.3	Section du spectrogramme contenant un sibilant polyphonique	9
2.1	Représentation de la totalité d'un signal dans l'espace de ses caractéristiques	14
2.2	Spectrogramme d'un son respiratoire normal	16
2.3	Spectrogramme du son respiratoire avec sibilants	16
2.4	Obtention du spectrogramme par la transformée de Fourier à court terme (<i>STFT</i>)	17
2.5	Résolution en temps et en fréquence	18
2.6	Résolution du spectrogramme	18
2.7	Caractéristiques tirées des coefficients obtenus suite à la transformée de Fourier à court terme (<i>STFT</i>)	20
2.8	Modélisation des conduits respiratoires	21
2.9	Principe de la séparation de la source et des conduits respiratoires par analyse cepstrale	22
2.10	Coefficients cepstraux	22
2.11	Banc de sept filtres espacés selon l'échelle de Mel	23
2.12	Principe du calcul des coefficients <i>MFCC</i>	24
2.13	Démarche suivie pour l'obtention de sept coefficients <i>MFCC</i>	25
2.14	Caractéristiques tirées des coefficients <i>MFCC</i>	25
2.15	Comparaison des méthodes d'estimation spectrale (<i>FFT-LPC</i>)	27
2.16	Comparaison des résolutions en temps et en fréquence obtenues par la transformée à court terme (<i>STFT</i>) et la transformée en ondelettes (<i>CWT</i>)	29
2.17	Algorithme de Mallat pour l'implantation de la décomposition en ondelettes discrète (<i>DWT</i>)	30
2.18	Exemple de la décomposition en ondelettes discrète (<i>DWT</i>)	31
2.19	Caractéristiques tirées des coefficients d'ondelettes (<i>DWTC</i>)	32
2.20	Transformée en paquets d'ondelettes (<i>WPD</i>)	33
2.21	Illustration en temps et en fréquence de la transformée en paquets d'ondelettes (<i>WPD</i>)	33
2.22	Recherche de la meilleure base de décomposition	35
2.23	Caractéristiques à base des meilleurs coefficients d'ondelettes <i>WBC</i> (1 de 2)	36

2.24	Caractéristiques à base des meilleurs coefficients d'ondelettes <i>WBC</i> (2 de 2)	37
2.25	Caractéristiques à base des coefficients <i>SBC</i>	38
2.26	Diagramme de calcul des coefficients <i>SBC</i>	39
2.27	Caractérisation par la méthode des coefficients de paquets d'ondelettes (<i>WPC</i>)	41
2.28	Prétraitement appelé "centrage" des données	42
2.29	Prétraitement en effectuant une analyse par composante principale (<i>PCA</i>)	43
2.30	Prétraitement appelé "blanchiment-1" effectué sur une classe de données	45
2.31	Prétraitement appelé "blanchiment-2" effectué sur l'ensemble des données	45
3.1	Diagramme bloc d'un système de classification	47
3.2	Exemple de données artificielles à discriminer appartenant à deux classes <i>R</i>	48
3.3	Exemple de modélisation des données selon différentes méthodes	49
3.4	Modélisation des données par la méthode de quantification vectorielle (<i>VQ</i>)	50
3.5	Cellules de Voronoï	53
3.6	Amélioration du discriminant par la méthode <i>VQ</i>	53
3.7	Algorithme <i>LBG</i>	54
3.8	Modélisation des données obtenue par mélange de deux gaussiennes	57
3.9	Discriminant obtenu par la méthode <i>GMM</i>	58
3.10	Modélisation des données obtenues par mélange de quatre gaussiennes	59
3.11	Amélioration du discriminant par la méthode <i>GMM</i>	59
3.12	Initialisation de l'algorithme EM utilisée pour la méthode <i>GMM</i>	61
3.13	Discriminant obtenu analytiquement	63
3.14	Modélisation des données par la méthode <i>SMM</i>	64
3.15	Réseau de neurones de type <i>MLP</i>	67
3.16	Discriminant obtenu par réseau <i>MLP</i>	68
3.17	Comparaison des discriminants pour différentes méthodes de modélisation	69
3.18	Désavantage d'un discriminant trop complexe	70
3.19	Principe du post traitement de la fonction distance	71
4.1	Méthode d'évaluation <i>leave one out</i>	74
4.2	Technique pour obtenir une courbe <i>ROC</i> : Cas continu	75
4.3	Technique pour obtenir une courbe <i>ROC</i> : Cas discret	76
4.4	Relation entre des aires sous la courbe (<i>AUC</i>) obtenues de deux différentes façons	77
4.5	Courbe de caractéristiques d'efficacité <i>ROC</i>	78

4.6	Résultats préliminaires de différents systèmes de classification	78
4.7	Variabilité des résultats de classification par <i>MLP</i>	80
4.8	Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation <i>VQ</i>	81
4.9	Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation <i>GMM</i>	81
4.10	Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation par réseau de neurones <i>MLP</i>	82
4.11	Effets du recouvrement sur la performance de la classification	83
4.12	Effets de la largeur de la fenêtre d'analyse sur la performance (AUC) de la classification	84
4.13	Effets du nombre d'itérations sur la performance (AUC) d'un système de classification <i>MFCC/VQ</i>	85
4.14	Effets du nombre d'itérations sur la performance (AUC) d'un système de classification <i>MFCC/GMM</i>	86
4.15	Effets du nombre d'itérations sur la performance (AUC) d'un système de classification <i>MFCC/MLP</i>	86
4.16	Effets du type d'initialisation et de matrice de covariance pour le système de classification <i>MFCC/GMM</i>	87
4.17	Effets de la dimension et du nombre de regroupements sur la performance (AUC) de la classification <i>MFCC/VQ</i>	88
4.18	Effets de la dimension et du nombre de regroupements sur la performance (AUC) d'un système de classification de combinaison <i>MFCC/GMM</i> et <i>SBC/GMM</i>	89
4.19	Effets de la dimension et des regroupements sur la performance (AUC) d'un système de classification de type <i>MFCC/MLP</i>	90
4.20	Amélioration des performances de la classification par filtrage (post-traitement) de la fonction distance	92
4.21	Effets du post-traitement sur les performances(AUC) d'un système de classification <i>MFCC/GMM</i>	93

LISTE DES TABLEAUX

1.1	Termes originaux de Laënnec	4
1.2	Définitions et appellations des sons adventices	5
1.3	Classification des sons respiratoires.	8
1.4	Combinaisons des différentes techniques utilisées pour la classification des sons respiratoires (voir la liste des abréviations).	10
1.5	Paramètres reliés aux techniques de modélisation pour la classification de sons respiratoires.	12
3.1	Nomenclature et symboles utilisés pour décrire les systèmes de classification	47
4.1	Signaux formant la base de données utilisée	73
4.2	Paramètres de l'expérimentation préliminaire	79
4.3	Paramètres optimaux pour les différents types de modélisation	91
4.4	Paramètres optimaux pour le système de classification <i>MFCC/GMM</i> . . .	91

LISTE DES ABRÉVIATIONS

<i>ANN</i>	<i>: Artificial Neural Network</i>	: Réseau de neurones artificielles
<i>AUC</i>	<i>: Area Under Curve</i>	: Aire sous la courbe
<i>CC</i>	<i>: Cepstral Coefficients</i>	: Coefficients cepstraux
<i>CWT</i>	<i>: Continuous Wavelet Transform</i>	: Transformée en ondelettes continue
<i>DCT</i>	<i>: Discrete Cosine Transform</i>	: Transformée en cosinus discrète
<i>DFT</i>	<i>: Discrete Fourier Transform</i>	: Transformée de Fourier discrète
<i>DWT</i>	<i>: Discrete Wavelet Transform</i>	: Transformée en ondelettes discrète
<i>DWTC</i>	<i>: Discrete Wavelet Transform Coefficients</i>	: Coefficients de la transformée en ondelettes discrète
<i>EM</i>	<i>: Expectation - Maximisation</i>	: Estimation - Maximisation
<i>FIR</i>	<i>: Finite Impulse Response</i>	: Réponse impulsionnelle finie
<i>FFT</i>	<i>: Fast Fourier Transform</i>	: Transformée de Fourier rapide
<i>GMM</i>	<i>: Gaussian Mixture Model</i>	: Mélange de distributions gaussiennes
<i>HIR</i>	<i>: Infinite Impulse Response</i>	: Réponse impulsionnelle infinie
<i>LBG</i>	<i>: Linde Buzo Gray</i>	<i>: Linde Buzo Gray</i>
<i>LPC</i>	<i>: Linear Predictive Coding</i>	: Codage linéaire prédictif
<i>LVQ</i>	<i>: Learning Vector Quantization</i>	: Quantification vectorielle supervisée
<i>MFCC</i>	<i>: Mel Frequency Cepstrum Coefficients</i>	: Coefficients cepstraux de Mel
<i>MLP</i>	<i>: Multi Layer Perceptron</i>	: Perceptron multi-couches
<i>PCA</i>	<i>: Principal Component Analysis</i>	: Analyse par composantes principales
<i>PDF</i>	<i>: Probability Density Function</i>	: Densité de probabilité (DDP)
<i>QMF</i>	<i>: Quadrature Mirror Filter</i>	: Filtres miroirs en quadrature
<i>RBF</i>	<i>: Radial Basis Function</i>	: Fonction radiale de base
<i>ROC</i>	<i>: Receiver Operating Characteristic</i>	: Courbe de caractéristique d'efficacité
<i>SBC</i>	<i>: Subband Based Coefficients</i>	: Coefficients tirés des sous-bandes
<i>SMM</i>	<i>: Student Mixture Model</i>	: Mélange de distributions de Student
<i>SOM</i>	<i>: Self Organization Map</i>	: Carte auto-organisatrice
<i>STFT</i>	<i>: Short Time Fourier Transform</i>	: Transformée de Fourier à court terme
<i>SVD</i>	<i>: Singular Value Decomposition</i>	: Décomposition en valeur singulière
<i>VQ</i>	<i>: Vector Quantization</i>	: Quantification vectorielle
<i>WBC</i>	<i>: Wavelet Best Coefficient</i>	: Meilleurs coefficients d'ondelettes
<i>WPC</i>	<i>: Wavelet Packet Coefficient</i>	: Coefficients de paquets d'ondelettes
<i>WPD</i>	<i>: Wavelet Packet Decomposition</i>	: Décomposition en paquets d'ondelettes

INTRODUCTION

L'asthme est l'une des maladies chroniques les plus répandues chez les enfants au Canada et pose également un grave problème chez les adultes. Selon l'Enquête Nationale sur la Santé de la Population (ENSP) de 1998-1999 [1], 2.5 millions de personnes souffrent de l'asthme (8% des adultes et 12% des enfants). Au niveau mondial, de 100 à 150 millions de personnes souffrent d'asthme et leur nombre est en augmentation. On enregistre plus de 180 000 décès par an dus à cette infection [2].

L'asthme est une maladie inflammatoire chronique des voies respiratoires. Il se manifeste par des crises d'essoufflement sifflantes dont le son caractéristique est nommé sibilant. Ces crises sont réversibles spontanément ou suite à un traitement. Le praticien spécialiste, dans ce type de maladie, se trouve confronté à une grande difficulté pour établir un diagnostic pertinent car la maladie d'asthme se caractérise le plus souvent par des crises nocturnes suivies de phases de répit, essentiellement diurnes. Donc, en général, le praticien voit son patient lors d'une phase de répit non significative de la gravité de la maladie.

Les systèmes de surveillance de la fonction respiratoire, semblables à ceux utilisés dans l'analyse du signal ECG (Électrocardiogramme), sont toujours en phase de développement [3-13]. La complexité de ce type de signaux et le manque de standardisation entre les équipes de recherche freinent considérablement la recherche sur les signaux respiratoires en général, et le développement des systèmes pour le diagnostic en particulier.

C'est pour répondre à cette problématique, qu'un projet de développement d'un outil d'identification des sibilants à domicile, particulièrement utile pendant les épisodes de crise d'asthme nocturne, a vu le jour au département de Mathématiques d'Informatique et de Génie (DMIG) de l'Université du Québec à Rimouski (UQAR).

Notre travail consiste à mettre en place un système fiable de détection automatique des sibilants à partir d'enregistrements acoustiques réalisés sur site, c'est-à-dire dans des conditions réelles, généralement bruitées. Ce processus de détection sera ramené à une problématique plus globale de reconnaissance de formes.

Dans ce mémoire nous présenterons donc, dans un premier chapitre, la problématique générale de la reconnaissance de formes ainsi que son utilisation dans la détection des sibilants. De plus, nous y introduirons les notions essentielles concernant les sons respiratoires tels que les appréhendent les praticiens. Dans le second chapitre, nous examinerons les techniques d'extraction des caractéristiques représentant le signal. Le troisième chapitre comprend la description des techniques de modélisation (classification) permettant d'obtenir le meilleur discriminant entre sons normaux et sons avec sibilants. Enfin, le dernier chapitre concernera la méthodologie de l'expérimentation ainsi que les résultats obtenus. Dans la conclusion, nous ferons le bilan de ce travail et proposerons des perspectives pour la prochaine étape de réalisation.

CHAPITRE I

RECONNAISSANCE DE FORMES POUR CLASSIFIER LES SONS RESPIRATOIRES

Dans ce chapitre, nous mettrons en exergue les définitions et les caractéristiques des différents bruits respiratoires, ainsi que l'origine de leurs productions. Nous décrirons, par la suite l'approche de la reconnaissance de formes pour détecter les sibilants.

1.1 Nomenclature et classification des sons respiratoires

Les qualificatifs de René Laënnec, médecin français inventeur du stéthoscope, donnés aux différents types de râles (humides, muqueuses, secs, grosses bulles...) n'étaient pas pathognomoniques et ne permettaient pas une distinction significative entre pneumonies, maladies pulmonaires chroniques et insuffisances cardiaques (voir le tableau 1.1). Les difficultés, voire les erreurs de traduction de la terminologie de Laënnec et l'introduction de nouveaux termes ou qualificatifs, ont entraîné une grande disparité, voire une confusion entre les écoles de médecine^[14]. Une normalisation de la terminologie des sons respiratoires a été proposée, en 1976, par l'*International Lung Sound Association (ILSA)*, et adoptée l'année suivante par l'*American Thoracic Society (ATS)*^[15]. Cette nouvelle nomenclature est résumée dans le tableau 1.2. Les sons respiratoires se divisent en deux grandes classes : les sons normaux et les sons adventices ou anormaux. Nous donnerons les définitions et les composantes de chacune des classes.

1.1.1 Sons respiratoires normaux

Les sons respiratoires normaux sont produits par le mouvement de l'air dans le conduit respiratoire lorsqu'une personne respire normalement. Selon la position où l'on mesure les

Tableau 1.1: Traduction des termes originaux de Laënnec pour la classification des sons respiratoires(Mikami et al. [16] selon [17]).

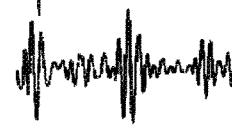
Description originale de Laënnec	Traduction anglaise
Râle humide ou crépitation	Wet rale, crepitation, or crackle
Râle muqueux ou gargouillement	Mucous or gurgling rale
Râle sec sonore ou ronflement	Dry or snoring rale
Râle sibilant sec ou sifflement	Dry whistling or wheezing rale
Râle crépitant sec à grosses bulles ou craquement	Dry crackling rale with large bubbles

sons respiratoires, on obtiendra les sons trachéo-bronchiques ou les murmures vésiculaires. Les sons trachéo-bronchiques sont obtenus près de la trachée, du larynx ou des grosses bronches alors que les murmures vésiculaires sont obtenus sur la surface thoracique.

1.1.2 Sons respiratoires adventices

Les sons adventices sont superposés aux sons respiratoires normaux ou les remplacent. Ces sons indiquent souvent des infections du système respiratoire. Ils sont répartis en deux classes, selon qu'ils soient continus ou discontinus [14]. Les sons adventices continus sont des bruits anormaux d'une durée supérieure à 250 ms. Dans cette classe, on trouve les sibilants et les ronchus. Un sibilant est défini comme étant une composante sonore à caractère musical surajouté au son respiratoire normal. Les sibilants sont principalement présents dans la phase d'expiration. On les dira monophoniques ou polyphoniques selon qu'ils sont composés d'un ou plusieurs harmoniques. Le mécanisme de production des sibilants au sein des bronches semble impliquer l'interaction des parois et du courant d'air. Pour Forgacs [19] (cité par Loudon et Murphy [18]), ces sons sont produits quand la section des voies aériennes est resserrée à tel point que les parois opposées sont en quasi-contact et que la vitesse du courant d'air atteint alors une valeur critique pour forcer cet obstacle, faisant ainsi vibrer les parois des bronches. Certains auteurs [20-22] définissent le sibilant comme étant un son continu, périodique, de fréquence fondamentale supérieure à 100 Hz et d'une durée supérieure à 100 ms. D'autres auteurs [16, 18] utilisent la définition de l'*American Thoracic Society* (tableau 1.2) qui le définit comme étant un son périodique de fréquence variable dans le temps mais supérieure à 400 Hz et de durée supérieure à 250 ms. Nous utiliserons cette dernière définition.

Tableau 1.2: Définitions et appellations des sons adventices selon l'American Thoracic Society (Tableau de [16, 18]).

G.B. et U.S.A.	ALLEMAGNE	FRANCE	Proposition de modification	Forme d'onde en temps étendu	
Discontinuous (< 250 ms)					
<u>Fine</u> (high pitched, low amplitude, short duration, IDW* = 0.92 ms, 2CD† = 6.02 ms)	Fine crackles	Feines Rasseln	Râles crépitants	Crépitements fins	
<u>Coarse</u> (low pitched, high amplitude, long duration, IDW = 1.25 ms, 2CD = 9.32 ms)	Coarse crackles	Grobes Rasseln	Râles bulleux ou sous-crépitants	Gros crépitements	
Continuous (> 250 ms)					
<u>High pitched</u> (dominant frequency > 400 Hz)	Wheezes	Pfeifen	Râles sibilants	Siflements	
<u>Low pitched</u> (dominant frequency <= 200 Hz)	Ronchi	Brummen	Râles ronflants	Ronchus	

* Initial Deflection Width (largeur de déflection initiale)

† Two Cycle Duration (durée des 2 premiers cycles)

1.2 Emplacement de la prise de son

La transmission des sibilants à travers les voies respiratoires est meilleure que la transmission à travers la cage thoracique. Takerawa et al. [23] (cité par Loudon et Murphy [18]) ont démontré que les composantes de fréquences élevées (1000 Hz et plus) étaient nettement mieux détectées au niveau de la trachée. Ces études renforcent l'importance d'écouter les sons respiratoires au niveau de la trachée pour étudier les malades asthmatiques. Il faut rappeler que les sons normaux captés à la trachée sont dits trachéo-bronchiques [14].

1.3 Base de données utilisée

Les sons respiratoires que nous avons utilisés dans ce projet, proviennent des bases de données ASTRA [24] et RALE [25] et de sites internet. Ces signaux sont captés à l'aide d'un microphone, placé au niveau de la trachée. Ils sont ensuite échantillonnés à une fréquence vérifiant le théorème de Shannon. Pour des considérations d'uniformisation et de quantité de données à traiter, nous avons sous-échantillonné les signaux pour faire en sorte qu'ils soient tous rammenés à un échantillonnage de 6000 Hz.

1.4 Reconnaissance de formes

Le problème de la reconnaissance de formes se subdivise, généralement, en deux phases : l'*apprentissage* et le *test*. La phase d'apprentissage consiste à créer des modèles en vue d'établir un *discriminant* entre les différentes classes, à partir de signaux dont on connaît la classe d'appartenance. Ceci s'effectue en trois étapes : 1) l'*extraction de caractéristiques* qui consiste à représenter chaque segment du signal temporel par une seule donnée dans l'espace des caractéristiques, 2) la *modélisation* qui permet de représenter l'ensemble des données par un nombre restreint de paramètres et 3) l'obtention d'une *fonction discriminante* séparant les différentes classes. Cette dernière est illustrée par la frontière séparant les zones R_1 et R_2 de la figure 1.1. Lors de la phase de test, les caractéristiques extraites d'un segment du signal sont classifiées selon leur emplacement par rapport à la fonction discriminante.

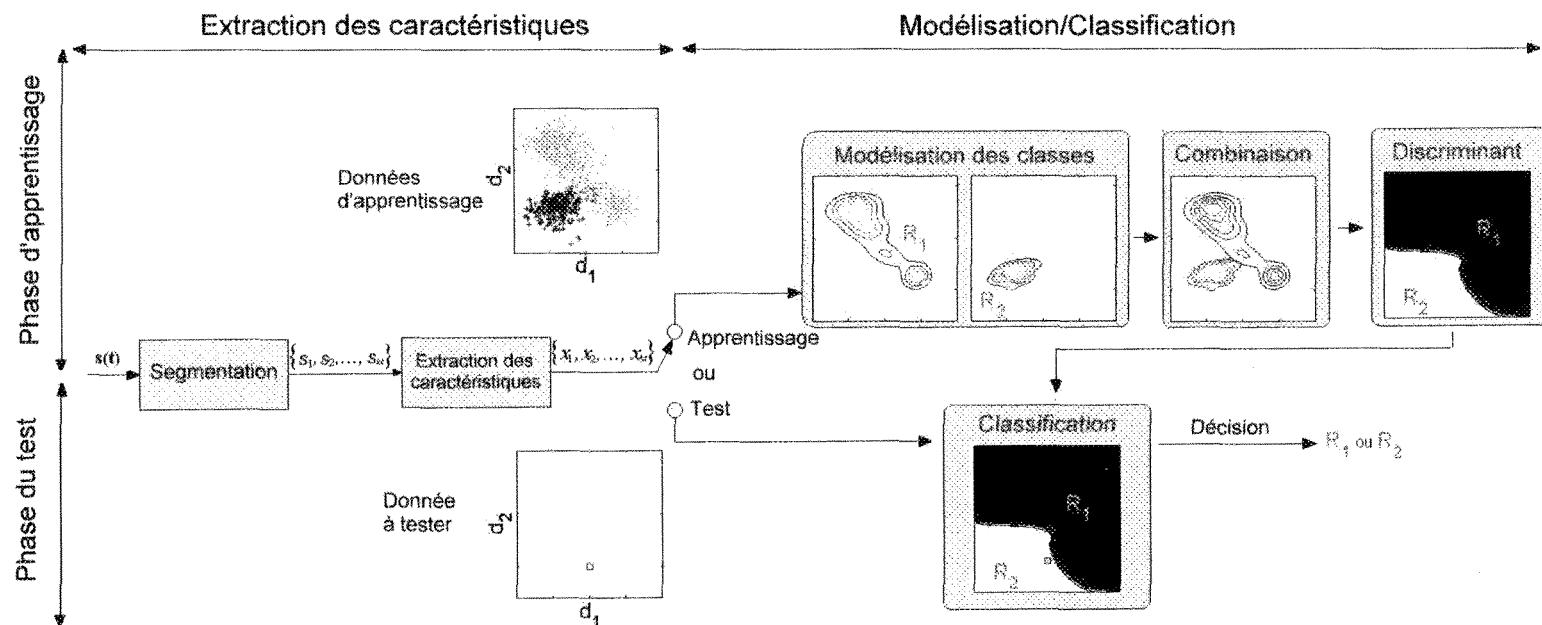


Figure 1.1: Principe de la reconnaissance de formes. Ici, nous avons deux types de données $(+)\in R_2$ et $(\cdot)\in R_1$ représentées en deux dimensions (d_1 et d_2). La modélisation permet d'obtenir une fonction discriminante (illustrée par la frontière entre la zone R_1 et R_2). Le carré (\square) représente une donnée à classifier.

Tableau 1.3: Classification des sons respiratoires.

Sons Respiratoires				
Normaux		Adventices		
Trachéo-Bronchiques	Murmures Vésiculaires	Continus $> 250ms$	Discontinus $< 250ms$	
		Sibilants $> 400Hz$	Ronchus $< 200Hz$	Gros Crépitements Crépitements Fins

1.5 Problématique

La problématique de recherche consiste à mettre en place un système de surveillance de la fonction respiratoire chez les malades asthmatiques en particulier. Cet outil permettra d'aider à faire un bon diagnostic, de suivre l'évolution de la maladie, et d'évaluer l'efficacité des traitements prescrits. La complexité de ce type de signaux et le manque de standardisation entre les équipes de recherche freinent considérablement la recherche sur les signaux respiratoires.

1.6 Objectifs

Ce projet de recherche consiste à mettre en place un dispositif permettant de distinguer deux classes de sons respiratoires ; en l'occurrence, les sons trachéo-bronchiques et les sibilants. Le tableau 1.3 situe les deux classes dans l'arbre de décomposition des différentes classes de sons respiratoires. Les figures 1.2 et 1.3 montrent le spectrogramme de deux cycles respiratoires. Le curseur dans la figure 1.2 montre un sibilant monophonique d'une fréquence d'environ 1000 Hz. Tandis que dans la figure 1.3, il indique un sibilant polyphonique aux fréquences d'environ 400 Hz et 1000 Hz.

1.7 Méthode

Pour atteindre les objectifs de ce projet nous faisons appel aux techniques de traitement de signaux liées à la problématique de la reconnaissance de formes. Ainsi, les performances du processus de séparation de classes dépendent énormément des techniques utilisées. Nous avons divisé ces techniques en deux groupes : 1) les méthodes d'extraction

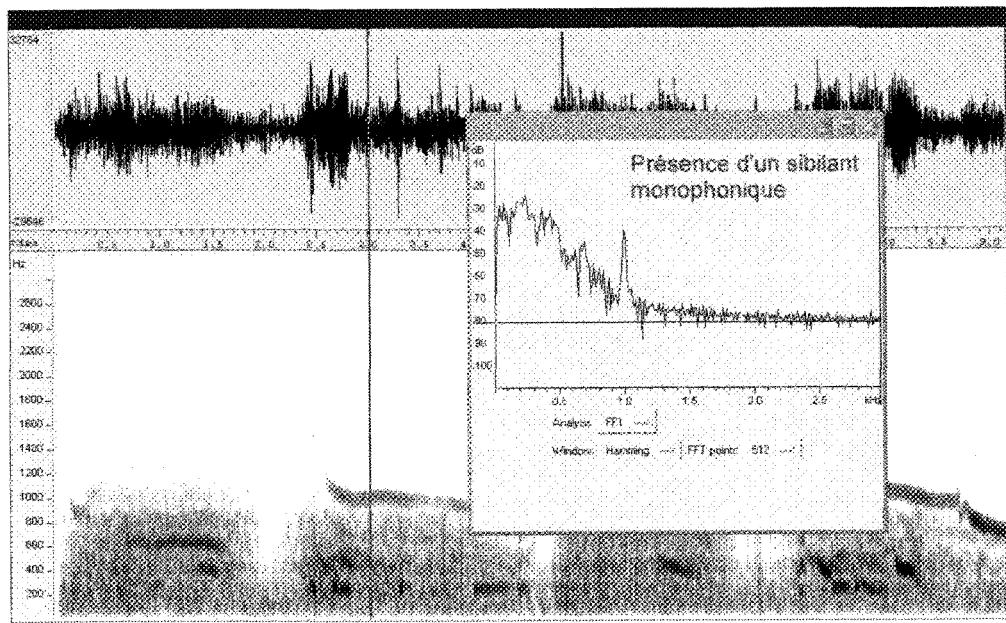


Figure 1.2: Section du spectrogramme d'un son respiratoire contenant un sibilant monophonique en utilisant le logiciel WaveSurfer [26].

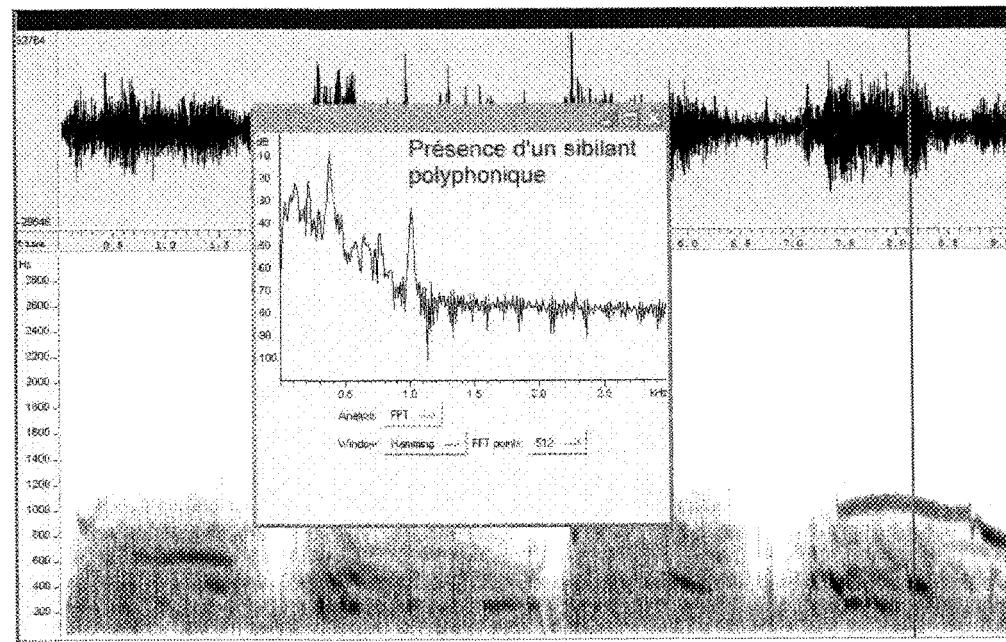


Figure 1.3: Section du spectrogramme d'un son respiratoire contenant un sibilant polyphonique en utilisant le logiciel WaveSurfer [26].

Tableau 1.4: Combinaisons des différentes techniques utilisées pour la classification des sons respiratoires (voir la liste des abréviations).

Extraction des caractéristiques	Modélisation \classification
<i>MFCC</i>	
<i>SBC</i>	
<i>FFT</i>	<i>VQ</i>
<i>LPC</i>	<i>GMM</i>
<i>CC</i>	<i>SMM</i>
<i>WBC</i>	<i>MLP</i>
<i>DWTC</i>	
<i>WPC</i>	

des caractéristiques et 2) les méthodes de modélisation/classification (voir la figure 1.1).

Notre approche pour mettre en place un tel système se résume en trois étapes :

1.7.1 Étude des différentes combinaisons des techniques d'extraction des caractéristiques et de modélisation

Cette étape consiste à trouver la meilleure combinaison entre les techniques d'extraction des caractéristiques et celles de la modélisation/classification. Pour la reconnaissance des sibilants, plusieurs combinaisons ont été proposées dans la littérature. La combinaison du spectre et des réseaux de neurones a été étudiée par Forkheim et al. [27], Rietveld et al. [9], et Waitman et al. [10]. La combinaison de la technique LPC et du plus proche voisin a été étudiée par Sankur et al. [28]. Des combinaisons de techniques par ondelettes et par réseaux de neurones ont été étudiées par Pesu et al. [8], Kandaswamy et al. [12]. Enfin, la combinaison des techniques MFCC et GMM est celle que nous avons proposée dans le cadre de ce projet [29,30] .

Afin d'atteindre ce premier objectif, Nous chercherons la combinaison optimale parmi un nombre restreint de techniques de caractéristiques et de méthodes modélisation/classification (voir le tableau 1.4).

1.7.2 Optimisation des paramètres des techniques d'extraction des caractéristiques et de modélisation

Cette étape consiste à trouver les paramètres optimaux sous-jacents à la technique d'extraction des caractéristiques et ceux sous-jacents à la technique de modélisation choisie.

Nous étudierons d'une part, l'effet des paramètres suivants lors de l'extraction des caractéristiques :

- la dimension des vecteurs caractéristiques ;
- la largeur de la fenêtre d'observation ;
- le taux de recouvrement des fenêtres.

D'autre part, nous étudierons l'effet des paramètres reliés à la modélisation/classification (voir le tableau 1.5). Nous verrons, à la section 3.8, que les paramètres "nombre de neurones de la couche cachée" et "nombre de regroupement", du tableau 1.5, sont similaires. Nous utiliserons alors "nombre de regroupements" pour décrire ces deux paramètres.

1.7.3 Étude de l'effet des méthodes de prétraitement des données

Cette étape consiste à déterminer si un prétraitement effectué sur les données extraites affectera les résultats de la classification. Les techniques de prétraitement étudiées sont les suivantes :

- la technique du centrage ;
- l'analyse par composantes principales (*PCA*) ;
- la technique du blanchiment.

Il faut rappeler que l'objectif principal de cette recherche est de déterminer les trois points, listés ci-dessous, permettant d'obtenir le meilleur taux de classification :

- trouver la meilleure combinaison des techniques d'extraction et de modélisation ;
- trouver le meilleur ajustement des différents paramètres de ces techniques ;
- déterminer si un prétraitement est nécessaire.

Tableau 1.5: Paramètres reliés aux techniques de modélisation pour la classification de sons respiratoires.

Paramètre	<i>VQ</i>	<i>GMM</i>	<i>MLP</i>	<i>SMM</i>
Nombre d'itération	✓	✓	✓	✓
Type de la matrice covariance		✓		✓
Type d'initialisation		✓		✓
Nombre de regroupements	✓	✓		✓
Nombre de neurones de la couches cachées			✓	

1.8 Hypothèses

Afin de comparer les résultats des différentes expérimentations, nous avons opté pour la mesure de l'aire sous la courbe (AUC - *Area Under Curve*) de la courbe de caractéristiques d'efficacité (ROC - *Receiver Operating Characteristic*) décrite à la section 4.3.

Les courbes ROC sont obtenues suite à l'utilisation de la technique “*leave one out*” effectuées sur le données listées à la section 4.1.

CHAPITRE II

EXTRACTION DES CARACTÉRISTIQUES

Ce chapitre traite de l'étape d'extraction des caractéristiques d'un signal. Il comprend les techniques rencontrées dans la littérature et celles que nous avons proposées dans ce projet.

2.1 Principe de l'extraction des caractéristiques d'un signal

L'extraction des caractéristiques est considérée comme un processus qui transforme des vecteurs de dimension très élevée vers des vecteurs de dimension moins élevée. Donc, c'est une projection $f : \mathbb{R}^N \rightarrow \mathbb{R}^D$, où $D \ll N$. L'extraction des caractéristiques peut être considérée comme un processus de réduction de dimension qui tente de capturer l'essentiel des caractéristiques du signal analysé avec peu de données.

La figure 2.1 illustre très bien ce processus. Le signal est divisé en segments temporels (sans recouvrement) dont chacun contient 512 échantillons. Chaque segment peut, par exemple, être transformé en une représentation à une seule composante (e.g. valeur moyenne des 512 échantillons), deux composantes (e.g. valeur moyenne et valeur maximale), etc. Le *nombre de composantes* choisi correspondra à la *dimension* de l'espace des caractéristiques. Nous verrons, plus loin, différentes techniques permettant d'obtenir ces vecteurs de caractéristiques.

Il est important de garder le nombre de dimensions le plus petit possible afin de pouvoir : 1) créer une *modélisation* robuste d'une classe de signaux et 2) réduire la complexité du calcul. Le concept de *modélisation* de classe sera étudié, en détails, au prochain chapitre. Pour l'instant, on définit une modélisation robuste comme étant une modélisation dont le nombre de données (nombre de segments) servant à la construction du modèle

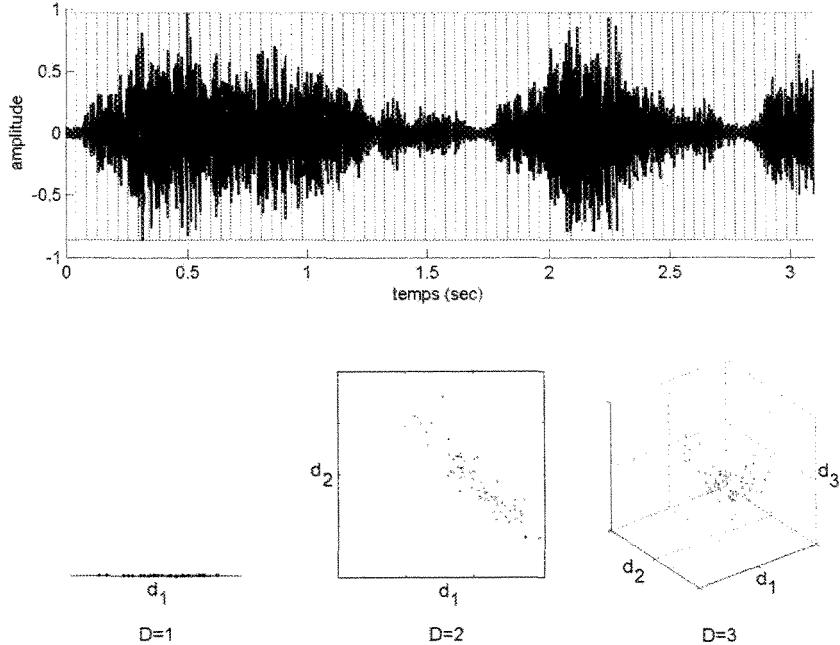


Figure 2.1: *Représentation de la totalité d'un signal dans l'espace de ses caractéristiques en le découplant en segments temporels. La dimension D de l'espace des caractéristiques est ici de un, deux et trois. Chaque point représente les caractéristiques tirées d'un segment. Dans le cas à une dimension, la seul caractéristique tirée d'un segment pourrait, par exemple, être la valeur moyenne du signal dans ce segment.*

est beaucoup plus large que la dimension des données. En effet, le nombre de données d'entraînement nécessaire à la modélisation augmente de façon exponentielle par rapport à la dimension. Ce phénomène est connu sous le nom de “*curse of dimensionality*” [31].

Les caractéristiques optimales, tirées des segments du signal, posséderaient les propriétés suivantes :

- grande variation lorsqu'elles sont issues de différentes classes ;
- variation minime à l'intérieur d'une même classe ;
- facile à mesurer ;
- insensibilité au bruit.

Dans ce chapitre, nous présenterons les techniques d'extraction rencontrées dans la littérature en vue de classifier les sons respiratoires. Nous présenterons aussi les méthodes que nous avons proposées pour caractériser efficacement les signaux respiratoires [29,32].

Ces différentes techniques peuvent être subdivisées en quatre groupes :

- analyse de Fourier ;

- prédition linéaire ;
- analyse cepstrale ;
- transformée en ondelettes.

Pour uniformiser la notation tout au long de ce document, nous allons garder \mathbf{x} pour désigner le vecteur des caractéristiques (figure 1.1) de dimension D.

$$\mathbf{x} = [x_1, x_2, \dots, x_D] \quad (2.1)$$

2.2 Analyse de Fourier

L'analyse de Fourier d'un signal consiste à calculer et représenter son contenu fréquentiel, appelé *spectre*. Pour le cas discret, la Transformée de Fourier Discrète (*DFT - Discrete Fourier Transform*), d'un signal discret $s[n]$, est décrite par l'équation 2.2.

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-\frac{2\pi nk}{N}} \quad k = 0, \dots, N-1 \quad (2.2)$$

où k est l'indice des fréquences et N est le nombre d'échantillons du signal $s[n]$.

L'analyse de Fourier s'avère intéressante dans le cas des sons respiratoires. En effet, la présence de sibilants est caractérisée, le plus souvent, par des pics en fréquence surajoutés aux spectres des sons normaux [22] ou bien par un décalage du spectre vers les fréquences les plus élevées. La figure 2.2 représente le spectrogramme d'un patient normal. Tandis que la figure 2.3 représente celui d'un asthmatique. On remarque que dans le cas du patient asthmatique le spectre s'étant jusqu'à 1200 Hz (figure 2.3) au lieu de 800 Hz (figure 2.2) pour le patient normal. De plus, le sibilant se manifeste par un pic (650 Hz) surajouté au spectre (voir la petite fenêtre de la figure 2.3).

2.2.1 Transformée de Fourier à court terme (*STFT*)

Malgré son succès fulgurant pour analyser une grande variété de signaux, la transformée de Fourier discrète (*DFT*) souffre de deux limitations : 1) elle n'est pas adaptée à l'analyse des signaux non-stationnaires et 2) elle ne donne pas l'instant précis d'un événement. Pour pallier à ces deux limitations, la transformée de Fourier à court terme

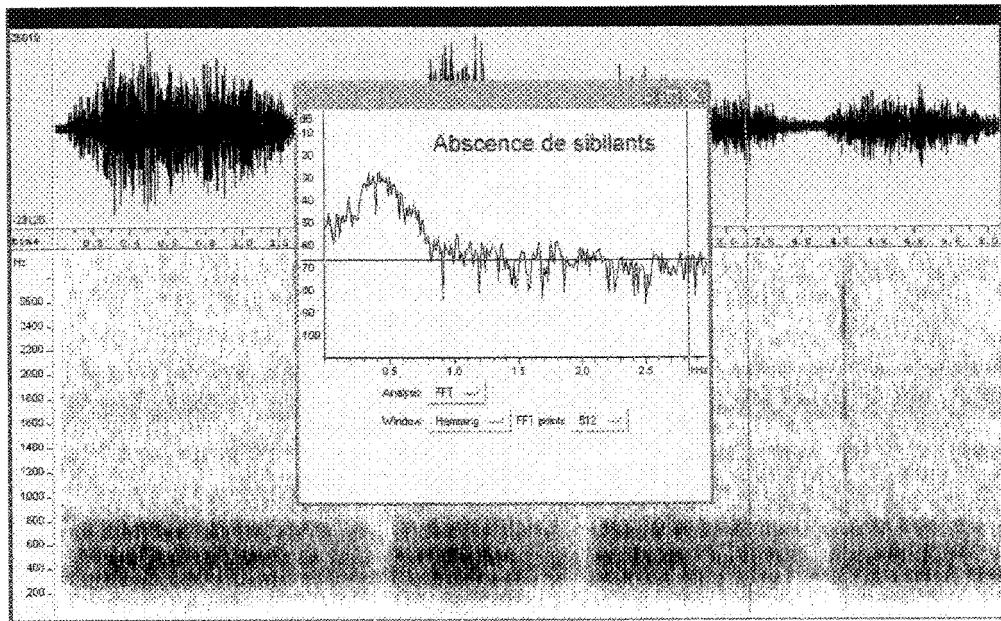


Figure 2.2: Spectrogramme du son respiratoire d'un patient normal obtenu avec le logiciel WaveSurfer [26].

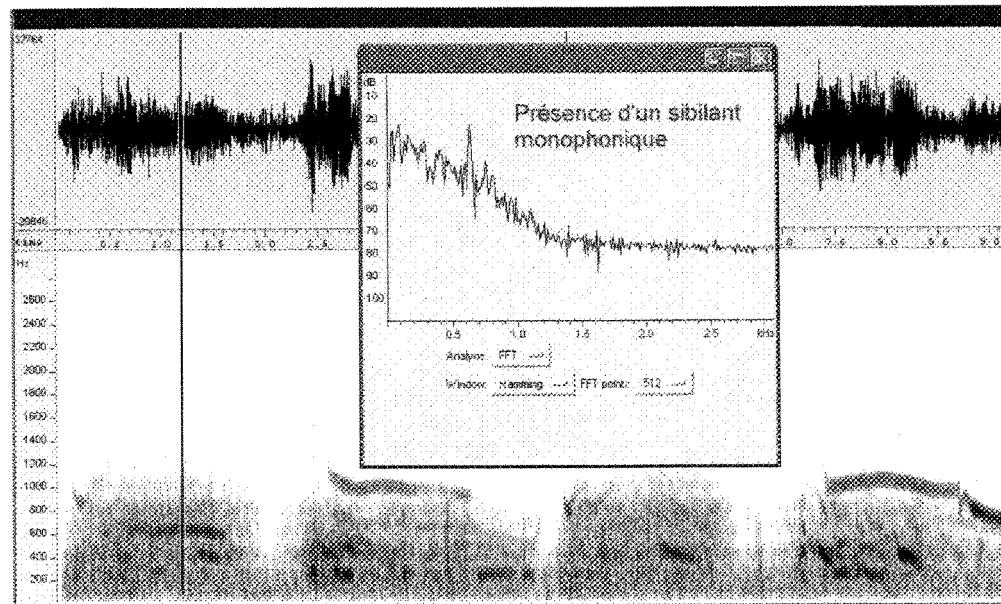


Figure 2.3: Spectrogramme du son respiratoire d'un patient asthmatique obtenu avec le logiciel WaveSurfer [26]. On remarque l'étendu du spectre jusqu'à 1200 Hz et la présence des sibilants autour de 600 Hz.

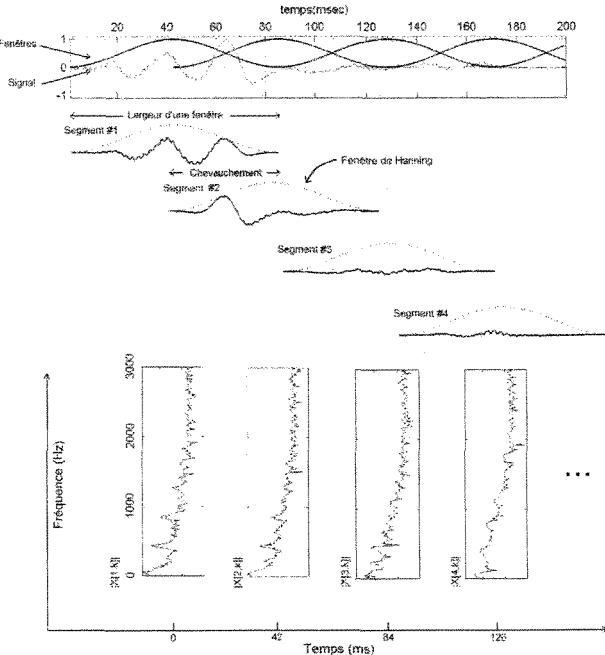


Figure 2.4: *Calcul du spectrogramme d'un son respiratoire par la transformée de Fourier à court terme (STFT).*

(*STFT - Short time Fourier Transform*) a été proposée. Elle consiste à multiplier le signal par une fenêtre glissante $g[n]$, dont la largeur L dépend de la résolution temporelle désirée Δt . Tout au long de l'expérimentation, nous utiliserons une fenêtre de type *Hanning*, tel que celle illustrée à la figure 2.4. Cette figure montre, de plus, le principe de la pondération par la fenêtre $g[n]$ permettant de calculer le spectrogramme d'un signal et suivre son évolution temporelle. La formulation mathématique de la transformée *STFT* est donnée par l'équation 2.3.

$$S_m[k] = \sum_{n=0}^{L-1} s[m+n]g[n]e^{-\frac{2j\pi nk}{L}} \quad n = 0, \dots, L-1 \quad (2.3)$$

où m est l'indice du segment analysé et k est l'indice des fréquences.

La largeur de la fenêtre utilisée dépend des résolutions temporelle ($\Delta t = L/f_e$) et fréquentielle ($\Delta f = f_e/L$) nécessaires à la bonne caractérisation de la forme du signal recherchée. On note que l'amélioration de la résolution temporelle se fait au détriment de

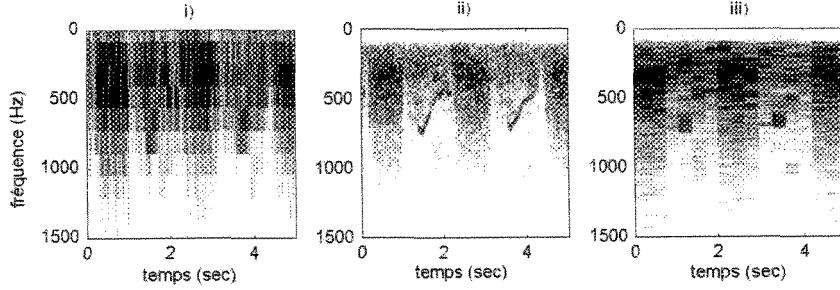


Figure 2.5: *Transformée de Fourier à court terme* i) excellente résolution temporelle $\Delta t = 6.25 \text{ ms}$, $\Delta f = 160 \text{ Hz}$; ii) bon compromis pour les deux résolutions $\Delta t = 62.5 \text{ ms}$, $\Delta f = 16 \text{ Hz}$; iii) excellente résolution fréquentielle $\Delta t = 625 \text{ ms}$, $\Delta f = 1.6 \text{ Hz}$. Le recouvrement est de 50 %.

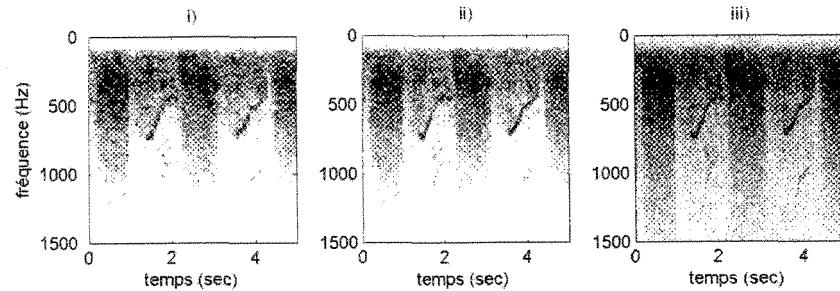


Figure 2.6: *Spectrogrammes à base de segments de 512 échantillons*. Ceci donne les résolutions : $\Delta f = 6000/512 = 11.7 \text{ Hz}$ et $\Delta t = 512/6000 = 85.3 \text{ ms}$. i) recouvrement de 0 % ; ii) recouvrement de 50 % ; iii) recouvrement de 99 %.

la résolution fréquentielle, et vice-versa. Les deux résolutions sont reliées par l'équation 2.4.

$$\Delta f = 1/\Delta t \quad (2.4)$$

Les spectrogrammes présentés à la figure 2.5 illustrent le compromis à faire entre la résolution en temps et la résolution en fréquence. En supposant que les détails du spectrogramme représentent la qualité de la transformée par *DFT*, nous pouvons voir qu'une résolution temporelle de l'ordre de 100 ms donne de meilleurs résultats dans ce cas, pour la détection des sibilants.

La revue de la littérature montre que la détection des sibilants se fait généralement dans la bande entre 100 à 3000 Hz. Nous considérons que la fréquence maximale f_{max} du spectre est de 3000 Hz afin d'inclure quelques harmoniques des sibilants. Pour satisfaire

ce dernier point, nous devons échantillonner le signal à la fréquence f_e de 6000 Hz, selon le critère de Nyquist : $f_e \geq 2 * f_{max}$.

À ce stade ci, il semble qu'un segment temporel de 512 échantillons donnerait un bon compromis entre les deux résolutions Δt et Δf . Un segment de 512 échantillons nous donne les résolutions : $\left(\Delta f = \frac{1}{f_e} = \frac{6000}{512} = 11.7 \text{ Hz} \right)$ et $\left(\Delta t = \frac{1}{\Delta f} = 85.3 \text{ ms} \right)$. Nous étudierons, au chapitre 4, l'effet de la largeur du segment du signal sur les performances du système de classification.

Le parcours progressif des segments temporels qui se recouvrent, tel qu'à la figure 2.4, permet un suivi pseudo-continu de l'évolution du spectre. En effet, le recouvrement augmente la résolution temporelle du spectrogramme (à ne pas confondre avec la résolution temporelle Δt). Il est à noter que la figure 2.5 a été tracée avec un recouvrement de 50%. La figure 2.6 illustre l'importance du recouvrement. Il faut noter, toutefois, que cette opération est pénalisante en terme de temps de calcul. Par exemple, un recouvrement de 50% doublera le nombre de segments à traiter. Nous étudierons l'effet du recouvrement au chapitre 4.

2.2.1.1 Vecteur des caractéristiques tiré de la transformée de Fourier à court terme (*STFT*)

À titre d'exemple, une fenêtre de 85.3 ms d'un signal échantillonné à 6000 Hz correspond à 512 échantillons. Dans ce cas, l'estimation spectrale obtenue par *FFT* sur 512 points donne un spectre non redondant (fréquences positives seulement) de 256 points. Ce nombre élevé de données donne un vecteur de caractéristiques de grande dimension (\mathbb{R}^{256}) qu'il faudra réduire. Rietveld et al. [9] proposent le découpage de la bande de fréquence en un nombre restreint de sous-bandes. Une seule valeur associée à une sous-bande j est obtenue en calculant la moyenne E_j de l'énergie des composantes du spectre contenues dans celle-ci (voir la figure 2.7). La première valeur (rectangle blanc) contient les composantes basse-fréquences du signal qui ne sont pas utilisées. Le vecteur des caractéristiques comprendra l'énergie des N sous-bandes sauf la première qui correspond à la composante continue du signal.

$$\mathbf{x} = [E_2, \dots, E_N] \quad (2.5)$$

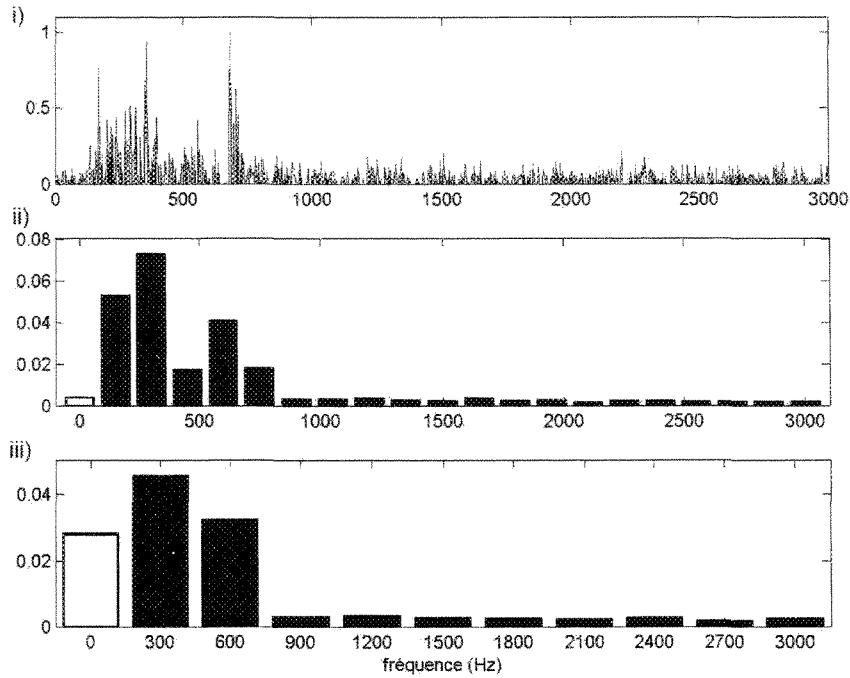


Figure 2.7: i) Spectre original comprenant les 256 valeurs obtenues par STFT sur un segment de 512 échantillons du signal, ii) 20 coefficients obtenus à partir du spectre moyen, iii) 10 coefficients obtenus à partir du spectre moyen. $f_e = 6000$ Hz. Les rectangles blancs contiennent les basses fréquences. Il est à noter que les basses fréquences ne sont pas significatives.

2.3 Analyse Cepstrale

2.3.1 Coefficients Cepstraux

En raison de l'origine de sa production, le son respiratoire $y[n]$ peut-être considéré comme issu de la convolution d'une source sonore $x[n]$, créée par le diaphragme, avec les conduits respiratoires dont la réponse impulsionnelle est $h[n]$.

$$y[n] = x[n] * h[n] \quad (2.6)$$

Étant donné que l'asthme aura pour effet de modifier les propriétés du conduit respiratoire, nous avons proposé l'utilisation de l'analyse cepstrale afin d'isoler la contribution du conduit de celle de la source. La figure 2.8 illustre cette approche de modélisation.

Dans le domaine des fréquences auquel on accède par la transformée de Fourier,

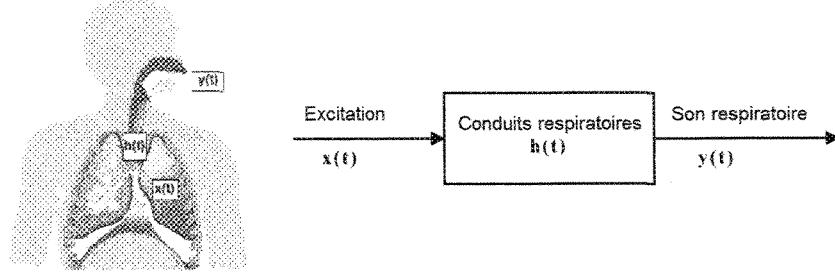


Figure 2.8: *Le son respiratoire est modélisé comme étant la convolution d'une certaine excitation avec les conduits respiratoires.*

l'opération de convolution devient une multiplication.

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega}) \quad (2.7)$$

Il est très difficile de séparer ces deux signaux que ce soit dans le domaine du temps ou celui des fréquences, sans connaissance à priori de $h[n]$ ou $x[n]$. Cependant, la transformation par la fonction logarithme dans le domaine des fréquences va nous permettre d'effectuer la séparation de $h[n]$ et de $x[n]$.

$$\log(Y(e^{j\omega})) = \log(X(e^{j\omega})) + \log(H(e^{j\omega})) \quad (2.8)$$

Pour la plupart des applications du traitement de signal, nous avons seulement besoin que des spectres d'amplitude. Par conséquent l'équation 2.8 s'écrit :

$$\log(|Y(e^{j\omega})|) = \log(|X(e^{j\omega})|) + \log(|H(e^{j\omega})|) \quad (2.9)$$

Les coefficients cepstraux sont alors donnés par la transformée de Fourier inverse où l'indice k correspond aux raies spectrales :

$$cc[k] = DFT^{-1}\log(|X(e^{j\omega})|) + DFT^{-1}\log(|H(e^{j\omega})|) \quad (2.10)$$

Les premières composantes du cepstre représentent la réponse en fréquence des conduits respiratoires. Les autres composantes représentent l'excitation. La figure 2.9 montre un exemple où, on observe que les premières composantes du cepstre, séparées par filtrage

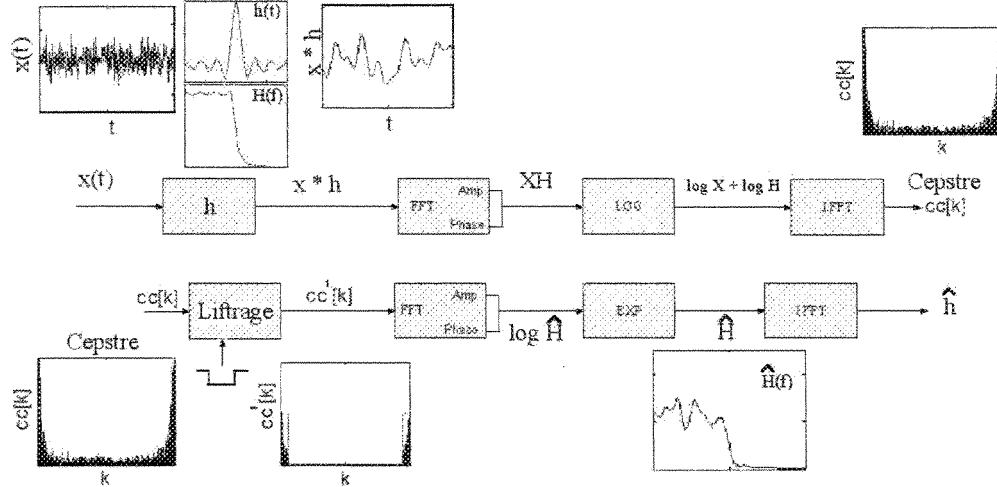


Figure 2.9: Séparation de la réponse en fréquence des conduits respiratoires et de l'excitation par analyse cepstrale. Les premières composantes du cepstre, séparées par filtrage cepstral, représentent la réponse fréquentielle des conduits respiratoires $H(f)$, tel qu'on peut le constater par la ressemblance entre $H(f)$ et $\hat{H}(f)$.

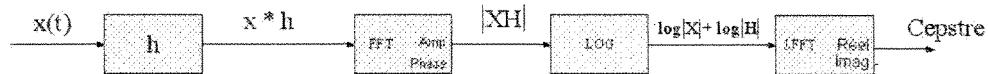


Figure 2.10: Obtention de coefficients cepstraux à partir d'un segment du signal.

cepstrale (aussi appelé liffrage), donnent une approximation grossière de la fonction de transfert du filtre équivalente aux conduits respiratoires.

2.3.1.1 Vecteur des caractéristiques tiré des coefficients cepstraux

Nous rappelons que l'asthme affecte les voies respiratoires en les resserrant par endroits. Pour mettre en évidence cet effet, il ne faut garder que les N premiers coefficients cepstraux correspondants au conduit respiratoire. Le calcul des coefficients cepstraux est illustré à la figure 2.10. Le vecteur des caractéristiques est donné par l'équation suivante :

$$\mathbf{x} = [cc_1, \dots, cc_N] \quad (2.11)$$

où le nombre N des coefficients retenus est déterminé expérimentalement.

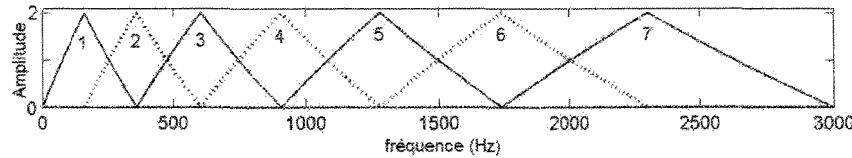


Figure 2.11: *Banc de sept filtres distribués selon l'échelle de Mel.*

2.3.2 Coefficients cepstraux à l'échelle de Mel (MFCC)

Les coefficients cepstraux à l'échelle de Mel (*MFCC - Mel Frequency Cepstrum Coefficients*) constituent un outil de caractérisation intensivement utilisé en reconnaissance automatique de la parole (RAP). Ces coefficients sont dérivés du spectre de puissance en appliquant un banc de filtres uniformément espacés sur une échelle fréquentielle modifiée, appelée échelle de Mel (voir l'exemple de la figure 2.11). L'échelle de Mel redistribue les fréquences en fonction de la fréquence perçue. Elle a été suggérée pour la première fois par Stevens et Volkman ^[33] en 1937 (cité par Shannon et Paliwal ^[34]) L'échelle de Mel est linéaire pour les fréquences inférieures à 1000 Hz et logarithmique pour les fréquences supérieures. Elle peut être approximée par l'équation 2.12.

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.12)$$

où f est la fréquence dans l'échelle linéaire et f_{mel} la fréquence perçue.

La figure 2.12 montre le principe du calcul des coefficients *MFCC*. Le spectre $X[k]$ d'un signal fenêtré $x[n]$ est calculé par la transformée de Fourier discrète (*DFT*). L'amplitude est pondérée par une série de réponses fréquentielles de L filtres. Le cepstre réel associé à l'énergie résultante $E[k]$ est appelé *MFCC* et est calculé par la formule suivante.

$$c[n] = \sum_{k=1}^L \log(E[k]) \cos \left(n(k - 0.5) \frac{\pi}{L} \right) \quad n = 0, 1, \dots, L - 1 \quad (2.13)$$

où L est le nombre de filtres distribués selon l'échelle de Mel.

Le choix d'utiliser les coefficients *MFCC* pour caractériser les sons respiratoires ^[32] nous a été motivé par le fait qu'ils s'inspirent du fonctionnement du système auditif et qu'ils décorrèlent les conduits respiratoires de l'excitation. L'algorithme est décrit en détails dans la section suivante.

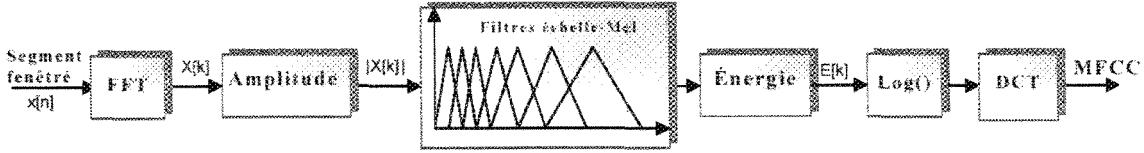


Figure 2.12: Obtention des coefficients MFCC à partir d'un segment du signal.

2.3.2.1 Vecteur des caractéristiques tiré des coefficients MFCC

La figure 2.13 illustre les différentes étapes de l'obtention de sept coefficients MFCC pour le signal $s(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t)$. Nous avons choisi ce signal afin de bien illustrer l'effet du banc de filtres. Nous pouvons voir :

- spectre d'un segment du signal, figure 2.13-i;
- banc de filtres distribués selon l'échelle perceptuelle de MEL, figure 2.13-ii;
- pondération du spectre par chacun des filtres, figure 2.13-iii ;
- logarithme de l'étape précédente, figure 2.13-iv ;
- transformée en cosinus discrète (*DCT - Discrete Cosine Transform*) de l'étape précédente, figure 2.13-v. La première valeur n'est pas utilisée car elle représente la valeur moyenne du signal de la figure 2.13-iv.

Le vecteur des caractéristiques à base de L coefficients MFCC est ainsi donné par :

$$\mathbf{x} = [c_1, \dots, c_{L-1}] \quad (2.14)$$

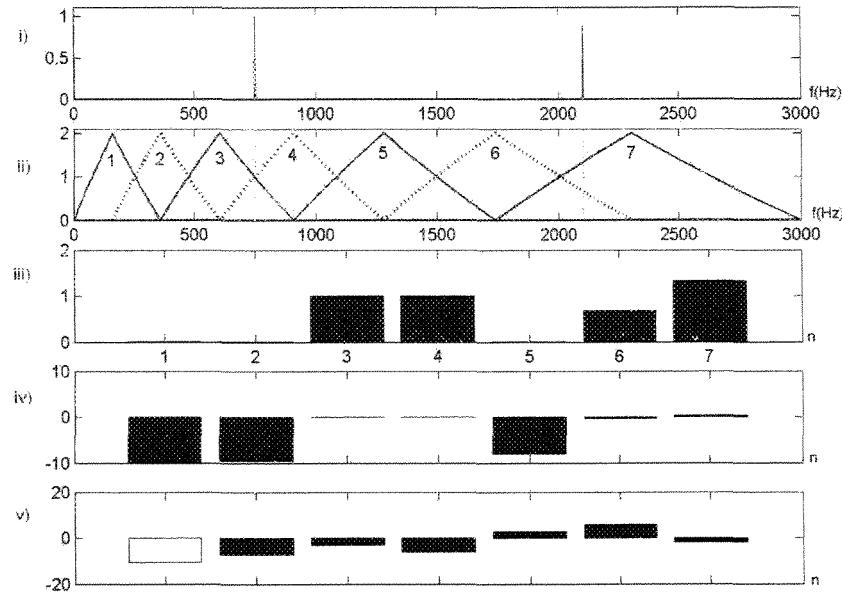


Figure 2.13: *Démarche suivie pour l'obtention de sept coefficients MFCC. i) valeur absolue de la STFT élevée au carré , ii) banc de filtres positionnés selon l'échelle de MEL, iii) multiplication de i par ii. La première raie spectrale n'affecte que les filtres 3 et 4, alors que l'autre affecte les filtres 6 et 7, iv) logarithme de iii, v) transformée discrète en cosinus de iv, le premier terme est éliminé car il correspond à la valeur moyenne du signal en iv.*

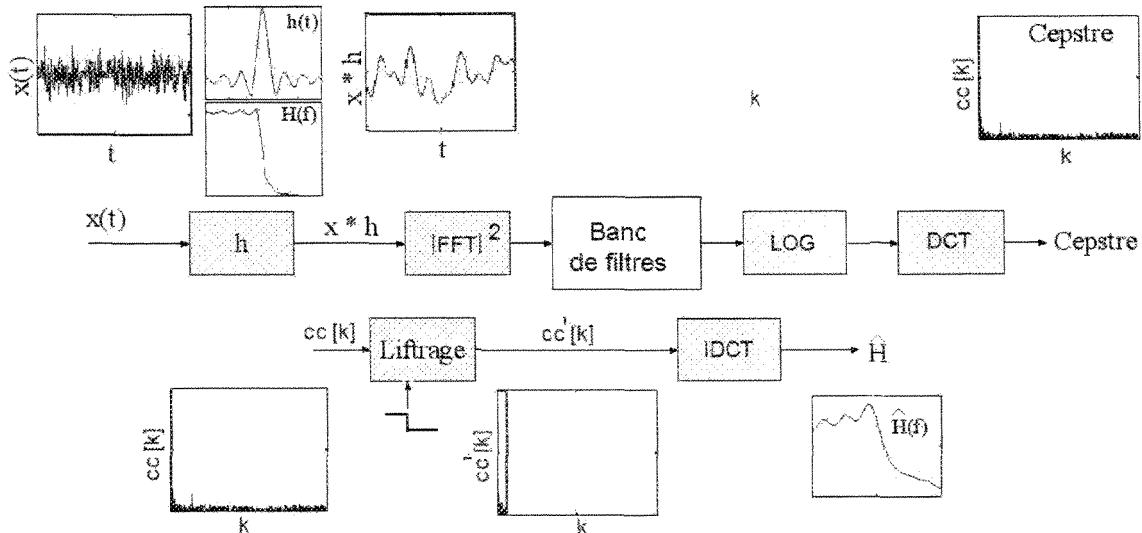


Figure 2.14: *Obtention de coefficients cepstraux par utilisation de la transformée DCT. Le Banc de filtres distribués selon l'échelle de MEL n'a pas été utilisée dans cette figure. La transformée DCT peut être utilisée pour obtenir des coefficients cepstraux permettant la séparation de l'excitation $x(t)$ et des conduits respiratoires $h(t)$. On remarque bien la ressemblance entre $H(f)$ et $\hat{H}(f)$.*

2.3.2.2 Obtention de coefficients cepstraux à partir de la transformation en cosinus discrète(*DCT*)

La figure 2.14 démontre qu'une transformée en cosinus discrète (*DCT*) permet d'obtenir des coefficients cepstraux caractérisant les conduits respiratoires. Le schéma de cette approche est tout à fait différent de celui de la figure 2.9.

2.4 Modélisation linéaire prédictive (*LPC*)

2.4.1 Coefficients de la prédiction linéaire

La modélisation linéaire prédictive (*LPC* - *Linear Predictive Coding*) suppose que chaque échantillon du signal peut être approximé par une combinaison linéaire d'un nombre réduit des échantillons précédents [35] :

$$\hat{s}[n] = \sum_{k=1}^P a_k s[n-k] \quad n = 1, 2, \dots, N \quad (2.15)$$

où $\hat{s}[n]$ est la prédiction du signal $s[n]$ et le vecteur $a = [a_1, \dots, a_P]$ forme les coefficients de la prédiction d'ordre P . L'erreur de prédiction $e[n]$ pour le $n^{\text{ème}}$ échantillon $s[n]$ est donnée par la différence entre l'échantillon actuel et sa valeur prédite.

$$e[n] = s[n] - \sum_{k=1}^P a_k s[n-k] \quad (2.16)$$

La méthode d'*autocorrelation* et la méthode de *covariance* sont les deux méthodes standards pour calculer les coefficients de prédiction. Les deux techniques sont basées sur la minimisation de la moyenne quadratique de l'erreur de prédiction.

$$E = \sum_{n=1}^N e[n]^2 = \sum_{n=1}^N \left(s[n] - \sum_{k=1}^P a_k s[n-k] \right)^2 \quad (2.17)$$

$$\frac{\partial E}{\partial a_k} = 0, \quad k = 1, \dots, P \quad (2.18)$$

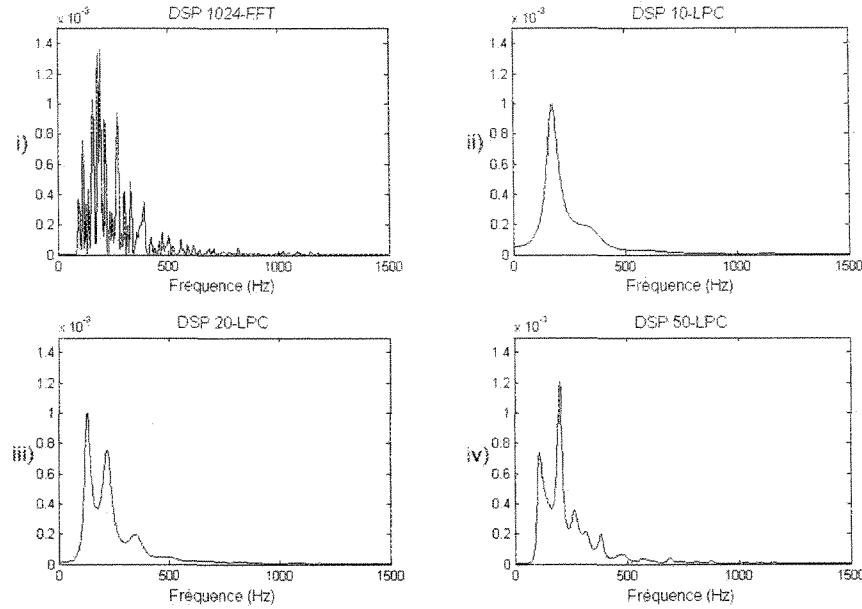


Figure 2.15: *Estimation spectrale d'un son normal en utilisant i) la méthode de Welsh à 1024-FFT, ii-iv) la méthode de Yule-Walker respectivement à 10, 20 et 50 coefficients LPC.*

Dans ce projet, nous utilisons la technique récursive de Levinson-Durbin afin de déterminer les coefficients *LPC*.

La figure 2.15 représente le spectre d'un son normal par la transformée *FFT* (figure 2.15-i) et par la modélisation *LPC* (figures 2.15-ii-iv). Il convient de souligner que la résolution fréquentielle dans le cas de l'estimation par *LPC* du spectre dépend de l'ordre du modèle et non pas de la durée du signal comme c'est le cas de l'estimation par la transformée *FFT*. À l'instar de l'analyse cepstrale, l'approche *LPC* adoucit le spectre du signal. Ceci est équivalent à une approximation de la fonction de transfert du conduit respiratoire.

2.4.1.1 Vecteur des caractéristiques tiré des coefficients *LPC*

Les N premiers éléments a_k , qui en quelque sorte modélisent le conduit, seront utilisés comme vecteur des caractéristiques. Nous y ajouterons aussi la variance de l'erreur de prédiction indiqué par E qui est obtenue également par l'algorithme de Levinson-Durbin.

$$\mathbf{x} = [a_1, a_2, a_3, \dots, a_N, E] \quad (2.19)$$

2.5 Analyse par ondelettes

La transformée de Fourier à court terme (*STFT*) a été proposée pour permettre l'analyse en temps et en fréquence d'un signal (calcul de spectrogramme). Cependant, cet outil souffre d'une limitation liée à la rigidité de ses résolutions temporelle (Δt) et fréquentielle (Δf). La figure 2.16-i montre un exemple de résolutions figées obtenues par la transformée *STFT*.

La transformée en ondelettes est un excellent outil d'analyse en temps et en fréquence proposé pour pallier aux inconvénients de la transformée *STFT*, en permettant d'obtenir des résolutions temporelle et fréquentielle variables. Tel qu'illustré à la figure 2.16-ii, aux hautes fréquences, la transformée en ondelettes donne une bonne résolution temporelle associée à une mauvaise résolution fréquentielle, alors qu'aux basses fréquences, elle donne une bonne résolution fréquentielle associée à une mauvaise résolution temporelle. Cette aptitude à modifier la résolution, permet d'obtenir une représentation différente de celle obtenue par *STFT*. À ce stade, nous ne pouvons pas conclure que l'une ou l'autre des méthodes représente mieux le signal analysé. Toutes les analyses par ondelettes ont été effectuées à l'aide des fonctions du logiciel, du domaine public, *WaveLab* [36].

2.5.1 Transformée en ondelettes continue (*CWT*)

La transformée en ondelettes continue (*CWT* - *Continuous Wavelet Transform*) d'un signal $x(t)$ est définie par :

$$CWT_x^\psi(a, b) = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}^*(t) dt \quad (2.20)$$

où $*$ désigne le complexe conjugué, le terme $a \in \mathbb{R}^+$ représente le paramètre échelle, et le terme $b \in \mathbb{R}$ représente le décalage temporel. Les fonctions $\psi_{a,b}^*$ sont obtenues par le décalage temporel d'une version compressée/décalée de l'ondelette mère (ou analysante) $\psi(t)$, et sont définies par :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \quad (2.21)$$

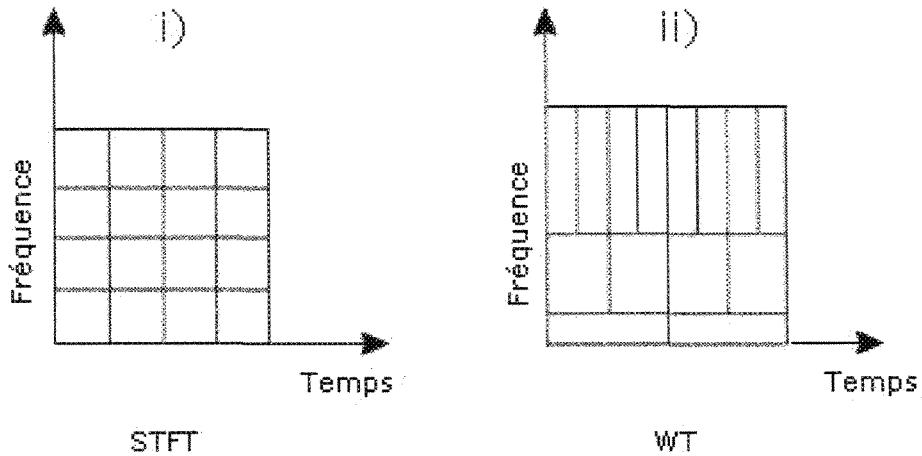


Figure 2.16: *Résolutions en temps et en fréquence obtenues par : i) transformée de Fourier à court terme (STFT) et ii) transformée en ondelettes (CWT). Il est à noter que la transformée en ondelettes donne une meilleure résolution temporelle en hautes fréquences et une meilleure résolution fréquentielle en basses fréquences.*

2.5.2 Transformée en ondelettes discrète (DWT)

La transformée en ondelettes discrète (DWT - *Discrete Wavelet Transform*) est la version discrète de la transformée en ondelettes continue (CWT). Elle est implantée à l'aide de l'algorithme de Mallat [37] en utilisant l'analyse à multirésolution. Cet algorithme est basé sur la définition d'une paire de filtres H et G , appelés aussi filtres à miroirs quadratiques (*QMF*), et dont les réponses impulsionales h et g doivent satisfaire certaines conditions. Plusieurs types d'ondelettes sont utilisés dans la littérature : *Haar, Beylkin, Coiflet, Daubechies, Symmlet, Vaidyanathan, Battle*, etc. Dans ce travail nous utiliserons seulement les ondelettes de *Daubechies* à dix moments nuls (*vanishing moments*).

Les filtres H et G constituent respectivement, un filtre passe-bas et un filtre passe-haut ayant la même fréquence de coupure (leurs bandes de fréquence sont complémentaires). Les sorties des filtres sont sous-échantillonnées par un facteur 2. Le filtre passe-haut fournit les coefficients de la décomposition en ondelettes discrète (DWT) ou les *détails* du signal à une échelle donnée. Le filtre passe-bas donne les coefficients de l'*approximation* du signal à la même échelle. On applique de nouveau, la même opération à l'approximation générant ainsi un autre détail et une nouvelle approximation [14]. Cet algorithme est illustré, de deux façons complémentaires, à la figure 2.17. Cette analyse est équivalente à

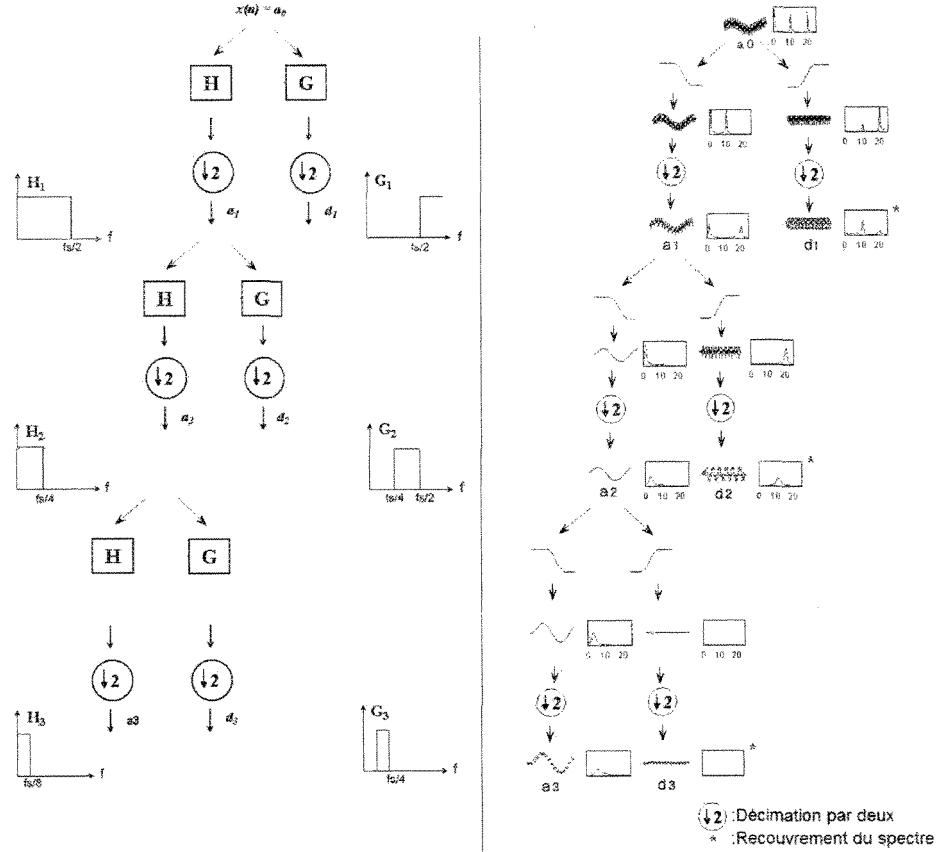


Figure 2.17: Deux façons complémentaires de représenter l'algorithme de Mallat pour l'implantation de la décomposition DWT. H est un filtre passe-bas et G un filtre passe-haut. Il est à noter que la décimation par deux double la largeur du spectre. Les coefficients de la décomposition DWT sont composés des coefficients d_1 (64 coefficients), d_2 (32 coefficients), d_3 (16 coefficients) et a_3 (16 coefficients), pour un total de 128 coefficients soit le nombre d'échantillons du signal.

une décomposition par une batterie de filtres passe-bandes dont les fréquences principales et les bandes passantes sont des sous-multiples de la fréquence d'échantillonnage [14]. Le diagramme gauche de la figure 2.17 est appelé aussi arbre de décomposition.

La figure 2.18 montre un exemple de la distribution temps-échelle du signal $s(t) = \sin(2\pi 0.5t) + \sin(2\pi 10t) + \sin(2\pi 20t)$ à l'aide d'une décomposition DWT de profondeur ($j = 3$). Chaque sous-bande représente l'évolution temporelle (amplitude codée en niveaux de gris) de la totalité du signal dans une plage de fréquence donnée.

La résolution en fréquence Δf est fonction de la profondeur (niveau j) de la décomposition,

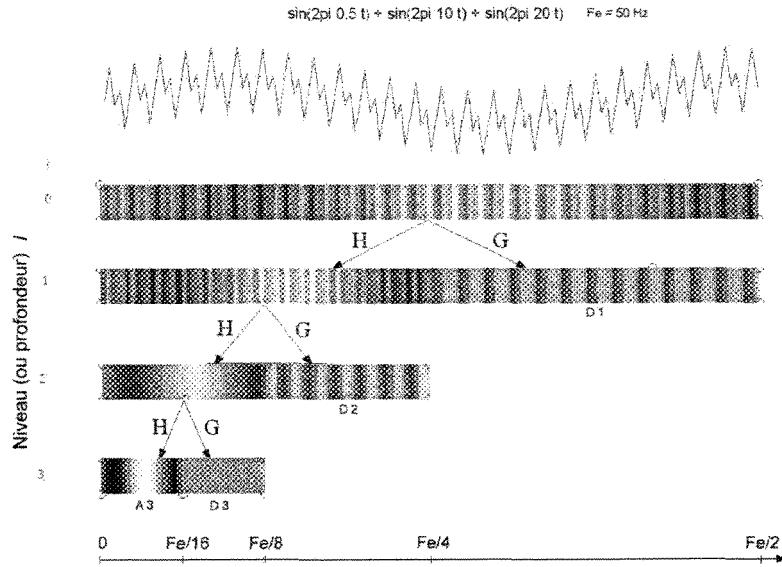


Figure 2.18: *Exemple de décomposition DWT.* Il est à noter que chaque sous-bande représente la totalité du signal temporel. Ceci est particulièrement apparent dans la sous-bande A_3 où l'on perçoit le creux de la composante $\sin(2\pi 0.5t)$ du signal sous forme de zone blanche. La composante $\sin(2\pi 10t)$ est détectée dans la sous-bande D_2 alors que la composante $\sin(2\pi 20t)$ apparaît dans la sous-bande D_1 .

selon l'équation 2.22, tandis que, la résolution temporelle Δt est donnée par l'équation 2.23.

$$\Delta f = \frac{f_e/2}{2^j} \quad (2.22)$$

$$\Delta t = \frac{2^j}{f_e/2} \quad (2.23)$$

où f_e est la fréquence d'échantillonnage et j est la profondeur de l'arbre de décomposition.

2.5.2.1 Vecteur des caractéristiques tiré des coefficients de la transformée en ondelettes discrète DWTC

L'utilisation des coefficients de la transformée en ondelettes discrète (DWTC - *Discrete Wavelet Transform Coefficients*) comme caractéristiques a été proposé par Kandaswamy et al. [12]. Ces auteurs ne s'intéressent qu'aux fréquences comprises entre 50 Hz et 1000 Hz. Étant donné que notre fréquence d'échantillonnage diffère de celle de leur article, nous avons dû sélectionner les sous-bandes qui comprennent les fréquences de 46.87 Hz à

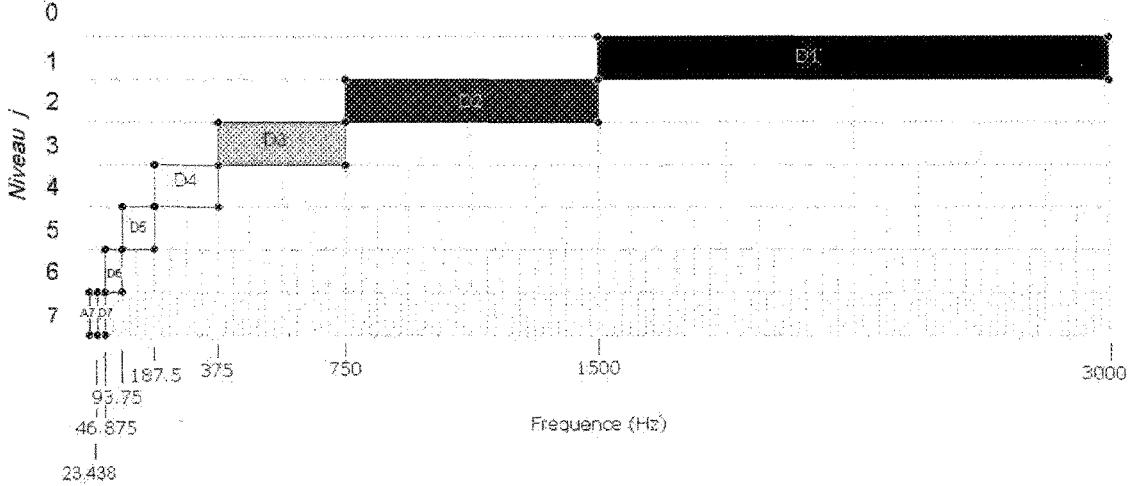


Figure 2.19: *Moyenne des coefficients obtenus par la transformation DWT à $fe=6000$ Hz. Les niveaux de gris indique la valeur relative de ces moyennes : noir étant une valeur nulle et blanc étant la valeur maximale.*

1500 Hz. La profondeur requise afin de récupérer le minimum de 46.87 Hz est de $j = 6$. Voir la figure 2.19, pour plus de détails.

Le vecteur des caractéristiques est composé des dix-neuf coefficients suivants :

- la moyenne (μ) de la valeur absolue des coefficients de D2, D3, D4, D5 et D6 ;
- la puissance moyenne (P) des coefficients de D2, D3, D4, D5 et D6 ;
- l'écart type (σ) des coefficients de D2, D3, D4, D5 et D6 ;
- le rapport des valeurs moyennes (R_μ) des sous-bandes adjacentes (D2, D3), (D3, D4), (D4, D5) et (D5, D6).

$$\mathbf{x} = [\mu_{D_i}, P_{D_i}, \sigma_{D_i}, R_\mu(D_i/D_{i+1})] \quad i = 2, \dots, 6 \quad (2.24)$$

2.5.3 Transformée en paquets d'ondelettes (WPT)

La transformée en paquets d'ondelettes (*WPT - Wavelet Packet Transform*), proposée par Coifman et Wickerhauser [38], est une généralisation de la transformée en ondelettes discrète (*DWT*). En effet, les filtres H et G sont appliqués simultanément à l'approximation et au détail de la résolution précédente, tel qu'illustré à la figure 2.20.

Les données à la sortie d'un déimateur forment un sous-groupe de coefficients d'ondelettes. Chaque sous-groupe est équivalent à la sortie d'un filtre passe-bande dont la

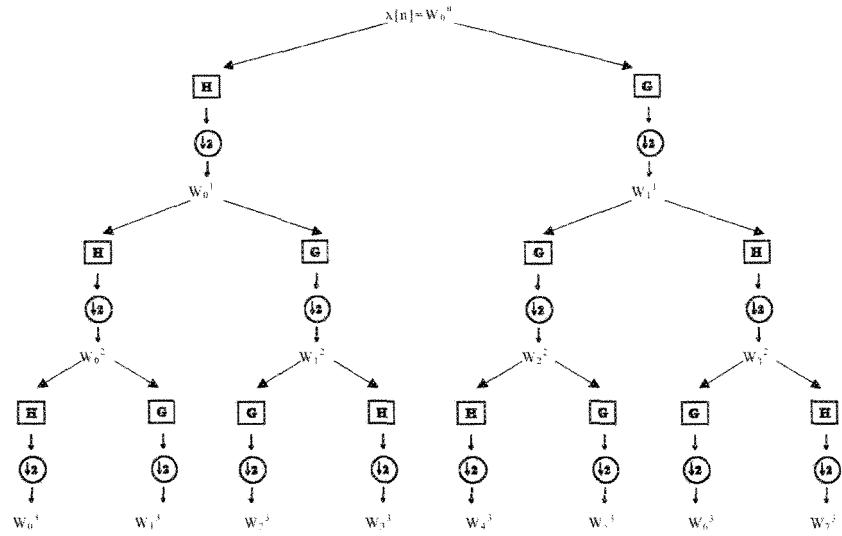


Figure 2.20: *Transformée en paquets d'ondelettes, où W_k^j représente les coefficients de la $k^{\text{ème}}$ sous-bande au niveau j .*

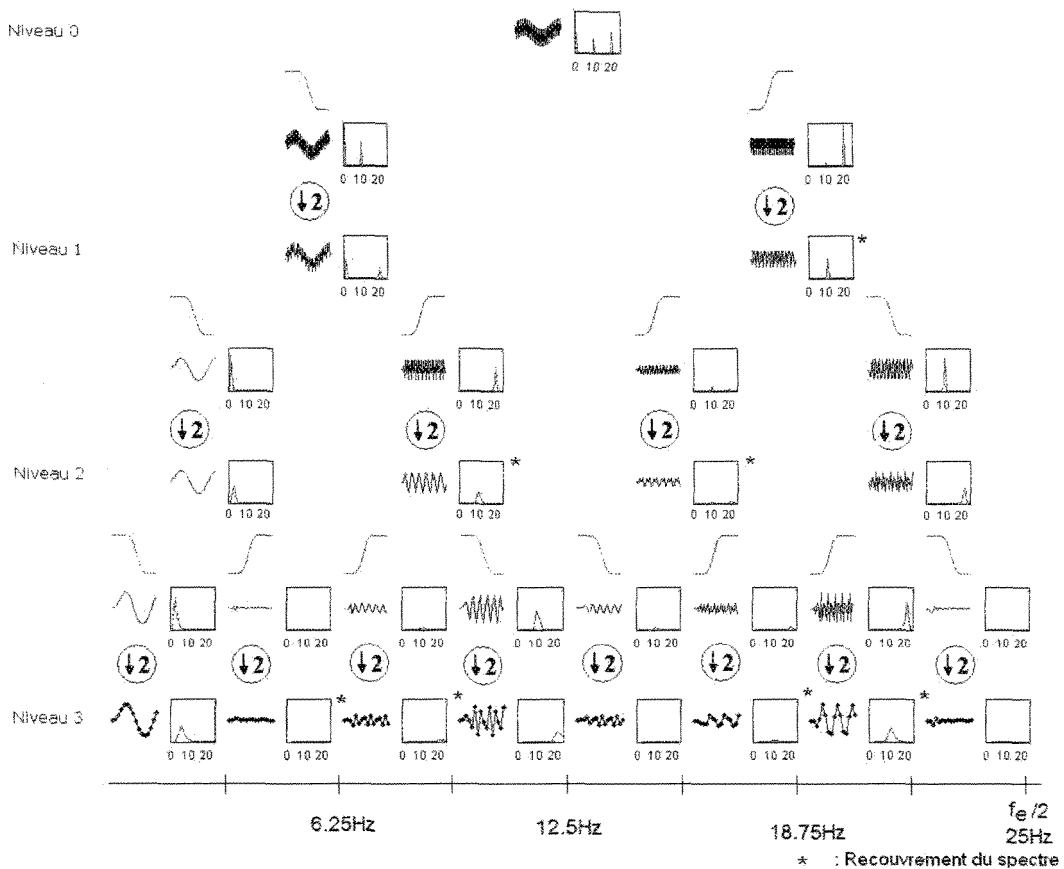


Figure 2.21: *Illustration en temps et en fréquence de la transformée en paquets d'ondelettes (WPD). Le recouvrement du spectre indiqué par * force un pivotement des filtres H et G. Les trois composantes du signal original $s(t) = \sin(2\pi 0.5t) + \sin(2\pi 10t) + \sin(2\pi 20t)$ apparaissent principalement à W_0^3 (0-3.125 Hz), W_3^3 (9.375-12.5 Hz) et W_6^3 (18.75-21.875 Hz).*

fréquence de coupure dépend de sa position dans l'arbre de décomposition (voir la figure 2.21). La meilleure résolution temporelle est obtenue au niveau $j = 0$, alors que la meilleure résolution en fréquence se trouve au dernier niveau, en l'occurrence pour $j = 3$.

Wickerhauser [39] a démontré que les sous-bandes de l'arbre des paquets d'ondelettes sont organisées d'une manière particulière qui n'est pas dans l'ordre naturel croissant des fréquences. L'analyse des sous-bandes nécessite un prétraitement, qui consiste à réarranger les filtres suivant la réorganisation dans l'ordre naturel [14]. Une autre façon de le voir est qu'il y a pivotement des filtres *QMF* à chaque fois que la décimation entraînera un recouvrement des spectres. Ces cas de pivotement sont indiqués par un astérisque à la figure 2.21.

2.5.3.1 Recherche de la meilleure base

L'analyse en paquets d'ondelettes, divise l'axe des fréquences en tranches régulières mais variant en fonction de la profondeur j de l'arbre de décomposition [14]. La base optimale ou "la meilleure base" peut être interprétée comme la meilleure segmentation de l'axe des fréquences [14]. Selon le sens du parcours de l'arbre dyadique (figure 2.22), deux techniques sont utilisées pour trouver la meilleure base. La première de type descendante est appelée méthode de division (*spliting*). La seconde de type ascendante est appelée méthode de regroupement (*merging*).

Dans la technique ascendante, on commence par une segmentation arbitraire de la sous-bande de fréquence en intervalles dyadiques de taille réduite, afin de prendre en compte les détails les plus fins apparaissant dans le signal. L'algorithme consiste alors à remplacer les intervalles contigus \mathbb{I}_k^j et \mathbb{I}_{k+1}^j (intervalles fils) par l'union des deux intervalles $\mathbb{I}_k^j \oplus \mathbb{I}_{k+1}^j$ (intervalle père), selon un critère d'optimisation. Le critère le plus utilisé dans la littérature est *l'entropie* du spectre [40], montré à l'équation 2.25 :

$$H(\mathbb{I}_k^j) = - \sum_{m=1}^{N_k} |w_{k,m}^j|^2 \log |w_{k,m}^j|^2 \quad (2.25)$$

où $w_{k,m}^j$ est le $m^{\text{ième}}$ coefficient de la $k^{\text{ième}}$ sous-bande au niveau j et où N_k est le nombre de coefficients dans la sous-bande.

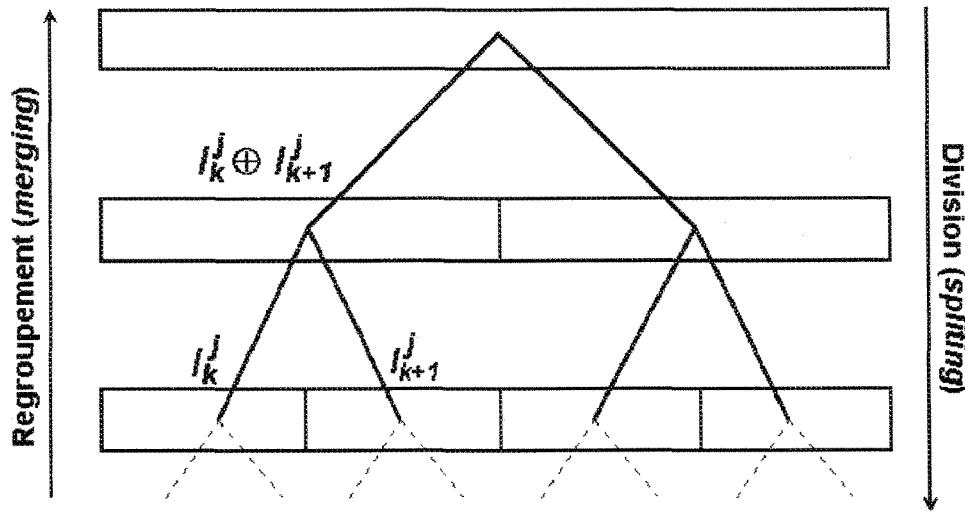


Figure 2.22: Recherche de la meilleure base de décomposition suite à la décomposition en paquets d'ondelettes.

Si l'entropie des coefficients de l'intervalle union (le père) est inférieur à la somme des entropies des deux intervalles contigus (les fils) alors l'intervalle père est retenu, sinon on garde les intervalles fils.

$$H(\mathbb{I}_k^j \oplus \mathbb{I}_{k+1}^j) \leq H(\mathbb{I}_k^j) + H(\mathbb{I}_{k+1}^j) \quad (2.26)$$

La figure 2.23-i, à titre d'exemple, montre sous forme de rectangle blanc, les meilleures sous-bases obtenues suite à la recherche de la meilleure base.

2.5.3.2 Vecteur des caractéristiques tiré des meilleurs coefficients d'ondelettes (*WBC*)

L'utilisation comme caractéristiques des meilleurs coefficients d'ondelettes (*WBC* - Wavelet Best Coefficient) a été proposée par Pesu et al. [8, 41]. Le vecteur des caractéristiques s'obtient en effectuant les étapes suivantes, pour chaque segment du signal :

- effectuer la décomposition en paquets d'ondelettes. Ceci donne comme exemple, les niveaux de gris de la figure 2.23-i;
- effectuer la recherche de la meilleure base. Ceci donnera les sous-bandes représentées par les rectangles blancs à la figure 2.23-i;
- calculer une valeur d'entropie associée à chacune des sous-bandes. Ceci est illustré

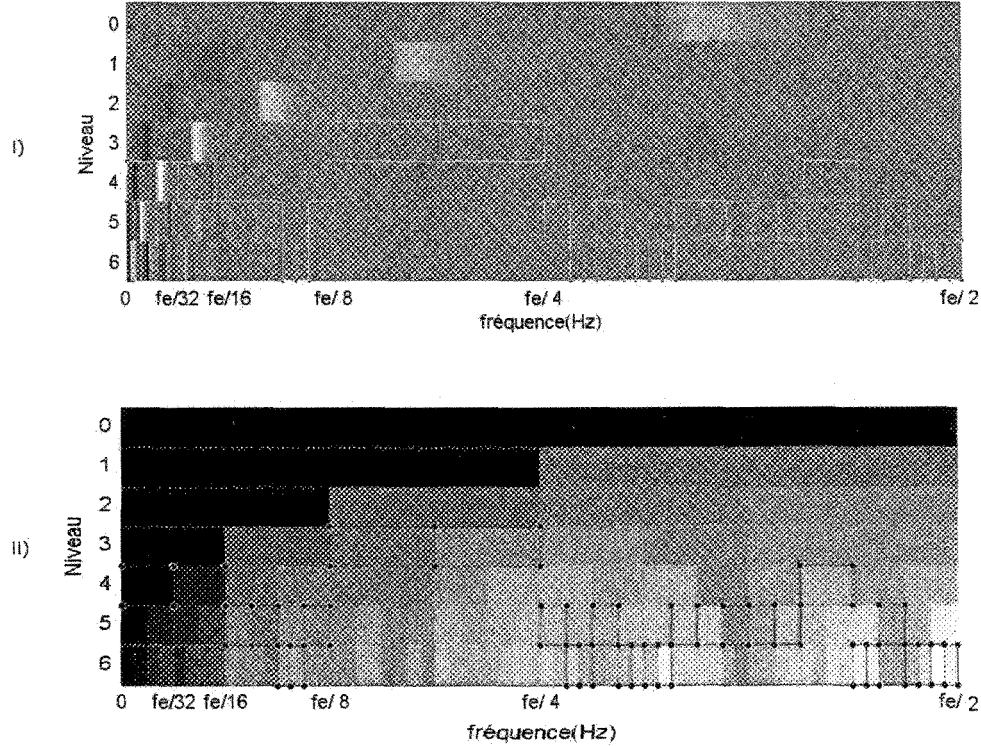


Figure 2.23: Caractéristiques à base des meilleurs coefficients d'ondelettes WBC d'un segment de signal d'un son respiratoire (1 de 2). i) meilleure base de décomposition indiquée par les rectangles blancs. ii) indice de l'entropie pour chaque sous-bande (Il est à noter qu'un ton foncé indique une plus grande entropie).

par un ton de gris pour chacune des sous-bandes à la figure 2.23-ii; Par la suite, on ne s'intéressera qu'aux sous-bandes obtenues suite à la recherche de la meilleure base, tel qu'illustré à la figure figure 2.24-i.

- établir sept zones situées dans la plage de fréquence à investiguer, soit de 187.5 Hz à 1500 Hz. Celles-ci sont illustrées à la figure 2.24-i par des lignes pointillées. Nous nous intéresserons qu'aux sous-bandes situées dans ces zones ;
 - déterminer sept nouvelles valeurs d'entropie associées aux sept zones de fréquence à partir de celles des sous-bandes se trouvant dans ces zones. Selon la base de décomposition, trois cas peuvent se présenter. Nous représenterons la longueur d'une zone ou d'une sous-bande par les symboles \blacktriangleleft \triangleright .
- 1) $(\blacktriangleleft \text{ zone } \triangleright) = (\blacktriangleleft \text{ sous-bande } \triangleright)$: la valeur d'entropie de la zone est égale à la valeur d'entropie de la sous-bande, tel que le niveau $j = 4$ de la figure 2.24-i ;
 - 2) $(\blacktriangleleft \text{ zone } \triangleright) < (\blacktriangleleft \text{ sous-bande } \triangleright)$: la valeur d'entropie de la zone est égale

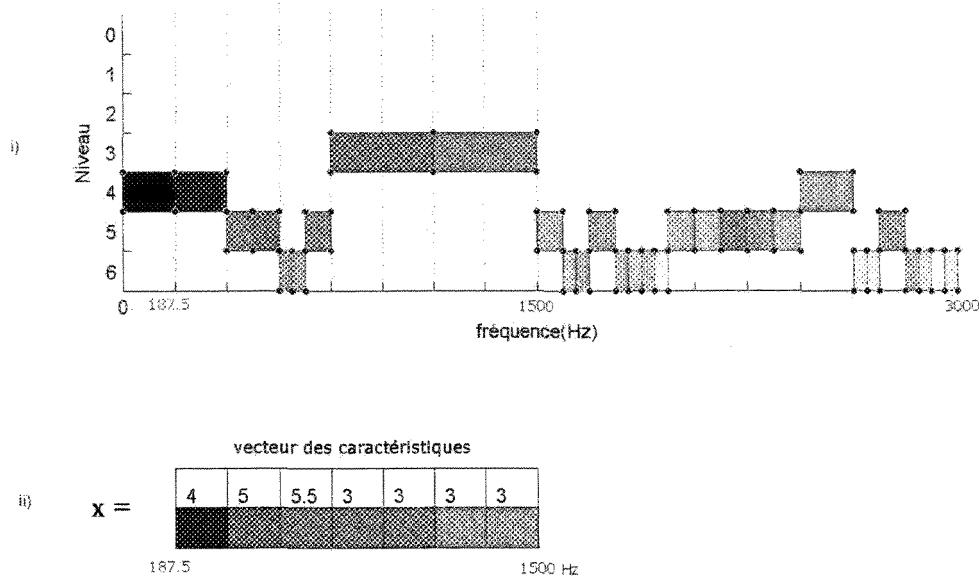


Figure 2.24: Caractéristiques à base des meilleurs coefficients d'ondelettes d'un segment d'un son respiratoire (2 de 2). i) valeurs d'entropie pour la meilleure base. On ne s'intéressera qu'aux valeurs d'entropie situées dans la plage de fréquence située dans les zones pointillées. ii) vecteur des caractéristiques comprenant, pour chaque zone, le niveau où les sous-bandes sont situées ainsi qu'une valeur calculée à partir des valeurs d'entropie des sous-bandes.

à la valeur d'entropie de la sous-bande divisée par le nombre de zones dans la sous-bande, tel que le niveau $j = 3$ de la figure 2.24-i ;

- 3) (\blacktriangleleft zone \blacktriangleright) $>$ (\blacktriangleleft sous – bande \blacktriangleright) : la valeur d'entropie de la zone est égale à la somme des valeurs d'entropie de toutes les sous-bande de la zone, tel que les niveaux $j = 5$ et $j = 6$ de la figure 2.24-i.

- obtenir le vecteur des caractéristiques composé des sept nouvelles valeurs d'entropie, ainsi que des sept niveaux où les sous-bandes étaient situées. (voir la figure 2.24-ii, pour plus de détails).

2.5.3.3 Vecteur des caractéristiques tiré de la méthode *SBC*

La méthode (*SBC - Subband Based Coefficients*) a été proposée initialement par Sarikaya et al. [42] pour caractériser le signal de parole dans le cadre de l'identification des locuteurs. Nous avons appliqué cette approche, avec succès, dans ce projet de re-

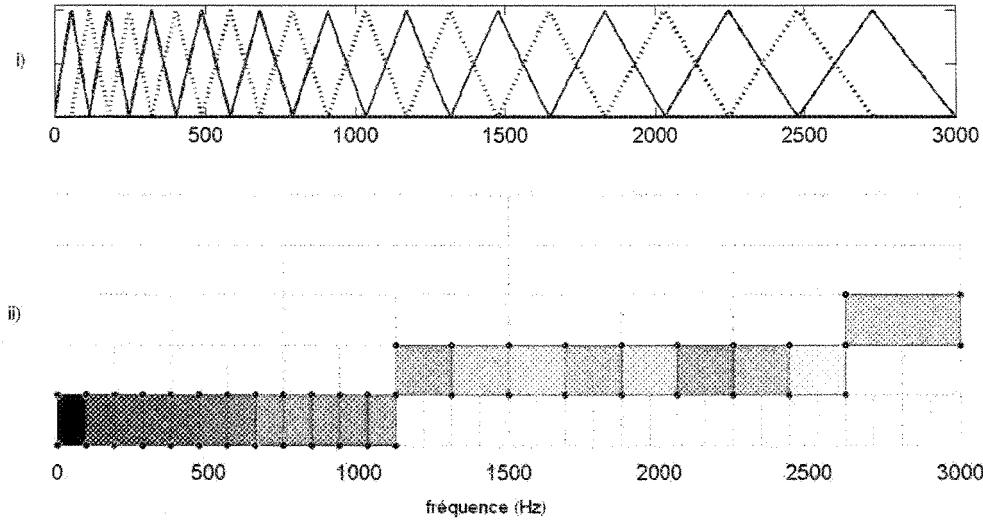


Figure 2.25: *Principe de la caractérisation par la méthode SBC. i) banc de filtres à l'échelle de Mel que l'on cherche à émuler. ii) base de décomposition des paquets d'ondelettes correspondant le mieux à ces filtres.*

cherche [29, 30]. Cette méthode consiste à émuler le banc de filtres de MEL utilisé lors du calcul des coefficients *MFCC* par une décomposition adéquate par paquets d'ondelettes, tel qu'illustré à la figure 2.25. Le diagramme de la figure 2.26 donne le principe de calcul des coefficients *SBC* dans le cas de 24 sous-bandes. La décomposition par paquets d'ondelettes perceptuelle est calculée selon l'arbre prédéfini. L'énergie de chaque sous-bande est calculée puis moyennée par le nombre de coefficients de cette sous-bande, selon l'équation 2.27.

$$S_k = \frac{1}{N_k} \sum_{m=1}^{N_k} [w_{k,m}^j]^2 \quad (2.27)$$

où $w_{k,m}^j$: le $m^{\text{ième}}$ coefficient d'ondelette de la $k^{\text{ième}}$ sous-bande du niveau j ,

k : index de la sous-bande fréquentielle, et

N_k : nombre de coefficients de la $k^{\text{ième}}$ sous-bande.

Finalement, les coefficients *SBC* sont déduits en appliquant la transformée en cosinus aux énergies des différentes sous-bandes, selon l'équation 2.28.

$$SBC(n) = \sum_{k=1}^L \log(S_k) \cos \left(n(i - 0.5) \frac{\pi}{L} \right) \quad (2.28)$$

où $n = 0, \dots, L - 1$.

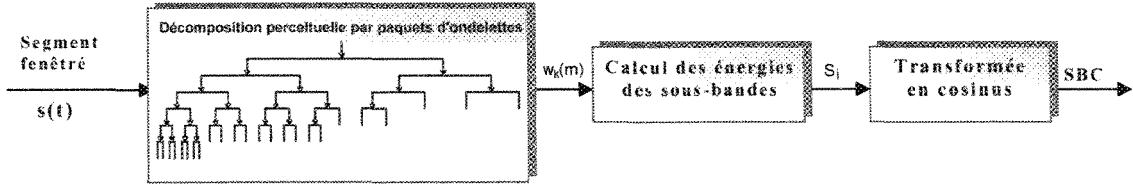


Figure 2.26: *Obtention des coefficients SBC à partir d'un segment du signal.*

Dans ce cas, le vecteur des caractéristiques est donné par l'équation 2.29 :

$$\mathbf{x} = [SBC(1), \dots, SBC(24)] \quad (2.29)$$

2.5.3.4 Vecteur des caractéristiques tiré de la méthode *WPC*

La méthode des coefficients de paquets d'ondelettes (*WPC* - *Wavelet Packet Coefficient*) a été proposée par Learned et Willsky [43]. Son but premier est de diminuer au maximum la dimension (D) du vecteur des caractéristiques en déterminant les sous-bandes les plus significatives suite à une décomposition en valeur singulière (*SVD* - *Singular Value Decomposition*) effectuée sur l'ensemble complet des caractéristiques tirées des segments (M) de chacune des classes. La décomposition en valeur singulière consiste à décomposer la matrice des caractéristiques \mathcal{M} , constituée des énergies des sous-bandes de M segments, sous la forme de l'équation 2.30. Il est à noter que la dimension de la matrice \mathcal{M} est de [Nombre de sous-bandes x Nombre de segments]

$$\mathcal{M}_{[D \times M]} = U_{[D \times D]} \ S_{[D \times M]} \ V_{[M \times M]}^T \quad (2.30)$$

où U est une matrice colonne des vecteurs propres de $\mathcal{M}\mathcal{M}^T$, V est une matrice colonne des vecteurs propres de $\mathcal{M}^T\mathcal{M}$ et S est une matrice diagonale contenant les valeurs singulières en ordre décroissant. On recrée alors une approximation de \mathcal{M} de dimension $[D \times M]$ en ne gardant que la première valeur de la matrice S correspondant à la première valeur singulière. Par la suite, on calcul la moyenne de chaque ligne obtenant ainsi une matrice de dimension $[D \times 1]$, ce pour chaque classe. Les sous-bandes significatives sont alors

celles dont les moyennes sont significatives par rapport aux autres. De plus les sous-bandes significatives totales sont composées des sous-bandes significatives obtenues pour chaque classe. La décomposition en valeur singulière peut même donner un vecteur de caractéristiques de seulement deux dimensions (D) tout en donnant de bons résultats (figure 2.27). Son désavantage réside dans le temps de calcul considérable que nécessite la décomposition. Nous en sommes donc réduit à sous-échantillonner les segments afin de bâtir le code. Voici en détail comment effectuer cette méthode :

- pour chaque classe, calculer la décomposition en paquets d'ondelettes *WPD* de chaque segment du signal ;
- calculer l'énergie de chaque sous-bande ;
- réorganiser les sous bandes d'énergie, pour tous les segments, sous forme de matrice $\mathcal{M}_{[D \times M]}$, une par classe ;
- calculer la décomposition en valeur singuliers (*SVD*) pour chaque classe ;
- garder seulement la première valeur singulière et recalculer $\hat{\mathcal{M}}_{[D \times M]} = U_{[D \times 1]} S_{[1 \times 1]} V_{[M \times 1]}^T$;
- calculer la moyenne de chaque ligne de $\hat{\mathcal{M}}_{[D \times M]}$ et déterminer ensuite les sous-bandes significatives ;
- s'intéresser lors de l'apprentissage et du test qu'aux sous-bandes significatives, quelque soit la classe, tel qu'illustré par les sous-bandes encadrées à la figure 2.27.

2.6 Prétraitement

Le prétraitement peut à la fois s'effectuer sur le signal et sur les caractéristiques. Il a comme but d'améliorer le taux de reconnaissance mais ne donne pas toujours les résultats escomptés.

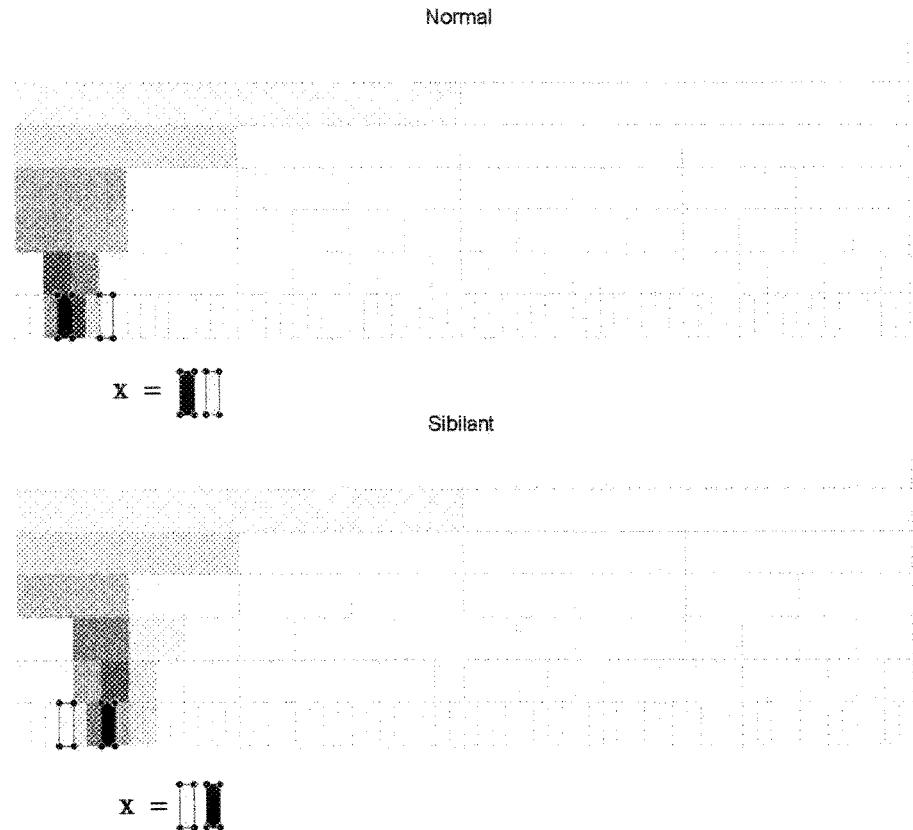


Figure 2.27: *Caractérisation par la méthode des coefficients de paquets d'ondelettes (WPC). Sous-bandes d'énergie tirées d'un seul segment suite à la décomposition en paquets d'ondelettes (WPD) pour un son normal et pour un sibilant. Les points noirs entourent les deux sous-bandes obtenues suite au calcul de la SVD sur l'ensemble des segments pour chaque classe. Ils correspondent bien aux sous-bandes ayant le plus d'énergie pour ce segment. Les sous-bandes significatives de chacune des classes sont choisies comme vecteur des caractéristiques.*

2.6.1 Prétraitement sur chaque segment du signal

2.6.1.1 Normalisation

La normalisation est un prétraitement effectué sur tous les segments temporels du signal. Elle consiste à normaliser l'amplitude du signal à la valeur unitaire.

$$s(n) = \frac{s[n]}{\max(|s[n]|)} \quad (2.31)$$

Tous les résultats de cette étude ont été obtenus en effectuant cette normalisation.

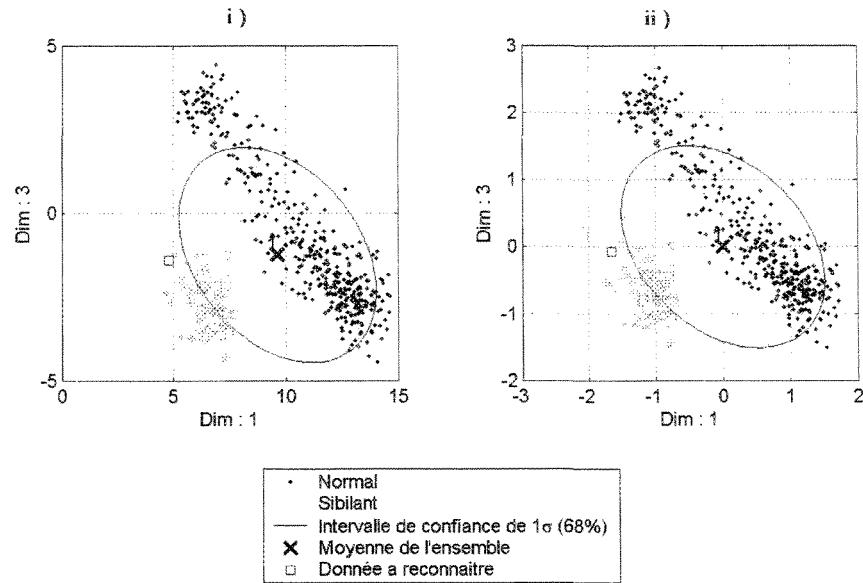


Figure 2.28: i) données avant le centrage. ii) données après le centrage regroupées autour de l'origine. Le carré représente une donnée à classifier.

2.6.2 Prétraitement sur le vecteur des caractéristiques

2.6.2.1 Centrage

Le but de ce prétraitement est de rendre la moyenne de l'ensemble des composantes du vecteur des caractéristiques égale à zéro et l'écart-type égal à l'unité, en effectuant l'opération de l'équation 2.32. La figure 2.28 illustre cette transformation. Ce prétraitement précèdera aussi d'autres types de prétraitement, comme nous le verrons plus loin.

$$\mathcal{X}' = \frac{\mathcal{X} - \mu}{\sigma} \quad (2.32)$$

2.6.2.2 Analyse par composantes principales (PCA)

L'analyse par composantes principales (*PCA-Principal Components Analysis*) transforme les caractéristiques du signal de façon à ce que les composantes du vecteur des caractéristiques soient non corrélées. Ceci permet de réduire la dimension des données en éliminant les composantes qui ne contribuent pas à plus qu'une certaine fraction de la variance des données. Nous avons utilisé le prétraitement appelé “centrage” avant l'utili-

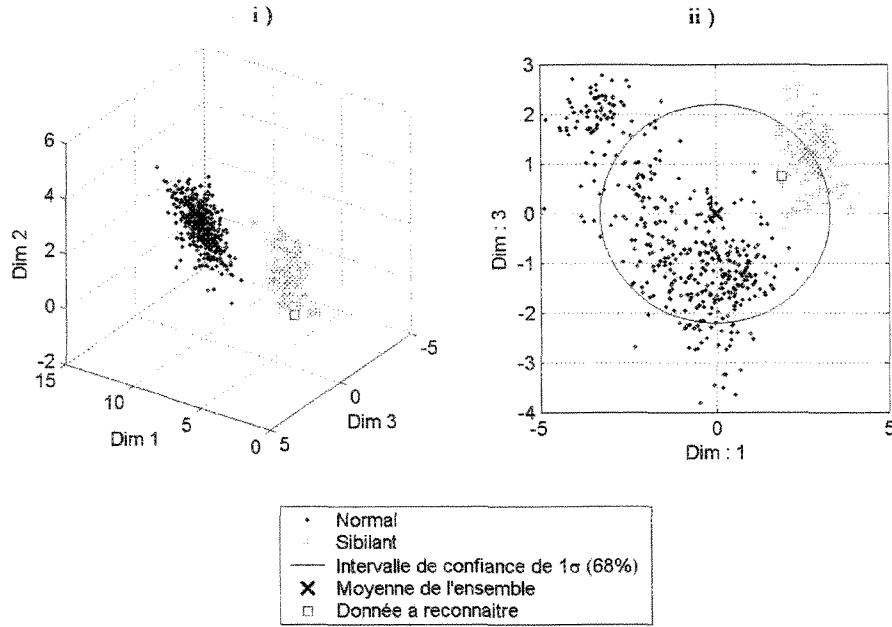


Figure 2.29: *i)* données avant l'application de la méthode PCA. *ii)* données après l'application de la méthode PCA, projetées dans un espace de dimension réduite.

sation de ce prétraitement [44].

Pour effectuer une réduction de la dimension par la méthode PCA, il faut effectuer une décomposition en valeur singulière (*SVD*) de la matrice de covariance. Pour ce faire, étant donné que la matrice est carrée, on peut dans un premier temps, obtenir les valeurs propres λ ainsi que les vecteurs propres ν de la matrice de covariance Σ de l'ensemble des données \mathcal{X} en résolvant l'équation 2.33.

$$(\Sigma - \lambda I) \nu = 0 \quad (2.33)$$

où : I est la matrice identité.

Par la suite, les valeurs propres et leurs vecteurs propres associés sont classés par ordre décroissant de valeurs propres. On forme ensuite une matrice A_p dont les colonnes sont les vecteurs propres associés seulement aux plus grandes valeurs propres. La représentation des données par la méthode PCA consiste alors à projeter les données \mathcal{X} sur le nouvel

espace à l'aide de l'équation 2.34.

$$\mathcal{X}' = A_p^T \mathcal{X} \quad (2.34)$$

La figure 2.30 illustre un exemple d'application de la méthode *PCA*.

2.6.2.3 Blanchiment d'un groupe de données : Blanchiment-1

La figure 2.30 montre les valeurs des déterminants des matrices de covariance. Comme nous le verrons ci-dessous, nous avons intérêt à éviter un déterminant s'approchant de zéro. La figure 2.30 illustre la transformation effectuée par l'équation 2.35, où les vecteurs propres ν et valeurs propres λ proviennent de la matrice de covariance prise sur le groupe de données donnant le plus petit déterminant (ici les +).

$$A_b = \frac{\nu}{\sqrt{[\lambda]}} \quad (2.35)$$

où $[\lambda]$ représente la matrice diagonale des valeurs propres.

La transformation s'effectue selon l'équation 2.36. Encore une fois, cette opération a été précédée du prétraitement de centrage.

$$\mathcal{X}' = A_b^T \mathcal{X} \quad (2.36)$$

2.6.2.4 Blanchiment sur l'ensemble des données : Blanchiment-2

Nous évaluons aussi l'effet du blanchiment sur l'ensemble des données tel qu'illustré à la figure 2.31. Ici, le prétraitement semble nuire à la classification étant donné qu'il distribue les points autour d'une valeur centrale et élimine ainsi les regroupements.

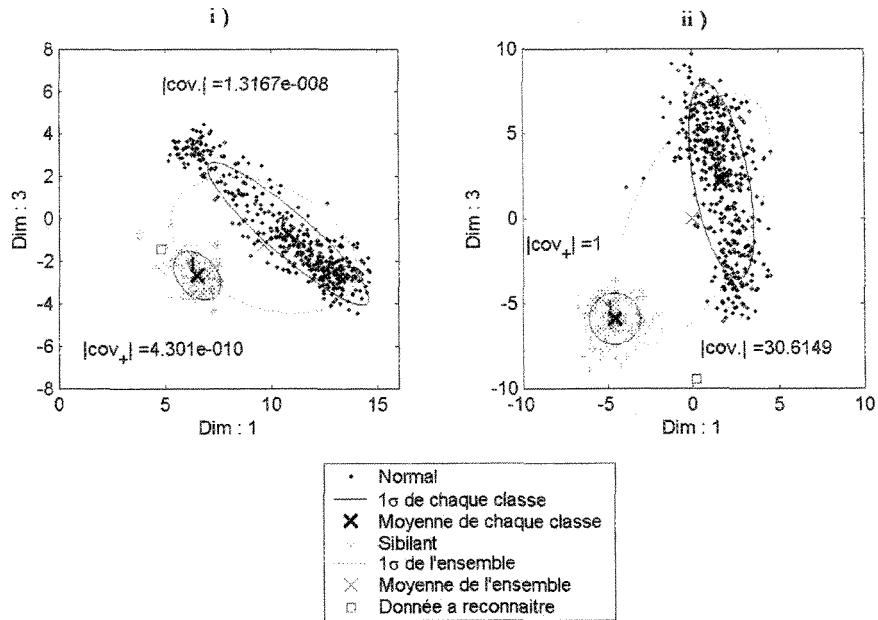


Figure 2.30: i) données avant le blanchiment-1. ii) données après le blanchiment-1. La matrice de transformation est déterminée à partir des données dont le $|\text{cov}(x_i)|$ est le plus près de zéro (ici les +). La transformation est, par la suite, appliquée aux données (+) et (·) ainsi qu'aux données à classifier (□).

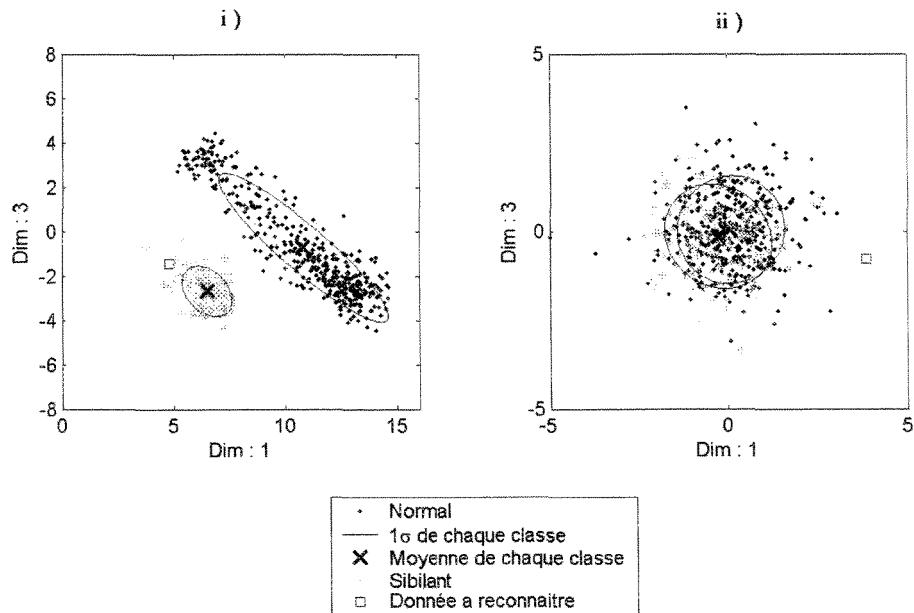


Figure 2.31: i) données avant le blanchiment-2. ii) données après le blanchiment-2. La matrice de transformation est déterminée à partir de toutes les données.

CHAPITRE III

MODÉLISATION ET CLASSIFICATION DES SIGNAUX

Ce chapitre traite des méthodes de classification utilisées dans ce projet. Ces méthodes fonctionnent en deux étapes : l'apprentissage et le test. Lors de la phase d'apprentissage, l'ensemble des vecteurs de caractéristiques provenant de chacune des classes est modélisé puis un discriminant est établi pour délimiter ces classes. Si lors de cette phase, l'appartenance à une classe est déjà connue, l'apprentissage est dit *supervisé*. C'est le cas des méthodes décrites dans ce chapitre. Lors de la phase de test, la position de nouvelles données par rapport au discriminant établit la classe d'appartenance de chaque nouvelle donnée.

3.1 Principe d'un système de classification

La figure 3.1 représente le diagramme bloc d'un système de classification. Le signal de son respiratoire $s(t)$ est divisé en segments $s_i(t)$ qui peuvent être entrelacés et pondérés par une fenêtre, avant d'en extraire les vecteurs des caractéristiques correspondant $\mathbf{x}_i(t)$. Lors de la phase d'apprentissage, des signaux de type normal et de type sibilant sont présentés au système. La totalité des vecteurs des caractéristiques appartenant à une même classe se retrouve alors modélisée par quelques paramètres. Les modèles ainsi obtenus forment une banque contenant un modèle par classe. Lors de la phase du test, un vecteur $\mathbf{x}_i(t)$, caractérisant un segment $s_i(t)$ du signal à tester, est présenté aux deux modèles. La décision d'appartenance à une classe de sons est alors prise selon la position de la donnée $\mathbf{x}_i(t)$ par rapport à la frontière entre les deux classes, aussi appelée “discriminant”.

Dans ce projet, nous avons expérimenté les quatre méthodes listées ci-dessous :

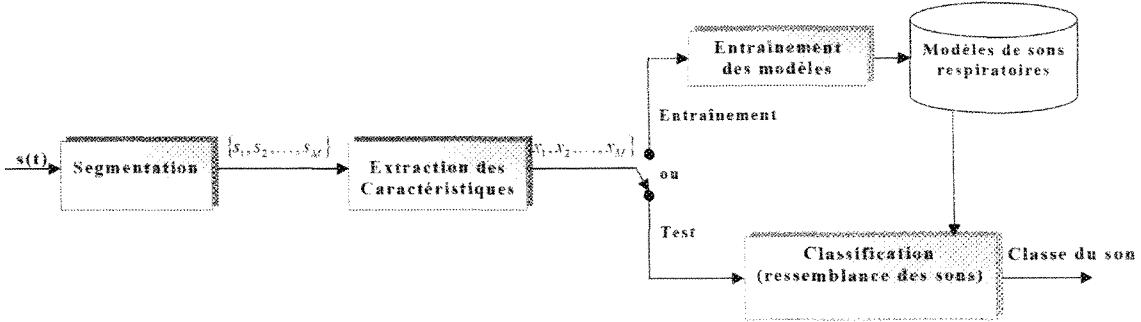


Figure 3.1: Diagramme bloc d'un système de classification [30]. Il est à noter que cette figure ci est une version simplifiée de la figure 1.1

- quantification vectorielle (*VQ* - *Vector Quantization*) ;
- modélisation par mélange de gaussiennes (*GMM* - *Gaussian Mixture Model*) ;
- modélisation par mélange de distributions de Student (*SMM* - *Student Mixture Model*) ;
- perceptron multi-couche (*MLP* - *Multi Layer Perceptron*).

Tableau 3.1: Nomenclature et symboles utilisés pour décrire les systèmes de classification

Symbol	Dimension	Définition	Commentaires
D		Dimension de l'espace des caractéristiques	(d_1, d_2, \dots, d_D)
M		Nombre de segments	
i		Indice du segment	
N		Nombre de regroupements	
j		Indice du regroupement	
x	Dx1	Vecteur des caractéristiques ou donnée	$\mathbf{x} = (x_1, x_2, \dots, x_D)^T$
X	DxM	Vecteurs des caractéristiques pour tous les segments	$\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$
c	Dx1	Centroïde (mot de code)	$\mathbf{c} = (c_1, c_2, \dots, c_D)^T$
C	DxN	Ensemble de <i>N</i> centroïdes (dictionnaire pour <i>VQ</i>)	$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N\}$
λ		Modèle de <i>GMM</i>	$\lambda = \{\mu_j, \Sigma_j, \omega_j\} \quad j = 1, \dots, N$
μ	Dx1	Moyenne	$\mu = (\mu_1, \mu_2, \dots, \mu_D)^T$
Σ	DxD	Matrice de covariance	
ω	Nx1	Poids	$\omega = (w_1, w_2, \dots, w_N)$
N		Distribution gaussienne ou Normal	
S		Distribution de Student	
R		Classe	R_1, R_2

3.2 Nomenclature

Le tableau 3.1 donne la nomenclature et les symboles utilisés pour représenter les différents paramètres dans ce chapitre. On trouve aussi les dimensions de ces paramètres.

Par souci de clarté, nous utiliserons à plusieurs reprises, les mêmes données bidimensionnelles de la figure 3.2 pour nos démonstrations. En effet, ces données simulées forment

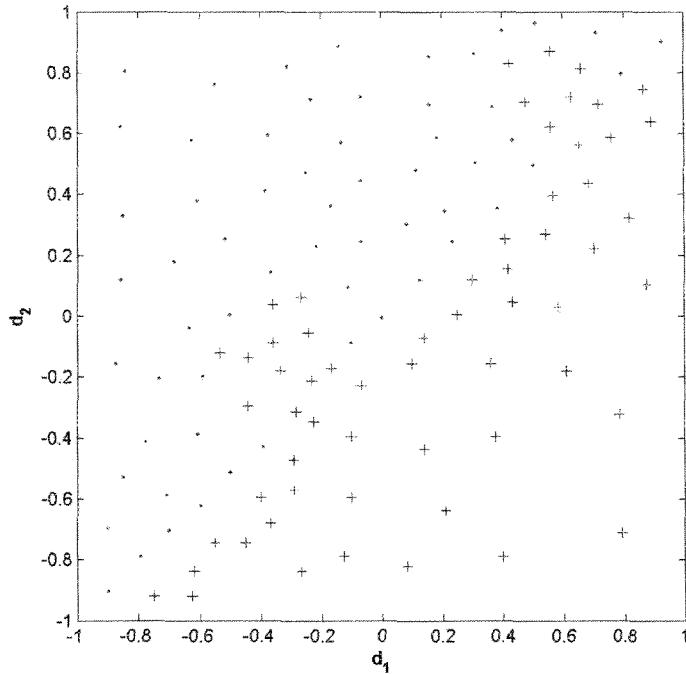


Figure 3.2: *Exemple de données artificielles à discriminer appartenant à deux classes R ; $(\cdot) \in R_1$ et $(+) \in R_2$. Nous utiliserons régulièrement ces données comme exemple, tout au long de ce chapitre.*

deux classes, complètement séparables, dans un espace à deux dimensions ; Ce qu'on n'a pu obtenir avec des données réelles. La classe R_1 contient les éléments représentés par le symbole (\cdot) et la classe R_2 contient les éléments représentés par le symbole $(+)$.

3.3 Modélisation

La modélisation consiste à représenter une quantité de données de grande taille, par un ensemble de taille très réduite, appelé dictionnaire (*Codebook*). La figure 3.3 montre un exemple du résultat de quelques méthodes de modélisation permettant d'obtenir un ensemble restreint à partir de 3294 données bidimensionnelles ($D=2$).

- Dans cet exemple, pour la méthode de quantification vectorielle (*VQ*) (figure 3.3-i), les centroïdes (\blacktriangledown), représentant les quatre sous-groupes, forment le dictionnaire :

$$\text{Dictionnaire} = [\mu_1, \mu_2, \mu_3, \mu_4] \quad (3.1)$$

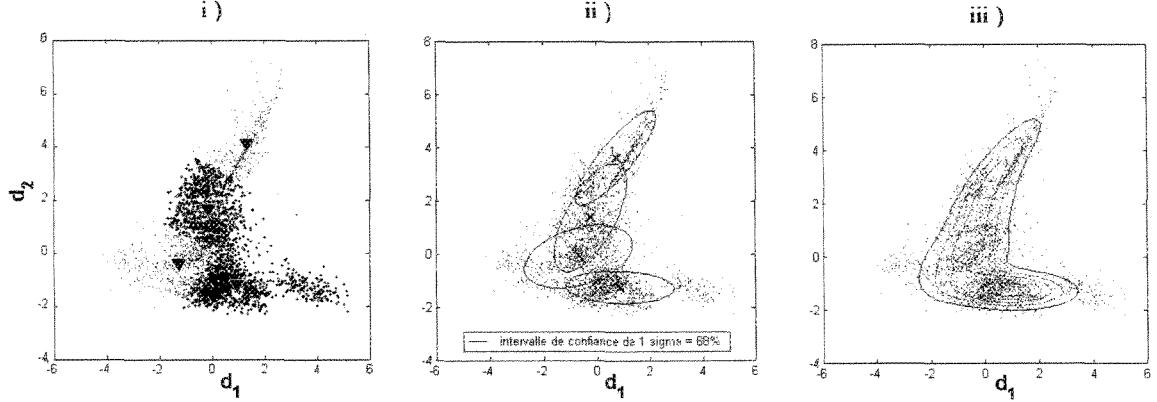


Figure 3.3: *Exemple de modélisation de 3294 données, provenant de 3294 segments d'un signal de son respiratoire, selon différentes méthodes : i) quantification vectorielle, ii) combinaison de gaussiennes, iii) modélisation par mélange de gaussiennes.*

où $\mu_j = (\mu_{j_{d1}}, \mu_{j_{d2}})^T$. Ceci donne un dictionnaire de huit éléments.

- Pour la combinaison de gaussiennes (figure 3.3-ii) des quatre sous-groupes, le dictionnaire est défini par :

$$\text{Dictionnaire} = [\mu_1, \mu_2, \mu_3, \mu_4, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4] \quad (3.2)$$

où $\Sigma_j = (\sigma_{j_{11}}, \sigma_{j_{12}}, \sigma_{j_{22}})^T$. Ceci donne un dictionnaire de vingt éléments.

- Pour la modélisation par mélange de gaussiennes (*GMM*) (figure 3.3-iii), le dictionnaire est défini par :

$$\text{Dictionnaire} = [\mu_1, \mu_2, \mu_3, \mu_4, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \omega] \quad (3.3)$$

où $\omega = (w_1, w_2, w_3, w_4)$ est un vecteur de scalaires représentant les poids des quatre gaussiennes. Ceci donnera un dictionnaire de vingt-quatre éléments.

Les explications seront détaillées un peu plus loin dans ce chapitre.

3.4 Quantification vectorielle (*VQ*)

La quantification vectorielle (*VQ*) est un processus permettant de projeter un grand ensemble de vecteurs de dimension D de l'espace \mathbb{R}^D vers un ensemble beaucoup moins. Ce processus permet de répartir l'ensemble des données de grande taille en un nombre

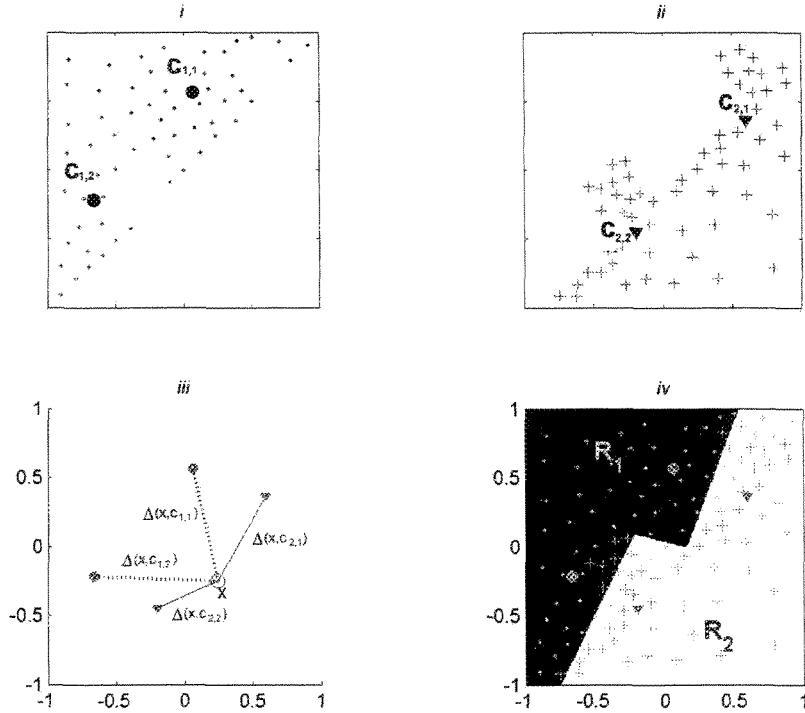


Figure 3.4: *La méthode de quantification vectorielle.* i) centroïdes de la classe R_1 , ii) centroïdes de la classe R_2 , iii) recherche des distances euclidiennes minimales pour chaque classe pour un segment à tester, iv) discriminant obtenu en effectuant l'opération décrite en iii pour tout les points du plan. Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

réduit de régions représentée seulement par leur centroïde \mathbf{c}_i appelé “mot de code” (*code-word*). Pour une classe de signaux, les mots de code résultant forment le “dictionnaire” \mathcal{C} (*codebook*). Les figures 3.4-i-ii illustrent deux centroïdes représentant l’ensemble des points pour chacune des deux classes.

La modélisation VQ présente deux phases : l’apprentissage et la classification. Dans la phase d’apprentissage, un modèle acoustique (mot de code) est créé pour chacune des classes et les modèles sont stockés dans une base de modèles. Dans la phase de classification, le signal inconnu à tester est classifié selon le modèle procurant la distance minimale, tel qu’illustiqué à la figure 3.4-iii.

3.4.1 Phase d'apprentissage

Dans cette phase, une séquence d'entraînement composée M données (vecteurs des caractéristiques) est utilisée :

$$\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\} \quad (3.4)$$

En général, la séquence d'entraînement est obtenue à partir d'une large base de données. On suppose que M est suffisamment grand pour que toutes les propriétés statistiques des sons soient captées par la séquence d'entraînement. Les vecteurs d'entraînement sont de dimension D .

$$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,D}), \quad i = 1, 2, \dots, M. \quad (3.5)$$

Le dictionnaire *codebook* \mathcal{C} contient N vecteurs représentant les N centroïdes. Chaque vecteur est de dimension D .

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N\} \quad (3.6)$$

$$\mathbf{c}_j = (c_{j,1}, c_{j,2}, \dots, c_{j,D}), \quad j = 1, 2, \dots, N. \quad (3.7)$$

Pour une séquence \mathcal{X} donnée et un nombre de centroïdes N défini d'avance, la phase d'apprentissage consiste à trouver le codebook \mathcal{C} qui minimise l'erreur quadratique moyenne donnée par l'équation 3.8 :

$$\varepsilon = \frac{1}{NM} \sum_{i=1}^M \sum_{j=1}^N \Delta(\mathbf{x}_i, \mathbf{c}_j)^2 \quad (3.8)$$

où $\Delta(\mathbf{x}_i, \mathbf{c}_j)$ est la distance euclidienne définie par l'équation 3.9 :

$$\Delta(\mathbf{x}_i, \mathbf{c}_j) = \sqrt{(\mathbf{x}_i - \mathbf{c}_j)^T (\mathbf{x}_i - \mathbf{c}_j)} \quad (3.9)$$

Rappelons que x_i est la $i^{\text{ème}}$ donnée et \mathbf{c}_j est le $j^{\text{ème}}$ centroïde (mot de code).

L'apprentissage est réalisé en utilisant l'algorithme *LBG* décrit à la section 3.4.3.

3.4.2 Phase de classification

Une nouvelle donnée \mathbf{x}_i , caractérisant une fenêtre temporelle du signal s_i , est présentée aux centroïdes de chacune des classes. Cette nouvelle donnée est illustrée par \mathbf{x} à la figure 3.4-iii. On recherche alors la distance euclidienne minimales, pour chaque classe. Dans cet exemple les distances euclidiennes minimales sont $\Delta(x_i, \mathbf{c}_{1,1})$ pour la classe R_1 et $\Delta(x_i, \mathbf{c}_{2,2})$ pour la classe R_2 . Le logarithme du rapport des distances euclidiennes minimales est ensuite calculé, selon l'équation 3.10 :

$$z(i) = \ln \left(\frac{\min_{j=1}^N (\Delta(x_i, \mathbf{c}_{1,j}))}{\min_{j=1}^N (\Delta(x_i, \mathbf{c}_{2,j}))} \right) \quad (3.10)$$

où la fonction $z(i)$ sera appelée tout simplement “distance”. Ces mesures de distance seront distribuées de part et d'autre de la valeur zéro qui correspond au discriminant.

Si $z(i) < 0$, le point appartient à la classe R_1 ; alors que si $z(i) > 0$, le point appartient à la classe R_2 (équation 3.12). La figure 3.4-iv illustre le résultat de cette équation pour plus de 40000 points représentés dans le plan $[d_1, d_2]$. Le noir représente un résultat négatif et le blanc un résultat positif. Une valeur nulle correspond au discriminant et établit la frontière entre les deux classes de la figure 3.4-iv.

Dans le cas où M segments du signal sont présentés simultanément, il y aura par conséquent, M points dans l'espace des caractéristiques qui seront testés. L'équation 3.10 est alors remplacée par l'équation 3.11.

$$z(i) = \ln \left(\frac{\frac{1}{M} \sum_{i=1}^M \min_{j=1}^N (\Delta(x_i, \mathbf{N}_{1,j}))}{\frac{1}{M} \sum_{i=1}^M \min_{j=1}^N (\Delta(x_i, \mathbf{c}_{2,j}))} \right) \quad (3.11)$$

La décision d'appartenance à une des deux classes est formulée comme suit :

$$z(i) \stackrel{\epsilon_{R_1}}{\leqslant} 0 \quad (3.12)$$

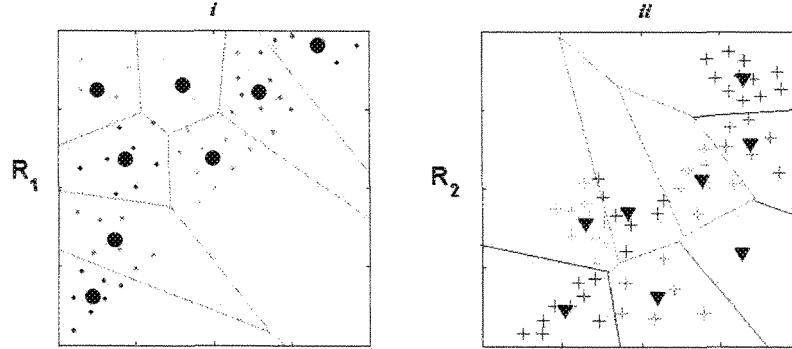


Figure 3.5: Cellules de Voronoï montrant les regroupements pour i) la classe R_1 et ii) la classe R_2 . Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

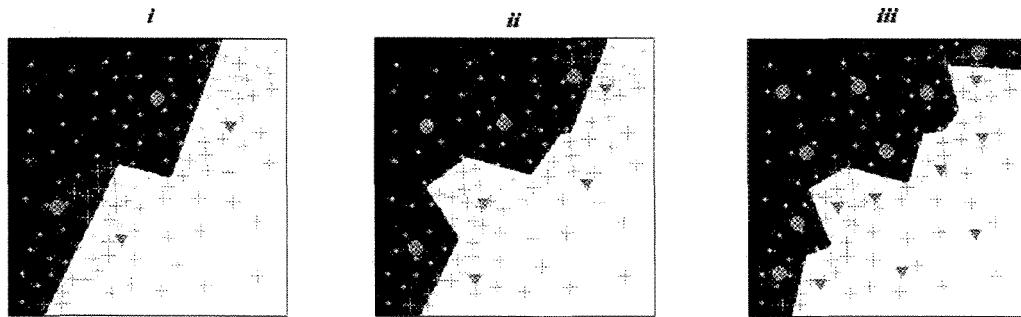


Figure 3.6: Discriminant obtenu pour : i) deux centroïdes par classe , ii) quatre centroïdes par classe et iii) huit centroïdes par classe. Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

Comme illustré à la figure 3.4-iv le discriminant n'épouse pas toujours bien les deux classes illustrées à la figure 3.2. La discrimination et la classification peuvent alors être améliorées en augmentant le nombre de centroïdes. La figure 3.5 montre les centroïdes obtenus pour huit regroupements par classe. Chaque regroupement est une cellule de Voronoï V_i de laquelle est tirée le centroïde \mathbf{c}_i . La cellule de Voronoï est définie par l'équation suivante :

$$V_i = \{x \in \mathbb{R}^D : \|x - \mathbf{c}_i\| \leq \|x - \mathbf{c}_j\|, \text{ pour } j \neq i\} \quad (3.13)$$

La figure 3.6 montre les discriminants obtenus par la modélisation VQ pour différents nombre de centroïdes et illustre le fait qu'un nombre plus élevé de regroupements améliore la capacité du discriminant à épouser la frontière entre les deux classes.

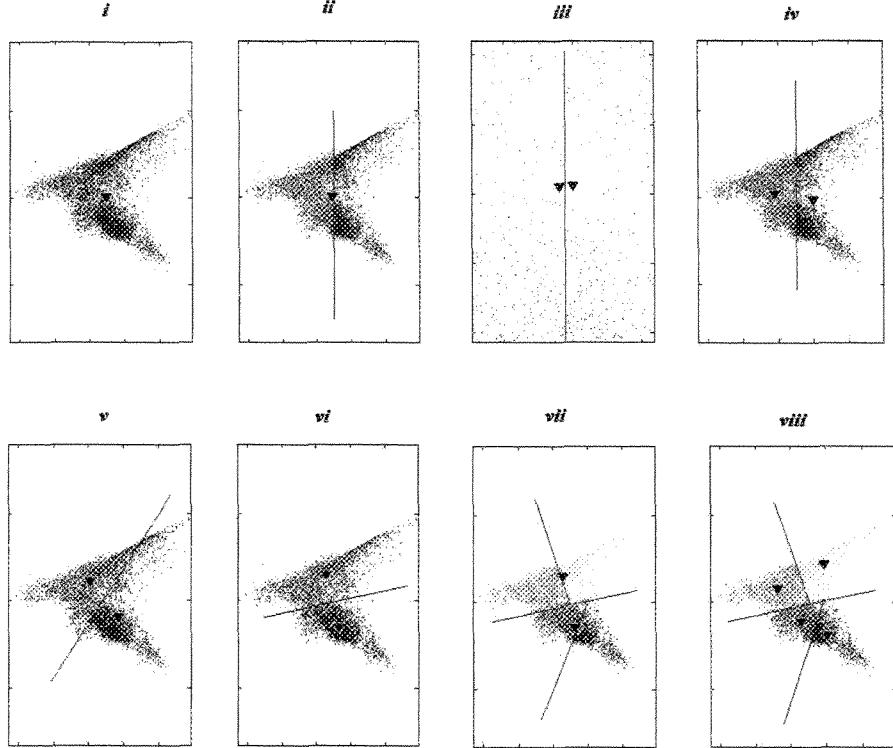


Figure 3.7: Illustration de l'algorithme *LBG*. *i*) centroïde pour l'ensemble des données d'une classe, *ii*) première séparation, *iii*) zoom de *ii*, *iv*) centroïde de chaque regroupement (itération 1), *v*) centroïde de chaque regroupement (itération 3), *vi*) centroïde de chaque regroupement (itération 6), *vii*) deuxième séparation et *viii*) centroïde de chaque nouveau regroupement (itération 1).

3.4.3 Algorithme *LBG*

L'algorithme *LBG*, du nom de ses inventeurs *Linde, Buzo et Gray* [45] est une technique itérative, de type nuée dynamique, qui permet l'implantation de la modélisation *VQ* (*Vector Quantization*) en minimisant l'erreur quadratique moyenne (équation 3.8) à chaque itération.

L'algorithme se résume en cinq étapes [46] :

- 1) créer un dictionnaire (*codebook*) de un centroïde ($N = 1$). Le centroïde est la moyenne de l'ensemble des M données, correspondant aux M vecteurs des caractéristiques (figure 3.7-i) ;

$$\mathbf{c}_1 = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i \quad (3.14)$$

- 2) doubler la taille du codebook en séparant chaque centroïde du présent codebook en deux nouveaux centroïdes :

$$\mathbf{c}_j^+ = \mathbf{c}_j(1 + \epsilon) \quad \mathbf{c}_j^- = \mathbf{c}_j(1 - \epsilon) \quad (3.15)$$

où $j = 1, 2, \dots, N$ et ϵ un paramètre de faible valeur (par exemple, $\epsilon = 0.01$).

Le dédoublement de la taille du codebook ($N = 2N$) divisera par deux le groupe original des données (figure 3.7-ii et iii) ;

- 3) calculer le centroïde de chaque nouveau groupe (figure 3.7-iv) ;
- 4) associer à nouveau les données au centroïde le plus près. Calculer à nouveau le centroïde de chaque groupe (figure 3.7-v). Répéter cette étape jusqu'à ce que le déplacement des centroïdes devient négligeable (figure 3.7-vi) ;
- 5) répéter à partir de l'étape 2 (figure 3.7-vii et viii) jusqu'à ce que le nombre désiré de centroïdes soit atteint.

3.5 Modélisation par mélange de gaussiennes (*GMM*)

La modélisation par mélange de gaussiennes (*GMM* - *Gaussian Mixture Model*) est une méthode statistique massivement utilisée dans les systèmes d'identification/vérification de locuteurs [47, 48]. Nous avons proposé l'utilisation de cette technique pour reconnaître les différentes classes de sons respiratoires [29, 30].

3.5.1 Principe de la modélisation par mélange de gaussiennes

La modélisation des données par une seule courbe gaussienne consiste à les représenter par les statistiques de leurs deux premiers moments (moyenne et covariance). L'utilisation des deux premiers moments suppose que les données suivent une distribution normale, ce qui n'est pas toujours le cas dans les problèmes pratiques. L'avantage d'utiliser une combinaison de gaussiennes réside dans le fait qu'une densité de probabilité (*ddp*), d'une forme quelconque, peut être approximée par une combinaison de gaussiennes (*GMM*). Sous forme mathématique, la densité de probabilité obtenue par la technique *GMM* s'écrira :

$$p(\mathbf{x}_i|\lambda) = \sum_{j=1}^N p(\mathbf{x}_i|\mathcal{N}_j) \omega_j \quad (3.16)$$

sous la contrainte : $\sum_{j=1}^N \omega_j = 1$

où λ est le modèle *GMM* (dictionnaire) d'une classe de sons respiratoires.

L'équation d'une courbe normale multi-dimensionnelle est décrite par l'équation 3.17.

$$p(\mathbf{x}_i|\mathcal{N}_j) = \frac{1}{2\pi^{\frac{D}{2}} \sqrt{|\Sigma_j|}} \exp\left(-\frac{\Delta_{ij}^2}{2}\right) \quad (3.17)$$

$$\Delta_{ij} = \sqrt{(\mathbf{x}_i - \mu_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \mu_j)}$$

Il est à noter que Δ_{ij} est appelée la distance de Mahalanobis.

Dans le cas bidimensionnel, en se référant à la figure 3.8, une donnée (+) est représentée par : $\mathbf{x} = (x_1 \ x_2)^T$, une moyenne par : $\mu = (\mu_1 \ \mu_2)^T$ et une matrice de covariance par : $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}$, où le terme σ_m représente l'écart-type selon la dimension m tandis que le terme ρ représente le coefficient de corrélation respectant le critère : $-1 \leq \rho \leq 1$. Il est à noter que les indices "1" et "2" correspondent aux dimensions d_1 et d_2 qui ont été omis pour ne pas encombrer les formules. Une variante de cette méthode de modélisation consiste à utiliser la matrice de covariance $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. Cette matrice de covariance est alors qualifiée de matrice *diagonale*.

Nous évaluerons aussi une autre variante qui consiste à prendre la moyenne des variances comme nouvelles variances, ainsi les nouvelles variances pour σ_1^2 ainsi que σ_2^2 seraient obtenues à partir des anciennes variances selon : $\frac{\sigma_1^2 + \sigma_2^2}{2}$. Cette matrice de covariance est qualifiée de matrice *sphérique*. Un ensemble de données, ayant une distribution statistique quelconque, pourra alors être modélisé par un ensemble de moyennes, de covariances et de poids. Le dictionnaire est définie comme :

$$\lambda = \{\mu_j, \Sigma_j, \omega_j\} \quad j = 1, \dots, N \quad (3.18)$$

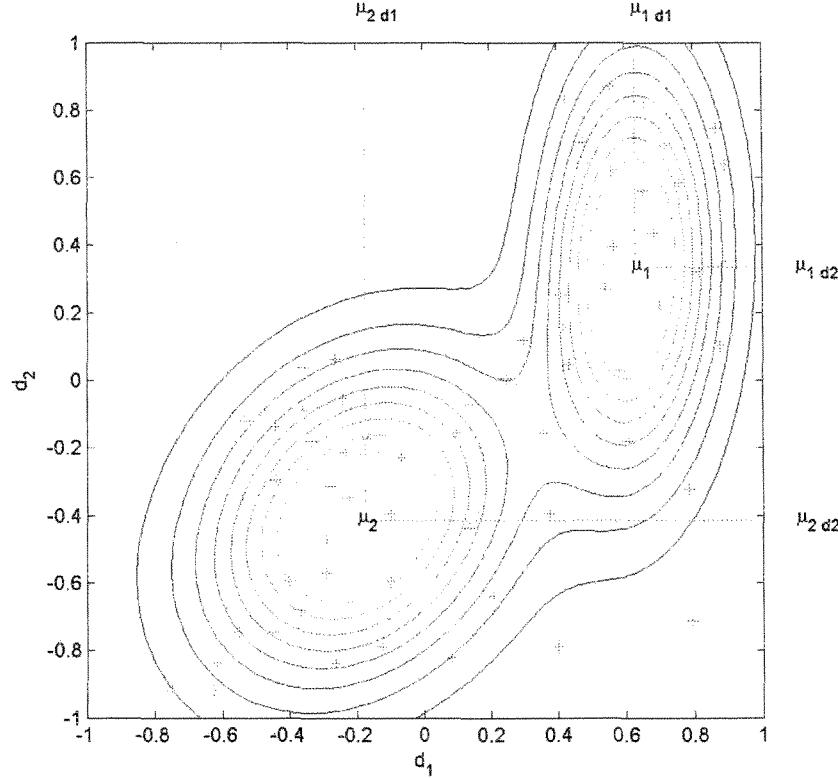


Figure 3.8: Modélisation des données obtenue par mélange de deux gaussiennes. $p(x|\lambda) = p(x|\mathcal{N}_1)\omega_1 + p(x|\mathcal{N}_2)\omega_2$. Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

Comme exemple, les données de la figure 3.2 pourront être modélisées par douze valeurs comprises dans les vecteurs, matrices et scalaires suivants : μ_1 (2 valeurs), μ_2 (2 valeurs), Σ_1 (3 valeurs), Σ_2 (3 valeurs) et ω (2 valeurs).

3.5.2 Phase d'apprentissage

La modélisation *GMM* caractérise un ensemble de points par une combinaison de gaussiennes. La figure 3.9-i-ii illustre la modélisation de chaque classe, par un mélange de deux gaussiennes formant le dictionnaire λ défini à l'équation 3.18. Les paramètres du dictionnaire sont estimés en utilisant la méthode du maximum de vraisemblance (*MLE - Maximum Likelihood Estimation*) via l'algorithme EM (*Expectation Maximization*) développé à la section 3.5.4.

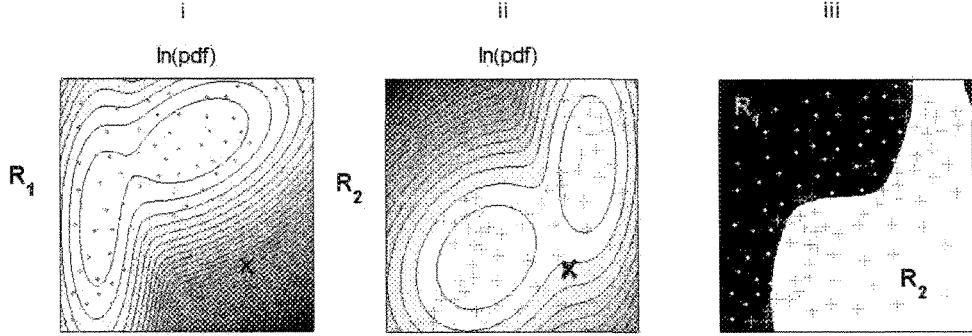


Figure 3.9: *i)* densité de probabilité pour la classe R_1 , *ii)* densité de probabilité pour la classe R_2 et *iii)* discriminant obtenu. Il est à noter que les échelles sont les mêmes qu'à la figure 3.2. Le \mathbf{X} représente une donnée à classifier.

3.5.3 Phase de classification

Une nouvelle donnée \mathbf{x}_i , illustrée par \mathbf{X} à la figure 3.9-i-ii, est présentée aux densités de probabilité de chacune des classes. Le calcul de la densité de probabilité, de chacune des classes, est effectué en ce point. Par la suite, les données peuvent être discriminées en utilisant la fonction distance $z(i)$, donnée par le logarithme du rapport des densités de probabilité, tel que décrit à l'équation 3.19. Selon l'équation 3.20, si $z(i) > 0$, le point appartient à la classe R_1 , si $z(i) < 0$, le point appartient à la classe R_2 . La figure 3.9-iii illustre le résultat de cette équation pour plus de 40000 points représentés dans le plan $[d_1, d_2]$. Le noir représente un résultat négatif et le blanc, le résultat positif. Une valeur nulle correspondrait au discriminant et serait la frontière séparant les deux zones de la figure 3.9-iii.

$$z(i) = \ln \left(\frac{p(x_i | \lambda_{R_1})}{p(x_i | \lambda_{R_2})} \right) \quad (3.19)$$

$$z(i) \begin{cases} \in R_1 \\ \in R_2 \end{cases} \gtrless 0 \quad (3.20)$$

Comme illustré sur la figure 3.9-iii, le discriminant n'épouse pas bien les deux classes illustrées à la figure 3.2. La discrimination peut alors être améliorée en augmentant le nombre de gaussiennes, pour chaque classe. La figure 3.10 illustre le mélange de gaussiennes pour quatre regroupements par classe.

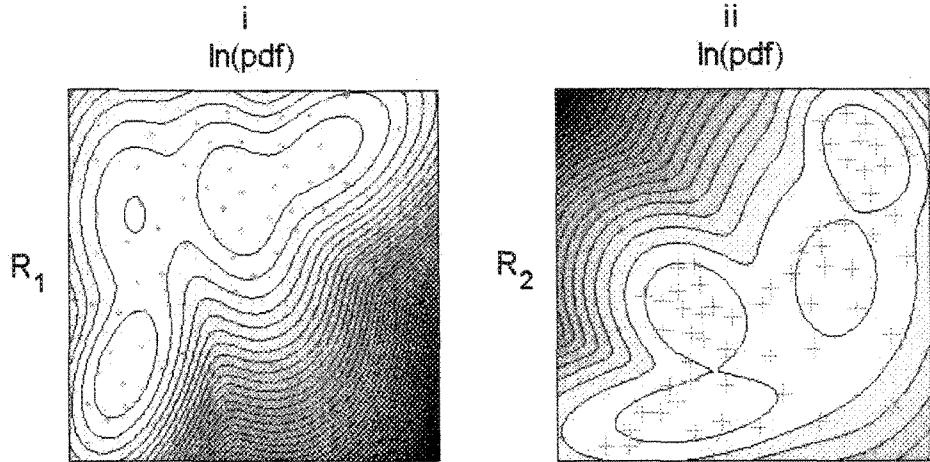


Figure 3.10: Modélisation des données obtenues par mélange de quatre gaussiennes pour i) classe R_1 et ii) classe R_2 . Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

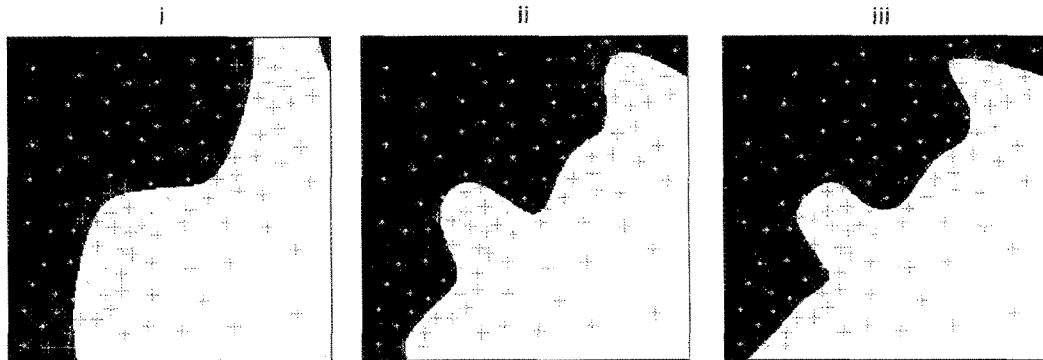


Figure 3.11: Discriminant obtenu pour différentes valeurs du nombre de gaussiennes i) deux par classe , ii) quatre par classe et iii) huit par classe. Il est à noter que les échelles sont les mêmes qu'à la figure 3.2.

La figure 3.11, illustre les discriminants obtenus par la modélisation *GMM* et montre que le nombre de gaussiennes améliore la capacité du discriminant à épouser la frontière entre les deux classes.

3.5.4 Algorithme EM (*Expectation Maximisation*)

L'algorithme de maximisation de la vraisemblance EM est un algorithme itératif qui permet de façon *non-supervisée* de trouver les paramètres du dictionnaire $\lambda = \{\mu_j, \Sigma_j, \omega_j\}$ modélisant un ensemble de données. On doit fournir à cet algorithme, le nombre N de gaussiennes désirées, un modèle $\lambda = \{\mu_j, \Sigma_j, \omega_j\}$ initial ainsi que l'ensemble des données

\mathcal{X} d'une même classe. À chaque itération, le modèle λ sera ajusté de façon à augmenter la vraisemblance (*Likelihood*) décrite par l'équation 3.21.

$$p(\mathcal{X}|\lambda) = \prod_{i=1}^M p(\mathbf{x}_i|\lambda) \quad (3.21)$$

$$\mathcal{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$$

La vraisemblance est un indice qui mesure l'efficacité du modèle λ à représenter les données \mathcal{X} . Dans le cas d'une seule gaussienne, par exemple, la valeur de la vraisemblance sera plus élevée lorsque la cloche de la courbe gaussienne sera située au dessus des données à modéliser. En pratique, nous nous intéresserons plutôt au logarithme de la vraisemblance mais l'idée reste la même (équation 3.22).

$$\ell = \ln p(\mathcal{X}|\lambda) = \ln \prod_{i=1}^M p(\mathbf{x}_i|\lambda) \quad (3.22)$$

$$\ell = \sum_{i=1}^M \ln p(\mathbf{x}_i|\lambda)$$

Nous traduirons les termes *Expectation* par “Estimation” et *Maximization* par “Maximisation”. L'algorithme *EM* se présente comme suit :

- 1) initialisation ;
- 2) boucle pour chaque itération $i = 1, 2, \dots, L$, comprenant les étapes :
 - estimation ;
 - maximisation ;

Nous allons maintenant décrire ces étapes en détails.

3.5.4.1 Initialisation

Le nombre N de gaussiennes et les valeurs initiales pour les paramètres μ , Σ et ω doivent être fournis à l'algorithme *EM* avant le lancement de celui-ci. Cette initialisation peut être effectuée de différentes façons. Notons que l'algorithme *LBG* vu lors la quantification vectorielle (*VQ*) peut être utilisée à cet effet. Ce dernier donne un nombre de

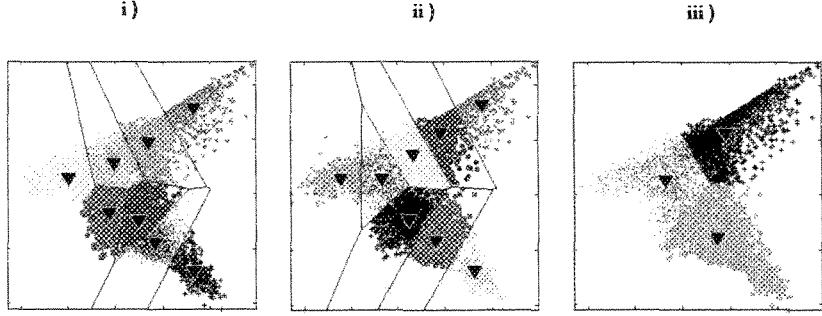


Figure 3.12: *Initialisation de l'algorithme EM par : i) méthode LBG (N=8) ; ii) méthode k-moyennes avec N=8 ; iii) méthode k-moyennes avec N=3.*

regroupement égal à une puissance de deux. Pour obtenir un nombre de regroupements quelconque, la méthode k-moyennes (*k-means*) peut être utilisée [49] (voir la figure 3.12).

Soulignons que pour un même nombre de regroupements, les deux méthodes d'initialisation proposées ci-dessus, ne donnent pas les mêmes centroïdes (voir la figure 3.12). Nous évaluerons l'effet de chacune de ces méthodes au chapitre 4.

3.5.4.2 Estimation (*Expectation*)

L'étape d'estimation consiste à estimer, pour chacune des données \mathbf{x}_i et chacune des gaussiennes \mathcal{N}_j , l'appartenance \hat{P} de la donnée à la gaussienne. Ceci peut être vu comme étant une probabilité, étant donnée que $(0 \leq \hat{P}(\mathcal{N}_j|\mathbf{x}_i) \leq 1)$. Une valeur de $\hat{P}(\mathcal{N}_j|\mathbf{x}_i) = 1$ signifie l'appartenance totale de \mathbf{x}_i à \mathcal{N}_j .

$$\hat{P}(\mathcal{N}_j|\mathbf{x}_i) = \frac{p(\mathbf{x}_i|\mathcal{N}_j)\hat{\omega}_j}{\sum_{m=1}^N p(\mathbf{x}_i|\mathcal{N}_m)\hat{\omega}_m} \quad (3.23)$$

3.5.4.3 Maximisation

L'étape de maximisation consiste à calculer les poids, les moyennes et les covariances suite aux résultats obtenus lors de l'estimation

$$\hat{\omega}_j = \frac{1}{M} \sum_{i=1}^M \hat{P}(\mathcal{N}_j|\mathbf{x}_i) \quad (3.24)$$

$$\hat{\mu}_j = \frac{1}{M\hat{\omega}_j} \sum_{i=1}^M \hat{P}(\mathcal{N}_j | \mathbf{x}_i) \mathbf{x}_i \quad (3.25)$$

$$\hat{\Sigma}_j = \frac{1}{M\hat{\omega}_j} \sum_{i=1}^M \hat{P}(\mathcal{N}_j | \mathbf{x}_i) [(\mathbf{x}_i - \hat{\mu}_j)] [(\mathbf{x}_i - \hat{\mu}_j)]^T \quad (3.26)$$

L'algorithme itère les sections Estimation et Maximisation (EM) jusqu'à ce qu'un certain critère de convergence soit rencontré. À chaque itération, les estimés $\hat{\omega}_j$ se retrouve directement dans l'équation 3.23, alors que les estimés $\hat{\Sigma}_j$ et $\hat{\mu}_j$ se retrouvent dans l'équation 3.23 à travers l'équation de la courbe normale décrite par l'équation 3.17.

3.5.4.4 Le discriminant

En développant la relation 3.19, nous pouvons établir l'équation analytique du discriminant séparant deux courbes gaussiennes, ce qui donne la relation 3.27 :

$$\begin{aligned} g(x_i) = & -\frac{1}{2} x_i^T [\Sigma_1^{-1} - \Sigma_2^{-1}] x_i + [\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1}] x_i \\ & + \frac{1}{2} [-\mu_1^T \Sigma_1^{-1} \mu_1 + \mu_2^T \Sigma_2^{-1} \mu_2 - \ln \frac{|\Sigma_1|}{|\Sigma_2|}] \end{aligned} \quad (3.27)$$

Laquelle est une équation quadratique de forme :

$$g(x_i) = -\frac{1}{2} x_i^T [E] x_i + [B] x_i + \frac{1}{2} [A] \quad (3.28)$$

Il est à noter que, lorsque le discriminant sépare une gaussienne d'un mélange de deux gaussiennes, il devient une combinaison de deux discriminants (voir la figure 3.13). Ceci complique de beaucoup la mise en équation. Pour cette raison, l'utilisation du critère de discrimination décrit par la relation 3.20 sera préféré à une équation analytique.

3.5.4.5 Contrainte reliée à la modélisation *GMM*

Dans le cas où une gaussienne caractérise un trop petit nombre de points, le déterminant de la matrice de covariance tendra vers zéro, ce qui amènera une valeur de $p(x | \mathcal{N}_j) \rightarrow \infty$, selon l'équation 3.17. Nous dirons dans ce cas que la matrice de covariance s'écrase ou

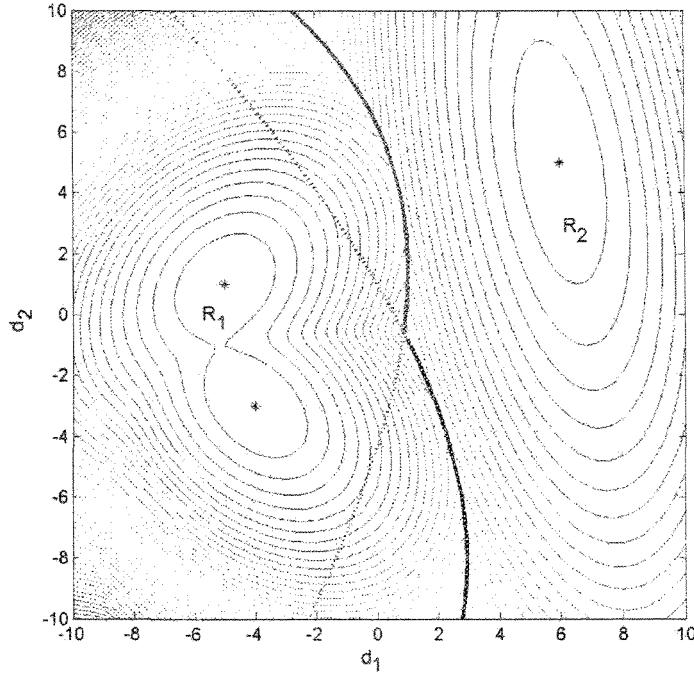


Figure 3.13: *Logarithme de la densité de probabilité d'une gaussienne (classe R_2) et d'une combinaison de deux gaussiennes (classe R_1). La ligne noire est le discriminant entre les deux classes R_1 et R_2 . Les lignes pointillées sont les prolongements des discriminants séparant chaque gaussienne formant R_1 et la gaussienne de la classe R_2 . Il est à noter que le discriminant global est une combinaison des ces deux discriminants.*

devient singulière. Ce problème pourra être contré, en diminuant le nombre de gaussiennes et ainsi augmenter la chance que les gaussiennes restantes regroupent plusieurs données. D'autres versions de l'algorithme proposent d'éliminer un regroupement lorsque celui-ci devient singulier. Nous n'évaluerons pas ces alternatives.

3.6 Mélange de distributions Student (SMM)

3.6.1 Principe

Cette méthode de modélisation a été introduite dans le but de modéliser les données de type *outliers* [50] qui sont des données éloignées des regroupements principaux et ainsi peu représentées par les modèles. Cette méthode est identique à la méthode GMM à l'exception que les distributions gaussiennes sont remplacées par des distributions de Student. La distribution de Student est définie par l'équation 3.29 [50] où le degré de

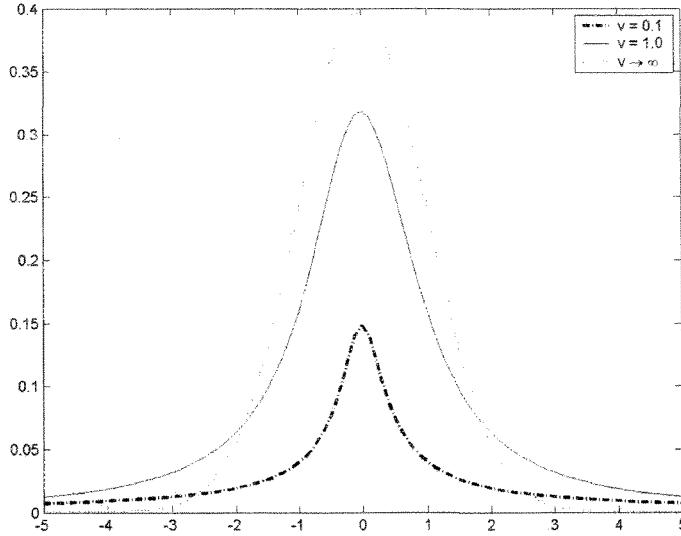


Figure 3.14: Distribution de Student à une dimension pour différentes valeurs du degré de liberté v , pour les paramètres μ et σ fixes. La limite $v \rightarrow \infty$ correspond à une gaussienne. Une valeur de v plus petite entraîne une traîne plus imposante permettant de modéliser les données éloignées (outliers).

liberté v permet de modifier la forme de la distribution. Une valeur $v \rightarrow \infty$ équivaut à une distribution gaussienne, alors que $v = 1$ correspond à une distribution de Cauchy qui a la caractéristique de posséder une traîne importante. La figure 3.14 montre qu'une distribution de Student modélise mieux les données situées loin de la moyenne, étant donné la possibilité d'obtenir une traîne plus large.

$$p(\mathbf{x}_i | \mathcal{S}_j) = \frac{\Gamma\left(\frac{v+D}{2}\right)}{\Gamma\left(\frac{v}{2}\right) (v\pi)^{\frac{D}{2}} \sqrt{|\Sigma_j|}} \times \frac{1}{\left(1 + \frac{\Delta_{ij}^2}{v}\right)^{(v+D)/2}}$$

ou Δ : Distance de Mahalanobis (3.29)

$$\Delta_{ij} = \sqrt{(\mathbf{x}_i - \mu_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \mu_j)}$$

Γ : Distribution Gamma

3.6.2 Algorithme

L'algorithme de maximisation de vraisemblance de type *EM* se prête aussi pour la détermination des paramètres des distributions de Student formant le mélange. Nous

utiliserons une valeur unique, du degré de liberté v , pour l'ensemble des distributions.

3.6.2.1 Initialisation

Cette phase est la même que celle du mélange de gaussiennes (section 3.5.4.1) à l'exception du degré de liberté v qui a été initialisé à 50. Nous avons choisi cette valeur car elle donnait les meilleurs résultats lors de la phase préliminaires d'expérimentation de la section 4.4.

3.6.2.2 Estimation (*Expectation*)

L'étape d'estimation consiste à estimer, pour chacune des données \mathbf{x}_i et chacune des distributions Student \mathcal{S}_j , l'appartenance P_{ij} de la donnée à la distribution. ($0 \leq \hat{P}_{ij} \leq 1$). Une valeur de $\hat{P}_{ij} = 1$ signifie l'appartenance totale de \mathbf{x}_i à \mathcal{S}_j [50].

Lors de cette étape, le degré de proximité u_{ij} de \mathbf{x}_i par rapport à \mathcal{S}_j est calculé. Il est à noter que les paramètres $u_{ij} \ll 1$ pour les données éloignées (*outliers*).

$$\hat{P}_{ij} = \frac{p(\mathbf{x}_i | \mathcal{N}_j) \hat{w}_j}{\sum_{m=1}^N p(\mathbf{x}_i | \mathcal{N}_m) \hat{w}_m} \quad (3.30)$$

$$\hat{u}_{ij} = \frac{D + v}{\Delta_{ij}^2 + v} \quad (3.31)$$

3.6.2.3 Maximisation

L'étape de maximisation consiste à calculer le paramètre v , les poids, les moyennes et les covariances suite aux résultats obtenus lors de l'estimation [50].

$$v = \frac{2}{y + \ln(y) - 1} + 0.0416 \left(1 + \operatorname{erf} \left(0.6594 * \ln \left(\frac{2.1971}{y + \ln(y) - 1} \right) \right) \right)$$

où : $y = -\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N \hat{P}_{ij} \left[\Psi \left(\frac{D + v}{2} \right) + \ln \left(\frac{2}{\Delta_{ij}^2 + v} \right) - \hat{u}_{ij} \right]$ (3.32)

erf : fonction d'erreur

Ψ : fonction polygamma

$$\hat{\omega}_j = \frac{1}{M} \sum_{i=1}^M \hat{P}_{ij} \quad (3.33)$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^n \hat{P}_{ij} \hat{u}_{ij} x_i}{\sum_{i=1}^n \hat{P}_{ij} \hat{u}_{ij}} \quad (3.34)$$

$$\hat{\Sigma}_j = \frac{\sum_{i=1}^n (\hat{P}_{ij} \hat{u}_{ij}) [(x_i - \hat{\mu}_j)] [(x_i - \hat{\mu}_j)]^T}{\sum_{i=1}^n \hat{P}_{ij} \hat{u}_{ij}} \quad (3.35)$$

3.7 Perceptron multi-couche (*MLP*)

3.7.1 Principe

Le perceptron multi-couche (*MLP - Multi-Layer Perceptron*) est un réseau de neurones amplément utilisé. Il comporte une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. La figure 3.15 illustre la structure interne d'un tel réseau. La couche d'entrée possède un nombre de neurones égale à la dimension des données (deux dans cet exemple). La fonction de transfert de cette couche est linéaire. La couche cachée est celle qui permet au discriminant d'épouser une forme complexe quelconque, comme le montre la figure 3.15, où le discriminant à la sortie $z(i)$ est une combinaison des sous-discriminants y_1 , y_2 et y_3 de la couche cachée. La fonction de transfert de cette couche doit être de type seuil, afin que le discriminant final soit un mélange des discriminants de la couche cachée. Le nombre de neurones de la couche de sortie est égale au nombre de classes à discriminer. Cependant, pour discriminer seulement deux classes, un seul neurone est suffisant. Nous avons utilisé, en sortie, une fonction de transfert sigmoïde tangentielle afin d'obtenir un meilleur contraste entre les classes. Il faut souligner que des fonctions du *Toolbox Matlab* ont été utilisées pour implanter les réseaux *MLP*.

3.7.2 Phase d'apprentissage

Lors de cette phase, le but est d'établir les poids W_{ij} et W_j' où i est l'indice du neurone de la couche d'entrée et j l'indice du neurone de la couche cachée. Les poids sont ajustés

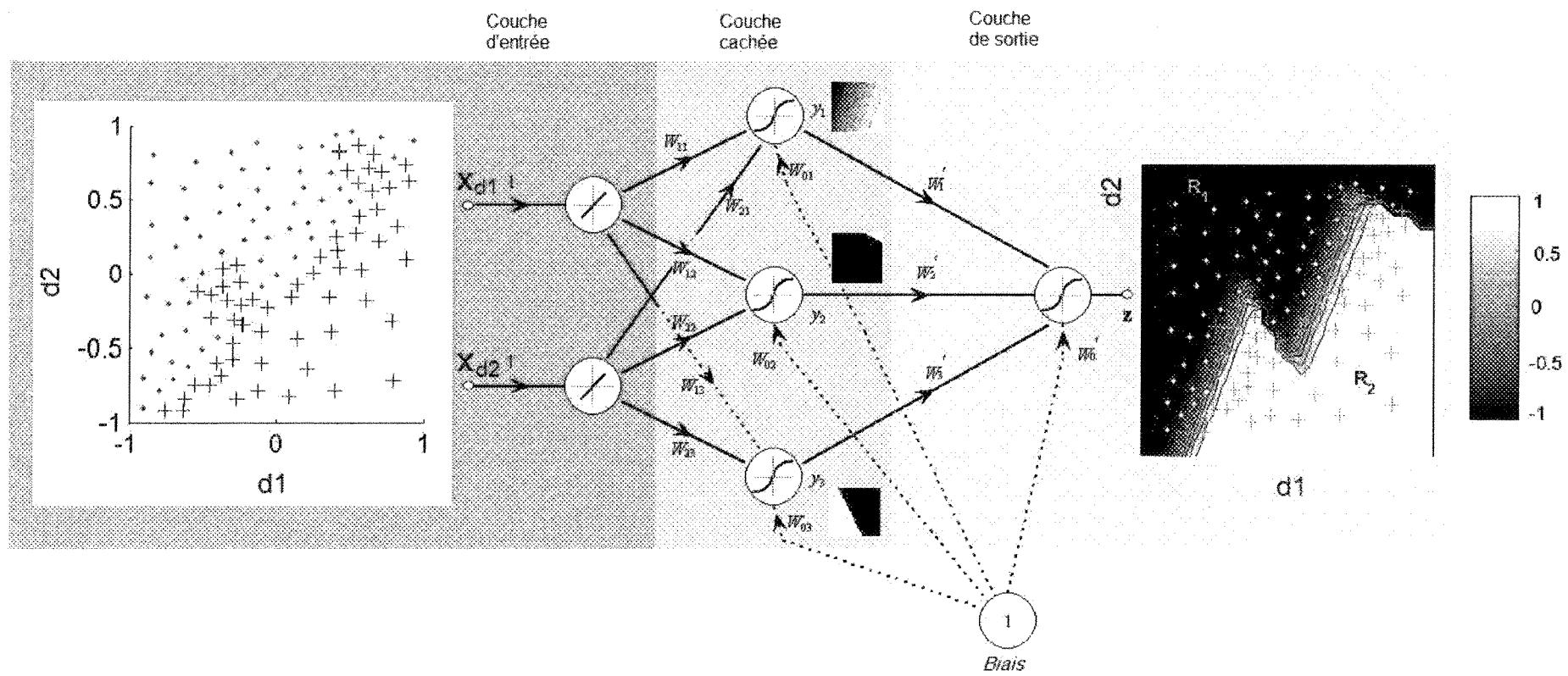


Figure 3.15: Réseau MLP possédant trois neurones dans la couche cachée. Chaque neurone de la couche cachée établit un discriminant qui contribue au discriminant final établi dans la couche de sortie.

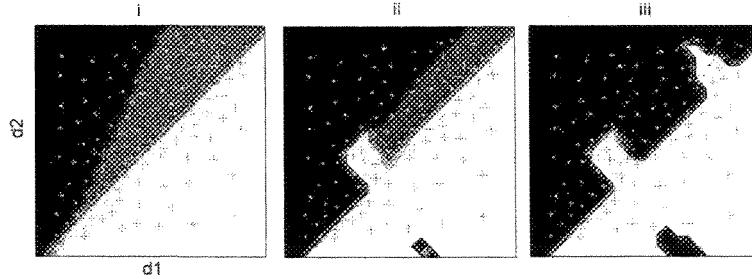


Figure 3.16: *Discriminant obtenu par réseau MLP* i) *deux neurones dans la couche cachée*, ii) *quatre neurones dans la couche cachée* et iii) *huit neurones dans la couche cachée*.

afin de discriminer les données d'entraînement. Cet ajustement des poids est effectué à l'aide d'un algorithme appelé “rétropropagation de l'erreur” qui itère jusqu'à ce qu'un certain critère de convergence soit rencontré.

3.7.3 Phase de classification

Une nouvelle donnée est présentée au réseau. Ceci équivaut à présenter la donnée à l'équation 3.36. La décision est alors prise à l'aide de la règle montrée à l'équation 3.37. Il est à noter que nous aurions pu utiliser deux neurones dans la couche de sortie et ainsi éviter l'emploi d'une telle règle de décision.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{tansig} \left(\begin{bmatrix} W_{11} & W_{21} \\ W_{12} & W_{22} \\ W_{13} & W_{23} \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} + \begin{bmatrix} W_{01} \\ W_{02} \\ W_{03} \end{bmatrix} [1] \right) \quad (3.36)$$

$$z(i) = \text{tansig} \left(\begin{bmatrix} W_1' & W_2' & W_3' \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} W_0' \end{bmatrix} [1] \right)$$

$$z(i) \stackrel{\in R_1}{\leqslant} 0 \quad (3.37)$$

La figure 3.16 illustre l'évolution du discriminant en fonction du nombre de neurones de la couche cachée.

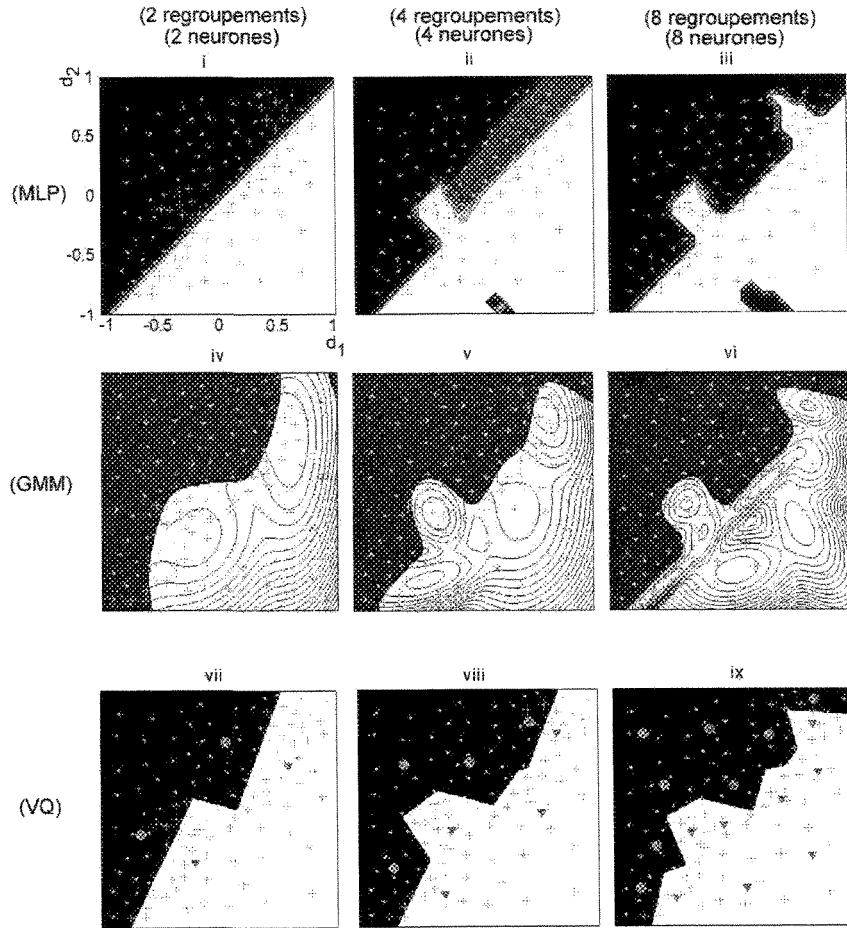


Figure 3.17: *Discriminants obtenus par les techniques MLP, GMM et VQ (selon la ligne), et pour 2, 4 et 8 regroupements ou du nombre de neurones dans la couche cachée (selon la colonne).*

3.8 Comparaison des discriminants pour différentes techniques de modélisation

La figure 3.17 montre que les trois méthodes des sections 3.4, 3.5 et 3.7 établissent en quelque sorte un discriminant semblable. Il est à noter que nous avons omis la modélisation *SMM* de la section 3.6 car, graphiquement, elle ne se distingue pas de la modélisation *GMM*. Cette figure illustre aussi que le nombre de neurones de la couche cachée d'un réseau *MLP* affecte le discriminant de la même façon que le nombre de regroupement des modélisations par *GMM* et *VQ*. Pour cette raison, pour la suite, nous parlerons du “nombre de regroupement” au lieu du “nombre de neurones dans la couche cachée”.

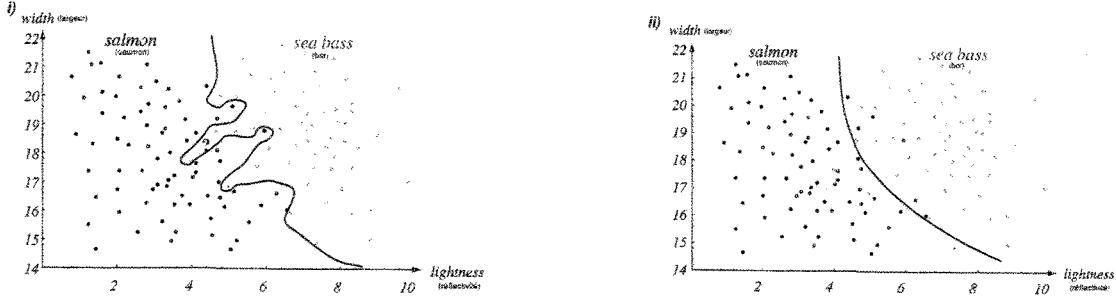


Figure 3.18: *i)* un discriminant d'ordre élevé classifie très bien les données d'entraînement, mais pas nécessairement les données en mode de reconnaissance. Par exemple, le nouveau point indiqué par ? sera classifié comme étant un “bar” alors qu'il a plus de chance d'appartenir à la classe “saumon”. *ii)* un discriminant quadratique est plus approprié dans ce cas pour la généralisation. Ces figures proviennent du manuel de Duda, Hart et Stork [49].

3.9 Contrainte d'un discriminant complexe

Il est à noter qu'il n'est pas nécessairement avantageux d'avoir une fonction discriminante trop complexe. Une fonction complexe discriminera avec succès les données d'entraînement mais pourrait donner de piétres performances en mode de reconnaissance. Ceci est illustré à la figure 3.18, tirée du manuel de Duda, Hart et Stork [49].

Le but ultime de l'entraînement est de trouver un discriminant assez complexe pour séparer ce qui caractérise les deux classes, mais pas trop complexe pour permettre la généralisation.

3.10 Post traitement

Les sibilants sont considérés comme des signaux continus d'une durée dépassant les 250 ms. Ils peuvent donc s'étendre sur plusieurs segments adjacents du signal respiratoire. Par conséquent, il devient intéressant de tenir compte de la variabilité de la fonction distance $z(i)$ en fonction du temps (en terme de segments). Ceci peut être réalisé en adoucissant la courbe de la fonction distance [29,30]. La fonction filtrée $z_f(i)$ est obtenue en convoluant la fonction originale $z(i)$ avec une fenêtre de Hamming $h(i)$, selon l'équation 3.38.

$$z_f(i) = z(i) * h(i) \quad (3.38)$$

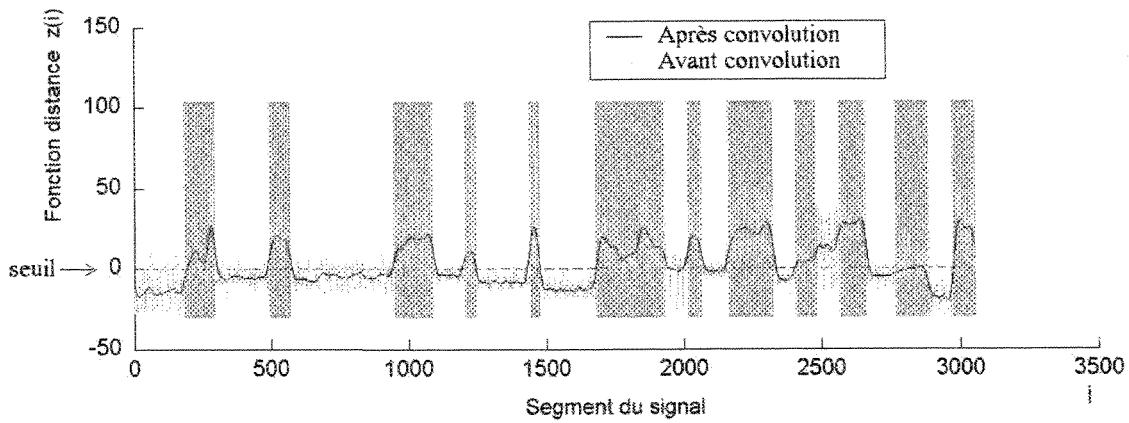


Figure 3.19: *La fonction distance $z(i)$ avant lissage (tracé gris) et après lissage (tracé noir). Les barres verticales délimitent les classes (les zones ombrées pour les sibilants et les zones blanches pour les sons normaux). Une valeur de la fonction $z(i)$ au dessus du seuil est détectée comme appartenant au sibilant.*

La figure 3.19 illustre la fonction distance avant et après le filtrage (lissage). La largeur de la fenêtre a été fixée à 51 échantillons. Nous n'analyserons pas l'effet de la largeur de la fenêtre, faute de temps.

CHAPITRE IV

MÉTHODOLOGIE ET RÉSULTATS

Les méthodes d'extraction des caractéristiques et de modélisation, exposées précédemment, ont été appliquées aux signaux respiratoires en vue de détecter les sibilants. Le présent chapitre décrit les protocoles d'expérimentation et les critères d'évaluation utilisés dans notre projet. Il présente également les résultats de nos différentes expérimentations.

4.1 Préparation de la base de données

Les données utilisées dans nos expérimentations proviennent de plusieurs sources : 1) la base de données ASTRA [24], 2) la base de données R.A.L.E. [25] et 3) de divers sites internet de laboratoires oeuvrant dans ce domaine. Nous les avons traités afin qu'une classe contienne seulement des sibilants et l'autre seulement des sons normaux. Nous savons que ces sons ont été enregistrés au niveau de la trachée mais nous ignorons toutes informations sur les sujets (âge, sexe, taille du patient, etc.) et les conditions d'enregistrement (type de capteurs, fréquence du filtre anti-repliement, nature du bruit si présent, etc.).

Nous rappelons que l'objectif du projet est la mise en place d'un système de classification des sons respiratoires fiable, robuste et insensible au bruit additionnel. Donc, le manque d'informations citées précédemment ne constitue pas un handicap parce que la classification ne doit surtout pas tenir compte de ces facteurs.

Ainsi, notre base de données regroupe deux classes : la classe des sons normaux et celle des sibilants avec les proportions respectives de 60% et 40%. Chacune des deux classes contient douze enregistrements (fichiers en format *.wav*) échantillonnés à la fréquence de 6000 Hz. Le tableau 4.1 donne la liste des fichiers des deux classes en précisant la durée (en seconde) et la taille (en échantillon) de chacun.

Tableau 4.1: Signaux formant la base de données utilisée

Sons respiratoires échantillonnés à 6000 Hz					
Normaux			Sibilants		
Nom du fichier	Nombre d'échantillons	Durée sec	Nom du fichier	Nombre d'échantillons	Durée sec
NN6000_01	94076	15.68	WW6000_01	55277	9.21
NN6000_02	102823	17.14	WW6000_02	38823	6.47
NN6000_03	194369	32.39	WW6000_03	71331	11.89
NN6000_04	60592	10.10	WW6000_04	21066	3.51
NN6000_05	101032	16.84	WW6000_05	16985	2.83
NN6000_06	106630	17.77	WW6000_06	126174	21.03
NN6000_07	46063	7.68	WW6000_07	25273	4.21
NN6000_08	49326	8.22	WW6000_08	83973	14.00
NN6000_09	41033	6.84	WW6000_09	40382	6.73
NN6000_10	43431	7.24	WW6000_10	48431	8.07
NN6000_11	54939	9.16	WW6000_11	59396	9.90
NN6000_12	47431	7.91	WW6000_12	40326	6.72
Total	941745	156.96	Total	627437	104.57
Proportion	60%		Proportion	40%	

4.2 Protocole d'expérimentation

Nous rappelons qu'un système de classification fonctionne en deux phases : la phase d'entraînement et la phase de test. Pour une évaluation rigoureuse de ces systèmes, les données servant à l'entraînement ne doivent pas être présentées lors du test.

La méthode “*leave one out*” est utilisée afin d'évaluer les différents systèmes de classification. Telle qu'illustré à la figure 4.1, cette méthode consiste à utiliser 11 fichiers de chacune des deux classes pour entraîner le système de classification et à prendre le fichier restant dans chacune des classes pour le tester. Par rotation, l'opération est répétée douze fois pour tester l'ensemble des 12 fichiers. Ainsi, chaque fichier est utilisé onze fois pour l'entraînement et une seule fois lors du test d'une classe.

4.3 Critères d'évaluation du système de reconnaissance

Pour la suite, les différents systèmes de classification seront identifiés de la façon suivante : [caractéristiques / modélisation]. De plus la terminologie suivante sera employée :

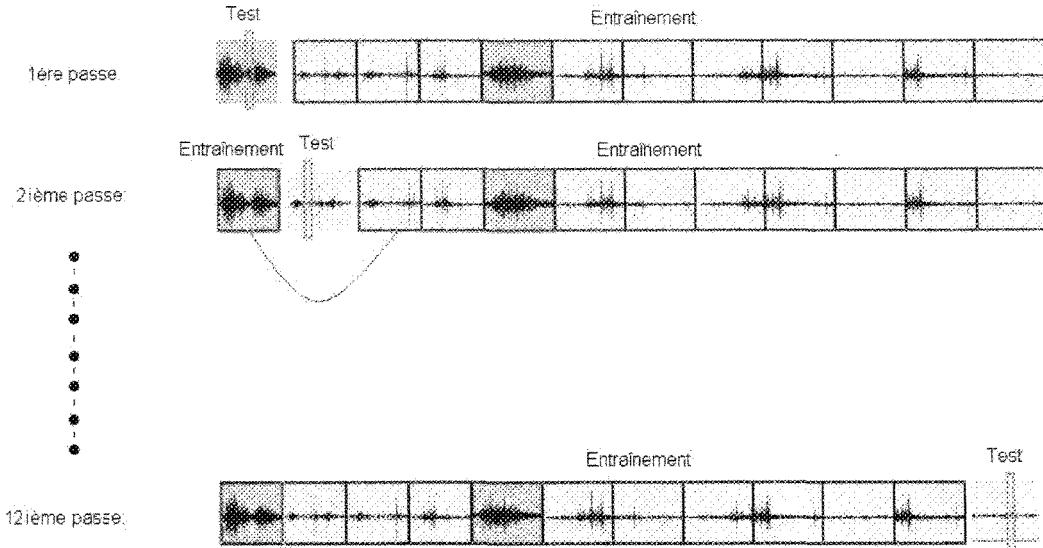


Figure 4.1: *Principe de la méthode “leave one out”, illustrée pour une seule classe. On a ici douze fichiers de classe normale dont onze servent à l’entraînement et l’autre au test. Ceci est répété jusqu’à ce que tous les fichiers aient servi de fichier test. Le rectangle apparaissant dans un fichier test indique le segment en train d’être testé.*

- VN : Vrai Normal : Nombre de segments de son normal étant classifiés normal ;
- VS : Vrai Sibilant : Nombre de segments de son sibilant étant classifiés sibilant ;
- FN : Faux Normal : Nombre de segments de son sibilant étant classifiés normal ;
- FS : Faux Sibilant : Nombre de segments de son normal étant classifiés sibilant.

Différents auteurs utilisent différents critères afin d’évaluer leurs systèmes de reconnaissance. Les critères de spécificité et de sensibilité sont souvent utilisés dans la littérature. La spécificité et la sensibilité sont respectivement définies par les équations 4.1 et 4.2

$$\text{Spécificité} = \% \text{ Vrai Sibilant} = \frac{\text{VS}}{\text{VS} + \text{FN}} \quad (4.1)$$

$$\text{Sensibilité} = \% \text{ Vrai Normal} = \frac{\text{VN}}{\text{VN} + \text{FS}} \quad (4.2)$$

Le critère de *spécificité* a été utilisé par Pesu et al. [8] pour un système de classification de type [WBC/VQ] et par Kandaswamy et al. [12] pour un système de type [DWTC/ANN]. Le critère incluant la *spécificité* et la *sensibilité* a été utilisé par Sankur et al. [28] pour un système de type [AR/kmeans], par Waitman et al. [10] pour un

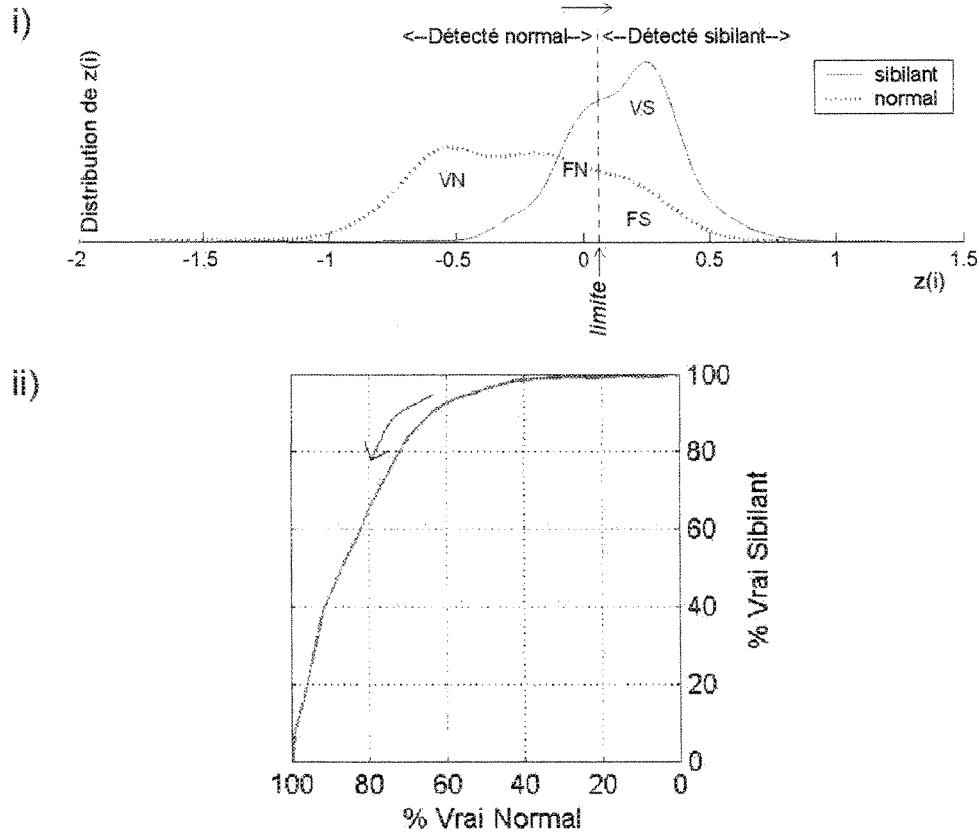


Figure 4.2: Courbe ROC obtenue à partir de la distribution de la fonction distance $z(i)$. VN : Vrai Normal, VS : Vrai Sibilant, FN : Faux Normal, FS : Faux Sibilant. La flèche indique le sens du balayage de la limite, lors du traçage de la courbe ROC.

système de type [FFT/ANN] et par Forkheim et al. [27] pour les systèmes de type [FFT/MLP, FFT/RBF, FFT/SOM et FFT/LVQ].

L'utilisation de la spécificité et de la sensibilité rend très difficile la comparaison de différents systèmes de classification étant donné que ces deux valeurs sont inversement reliées l'une à l'autre. Nous avons donc opté pour la représentation sous forme de courbe de caractéristiques d'efficacité (ROC - Receiver Operating Characteristic) car cela nous permettra d'utiliser l'aire sous la courbe (AUC - Area Under Curve) comme technique de comparaison entre les différents systèmes de classification.

Voyons comment obtenir une courbe ROC à partir des distributions de distances $z(i)$ obtenues en testant tous les segments des deux classes. La figure 4.2-i montre la distribution des distances $z(i)$, provenant des équations 3.12, 3.20 ou 3.37. Ces distances $z(i)$ sont obtenues pour l'ensemble des segments $s_i(n)$ issus de chacune des deux classes,

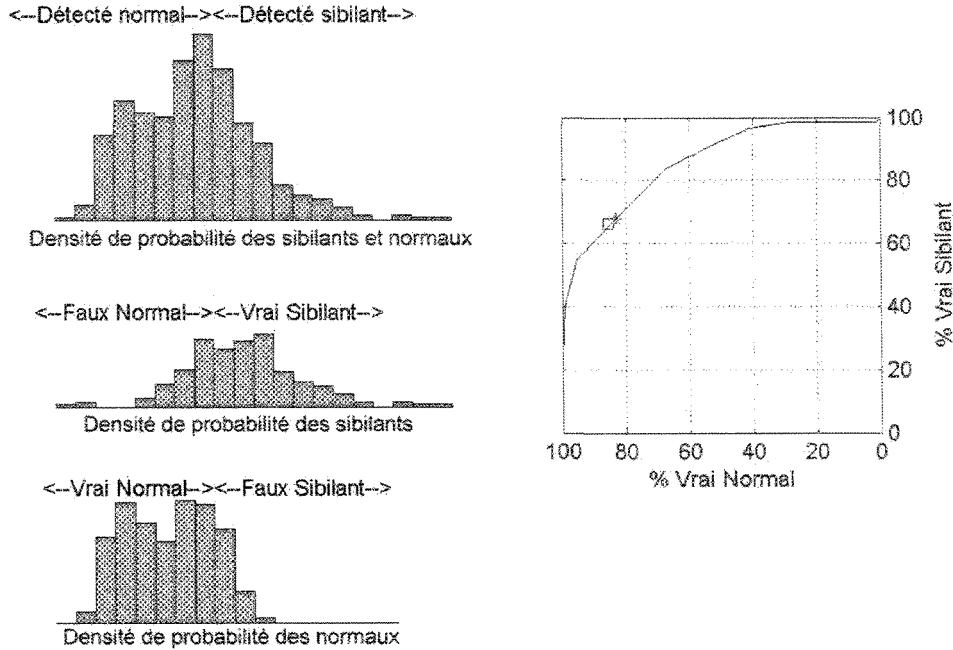


Figure 4.3: Courbe ROC obtenue à partir de la distribution distance $z(i)$ sous forme d'histogramme. L'astérisque (*) marque l'endroit où la distance $z(i)=0$. Le carré (□) indique le résultat obtenu selon un critère basé sur la spécificité et la sensibilité. L'astérisque et la carré se superposeraient, dans un cas où la courbe ROC serait tracée à partir de densités continues.

lors de la phase de test. La figure 4.2 illustre la façon dont la courbe ROC est obtenue. En faisant glisser la limite entre les deux courbes de la figure 4.2-i de gauche à droite, on obtient la courbe de la figure 4.2-ii, suite à l'utilisation des équations 4.1 et 4.2.

En pratique, les distributions de la fonction distance $z(i)$ sont discrètes, telles que représentées sous forme d'histogrammes à la figure 4.3. Le nombre d'intervalle d'une distribution pourra affecter l'allure de la courbe ROC. Nous avons choisi d'établir le nombre d'intervalle selon la règle de Sturge [51] décrite par l'équation 4.3 :

$$\text{Nombre d'intervalle} = [1 + 3.3 * \log_{10}(\text{Nombre de données})] \quad (4.3)$$

Nous avons remarqué que le critère regroupant la spécificité et la sensibilité donne toujours un point, tel que le carré (□) de la figure 4.3, situé quelque part sur la courbe ROC. L'astérisque, positionné sur la même figure, indique la valeur de spécificité et de sensibilité lorsque la limite est positionnée à $z(i) = 0$, ceci correspond au discriminant.

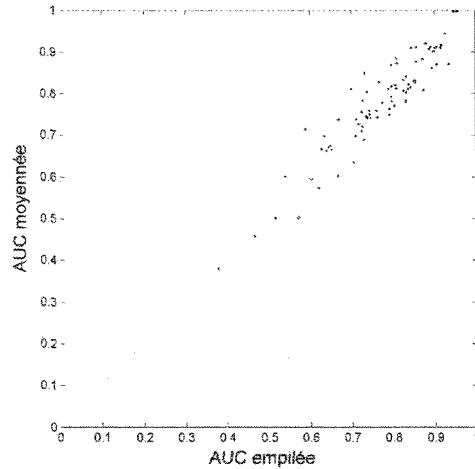


Figure 4.4: *Les approches par aire sous la courbe empilée et aire sous la courbe moyennée donnent des résultats semblables. Chaque point de la figure, déterminé par les valeurs des deux aire sous la courbe (AUC) représente le résultat d'un test de classification par la méthode “leave one out”.*

Le carré et l'astérisque seraient en fait superposés si les distributions étaient continues au lieu d'être discrètes.

La technique *leave one out* que nous avons utilisée, nous donne douze courbes *ROC* relatives aux tests des 24 fichiers des sons respiratoires. La $k^{\text{ième}}$ courbe *ROC* provient du $k^{\text{ième}}$ fichier normal et du $k^{\text{ième}}$ fichier sibilant. La question qui se pose est la suivante : serait-il possible de représenter les résultats des 24 fichiers par une seule courbe *ROC* ? Deux solutions peuvent être envisagées [52]. Soit qu'on compile toutes les distances $z(i)$ obtenues lors des douze passes afin de tracer une seule courbe *ROC* ou soit qu'on calcule la moyenne des douze courbes *ROC* obtenues séparément. La figure 4.4 donne les résultats de classification sous forme de “AUC empilée” et “AUC moyennée” pour différents tests réalisés avec les 24 fichiers. Ces tests diffèrent par le type du système de classification employé [caractéristiques/modélisation], la dimension des caractéristiques (D), le nombre de regroupement (N) et autres paramètres spécifiques. Cette figure montre que les deux méthodes donnent des résultats semblables. Pour mieux illustrer le principe des deux méthodes, nous présentons à la figure 4.5 les résultats d'un test *leave one out* de sons respiratoires provenant de 4 fichiers seulement (pour des raisons de lisibilité). Pour la suite du projet, nous opterons pour la méthode “AUC empilée”.

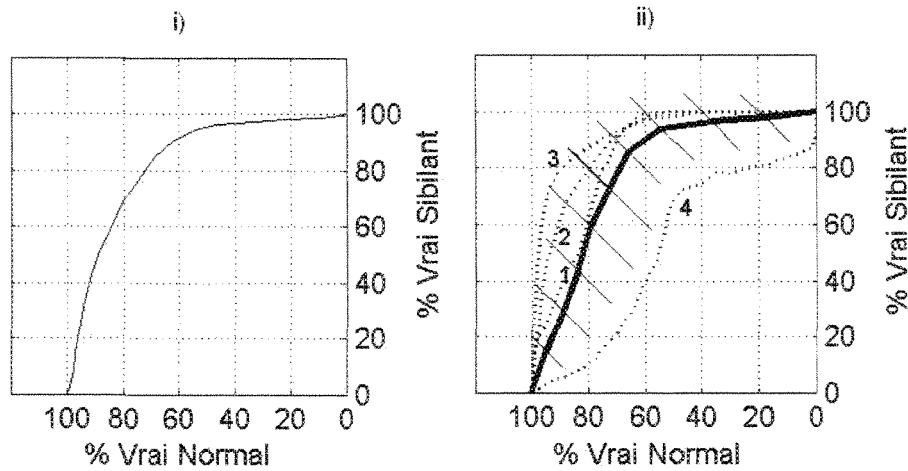


Figure 4.5: Courbes de caractéristiques d'efficacité ROC i) obtenue en empilant toutes les distances $z(i)$ provenant de quatre fichiers, ii) même courbe obtenue en moyennant quatre courbes des distances $z(i)$ provenant des mêmes fichiers. Les barres d'erreur indiquent l'écart-type de part et d'autres de la moyenne.

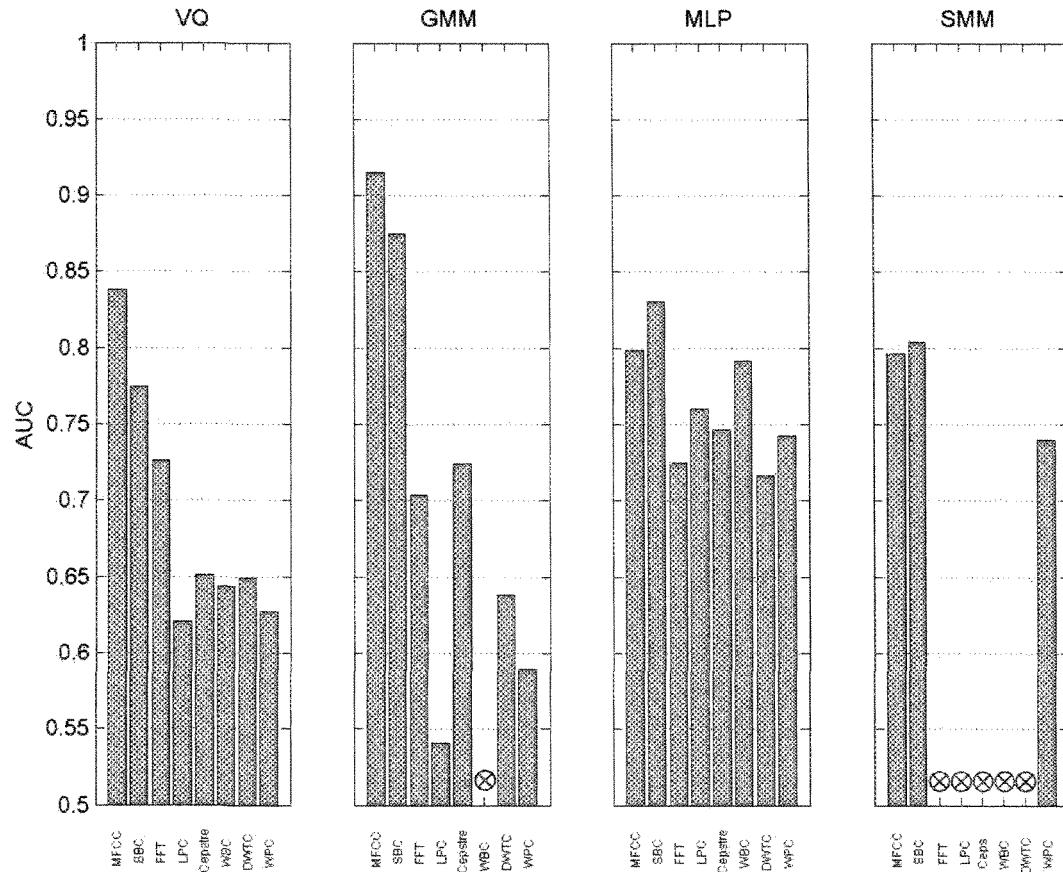


Figure 4.6: Aire sous la courbe (AUC) des différentes combinaisons [caractéristiques/modélisation]. L'absence de résultat, remplacée par \otimes , indique qu'une matrice de covariance est devenue singulière et que le programme ne s'est pas terminé normalement.

Tableau 4.2: Paramètres de l'expérimentation préliminaire

	<i>VQ</i>	<i>GMM</i>	<i>MLP</i>	<i>SMM</i>
Dimension	24	24	24	24
Largeur des segments	512	512	512	512
Recouvrement (%)	50	50	50	50
Nombre de regroupements	8	8	30	8
Nombre d'itérations	50	15	50	15
Matrice de covariance	-	diagonale	-	pleine
Type d'initialisation	-	LBG	-	LBG
Prétraitement	normalisation	normalisation	normalisation	normalisation
Post traitement	aucun	aucun	aucun	aucun

4.4 Résultats préliminaires

Nous allons analyser différents systèmes de classification obtenus par différentes combinaisons de techniques d'extraction des caractéristiques et de modélisation. Vu le nombre important de combinaisons et de paramètres sous-jacents, nous procéderons par élimination. En effet, nous abandonnerons, tout au long de l'expérimentation, les techniques qui donnent de mauvais résultats et ne garderons que celles qui semblent prometteuses. Nous sommes conscient du fait qu'un système [caractéristiques/modélisation] pourrait donner des résultats insignifiants pour certaines valeurs de paramètres mais pourrait s'avérer bénéfique en optimisant ces valeurs.

Pour débuter notre analyse, nous avons utilisé les valeurs listées au tableau 4.2 pour obtenir les premiers résultats. Il faut souligner que la dimension D de l'espace des caractéristiques, initialement fixée à 24, peut être réduite selon la technique de caractérisation utilisée. Par exemple, dans le cas le plus extrême, la technique *WPC* ramène la dimension du vecteur des caractéristiques à $D=2$.

La figure 4.6 présente les résultats préliminaires que nous avons obtenus. Les performances des différentes combinaisons sont représentées par les barres verticales indiquant la valeur de l'aire sous la courbe (*AUC*). Il faut noter que l'absence de résultats, remplacée par le symbole \otimes , indique que la matrice de covariance est devenue singulière et que le programme ne s'est pas terminé normalement, tel que discuté à la section 3.5.4.5. À la suite de ces premiers résultats, il nous paraît logique d'abandonner la modélisation *SMM* car elle est trop instable et ne donne pas des résultats comparables à ceux obtenus par

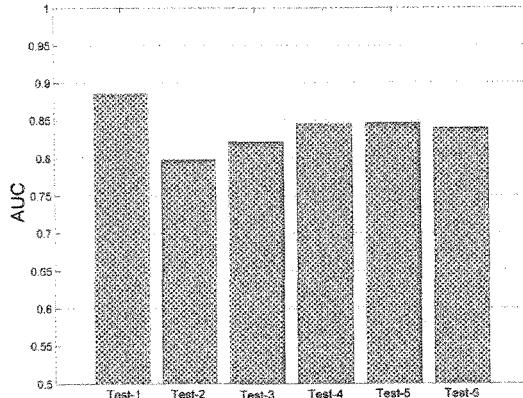


Figure 4.7: *Variabilité des résultats (AUC) de classification MLP obtenus de six tests successifs effectués sur le même groupe de fichiers, selon la méthode “leave one out”. Il est à noter que le réseau est réinitialisé puis entraîné à chacun des tests.*

les modélisations *GMM*, *MLP* et *VQ*.

Pour ce qui a trait aux techniques d'extraction des caractéristiques, la figure 4.6 montre que les performances obtenues par la technique *MFCC* et la technique *SBC* dépassent de loin celles obtenues par les autres techniques et ce peu importe le type de modélisation employé. Par conséquent, nous retiendrons uniquement ces deux méthodes d'extraction des caractéristiques.

4.5 Non-reproductibilité des résultats du réseau *MLP*

Contrairement aux méthodes de modélisation *VQ* et *GMM*, la méthode *MLP* donne des résultats sensiblement différents à chaque test effectué selon la méthode *leave one out*. Ce manque de reproductibilité des résultats est causé par l'initialisation aléatoire des poids du réseau lors de la phase d'apprentissage relative à chacun des tests. Ceci rend son évaluation plus délicate. Il est à noter que même un nombre d'itération très élevé (500 à 1000) ne règle en rien ce problème. La figure 4.7 montre les performances, sensiblement différentes, obtenues suite à six différents tests. Nous devons être conscient, lors des comparaisons qui suivent, que la performance obtenue, suite à une classification par *MLP*, ne constitue qu'une expérience d'une variable aléatoire. Les calculs sont très laborieux pour envisager de prendre la moyenne des tests effectués dans des conditions similaires.

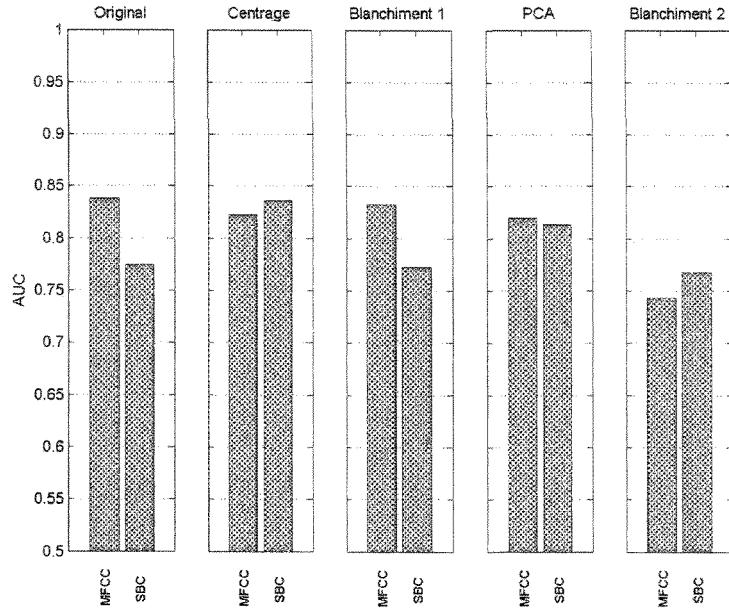


Figure 4.8: *Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation VQ. Il est à noter que la performance est détériorée pour les quatres méthodes de prétraitement, dans le cas de la caractérisation par la technique MFCC. Par contre, il y a amélioration par les méthodes centrage et PCA, dans le cas de la caractérisation par la technique SBC.*

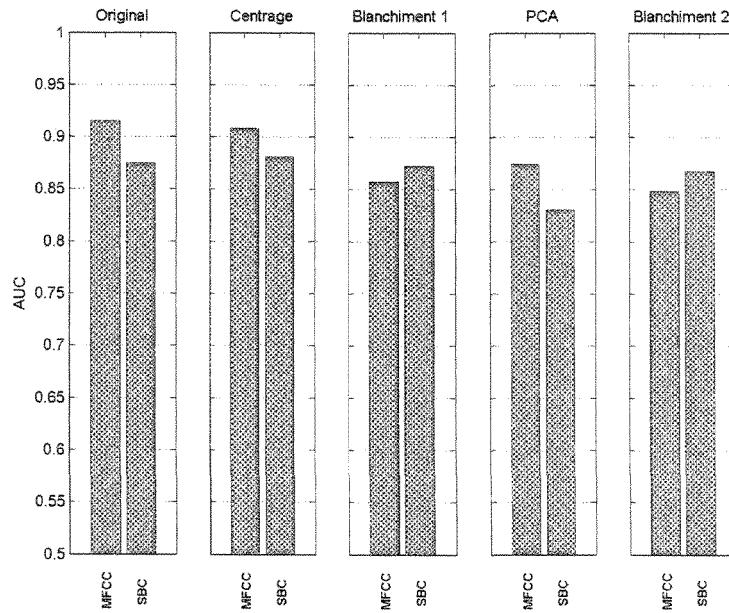


Figure 4.9: *Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation GMM. Il est à noter que la performance est détériorée pour les quatres méthodes de prétraitement, dans le cas de la caractérisation par la technique MFCC. Par contre, il y a amélioration par la méthode centrage, dans le cas de la caractérisation par la technique SBC.*

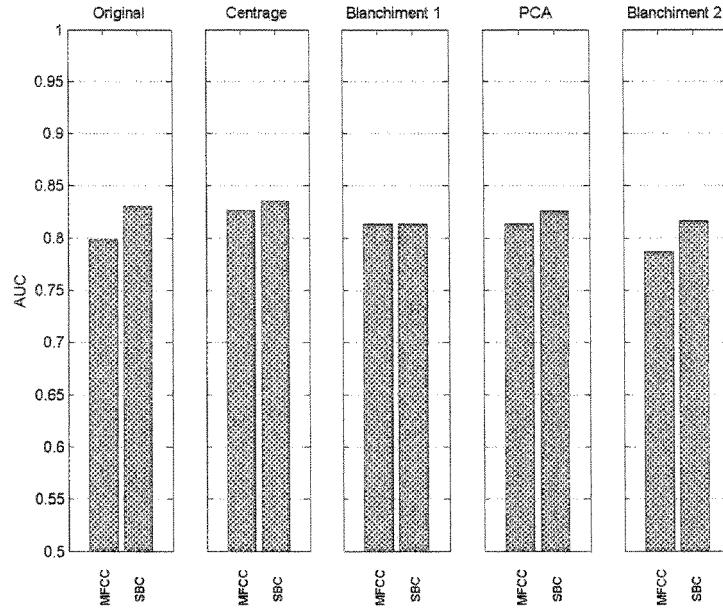


Figure 4.10: *Effets du prétraitement sur la performance (AUC) d'un système de classification basé sur la modélisation par réseau de neurones MLP. Il est à noter que la performance est améliorée pour les méthodes centrage, blanchiment-1 et PCA, dans le cas de la caractérisation par la technique MFCC. Par contre, il y a amélioration seulement par la méthode centrage, dans le cas de la caractérisation par la technique SBC.*

4.6 Effets de différents paramètres

4.6.1 Effets du prétraitement

L'opération de normalisation, telle qu'introduite à la section 2.6.1.1, a été appliquée à tous les signaux de la base de données avant l'extraction des caractéristiques. Les effets des techniques de prétraitements appliquées aux vecteurs des caractéristiques sont étudiés dans cette section. Ces techniques de prétraitements, décrites à la section 2.6, sont les suivantes : centrage, blanchiment-1, blanchiment-2 et l'analyse par composantes principales (*PCA*). Les figures 4.8, 4.9 et 4.10 donnent les résultats obtenus suite à l'utilisation des différentes techniques de prétraitements pour les modélisations *VQ*, *GMM* et *MLP*. Le prétraitements par centrage améliore la performance des caractéristiques *SBC* pour toutes modélisations. Cette technique semble prometteuse mais nous ne l'avons pas utilisée dans ce qui suit car l'amélioration apportée ne dépasse jamais les meilleurs résultats obtenus avec la technique d'extraction *MFCC*. La figure 4.10 illustre que les prétraitements par

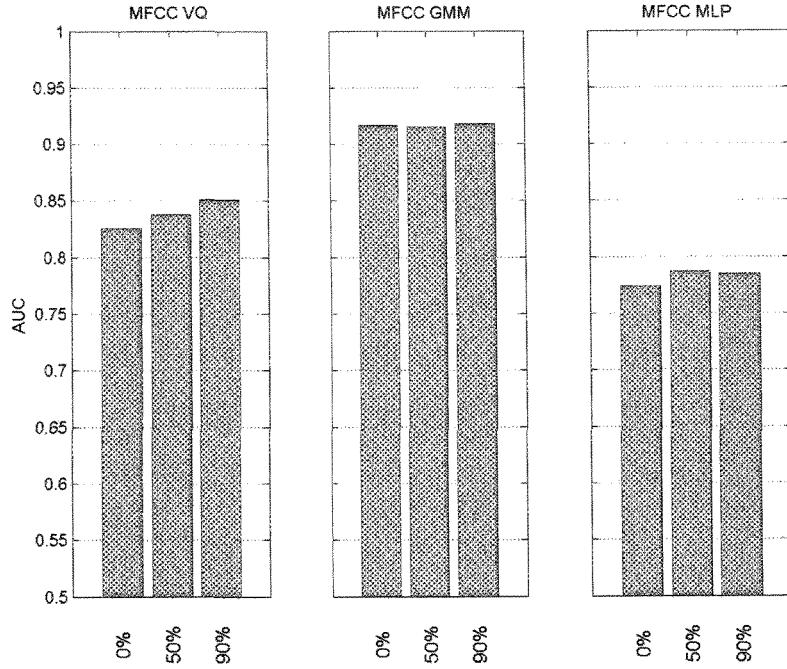


Figure 4.11: *Effets du recouvrement sur la performance (AUC) de la classification, dans le cas d'une caractérisation par la méthode MFCC. Un taux de recouvrement élevé est bénéfique mais au détriment du temps de calcul.*

centrage et par la technique *PCA* sont utiles pour le réseau de neurone de type *MLP*. À la lumière de ces résultats, nous avons décidé de poursuivre notre analyse en utilisant que le prétraitement par normalisation du signal avant extraction des caractéristiques.

4.6.2 Effets du recouvrement

Trois valeurs de taux de recouvrement ont été testées dans ce projet, afin d'évaluer l'effet du recouvrement sur les performances des systèmes de classification, soit 0%, 50% et 90%. De plus, nous nous sommes limités à la technique d'extraction des caractéristiques *MFCC*, pour l'évaluation. La figure 4.11 montre que le recouvrement tend à améliorer la performance mais ceci au détriment du temps de calcul. En effet, le nombre de segments traités est proportionnel au taux de recouvrement. Nous opterons pour un recouvrement de 50% comme compromis entre les performances de la classification et le temps de calcul.

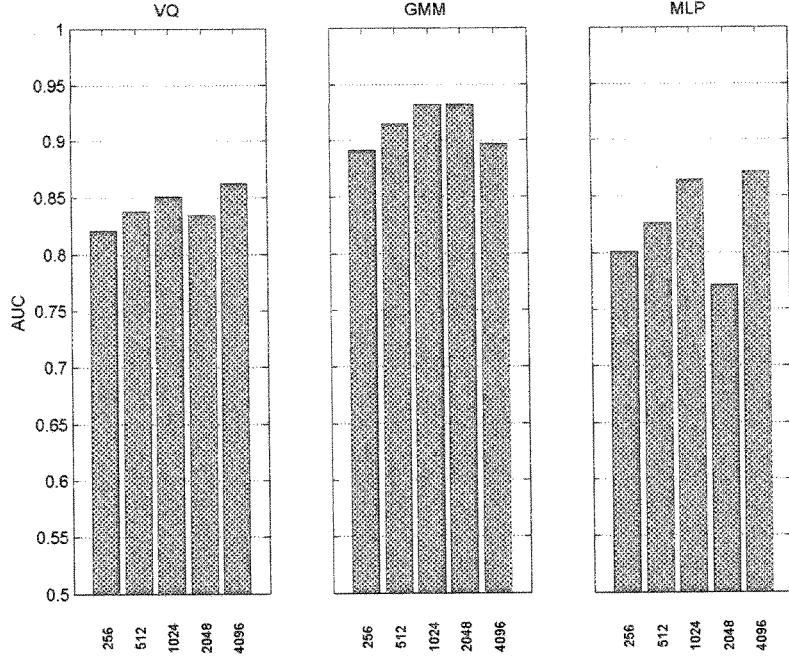


Figure 4.12: *Effets de la largeur de la fenêtre d'analyse sur la performance (AUC) de la classification, dans le cas d'une caractérisation par la méthode MFCC. Un segment de 1024 échantillons augmente la performance pour les trois types de modélisation. Il est à noter que les nombres entiers en abscise représentent la largeur de la fenêtre d'analyse en échantillons.*

4.6.3 Effets de la largeur de la fenêtre d'analyse

Différentes largeurs de la fenêtre d'analyse ont été testées, soit 256, 512, 1024, 2048 et 4096 échantillons. Dans cette expérimentation, nous n'avons évalué que la technique d'extraction des caractéristique *MFCC*. La figure 4.12 montre qu'un segment de 1024 échantillons donne le premier maximum local, pour les trois méthodes de modélisation. Nous fixerons donc la largeur du segment d'analyse à 1024 échantillons. Ceci donnera les résolutions suivantes :

$$\Delta f = \frac{1}{\frac{L}{f_e}} = \frac{f_e}{L} = \frac{6000}{1024} = 5.86 \text{ Hz} \quad (4.4)$$

$$\Delta t = \frac{1}{\Delta f} = 171 \text{ ms} \quad (4.5)$$

où L est la largeur de la fenêtre et f_e est la fréquence d'échantillonnage. Il est à noter que nous sommes dans l'ordre de grandeur estimée à la section 2.2.1 où l'on avait déduit une résolution fréquentielle (Δf) de 11.7 Hz et une résolution temporelle (Δt) de 85.3 ms.

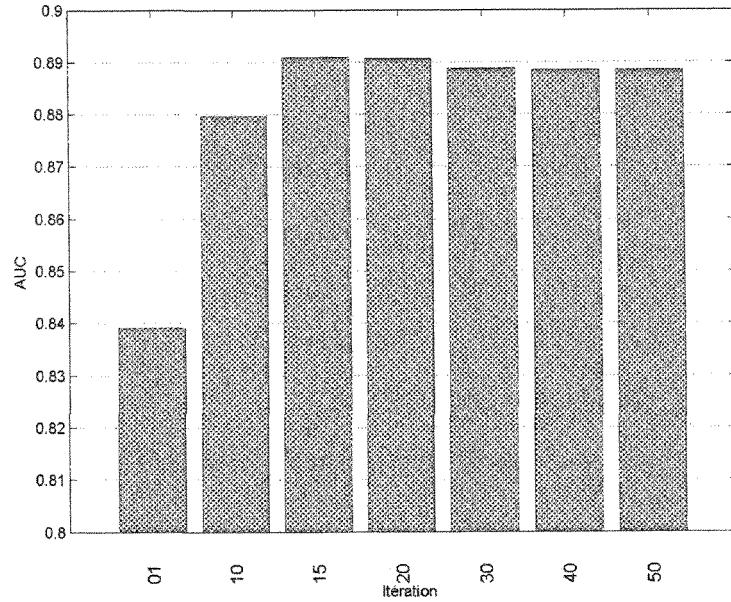


Figure 4.13: *Effets du nombre d’itérations sur la performance (AUC) d’un système de classification MFCC/VQ. Les paramètres utilisés sont : $M=1024$ $D=24$ et $N=8$.*

4.6.4 Effets du nombre d’itérations

Les figures 4.13, 4.14 et 4.15 montrent les performances obtenues selon le nombre d’itérations effectuées lors de l’entraînement. Encore une fois, nous n’avons utilisé que la technique d’extraction des caractéristiques *MFCC*. À la suite de ces expérimentations, les valeurs finales retenues sont de 15 itérations pour la modélisation *VQ*, 10 itérations pour la modélisation *GMM* et 20 itérations pour le réseau *MLP*. Il est intéressant de constater que les performances pour une seule itération sont tout de même intéressantes. Il est à noter que la méthode *GMM* contient deux types d’itérations : celle de l’initialisation, que nous avons fixé à 15, et celle de l’entraînement, qui est présentée à la figure 4.14. Nous avons vu, à la section 3.9 qu’un nombre d’itérations élevé contribue à surentraîner le modèle le rendant donc moins apte à la généralisation. En effet, la figure 4.14 montre la déterioration des performances suite à une augmentation du nombre d’itérations.

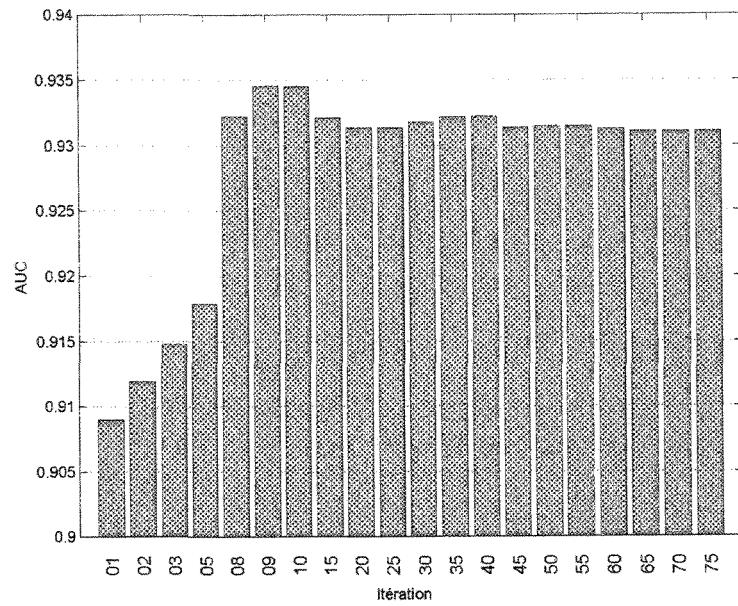


Figure 4.14: *Effets du nombre d'itérations sur la performance (AUC) d'un système de classification [MFCC/GMM]. Les paramètres utilisés sont : $M=1024$, $N=8$, $D=24$, init=LBG et Cov=diag.*

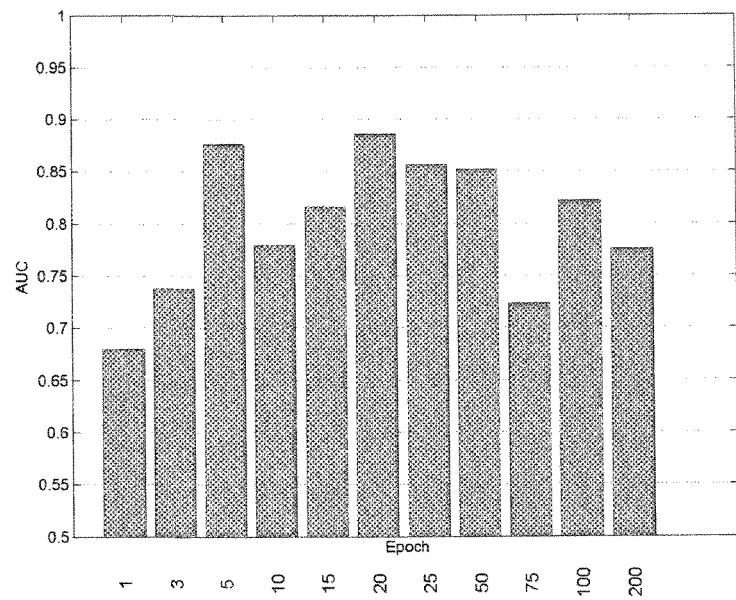


Figure 4.15: *Effets du nombre d'itérations sur la performance (AUC) d'un système de classification [MFCC/MLP]. Les paramètres utilisés sont : $M=1024$, $N=8$ et $D=24$.*

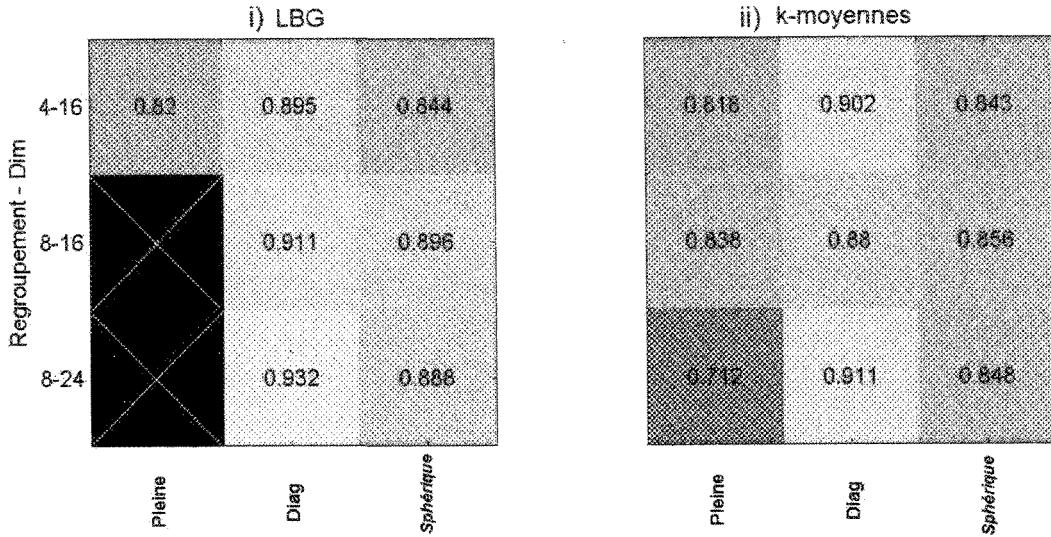


Figure 4.16: Aires sous la courbe (AUC) obtenues pour le système de classification [MFCC/GMM], selon que le type d'initialisation soit par LBG ou bien par k-moyennes. L'effet du type de matrice de covariance est aussi représenté. On retiendra qu'une initialisation par LBG ainsi que l'utilisation d'une matrice de covariance diagonale donne de meilleurs résultats. Les paramètres utilisés sont : $M=1024$ et $itération=15$.

4.6.5 Effets du type d'initialisation et du type de matrice de covariance sur la classification par la méthode **GMM**

Telle qu'illustré à la figure 4.16, l'utilisation d'une matrice de covariance diagonale donne de meilleurs résultats, soit une aire sous la courbe (AUC) de 0.932, que l'utilisation d'une matrice de covariance pleine qui elle donne une aire sous la courbe (AUC) de 0.838. De plus, cette dernière est beaucoup plus complexe à programmer et plus vorace en temps de calcul. Telle que vue à la section 3.5.1, une matrice de covariance pleine épouse mieux les données d'entraînement qu'une matrice diagonale menant ainsi à un surentraînement. Cette figure illustre aussi que l'initialisation par l'algorithme *LBG* donne de meilleurs résultats.

4.6.6 Effets de la dimension et du nombre de regroupements

Nous étudierons ici l'effet combiné de la dimension (D) des caractéristiques et du nombre de regroupements (N) du modèle, étant donné que ces deux paramètres semblent être intimement liés. Le terme regroupement signifie : la taille du dictionnaire pour la

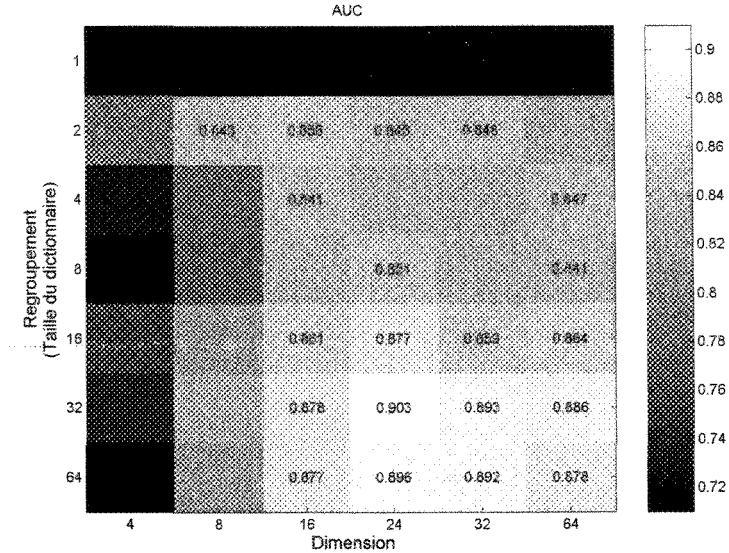


Figure 4.17: *Effets de la dimension et du nombre de regroupements sur la performance (AUC) de la classification [MFCC/VQ]. Les paramètres utilisés sont : $M=1024$ et $itération=50$.*

modélisation *VQ*, le nombre de gaussiennes pour la modélisation *GMM*, ou le nombre de neurones de la couche cachée pour le réseau *MLP*. Notons que cette expérimentation n'a pas été effectuée avec les valeurs d'itération optimales établies précédemment. Les paramètres utilisés sont listés sous chaque figure.

La figure 4.17 représente, au moyen du niveau de gris, les performances (AUC) du système de classification [MFCC/VQ] pour différentes dimensions du vecteur des caractéristiques et différentes valeurs du nombre de regroupements. Nous retiendrons, de cette figure, que le meilleur résultat ($AUC=0.903$) est obtenu avec la combinaison de 24 dimensions et de 32 regroupements. De plus, il est intéressant de noter que l'utilisation de seulement deux regroupements donne de bons résultats qui sont supérieurs à 0.843.

La figure 4.18 donne les performances de la modélisation *GMM* pour deux types de vecteur de caractérisatiques, soit la technique *MFCC* et la technique *SBC*. Dans le cas de la caractérisation par la technique *MFCC* (voir la figure 4.18-i), nous retiendrons que le meilleur résultat ($AUC=0.932$) est obtenu avec le système de classification comprenant la combinaison de 24 dimension et de 8 gaussiennes. Comme dans le cas précédent, une combinaison comprenant seulement deux regroupements, par exemple la combinaison de 24 dimensions et de 2 gaussiennes donne des résultats intéressants ($AUC=0.931$). Par contre,

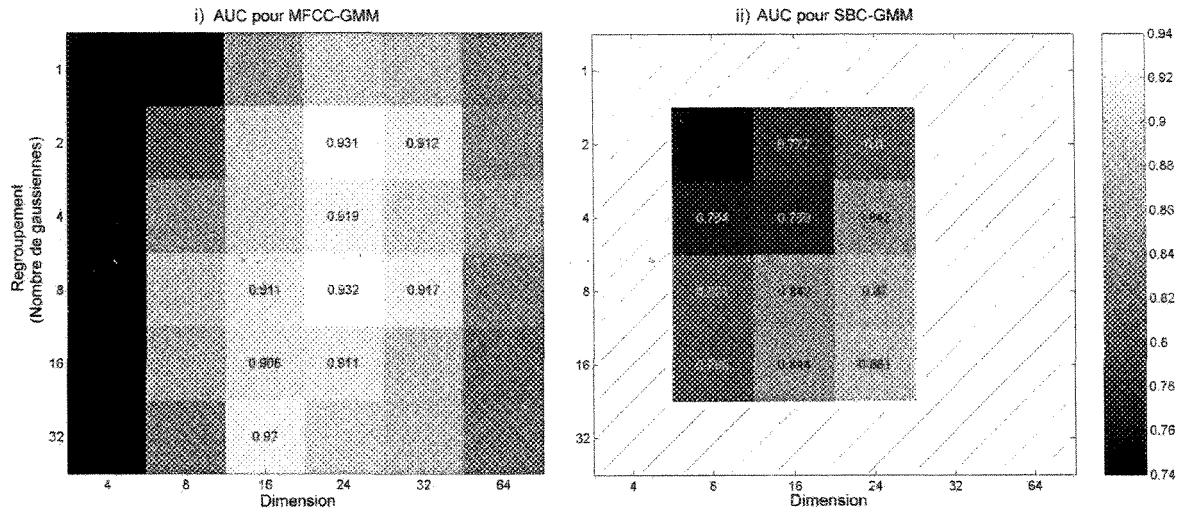


Figure 4.18: *Effets de la dimension et des regroupements sur la performance (AUC) de systèmes de classifications de combinaisons [MFCC/GMM] et [SBC/GMM]. Il est à noter que la zone hachurée indique des combinaisons non-testées.* i) Pour la combinaison [MFCC/GMM], les paramètres utilisés sont : $M=1024$, $init=LBG$, $cov=diag$ et $itération=15$, ii) Pour la combinaison [SBC/GMM], les paramètres utilisés sont : $M=512$, $init=LBG$, $cov=diag$ et $itération=10$.

la caractérisation par la technique *SBC* (voir la figure 4.18-ii), est moins performante que par la technique *MFCC*. Le meilleur résultat (AUC=0.881) est obtenu avec la combinaison de 24 dimensions et de 16 gaussiennes. Étant donné l'énorme temps de calcul, cette combinaison a été évaluée seulement pour les valeurs 8, 16 et 24 dimensions et les valeurs 2, 4, 8 et 16 pour le nombre de regroupement (nombre de gaussiennes). Donc les valeurs manquantes (zones hachurées) de la figure 4.18-ii n'ont pas de signification pour la palette de couleur, d'où les zones sans résultat de la figure 4.18-ii. Pour la suite, nous laisserons tomber les vecteurs de caractéristiques de type *SBC*, étant donné que ses performances sont inférieures à celles obtenues par la technique *MFCC*.

Finalement, la figure 4.19 donne les performances du système de classification [MFCC/MLP]. Le meilleur résultat (AUC=0.878) a été obtenu avec la combinaison de 24 dimensions et 64 neurones dans la couche cachée.

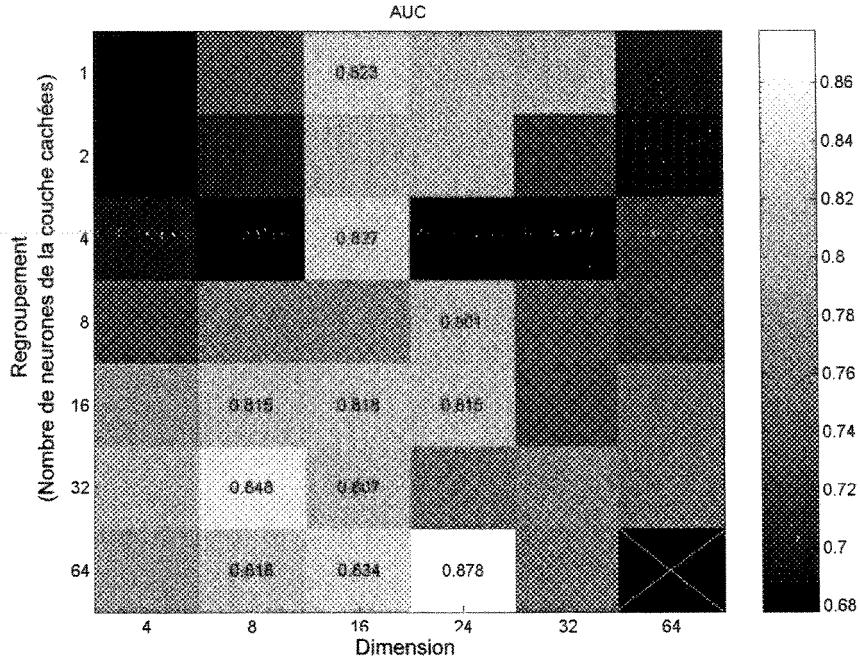


Figure 4.19: *Effets de la dimension et des regroupements sur la performance (AUC) d'un système de classification de type [MFCC/MLP].* Les paramètres utilisés sont : $M=1024$ et itération=20.

4.7 Résultats optimaux pour chaque type de modélisation

À titre de comparaison entre les différents types de modélisation, le tableau 4.3 résume les résultats obtenus selon les paramètres trouvés optimaux lors de l'analyse des sections précédentes. Nous constatons que le système de classification de type [MFCC/GMM] élaboré avec les paramètres listés au tableau 4.3, donne le meilleur résultat soit une valeur *AUC* de 0.935 (cette valeur est tirée de la figure 4.14).

4.7.1 Effets du post-traitement

Le principe de post traitement, discuté à la section 3.10, consiste à filtrer la fonction distance $z(i)$ de chacune des deux classes avant de tracer la courbe *ROC* duquel on calcul l'aire sous la courbe (AUC). La figure 4.20-i montre la fonction distance $z(i)$ avant et après traitement (filtrage). Les distributions correspondantes sont reportées respectivement aux

Tableau 4.3: Paramètres optimaux pour les différents types de modélisation

	<i>VQ</i>	<i>GMM</i>	<i>MLP</i>
Caractéristiques	<i>MFCC</i>	<i>MFCC</i>	<i>MFCC</i>
Prétraitement	normalisation	normalisation	normalisation
Segment	1024	1024	1024
Recouvrement	50%	50%	50%
Dimension	24	24	24
Regroupement	32	8	64
Nombre d'itérations	50	10	20
Type de matrice de covariance	-	diagonale	-
Type d'initialisations	-	<i>LBG</i>	-
Aire sous la courbe (<i>AUC</i>) obtenue	0.903	0.935	0.878

Tableau 4.4: Paramètres optimaux

Vecteur des caractéristiques	<i>MFCC</i>
Largeur d'un segment	1024 échantillons
Recouvrement	50%
Dimension	16 caractéristiques
Modélisation	<i>GMM</i>
Type d'initialisation	<i>LBG</i>
Itération lors de l'initialisation	15
Regroupement	8 gaussiennes
Type de matrice de covariance	diagonale
Itération de l'algorithme <i>EM</i>	10
Prétraitement	normalisation
Post traitement	oui

figures 4.20-ii et 4.20-iii. Le taux de recouvrement entre les deux distributions est réduit (voir la figure 4.20-iii) et offre ainsi une meilleure discrimination entre les deux classes. Cette amélioration peut être constatée par l'augmentation de l'aire sous les courbes (*AUC*) de la figure 4.20-iv.

Nous reprenons ici les résultats affichés à la figure 4.16 et nous y effectuons le post-traitement. L'effet du post-traitement est presque exclusivement positif (voir la figure 4.21). À titre d'exemple, le résultat de la combinaison dont les paramètres sont : cov=diag, D=8, N=16 et init=LBG est passé d'une valeur *AUC* de 0.911 à une valeur de presque 0.950. Par contre, le résultat de notre meilleur candidat avant post-traitement dont les paramètres sont : cov=diag, D=8, N=24 et init=LBG, n'est pas de beaucoup amélioré. Le tableau 4.4 liste les paramètres donnant le meilleur résultat soit une valeur AUC de 0.950.

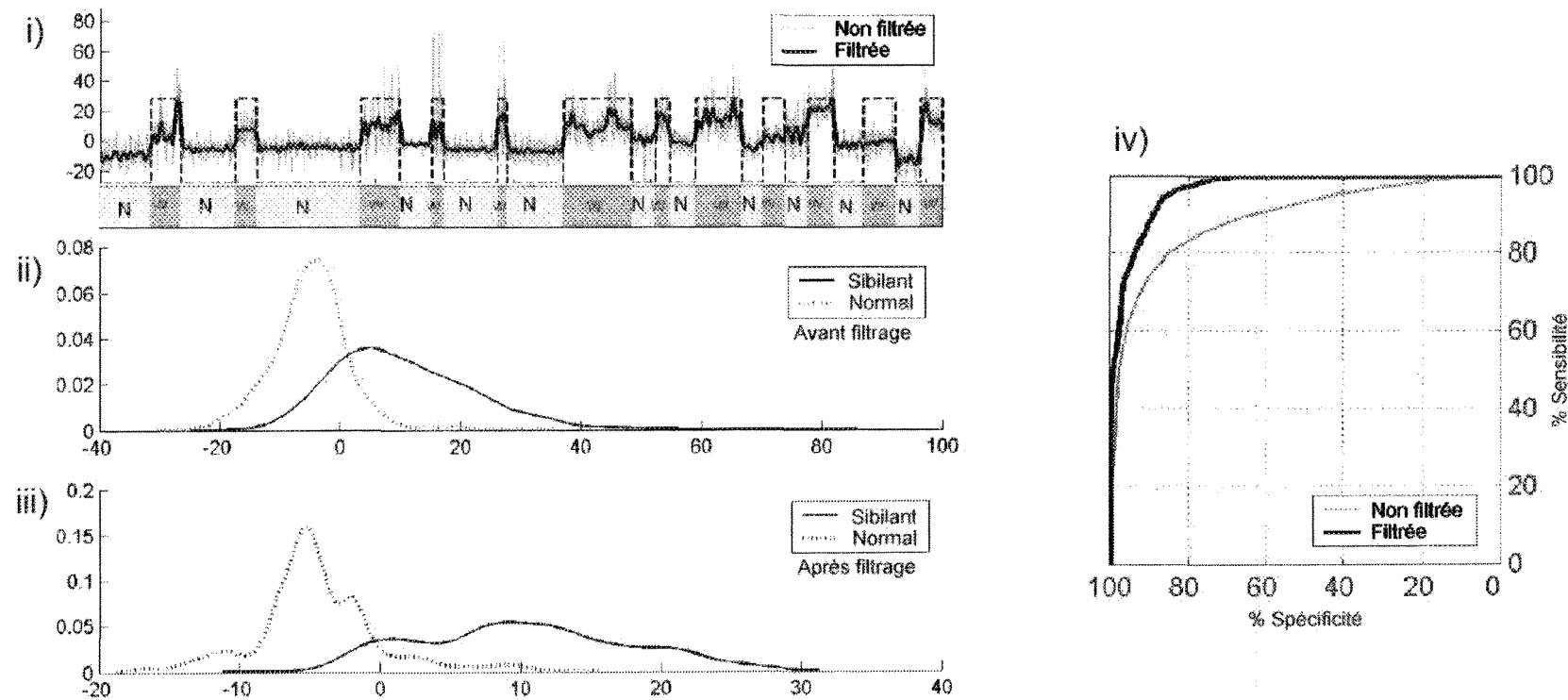


Figure 4.20: Amélioration des performances de la classification par filtrage (post-traitement) de la fonction distance *i*) fonction distance $z(i)$ pour des fichiers alternés de sons normaux (*N*) et des sons avec sibilants (*W*), ii) distribution des distances $z(i)$ avant filtrage, iii) distribution des distances $z(i)$ après filtrage, et iv) courbes ROC correspondantes (pour le système de classification [MFCC/GMM]) [30].

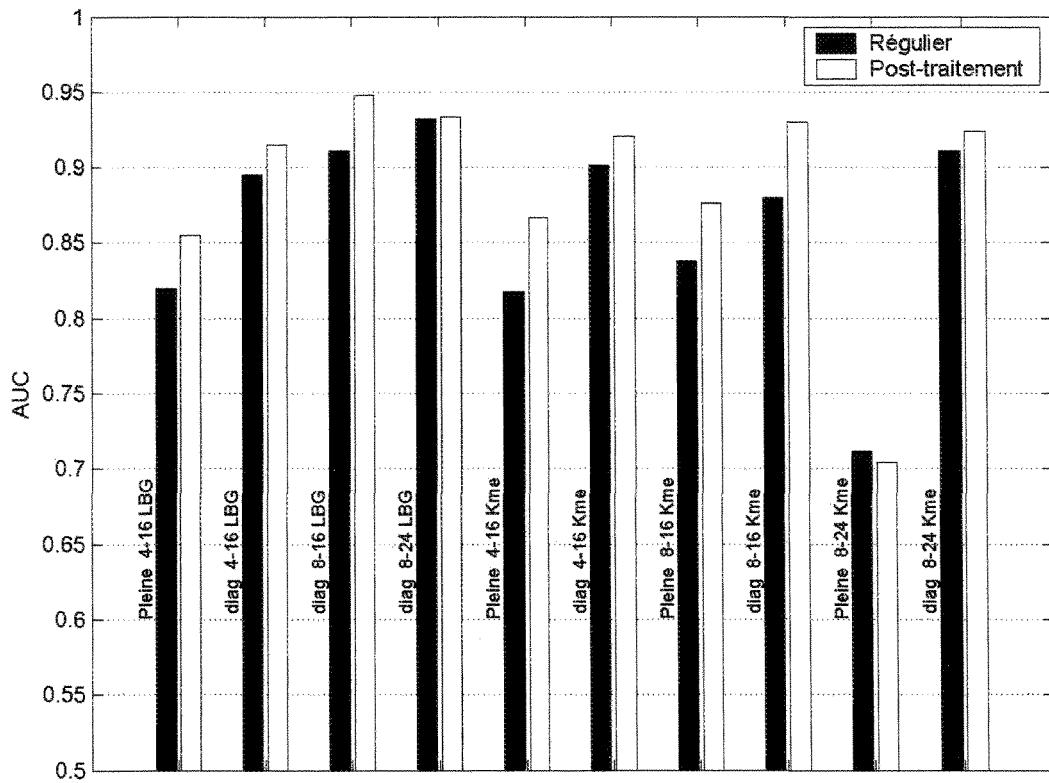


Figure 4.21: *Effets du post-traitement sur les performances (AUC) d'un système de classification [MFCC/GMM]. Il est à noter que les indices sur les barres verticales indiquent : [matrice de covariance, regroupement(N), dimension(D), initialisation]*

CONCLUSION

Ce travail consiste à étudier et comparer différents systèmes de classification afin de sélectionner celui qui est le plus apte à détecter les sons respiratoires de type sibilants produits lors d'un épisode de crise d'asthme. Ce travail s'inscrit dans un projet plus vaste, ayant cours au Département de Mathématiques, d'Informatique et de Génie (DMIG) de l'Université du Québec à Rimouski (UQAR), consistant à développer un outil d'identification de crise d'asthme à domicile.

Les systèmes de classifications que nous avons utilisés fonctionne en deux phases : une phase d'entraînement et une phase de test. L'entraînement et le test s'opèrent sur des caractéristiques extraites d'un ensemble de segments temporels des signaux acoustiques respiratoires. Nous nous intéressons à deux classes de signaux : les sons respiratoires normaux et ceux de type sibilants. Lors de la phase d'entraînement, les caractéristiques extraites des deux classes sont indépendamment modélisées, afin de tirer un discriminant entre les deux classes de son respiratoire. Lors de la phase de test, la *distance* entre la position des caractéristiques extraites d'un segment sonore et le discriminant est calculée, afin d'établir la classe d'appartenance du segment testé.

L'expérimentation consiste à comparer différentes techniques d'extraction de caractéristiques combinées à différents types de modélisation, afin de déterminer la combinaison la plus apte à détecter les sibilants. Nous utilisons les techniques d'extraction de caractéristiques de types : Coefficients cepstraux de Mel (*MFCC - Mel Frequency Cepstrum Coefficients*), Coefficients tirés des sous-bandes (*SBC - Subband Based Coefficients*), Transformée de Fourier rapide (*FFT - Fast Fourier Transform*), Codage linéaire prédictif (*LPC - Linear Predictive Coding*), Coefficients cepstraux (*CC - Cepstral Coefficients*), Meilleurs coefficients d'ondelettes (*WBC - Wavelet Best Coefficient*), Coefficients de la transformée en ondelettes discrète (*DWTC - Discrete Wavelet Transform Coefficients*) et Coefficients de paquets d'ondelettes (*WPC - Wavelet Packet Coefficient*) combinées aux types de modélisation : Quantification vectorielle (*VQ - Vector Quantization*), Mélange de distributions gaussiennes (*GMM - Gaussian Mixture Model*), Mélange de distributions de Student (*SMM - Student Mixture Model*) et le réseau de neurones de type perceptron multi-couches (*MLP - Multi Layer Perceptron*). De plus, ces combinaisons s'élaborent sur

plusieurs paramètres sous-jacents. Pour les techniques d'extraction, ces paramètres sont : la dimension des caractéristiques (D), le largeur des segments et le taux de recouvrement des segments adjacents. Pour les types de modélisation, les paramètres sont : le type d'initialisation, le nombre d'itération lors de l'initialisation, le nombre de regroupements, le type de matrice de covariance et le nombre d'itération permettant de construire le modèle. Nous analysons, en plus l'effet des techniques de prétraitement suivantes : normalisation, centrage, analyse en composantes principales (*PCA - Principal Component Analysis*) et le blanchiment ainsi qu'une technique de post-traitement basée sur le lissage des *distances*.

L'expérimentation a été effectuée en utilisant vingt-quatre enregistrements de sons respiratoires obtenus à l'aide de microphone placé à la trachée. Douze enregistrements sont des enregistrements de respiration normal alors que les douze autres contiennent des sibilants. L'utilisation d'une approche de test de type "*leave one out*" combinée à la mesure de l'aire sous la courbe (*AUC - Area Under Curve*) des courbes de caractéristiques d'efficacité (*ROC - Receiver Operating Characteristic*) obtenues permet de comparer les différentes combinaisons entre elles.

Il est difficile d'évaluer toutes les combinaisons possibles de techniques d'extraction de caractéristiques, de modélisation et des différents paramètres sous-jacents. Notre méthodologie consiste à abandonner, le long de ce projet, les approches qui semblent les moins prometteuses.

À la suite de cette expérimentation, la technique de reconnaissance de formes associant la modélisation par *GMM* à l'extraction de caractéristiques par coefficients cepstraux de type *MFCC* a fourni les meilleurs résultats, en donnant une aire sous la courbe (*AUC*) de 0.950. Le tableau 4.4 liste les paramètres utilisés permettant d'obtenir ce résultat.

Pour la suite, les recherches effectuées par Messieurs Mohammed Bahoura et Xiaoguang Lu se concentrent sur la séparation et la détection des sons adventices discontinus (les crépitants) afin de pouvoir incorporer cette classe de signaux à l'outil d'identification des pathologies respiratoires.

RÉFÉRENCES

- [1] Santé Canada. Les maladies respiratoires au canada. <http://www.hc-sc.gc.ca/pphb-dgspsp/publicat/rdc-mrc01/pdf/mrc0901f.pdf>, Septembre 2001.
- [2] Organisation Mondiale de la Santé (OMS). Les maladies respiratoires chroniques. http://www.who.int/mip/2003/other_documents/en/, 2003.
- [3] E. Lens, G. Postiaux, and P. Chappelle. Application in bedside medicine of automated spectral analysis of breath sound, wheezing and crackles. In *Proceedings of the 12th International Lungs Sounds Conference*, CNRS, Paris, 15-19 February 1987.
- [4] M. Nissan and N. Gavriely. A microcomputer based lung sounds analysis. *Comput. Methods Programs Biomed.*, 40(1) :7–13, 1993.
- [5] X. Sun, B.M. Cheetham, and J.E Earis. Real time analysis of lung sounds. *Technology and Health Care*, 6(1) :3–10, 1998.
- [6] A.R.A. Sovijärvi, P. Helistö, L.P. Malmberg, K. Kallio, E. Paajanen, A. Saarinen, P. Lippinen, S. Haltsonenb, L. Pekkanen, P. Piirilä, L. Näveri, and T. Katila. A new versatil pc-based lung sound analyzer with automatic crackle analysis (HeLSA). *Technology and Health Care*, 6(1) :11–22, 1998.
- [7] M. Waris, P. Helisto, S. Haltsonen, A. Saarinen, and A.R. Sovijarvi. A new method for automatic wheeze detection. *Technol Health Care*, 6(1) :33–40, 1998.
- [8] L. Pesu, P. Helisto, E. Adomovic, J.C. Pesquet, A. Saarinen, and A.R.A. Sovijarvi. Classification of respiratory sounds based on wavelet packet decomposition and learning vector quantization. *Technology and Health Care*, 6(1) :65–74, 1998.
- [9] S. Rietveld, M. Oud, and E.H. Dooijes. Classification of asthmatic breath sounds : preliminary results of the classifying capacity of human examiners versus artificial neural networks. *Comput. Biomed. Res.*, 32 :440–448, 1999.
- [10] L.R. Waitman, K.P. Clarkson, J.A. Barwise, and P.H. King. Representation and classification of breath sounds recorded in an intensive care setting using neural networks. *Journal of Clinical Monitoring and Computing*, 16 :95–105, 2000.

- [11] Z. Dokur and T. Olmez. Classification of respiratory sounds by using an artificial neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(4) :567–580, 2003.
- [12] A. Kandaswamy, C. Sathish Kumar, Rm. Pl. Ramanathan, S. Jayaraman, and N. Malmurugan. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34 :523–537, 2004.
- [13] E.C. Guler, B. Sankur, Y.P. Kahya, and S. Raudys. Two-stage classification of respiratory sound patterns. *Computers in Biology and Medicine*, 35(1) :67–83, 2005.
- [14] M. Bahoura. *Analyse des Signaux Acoustiques Respiratoires : Contribution a la Détection Automatique des Sibilants par Paquets d'Ondelettes*. PhD thesis, Université de Rouen, Janvier 1999.
- [15] American Thoracic Society Ad Hoc Committee on Pulmonary Nomenclature. Updated nomenclature for membership reaction. *ATS News Fall*, 3 :5–6, 1977.
- [16] R. Mikami, M. Murao, D.W. Cugell, J. Chretien, P. Cole, J. Meier-Sydw and R.L. Murphy, and R.G. Loudon. International symposium on lung sounds. synopsis of proceedings. *Chest*, 92(2) :342–5, 1987.
- [17] P. Marquet. Les bruits d'origine ventilatoire. *Rev. Mal. Resp.*, 10 :87–101, 1995.
- [18] R.G. Loudon and R.L. Murphy. State of the art : lung sounds. *Am Rev Respir Dis*, 130(4) :663–73, 1984.
- [19] P. Forgacs. *Lung Sounds*. Cassell and Coller MacMillan Publishers Ltd., London, 1978.
- [20] ARA. Sovijärvi, F. Dalmasso, J. Vanderschoot, LP. Malmberg, G. Righini, and SAT. Stoneman. Definition of terms for applications of respiratory sounds. *European Respiratory Review*, 10 :597–610, 2000.
- [21] G. Charbonneau, E. Ademovic, BMG. Cheetham, LP. Malmberg, J. Vanderschoot, and ARA. Sovijarvi. Basic techniques for respiratory sound analysis. *European Respiratory Review*, 10 :625–635, 2000.
- [22] S.S. Kraman, H. Pasterkamp, M. Kompis, M. Takase, and G.R. Wodicka. Effects of breathing pathways on tracheal sound spectral features. *Respiration Physiology*, 111 :295–300, 1998.

- [23] Y. Takezawa, F. Shirai, and S. Sawaki et al. Comparaison of wheezes over the trachea and on the chest wall. In *the Fifth International Lung Sounds Conference*, London, England. International Lung Sound Association.
- [24] CD audio des Laboratoires Astra France. L'auscultation pulmonaire. <http://www.ifrance.com/y-detauriac/cdrom/pneumo.htm> ?
- [25] The R.A.L.E Lung Sound 3.0. <http://www.rale.ca>, 2003.
- [26] K. Sjolander and J. Beskow. Wavesurfer 1.8.1. <http://www.speech.kth.se/wavesurfer/index.html>, Department of Speech, Music and Hearing. The Royal Institute of Technology in Stockholm.
- [27] K.E Forkheim, D. Scuse, and H. Pasterkamp. A comparison of neural network models for wheeze detection. *IEEE WESCANEX 95 proceedings*, 1 :214–219, 1995.
- [28] B. Sankur, Y.P. Kahya, E.C. Guler, and T. Engin. Comparison of AR-based algorithms for respiratory sound classification. *Comput. Biol. Med.*, 24 :67–76, 1994.
- [29] M. Bahoura and C. Pelletier. Respiratory sounds classification using gaussian mixture models. In *Electrical and Computer Engineering, 2004. IEEE CCECE 2004. Canadian Conference on*, Niagara Falls, Canada, May 2-5, 2004.
- [30] M. Bahoura and C. Pelletier. Respiratory sounds classification using cepstral analysis and gaussian mixture models. In *26th Annual Conference of the IEEE EMBS*, pages 9–12, San Francisco, California, September 1-5, 2004.
- [31] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition : A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), January 2000.
- [32] M. Bahoura and C. Pelletier. New parameters for respiratory sound classification. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, volume 3, pages 1457–1460, Montreal, Canada, May 4-7, 2003.
- [33] S. Stevens and J. Volkman. The relation of pitch to frequency : A revised scale. *American Journal of Psycholinguistics*, 53 :329–353, 1940.
- [34] B.J. Shannon and K.K. Paliwal. A comparative study of filter bank spacing for speech recognition. In *Microelectronic Engineering Research Conference 2003*, pages 1–3, 1999.

- [35] T. Phan and T. Soong. Text-independent speaker identification. Technical report, The University of Texas at Austin, Electrical and Computer Engineering, 1999.
- [36] D. Donoho, M.R. Duncan, X. Huo, O. Levi, J. Buckheit, M. Clerc, J. Kalifa, S. Mallat, and T. Yu. Wavelab 802 for matlab5.x. <http://www-stat.stanford.edu/~wavelab/>, Stanford Statistics Department, Stanford University.
- [37] S. Mallat. Multiresolution signal decomposition : The wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(7) :674–693, July 1989.
- [38] R.R. Coifman and M.V. Wickerhauser. Best-adapted wave packet bases. Technical report, Yale University, 1990.
- [39] M. V. Wickerhauser. INRIA lectures on wavelet packet algorithms. In Pierre-Louis Lions, editor, *Problèmes Non-Linéaires Appliqués, Ondelettes et Paquets D’Ondes*, pages 31–99. INRIA, Roquencourt, France, 17–21 June 1991. Minicourse lecture notes.
- [40] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithm for best basis selection. *IEEE Trans. Inform. Theory*, 38 :713–718, 1992.
- [41] L. Pesu, P. Helisto, E. Ademovic, and J.C. Pesquet. Wavelet packet based respiratory sound classification. In *Proceedings of IEEE Symposium on Time-Frequency and Time-Scale Analysis*, pages 377–380, Paris, France., June 18-21 1996.
- [42] R. Sarikaya, B. Pellom, and J.H.L. Hansen. Wavelet packet transform features with application to speaker identification. In *NORSIG-98 IEEE Norsic Signal Processing Symposium*, pages 81–84, Vigso, Denmark, June 1998.
- [43] R.E. Learned and A.S. Willsky. A wavelet packet approach to transient signal classification. *Applied and computational harmonic analysis*, 2 :265–278, 1995.
- [44] H. Demuth and M. Beale. *Neural Network Toolbox For Use with MATLAB*. 3, Apple Hill Drive Natick, MA 01760-2098, 4 edition, 2004.
- [45] Y. Linde, A. Buzo, and R.M Gray. An algorithm for vector quantizer design. *IEEE Trans. Comm.*, 28(1) :84–95, January 1980.
- [46] M.N. Do. Digital signal processing mini-project "an automatic speaker recognition system". Technical report, Audio Visual Communications Laborator Swiss Federal Institute of Technology, Lausanne, Switzerland, 2000.

- [47] D. Reynolds and R. Rose. Robust test-independent speaker identification using gaussian mixture speaker models. *IEEE Trans. Speech Audio Processing*, 3 :72–83, 1995.
- [48] D. Reynolds. Speaker identification and verification using gaussian mixture speaker models. *Speech Communication*, 17 :91–108, 1995.
- [49] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-InterScience, second edition, 2001.
- [50] S. Shoham. Robust clustering by deterministic agglomeration EM of mixtures of multivariate t-distributions. *Pattern Recognition*, 35 :1127–1142, 2002.
- [51] B. Scherrer. *Biostatistique*. Gaëtan Morin Éditeur, Chicoutimi, Canada, 1984.
- [52] T. Fawcett. ROC graphs : Notes and practical considerations for researchers. HP Laboratories, March 2004.

Index

- Aire sous la courbe, voir AUC, 75
- Algorithme
- Back propagation, 68
 - EM, 59, 64
 - LBG, 54
- Analyse Cepstrale, 20
- Analyse de Fourier, 15
- Analyse en composantes principales, 42
- Analyse multirésolution, 29
- Apprentissage supervisé, 46
- Asthme, 1
- AUC, 75, 79, 88
- Base de données, 6
- Blanchiment, 44, 82
- Calcul de spectrogrammes, 17
- Caractéristiques, 19, 22, 24, 27, 31, 35, 37
- Cellule de Voronoi, 53
- Centrage, 42, 82
- Centroïde, 50
- Classification des signaux, 46
- Coefficients Cepstraux, 20
- Coefficients MFCC, 23
- Convolution, 20, 70
- Couche cachée, 66
- Covariance, 55, 56, 61, 65, 87
- Critère de Nyquist, 19
- Décimation, 34
- Ddp, 55, 58
- Densité de probabilité, voir ddp, 55
- Dictionnaire, 48, 50
- GMM, 56
- VQ, 51
- Discriminant
- GMM, 58, 62
 - MLP, 68
 - VQ, 53
- Distance de Mahalanobis, 56
- Distance euclidienne, 51
- Dyadique, 34
- Echelle de Mel, 23
- EM, 59, 64
- Entropie, 34
- Erreur quadratique moyenne, 51, 54
- Estimation Maximisation, 59, 64
- Extraction des caractéristiques, 13
- Fenêtre de Hamming, 70
- Fenêtre de pondération, 17
- Filtres à miroirs quadratiques, 29
- Fonction respiratoire, 1
- Formule de Sturge, 76
- Fréquence d'échantillonnage, 6, 19

- GMM, 55, 78
- Initialisation, 87
 - GMM, 60
 - SMM, 65
- Itération, 85
- k-means, 61
- Largeur de la fenêtre, 84
- LBG, 54
- Levinson-Durbin, 27
- LPC, 26
- Maladie d'asthme, 1
- Maladies pulmonaires chroniques, 3
- Matrice de covariance diagonale, 56, 87
- Matrice de covariance pleine, 87
- Maximum de vraisemblance, 57
- Meilleure base de décomposition, 34, 36
- Mel, 23
- Microphone, 6
- MLE, 57
- MLP, 66, 78
- Modélisation, 48
- Modélisation linéaire prédictive, 26
- Modélisation par Mélange de
 - Distributions de Student, 63
 - gaussiennes, 55
- Mot de code, 50
- Moyenne, 55, 61, 65
- MSE, 51
- Multirésolution, 29
- Normalisation, 41, 82
- Ondelettes, 28
- Outliers, 63
- Paquets d'ondelettes
- Recherche de la meilleure base, 34
- Transformée (WPT), 32
- PCA, 42, 82
- Perceptron Multi-couche, 66
- Phase d'apprentissage
 - GMM, 57
 - MLP, 66
 - VQ, 51
- Phase de classification/test
 - GMM, 58
 - MLP, 68
 - VQ, 52
- Post traitement, 70
- Prétraitement, 82
- QMF, 29, 34
- quantification vectorielle, 49
- Réseaux de neurones, 66
- Résolution fréquentielle, 18, 84
- Résolution temporelle, 18, 84
- Règle de Sturge, 76
- Recherche de la meilleure base, 34
- Reconnaissance de formes, 6
- Recouvrement, 13, 19, 34, 83
- Regroupement, 50, 87
- ROC, 75–77
- SBC, 37
- Sensibilité, 73

Sibilants
 polyphoniques, 4
 monophoniques, 4
SMM, 63, 78
Sons respiratoires adventices
 Ronchus, 4
 Sibilants, 4
Sons respiratoires normaux
 Murmures vésiculaires, 4
 Trachéo-bronchiques, 4
Spécificité, 73
STFT, 15
Systèmes de surveillance, 1

Terminologie originale de Laënnec, 3
Trachée, 6
Transformée de Fourier à court terme, 15
Transformée en ondelettes
 continue (CWT), 28
 discrete (DWT), 29

Voronoi, 53
VQ, 49, 78
Vraisemblance, 60