

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOULOUD MAMMERI - TIZI-OUZOU

Faculté de Génie Electrique et d'Informatique

Département d'Electronique

THÈSE

pour obtenir le titre de

Doctorat

Option : Electronique

Présentée et soutenue par

Zahia Zidelmal épouse Amirou

Reconnaissance d'arythmies cardiaques par Support Vector Machines (SVMs)

Thèse dirigée par Adel Belouchrani

soutenue le .. /...../2012 devant le Jury :

<i>Président :</i>	Soltane Ameer	- Pr, UMM (Tizi-Ouzou)
<i>Directeur de Thèse :</i>	Adel Belouchrani	- Pr, ENP (Alger)
<i>Examineurs :</i>	Azeddine Baghdadi	- Pr, Univ-Paris-13 (France)
	Latifa Hamami	- Pr, ENP (Alger)
	Hacen Belbachir	- MCA, USTHB (Alger)
	Salah Haddab	- MCA, UMM (Tizi-Ouzou)
<i>Invité :</i>	Brahim Kichou	- Cardiologue, CHU (Tizi-Ouzou)

À mes parents

Remerciements

Avant tout, je remercie Dieu de m'avoir donné la force d'aller jusqu'au bout

Je tiens à remercier le Professeur Soltane Ameur d'avoir accepté de présider ce jury de thèse, ainsi que le professeur Latifa Hamami, le professeur Azeddine Baghdadi, le Docteur Hacem Belbachir, le Docteur Salah Haddab et le Docteur Brahim Kichou d'avoir accepté de faire partie du jury.

Je tiens à exprimer ma profonde gratitude au professeur Adel Belouchrani de m'avoir fait l'honneur de diriger mes travaux durant ces années de doctorat en étant toujours disponible et encourageant. Ses conseils, sa rigueur scientifique et ses valeurs humaines m'ont bien guidée.

Le Docteur Mourad Adnane m'a beaucoup apporté grâce à son expérience de recherche dans le domaine du génie biomédical. Qu'il trouve ici l'expression de mes vifs remerciements.

Merci à Monsieur Ahmed Amirou, enseignant en recherche opérationnelle pour sa précieuse aide concernant les problèmes d'optimisation et d'apprentissage.

Je remercie également le Docteur Rachid Achaibou, spécialiste en cardiologie pour le temps qu'il m'a consacré et tous les entretiens qu'il m'a accordés lors de la préparation de la présente thèse.

Je n'oublierai pas de remercier Boualem et Nouredine Sedkaoui pour leur précieuse aide.

Le Professeur Stéphane Canu, Directeur du laboratoire LITIS (INSA, Rouen) et le Professeur Yves Grandvalet m'ont reçue plusieurs fois dans leur unité de recherche. Je les remercie vivement pour leurs directives concernant les méthodes d'apprentissage statistique et décision.

Ma gratitude s'adresse aussi au Docteur Djaffar Ould-Abdesslam, au professeur Jean Markle, au professeur Jean-Philippe Hurbin, au Docteur Ali Mokadem ainsi qu'aux autres membres du laboratoire MIPS qui ont marqué un intérêt particulier à mon travail et m'ont permis de l'ouvrir sur des perspectives très intéressantes.

En fin un grand merci à Ahmed et mes enfants qui ont été très patients avec moi.

Liste des Abréviations

SA	Sino-Atrial
AV	AtrioVentriculaire
PRA	Période Refractaire Absolue
PRR	Période Refractaire Relative
PRE	Période Refractaide Effective
ECG	ElectroCardioGramme
BSA	Bloc Sino-Auriculaire
BAV	Bloc Auriculo-Ventriculaire
RBBB	Right Bundle Branch Block
LBBB	Left Bundle Branch Block
PVC	Prematur Ventricular Contraction
ESV	ExtraSystole Ventriculaire
AHA	American Hearth Association
AAMI	American Association of Medical Instrumentation
RII	Réponse Impulsionnelle Infinie
AMR	Analyse MultiRésolution
DSP	Densité Spectrale de Puissance
FFT	Fast Fourier Transform
FN	Faux Négatif
FP	Faux positif
TP	True Positif (Vrai Positif)
KLT	Karhunen Loève Transform
AR	AutoRégressive
MSE	Mean Square Error
SVM	Support Vector Machines
VC	Vapnik-Chervonenkis
SMO	Sequentiel Minimal Optimization
SRM	Minimisation du Risque Structurel

Table des figures

2.1	<i>Les phases du potentiel d'action d'une cellule du muscle cardiaque.</i>	7
2.2	<i>Coupe longitudinale du coeur présentant : 1 : Nœud sinusal; 2 : Oreillettes; 3 : Nœud AV; 4 : Faisceau de His; 5 : Branches; 6 : Réseau de Purkinje.</i>	9
2.3	<i>Allure d'un électrocardiogramme normal (Segment de l'enregistrement MITDB : 100).</i>	10
2.4	<i>Dérivations bipolaires d'Einthoven.</i>	12
2.5	<i>Dérivations unipolaires des membres.</i>	12
2.6	<i>Dérivations précordiales de Wilson.</i>	13
2.7	<i>Signal électrocardiographique perturbé par le secteur. (Segment de l'enregistrement AHA : 0201).</i>	15
2.8	<i>Bruit dû au mauvais contact électrode-peau (MITDB : 101).</i>	15
2.9	<i>Ondulations de la ligne de base (MITDB : 119).</i>	16
2.10	<i>Exemple de bruits musculaires sur un segment de l'enregistrement MITDB : 106.</i>	17
2.11	<i>Rythme sinusal, segment de l'enregistrement AHA : 0201.</i>	18
2.12	<i>Exemple de Bloc de Branche Gauche (MITDB : 111).</i>	19
2.13	<i>Exemple de Bloc de Branche Droite (MITDB : 124).</i>	20
2.14	<i>Exemple d'extrasystole ventriculaire (MITDB : 119).</i>	21
2.15	<i>Exemple de doublet ventriculaire (MITDB : 208).</i>	21
2.16	<i>Exemple de bigéminisme (MITDB : 200).</i>	22
2.17	<i>Exemple de trigéminisme (MITDB : 119).</i>	22
3.1	<i>Arbre d'analyse multirésolution.</i>	31
3.2	<i>Décomposition dyadique d'un signal en bandes de fréquences.</i>	33
3.3	<i>a : Signal ECG ; b : analyse en ondelettes non redondantes du signal ECG (les différents niveaux de résolution ont été mis bout à bout, ($d_n^1 : [2049, 4096]$; $d_n^2 : [1025, 2048]$; $d_n^3 : [513, 1024]$; $a_n^3 : [1, 512]$). Du fait des changements d'échelle, les variations des petits coefficients ne sont pas visibles ; c : Mise en évidence du bruit dans les derniers coefficients de détail de haute résolution.</i>	36
3.4	<i>Comparaison des deux types de seuillage dur et doux : en abscisse, on présente les valeurs possibles des coefficients d'ondelettes (d_n^j) et en ordonnées les valeurs correspondantes qu'ils prennent à l'issue du seuillage(d_{ns}^j)).</i>	38
3.5	<i>Résultat de l'algorithme de filtrage des hautes fréquences appliqué à l'enregistrement MITDB : 117.</i>	39

3.6	<i>Résultat de l'algorithme de filtrage des hautes fréquences appliqué à l'enregistrement MITDB : 104 entaché d'un bruit musculaire non stationnaire.</i>	39
3.7	<i>Ondulation de la ligne de base.</i>	40
3.8	<i>a : Signal ECG ; b : analyse en ondelettes non redondantes du signal ECG (les différents niveaux de résolution ont été mis bout à bout, (d_n^1 : [2049, 4096] ; d_n^2 : [1025, 2048] ; d_n^3 : [513, 1024] ; d_n^4 : [257, 512] ; d_n^5 : [129, 256] ; d_n^6 : [65, 128] ; a_n^6 : [1, 64]). Du fait des changements d'échelle, les variations des petits coefficients ne sont pas visibles ; c : Mise en évidence de l'ondulation de la ligne de base sur les coefficients d'échelle de plus basse résolution.</i>	41
3.9	<i>Suppression de la ligne de base sur un segment de l'enregistrement MITDB : 203.</i>	42
3.10	<i>Suppression de la ligne de base sur un segment de l'enregistrement MITDB : 210. Haut : signal original ; Bas : signal filtré.</i>	42
3.11	<i>Signal original et signal reconstruit avec filtrage et suppression de la dérive de la ligne de base. A gauche, le segment contient un battement systolique ; A droite, le segment contient un battement ectopique.</i>	43
3.12	<i>A gauche : Segment MITDB : 100 [0s, 4s]. A droite : Segment MITDB : 117 [0s, 4s] où l'onde T est aussi importante en amplitude que l'onde R. En revanche, quelque soit le patient, les ondes R varient plus rapidement que les autres.</i>	44
3.13	<i>Segment MITDB : 203 [90s, 110s] présentant une variété d'ondes R avec diverses amplitudes et morphologies.</i>	44
3.14	<i>Réponse en fréquence du filtre P-Bas proposé par Pan et Tompkins.</i>	46
3.15	<i>Réponse en fréquence du filtre dérivateur représentée à l'échelle logarithmique.</i>	47
3.16	<i>ECG filtré (a) ; sortie du filtre dérivateur (b) et sortie du filtre intégrateur à fenêtre glissante (c).</i>	47
3.17	<i>Détection des pics R par l'algorithme de Pan et Tompkins sur un segment de l'enregistrement MITBD : 100.</i>	49
3.18	<i>Segment de l'enregistrement MITDB : 208 [0s, 8s). Avec un seuil de 30 %, la plupart des battements PVC ne sont pas détectés par la méthode ; leurs dérivées sont nettement inférieures à celles des battements normaux.</i>	49
3.19	<i>Densité Spectrale de Puissance des quatre types de complexes QRS.</i>	52
3.20	<i>Densité spectrale de puissance des battements normaux (figure de gauche) et battements PVC (figure de droite). Pour une meilleure comparaison, la même échelle est utilisée à chaque niveau. Les détails d_n^1 et d_n^2 ne sont pas considérés car ils sont en dehors de la bande d'intérêt.</i>	54

3.21	<i>Décomposition d'un segment du signal AHA : 0201 [1636 s, 1660 s] en utilisant l'ondelette de Haar. L'effet des ondes P et T appartenant à la bande de fréquence [0 Hz, 5 Hz] est nettement atténué. . .</i>	55
3.22	<i>Détection des complexes QRS par notre algorithme. Toutes les ondes R sont détectées en utilisant le seuil primaire.</i>	57
3.23	<i>Détection des complexes QRS par l'algorithme Tompkins (seuil primaire).</i>	57
3.24	<i>Exemples de battements segmentés sur 72 échantillons.</i>	58
4.1	<i>Différentes formes de battements cardiaques considérés.</i>	60
4.2	<i>Intervalles temporels entre des pics R successifs dans le cas d'un rythme normal (segment de l'enregistrement MITDB : 100).</i>	61
4.3	<i>Intervalles temporels entre des pics R successifs dans le cas de présence de battements ectopics (segment de l'enregistrement MITDB : 208.</i>	62
4.4	<i>Intervalles RR réguliers sur l'enregistrement MITDB : 100 normal (Gauche); Histogramme des intervalles RR (Droite).</i>	63
4.5	<i>Histogrammes RR pour les quatre types de battements considérés . .</i>	63
4.6	<i>Varianace de l'erreur de prédiction en fonction de l'ordre p du modèle.</i>	65
4.7	<i>Résultat de la prédiction appliquée à un battement normal (gauche) et un battement ectopic(droite)</i>	66
4.8	<i>Bloc de branche MITDB : 109 [11 s, 16.7 s]. Il y a ici un problème de conduction de l'impulsion électrique dans le faisceau de His. La contraction des deux ventricules n'est donc pas parfaitement simultanée. La durée totale de l'onde R est ici plus longue que lors des battements normaux.</i>	67
4.9	<i>Détection des segments QS sur l'enregistrement MITDB : 100. . . .</i>	67
4.10	<i>Détection des segments QS sur l'enregistrement MITDB : 119 (haut); zoom sur le segment [1 s, 3 s] contenant un battement normal et un PVC (bas).</i>	68
4.11	<i>Niveaux de puissance en fonction de la fréquence, correspondant aux quatre types de battements considérés (N, V, RBBB et LBBB). . . .</i>	69
4.12	<i>Relation entre les niveaux de puissance normalisée (0.25, 0.50, 0.75, 0.95), occupés par les quatre types de battements (N, V, RBBB, et LBBB) et le contenu fréquentiel.</i>	70
5.1	<i>Illustration de l'inégalité (5.4). La courbe croissante, appelée confiance, correspond à la borne supérieure du terme de complexité. Les comportements du terme de complexité et de l'erreur empirique sont clairement opposés. On recherche donc le meilleur compromis entre complexité et erreur empirique.</i>	76
5.2	<i>Principe d'hyperplan séparateur, il en existe plusieurs. Celui qui correspond au minimum d'erreurs est l'hyperplan optimal.</i>	77

5.3	<i>L'hyperplan optimal H (en gras) avec la marge maximale d.</i>	77
5.4	<i>Données linéairement séparables (a); données non linéairement séparables (b)</i>	78
5.5	<i>Exemple d'espace d'entrée $\mathcal{X}$, (a) et d'espace caractéristique $\mathcal{F}$, (b)</i>	79
5.6	<i>Illustration de la marge et des Vecteurs Support.</i>	82
5.7	<i>Représentation du compromis entre la largeur de la marge souple et le coût d'une erreur.</i>	85
5.8	<i>Approximations de la fonction de perte 0, 1 (vert), par les fonctions coude ou hinge loss (bleu) et la fonction logistique (rouge). L'axe des abscisses correspond à la quantité $yf(x)$ qui est négative si l'exemple x est mal classé par f.</i>	87
6.1	<i>Coût de chaque classification en fonction de la probabilité a posteriori P. Les courbes a, b et c représentent respectivement : $E_y = c_+ P$, $E_y = c_- (1 - P)$ et $E_y = r_+ P + r_- (1 - P)$.</i>	91
6.2	<i>Illustration de la règle de Chow.</i>	92
6.3	<i>Fonction de coût pour les exemples positifs (gauche) et pour les exemples négatifs (droite) avec $p_- = 0.35$ and $p_+ = 0.6$ (trait continu : double hinge, trait interrompu : logistique).</i>	94
6.4	<i>Correspondance entre les vecteurs d'apprentissage et les pertitions de la fonction de coûts (Double hinge loss).</i>	97
6.5	<i>Résultat obtenu en appliquant l'algorithme de classification symétrique avec rejet sur une base de données générée sous Matlab.</i>	100
6.6	<i>Résultat obtenu en appliquant l'algorithme de classification asymétrique avec rejet sur une base de données générée sous Matlab.</i>	100
7.1	<i>Segment de l'enregistrement MITDB : 107 [0 s, 16.6 s] avec les points rouges représentant les positions des pics R détectés (haut); produit h des coefficients détails avec son seuillage (bas).</i>	105
7.2	<i>Détetion des pics R sur un segment de l'enregistrement MITDB : 222 [1102 s, 1112 s].</i>	105
7.3	<i>détection des ondes R (points rouge sur un segment de l'enregistrement MITDB : 203 [598 s, 608 s] présentant des battements et des inter-battements atypiques.</i>	106
7.4	<i>Résultat de la détection des pics R sur le segment MITDB : 104 [310s, 320s].</i>	107
7.5	<i>Segment de l'enregistrement MITDB : 101 [174s, 180s] indiquant un cas de chute brusque de la ligne de base.</i>	107
7.6	<i>Résultats de la détection des pics R sur le segment AHA : 0001 [1628 s, 1660 s]. Nous remarquons que les trois battements PVC sont correctement détectés malgré qu'ils sont suivis chacun d'une onde T plus importante car l'effet de cette dernière est très amorti sur le signal de localisation h.</i>	108

7.7	<i>Résultats de la détection des pics R sur le segment AHA : 0201 [1828 s, 1860 s]. Nous remarquons que les deux battements PVC sont inversés mais correctement détectés.</i>	108
7.8	<i>Procédure de validation croisée avec $k = 3$.</i>	112
7.9	<i>Région de rejet induite par les seuils de rejet fixés avec les coûts des rejets ($r = 0.35$ et le rapport des coûts des fausses classifications $\beta = 1$). Les exemples représentés par (o rouge) représentent la classe négative. Les exemples représentés par (* bleue) représentent la classe positive.</i>	113
7.10	<i>région de rejet induite par les seuils de rejet fixés avec les coûts asymétriques ($r = 0.35$ et $\beta = 1.2$). Les exemples représentés par o appartiennent à la classe négative. Les exemples représentés par * appartiennent à la classe positive.</i>	114
7.11	<i>Courbe de compromis erreur-rejet (haut); Pente de la courbe de compromis (bas).</i>	115
7.12	<i>Coût global de la classification en fonction du taux de rejet (haut); Sensitivité et Prédicativité positive du classifieur en fonction du taux de rejet (bas).</i>	116

Liste des tableaux

3.1	<i>Correspondance entre détails et bandes de fréquences</i>	53
5.1	<i>Quelques noyaux usuels</i>	81
5.2	<i>Lien entre les contraintes primales et duales</i>	86
6.1	<i>Coût associé à chaque décision</i>	90
6.2	<i>Partition de l'ensemble d'apprentissage</i>	96
6.3	<i>Situation des contraintes pour les cinq types d'exemples</i>	98
7.1	<i>Comparaison avec d'autres algorithmes</i>	110
7.2	<i>Résultats de la détection par notre méthode, testée sur la base MITDB</i>	118

Table des matières

1	Introduction	1
1.1	Présentation	1
1.2	Problématique	2
1.3	Contributions personnelles	3
1.4	Organisation du manuel	3
2	Notions d'Electrocardiographie et reconnaissance d'arythmies	5
2.1	Introduction	5
2.2	le système cardiovasculaire	5
2.2.1	Activité mécanique cardiaque	6
2.2.2	Activité électrique du cœur	6
2.2.3	Automatisme cardiaque	8
2.3	Electrocardiographie de surface	9
2.3.1	L'électrocardiogramme	9
2.3.2	Techniques d'enregistrement	11
2.3.3	Les systèmes de dérivations	11
2.3.4	Conditions d'enregistrement	13
2.3.5	Le Holter	13
2.4	Artefacts visibles sur l'ECG	14
2.4.1	Bruits techniques	14
2.4.2	Bruits physiques	16
2.5	Le rythme cardiaque sinusal	17
2.6	Troubles du rythme et de la conduction cardiaque	17
2.6.1	Les blocs cardiaques	18
2.6.2	Les arythmies ventriculaires	20
2.7	Reconnaissance automatique d'arythmies : Etat de l'art	22
2.7.1	Paramètres descriptifs d'un battement cardiaque	23
2.7.2	Algorithmes d'apprentissage et de décision	24
2.7.3	Problèmes liés à la reconnaissance automatique d'arythmies	24
2.8	Conclusion	24
3	Segmentation de battements cardiaques	27
3.1	Introduction	27
3.2	Analyse multirésolution	27
3.2.1	La fonction Ondelette	27
3.2.2	La fonction d'échelle	28
3.2.3	Décomposition du signal (Analyse)	29
3.2.4	Algorithme d'analyse	31
3.2.5	Reconstruction du signal décomposé (Synthèse)	33

3.3	Filtrage du signal	34
3.3.1	Suppression du bruit de haute fréquence	34
3.3.2	Filtrage par seuillage des coefficients de détail	35
3.3.3	Correction des dérives de la ligne de base	38
3.4	Détection des complexes QRS	43
3.4.1	Méthodes existantes pour la détection des QRS	45
3.4.2	Méthode de Pan et Tompkins	45
3.4.3	Méthode proposée	50
3.5	Segmentation des battements	58
3.6	Conclusion	58
4	Extraction des paramètres discriminants	59
4.1	Introduction	59
4.2	Description des classes considérées	59
4.2.1	La classe V	59
4.2.2	La classe N	60
4.3	Calcul des paramètres discriminants	61
4.3.1	Mesure de régularité	61
4.3.2	Morphologie du battement	62
4.3.3	Largeur du battement	65
4.3.4	Paramètres fréquentiels	68
4.3.5	Les vecteurs de caractéristiques	70
4.4	Conclusion	71
5	Support Vector Machines(SVMs)	73
5.1	Introduction	73
5.2	Apprentissage statistique	74
5.3	Théorie de Vapnik-Chervonenkis	75
5.4	Objectifs d'un SVM	76
5.5	Linéarité et non linéarité	78
5.6	L'espace augmenté	78
5.6.1	Fonctions noyaux et similarité	79
5.6.2	Choix de la fonction noyau	80
5.7	Fondement mathématique des SVMs	81
5.7.1	Principe général	81
5.7.2	Cas linéairement séparable	81
5.7.3	Cas non linéairement séparable	84
5.8	Fonctions de coût d'un SVM	87
5.9	Algorithmes d'apprentissage des SVMs	88
5.10	Conclusion	88

6	Classification avec rejet d'ambigüités	89
6.1	Introduction	89
6.2	Règle de décision de Bayes	90
6.3	Formulation du problème	92
6.3.1	Fonction de coût du SVM avec rejet	92
6.3.2	Formulation du problème primal	94
6.3.3	Formulation du problème dual	95
6.4	Résolution du problème	95
6.4.1	Partitions de l'ensemble d'apprentissage	96
6.4.2	Calcul des variables duales	96
6.4.3	Algorithme d'apprentissage	97
6.5	Validation de l'algorithme	99
6.6	Conclusion	99
7	Validation expérimentale	103
7.1	Introduction	103
7.2	Base de données	103
7.3	Résultats de la détection des complexes QRS	104
7.3.1	Détection de battements atypiques avec artefacts	104
7.3.2	Résultats statistiques de la détection	109
7.4	Résultats de la classification	109
7.4.1	Base d'apprentissage	110
7.4.2	Base de test	110
7.4.3	Protocole expérimental	111
7.4.4	Classification à coûts symétriques	112
7.4.5	Classification à coûts asymétriques	113
7.5	Conclusion	117
8	Conclusion et perspectives	119
A	Bilan scientifique relatif à la thèse	121
	Bibliographie	123

CHAPITRE 1

Introduction

Sommaire

1.1	Présentation	1
1.2	Problématique	2
1.3	Contributions personnelles	3
1.4	Organisation du manuel	3

1.1 Présentation

Les maladies cardiovasculaires constituent un problème majeur de santé publique. Elles viennent en tête des causes médicales de décès en Algérie, environ 19,7% en 2008 selon l'OMS (Organisation Mondiale de la Santé). Les facteurs de risques sont multiples : tabac, sédentarité, obésité, hypertension artérielle, diabète et facteurs génétiques.

Le cœur, organe central du système cardiovasculaire, peut être affecté par de nombreuses pathologies qui peuvent être soit bénignes, comme certaines tachycardies par exemple, soit très sérieuses, comme l'infarctus du myocarde qui cause 10% de décès dans le monde. En raison de l'ampleur du problème, le suivi des patients à risques devient primordial. Les arythmies mineures informent le médecin sur l'état cardiaque du patient. Elles doivent être détectées notamment pour prévenir une dégénérescence possible en arythmies sévères.

Avec l'évolution des techniques, les médecins disposent aujourd'hui d'outils performants pour observer le fonctionnement du muscle cardiaque et poser leur diagnostic. Parmi les examens cardiologiques possibles, l'ElectroCardioGramme (ECG) est l'examen le plus couramment effectué, car il est rapide à mettre en place, peu coûteux et surtout non invasif donc très peu contraignant pour le patient.

L'électrocardiogramme est une représentation graphique du potentiel électrique qui commande l'activité musculaire du cœur. Ce potentiel est recueilli par des électrodes posées à la surface du corps. Il se présente comme une suite de déflexions (ondes) répétitives représentant chacune une phase de fonctionnement du cœur. Chaque déformation visible sur ces ondes peut être attribuée à un dysfonctionnement cardiaque ou arythmie.

Les troubles du rythme cardiaque trouvent principalement leurs origines dans la naissance du stimulus cardiaque ou encore dans la conduction de l'onde de

dépolarisation à travers le myocarde (chemin suivi par l'onde de dépolarisation à partir de son point d'activation électrique) .

Une manière de détecter des troubles cardiaques, consiste à recueillir le signal électrique cardiaque (ECG) par des capteurs, puis l'analyser. Cette analyse présente des enjeux à la fois pratiques et théoriques pour les recherches actuelles en reconnaissance de formes et en médecine. L'objectif poursuivi à travers cette thèse est de proposer de nouvelles méthodes de reconnaissance d'arythmies cardiaques afin d'aider le médecin à lire les enregistrements de longue durée.

1.2 Problématique

L'ECG est très souvent complété par un examen similaire d'une durée de 24 heures appelé "Holter", examen au cours duquel le patient peut vaquer à ses occupations habituelles. Le principal avantage de l'enregistrement Holter par rapport à l'ECG de courte durée est qu'il permet la détection d'événements sporadiques qui n'interviennent pas nécessairement au cours des quelques secondes d'enregistrement ECG réalisé en milieu hospitalier, lorsque le patient est au repos.

L'analyse de tels enregistrements nécessite l'utilisation d'outils de lecture automatique du signal car la quantité d'informations enregistrées en 24 heures est très importante : elle correspond à environ 100 000 battements cardiaques sur 2 ou 3 voies d'enregistrements. Ces outils de lecture doivent permettre le repérage d'informations pertinentes, d'éventuelles arythmies ainsi que leur fréquence d'occurrence. Cette reconnaissance automatique est rendue possible grâce aux outils de traitement du signal ainsi qu'aux techniques d'apprentissage.

De manière générale, l'analyse d'un signal est composée de deux principales étapes : La première consiste à trouver une représentation du signal adaptée à l'information recherchée ; on exprime alors le signal original à l'aide de descripteurs. La deuxième étape consiste à analyser la description obtenue pour déduire les propriétés recherchées. Notons que chaque étape nécessite une étude détaillée du signal considéré : sa morphologie, son contenu fréquentiel ainsi que les zones d'intérêt.

Dans le cas de la reconnaissance d'arythmies cardiaques, la première étape consiste à traiter le signal brut provenant d'un enregistrement Holter souvent bruité : à le filtrer pour extraire le signal utile, à le transformer si nécessaire dans le domaine fréquentiel ou dans le domaine temps échelle afin de localiser et de segmenter les zones d'intérêt. Ces dernières nous permettent la quantification et la description de chaque battement. Notons que ces descripteurs sont ceux habituellement utilisés par le cardiologue pour poser son diagnostic. A ce stade, chaque battement doit être représenté par un vecteur de caractéristiques soigneusement choisies pour la discrimination entre un battement normal et un battement pathologique.

Dans la seconde étape, on considère que tous les battements de la base d'étude

sont des vecteurs de même dimension à introduire à l'entrée d'un classifieur. L'apprentissage du classifieur se fait sur un sous ensemble de battements préalablement identifiés par un médecin. On comprend bien que plus la première étape est effectuée rigoureusement ou, en d'autres termes, plus les descripteurs sont pertinents pour la propriété recherchée, plus la seconde étape se voit efficace et simplifiée.

1.3 Contributions personnelles

Cette thèse a été entamée avec l'idée de faire une classification de battements cardiaques par SVMs en les localisant par des méthodes classiques essentiellement basées sur les filtres et les dérivées comme la méthode publiée dans [Pan 1985]. Très rapidement nous avons été confrontés aux limites de ces méthodes lorsqu'elles sont appliquées à des enregistrements contenant une variété de battements atypiques et affectés par différents artefacts. Nous avons alors proposé une nouvelle méthode basée sur les caractéristiques spectrales de chaque type de battements [Zidelmal 2012a].

En ce qui concerne la classification, nous avons soulevé deux problèmes : le premier est que la plupart des classifieurs exposés dans la littérature considéraient que les deux types d'erreurs de classification avaient des coûts symétriques, ce qui est incompatible avec les principes d'un diagnostic médical. Nous avons alors introduit des coûts de classification non balancés. Le deuxième problème auquel nous avons été confrontés est l'application de la règle de décision de Bayes dans le cas des battements ambigus dont la probabilité d'appartenance à une classe est autour de 0.5. Comme il s'agit d'une aide au diagnostic médical, nous avons préféré introduire une option de rejet d'ambiguïtés dans le SVM afin de minimiser le coût de la classification. Cette nouvelle approche a été publiée en 2009 dans [Zidelmal 2009] puis améliorées à travers le travail publié dans [Zidelmal 2012b].

1.4 Organisation du manuel

Allant de l'étude de l'ECG à la reconnaissance automatique de battements, ce document est structuré comme suit :

Après un chapitre d'introduction, le second chapitre présente sommairement le fonctionnement du système cardiovasculaire. Il permet en particulier de comprendre l'origine des signaux électriques enregistrés par l'électrocardiogramme et les différents types d'artefacts qui l'affectent ainsi que quelques pathologies cardiaques typiques illustrées chacune par un enregistrement ECG correspondant.

La détection des battements cardiaques constitue le préalable de toute analyse automatique de l'ECG. Le battement cardiaque est l'onde ventriculaire de plus grande amplitude de chaque cycle. Elle constitue la zone pertinente révélatrice du fonctionnement cardiaque. Sa segmentation fait l'objet du troisième chapitre. Le

choix et le calcul des paramètres caractérisant chaque battement segmenté sont exposés à travers le quatrième chapitre.

A ce stade, chaque battement est représenté par un vecteur appelé vecteur de caractéristiques. Parmi ces vecteurs, nous en avons choisi un ensemble pour l'apprentissage du classifieur SVM dont le principe est donné en chapitre 5.

L'objectif de tout classifieur est de classer la majorité des échantillons de manière automatique. Néanmoins, dans certaines applications comme les diagnostics médicaux, il serait préférable de rejeter les points ambigus car le coût d'une mauvaise classification est élevé. Le SVM standard est alors modifié en y introduisant une option de rejet. Le problème d'optimisation correspondant est développé au sixième chapitre.

Pour pouvoir comparer les algorithmes que nous avons développés avec ceux exposés dans la littérature, nous les avons appliqués à des bases de données standards AHA (American Heart Association) database et MIT-BIH Arrhythmia Database [Mark 1988] habituellement utilisées pour la localisation de battements cardiaques et la reconnaissance automatique d'arythmies. Les résultats que nous avons obtenus avec de telles bases d'étude sont exposés et commentés à travers le septième chapitre qui s'attache particulièrement à décrire l'interaction entre le coût de la classification et les seuils de décision.

Enfin, nous concluons sur ce travail en proposant quelques pistes de recherche sur lesquels s'ouvre cette thèse.

Notions d'Electrocardiographie et reconnaissance d'arythmies

2.1 Introduction

Ce chapitre présente le contexte d'étude : la reconnaissance d'arythmies cardiaques. Il présente le mécanisme à l'origine de l'activité cardiaque. L'électrocardiogramme de surface et les problèmes accompagnant son enregistrement sont introduits. Les battements cardiaques dont la morphologie et la fréquence sont relatifs aux principales arythmies, sont passés en revue pour permettre de comprendre les problèmes de quantification de battements ainsi que leur reconnaissance automatique. Les différents blocs de base sont détaillés au regard des différents travaux de la littérature. Pour illustrations, nous considérons les enregistrements recueillis de la base de données MIT-BIH Arrhythmia (MITDB) ou encore de la base de données AHA (*American Heart Association*).

2.2 le système cardiovasculaire

Le système cardio-vasculaire assure la circulation du sang qui permet les échanges respiratoires et nutritifs indispensables à la vie. Depuis la découverte fondamentale par William Harvey en 1615 de l'existence de la petite circulation (pulmonaire) et de la grande circulation (systémique), de nombreux travaux sont venus enrichir la connaissance de la circulation sanguine. La grande circulation représente la vascularisation (ou perfusion) de toutes les cellules du corps, hormis les poumons. La petite circulation est celle qui concerne uniquement les poumons. Dans la grande circulation, l'oxygène du sang artériel est consommé par les cellules, ce qui produit le sang veineux. A travers les capillaires des poumons, le sang veineux se recharge en oxygène et s'artérialise.

La circulation proprement dite est assurée par le cœur, organe ayant le rôle de pomper et distribuer le sang. Il assure un échange régulier entre la petite et la grande circulation suivant un cycle bien précis. La régularité de ces échanges est commandée par un stimulus électrique. Ce dernier dépend du système nerveux autonome¹. Il parcourt le cœur du nœud sinusal à l'apex² pour déclencher la

¹Partie du système nerveux innervant notamment le cœur, les poumons et le tube digestif.

²extrémité inférieure du cœur.

contraction des différentes chambres qui constituent le cœur. Ce stimulus peut être observé en mesurant les différences de potentiel de plusieurs électrodes à la surface du corps. L'interprétation de ces différences de potentiel relève du domaine de l'électrocardiographie. Le cycle cardiaque ne tolère aucune interruption car certaines cellules meurent lorsqu'elles ne sont plus alimentées en sang artériel. C'est notamment le cas des cellules du cerveau, c'est pourquoi les défaillances cardiaques sont si fatales. Dans ce qui suit, l'activité du cœur est présentée selon deux axes : l'activité mécanique et l'activité électrique.

2.2.1 Activité mécanique cardiaque

Le cycle de la circulation sanguine se répète constamment et comprend deux périodes : la systole et la diastole. La systole est la période correspondant à l'éjection du sang dans la grande et la petite circulation. Elle est décomposée en trois phases : la systole auriculaire ³, la systole ventriculaire (contraction isovolumique et éjection) et la diastole, phase pendant laquelle le cœur se remplit de sang. Cette phase est composée de deux étapes : la relaxation isovolumique et le remplissage.

2.2.2 Activité électrique du cœur

Le mécanisme cardiaque, qui comprend l'expulsion du sang et l'ouverture-fermeture des valves, fonctionne uniquement grâce aux contractions du myocarde. Ces contractions sont déclenchées par la propagation de proche en proche d'un potentiel d'action.

Lorsqu'on parle du fonctionnement électrique du cœur, il faut revenir au niveau cellulaire et se rappeler qu'il existe une polarisation naturelle de la cellule. Chaque cellule myocardique réagit à un stimulus électrique ⁴ grâce à une membrane permettant une perméabilité sélective aux ions. Au repos, l'intérieur de la membrane cellulaire est chargé négativement par rapport à l'extérieur qui est pris comme référence. Dans cet état électrique stable, on dit que la cellule est polarisée.

Lorsque la cellule est stimulée électriquement, les propriétés de la membrane sont modifiées et sa perméabilité aux ions augmente. Les échanges ioniques à travers la membrane des cellules myocardiques donnent naissance au potentiel d'action. La figure (2.1) montre l'effet des échanges ioniques transmembranaires sur le potentiel d'action.

- **La phase 0** : c'est la phase de dépolarisation rapide. Elle est caractérisée par l'irruption des ions de sodium Na^+ à l'intérieur de la cellule et concourent

³Ejection du sang à partir des oreillette

⁴Le tissu nodal possède un centre d'automatisme au niveau du nœud sinusal qui se dépolarise spontanément et envoie des impulsions électriques de façon rythmique et sans stimulation externe.

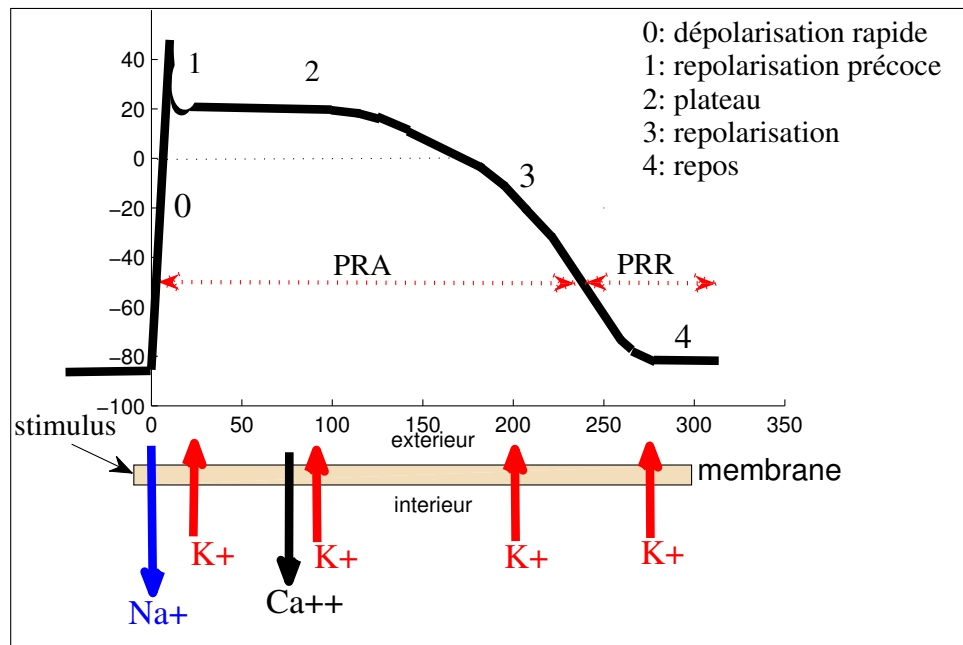


FIG. 2.1 – Les phases du potentiel d'action d'une cellule du muscle cardiaque.

à l'établissement d'un déséquilibre électrique entre l'intérieur et l'extérieur. Cette phase correspond à la dépolarisation rapide.

- **La phase 1** : c'est la phase de repolarisation précoce. Elle correspond à la fermeture des canaux sodium et l'ouverture des canaux potassium K^+ . Cette ouverture permet aux ions K^+ de suivre le gradient électrostatique.
- **La phase 2** : la fuite de potassium se poursuit mais l'entrée des ions de calcium Ca^{++} à l'intérieur de la membrane permet le maintien de la dépolarisation en plateau car il y'a un équilibre entre charges entrantes Ca^{++} et sortantes K^+ .
- **La phase 3** : l'accroissement de conductance au potassium est responsable des charges négatives intracellulaires, et donc de la repolarisation finale.
- **La phase 4** : elle correspond à la phase de repos. On obtient l'équilibre avec une différence de potentiel négative.

Remarque : Grâce à une enzyme appelée "pompe sodium-potassium", l'équilibre ionique initial est rétabli après un potentiel d'action.

L'excitabilité de la cellule myocardique, sa capacité à générer un potentiel d'action en réponse à une stimulation peut être décomposée en trois phases spécifiques :

- La période réfractaire absolue (P.R.A.) : période pendant laquelle tout sti-

mulus externe n'a aucun effet sur la cellule (aucune excitation possible).

- La période réfractaire effective (P.R.E.) : période incluant la P.R.A., on y ajoute une phase pendant laquelle la cellule peut être stimulée mais ne conduit pas.
- La période réfractaire relative (P.R.R.) : période pendant laquelle seul un stimulus puissant peut générer un potentiel d'action.

Le stimulus originel doit être supérieur au seuil d'excitabilité pour pouvoir déclencher le processus. Il doit être plus puissant encore en période réfractaire relative. On voit que la commande de la contraction des cellules myocardiques est temporisée par les périodes réfractaires. La période globale du cycle cardiaque est ainsi soumise aux périodes réfractaires de chaque cellule.

L'activation de proche en proche des cellules ne se fait que dans un certain sens (des cellules nodales jusqu'aux cellules myocardiques les plus éloignées) établissant un front directionnel de la dépolarisation du myocarde. La période réfractaire, jouant le rôle tampon pour empêcher un retour de stimulation dans le sens inverse, contribue à stabiliser l'activité électrique myocardique.

Séquence normale d'activation cardiaque : Dans une séquence normale d'activation cardiaque telle que décrite dans [Tawara 2000], le stimulus de départ du cycle cardiaque est généré par les cellules nodales. En suivant le parcours de cette onde électrique, on peut faire le lien avec le cycle mécanique. Tout d'abord les cellules des oreillettes sont dépolarisées, ce qui provoque leur contraction (systole auriculaire), l'onde traverse le nœud auriculo-ventriculaire et le septum inter-ventriculaire, puis les deux ventricules se dépolarisent (systole ventriculaire). Enfin, les cellules du myocarde entrent en repolarisation (relaxation ventriculaire) puis stabilisation (phase de repos).

2.2.3 Automatisme cardiaque

Il existe un automatisme cardiaque, c'est à dire, le cœur génère sa propre activité. Différents centres au niveau du cœur permettent d'avoir cet automatisme :

- Le nœud Sino-Atrial (SA)
- Le nœud AtrioVentriculaire (AV)
- Le faisceau de Purkinje.

Ces différentes structures permettent un automatisme avec des fréquences variables. Le nœud SA ayant la plus élevée, bat à une fréquence allant de 60 à 80 battements par minute. La progression de cette activité se produit le long de l'oreillette pour se propager au faisceau de His, aux branches de division de ce faisceau puis à l'ensemble du myocarde. Les parties du cœur participant à la propagation cette activité sont visibles sur la coupe longitudinale du cœur présentée sur la figure 2.2.

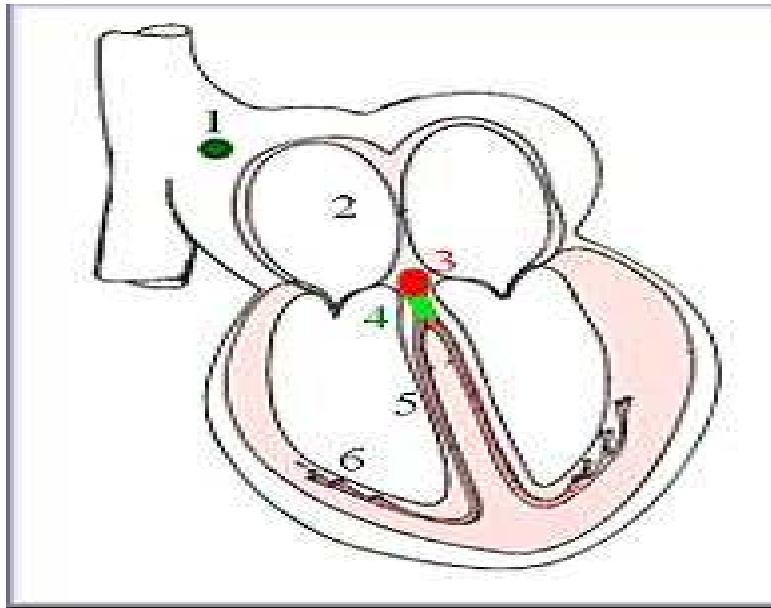


FIG. 2.2 – Coupe longitudinale du cœur présentant : 1 : Nœud sinusal ; 2 : Oreillettes ; 3 : Nœud AV ; 4 : Faisceau de His ; 5 : Branches ; 6 : Réseau de Purkinje.

2.3 Electrocardiographie de surface

Le corps humain étant électriquement conducteur, les potentiels d'actions générés lors de l'activité électrique cardiaque peuvent être recueillis par des électrodes placées sur le corps. L'enregistrement de l'activité électrique du cœur se fait sur un plan frontal ou sur un plan horizontal selon l'emplacement des électrodes. La représentation graphique d'une telle activité est appelée : ElectroCardioGramme (ECG).

L'ECG est un outil de diagnostic permettant de détecter toute pathologie liée à l'activité du cœur. Il nous permet de mesurer des grandeurs anatomiques et fonctionnelles du cœur :

- Orientation anatomique
- Taille
- Variation de conduction
- Arythmies et atteintes ischémiques
- Effet de variation des électrolytes
- Influence des thérapeutiques

2.3.1 L'électrocardiogramme

L'ECG se présente comme une suite de déflexions (ondes) correspondant chacune à une phase de fonctionnement du cœur (voir figure 2.3)

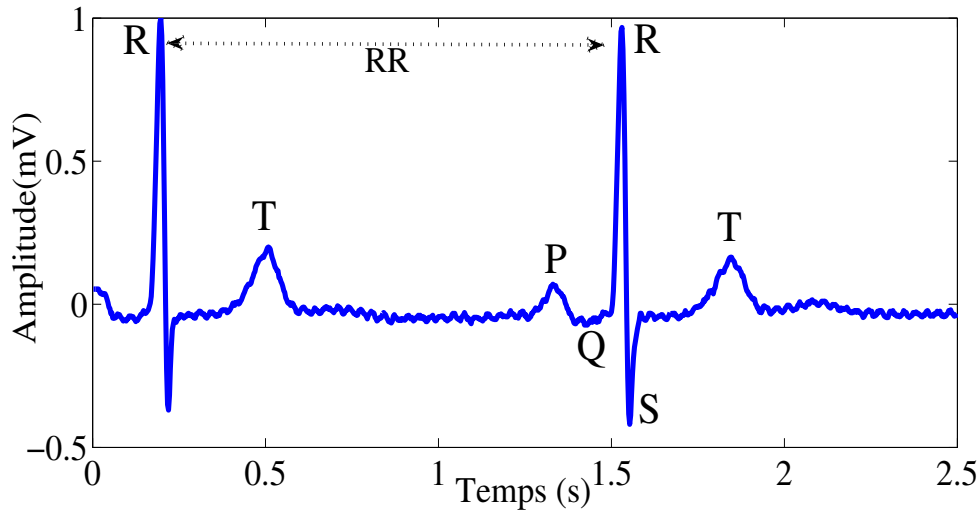


FIG. 2.3 – Allure d'un électrocardiogramme normal (Segment de l'enregistrement MITDB : 100).

- **Onde P** : elle est liée à la dépolarisation auriculaire. C'est une onde arrondie de petite amplitude (inférieure ou égale à 0,2 mV), sa durée est d'environ 120 ms. Les ondes P auriculaires précèdent régulièrement les complexes ventriculaires (QRS). La repolarisation auriculaire n'est pas visible sur l'ECG normal car elle est masquée par la dépolarisation ventriculaire.
- **Intervalle P-R** : c'est un court segment isoélectrique qui sépare l'onde P du complexe ventriculaire (QRS), il correspond à la conduction auriculo-ventriculaire. Sa durée est de l'ordre de 200 ms.
- **Complexe QRS** : C'est l'onde la plus pertinente, elle correspond à l'activation et la dépolarisation ventriculaire. Cette onde est couramment appelée : battement cardiaque, sa durée est de l'ordre de 80 ms.
- **Segment ST** : Il correspond au début de la repolarisation ventriculaire. Il est généralement isoélectrique et suit horizontalement la ligne de base.
- **Intervalle QT** : cet intervalle va du début du complexe QRS à la fin de l'onde T, il représente la durée de l'activation ventriculaire. Sa durée est inversement proportionnelle à la rapidité du rythme cardiaque soit, plus le rythme est rapide, plus QT est court.
- **Onde T** : elle est le témoin électrique de la repolarisation du muscle myocardique. Sa durée est imprécise du fait de sa fin progressive. Elle est généralement dirigée dans le même sens que le complexe QRS. Sa forme est asymétrique, avec un premier versant en pente faible, et un deuxième versant descendant en pente rapide.

2.3.2 Techniques d'enregistrement

Les modalités d'enregistrement sont variées. Elles se distinguent selon l'emplacement des électrodes sur la surface du corps. L'enregistrement de plusieurs tracés (projections du signal sur diverses lignes du corps est appelé : système de dérivations ECG). Un enregistrement effectué au moyen d'électrodes cutanées placées sur les membres et le thorax chez un sujet allongé sur le dos, est le plus habituel. Il définit l'ECG de surface standard. Ce même enregistrement effectué chez un sujet qui pédale par exemple une bicyclette ergométrique ou marche sur un tapis roulant représente l'ECG d'effort.

2.3.3 Les systèmes de dérivations

On appelle dérivation, un circuit électrique déterminé par un couple d'électrodes reliées à un appareil de mesure. L'élément de base d'une électrode est une plaque d'argent revêtue d'une couche de chlorure d'argent (gel insoluble), ce qui la rend très conductrice.

Ces électrodes sont positionnées à des endroits bien définis sur corps du patient afin d'explorer la quasi-totalité du champ électrique cardiaque en offrant un ensemble cohérent de dérivations non redondantes (six périphériques et six précordiales).

a. Dérivations standard d'Einthoven : Ce sont des dérivations bipolaires déterminées à partir de deux électrodes exploratrices. Ces électrodes sont placées à des endroits bien déterminés (voir figure 2.4) : Le poignet droit (R), le poignet gauche (L) et la cheville gauche (F). La cheville droite est généralement reliée à la masse en vue d'atténuer les bruits de mode commun. Ces trois dérivations nous permettent l'enregistrement des différences de potentiels suivantes :

$$\begin{cases} D_1 = V_L - V_R , \\ D_2 = V_F - V_R , \\ D_3 = V_F - V_L . \end{cases} \quad (2.1)$$

b. Dérivations unipolaires : Ces dérivations sont indiquées par la figure 2.5. Elles permettent d'obtenir des signaux de plus grande amplitude. Chaque signal enregistré représente la différence entre le potentiel d'une électrode et la moyenne des potentiels recueillis par les deux autres électrodes :

$$\begin{cases} aVR = V_R - \frac{V_L + V_F}{2} , \\ aVL = V_L - \frac{V_R + V_F}{2} , \\ aVF = V_F - \frac{V_L + V_R}{2} . \end{cases} \quad (2.2)$$

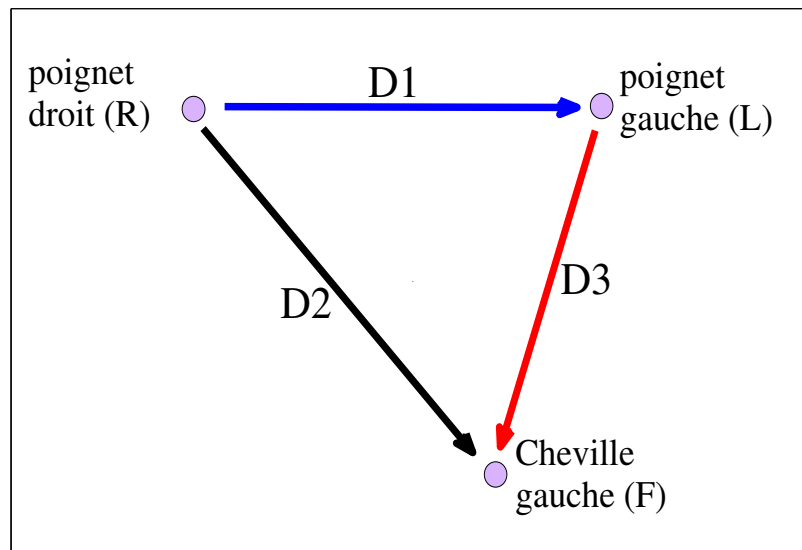


FIG. 2.4 – *Dérivations bipolaires d'Einthoven.*

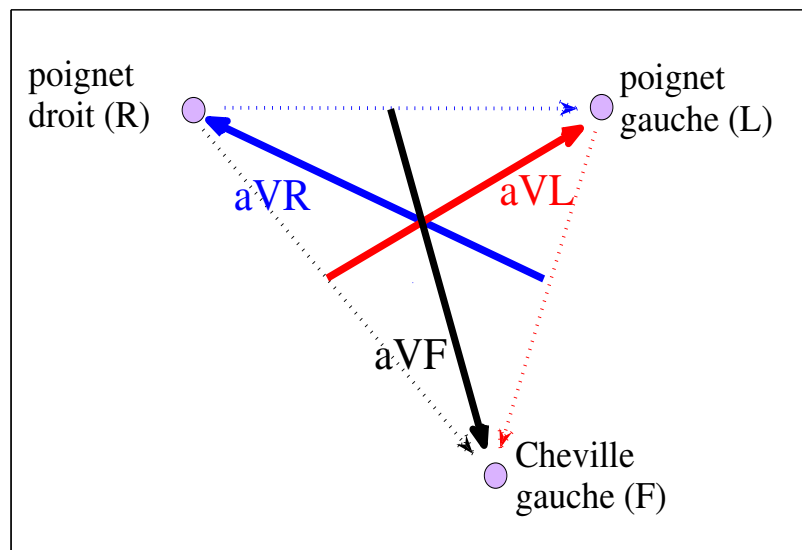


FIG. 2.5 – *Dérivations unipolaires des membres.*

c. Dérivations précordiales : Ce sont des dérivations unipolaires mises au point par N.F.Wilson en 1934 [Wilson 1944]. Les électrodes sont posées sur le thorax. Ces dérivations sont désignées par la lettre V suivie du numéro de leur emplacement. Le potentiel de l'électrode exploratrice est pris par rapport à la moyenne des potentiels V_R , V_L et V_F . Les Six points, définis par Wilson permettent d'obtenir les dérivations V_1 à V_6 . (voir figure 2.6). Les dérivations précordiales ont deux caractéristiques qui les distinguent fondamentalement des dérivations des membres : elles mesurent l'activité électrique cardiaque sur le plan horizontal et les électrodes sont posées à proximité du cœur.

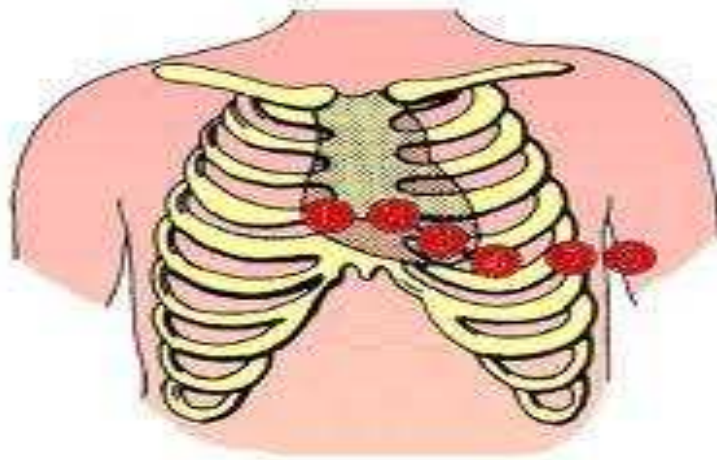


FIG. 2.6 – Dérivations précordiales de Wilson.

2.3.4 Conditions d'enregistrement

Pour un enregistrement de bonne qualité, le respect de certaines conditions est nécessaire. Le patient doit être couché sur le dos, en résolution musculaire complète, dans une position confortable et protégé du froid afin d'éliminer au maximum les ondulations de la ligne de base et les parasites dûs aux tremblements musculaires ou au mauvais contact électrode-peau.

2.3.5 Le Holter

C'est une technique d'enregistrement de l'activité cardiaque d'un sujet pendant 24 ou 48 heures, lui permettant de continuer ses activités normalement, sans

alitement ni hospitalisation. L'enregistrement se fait sur une bande magnétique ou un support numérique. Cet examen permet l'analyse du rythme cardiaque d'un individu de façon à y déceler d'éventuelles pathologies cardiovasculaires.

Le Holter est particulièrement indiqué dans le cas de certaines pathologies cardiaques mais aussi en prévention, chez des individus porteurs d'un stimulateur cardiaque qu'il faut surveiller. Son utilisation est également recommandée pour la surveillance d'arythmies cardiaques (extrasystoles entre autres) ainsi que dans les cas où le diagnostic ne peut être posé avec certitude ; c'est le cas de certaines angines de poitrine par exemple, ou lorsque l'épreuve d'effort est impossible ou non significative, pour dépister une ischémie myocardique silencieuse (c'est-à-dire sans douleur thoracique) chez des sujets à facteurs de risque élevés (tabac, alcool, sucre, athérome, hypertension artérielle, etc).

2.4 Artefacts visibles sur l'ECG

Sur tout enregistrement électrocardiographique, des événements indésirables appelés artefacts peuvent apparaître et brouiller le tracé. Le problème est surtout posé dans le cas d'un traitement automatique où la présence de ces bruits peut induire en erreur le diagnostic final. Ces perturbations ont fait l'objet de plusieurs études [Moody 1984, Mark 1988, Chen 2006, Borries 2005] mais certaines d'entre elles restent encore difficiles à traiter de manière automatique.

Ces bruits peuvent avoir plusieurs origines : techniques, physiques, pathologiques, ou pharmacologiques. Nous allons surtout développer l'aspect technique et physique des artefacts présents sur les tracés électrocardiographiques notamment sur les tracés Holter particulièrement bruités.

2.4.1 Bruits techniques

Le matériel utilisé lors de l'enregistrement doit être manipulé avec précaution car il peut être source de bruits dont les plus courants sont présentés ci-dessous :

a. Bruits dûs au secteur : Le réseau de distribution électrique peut parfois brouiller le signal électrocardiographique avec une onde dont l'harmonique fondamentale est à 50 Hz (voir Figure 2.7). Ce bruit apparaît sur tous les enregistrements. Il peut être assez fort mais s'élimine facilement avec un filtre sélectif car c'est un bruit de haute fréquence à bande étroite.

b. Bruits dûs au mauvais contact électrode-peau : Lorsque les électrodes sont incorrectement connectées, des sauts brusques de la ligne de base apparaissent (voir Figure 2.8). L'effet sur le tracé peut aller de la simple diminution d'amplitude à l'apparition de pics lorsque les électrodes sont en contact intermittent avec la peau. Ces pics peuvent parfois être confondus avec les ondes du tracé normal.

Ce type de bruit s'élimine difficilement car sa puissance se trouve dans la même gamme de fréquence que celle des complexes QRS.

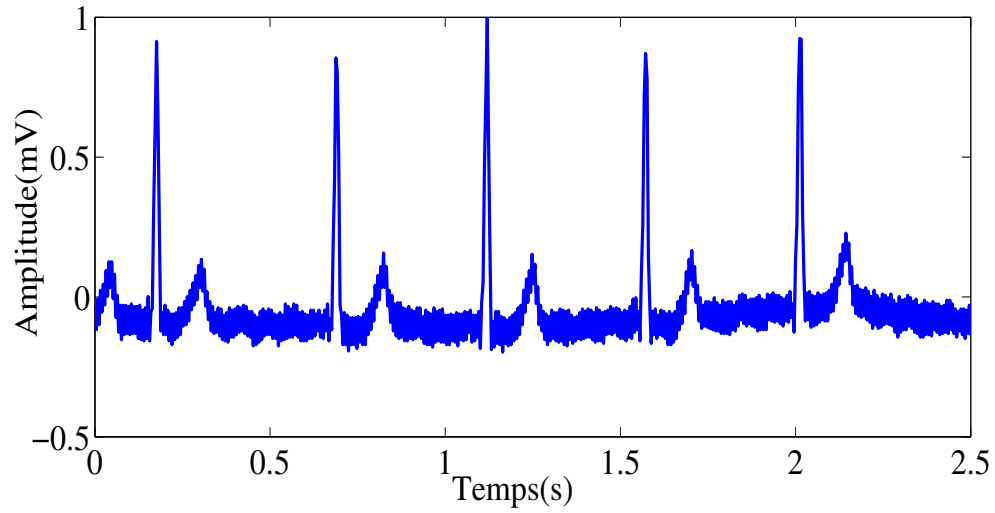


FIG. 2.7 – *Signal électrocardiographique perturbé par le secteur. (Segment de l'enregistrement AHA : 0201).*

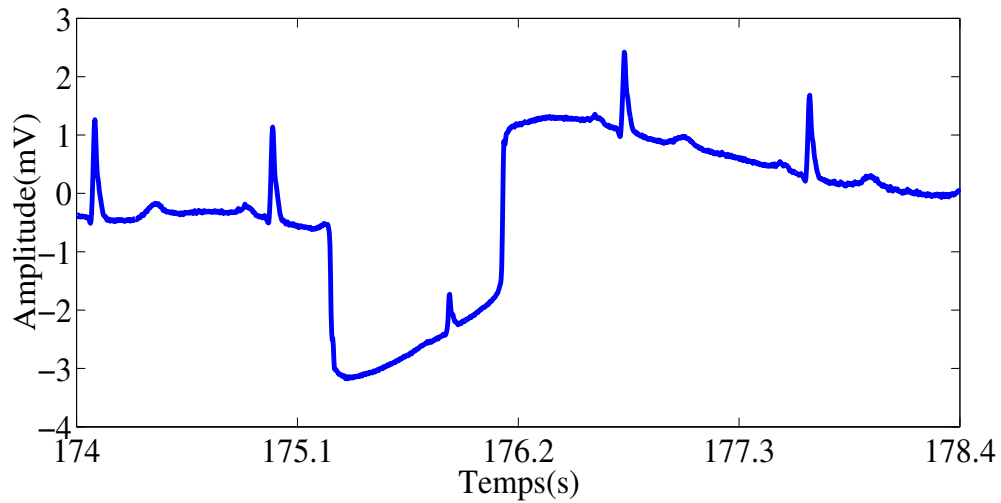


FIG. 2.8 – *Bruit dû au mauvais contact électrode-peau (MITDB : 101).*

c. Autres bruits courants : Parmi les bruits courants, on peut citer les artefacts dûs aux mouvements des câbles électriques, à la saturation des instruments de mesure, aux mauvais câblages, au port de vêtements synthétiques, etc.

2.4.2 Bruits physiques

Les artefacts physiques sont dûs aux activités électriques du corps humain telles que les contractions des muscles ou encore la respiration.

a. Dérives de la ligne de base : Lors de l'enregistrement de l'ECG, l'activité respiratoire peut faire osciller la ligne de base du signal à un rythme régulier. D'autres perturbations peuvent avoir pour effet de déplacer temporairement la ligne de base comme, par exemple, les mouvements du patient. Un tel bruit est visible sur la figure 2.9. Ces perturbations sont généralement peu gênantes pour l'analyse de l'ECG et peuvent être en grande partie filtrées car leur puissance se situe dans une bande de fréquence basse et empiète peu sur celle de l'ECG.

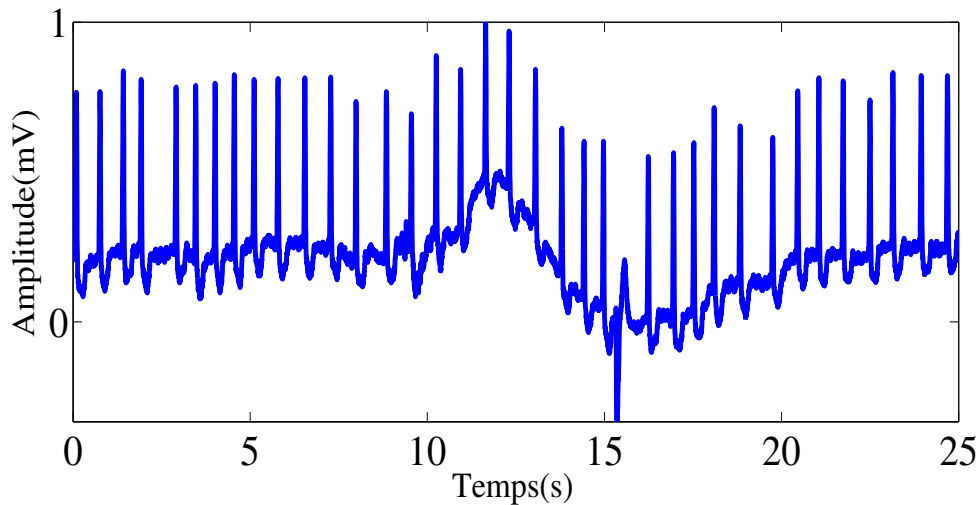


FIG. 2.9 – Ondulations de la ligne de base (MITDB : 119).

b. Bruits myoélectriques ou tremblements somatiques : La contraction d'un muscle est commandée par une dépolarisation des cellules musculaires. Bien que les électrocardiographes soient construits pour être surtout sensibles aux fréquences du myocarde, l'ECG enregistre aussi les contractions des muscles squelettiques. L'aspect le plus courant de ce bruit est une oscillation de haute fréquence et non stationnaire. Elle est liée à un tremblement musculaire d'un sujet qui n'est pas convenablement détendu. Un exemple de tel bruit est représenté sur la figure 2.10. Ces perturbations sont assez gênantes lorsque le patient bouge beaucoup ou lorsqu'il frissonne, elles peuvent noyer les ondes P et T et empêcher parfois la détection des pics R. L'apparition de ces perturbations dépend de l'état du patient, s'il est très tendu ou atteint de maladies (Parkinson par exemple), l'enregistrement peut être de mauvaise qualité sur toutes les voies de l'ECG.

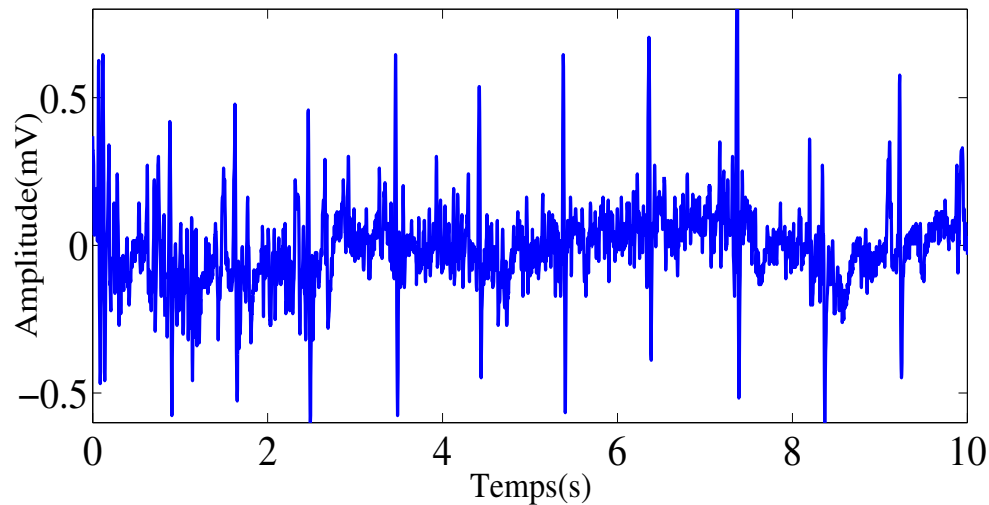


FIG. 2.10 – *Exemple de bruits musculaires sur un segment de l'enregistrement MITDB : 106.*

c. Autres artefacts altérant l'ECG : Certaines maladies généralisées peuvent affecter le tracé électrocardiographique. L'hyperthyroïdie, l'ischémie, l'hypokaliémie (prolongement de l'intervalle QT, onde T aplatie), modifient l'électrocardiogramme. L'usage de médicament, notamment la digoxine qui bloque la conduction AV et ralentit la fréquence cardiaque. La digitaline provoque un abaissement du segment ST avec inversion des ondes T et tend à raccourcir l'intervalle QT.

2.5 Le rythme cardiaque sinusal

En absence de toute pathologie, le rythme est sinusal. Le rythme sinusal est le rythme cardiaque normal. Il correspond à une activation physiologique des oreillettes, puis des ventricules à partir du nœud sinusal. Son rythme est compris entre 60 et 80 battements par minute avec un intervalle régulier entre des battements normaux (voir figure 2.11). Le cœur s'accélère normalement lors d'une activité physique, dans des circonstances physiologiques qui exigent un surcroît de demande métabolique ou sous l'effet des émotions ou d'excitants tels que le café et le tabac.

2.6 Troubles du rythme et de la conduction cardiaque

Sous cette dénomination, nous regroupons les arythmies cardiaques et les blocs cardiaques. Le meilleur outil pour diagnostiquer une arythmie est l'électrocardio-

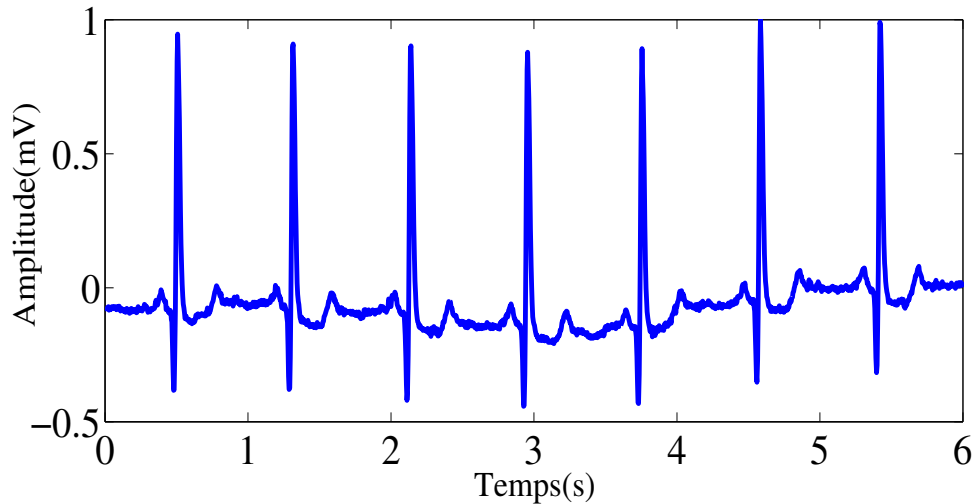


FIG. 2.11 – *Rythme sinusal, segment de l'enregistrement AHA : 0201.*

gramme. Dans l'analyse de l'ECG, les pathologies sont détectées et classées en fonction de leur déviation par rapport au rythme idéal qu'est le rythme sinusal. Chaque déviation visible sur l'ECG peut être attribuée à une anomalie physiologique [Lake 1990]. Ainsi, les blocs cardiaques sont dus à un défaut de conduction de l'onde de dépolarisation à travers le myocarde et les arythmies sont générées par un foyer ectopique prenant le relais ou supplantant le nœud sinusal. Ces pathologies ne sont pas exclusives, un patient peut être atteint d'arythmies et de blocs cardiaques. Si certaines arythmies passent inaperçues, d'autres peuvent donner lieu à une sensation de palpitations, de malaise, d'essoufflement ou même conduire à l'évanouissement. Une lecture complète sur les arythmies peut être faite dans [Beasley 2003].

Les paragraphes suivants détaillent les blocs et certaines arythmies cardiaques que nous envisageons de reconnaître de façon automatique.

2.6.1 Les blocs cardiaques

Les blocs cardiaques sont dus à une rupture de conduction dans le myocarde, ce qui altère sa dépolarisation. Ces ruptures peuvent être plus ou moins sévères : freinantes (allongement du temps de parcours), intermittentes (un stimulus sur 2 ou 3 est conduit), ou complète (aucune conduction). On distingue quatre types de blocs cardiaques.

a. Bloc Sino-Auriculaire (BSA) : Le nœud sinusal peut ne pas transmettre de stimulus aux cellules des oreillettes. La conséquence est qu'au moins un cycle complet n'est pas effectué. Après la pause, due au bloc, le cycle reprend normalement si aucun autre foyer ectopique n'a déclenché de contraction.

b. Blocs Auriculo-Ventriculaire (BAV) : On appelle BAV l'altération de la conduction du stimulus de dépolarisation entre les oreillettes et les ventricules. On distingue trois degrés de sévérité.

- Les BAV de premier degré provoquent l'allongement du segment PR de chaque cycle.
- Les BAV de deuxième degré traduisent l'absence momentanée d'onde QRS après une onde P normale. Lorsque les segments PR précédents sont normaux, on parle de Mobitz de type II. Lorsque les segments précédents sont rallongés, on parle de Mobitz de type I.
- Les BAV de troisième degré sont dit complets, c'est-à-dire qu'aucune dépolarisation auriculaire ne parvient aux ventricules. Un foyer ectopique ventriculaire ou jonctionnel joue alors le rôle de pacemaker. Le foyer est identifiable par la forme et la fréquence des battements. Les activités auriculaire et ventriculaire sont complètement dissociées.

c. Blocs de branches : Un bloc de branche est dû au blocage de la dépolarisation dans l'une des branches du faisceau de His. Un bloc dans l'une ou l'autre branche provoque un retard dans la dépolarisation du ventricule auquel elle appartient. La dépolarisation des ventricules est désynchronisée et le complexe QRS se voit élargi. Les figures 2.12 et 2.13 indiquent respectivement un bloc de branche gauche ou *Left Bundle Branch Block*(LBBB) et un bloc de branche droite ou *Right Bundle Branch Block*(RBBB).

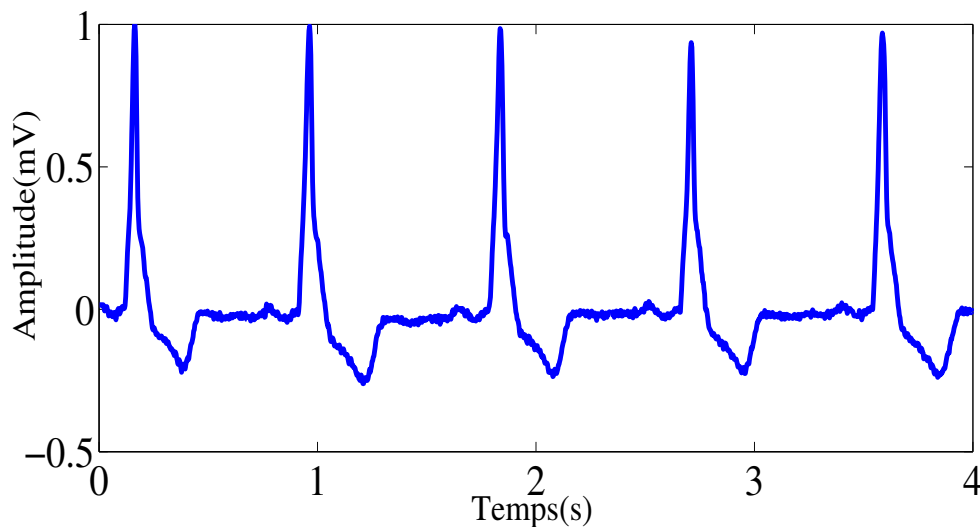


FIG. 2.12 – Exemple de Bloc de Branche Gauche (MITDB : 111).

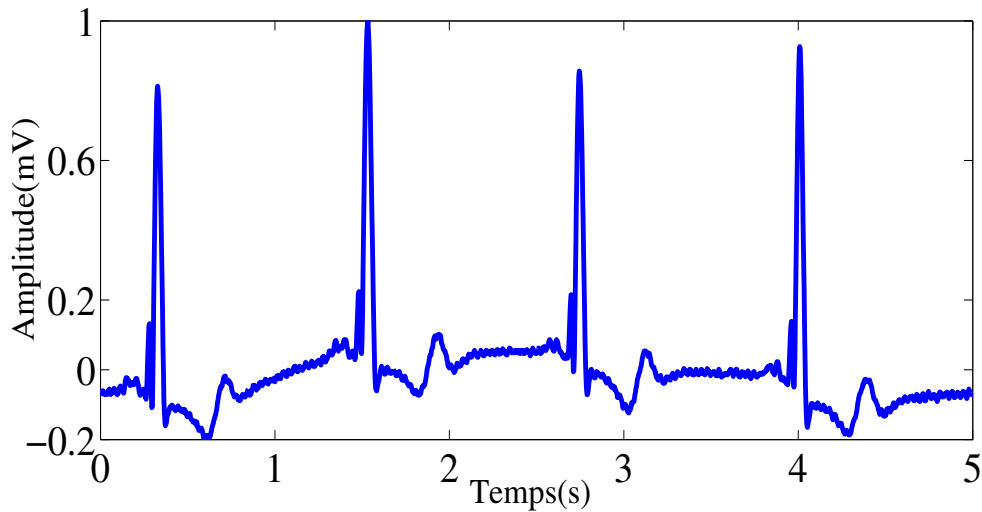


FIG. 2.13 – Exemple de Bloc de Branche Droite (MITDB : 124).

2.6.2 Les arythmies ventriculaires

Les arythmies proprement dites relèvent de l'entrée en jeu d'un foyer ectopique qui peut se situer dans n'importe quelle portion du cœur, ou de la formation d'un circuit électrique (appelé réentrée) dont la localisation peut être auriculaire, jonctionnelle (entre oreillettes et ventricules) ou ventriculaire.

a. L'arythmie extrasystolique : C'est la plus fréquente. Les extrasystoles ESV ou PVC sont des battements ectopiques, uniques ou répétés, provenant d'un seul ou de plusieurs foyers qui peuvent entraîner des sensations désagréables, de coups dans la poitrine, d'arrêts du cœur ou de palpitations. Les extrasystoles sont des phases systoliques en trop. Elles apparaissent sur l'ECG comme complexes QRS larges (environ 120 ms) et prématurés. Un exemple type de battement PVC est illustré par la figure 2.14 présentant un segment de l'enregistrement MITDB : 119.

Les extrasystoles ne constituent habituellement pas en elles-mêmes un facteur de gravité, leur pronostic dépend de l'état cardiaque du patient. Elles peuvent être absolument normales (extrasystoles dites bénignes) comme elles peuvent être pathologiques. Lorsqu'il existe un double foyer ventriculaire, on parle de doublet ventriculaire (voir Figure 2.15).

b. Les bigéminismes et trigéminismes : Ce sont des rythmes à deux commandes. La commande de base (généralement sinusale) est interrompue par des battements d'origine ectopique. Lorsqu'on se trouve en présence d'un bigéminisme, les QRS qui appartiennent au rythme de base sont suivis d'un QRS d'origine ectopique avec une succession de 1/1. On parle de trigéminisme lorsqu'on est en présence d'une succession 2/1. Les figures 2.16 et 2.17 présentent respectivement

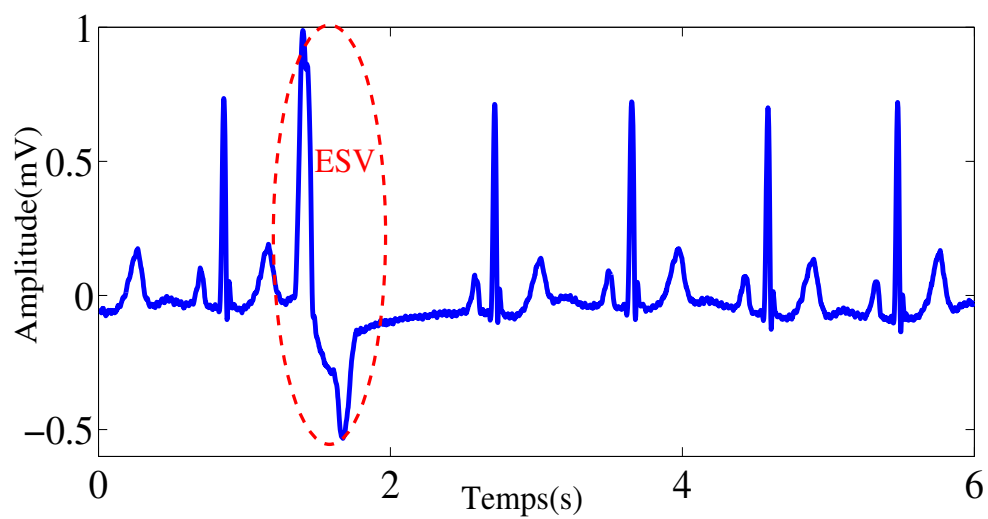


FIG. 2.14 – Exemple d'extrasystole ventriculaire (MITDB : 119).

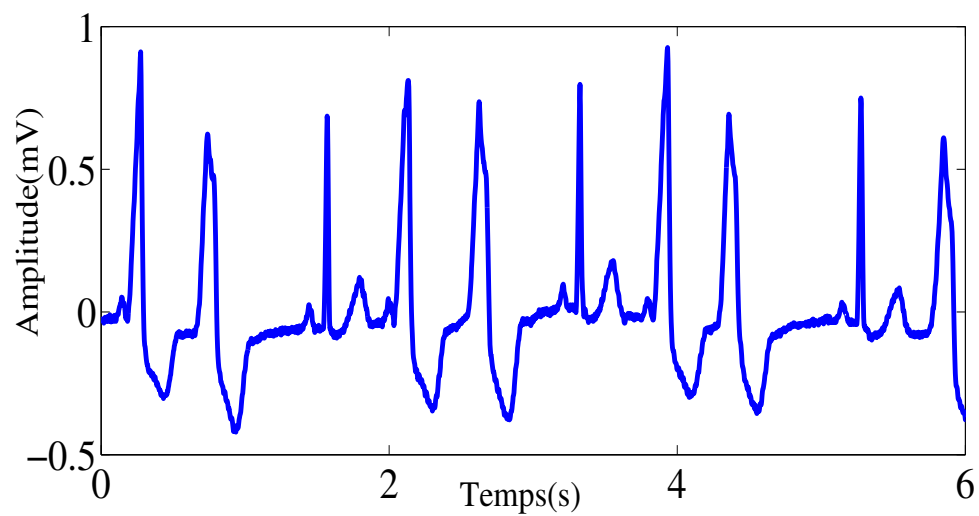


FIG. 2.15 – Exemple de doublet ventriculaire (MITDB : 208).

un exemple de bigéminisme et de trigéminisme ventriculaires.

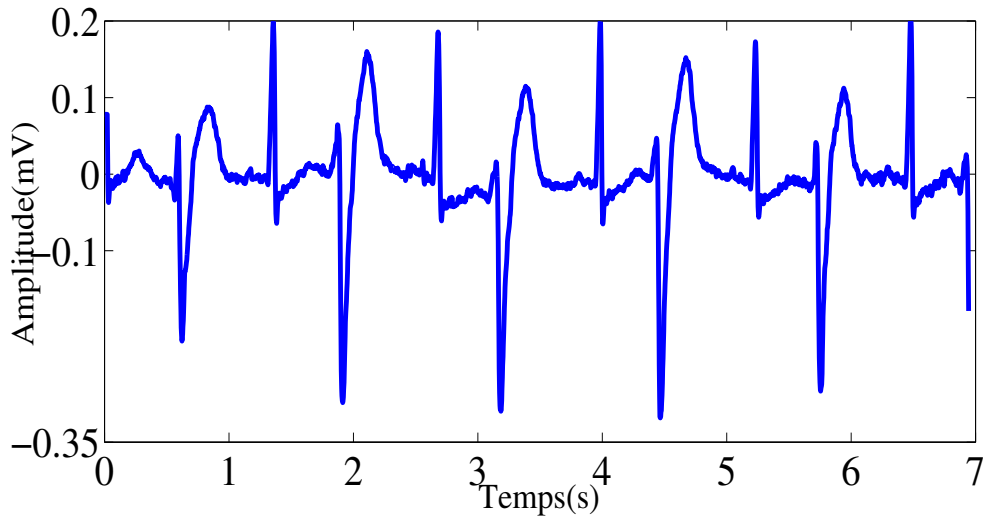


FIG. 2.16 – Exemple de bigéminisme (MITDB : 200).

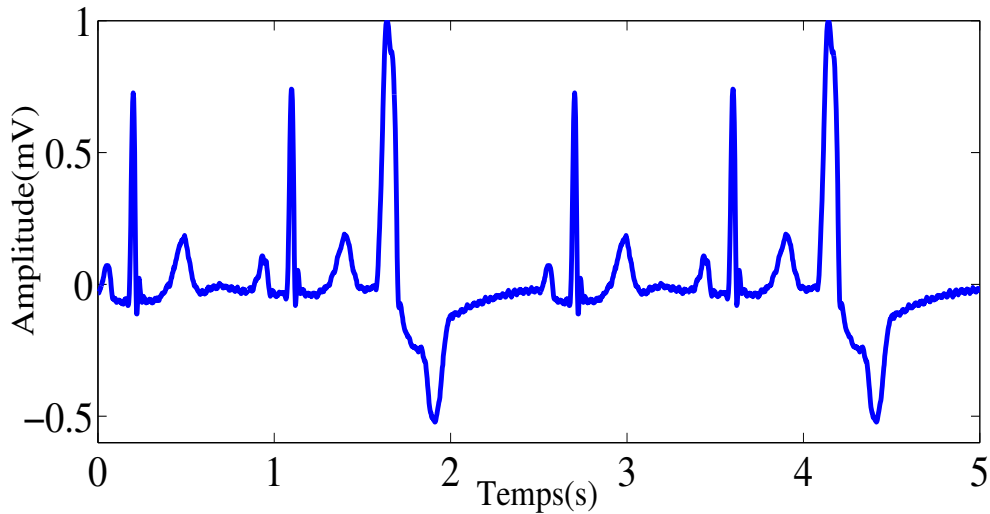


FIG. 2.17 – Exemple de trigéminisme (MITDB : 119).

2.7 Reconnaissance automatique d'arythmies : Etat de l'art

Depuis quelques décénies, la reconnaissance automatique d'arythmies cardiaques est devenue un domaine de recherche très actif. Il représente un carrefour

entre plusieurs disciplines comme la médecine et la reconnaissance de formes. La classification automatique d'arythmies cardiaques passe par deux étapes :

- Description et quantification des motifs à reconnaître.
- Apprentissage et décision.

2.7.1 Paramètres descriptifs d'un battement cardiaque

Le calcul de ces paramètres se fait sur le signal ECG dans le domaine temporel, dans le domaine fréquentiel ou dans le domaine temps-fréquence selon l'information recherchée. Les performances d'un descripteur de battements cardiaques dépendent de plusieurs facteurs comme la qualité de l'enregistrement, les prétraitements, la base d'étude ainsi que le nombre de dérivations étudiées.

Avant de procéder à la quantification des battements cardiaques, leur localisation est nécessaire. La détection des complexes QRS constitue le préalable de toute analyse automatique de l'ECG. Durant ces 30 dernières années, plusieurs chercheurs se sont penchés sur ce problème en posant divers algorithmes basés pour la plupart sur la dérivée de l'onde R ou sur le contenu fréquentiel des complexes QRS. Les différentes méthodes ont été récapitulées et regroupées par type dans [Kohler 2002].

La qualité d'un classifieur automatique de battements cardiaques est étroitement liée au choix des paramètres descriptifs. Ces derniers doivent être les plus discriminants possibles sans s'éloigner de ceux habituellement utilisés par les médecins. Chaque battement cardiaque est principalement décrit par sa morphologie. Certains auteurs l'ont simplement représentée en sous-échantillonnant le segment QRS [De-Chazal 2004, De-Chazal 2006], d'autres ont utilisé des outils de compression en appliquant les fonctions d'Hermite [Lagerholm 2000], ou la Transformée de Karhunen Loève. Cette idée fût proposée par [Moody 1989] puis reprise par [Gomez 2006]. Ces méthodes de compression permettent de réduire relativement le nombre d'échantillons en sauvegardant les caractéristiques temporelles et fréquentielles du battement ainsi que sa morphologie. Pour distinguer un battement normal d'un battement anormal, V. Krasteva et al. ont procédé au calcul des coefficients d'inter-corrélation entre les battements de la base d'étude et un battement de référence convenablement choisi dans chaque classe considérée [Krasteva 2007].

Comme il a été exposé précédemment, les battements anormaux proviennent d'un foyer ectopique ou d'un pacemaker. Ils peuvent aussi être dûs à la rencontre d'un bloc de branche droit (RBBB) ou gauche (LBBB). Ces battements sont connus pour leur largeur et leur lenteur. A partir de là, certains auteurs ont pensé à analyser le contenu fréquentiels des battements par la Transformée de Fourier [Minami 1999] ou par bancs de filtres [Afonso 1999]. La transformée en ondelettes offrant les possibilités d'analyser séparément les composantes fréquentielles s'avère de plus en plus appropriée pour le traitement des signaux biologiques et notamment pour cette application où les descripteurs temps-fréquences s'avèrent efficaces [Christov 2006].

L'étude de la régularité des battements cardiaques est incontournable lors de l'établissement d'un diagnostic. L'intervalle RR est un paramètre classiquement utilisé par les cardiologues et par tous les chercheurs qui travaillent sur la reconnaissance automatique d'arythmies [De-Chazal 2004, De-Chazal 2006, Krasteva 2007].

2.7.2 Algorithmes d'apprentissage et de décision

Le score d'un classifieur dépend étroitement du choix et du calcul des paramètres caractéristiques de la forme à reconnaître. Cependant, le choix de la méthode de classification n'influence pas moins le résultat et le taux de bonne classification. La littérature propose une variété de règle d'apprentissage : Les réseaux de neurones ont été appliqués avec succès pour la discrimination de battements cardiaques [Hu 1997, Yeap 1990, De-Chazal 2004, De-Chazal 2006, Patra 2010], d'autres travaux ont utilisé les K plus proches voisins [Christov 2005]. A coté de ces méthodes d'apprentissage, les Support Vector Machines (SVMs) connus pour leur excellent pouvoir de généralisation en classification et en régression ont conduit à des résultats prometteurs dans l'aide au diagnostic médical [Osowski 1997, Zidelmal 2009].

2.7.3 Problèmes liés à la reconnaissance automatique d'arythmies

D'après la littérature, plusieurs problèmes sont souvent rencontrés lors de la classification des désordres cardiaques. Parmi ces problème, on cite :

- Le signal ECG présente une variabilité élevée inter et intra-patient, dans la morphologie et la synchronisation. Par conséquent, la réalisation d'un classificateur globale est un problème très difficile.
- Les paramètres extraits à partir du signal d'ECG sont très susceptibles aux variations de la morphologie de l'ECG qui est non stationnaire ainsi qu'aux artefacts qui l'affectent.
- Les processus d'extraction de paramètres sont confrontés à la question du choix des paramètres les plus appropriés (les plus discriminants) tout en réduisant les vecteurs de caractéristiques de façon optimale.

2.8 Conclusion

A travers cette section, nous avons introduit les éléments de base de l'électrophysiologie du cœur, de son fonctionnement et des différents aspects de l'ECG liés à son enregistrements ou à la présence d'arythmies. Ces éléments sont jugés nécessaires pour bien mener les prochains chapitres.

Après une introduction sur l'anatomie du cœur humain, nous avons décrit brièvement ses deux activités électrique et mécanique ainsi que le lien entre elles.

Dans une deuxième partie de ce chapitre nous avons exposé les notions d'électrocardiographie. Quelques pathologies cardiaques typiques ont été présentées et illustrées chacune par un enregistrement ECG correspondant. Les différents types d'artefacts qui peuvent affecter le signal ECG enregistré et contraindre son traitement sont également exposés. Nous avons terminé ce chapitre en exposant quelques algorithmes de reconnaissance automatique d'arythmies cardiaques ainsi que les contraintes souvent rencontrées.

Segmentation de battements cardiaques

3.1 Introduction

La segmentation d'un signal ECG consiste à le partitionner en trames successives et homogènes définissant chacune un cycle cardiaque. Dans notre cas, l'objectif que nous poursuivons est d'isoler sans intervention humaine les complexes QRS de chaque cycle cardiaque afin d'en extraire les caractéristiques qui nous permettraient de reconnaître un battement normal et un battement pathologique.

Cette opération doit être la plus efficace possible. Pour cela, plusieurs étapes sont nécessaires : tout d'abord un ensemble de prétraitements dont le rôle est d'éliminer les artefacts qui nuiraient à la segmentation ; ensuite la détection des pics des ondes R et la segmentation des complexes QRS.

Notre base d'étude est constituée de signaux fortement bruités et non stationnaires, surtout en phases d'arythmies. En plus des artefacts qui sont d'origines et de natures diverses, la base présente une grande variabilité de la forme du signal, d'un sujet à un autre, et même chez le même sujet, d'un cycle à un autre. Pour l'analyse de tels signaux, l'outil que nous adoptons est l'analyse multirésolution car elle permet la caractérisation du signal sur différentes bandes de fréquences.

3.2 Analyse multirésolution

Le principe de base de l'Analyse MultiRésolution (AMR) consiste à séparer de façon itérative, le signal en deux composantes, l'une représentant l'allure du signal (approximation) et l'autre, ses détails. Cette opération est réalisée grâce à la projection du signal sur deux sous espaces vectoriels orthogonaux et complémentaires, l'un appelé espace des approximations et l'autre espace des détails.

3.2.1 La fonction Ondelette

Nous commençons d'abord par définir l'ondelette ψ , une fonction à support compact, oscillante, de moyenne nulle et de carré sommable. On définit une famille $\psi_{a,b}$ d'ondelettes à partir d'une ondelette mère ψ par sa dilatation ou compression

et sa translation :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (3.1)$$

où a est le facteur d'échelle (dilatation, compression) et b , le facteur de translation, paramètre de localisation temporelle. La pondération en $1/\sqrt{a}$ permet d'avoir des fonctions analysantes de même norme : toutes les ondelettes de la même famille ont la même énergie.

$$\forall a > 0, \forall b \in \mathbb{R}, \int_{-\infty}^{+\infty} |\psi_{a,b}(t)|^2 dt = \|\psi\|^2 = 1$$

Il existe plusieurs familles d'ondelettes decrites dans tous les ouvrages traitant la théorie des ondelettes [Truchetet 1998, Mallat 2009].

Dans l'expression $\psi(\frac{t-b}{a})$, le pas de translation à l'échelle a est b/a . On pose :

$$a = a_o^j \text{ et } b = nb_o a_o^j$$

avec $a_o, b_o \in \mathbb{Z}$. Si on choisit $a_o = 2$ et $b_o = 1$, on parle alors de **fonction ondelette dyadique**. Pour un niveau de résolution j , on a :

$$\psi_{j,n}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - n) \quad (3.2)$$

Les fonctions $\psi_{j,n}$ forment alors une base orthonormée par translation et dilatation ou compression :

Relation entre fonctions ondelettes : Soit $\psi_{0,0}(t)$, la fonction ondelette au niveau 0. On a :

- $\psi_{0,0}(t) = \psi(t) \mapsto \psi_{j,n}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - n)$ (dilatation de rapport 2^j)
- $\psi_{0,0}(t) = \psi(t) \mapsto \psi_{-j,n}(t) = 2^{\frac{j}{2}} \psi(2^j t - n)$ (compression de rapport 2^j)

3.2.2 La fonction d'échelle

L'apparition de l'analyse multirésolution coïncide avec l'introduction d'une seconde fonction appelée : fonction d'échelle φ : *le père des ondelettes* telle que la famille :

$$\varphi_{j,n} \bigcup \psi_{j,n}, \quad j \geq 0 \quad n \in \mathbb{Z}$$

forme une base orthonormée. Les fonctions $\varphi_{j,n} \in L^2(\mathbb{R})$ sont construites suivant la relation :

$$\varphi_{j,n}(t) = 2^{-\frac{j}{2}} \varphi(2^{-j} t - n) \quad (3.3)$$

La fonction ondelette et la fonction d'échelle sont étroitement liées. A chaque niveau de résolution, la fonction ondelette est une combinaison linéaire de sa fonction

d'échelle. On associe par exemple à la fonction ondelette de Haar (fonction en escalier) une fonction échelle (fonction porte). Ces deux fonctions sont respectivement données par :

$$\text{La fonction ondelette : } \begin{cases} \psi_{0,0}(t) = +1 & \text{si } 0 \leq t \leq 1/2, \\ \psi_{0,0}(t) = -1 & \text{si } 1/2 \leq t \leq 1, \\ \psi_{0,0}(t) = 0 & \text{ailleurs.} \end{cases} \quad (3.4)$$

$$\text{La fonction échelle : } \begin{cases} \phi_{0,0}(t) = 1 & \text{si } t \in [0, 1], \\ \phi_{0,0}(t) = 0 & \text{ailleurs.} \end{cases} \quad (3.5)$$

On peut également poser : $\psi_{0,0} = \varphi_{-1,0} + \varphi_{-1,1}$

3.2.3 Décomposition du signal (Analyse)

La décomposition d'un signal sur un niveau de résolution consiste en sa projection sur deux sous espaces vectoriels :

3.2.3.1 Espace d'approximation

Nous nous plaçons dans l'espace $L^2(\mathbb{R})$ de fonctions continues à variable réelle et à énergie finie. Une analyse à la résolution j d'une fonction $f(t)$ est obtenue par action d'un projecteur linéaire A_j sur $f(t)$. L'approximation de $f(t)$ sur ce niveau est :

$$a_j(t) \in V_j$$

V_j étant un sous espace de L^2 . On construit une analyse multirésolution à l'aide de sous espaces V_j emboîtés tels que le passage de l'un à l'autre soit le résultat d'un changement d'échelle. Par exemple, dans le cas dyadique :

$$a_j(t) \in V_j \Leftrightarrow a_j\left(\frac{t}{2}\right) \in V_{j+1}$$

ce qui correspond à une dilatation d'un facteur 2. L'espace V_{j+1} contient des signaux plus grossiers (de plus basse fréquence) que V_j , autrement dit : $V_{j+1} \subset V_j$. Pour un sous ensemble de L^2 , on a :

$$\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \dots \subset V_{j+1} \subset V_j \subset \dots \quad (3.6)$$

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \quad (3.7)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad (3.8)$$

Si la fonction d'échelle $\varphi_{j, n}$ engendre une base orthonormée V_j , la projection par A_j d'une fonction $f(t)$ sur cette base fournira les coefficients de cette décomposition. Ces coefficients qui décrivent l'approximation de $f(t)$ à l'échelle j sont appelés coefficients d'échelle et sont donnés par :

$$a_n^j = \langle f(t), \varphi_{j, n}(t) \rangle \quad (3.9)$$

et

$$a_j(t) = \sum_n \langle f(t), \varphi_{j, n}(t) \rangle \varphi_{j, n}(t) = \sum_n a_n^j \varphi_{j, n}(t) \quad (3.10)$$

La base étant orthonormée, $\|a_j(t)\|^2 = \sum_{n=-\infty}^{+\infty} |a_n^j|^2$

3.2.3.2 Espace des détails

On vient de voir que $V_j \subset V_{j-1}$, on peut alors définir pour chaque V_j , son complément orthogonal W_j dans V_{j-1} qui nous permet de récupérer les détails perdus en passant de V_{j-1} à V_j tel que :

$$V_{j-1} = V_j \oplus W_j$$

Comme W_{j-1} est orthogonal à V_{j-1} , W_{j-1} est orthogonal à W_j . Cette propriété s'écrit :

$$W_j \perp W_k, \quad \forall j \neq k$$

La base W_j est engendrée par la fonction ondelette $\psi_{j, n}$. L'approximation à l'échelle immédiatement plus fine pourra être reconstruite en utilisant les détails du signal fournis par sa projection sur la base W_j suivant la relation :

$$a_{j-1}(t) = a_j(t) + \sum_n \langle f(t), \psi_{j, n}(t) \rangle \psi_{j, n}(t) = a_j(t) + d_j(t)$$

$d_j(t)$ est la projection de $f(t)$ sur W_j . Le signal détail est décrit par les coefficients de détails notés : $d_n^j = \langle f(t), \psi_{j, n}(t) \rangle$ tel que :

$$d_j(t) = \sum_n d_n^j \psi_{j, n}(t)$$

Remarque : A chaque itération, seuls les signaux d'approximation sont à nouveau décomposés (voir figure 3.1). Dans la pratique, nous choisissons un nombre approprié de niveaux en nous basant sur la nature du signal, sa fréquence d'échantillonnage et l'application envisagée.

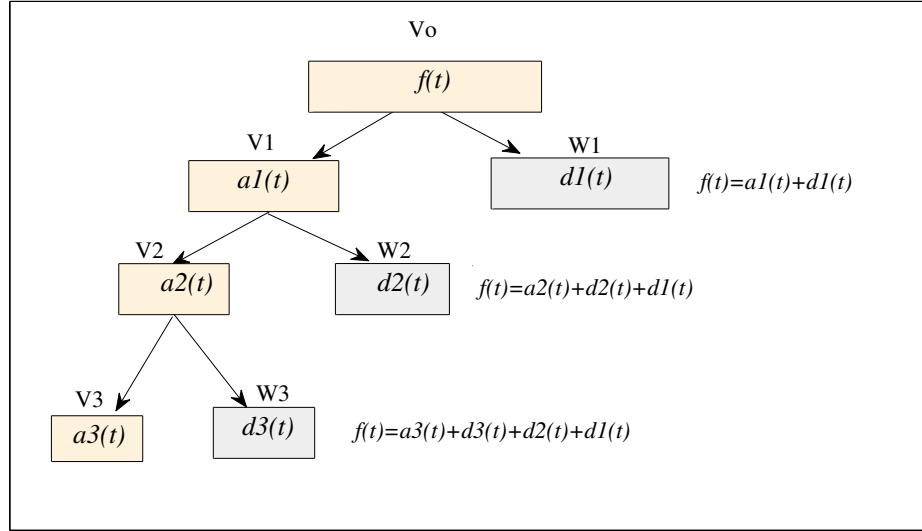


FIG. 3.1 – Arbre d'analyse multirésolution.

3.2.4 Algorithme d'analyse

Il existe une variété d'algorithmes d'analyse basés sur le même principe. Parmi ces algorithmes, celui de S.Mallat. Cet algorithme de décomposition est récursif. A chaque niveau de résolution, il permet de trouver l'approximation et les détails à la résolution immédiatement inférieure.

3.2.4.1 Calcul des coefficients d'approximation

Les points clés pour un tel calcul sont :

$$a_n^j = \langle f(t), \varphi_{j,n}(t) \rangle, \quad a_n^{j-1} = \langle f(t), \varphi_{j-1,n}(t) \rangle \text{ et } V_0 \subset V_{-1}$$

$\varphi(t)$ étant une fonction de V_0 , peut se décomposer sur V_{-1} alors, $\exists h[n]$, suite numérique avec $n \in \mathbb{Z}$ telle que :

$$\varphi(t) = \sum_n h[n] \varphi_{-1,n}(t) \quad (3.11)$$

La suite $h[n]$ peut être définie comme : $h(n) = \langle \varphi(t), \varphi_{-1,n}(t) \rangle$ avec, conformément à (3.3), $\varphi_{-1,n}(t) = 2^{\frac{1}{2}} \varphi(2t - n)$. On alors :

$$\varphi(t) = \sum_n h[n] 2^{\frac{1}{2}} \varphi(2t - n)$$

La suite $h[n]$ sera considérée comme la réponse impulsionnelle d'un filtre numérique. La décomposition précédente peut être généralisée pour des échelles quelconques :

$$\varphi_{j,n}(t) = 2^{-\frac{j}{2}} \sum_k h[k] 2^{\frac{1}{2}} \varphi(2(2^{-j}t - n) - k) \quad (3.12)$$

En regroupant les indices et les exposants, on obtient :

$$\varphi_{j, n}(t) = \sum_k h[k] \varphi_{j-1, k+2n}(t) \quad (3.13)$$

En posant $\ell = k + 2n$, les coefficients d'approximation $a_n^j = \langle f(t), \varphi_{j, n}(t) \rangle$ à la résolution j seront donnés par :

$$a_n^j = \sum_n h[\ell - 2n] \langle f(t), \varphi_{j-1, \ell}(t) \rangle \quad (3.14)$$

Si en plus on note : $\tilde{h}[n] = h[-n]$, on obtient :

$$a_n^j = \sum_n \tilde{h}[\ell - 2n] \langle f(t), \varphi_{j-1, \ell}(t) \rangle = \sum_n \tilde{h}[\ell - 2n] a_\ell^{j-1} \quad (3.15)$$

L'approximation $a_j(t)$ du signal $f(t)$ est donc le filtrage de l'approximation $a_{j-1}(t)$ par le filtre de réponse impulsionnelle $\tilde{h}$ suivi par un sous échantillonnage de rapport 2.

3.2.4.2 Calcul des coefficients de détail

Dans ce cas aussi, on part du fait que $W_0 \subset V_{-1}$. La fonction ondelette $\psi(t)$ peut alors être décomposée sur la base V_{-1} . Il existe alors une suite numérique $g(n)$ telle que :

$$\psi(t) = \sum_n g[n] \varphi_{-1, n}(t) \quad (3.16)$$

ou encore :

$$\psi(t) = \sum_n g[n] 2^{\frac{1}{2}} \varphi(2t - n)$$

La suite numérique $g[n]$ qui sera également considérée comme la réponse impulsionnelle d'un filtre numérique peut alors être construite en partant de l'expression précédente.

$$g[n] = \langle \psi(t), \varphi_{-1, n}(t) \rangle$$

Un calcul analogue en tous points au calcul des coefficients d'approximation, nous permet d'écrire les coefficients de détail à la résolution j comme suit :

$$d_n^j = \sum_n \tilde{g}[\ell - 2n] \langle f(t), \varphi_{j-1, \ell}(t) \rangle = \sum_n \tilde{g}[\ell - 2n] a_\ell^{j-1} \quad (3.17)$$

Le signal de détail $d_j(t)$ du signal $f(t)$ est donc le filtrage de l'approximation $a_{j-1}(t)$ par le filtre de réponse impulsionnelle $\tilde{g}$ suivi par un sous-échantillonnage de rapport 2.

Remarque : Certains algorithmes de décomposition sont complétés par un sur-échantillonnage de rapport 2 afin de garder le même nombre d'échantillons dans tous les vecteurs à la sortie des filtres.

Conclusion : La décomposition d'un signal par analyse multirésolution consiste à l'injecter dans deux filtres miroirs (un P-Bas et un P-Haut). A chaque itération, l'approximation du signal est à nouveau décomposée (voir figure 3.2).

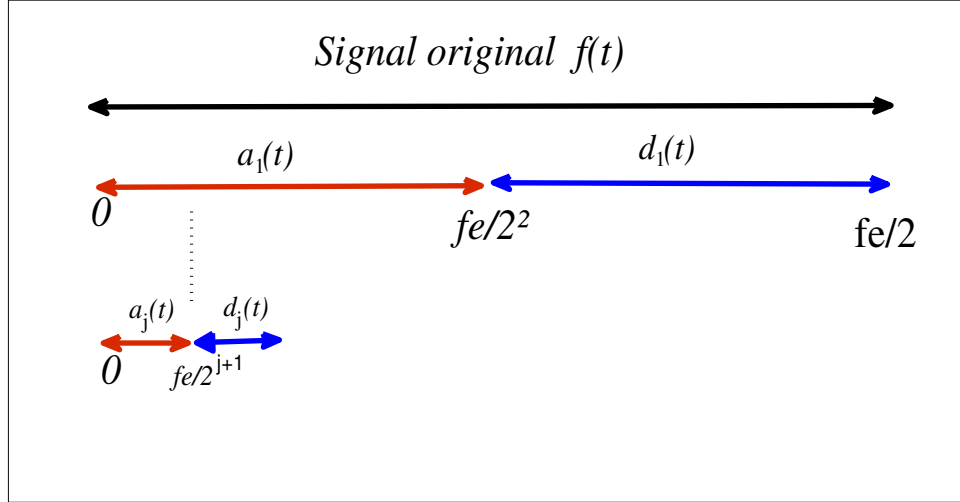


FIG. 3.2 – Décomposition dyadique d'un signal en bandes de fréquences.

3.2.5 Reconstruction du signal décomposé (Synthèse)

D'après les paragraphes précédents, on a :

$$a_{j-1}(t) = \sum_n a_n^j \varphi_{j, n}(t) + \sum_n d_n^j \psi_{j, n}(t) \quad (3.18)$$

A l'échelle $j - 1$, nous avons : $a_{j-1}(t) = \sum_n a_n^{j-1} \varphi_{j-1, n}(t)$ avec :

$$a_n^{j-1} = \langle f(t), \varphi_{j-1, n}(t) \rangle \text{ alors, } a_{j-1}(t) = \sum_n \langle f(t), \varphi_{j-1, n}(t) \rangle \varphi_{j-1, n}(t)$$

or la projection de la fonction $f(t)$ sur la base $\varphi_{j-1, n}(t)$ est égale à la projection de son approximation $a_{j-1}(t)$ sur la même base :

$$a_{j-1}(t) = \sum_n \langle a_{j-1}(t), \varphi_{j-1, n}(t) \rangle \varphi_{j-1, n}(t)$$

En remplaçant le deuxième terme $a_{j-1}(t)$ de l'expression précédente, par l'expression (3.18), on obtient :

$$a_n^{j-1} = \sum_k a_k^j < \varphi_{j,k}(t), \varphi_{j-1,n}(t) > + \sum_k d_k^j < \psi_{j,k}(t), \varphi_{j-1,n}(t) > \quad (3.19)$$

par ailleurs, nous avons vu dans la décomposition que :

$$\varphi_{j,k}(t) = \sum_{\ell} h[\ell] \varphi_{j-1,\ell+2k}(t)$$

On peut alors écrire le produit scalaire de deux fonctions d'échelle à deux résolutions successives comme suit :

$$< \varphi_{j,k}(t), \varphi_{j-1,n}(t) > = h(n - 2k) \quad (3.20)$$

et pour les fonctions ondelettes :

$$< \psi_{j,k}(t), \varphi_{j-1,n}(t) > = g(n - 2k) \quad (3.21)$$

L'équation de reconstruction est alors :

$$a_n^{j-1} = \sum_k a_k^j h(n - 2k) + \sum_k d_k^j g(n - 2k) \quad (3.22)$$

3.3 Filtrage du signal

Dans le cas des enregistrements Holter, nous retrouvons des bruits d'origines diverses (voir chapitre 2) car les conditions d'enregistrement ne sont pas contrôlées. Ces bruits sont présents parfois de façon transitoire, parfois de façon plus persistante avec une puissance plus ou moins élevée.

Pour pouvoir segmenter les battements cardiaques de manière efficace, nous sommes amenés à réaliser un certain nombre de prétraitements à savoir :

- La suppression des bruits de haute fréquence.
- La suppression des ondulations de la ligne de base.

3.3.1 Suppression du bruit de haute fréquence

Sans hypothèse précise sur ce bruit ajouté au signal, il a été supposé dans toute la littérature qu'il est blanc, gaussien, additif et très apparent sur les segments d'inactivité [Sornmo 2006, Pan 1985, Borries 2005, Lepage 2003]. Nous savons en outre, que du bruit musculaire dû à l'activité des autres muscles vient s'ajouter au signal. Malheureusement, ce bruit est beaucoup plus difficile à supprimer à cause du chevauchement entre son contenu fréquentiel et celui de l'ECG [Sornmo 2006]. En considérant que le bruit est additif, le signal ECG peut être exprimé par une fonction $f(t)$ telle que :

$$f(t) = u(t) + b(t) \quad (3.23)$$

où $u(t)$ est le signal utile et $b(t)$, le bruit.

Pour supprimer le bruit $b(t)$, plusieurs méthodes reportées dans la littérature proposent des performances satisfaisantes. Les premières méthodes étaient basées sur un calcul de moyenne sur plusieurs battements successifs du signal [Evanich 1972, Sornmo 2006] en supposant que le bruit est aléatoire et stationnaire. La réduction du bruit est alors proportionnelle au nombre de battements utilisés pour ce moyennage ; toutefois, un calcul sur un nombre trop important de battements peut causer des erreurs importantes à cause de la dispersion du signal entre les battements (notamment à l'effort). Ce type de filtre n'est pas applicable lorsqu'on cherche des informations sur les ondes, battement par battement. Les travaux effectués par la suite utilisaient des filtres adaptatifs ou des filtres à Réponse Impulsionnelle Infinie (RII) en choisissant une bande passante relative à l'information recherchée [Sornmo 2006, Pan 1985, Thakor 1984] ; néanmoins, ces méthodes conduisent souvent à une distorsion des ondes d'activation.

Le signal ECG étant un signal multi-composantes, non stationnaire affecté par un bruit dont certaines composantes comme le bruit musculaire sont corrélées, la représentation temps-échelle s'avère être un outil plus adapté à son traitement. La plupart des travaux récents proposent un filtrage par ondelettes [Lepage 2003, Borries 2005, Sornmo 2006].

3.3.2 Filtrage par seuillage des coefficients de détail

Dans la présente application, nous utilisons la méthode de filtrage proposée par Donoho et Johnstone dans [Donoho 1995, Johnstone 1996] utilisant le seuillage des coefficients de détails après la décomposition du signal par analyse multirésolution.

Les coefficients au dessus du seuil, caractéristiques de ruptures dans le signal, sont suffisants pour le reconstituer par synthèse. Les faibles coefficients en dessous du seuil, correspondent à du bruit et peuvent être supprimés. La figure 3.3 montre la décomposition d'un signal ECG sur trois niveaux de résolution. La décomposition jusqu'au niveau 3 a été choisie dans le but de ne pas dégrader les complexes QRS après le seuillage. Sur l'ensemble des coefficients, on remarque que ceux de faible amplitude ou de faible variance correspondent plutôt aux régions d'inactivité du signal ECG ne comportant que du bruit. Pour seuiller ces coefficients de détail, Donoho a proposé un seuil universel T défini comme suit :

$$T = \hat{\sigma} \sqrt{2 \log(N)} \quad (3.24)$$

où N est le nombre d'échantillons de la séquence à seuiller et $\hat{\sigma}$, l'estimation de l'écart type du bruit à la résolution la plus fine ($j = 1$). Ce seuil a été amélioré par Johnstone et Silverman afin de l'adapter à chaque niveau de résolution j et tenir compte ainsi des bruits non stationnaires et corrélés [Johnstone 1996]. On calcule alors un seuil à chaque niveau de résolution par :

$$T_j = \hat{\sigma}_j \sqrt{2 \log(N_j)} \quad (3.25)$$

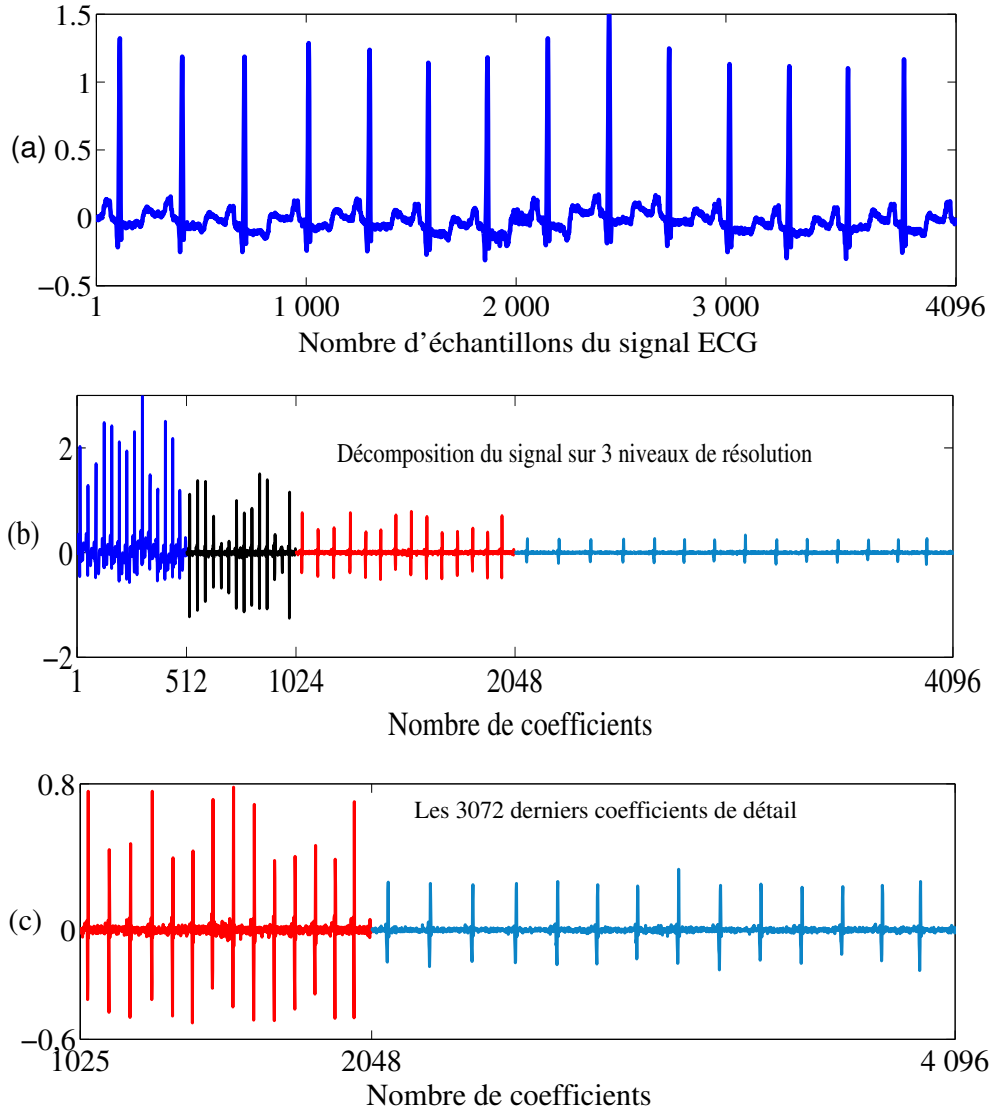


FIG. 3.3 – *a* : Signal ECG; *b* : analyse en ondelettes non redondantes du signal ECG (les différents niveaux de résolution ont été mis bout à bout, (d_n^1 : [2049, 4096]; d_n^2 : [1025, 2048]; d_n^3 : [513, 1024]; a_n^3 : [1, 512])). Du fait des changements d'échelle, les variations des petits coefficients ne sont pas visibles; *c* : Mise en évidence du bruit dans les derniers coefficients de détail de haute résolution.

où N_j est le nombre de coefficients dans le vecteur d_n^j et $\hat{\sigma}_j$, l'estimation de l'écart type du bruit à la résolution (j) .

3.3.2.1 Estimation de l'écart type du bruit

L'écart type σ du bruit est estimé par une méthode statistique utilisant la valeur médiane de la déviation absolue (*Median Absolute Deviation* ou MAD) qui représente un estimateur robuste. Pour un vecteur X , cette valeur est donnée par :

$$MAD(X) = \text{med}[|X - \text{med}(X)|]$$

où $\text{med}(X)$ représente la médiane ou la valeur centrale du vecteur X . L'écart type d'un bruit gaussien de moyenne nulle sur X peut être estimé par :

$$\hat{\sigma} = Q \cdot MAD(X)$$

Q est une constante qui dépend de la distribution du bruit. Dans le cas d'un bruit gaussien, $Q = F^{-1}(0.75) = 1.4826$, avec F^{-1} , l'inverse de la fonction de répartition de la distribution gaussienne. Dans ce cas, on a :

$$\hat{\sigma} = \frac{MAD(X)}{0.6745}$$

Plusieurs types de seuillage ont été proposés :

3.3.2.2 Seuillage dur

Soit d_n^j les coefficients de détail issus du signal $f(t)$ à la résolution j , et T_j la valeur du seuil de sélection correspondant. A l'issue du seuillage, on obtient le vecteur de coefficients seuillés d_{ns}^j tels que :

$$d_{ns}^j = \begin{cases} d_n^j & \text{si } |d_n^j| > T_j \\ 0 & \text{si } |d_n^j| \leq T_j \end{cases} \quad (3.26)$$

3.3.2.3 Seuillage doux

Le seuillage doux est similaire au seuillage dur sauf qu'il diminue de T_j la valeur des coefficients d_n^j se trouvant au-dessus du seuil :

$$d_{ns}^j = \begin{cases} d_n^j - T & \text{si } d_n^j > T_j \\ d_n^j + T & \text{si } d_n^j < -T_j \\ 0 & \text{si } |d_n^j| \leq T_j \end{cases} \quad (3.27)$$

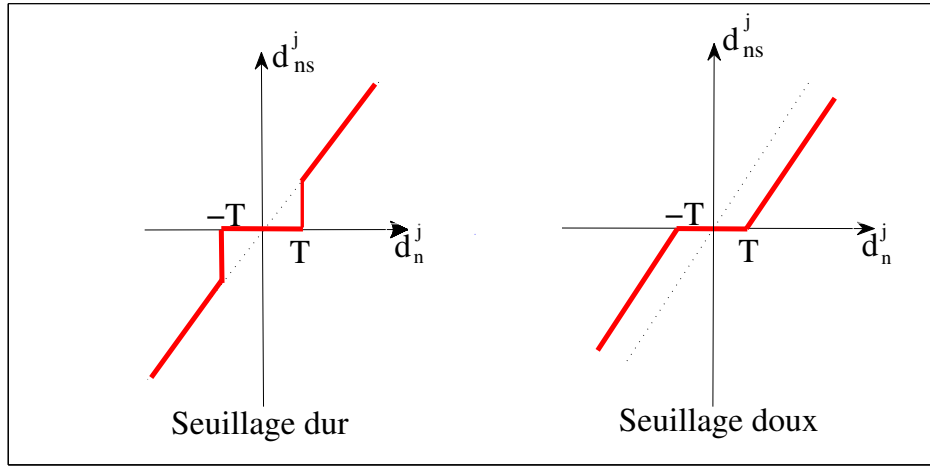


FIG. 3.4 – Comparaison des deux types de seuillage dur et doux : en abscisse, on présente les valeurs possibles des coefficients d'ondelettes (d_n^j) et en ordonnées les valeurs correspondantes qu'ils prennent à l'issue du seuillage (d_{ns}^j).

Une comparaison de l'effet des seuillages durs et doux sur les coefficients de détail peut être observée sur la Figure 3.4. Le seuillage dur conserve les coefficients de détail supérieurs aux seuils, contrairement au seuillage doux qui les atténue. De tels coefficients correspondent aux fortes ruptures dans le signal qui sont souvent les complexes QRS. Nous considérons alors le seuillage dur afin de sauvegarder les complexes QRS.

Les résultats de filtrage de deux segments d'ECG faiblement et fortement bruités sont respectivement donnés en figures 3.5 et 3.6.

3.3.3 Correction des dérives de la ligne de base

On appelle ligne de base, la ligne isoélectrique du cœur. Elle correspond au tracé qui serait observé sur un ECG si le cœur n'avait aucune activité électrique. Lorsque l'ECG est effectué en cabinet, ou pendant les périodes d'enregistrement nocturne du Holter, cette ligne est souvent horizontale car le patient n'effectue aucun mouvement. En revanche, dans le cas d'un enregistrement de longue durée (Holter par exemple), les mouvements et la respiration du patient modifient les positions relatives des électrodes, de sorte que cette ligne présente un tracé ondulé (voir figure 3.7).

Pour l'analyse d'un enregistrement ECG, un œil exercé fait abstraction de cette ligne : elle est prise comme référence pour étudier la forme et la hauteur des différentes ondes cardiaques. Dans le cas du traitement automatique, il est impératif de la repérer précisément pour fixer la référence des tensions dans l'algorithme global.

Puisque la dérive de la ligne de base est un phénomène s'illustrant dans

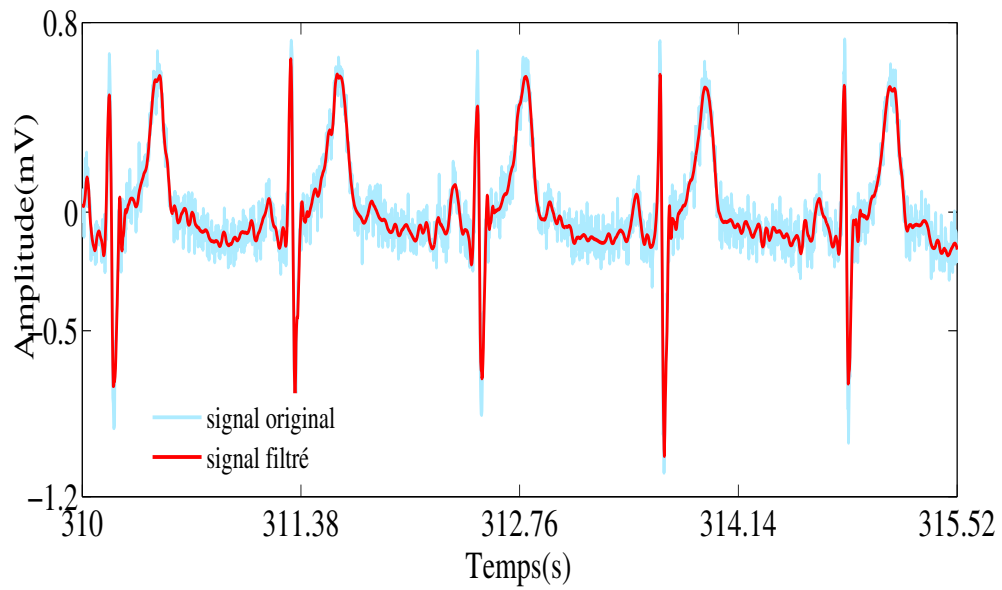


FIG. 3.5 – Résultat de l'algorithme de filtrage des hautes fréquences appliqué à l'enregistrement MITDB : 117.

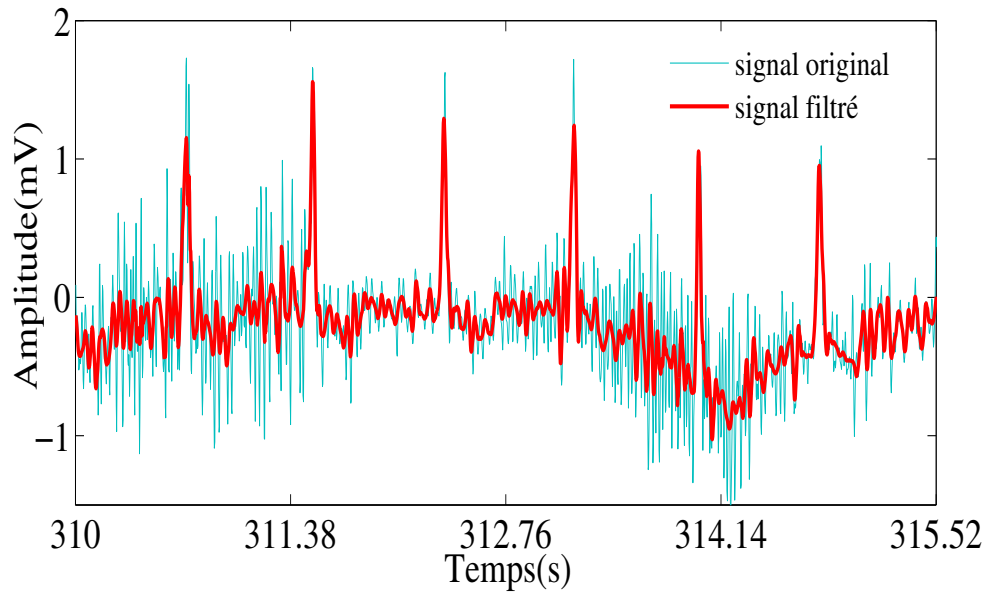


FIG. 3.6 – Résultat de l'algorithme de filtrage des hautes fréquences appliqué à l'enregistrement MITDB : 104 entaché d'un bruit musculaire non stationnaire.

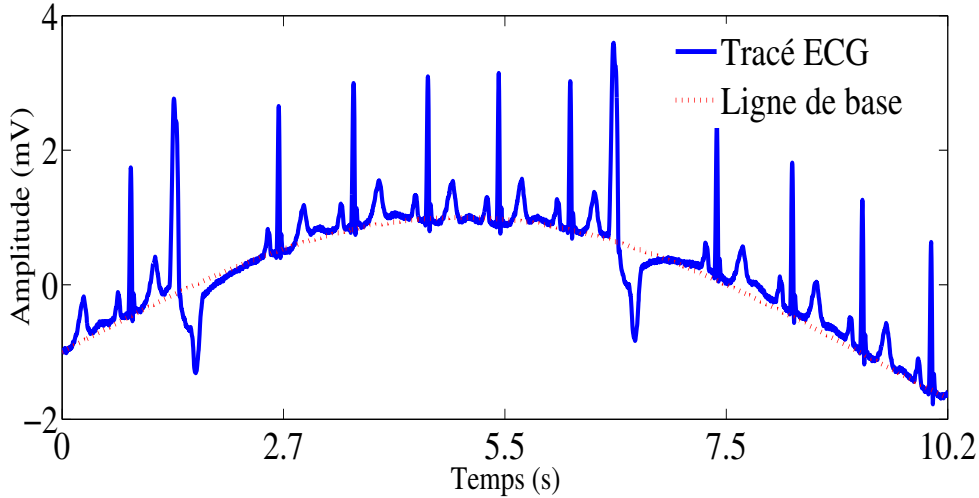


FIG. 3.7 – Ondulation de la ligne de base.

les basses fréquences du signal ECG, nous pouvons appliquer un filtrage passe-haut pour la supprimer. L'analyse multirésolution réalise un découpage itératif en bandes de fréquences. Elle s'avère alors utile pour corriger cette dérive [Arzeno 2008, Borries 2005].

La procédure consiste à décomposer le signal jusqu'à un niveau de résolution j suffisant pour séparer la ligne de base du reste du signal ECG. Nous reconstruisons alors le signal utile en annulant l'élément de plus basse fréquence. En tenant compte de la fréquence d'échantillonnage f_e des enregistrements utilisés et en nous référant aux travaux publiés dans [Thakor 1984, Pan 1985] où il a été montré que la puissance des complexes QRS appartient à une bande de fréquences supérieures à 5 Hz, nous avons fixé le niveau j à 6. La procédure équivaut alors à un filtrage P-Haut avec une fréquence de coupure de 2.8 Hz.

La figure 3.8 indique la décomposition jusqu'au niveau 6 d'un signal ECG avec dérive de la ligne de base. Nous constatons que ce bruit n'est représenté que par les coefficients d'échelle a_n^6 correspondant à une bande de fréquence $[0, 2.8\text{Hz}]$. Des résultats de suppression de la ligne de base sont donnés en figure 3.9 et en figure 3.10.

Nous tenons à noter que ces méthodes de filtrage utilisées n'affectent ni la morphologie ni le contenu fréquentiel des QRS qui constituent la zone d'intérêt de notre étude. Ces deux paramètres font partie des principales caractéristiques discriminantes des battements cardiaques. La figure 3.11 indique deux types de battements avant et après filtrage.

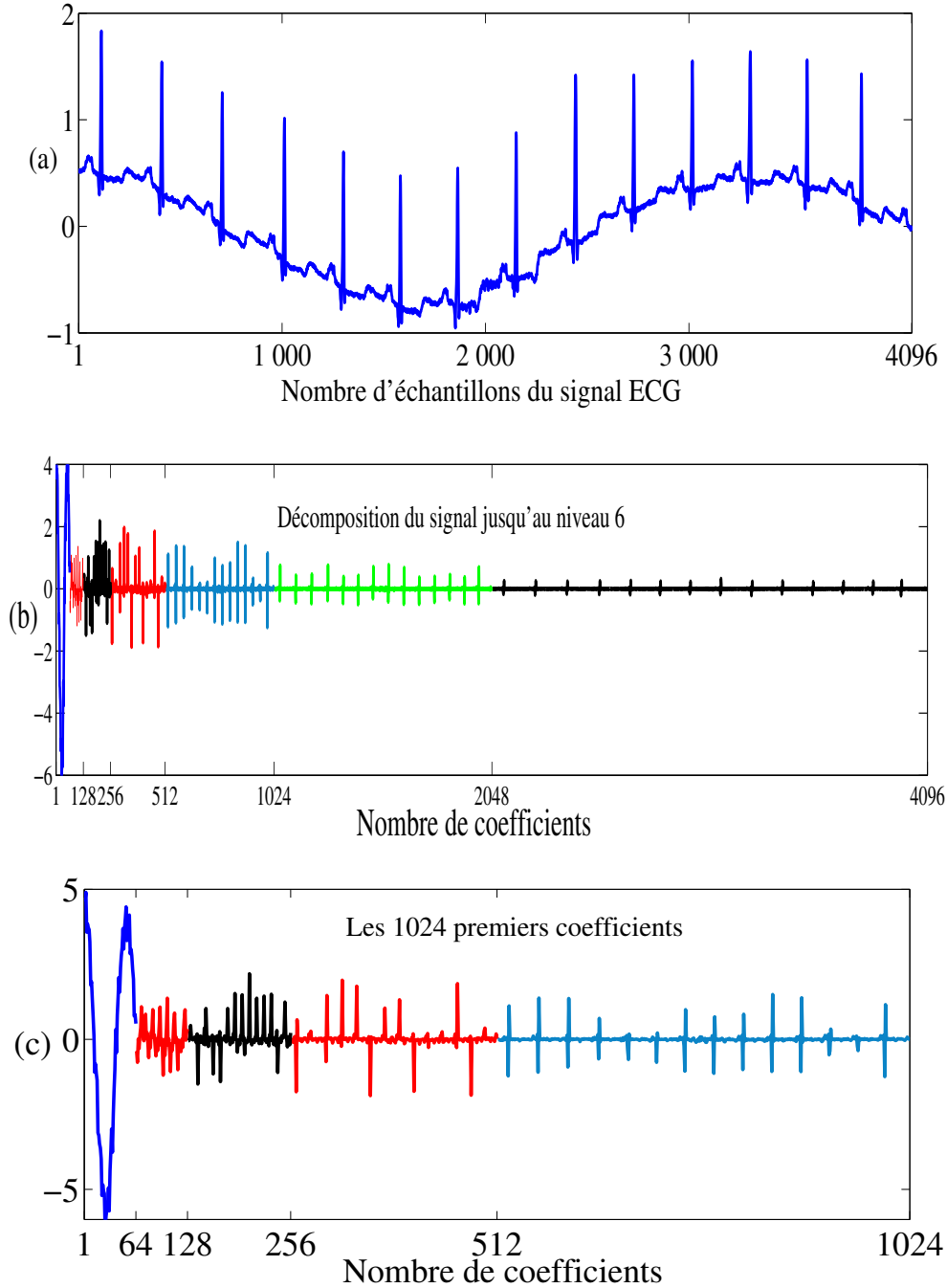


FIG. 3.8 – a : *Signal ECG*; b : *analyse en ondelettes non redondantes du signal ECG* (les différents niveaux de résolution ont été mis bout à bout, (d_n^1 : [2049, 4096] ; d_n^2 : [1025, 2048] ; d_n^3 : [513, 1024] ; d_n^4 : [257, 512] ; d_n^5 : [129, 256] ; d_n^6 : [65, 128] ; a_n^6 : [1, 64]). Du fait des changements d'échelle, les variations des petits coefficients ne sont pas visibles; c : *Mise en évidence de l'ondulation de la ligne de base sur les coefficients d'échelle de plus basse résolution.*

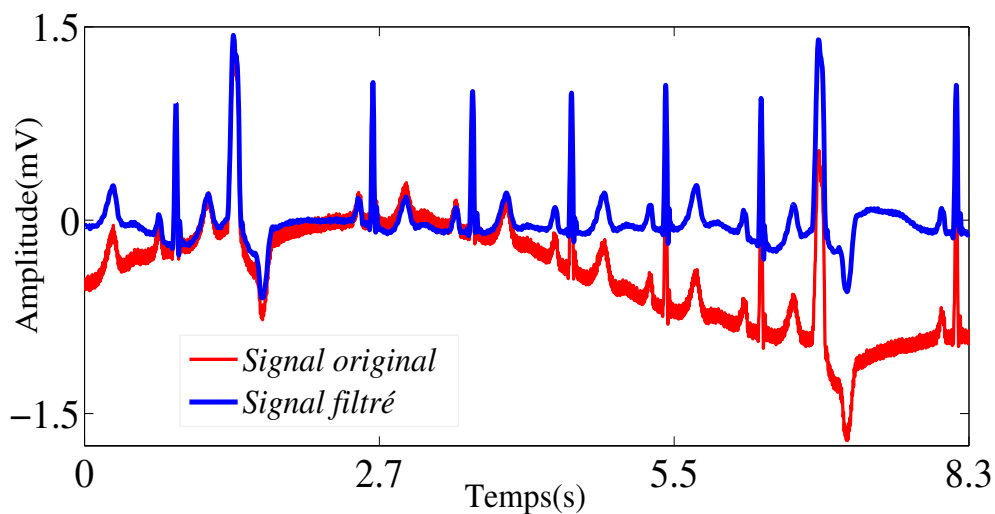


FIG. 3.9 – *Suppression de la ligne de base sur un segment de l'enregistrement MITDB : 203.*

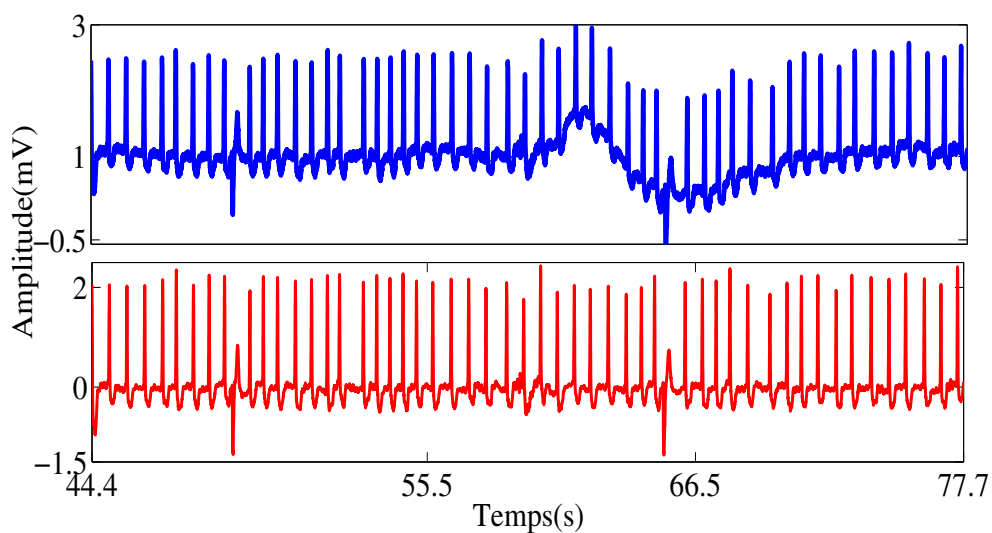


FIG. 3.10 – *Suppression de la ligne de base sur un segment de l'enregistrement MITDB : 210. Haut : signal original; Bas : signal filtré.*

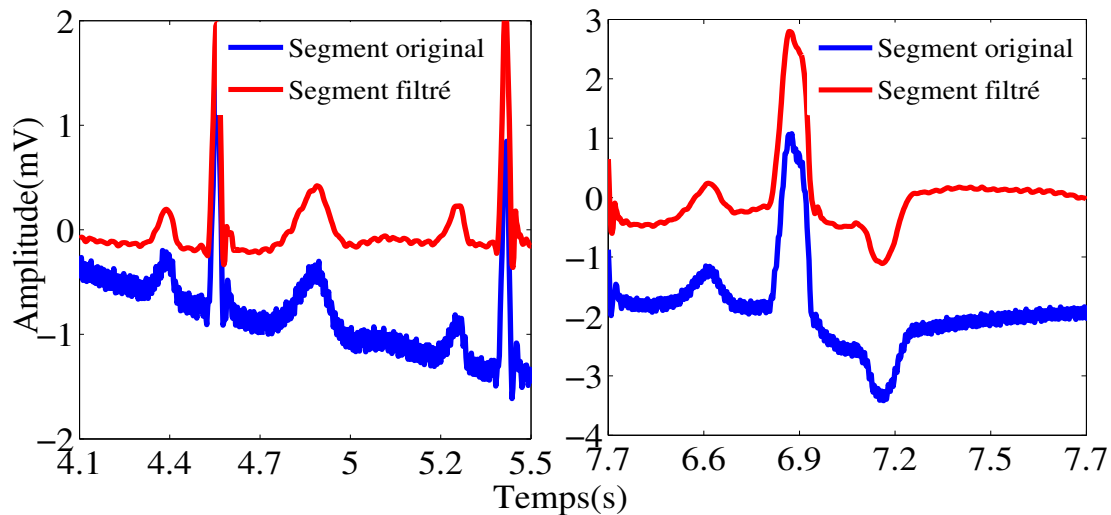


FIG. 3.11 – *Signal original et signal reconstruit avec filtrage et suppression de la dérive de la ligne de base. A gauche, le segment contient un battement systolique ; A droite, le segment contient un battement ectopique.*

3.4 Detection des complexes QRS

La détection des complexes QRS (localisation des battements) est la première étape de toute analyse automatique de l'ECG. À première vue, cette détection semblerait pouvoir être effectuée par un simple seuillage du signal (figure 3.12 .a), car les ondes R sont en général de plus grande amplitude que les autres. Mais ce n'est pas le cas chez tous les patients : parfois, l'onde T est d'amplitude comparable à celle de l'onde R, ce qui pourrait constituer une sérieuse cause d'erreur (voir figure 3.12 .b). La figure 3.13 montre que dans certains cas pathologiques comme chez le patient 203, l'onde R peut parfois avoir une très faible amplitude et une morphologie très variable d'un cycle cardiaque à un autre.

Une bonne détection des complexes QRS nécessite donc un traitement du signal très adéquat. Cette tâche est très difficile. Ceci est dû principalement à :

- La morphologie du complexe QRS qui change de manière significative d'un cas sain à un cas malade.
- Une onde T élevée peut être morphologiquement semblable à un complexe QRS, notamment après un battement PVC.
- Les divers types de bruit (provoqués par un stimulateur, une interférence de ligne de puissance, mouvement du patient,...) interfèrent dans la détection du QRS.
- La qualité et la largeur de la bande de fréquence du signal ECG enregistré peuvent changer de manière significative, celle-ci dépend fortement des conditions d'enregistrement (diagnostic ou surveillance) et des conditions d'enregistrement (pendant l'exercice, au repos, ...).

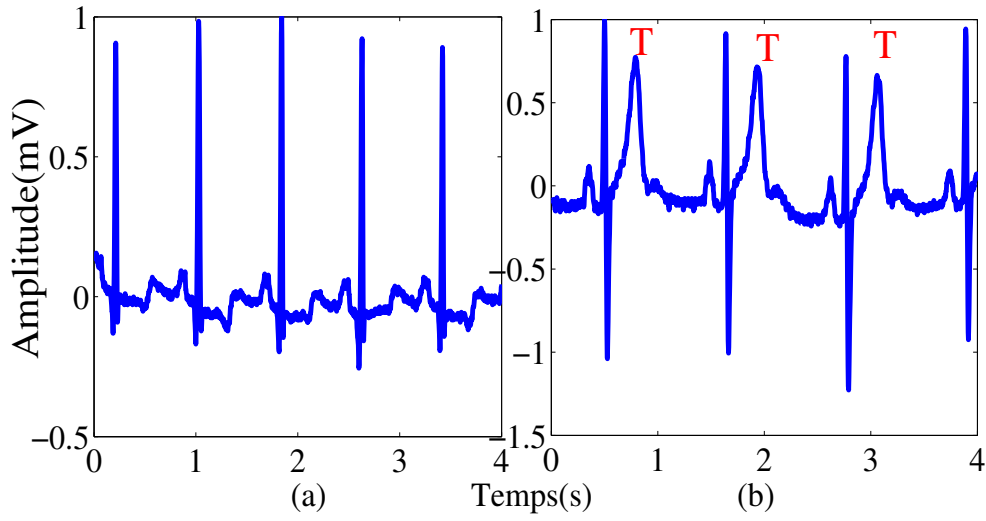


FIG. 3.12 – A gauche : Segment MITDB : 100 [0s, 4s]. A droite : Segment MITDB : 117 [0s, 4s] où l'onde T est aussi importante en amplitude que l'onde R. En revanche, quelque soit le patient, les ondes R varient plus rapidement que les autres.

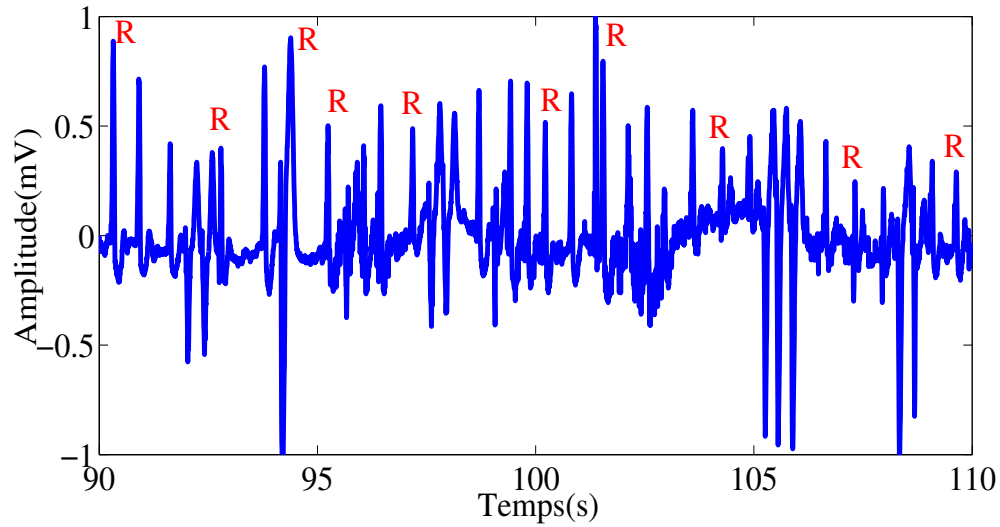


FIG. 3.13 – Segment MITDB : 203 [90s, 110s] présentant une variété d'ondes R avec diverses amplitudes et morphologies.

3.4.1 Méthodes existantes pour la détection des QRS

Depuis une trentaine d'années, la détection des complexes QRS a fait l'objet de nombreux travaux et continue à être un champ de recherche très actif. Le travail publié dans [Kohler 2002] récapitule les différentes méthodes par catégories. La comparaison de leurs performances y est également exposée. L'idée générale de toutes les méthodes est d'exploiter non seulement la grande dynamique des ondes R, mais aussi une propriété qui leur est propre : leurs variations rapides (leurs mobilités) et par conséquent, leurs contenus fréquentiels.

La majorité des méthodes utilisent le calcul de la dérivée, puis un seuillage [Pan 1985, Arzeno 2008]. De nombreuses variantes existent, faisant intervenir d'autres caractéristiques (dérivée seconde, amplitude). Ces multiples méthodes se différencient par leur plus ou moins grande sensibilité aux bruits et aux dérives de la ligne de base. D'autres méthodes plus récentes ont vu le jour depuis une dizaine d'années. Elles sont basées pour la plupart, sur l'utilisation des réseaux de neurones [Adnane 2010] ou des ondelettes [Lepage 2003, Ruchita 2010, Dinh 2002, Chen 2006].

3.4.2 Méthode de Pan et Tompkins

Le travail de Pan et Tompkins dans leur publication [Pan 1985], est focalisé essentiellement sur la détection des complexes QRS. Les auteurs ont alors proposé des filtres qui doivent préserver la densité spectrale de puissance des segments QS.

Le filtre P-Bas proposé est un Filtre à Réponse Impulsionnelle Infinie (RII) régi par l'équation aux différences suivante :

$$y(n) = 2y(n-1) - y(n-2) + x(n) - 2x(n-6) + x(n-12)$$

La fonction de transfert correspondante est alors donnée par :

$$H(z) = \frac{(1 - z^{-6})^2}{(1 - z^{-1})^2} \quad (3.28)$$

La fréquence de coupure et le gain linéaire de ce filtre sont respectivement $f_c = 18 \text{ Hz}$ et $G = 36 \text{ dB}$. La courbe de réponse d'un tel filtre est présentée en figure 3.14

L'ECG débruité est ensuite injecté dans un filtre dérivateur afin de mettre en évidence les fortes pentes qui distinguent les complexes QRS des composantes ECG de basses fréquences telles que les ondes P et T. Un tel filtre est géré par l'équation aux différences finies :

$$y(n) = \frac{1}{8}[-x(n-2) - 2x(n-1) + 2x(n+1) + x(n+2)]$$

La fonction de transfert d'un tel filtre est :

$$H(z) = \frac{1}{8}(-z^{-2} - 2z^{-1} + 2z^1 + z^2) \quad (3.29)$$

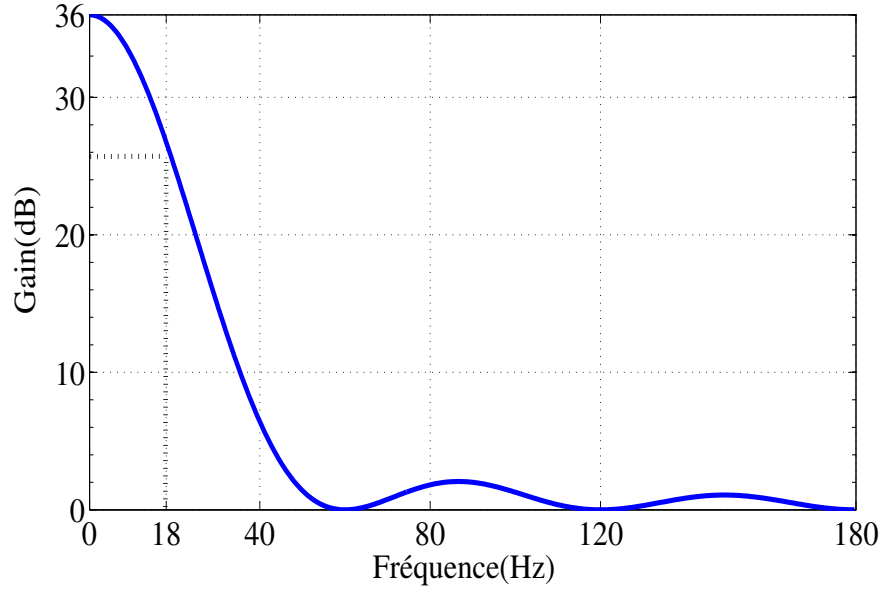


FIG. 3.14 – Réponse en fréquence du filtre P-Bas proposé par Pan et Tompkins.

sa réponse en amplitude est :

$$|H(f)| = \frac{1}{4}[\sin(4\pi f) + \sin(2\pi f)] \quad (3.30)$$

La figure 3.15 montre que la réponse en fréquence de ce filtre est linéaire entre 0 Hz et 30 Hz. Il approxime donc une dérivée idéale dans cette gamme de fréquence, gamme à laquelle appartiennent les complexes QRS.

L'opération suivante est la quadrature de la sortie $y(n)$ du dérivateur pour obtenir $z(n) = [y(n)]^2$ afin d'éviter le signe et de ressortir les valeurs les plus hautes qui correspondraient aux complexes QRS.

Le signal $z(n)$ est ensuite envoyé dans un filtre intégrateur à fenêtre glissante géré par l'équation aux différences suivante :

$$g(n) = \frac{1}{N}[z(n - (N - 1)) + z(n - (N - 2)) + \dots + z(n)] \quad (3.31)$$

La longueur N de la fenêtre a été choisie égale à 54 échantillons pour une fréquence d'échantillonnage de 360 Hz. Elle correspond au complexe QRS le plus large possible. La sortie de ce filtre permet en plus de la détection des pics R, la mesure de la durée des ondes QRS. Les signaux $x(n)$, $z(n)$ et $g(n)$ sont présentés sur la figure 3.16

L'algorithme de Pan et Tompkins a été amélioré par Fokapu et Girard [Fokapu 1993]. Le travail présenté dans cet article repose sur l'utilisation des dérivées première et seconde de l'électrocardiogramme $x(n)$. Les étapes de la

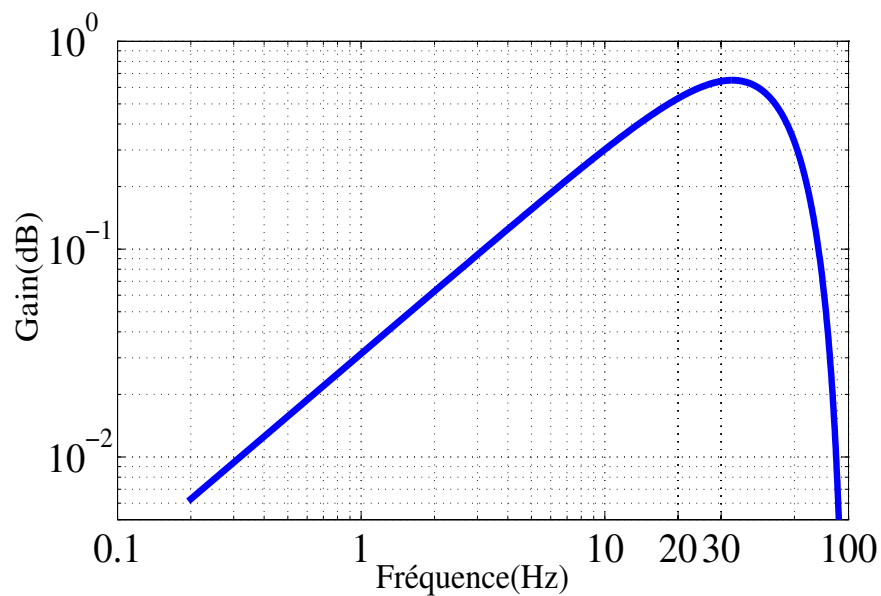


FIG. 3.15 – Réponse en fréquence du filtre dérivateur représentée à l'échelle logarithmique.

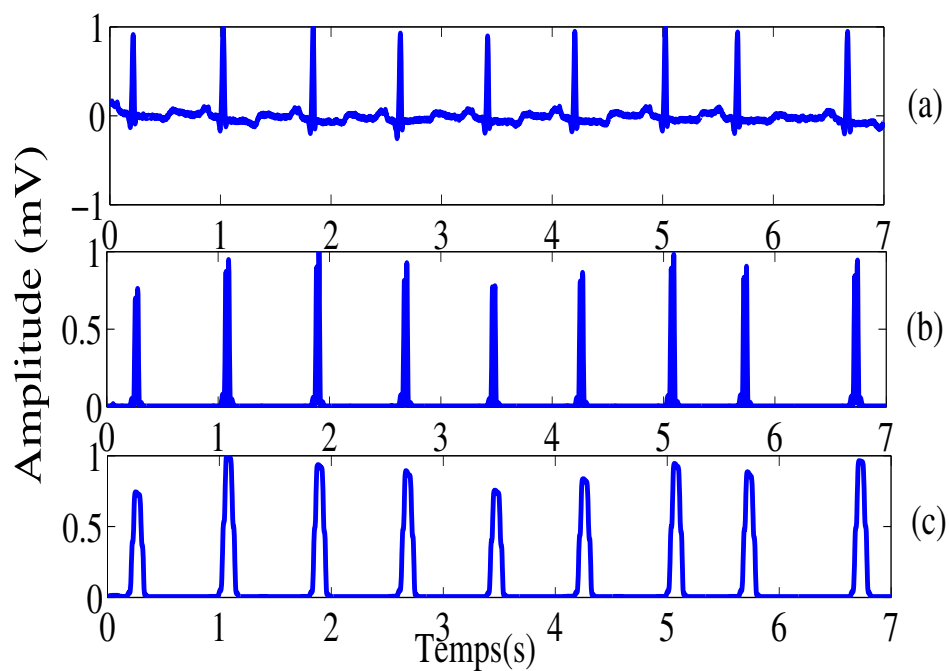


FIG. 3.16 – ECG filtré (a); sortie du filtre dérivateur (b) et sortie du filtre intégrateur à fenêtre glissante (c).

méthode sont :

Calcul des dérivées :

$$\begin{cases} \text{première :} & y_1(n) = x(n+1) - x(n-1) \\ \text{seconde :} & y_2(n) = x(n+2) - 2x(n) + x(n-2) \end{cases} \quad (3.32)$$

Lissage des dérivées par une moyenne mobile :

$$y_{liiss} = \frac{1}{4}[y(n+1) + 2y(n) + y(n-1)] \quad (3.33)$$

Normalisation :

$$y_a(n) = a_1 \cdot [y_{1liiss}(n)]^2 + a_2 [y_{2liiss}(n)]^2 \quad (3.34)$$

Où a_1 et a_2 sont des coefficients de normalisation tels que :

$$a_1 = \frac{1}{\max(y_{1liiss})} \quad \text{et} \quad a_2 = \frac{1}{\max(y_{2liiss})}$$

La localisation des pics R se fait grâce à un seuil originellement fixé à 30 % de la valeur maximale de y_a . Néanmoins, cette localisation doit être complétée par la recherche des maximums locaux car nous avons besoin de la position exacte du pic R pour définir *a priori* une fenêtre contenant l'intervalle QRS. La figure 3.17 montre un exemple de détection des pics R par cette méthode en utilisant un segment de l'enregistrement 100 et un seuil de 30%.

La méthode de Pan et Tompkins a fait ses preuves dans de nombreuses applications citées dans la littérature ouverte. Elle possède de bonnes caractéristiques comme la simplicité de sa mise en œuvre. Elle représente la méthode la plus citée dans la littérature. Néanmoins, cet algorithme basé sur les dérivées ne garantit pas la détection des pics R dans le cas des complexes QRS larges liés à certaines anomalies ventriculaires telles que les extrasystoles. Ces battements sont très atténués par les filtres dérivateurs car ils sont de basse fréquence. Leur détection nécessite toujours des seuils adaptatifs, ce qui consomme plus de temps. Pour illustration, nous avons appliqué la méthode à un segment de l'enregistrement 208 contenant des battements PVC représentés par des QRS larges. Le résultat indiqué par la figure 3.18 montre que la plupart des battements PVC ne sont pas détectés avec le seuil primaire de 30%.

Pour résoudre le problème des formes particulières de battements, plusieurs auteurs ont fait appel à la transformée en ondelettes et à l'analyse multirésolution afin de tenir compte de la différence entre les contenus fréquentiels des battements normaux et pathologiques. Dans [Ruchita 2010], les auteurs ont utilisé une analyse multirésolution avec des ondelettes dyadiques, d'autres [Dinh 2002] ont examiné le choix de l'ondelette mère et l'impact de ses propriétés de linéarité et de localisation

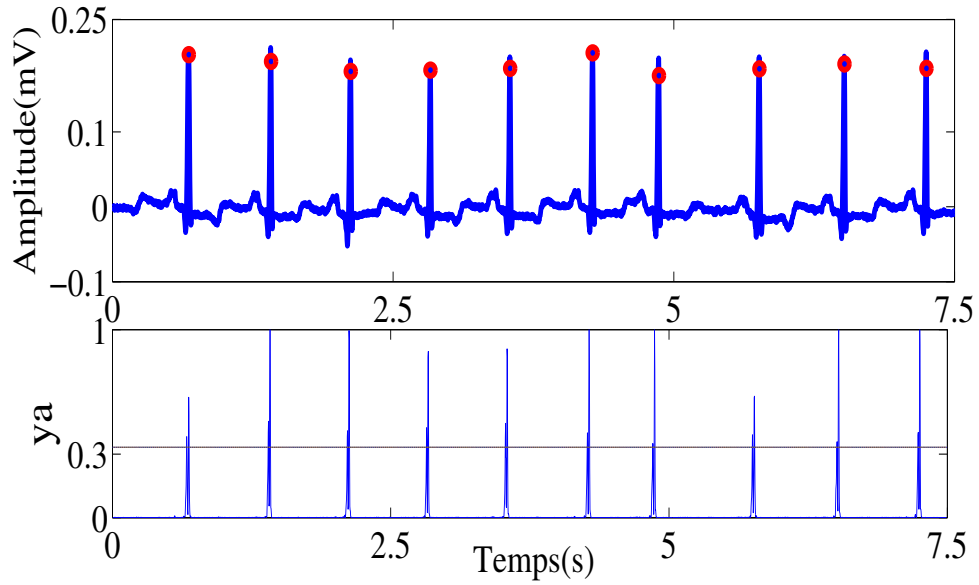


FIG. 3.17 – Détection des pics R par l'algorithme de Pan et Tompkins sur un segment de l'enregistrement MITBD : 100.

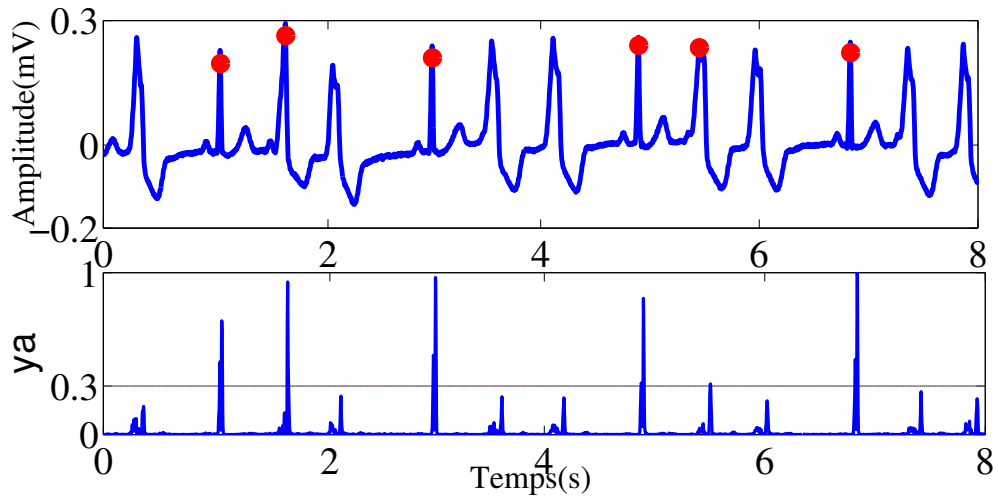


FIG. 3.18 – Segment de l'enregistrement MITDB : 208 [0s, 8s). Avec un seuil de 30 %, la plupart des battements PVC ne sont pas détectés par la méthode ; leurs dérivées sont nettement inférieures à celles des battements normaux.

fréquentielle sur le taux de bonne détection. Dans [Lepage 2003], l’auteur a combiné deux ondelettes différentes, celle de Haar (qui possède de bonnes propriétés de localisation temporelle), et celle de Coiflets (qui possède de bonnes propriétés de localisation fréquentielle). Il effectue ensuite le produit des coefficients d’ondelettes du niveau de basse résolution de Coiflet, qui présente peu de bruit, avec un niveau de haute résolution de Haar où le bruit est plus élevé. On réalise ainsi un lissage de la haute résolution dans le domaine des fréquences, ce qui conduit à diminuer le bruit présent sur le niveau de haute résolution.

3.4.3 Méthode proposée

La méthode que nous proposons [Zidelmal 2012a] est basée sur l’étude des paramètres spectraux des complexes QRS et d’une analyse multirésolution qui nous permet de considérer le contenu fréquentiel des battements normaux et des battements pathologiques.

3.4.3.1 Estimation de la DSP des complexes QRS

Pour analyser le contenu fréquentiel des complexes QRS, nous avons sélectionné quatre types de battements appartenant aux classes les plus dominantes de la base MITDB : les battements normaux (N), les extrasystoles Ventriculaires (V), les blocs de branches droites et gauches (RBBB et LBBB). Chaque classe est représentée par 200 battements sélectionnés dans la base MITDB. Les complexes QRS sont extraits grâce à une fenêtre de longueur 180 ms autour des positions annotées des ondes R (50 ms avant le pic R et 130 ms après le pic R). Chaque fenêtre contient alors 64 échantillons pour la fréquence d’échantillonnage 360 Hz.

Chaque segment contenant le complexe QRS est pondéré par la fenêtre de Blackman donnée par l’équation (3.35) afin de forcer le début et la fin du segment à zéro et supprimer ainsi les discontinuités dues à d’éventuelles adjacences des ondes P et T. La fenêtre de blackman est également choisie pour ses lobes secondaires très atténués dans le domaine fréquentiel (-74dB).

$$W(k) = 0.42 - 0.5\cos\left(\frac{2\pi k}{K-1}\right) + 0.08\cos\left(\frac{2\pi k}{K-1}\right); \quad 0 \leq k \leq K-1 \quad (3.35)$$

où K est le nombre d’échantillons du segment considéré.

La TFD connue comme outil de mesure du spectre d’amplitude et de phase des signaux déterministes, ne peut pas s’appliquer directement sur ces segments d’ECG considérés comme signaux aléatoires. La représentation fréquentielle de ces derniers fait appel à une description statistique appelée : fonction d’autocorrélation. Dans ce cas, l’analyse spectrale devient un problème d’estimation.

Estimation de la fonction d'autocorrélation : La fonction d'autocorrélation d'un signal aléatoire $x(k)$ stationnaire est définie par :

$$R_{xx}(n) \triangleq E[x(k)x(k+n)] = E[x(k-n)x(k)] = R_{xx}(-n) \quad (3.36)$$

La grandeur R_{xx} peut être approchée en utilisant des estimateurs. Notons que chaque estimateur d'une grandeur donnée, a , est caractérisé par :

- son biais : $B_{\hat{a}} = E[\hat{a}] - a$
- sa variance : $V_{\hat{a}} = E[(\hat{a} - E[\hat{a}])^2]$

1. Estimateur non biaisé : si on dispose de K valeurs de $x(k)$, alors pour chaque n , la valeur moyenne de $x(k)x(k+n) = u_n(k)$ ne peut être estimée qu'à partir de $K - n$ valeurs $u_n(k)$, on obtient alors :

$$\hat{R}_{xx}(n) = \frac{1}{K - |n|} \sum_{k=0}^{K-|n|-1} x(k)x(k+n) \quad (3.37)$$

Le biais $B_{\hat{R}_{xx}}$ de cet estimateur peut être calculé aisément. On a :

$$E[\hat{R}_{xx}(n)] = \frac{1}{K - |n|} \sum_{k=0}^{K-|n|-1} E[x(k)x(k+n)] \quad (3.38)$$

et

$$\frac{1}{K - |n|} \sum_{k=0}^{K-|n|-1} R_{xx}(n) = R_{xx}(n) \quad (3.39)$$

$B_{\hat{R}_{xx}} = E[\hat{R}_{xx}(n)] - R_{xx}(n) = 0$, l'estimateur est donc non biaisé. Cet estimateur est consistant mais présente une variance importante lorsque la moyenne (3.38) est calculée sur peu de termes (n proche de K).

2. Estimateur biaisé : Un autre estimateur très souvent utilisé est donné par :

$$\tilde{R}_{xx}(n) = \frac{1}{K} \sum_{k=0}^{K-|n|-1} x(k)x(k+n) \quad (3.40)$$

Cet estimateur diffère de l'estimateur non biaisé par un facteur multiplicatif.

$$\tilde{R}_{xx}(n) = \frac{K - |n|}{K} \hat{R}_{xx}(n)$$

L'estimateur biaisé possède une erreur systématique mais sa variance est plus faible que celle de l'estimateur non biaisé quand n est proche de K .

Plus de détails sur la théorie de l'estimation peuvent être trouvés dans [Kunt 1999].

Théorème de Wiener-Kinchine : Ce théorème stipule que la densité spectrale de puissance d'un signal aléatoire $x(t)$ est estimée par celle de sa fonction d'autocorrélation :

$$\hat{P}_x(f) = \int_{-\infty}^{+\infty} C_{xx}(\tau) e^{-j2\pi f\tau} d\tau \quad (3.41)$$

Dans le cas d'un signal discret $x(k)$ de longueur K , l'estimateur simple de sa DSP, appelé aussi périodogramme simple est donné par :

$$\hat{P}_x(f) = \sum_{-K+1}^{K-1} \hat{R}_{xx}(n) e^{-j2\pi fn} \quad (3.42)$$

La Densité Spectrale de Puissance d'un complexe QRS représenté par un segment $s(k)$ a été donc évaluée par un périodogramme adouci en utilisant sa fonction d'autocorrélation estimée $\hat{R}_{ss}(n)$. Afin d'avoir une bonne précision fréquentielle, la (*FFT*) a été réalisée sur 512 points en introduisant le *zero-padding*. Les DSP relatives aux quatre types de complexes QRS considérés sont illustrées sur la figure 3.19.

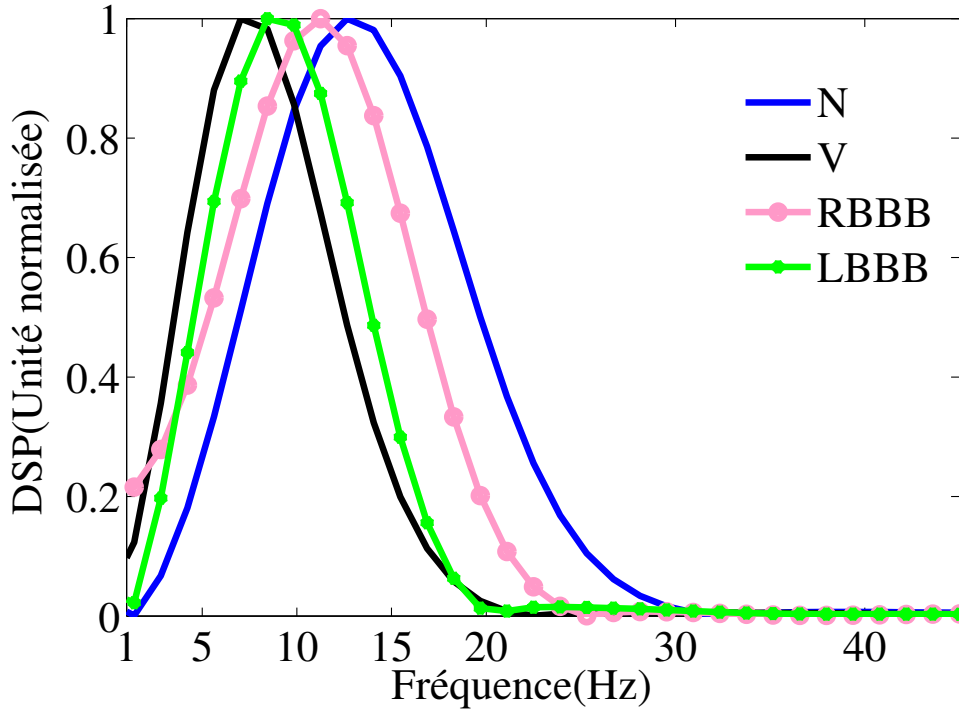


FIG. 3.19 – Densité Spectrale de Puissance des quatre types de complexes QRS.

TAB. 3.1 – Correspondance entre détails et bandes de fréquences

Détail d_n	Bande de fréquence
d_n^1	[90 Hz, 180 Hz]
d_n^2	[45 Hz, 90 Hz]
d_n^3	[22.5 Hz, 45 Hz]
d_n^4	[11.25 Hz, 22.5 Hz]
d_n^5	[5.62 Hz, 11.25 Hz]

3.4.3.2 Décomposition des QRS en bandes de fréquences

L'analyse multirésolution décompose le signal en différentes bandes de fréquences. Ainsi les hautes fréquences et les basses fréquences pourront être analysées séparément. Les niveaux de décomposition sont fixés en fonction de la fréquence d'échantillonnage et de la bande de fréquence recherchée. Lors de cette décomposition, le choix de l'ondelette analysante est très important, cependant, il n'existe pas de règle générale permettant la sélection d'une ondelette mère. Ce choix dépend de la nature du signal à traiter ainsi que de l'application envisagée. Dans le cas de la présente application, l'ondelette de Haar donne de meilleurs résultats car elle est compacte et offre une meilleure localisation temporelle.

Le signal ECG est décomposé jusqu'au niveau 5. Nous avons choisi ce niveau en nous basant d'une part, sur les densités spectrales de puissance des complexes QRS données en figure 3.19 et d'autre part, sur la fréquence d'échantillonnage qui est de 360 Hz. En tenant compte du théorème de Shannon, la bande de fréquence du signal est [0 Hz, 180 Hz] où 180 Hz représente la fréquence de Nyquist. La correspondance entre les coefficients de détail obtenus et les bandes de fréquences est représentée sur la table 3.1.

3.4.3.3 Sélection des bandes de fréquences

Le résultat donné en figure 3.19 montre que la puissance des battements normaux et celle des battements pathologiques considérés sont concentrées entre 5Hz et 22Hz. Par ailleurs, les travaux de Thakor et al. sur la densité spectrale de puissance du signal ECG et celle des complexes QRS ont montré que la puissance d'un segment ECG est principalement concentrée dans les ondes QRS qu'il contient [Thakor 1984].

Pour examiner et illustrer les niveaux de puissance des battements cardiaques sur différents niveaux de résolution, nous avons sélectionné deux segments de la base d'étude : un segment contenant deux battements normaux (plus grande fréquence) pris de l'enregistrement MITDB : 100 et un segment contenant deux

battement PVC (plus basse fréquence) pris de l'enregistrement MITDB : 208. Ces deux segments sont décomposés en 5 niveaux de résolution. La densité spectrale de puissance correspondant à chaque niveau est représentée en figure 3.20. Cette

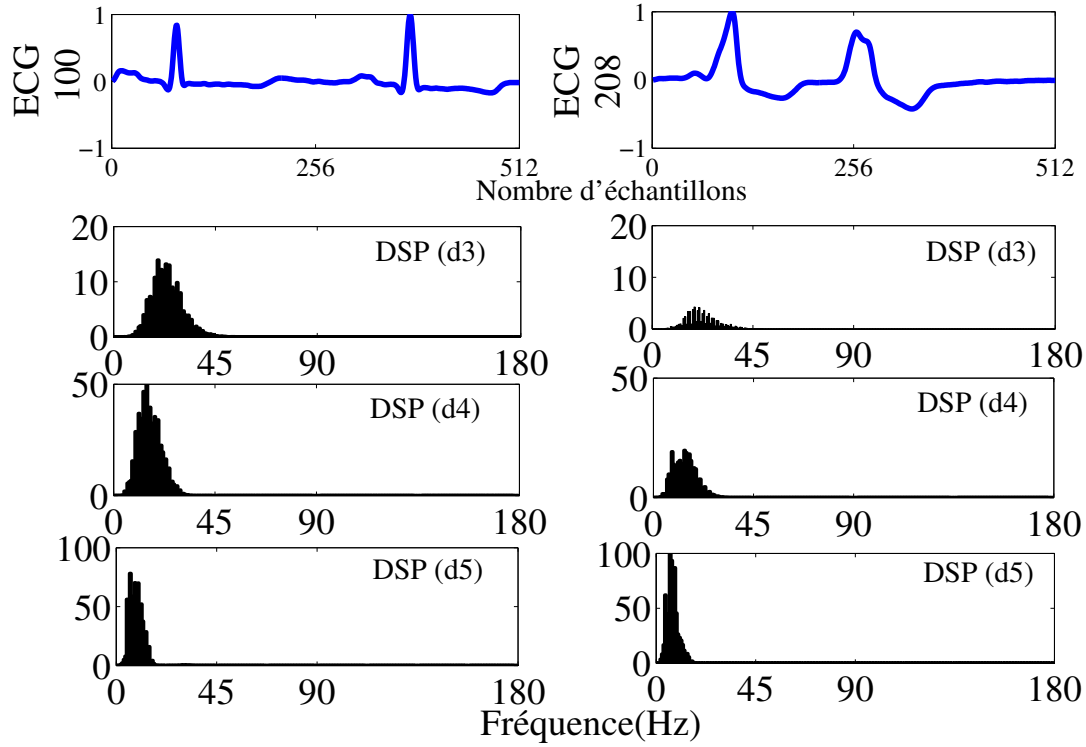


FIG. 3.20 – Densité spectrale de puissance des battements normaux (figure de gauche) et battements PVC (figure de droite). Pour une meilleure comparaison, la même échelle est utilisée à chaque niveau. Les détails d_n^1 et d_n^2 ne sont pas considérés car ils sont en dehors de la bande d'intérêt.

figure montre que les battements normaux sont plus significatifs sur le niveau de résolution $j = 4$ tandis que les battements PVC sont plus significatifs sur le niveau $j = 5$. Les coefficients de détail d_n^4 et d_n^5 sont donc sélectionnés pour mettre au point notre algorithme de détection. Leur produit est utilisé pour localiser les complexes QRS. La figure 3.21 montre que sur le signal h (produit des deux vecteurs d_n^4 et d_n^5) choisi comme signal de localisation, les ondulations de la ligne de base ainsi que l'effet des ondes P et T sont supprimés.

Sur le signal ECG, nous constatons l'apparition de deux battements PVC (l'un à 1640 s et l'autre à 1657 s). Ces deux battements sont parfaitement localisés malgré qu'ils sont suivis chacun d'une onde T très importante.

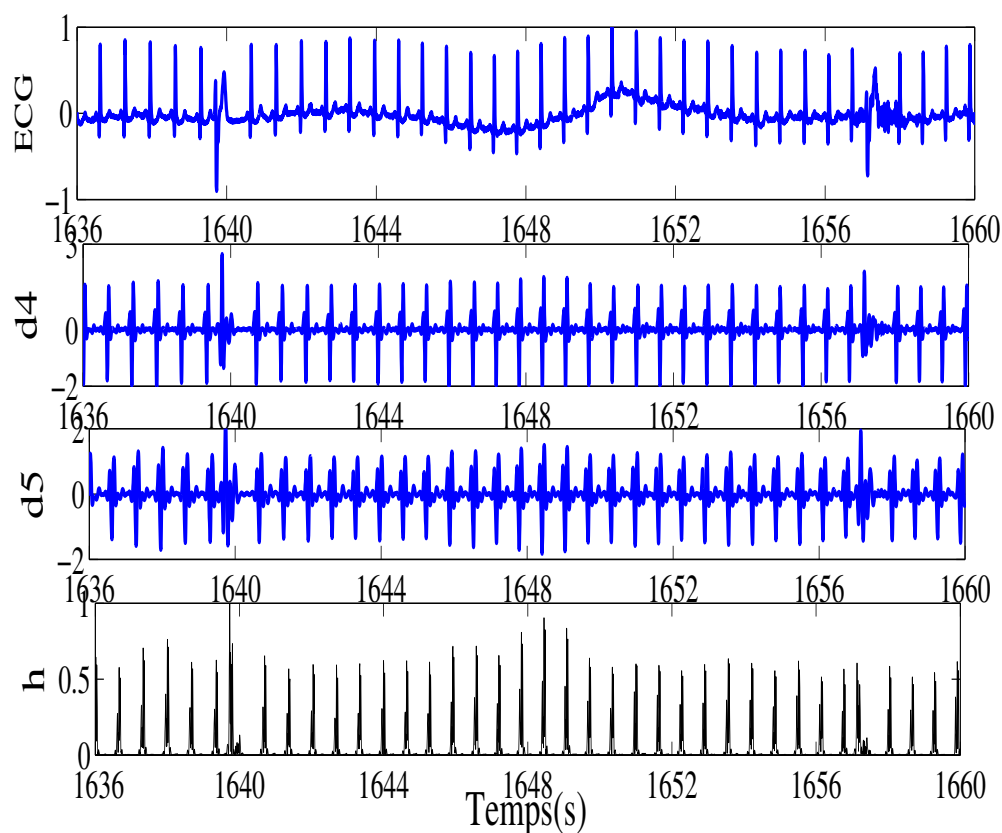


FIG. 3.21 – Décomposition d'un segment du signal AHA : 0201 [1636 s, 1660 s] en utilisant l'ondelette de Haar. L'effet des ondes P et T appartenant à la bande de fréquence [0 Hz, 5 Hz] est nettement atténué.

3.4.3.4 Algorithme de détection

1. Décomposer le signal ECG $f(t)$ jusqu'à 5 niveaux de résolution .

2. Calculer $h = \left| \prod_{j=4}^5 (d_n^j) \right|$

3. Localisation des QRS :

$$\begin{cases} \text{si } h(n) \geq 0,3 \cdot \max(h) \text{ alors } n \Rightarrow QRS \text{ candidat} \\ \text{sinon } n \not\Rightarrow QRS. \end{cases}$$

4. soient n et n' deux positions consécutives sélectionnées, alors :

$$\begin{cases} \text{si } |n - n'| < 36 \text{ alors } n \text{ et } n' \Rightarrow \text{même QRS} \\ \text{sinon } n \text{ et } n' \not\Rightarrow \text{même QRS}. \end{cases}$$

(36 $f_e = 100$ ms est la durée standard d'un complexe QRS)

5. Une détection multiple dans un intervalle de 200 ms doit être supprimée. Cette contrainte a un sens physiologique : la période réfractaire est nettement supérieure à 200 ms.

6. Recherche des battements omis : Si aucun battement n'est détecté dans une période égale à 1.5 fois l'intervalle RR courant, le seuil de détection est divisé par 2 pour rechercher un éventuel QRS omis. Cette période de recherche déjà utilisée dans [Hamilton 1986] a un sens physiologique : un inter-battement ne change pas aussi vite.

3.4.3.5 Comparaison avec l'algorithme de Tompkins

Pour situer notre algorithme par rapport à ceux publiés dans la littérature, nous l'avons comparé à l'algorithme le plus connu et le plus cité dans la littérature, l'algorithme de Tompkins [Pan 1985]. Les deux algorithmes sont appliqués à un segment de l'enregistrement MITDB : 208 présentant des battements normaux et des battements PVC (les plus larges de la base d'étude).

La figure 3.23 montre qu'avec la stratégie utilisée dans [Pan 1985], la dérivée du signal amplifie les hautes fréquences (pentes raides), caractérisant les battements N et atténue les basses fréquences (pentes lentes), caractérisant les battements PVC. La plupart de ces derniers ne sont pas détectés. Leur détection nécessiterait alors un seuillage adaptatif. La figure 3.22 montre que notre algorithme, tenant compte du contenu fréquentiel de chaque type de battement détecte toutes les ondes R en utilisant uniquement le seuil primaire $0.3\max(h)$.

Les résultats de détection de certaines formes particulières de battements sévèrement bruités seront exposés plus loin.

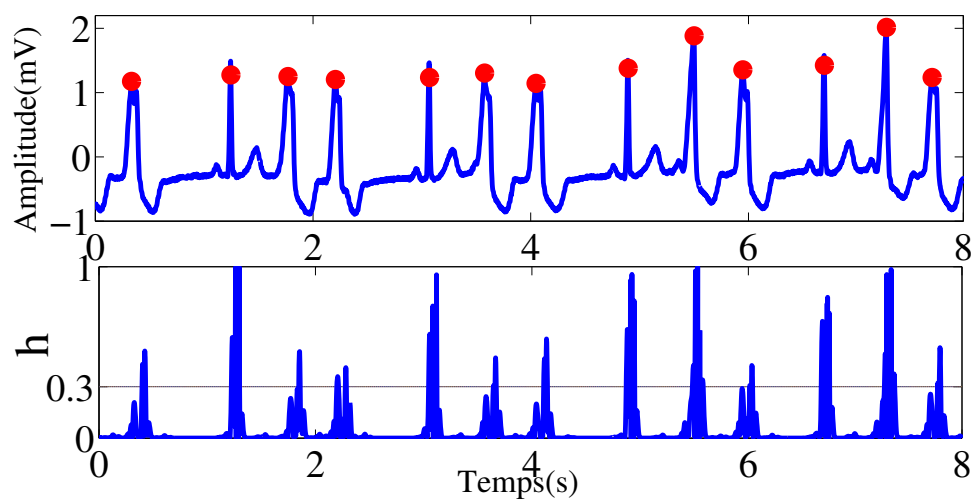


FIG. 3.22 – Détection des complexes QRS par notre algorithme. Toutes les ondes R sont détectées en utilisant le seuil primaire.

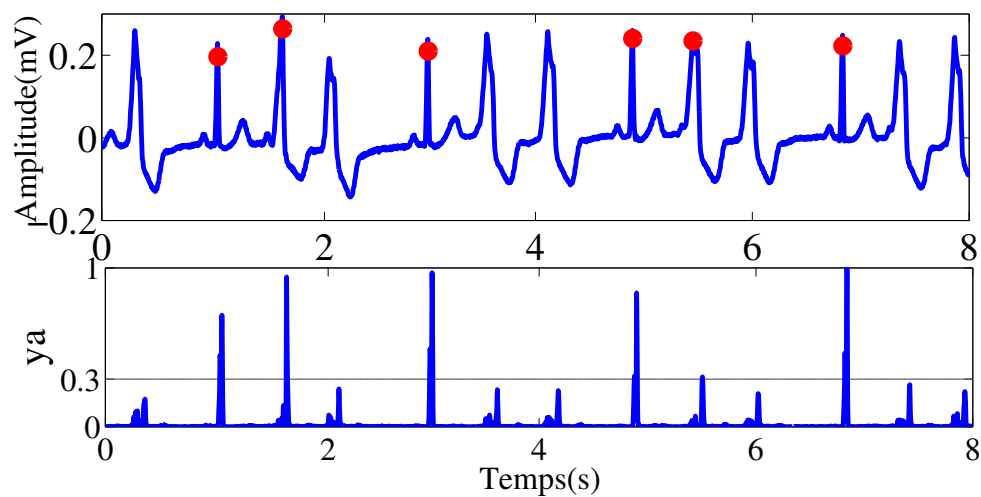


FIG. 3.23 – Détection des complexes QRS par l'algorithme Tompkins (seuil primaire).

3.5 Segmentation des battements

Après la détection des pics R, chaque battement est segmenté puis rangé dans un vecteur sur lequel sera effectué un calcul de caractéristiques. Ces vecteurs doivent impérativement avoir la même dimension pour que les caractéristiques des battements correspondants soient calculées dans les mêmes conditions.

La représentation temporelle de chaque segment contenant le battement est obtenue selon la procédure développée dans [Moody 1989]. Dans le cas d'un ECG échantillonné à 360 Hz, ce segment est représenté par 200 ms soit, environ 60ms avant le sommet R et 140 ms après. Chaque battement est donc pondéré avec une fenêtre rectangulaire de 72 échantillons. La figure 3.24 montre quelques complexes QRS segmentés.

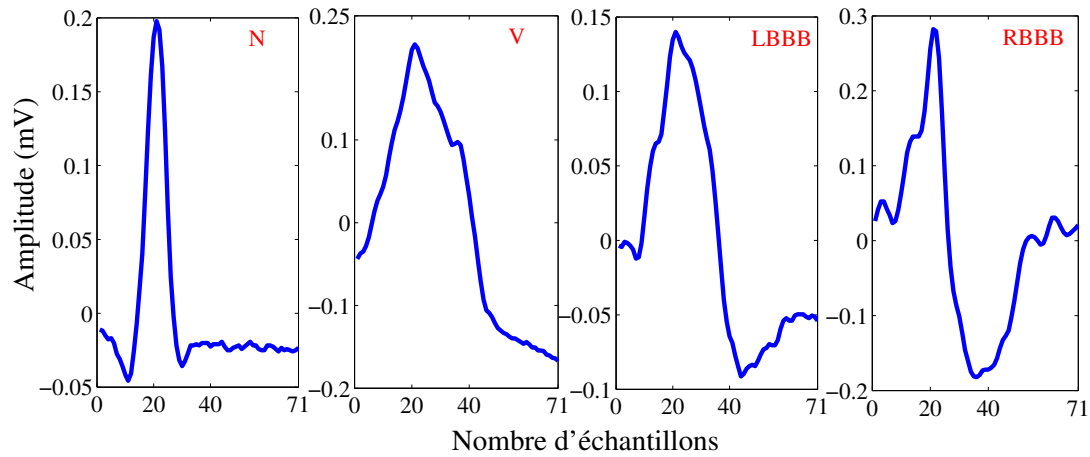


FIG. 3.24 – Exemples de battements segmentés sur 72 échantillons.

3.6 Conclusion

Dans ce chapitre, nous avons présenté le prétraitement du signal ECG suivi de la détection des complexes QRS qui nous permet de segmenter les battements. L'algorithme de détection proposé est testé en utilisant les 48 enregistrements de la base de données MITDB et quelques enregistrements de la base de données AHA. Cet algorithme permet la détection de l'onde R des battements normaux et d'une grande variété de battements pathologiques malgré qu'ils sont parfois noyés dans différents types de bruits. Les bandes de fréquences utilisées dans l'algorithme permettent de prendre en considération tous les types d'ondes QRS en atténuant les bruits et l'effet des ondes P et T qui sont parfois du même ordre de grandeur que l'onde R. Les résultats auxquels nous avons abouti seront présentés plus loin.

Extraction des paramètres discriminants

4.1 Introduction

Dans ce chapitre, nous abordons le problème de la reconnaissance automatique de battement cardiaques. Nous nous intéressons dans cette étude aux quatre classes de battements les plus dominantes de la base d'étude : la classe N représentant les battements normaux (systoliques) ; la classe V représentant les battements ectopics (ESV) ; la classe RBBB représentant les blocs de branche droite et la classe LBBB représentant les blocs de branche gauche. Après la description des classes, nous allons choisir et extraire les paramètres les plus discriminants. Ces paramètres doivent être ceux habituellement utilisés par les cardiologues.

4.2 Description des classes considérées

L'objectif de la présente étude est d'évaluer les performances d'un classifieur binaire avec rejet d'ambiguïtés dans le cas d'une aide au diagnostic médical. Nous suivons alors les recommandations pratiques AAMI [Mark 1988] pour former deux classes de battement en respectant les annotations de "American Heart Association (AHA)" que nous trouvons aussi dans [Mark 1988] .

4.2.1 La classe V

La classe positive "V" représente les battements ectopics. Elle comprend les Contractions Ventriculaires Prématuurées (PVC ou ESV), et les battements inconnus (minoritaires dans la base).

Un battement ectopic est un battement anormal suffisamment répandu pour être pris en considération et pour estimer son occurrence sur un enregistrement de longue durée. Les ESV s'observent sur quasiment tous les enregistrements, principalement en période de récupération après un effort. Bien que leur présence n'indique aucune pathologie particulière, si, de façon récurrente, leur nombre par minute est supérieur à 6, les extrasystoles peuvent être un signe précurseur d'une tachycardie ventriculaire qui, elle, constitue une pathologie majeure.

Contrairement aux battements normaux qui ont pour origine la dépolarisation des cellules sinusales (voir section 2.6), l'ESV naît de la dépolarisation spontanée

d'un petit groupe de cellules ventriculaires, appelé alors foyer ectopique ventriculaire. L'impulsion électrique créée n'emprunte pas la voie normale de conduction (faisceau de His). Elle se propage plus lentement dans les ventricules. La contraction ventriculaire est ainsi étalée dans le temps et perd de son efficacité.

Le tracé d'un battement ESV est caractérisé par certaines propriétés dont il faut tenir compte pour le détecter :

- L'onde R n'est pas précédée d'une onde P, puisqu'il n'y a pas eu d'activité auriculaire préalable.
- La durée du complexe QRS est supérieure à la durée d'un complexe QRS normal car l'impulsion électrique n'emprunte pas la voie normale de conduction et se propage donc lentement dans les ventricules.
- Le complexe QRS présente une forme bizarre (atypique).
- Ces battements sont prématurés.

4.2.2 La classe N

La classe négative "N" représente la classe normale. Elle comprend tous les battements systoliques normaux (environ 70 % de la base MITDB) et les battements réguliers provenant des blocs de branches droites et gauches notés respectivement LBBB et RBBB. Ces battements sont aussi très représentatifs dans la base d'étude. Les différents battements considérés sont illustrés par la figure 4.1.

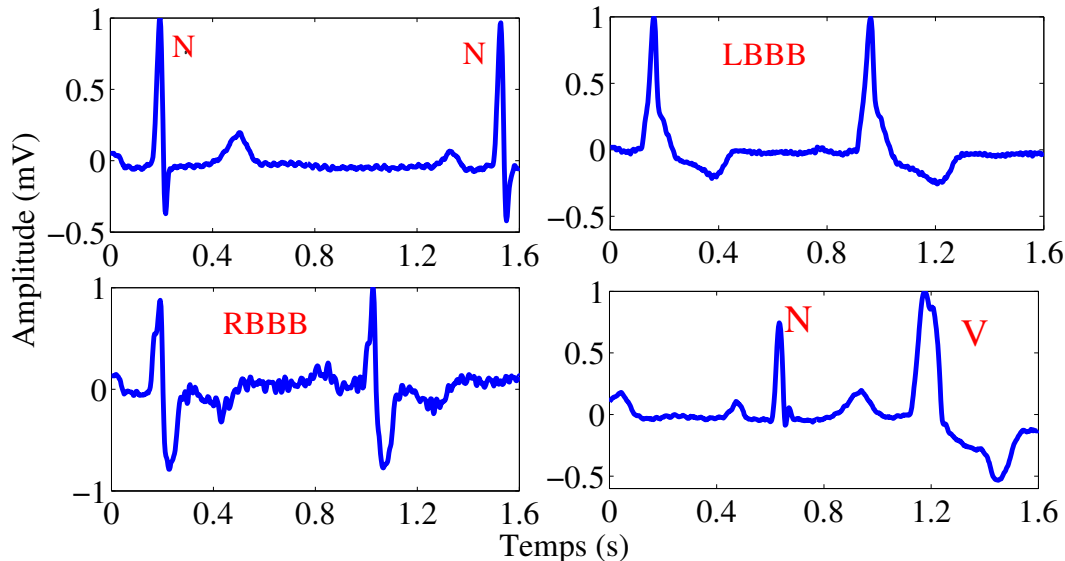


FIG. 4.1 – Différentes formes de battements cardiaques considérés.

4.3 Calcul des paramètres discriminants

Les paramètres retenus sont similaires à ceux habituellement utilisés par les cardiologues :

4.3.1 Mesure de régularité

L'absence de régularité des battements cardiaques est une caractéristique d'arythmie très importante pour le diagnostic. Elle est souvent associée à un trouble de la production ou de la conduction de l'impulsion électrique (foyers ectopiques, blocs, boucles, etc). Les arythmies permanentes ou sporadiques nécessitent un suivi médical et sont des indications typiques à la pose régulière d'un enregistreur Holter.

La mesure de la régularité des battements est souvent effectuée par la mesure des intervalles RR . On appelle intervalle RR l'écart temporel entre le pic R du battement considéré et le pic R du battement précédent (voir figure 4.2). L'étude

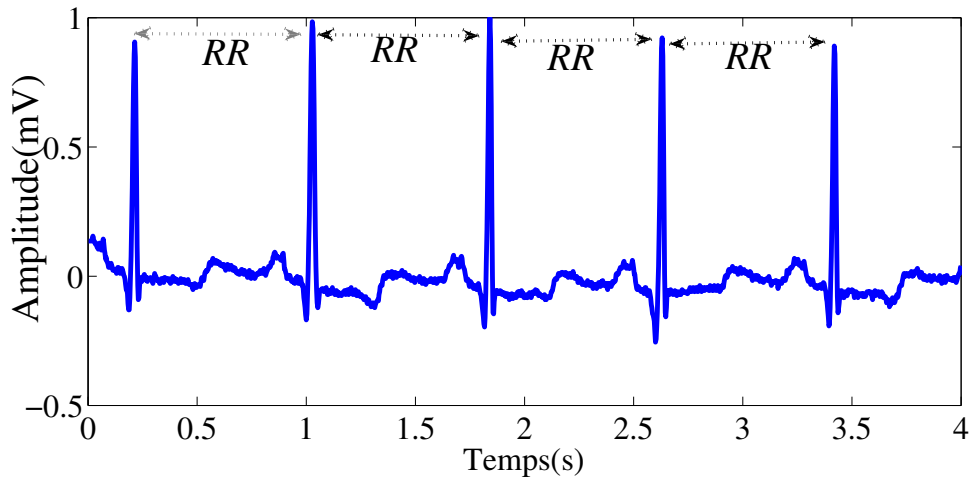


FIG. 4.2 – Intervalles temporels entre des pics R successifs dans le cas d'un rythme normal (segment de l'enregistrement MITDB : 100).

de différentes bases de données (MITDB, AHA) nous a permis de mettre en évidence le fait que la distance RR était un élément très utile pour la caractérisation d'une famille de battements. Nous avons constaté que les battements normaux et les blocs de branches sont réguliers tandis que les battements extrasystoliques surviennent de manière précoce et irrégulière (voir figure 4.3).

Pour quantifier la régularité des battements, nous avons procédé comme suit : Soit n l'indice du battement R considéré, l'intervalle RR correspondant est donné par : $RR_n = R_n - R_{n-1}$. Nous avons ensuite normalisé cette grandeur afin de la rendre indépendante du rythme cardiaque. L'intervalle RR normalisé est défini comme suit :

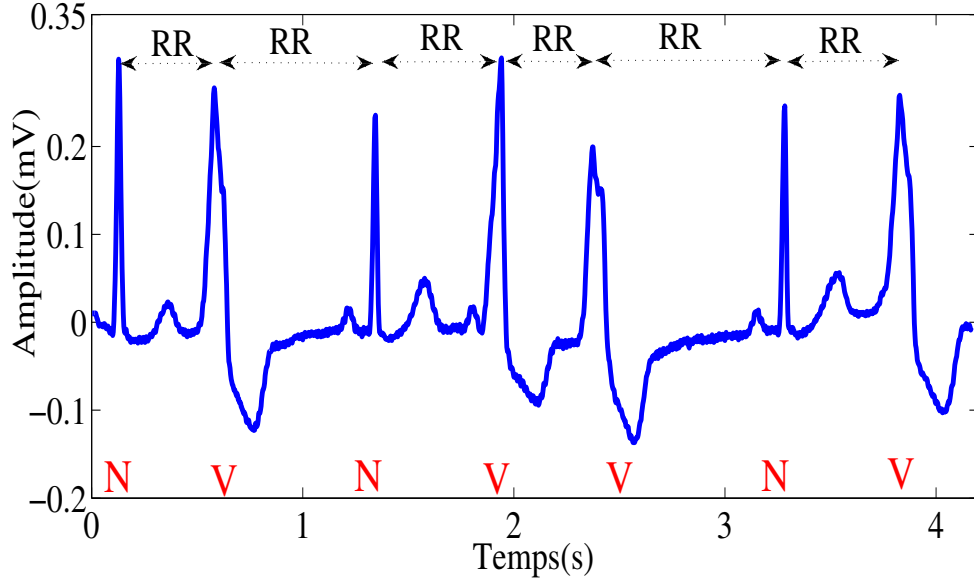


FIG. 4.3 – Intervalles temporels entre des pics R successifs dans le cas de présence de battements ectopics (segment de l'enregistrement MITDB : 208).

$$RR_{norm}(n) = \frac{RR_n - E[RR]}{\max(RR)} \quad (4.1)$$

où $E[RR]$ est la moyenne des intervalles RR du même enregistrement. La figure 4.4 montre la régularité des battements cardiaques dans le cas d'un électrocardiogramme normal (enregistrement 100). Sur la même figure, nous présentons l'histogramme normalisé des intervalles RR indiquant un très faible écart type (écart par rapport à la moyenne).

La régularité des quatre types de battements considérés est récapitulée sur la figure 4.5.

4.3.2 Morphologie du battement

Les battements ectopics ont une morphologie nettement différente de celle des battements systoliques. Elle constitue un bon paramètre discriminant. Ce paramètre est retenu par la majorité des auteurs ayant travaillé sur la reconnaissance automatique de battements cardiaques [Moody 1989, Gomez 2006, Krasteva 2007, De-Chazal 2006, Patra 2010]. La morphologie d'un battement peut être quantifiée de plusieurs manières :

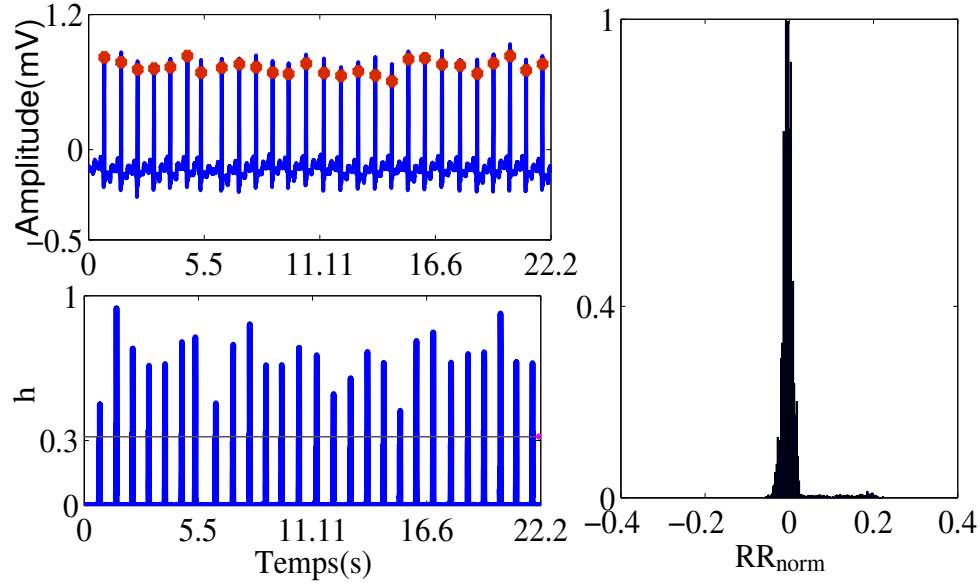


FIG. 4.4 – *Intervalles RR réguliers sur l'enregistrement MITDB : 100 normal (Gauche) ; Histogramme des intervalles RR (Droite).*

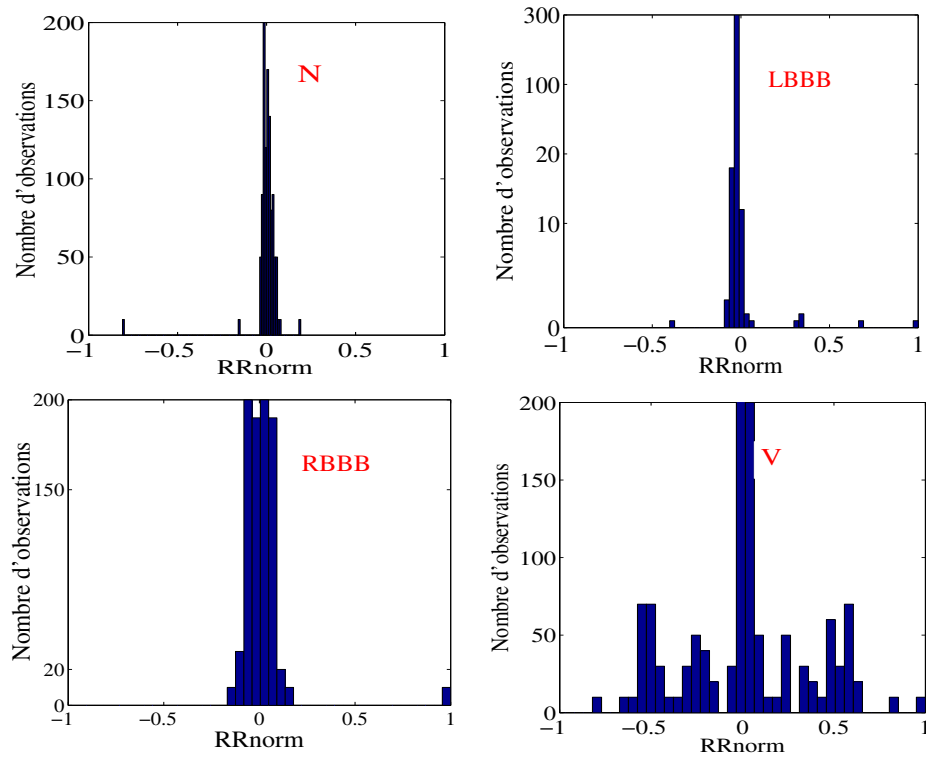


FIG. 4.5 – *Histogrammes RR pour les quatre types de battements considérés*

4.3.2.1 Méthodes existantes

Par soucis d'avoir des vecteurs de caractéristiques de dimension importante ainsi qu'une redondance d'informations, chaque auteur a utilisé une méthode appropriée permettant d'optimiser la quantification et la représentation du segment contenant le battement à caractériser.

Certains auteurs ont utilisé l'intercorrélation. La méthode consiste à comparer point par point chaque battement de l'enregistrement considéré à un ensemble de battements de référence (battements normaux et battements ectopics) choisis dans la base d'apprentissage. Les coefficients d'intercorrélation obtenus sont comparés à des seuils déterminés de manière empirique [Krasteva 2007]. D'autres auteurs ont procédé par sous échantillonnage des battements pour réduire la taille du vecteur [De-Chazal 2006]. Des outils de compression plus performants sont également utilisés pour la même application : la Transformée de Karhunen-Loève (KLT) est originalement utilisée par G.Moody en 1989 [Moody 1989] puis reprise en 2006 par dans [Gomez 2006].

Des méthodes statistiques sont également utilisées. On cite : l'Analyse en Composantes Indépendante (ICA) [Zhao 2008, Sung-Nien 2007] et l'Analyse en Composantes Principales (PCA) [Patra 2010]. Toujours, par soucis de quantifier le battement avec un minimum de paramètres, des coefficients d'ondelettes ont été utilisés dans [Thakor 1993], et les coefficients d'un prédicteur linéaire dans [Wang 1994]. Pour la présente application, nous avons retenu la prédiction linéaire afin de modéliser les complexes QRS car elle est efficace, elle opère dans le domaine temporel et engendre moins de paramètres que les autres méthodes.

4.3.2.2 Modélisation des battements par Prédiction Linéaire

La prédiction linéaire est une méthode de modélisation AutoRegressive (AR). Cette technique est utilisée en vue de compresser un signal. Elle remplace alors ce dernier par les paramètres de prédiction.

L'idée de base de la méthode est que l'échantillon d'un signal peut être obtenu par une combinaison linéaire des échantillons passés de ce signal en minimisant l'erreur de prédiction au sens des moindres carrés. L'échantillon prédit est calculé par la relation :

$$\tilde{y}(n) = \sum_{k=1}^p a_k y(n-k) \quad (4.2)$$

où p est l'ordre du prédicteur, a_k sont les coefficients du prédicteur. L'erreur de prédiction appelée aussi innovation est donnée par :

$$e(n) = \tilde{y}(n) - y(n) \quad (4.3)$$

et l'erreur quadratique moyenne par :

$$MSE = E[e^2(n)] \quad (4.4)$$

L'algorithme de calcul des coefficients a_k consiste à trouver les valeurs qui minimisent la puissance dans $e(n)$.

Sélection de l'ordre du prédicteur : La sélection de l'ordre du modèle se fait généralement par une analyse de la variance de l'erreur de prédiction. Nous avons considéré l'estimation cette variance en fonction de l'ordre p du prédicteur pris entre 1 et 20. Nous constatons sur la figure 4.6 que cette variance diminue jusqu'à $p = 4$, valeur à partir de laquelle elle ne décroît plus de façon significative. Avec

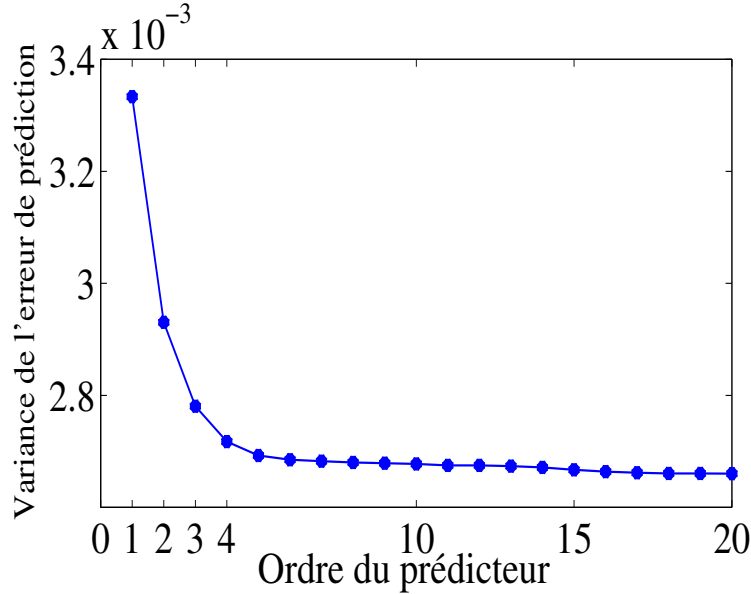


FIG. 4.6 – *Variance de l'erreur de prédiction en fonction de l'ordre p du modèle.*

un prédicteur d'ordre 4, nous retenons trois paramètres a_1, a_2 et a_3 ; le coefficient a_0 étant toujours égal à 1.

La figure 4.7 montre que les segments prédits sont presque identiques aux segments originaux quand la prédiction linéaire d'ordre 4 est appliquée à un battement normal et à un battement ectopie pondérés chacun avec une fenêtre rectangulaire de 72 échantillons (voir section 3.5).

Nous constatons sur les figures précédentes que l'erreur de prédiction est un bruit blanc de très faible écart type (respectivement $\sigma_1 = 0.0153$ et $\sigma_2 = 0.0195$).

4.3.3 Largeur du battement

L'une des principales caractéristiques des battements ectopiques est qu'ils sont plus larges que les battements sinusoïdaux. Cette propriété souvent retenue par les médecins, s'avère une caractéristique fondamentale dans la classification de battements cardiaques. Elle est également retenue comme paramètre discriminant

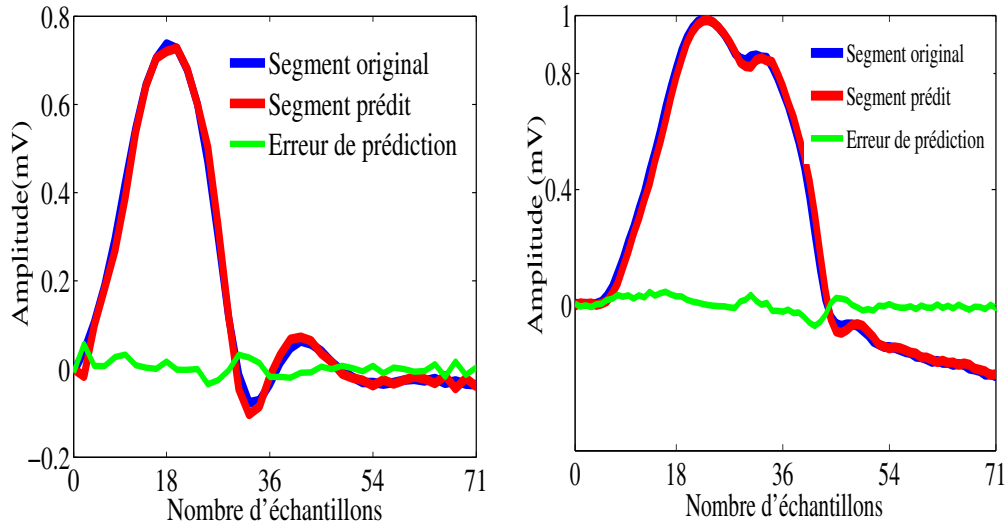


FIG. 4.7 – Résultat de la prédiction appliquée à un battement normal (gauche) et un battement ectopic (droite)

dans la plupart des classifieurs automatiques [De-Chazal 2006, De-Chazal 2004, Krasteva 2007, Zhao 2008].

Dans le ventricule ayant une branche bloquée, l'impulsion électrique arrive en retard par rapport à l'autre : on observe alors sur le tracé ECG une onde R plus large, correspondant aux activités successives et non simultanées des deux ventricules, en revanche le rythme est régulier car il reste contrôlé par le sinus (voir figure 4.8). Un bloc de branche peut-être qualifié de partiel ou total en fonction du degré d'inefficacité de la transmission le long de la branche en question. Notons qu'indépendamment de l'absence de distinction entre les deux blocs de branches gauche et droit lors d'un enregistrement Holter, il existe un risque de confusion entre une onde R avec bloc de branche et une onde R d'extra-systole ventriculaire (ESV) à cause de l'élargissement de l'onde observée dans les deux cas. Les algorithmes d'analyse de la forme de l'onde R permettent en général de différencier efficacement les complexes QRS normaux (y compris avec une variante de type "blocs de branches") des extrasystoles ventriculaires (ESV) car le paramètre largeur de l'onde R est toujours complété par le calcul des intervalles RR .

Dans la présente étude, la largeur du complexe QRS est représentée par celle du segment QS. La détection des pics R (voir section 3.4) est suivie d'une recherche de minima locaux au long des deux pentes positives et négatives de part et d'autre des pics R. Notons que cette recherche est effectuée sur une fenêtre de 180 ms, ce qui correspond au double de la largeur d'un QRS moyen. Les figures 4.9 et 4.10 montrent des exemples de détection des ondes Q, R et S dans le cas d'un ECG normal et dans le cas d'un ECG avec battements PVC.

Sur la figure 4.10, nous constatons que le battement ectopic (PVC) est nette-

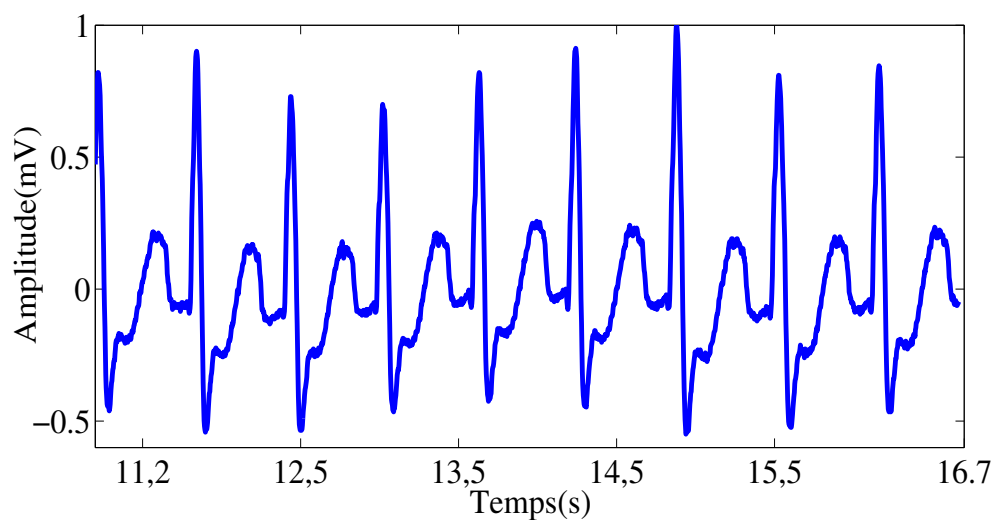


FIG. 4.8 – Bloc de branche MITDB : 109 [11 s, 16.7 s]. Il y a ici un problème de conduction de l'impulsion électrique dans le faisceau de His. La contraction des deux ventricules n'est donc pas parfaitement simultanée. La durée totale de l'onde R est ici plus longue que lors des battements normaux.

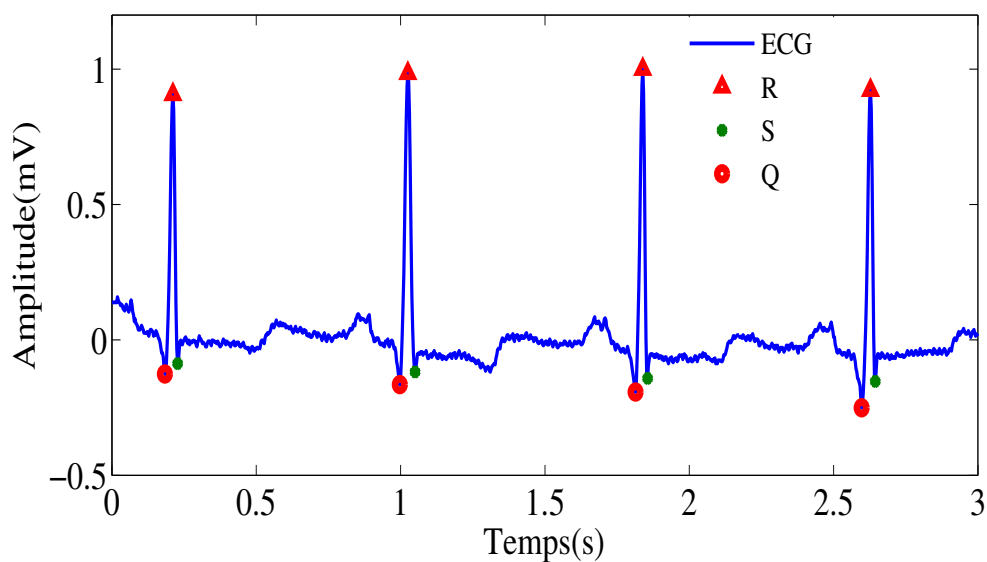


FIG. 4.9 – Détection des segments QS sur l'enregistrement MITDB : 100.

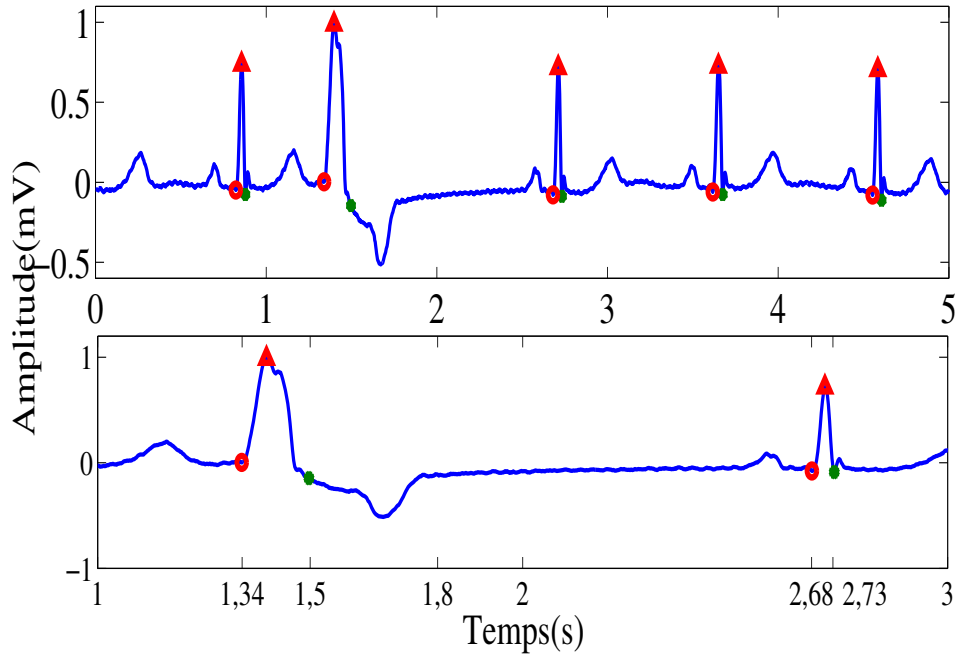


FIG. 4.10 – Détection des segments QS sur l'enregistrement MITDB : 119 (haut) ; zoom sur le segment $[1\text{ s}, 3\text{ s}]$ contenant un battement normal et un PVC (bas).

ment plus large que le battement normal. Leurs durées respectives obtenues sont 160 ms et 50 ms.

4.3.4 Paramètres fréquentiels

Les arythmies sont généralement dues à l'origine du battement qui naît dans un foyer autre que le foyer sinusal ou à un blocage dans les branches de conduction. Dans les deux cas, l'impulsion électrique n'emprunte pas la voie de conduction normale. Les battements cardiaques résultants sont étalés dans le temps, leurs contenus fréquentiels sont alors affectés. Depuis les travaux de V.V.Thakor [Thakor 1984] présentant l'analyse spectrale de l'ECG et celle des complexes QRS isolés, plusieurs chercheurs ont utilisé le contenu fréquentiel des battements comme paramètre discriminant [Minami 1999, Krasteva 2007, Afonso 1999].

En nous basant sur la densité spectrale de puissance (DSP) que nous avons obtenue dans la section 3.4, nous présentons en figure 4.11 les niveaux de puissance occupés par chaque type de battements.

Le lien entre les fréquences et les niveaux de puissance est plus visible sur la figure 4.12. Cette figure montre clairement que les battements V virent vers les basses fréquences et se distinguent parfaitement des N et des RBBB sur tous les niveaux de puissance. Les battements LBBB et V ont des spectres très voisins mais la confusion pouvant provenir de ce paramètre est écartée par la mesure de

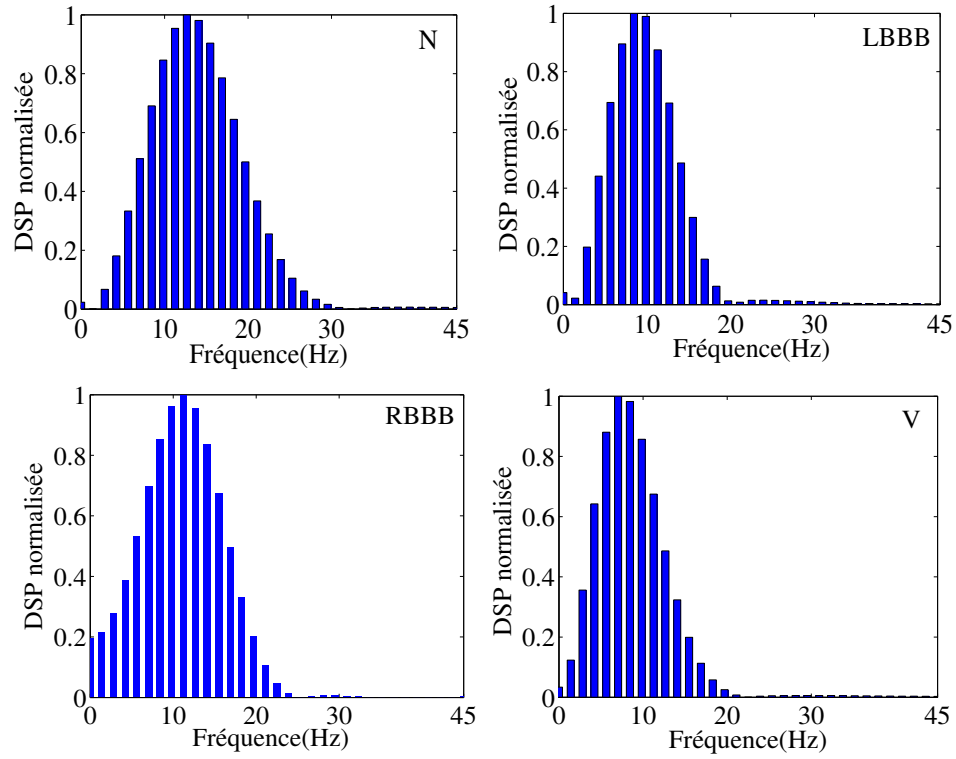


FIG. 4.11 – Niveaux de puissance en fonction de la fréquence, correspondant aux quatre types de battements considérés (N, V, RBBB et LBBB).

la régularité des intervalles RR qui, elle, sépare parfaitement ces deux classes (voir figure 4.5).

Nous avons complété la DSP des complexes QRS par un autre paramètre fré-

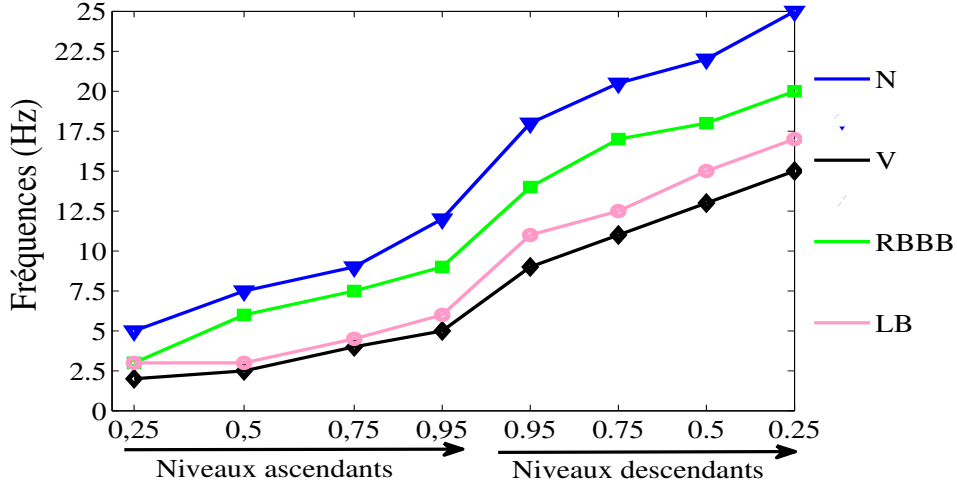


FIG. 4.12 – Relation entre les niveaux de puissance normalisée (0.25, 0.50, 0.75, 0.95), occupés par les quatre types de battements (N, V, RBBB, et LBBB) et le contenu fréquentiel.

quentiel appelé mobilité. Ce paramètre utilisé dans [Ramaswamy 2004] pour la classification de battements cardiaques est défini comme suit :

$$MB(x) = \sqrt{\frac{var(x')}{var(x)}} \quad (4.5)$$

Où x représente le segment QS, $var(x)$, la variance de x et x' la première dérivée de x . MB représente la puissance de la dérivée sur la puissance du signal. Les QRS ectopiques étant plus larges que les QRS normaux, la puissance de leurs dérivées est plus faible par conséquent, la mobilité d'un battement normal est plus élevée que celle d'un battement ectopique.

4.3.5 Les vecteurs de caractéristiques

Sur les enregistrements ECG de notre base d'étude, chaque battement " i " est représenté par un vecteur de caractéristiques " x_i " à 12 éléments : Deux paramètres temporels qui sont l'intervalle RR normalisé et la largeur du segment QS, trois coefficients de prédiction linéaire a_1 , a_2 et a_3 décrivant la morphologie du battement et sept paramètres fréquentiels dont six sont les niveaux de puissance correspondant aux fréquences (7.5 Hz, 10 Hz, 12.5 Hz, 15 Hz, 17.5 Hz et 20 Hz) et

le septième représente la mobilité du battement. Chaque vecteur " x_i " est complété par un treizième élément " y_i " correspondant au label (la classe) du battement " i ". Ces labels sont obtenus à partir de ceux accompagnant chaque enregistrement de la base MITDB.

4.4 Conclusion

A travers ce chapitre, nous avons montré que l'analyse automatique d'un enregistrement ECG peut mettre en évidence un grand nombre de pathologies cardiaques souvent liées à la forme des battements, à leur régularité ou encore à leurs contenus fréquentiels. Cette analyse doit être basée sur un ensemble de points à élaborer soigneusement :

- Représenter chaque QRS par un vecteur de caractéristiques.
- Les caractéristiques doivent être les plus discriminantes possibles sans s'éloigner de celle utilisées par les médecins.
- Les vecteurs doivent être aussi réduits que possible sans perdre les principales informations.

Ces vecteurs de caractéristiques doivent être classifiés de manière automatique. Pour ce faire, nous avons choisi d'utiliser un algorithme nommé : "Support Vector Machines (SVM)" connu pour son efficacité et dont la convergence vers une solution unique est garantie. Cet algorithme fera l'objet du chapitre suivant.

Support Vector Machines(SVMs)

5.1 Introduction

Les "*Support Vector Machines*", ou Séparateurs à Vaste Marge (SVMs) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVMs sont une généralisation des classifieurs linéaires. Ils ont été développés dans les années 1990 à partir des considérations théoriques de Vladimir Vapnik sur le développement d'une théorie statistique de l'apprentissage appelée Théorie de Vapnik-Chervonenkis. Les SVMs ont rapidement été adoptés pour leur capacité de travailler avec des données de grande dimension, leur faible nombre d'hyper paramètres à régler, le fait qu'ils soient bien fondés théoriquement et leur pouvoir de généralisation.

Les SVMs reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau. Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVMs. L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner [Vapnik 1963], et en 1973 par Richard Duda et Peter Hart dans leur livre "*Pattern Classification and scene analysis*" [Richard 1973]. Les fondations théoriques des SVMs ont été explorées par V.Vapnik et ses collègues dans les années 70 avec le développement de la théorie de Vapnik-Chervonenkis, et la théorie de l'apprentissage.

L'idée des fonctions noyaux n'est pas non plus nouvelle : le théorème de Mercer date de 1909 [Mercer 1909]. L'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozoener. Ce n'est toutefois qu'en 1992 que ces idées furent bien comprises et rassemblées par Bosser, Guyon et Vapnik dans un article fondateur des séparateurs à vaste marge [Bosser 1992]. Les variables ressorts, qui permettent de résoudre certaines limitations pratiques importantes ne furent introduites qu'en 1995. À partir de cette date, qui correspond à la publication du livre de V. Vapnik [Vapnik 1995], les SVMs gagnent en popularité. Ils sont actuellement appliqués dans de très nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finance, etc). Selon les données, la performance des SVMs est de même ordre, ou même supérieure, à celle des réseaux de neurones ou d'autres méthodes de classification.

5.2 Apprentissage statistique

Effectuer une classification consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi plusieurs. Le cas le plus simple consiste à discriminer deux classes. D'une manière plus formelle, la classification bi-classe revient à estimer une fonction $f : \mathcal{X} \mapsto \{+1, -1\}$ à partir d'un ensemble d'apprentissage constitué de couples (x_i, y_i) . On notera ici une hypothèse fondamentale pour toute la théorie statistique de l'apprentissage, à savoir que les exemples sont tirés indépendamment les uns des autres, selon une même distribution de probabilités $P(x, y)$ inconnue, tels que :

$$(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}; \quad i = 1, \dots, N; \quad \mathcal{Y} = \{+1, -1\}$$

De sorte à ce que f classifie correctement des exemples inconnus (x_t, y_t) . Par exemple, nous pouvons assigner x_t à la classe $(+1)$ si $f(x_t) \geq 0$ et à la classe (-1) sinon. Les exemples inconnus sont supposés suivre la même distribution de probabilité $P(x, y)$ que ceux de l'ensemble d'apprentissage. La meilleure fonction f est celle obtenue en minimisant le risque :

$$R(f) = \int L[f(x), y] dP(x, y) \quad (5.1)$$

Où L désigne une fonction de coût comme :

$$L[f(x), y] = [f(x) - y]^2 \quad (5.2)$$

Le risque (5.1) ne peut pas être directement minimisé dans la mesure où la distribution de probabilité sous-jacente $P(x, y)$ est inconnue. Aussi, il faut chercher une fonction de décision proche de l'optimale, à partir de l'information dont on dispose, c'est à dire l'ensemble d'apprentissage et la classe de fonctions F à laquelle la solution f appartient.

Pour ce faire, on approxime le minimum du risque théorique par le minimum du risque empirique. On appelle cette mesure un risque empirique car elle est mesurée empiriquement sur les données d'apprentissage. Ce risque est la moyenne des coûts mesurés pour chaque exemple d'apprentissage. Il prend alors la forme :

$$R_{emp} = \frac{1}{N} \sum_{i=1}^N [f(x_i) - y_i]^2 \quad (5.3)$$

où N est le nombre de vecteurs d'apprentissage. Il est possible de donner des conditions au classifieur pour qu'asymptotiquement (*quand* $N \rightarrow \infty$), le risque empirique (5.3) converge vers le risque (5.1).

Une fonction de décision simple (la classe la plus simple est constituée de fonctions linéaires) capable de discriminer correctement les données est préférable à une fonction complexe. Pour cela, on introduit un terme de régularisation pour limiter la complexité des fonctions de F .

5.3 Théorie de Vapnik-Chervonenkis

Une manière de contrôler la complexité d'une classe de fonctions est donnée par la théorie de Vapnik-Chervonenkis (VC) et le principe de Minimisation du Risque Structurel (MRS). Ici, le concept de complexité de la fonction de décision f s'exprime par la dimension VC (notée h) de la classe de fonctions F à laquelle appartient f . Grossièrement, la dimension VC mesure combien d'échantillons de l'ensemble d'apprentissage peuvent être séparés par toutes les classifications possibles issues des fonctions de la classe. Elle mesure la capacité du classifieur.

Considérons une famille imbriquée de classes de fonctions,

$$F_1 \subset F_2 \dots \subset F_k$$

Avec une dimension VC non décroissante et $f_1, \dots, f_k$, les fonctions minimisant le risque empirique dans chacune de ces classes.

La minimisation du risque structurel consiste à choisir la classe F_i (et la fonction f_i) de sorte à ce qu'une borne supérieure de l'erreur de généralisation puisse être minimisée.

Théorème : Soit h la dimension VC d'une famille de fonctions et $R_{emp}(f)$, le risque empirique défini par (5.3). Avec une fonction de perte binaire, (i.e. $L[f(x), y] = H[-y \cdot f(x)]$ où H désigne l'échelon de Heaviside. Il existe $\delta \in [0, 1]$ et $f \in F$, pour lesquels l'inégalité (bornant le risque) donnée par :

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h(\ln(\frac{2N}{h}) + 1) - \ln(\frac{\delta}{4})}{N}} \quad (5.4)$$

est vraie avec une probabilité d'au moins $(1 - \delta)$ pour $N > h$.

Cette borne n'est qu'un exemple. Des formulations du même type ont été démontrées pour d'autres fonctions de perte et d'autres mesures de complexité. Le but recherché ici est de minimiser l'erreur de généralisation $R(f)$ en obtenant un faible risque empirique $R_{emp}(f)$ tout en gardant la plus petite classe de fonctions possible. L'inégalité (5.4) fait apparaître deux cas extrêmes :

- une très petite classe de fonctions (par exemple F_1) fait décroître rapidement le terme de complexité (celui sous la racine carrée), mais le risque empirique demeure grand.
- une très grande classe de fonctions (par exemple F_k) implique un risque empirique petit, mais le terme de complexité explose.

La meilleure classe de fonctions est généralement intermédiaire entre la plus petite et la plus grande, puisque l'on cherche une fonction qui explique au mieux les données tout en préservant un faible risque empirique (voir figure 5.1).

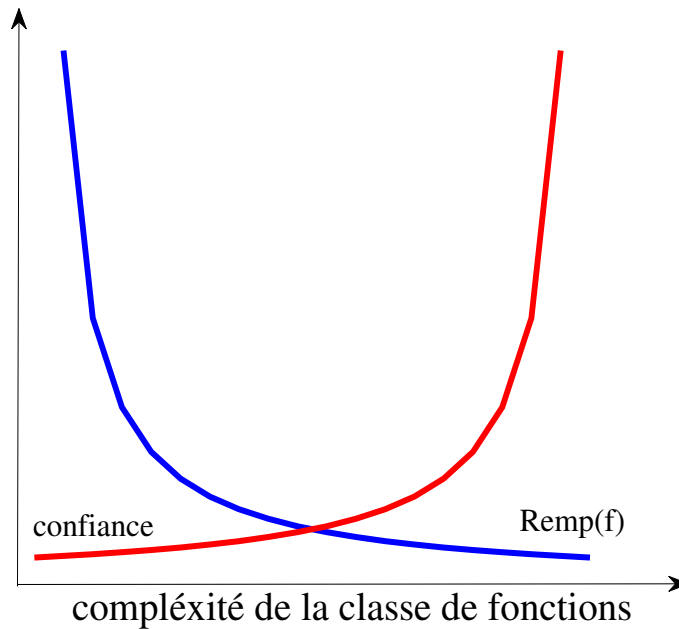


FIG. 5.1 – *Illustration de l'inégalité (5.4). La courbe croissante, appelée confiance, correspond à la borne supérieure du terme de complexité. Les comportements du terme de complexité et de l'erreur empirique sont clairement opposés. On recherche donc le meilleur compromis entre complexité et erreur empirique.*

5.4 Objectifs d'un SVM

Soit un nuage de points de natures différentes (points rouges, points bleus). L'objectif recherché est de trouver une frontière de décision (hyperplan séparateur) qui puisse séparer le nuage de points en deux régions en commettant un minimum d'erreurs, c'est à dire, (trouver l'hyperplan optimal). La figure 5.2 montre qu'il existe en effet plusieurs d'hyperplans séparateurs dont les performances en apprentissage sont identiques (le risque empirique est le même), mais dont les performances en généralisation peuvent être très différentes. Pour résoudre ce problème, il a été montré [Vapnik 1995] qu'il existe un unique hyperplan (l'optimal), défini comme l'hyperplan maximisant la marge entre les échantillons.

Un autre objectif des séparateurs à vaste marge comme le terme l'indique, est de repousser le plus possible les deux classes l'une de l'autre. Ceci revient à maximiser la distance entre les points les plus proches du plan séparateur H (voir

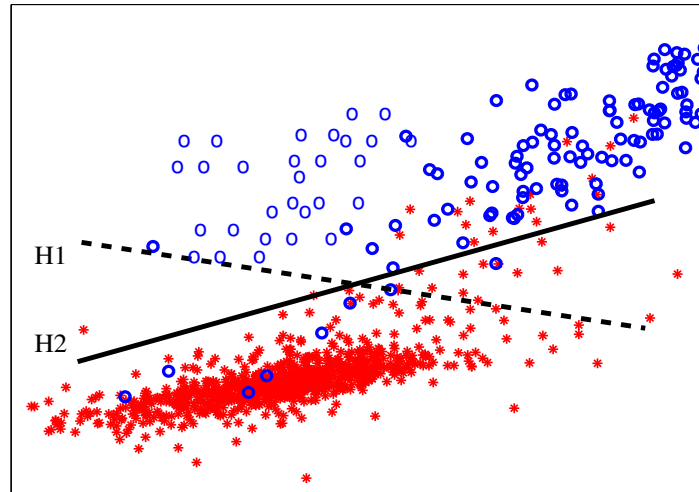


FIG. 5.2 – Principe d'hyperplan séparateur, il en existe plusieurs. Celui qui correspond au minimum d'erreurs est l'hyperplan optimal.

figure 5.3). Cette distance est appelée "Marge d ". Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsqu'on classe un nouvel exemple. De plus, si on trouve un classifieur qui se comporte le mieux vis-a-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer de nouveaux exemples.

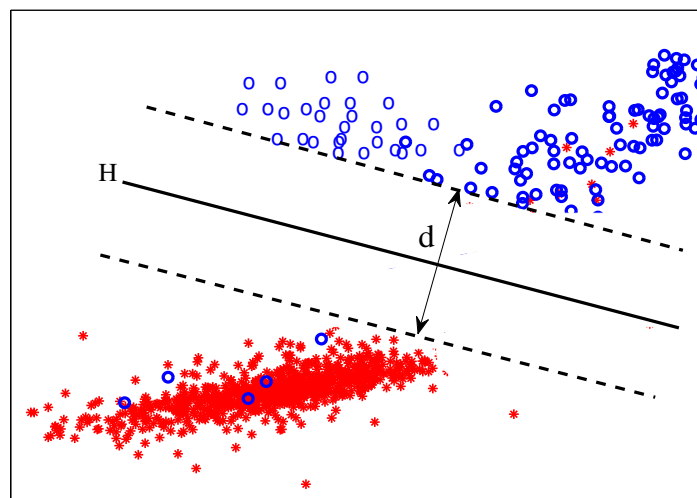


FIG. 5.3 – L'hyperplan optimal H (en gras) avec la marge maximale d .

5.5 Linéarité et non linéarité

Parmi les modèles des SVM, on constate les cas linéairement séparables et les cas non linéairement séparables. Les premiers sont les plus simples car ils permettent de trouver facilement le classifieur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données. Le classifieur de marge maximale ne peut pas être utilisé dans ces cas car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

Pour illustration, la figure 5.4.(a) indique un plan (espace à deux dimensions) dans lequel sont répartis deux groupes de points : les points (+) pour $y > x$ et les points (-) pour $y < x$. On peut trouver un séparateur linéaire évident dans cet exemple, la droite d'équation $y = x$. Le problème est dit linéairement séparable.

La figure 5.4.(b) montre un plan dans lequel les points (-) sont regroupés à l'intérieur d'un cercle, avec des points (+) tout autour : aucun séparateur linéaire ne peut correctement séparer les deux groupes : ce problème n'est pas linéairement séparable.

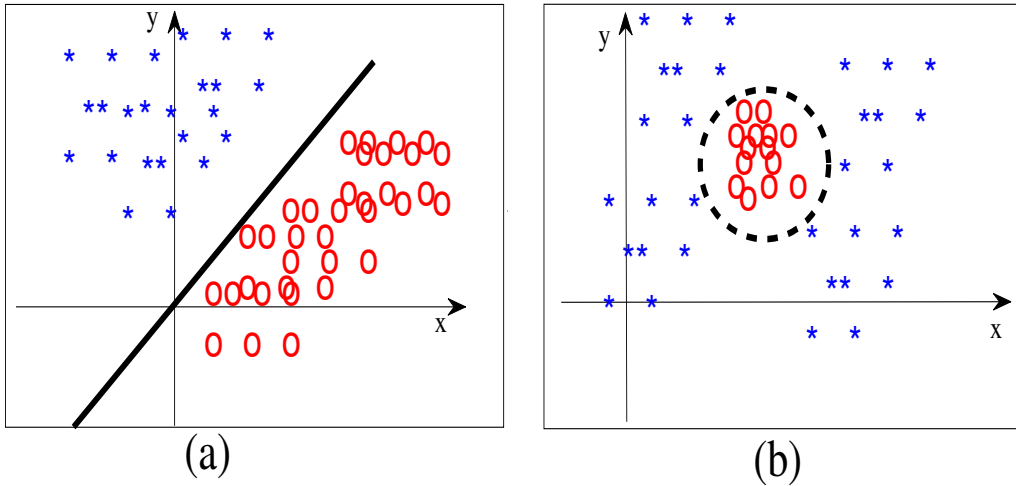


FIG. 5.4 – Données linéairement séparables (a); données non linéairement séparables (b)

5.6 L'espace augmenté

Choisir des frontières de décision linéaires semble être un facteur limitant. Cependant, de tels modèles peuvent être considérablement enrichis en projetant les données non linéairement séparables dans un espace caractéristique $\mathcal{F}$ (*feature space*) de plus grande dimension permettant d'augmenter la séparabilité des données (voir figure 5.5). On peut alors appliquer le même algorithme dans ce nouvel

espace, ce qui se traduit par une frontière de décision non linéaire dans l'espace initial.

Considérons l'application non linéaire définie par :

$$\phi: \mathcal{X} \mapsto \mathcal{F}$$

$$x \mapsto \phi(x)$$

Il suffit alors d'appliquer l'algorithme d'apprentissage dans $\mathcal{F}$ et non dans $\mathcal{X}$ en considérant l'ensemble $(\phi(x_i), y_i) \in \mathcal{F} \times \mathcal{Y}$ avec $i = 1, \dots, N$ et $\mathcal{Y} = \{+1, -1\}$.

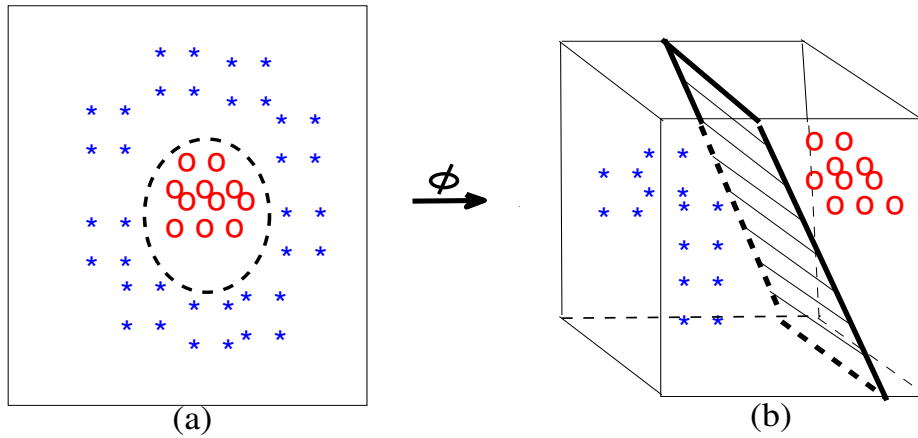


FIG. 5.5 – Exemple d'espace d'entrée $\mathcal{X}$, (a) et d'espace caractéristique $\mathcal{F}$, (b)

5.6.1 Fonctions noyaux et similarité

L'étape de généralisation de toute méthode d'apprentissage consiste à mettre en relation une nouvelle donnée à classer avec une base d'apprentissage par l'intermédiaire d'informations extraites de ces données. Cette prise de décision est basée sur une mesure de similarité. La classe d'une nouvelle donnée est alors décidée comme étant celle qui présente le plus de similarité.

Si on considère deux données x et $x' \in \mathbb{R}^N$, leur produit scalaire est donné par : $\langle x, x' \rangle = \sum_{i=1}^N x_i \cdot x'_i$. D'un point de vue géométrique, ce produit scalaire correspond au cosinus de l'angle entre ces vecteurs normalisés à 1. Cette opération mesure alors le degrés de similarité entre les données puisque, plus elles sont similaires, plus l'angle qu'elles décrivent est faible et leur produit scalaire est important. A l'inverse, si elles tendent à être orthogonales, leur produit scalaire tend vers zéro. Ce produit scalaire permet en plus de doter l'espace considéré d'une métrique définie par : $\|x\| = \sqrt{\langle x, x \rangle}$.

L'algorithme SVM requiert que les données soient représentées dans un espace doté d'un produit scalaire. Pour garantir cela, on introduit un projecteur ϕ de l'espace d'origine $\mathcal{X}$ dans un espace $\mathcal{F}$ (*the feature space*) qui sera lui, doté d'un produit scalaire. Il suffit alors de remplacer chaque produit scalaire $\langle x_i, x_j \rangle$ par $\langle \phi(x_i), \phi(x_j) \rangle$. Cependant, pour certains espaces $\mathcal{F}$, il arrive que la projection ϕ ne soit pas calculable directement. Pour pallier à ce problème, on a recours à une famille de fonctions permettant de réaliser implicitement les produits scalaires dans $\mathcal{F}$. On appelle ces fonctions, fonctions noyaux (*Kernel*) et on les définit comme :

$$k(x, x') = \langle \phi(x), \phi(x') \rangle \quad (5.5)$$

Ainsi, chaque fonction noyau correspond à une projection ϕ et par conséquent, à un espace augmenté $\mathcal{F}$. En pratique, la transformation ϕ n'a pas besoin d'être connue explicitement, seule la fonction noyau intervient dans les calculs. On peut donc envisager des transformations complexes, et même des espaces de redescription de dimension infinie.

La question est de savoir quelles sont les fonctions k qui admettent une telle représentation, c'est-à-dire pour lesquelles on peut trouver un espace $\mathcal{F}$ et une projection ϕ . Cette question a suscité un certain nombre de travaux, tant au sein du domaine de l'apprentissage statistique que de l'analyse fonctionnelle.

Théorème de Mercer : Une fonction $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ est un noyau valide si elle est symétrique et définie positive.

Sous cette condition, le noyau k définit bien un espace de Hilbert $\mathcal{H}$ ¹. Lorsqu'on considère une base d'apprentissage, soit, un sous-ensemble discret de l'espace $\mathcal{X}$, on peut considérer de manière équivalente la matrice de Gram appelée aussi matrice de similarité définie par :

$$G(i, j) = k(x_i, x_j) \quad (5.6)$$

Cette matrice est de dimension $N \times N$. N étant la taille de la base d'apprentissage. Les conditions précédentes s'écrivent alors comme suit :

- $G(i, j) = G(j, i)$
- $c^T G c \geq 0 \quad \forall c \in \mathbb{R}^N$

Cette dernière condition se traduit par le fait que toutes les valeurs propres de la matrice de Gram doivent être strictement positives.

5.6.2 Choix de la fonction noyau

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur des SVMs d'effectuer des tests pour déterminer celle

¹L'espace de Hilbert est un espace vectoriel muni d'un produit scalaire dont l'espace normé associé est complet.

TAB. 5.1 – Quelques noyaux usuels

Noyau linéaire	$k(x_i, x_j) = x_i^T x_j$
Noyau gaussien	$k(x_i, x_j) = \exp(-\frac{\ x-y\ ^2}{2\sigma^2})$
Noyau polynomial	$k(x_i, x_j) = (x_i^T x_j + 1)^d$

qui convient le mieux pour son application. La table 5.1 indique quelques noyaux usuels. Les paramètres du noyau choisi doivent être déterminés en fonction de la base d'apprentissage par des méthodes statistiques comme la validation croisée.

5.7 Fondement mathématique des SVMs

Le fondement mathématique des Séparateurs à Vaste Marge est expliqué dans plusieurs ouvrages comme [Christopher 1998, Vapnik 1995, Loosli 2005]

5.7.1 Principe général

Les SVMs peuvent être utilisés pour résoudre des problèmes de discrimination binaire, c'est-à-dire, décider à quelle classe appartient un échantillon. La résolution de ce problème passe par la construction d'une fonction f qui, à un vecteur d'entrée $x \in X$ fait correspondre une sortie $f(x)$: Il est alors décidé que x est de classe $+1$ si $f(x) > 0$ et de classe -1 si $f(x) < 0$. C'est un classifieur linéaire. La frontière de décision $f(x) = 0$ est un hyperplan séparateur.

Soit H un hyperplan, w son vecteur normal et b , son décalage par rapport à l'origine (voir figure 5.6). L'hyperplan H est alors donné par :

$$f(x) = w^T x + b = 0$$

Le but de l'algorithme d'apprentissage d'un SVM est de trouver les paramètres w et b du meilleur hyperplan par le biais d'un ensemble d'apprentissage :

$$\mathcal{X} \times \mathcal{Y} = \{(x_1, y_1), \dots, (x_i, y_i)\} \in \mathbb{R}^N \times \{+1, -1\}$$

où les y_i sont les labels respectifs des x_i , N la taille de l'ensemble d'apprentissage.

5.7.2 Cas linéairement séparable

On se place ici dans le cas où le problème est linéairement séparable. Même dans ce cas simple, le choix de l'hyperplan séparateur n'est pas évident car il existe en effet une infinité d'hyperplans séparateurs. Pour résoudre ce problème, il a été montré qu'il existe un unique hyperplan optimal, défini comme étant celui qui maximise la marge entre les échantillons et l'hyperplan séparateur [Vapnik 1995].

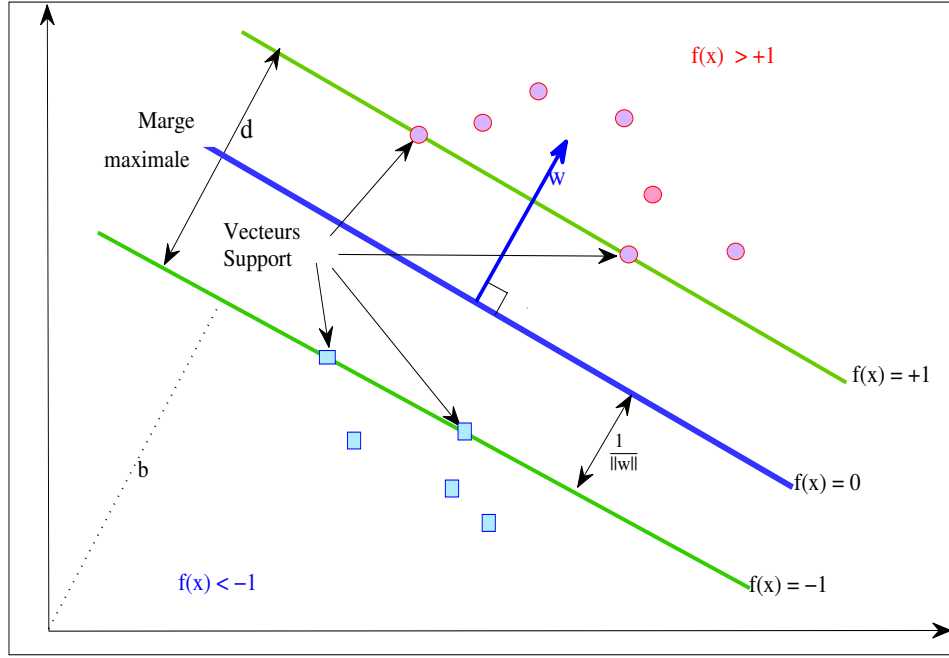


FIG. 5.6 – Illustration de la marge et des Vecteurs Support.

5.7.2.1 Formulation du problème d'optimisation primal

La marge est la distance entre deux points, les plus proches de l'hyperplan mais appartenant à des classes différentes (voir figure 5.6). Ces derniers sont appelés : **Vecteurs Support (VS)**. Il s'agit alors de trouver le couple (w, b) qui maximise la marge afin de déterminer l'équation de l'hyperplan optimal H . Ce couple est défini par :

$$\arg \max_{w, b} \min_i \|x - x_i\| : x \in \mathcal{X}, (w^T x + b) = 0, i = 1, \dots, N$$

Soient x^+ et x^- deux points de classes différentes situés respectivement sur les frontières positive et négative délimitant la marge maximale. Pour simplifier le problème d'optimisation, on considère que x^+ et x^- sont situés sur les hyperplans canoniques tels que $f(x^+) = +1$ et $f(x^-) = -1$, c'est à dire $w^T x^+ + b = +1$ et $w^T x^- + b = -1$.

On sait que la distance d'un point quelconque x à H est définie par :

$$d_{x,H} = \frac{|w^T x + b|}{\|w\|}$$

La distance entre chacun des deux points x^+ et x^- et H est alors $1/\|w\|$. Dans ce cas, la marge est :

$$d = \frac{w^T}{\|w\|} (x^+ - x^-) = \frac{2}{\|w\|}$$

Nous déduisons à partir de là que maximiser la marge revient à minimiser $\|w\|$ sous contraintes que $y_i(w^T x_i + b) \geq 1$. Cette contrainte signifie que le SVM tient compte non seulement de la position des exemples par rapport à l'hyperplan ($\text{signe}(f(x))$), mais aussi de leurs distances par rapport à cet hyperplan. Le problème d'optimisation est alors posé comme suit :

$$\begin{cases} \min_{w,b} & \frac{1}{2}\|w\|^2, \\ \text{S.c} & y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, N. \end{cases} \quad (5.7)$$

Notons qu'il est plus aisé de minimiser $\|w\|^2$ plutôt que $\|w\|$.

5.7.2.2 Formulation du problème d'optimisation dual

La résolution du problème quadratique (5.7) revient à résoudre son problème dual. Son Lagrangien est :

$$\mathcal{L}(w, b) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^N \alpha_i [y_i(w^T x_i + b) - 1] \quad (5.8)$$

où les α_i sont les coefficients de Lagrange qui doivent être positifs ou nuls. Le problème (5.7) doit satisfaire les conditions de KKT (Karush-Kuhn-Tucker) qui consistent à annuler les dérivées partielles du Lagrangien 5.8 par rapport aux variables primales w, b .

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0, \\ \frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i y_i x_i. \end{cases} \quad (5.9)$$

Le problème $\frac{1}{2}\|w\|^2$ étant convexe, $w(\alpha)$ est unique. En réinjectant les valeurs obtenues par les conditions KKT dans l'équation (5.8), nous obtenons la forme duale du problème (5.7) comme suit :

$$\begin{cases} \max_{\alpha} & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j y_i y_j (x_i^T x_j), \\ \text{S.c} & \sum_{i=1}^N \alpha_i y_i = 0, \\ & \alpha_i \geq 0; \quad i = 1, \dots, N. \end{cases} \quad (5.10)$$

C'est un problème quadratique de dimension N (taille de l'ensemble d'apprentissage). Sa résolution revient à chercher les indices i des α_i^* positifs correspondant aux points x_i^* qui sont Vecteurs Supports (VS). Pour chaque nouveau point x à classer, la fonction de décision sera donnée par :

$$f(x) = \sum_{i=1}^N \alpha_i^* y_i^* (x_i^* \cdot x) + b \quad (5.11)$$

avec, $\sum_{i=1}^N \alpha_i^* y_i^* x_i^* = w$. La valeur de b peut être calculée en utilisant n'importe quel vecteur support x^* en posant : $\alpha^*[y^*(w \cdot x^* + b) - 1] = 0$.

5.7.2.3 Conséquences

Il y a trois remarques intéressantes à faire à propos du résultat précédent :

1. La première découle de l'une des conditions de KKT, qui donne $\alpha_i[y_i f(x_i) - 1] = 0$ pour $i = 1, \dots, N$ d'où $\alpha_i = 0$ ou bien $y_i f(x_i) = 1$. Les seuls points pour lesquels les contraintes du lagrangien sont actives sont donc les points pour lesquels $y_i f(x_i) = 1$. Ces points sont situés sur les hyperplans canoniques. En d'autres termes, seuls les vecteurs supports participent à la définition de l'hyperplan optimal.
2. La deuxième remarque découle de la première. Seul un sous-ensemble restreint de points est nécessaire pour le calcul de la solution. Ceci est donc efficace au niveau de la complexité.
3. La dernière remarque est que l'hyperplan solution ne dépend que du produit scalaire entre le vecteur d'entrée et les vecteurs supports. Cette remarque est à l'origine de la deuxième innovation majeure des SVMs : le passage à un espace de caractéristiques $\mathcal{F}$ grâce à la fonction noyau.

5.7.3 Cas non linéairement séparable

Les données d'apprentissage peuvent être bruitées et non séparables, même dans l'espace $\mathcal{F}$. Il faut alors trouver un bon compromis entre le risque empirique et complexité (cf. Figure 5.1). En 1995, Corinna Cortes et Vladimir Vapnik proposèrent une technique dite de marge souple en introduisant des variables ressort ξ_i (*slack variables*) pour relâcher sensiblement les contraintes sur la marge. Ces dernières deviennent alors :

$$y_i f(x_i) \geq 1 - \xi_i$$

5.7.3.1 Formulation du problème primal

Les variables de relaxation autorisent quelques erreurs de classification lors de l'apprentissage. Le problème d'optimisation (5.7) devient alors :

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N |1 - y_i f(x_i)|_+ \quad (5.12)$$

avec où $|\cdot|_+ = \max(\cdot, 0)$ Le problème (5.12) est souvent exprimé en fonction des variable d'écart ξ_i comme suit :

$$\begin{cases} \min_{w,b,\xi} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i , \\ \text{S.c} & y_i (w^T x_i + b) \geq 1 - \xi_i , \\ & \xi_i \geq 0, \quad i = 1, \dots, N . \end{cases} \quad (5.13)$$

La constante C est un paramètre déterminant la tolérance du SVM aux exemples mal classés. Elle permet de contrôler le compromis entre nombre d'erreurs de classement et la largeur de la marge (voir figure 5.7). Plus C est grand, plus on pénalise les mauvaises classifications et la complexité de la classe de fonctions de décision devient grande. Le choix automatique de ce paramètre de régularisation est un problème statistique majeur.

A travers le problème (5.12), on cherche à maximiser la marge et à minimiser la fonction de pertes (fonction de coût) définie par :

$$\ell(y_i, f(x_i)) = C \sum_{i=1}^N |1 - y_i f(x_i)|_+ = C \sum_{i=1}^N \xi_i \quad (5.14)$$

Cette fonction couramment appelée "*hinge loss*" est une fonction convexe. Elle garantit une solution unique au problème [Vapnik 1995].

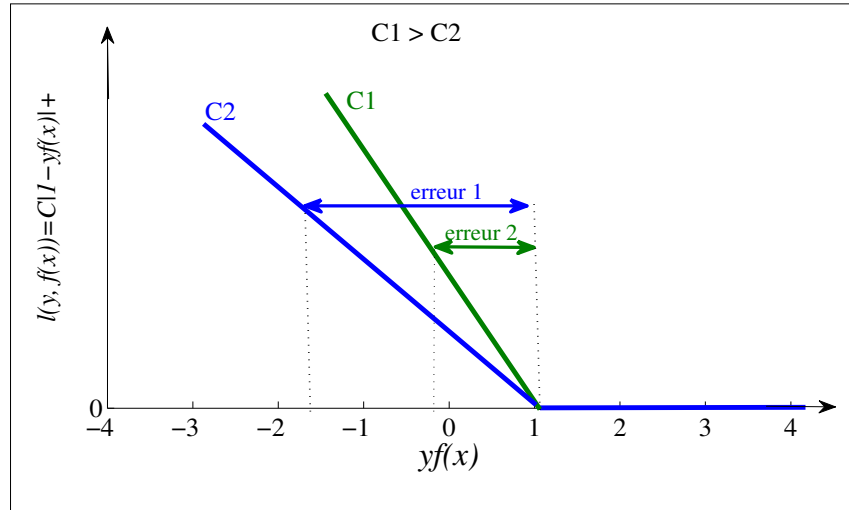


FIG. 5.7 – Représentation du compromis entre la largeur de la marge souple et le coût d'une erreur.

5.7.3.2 Formulation du problème dual

La solution du problème (5.13) est aussi le point selle de son Lagrangien :

$$\mathcal{L}(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^N \beta_i \xi_i \quad (5.15)$$

Avec α et β , des coefficients de Lagrange positifs ou nuls. Les conditions d'optimalité fournies à l'égard du Lagrangien (5.15) se traduisent par ses dérivées partielles

TAB. 5.2 – Lien entre les contraintes primales et duales

$y_i f(x_i) > 1$ et $\xi = 0$	$\alpha_i = 0$
$y_i f(x_i) = 1$ et $\xi = 0$	$0 < \alpha_i < C$
$y_i f(x_i) < 1$ et $\xi > 0$	$\alpha_i = C$

nulles par rapport aux variables primales w , b , ξ (conditions de KKT).

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}(w, b, \xi, \alpha, \beta)}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \ , \\ \frac{\partial \mathcal{L}(w, b, \xi, \alpha, \beta)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i y_i x_i \\ \frac{\partial \mathcal{L}(w, b, \xi, \alpha, \beta)}{\partial \xi} = 0 \Rightarrow \beta_i = C - \alpha_i \end{array} \right. \quad (5.16)$$

Injectées dans l'expression du Lagrangien (5.15), ces relations fournissent le problème dual à résoudre :

$$\left\{ \begin{array}{l} \max_{\alpha} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i^T x_j) \ , \\ \text{S.c} \quad \sum_{i=1}^N \alpha_i y_i = 0 \ , \\ \quad \quad 0 \leq \alpha_i \leq C; \quad i = 1, \dots, N \ . \end{array} \right. \quad (5.17)$$

Ce problème est similaire au problème (5.10) avec une contrainte supplémentaire sur les coefficients de Lagrange α_i et la matrice Hessienne $(x_i^T x_j)$ qui est remplacée par la matrice de Gram $G(i, j)$ car l'espace d'entrée est projeté vers l'espace augmenté. La plupart des méthodes d'optimisation sont basées sur des conditions d'optimalité du second ordre (contrainte duales). La correspondance entre les contraintes primales et les contraintes duales du problème précédent est récapitulée sur la table 5.2.

Les seuls coefficients α_i non nuls sont ceux associés aux vecteurs x_i^* de la base d'apprentissage qui sont sur les deux frontières délimitant la marge ($y_i^* f(x_i^*) = 1$ et $0 < \alpha_i^* < C$) et ceux à l'intérieur de la marge ($y_i^* f(x_i^*) < 1$ correspondant à $\alpha_i^* = C$). Ces points sont les vecteurs supports recherchés. La fonction de décision pour classer un nouveau point x est alors :

$$f(x) = \sum_{i=1}^N \alpha_i^* y_i^* k(x_i^*, x) + b \quad (5.18)$$

5.8 Fonctions de coût d'un SVM

Nous avons supposé jusqu'ici qu'il existait un hyperplan (éventuellement dans l'espace de redescription) permettant de séparer les exemples des deux classes. Or, d'une part, il n'est pas nécessairement souhaitable de rechercher absolument un tel hyperplan, cela peut en effet conduire à une suradaptation aux données, d'autre part, il se peut que du bruit ou des erreurs dans les données ne permettent tout simplement pas de trouver un tel hyperplan. Pour ces raisons, une version moins contraignante du problème de recherche d'une séparatrice à vaste marge est le plus souvent considérée.

L'idée est de pénaliser les séparatrices admettant des exemples qui ne sont pas du bon côté des marges, sans cependant interdire une telle possibilité. On définit pour ce faire une fonction de coût particulière introduisant une pénalité pour tous les exemples mal classés et qui sont à une distance de la marge, qu'ils devraient respecter. On considère généralement des fonctions de coût qui soient compatibles avec la fonction de perte classique $(0, 1)$ perte qui compte un coût de 1 pour chaque exemple mal classé et un coût nul pour un exemple correctement classé. En particulier, on cherche des fonctions de coût qui conduisent à un critère inductif compatible avec le critère classique de minimisation du nombre d'exemples mal classés et donc à des solutions optimales compatibles. On parle de fonctions de coût de substitution (*surrogate loss functions*). Plusieurs fonctions sont possibles. Les plus utilisées sont représentées sur la figure 5.8 .

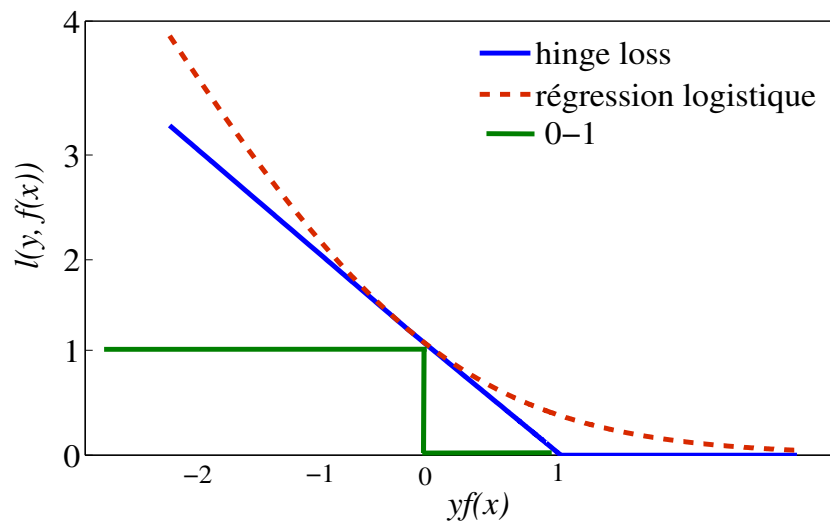


FIG. 5.8 – Approximations de la fonction de perte 0, 1 (vert), par les fonctions coude ou hinge loss (bleu) et la fonction logistique (rouge). L'axe des abscisses correspond à la quantité $yf(x)$ qui est négative si l'exemple x est mal classé par f .

5.9 Algorithmes d'apprentissage des SVMs

L'apprentissage d'un SVM se ramène essentiellement à résoudre un problème d'optimisation impliquant un système de résolution de programmation quadratique dans un espace de dimension conséquente. C'est pourquoi ces programmes utilisent des méthodes spéciales pour y parvenir de manière efficace. Le succès des SVMs a entraîné le développement de nombreux algorithmes permettant leur mise en œuvre. Parmi ces méthodes, l'algorithme SMO (Sequential Minimal Optimization) posé par J.C.Platt en 1998 [J.platt, 1998] . Cet algorithme d'apprentissage pour SVMs est généralement rapide, simple à implémenter et nécessite un espace mémoire réduit. Un autre algorithme aussi bien connu et très utilisé est le "*SVM^{light}*" décrit dans [Joachim 2002] et disponible sur le lien (www.download.joachims.org/SVM^{light}). L'algorithme *SimpleSVM* basé sur la méthode des contraintes actives fût proposé par Vishwanathan en 2003[Vishwanathan 2003] puis repris et amélioré par G.Loosli et S.Canu en 2005 [Loosli 2005]. Dans cet article, les auteurs montrèrent que l'algorithme *SimpleSVM* offre une meilleure rapidité de convergence et une meilleure complexité algorithmique ($O^{1.2}$) que l'algorithme SMO.

5.10 Conclusion

Dans ce chapitre, nous avons présenté de manière simple et complète la méthode d'apprentissage introduite par Vladimir Vapnik, les " Support Vector Machines ". Nous avons donné une vision générale et le fondement mathématique des SVM. Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des classes de données. Nous avons exposé les cas linéairement séparable et les cas non linéairement séparables qui nécessitent l'utilisation de fonctions noyaux (kernel) pour changer d'espace. Cette méthode est applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multi-classe. Les SVMs représentent aujourd'hui l'une des méthodes les plus utilisées grâce à leur pouvoir de généralisation. Ils sont fondés rigoureusement et simplement. L'utilisateur peut alors porter des modifications selon l'objectif recherché. Nous verrons dans le chapitre suivant comment nous avons introduit le rejet d'ambiguïtés pour la classification des battements cardiaques.

Classification avec rejet d'ambigüités

6.1 Introduction

A travers ce chapitre, nous allons modifier l'algorithme SVM afin de développer une méthode de classification qui permettra la reconnaissance d'arythmies cardiaques avec un coût optimal de classification. En effet, malgré la grande capacité de généralisation des SVMs, l'erreur de classification ne peut pas être complètement éliminée et peut alors engendrer des conséquences très graves ; notamment dans le cas d'un diagnostic médical où le coût d'une erreur de classification peut être très élevé. Dans ce cas, il est souhaitable de reporter le classement d'un battement dont la probabilité d'appartenance à une classe est autour de 0.5 . Ceci nous a motivés à introduire l'option de rejet dans le SVM.

Depuis les publications de Chow [Chow 1957, Chow 1970] sur le compromis erreur-rejet, cette approche n'a pas bénéficié d'un grand intérêt de la part de la communauté du "*machine learning*". Nous notons cependant quelques remarquables travaux [Kwok 1999, Tortorella 2004] où les auteurs proposèrent une méthode de rejet explicitement pour les SVMs. La zone de rejet est délimitée par deux hyperplans parallèles à l'hyperplan optimal. Les seuils de décision sont calculés sur les scores du SVM. En 2002, Un algorithme d'apprentissage SVM avec rejet fut proposé par [Fumera 2002] où la zone de rejet est délimitée durant la phase d'apprentissage. Néanmoins, cette méthode manque de convexité et impose un calcul très lourd. D'autres classifieurs SVMs avec option de rejet et basés sur la fonction de pertes "*double hinge loss*" sont récemment proposés dans [Herbei 2006, Bartlett 2008, Grandvalet 2009].

La méthode de rejet que nous utilisons dans cette thèse est basée sur l'interprétation probabilistique des SVMs introduite par Yves Grandvalet en 2006 [Grandvalet 2006]. Nous introduisons une fonction de perte à double charnière (*double hinge loss*) dédiée à un problème de classification avec option de rejet généralisant celle proposée par [Bartlett 2008] à une classification asymétrique. L'approche probabiliste nous permet d'aboutir à une règle de décision donnant de meilleurs résultats comme nous le verrons plus loin.

6.2 Règle de décision de Bayes

L'objectif de la classification est de prédire le 'label' $y \in \mathcal{Y}$ d'une observation $\mathbf{x} \in \mathcal{X}$. pour ce faire, nous construisons une règle de décision ' d ' qui affecte une étiquette à tout $\mathbf{x} \in \mathcal{X}$. Dans un problème de classification binaire, où les classes sont étiquetées par $+1$ ou -1 , les deux types d'erreurs possibles sont : faux positifs (FP) où un exemple négatif est classé comme positif entraînant un coût c_- ; faux négatifs (FN), où un exemple positif est classé comme négatif entraînant un coût c_+ . Nous considérons ici un problème où certains exemples ne doivent pas être classés. La décision ' d ' du classifieur a l'option de rejeter les exemples dont la probabilité d'appartenance à chacune des deux classes est proche de 0.5 . Cette troisième décision d'abstention notée 0, entraine un coût r_- et r_+ respectivement pour les exemples étiquetés -1 et $+1$.

Les coûts engendrés par chacune des décisions possibles sont récapitulés dans la table 6.1. Dans ce qui suit, nous considérons :

TAB. 6.1 – Coût associé à chaque décision

		y		$E_y[d(\mathbf{x})]$
		$+1$	-1	
$d(\mathbf{x})$	$+1$	0	c_-	$c_- (1 - P(Y = 1 \mathbf{x}))$
	0	r_+	r_-	$r_+ P(Y = 1 \mathbf{x}) + r_- (1 - P(Y = 1 \mathbf{x}))$
	-1	c_+	0	$c_+ P(Y = 1 \mathbf{x})$

$$c_- > 0, c_+ > 0, r_- > 0, r_+ > 0. \quad (6.1)$$

En posant $P = P(Y = 1|\mathbf{x})$, probabilité d'avoir un point positif et $1 - P$, la probabilité d'avoir un point négatif, les coûts relatifs à chaque décision sont illustrés par la figure 6.1. L'option de rejet introduite n'a un sens que si le coût d'un exemple rejeté est inférieur au coût d'un exemple mal classé et ce, entre les seuils p_- et p_+ . Sur la figure 6.1, le point G, intersection des droites (a) et (b) doit être au dessus du segment [AB], c'est à dire :

$$c_+ P_G > r_+ P_G + r_- (1 - P_G)$$

où $P_G = \frac{c_-}{c_+ + c_-}$ est l'abscisse du point G. A partir de là, on estime que le rejet est valable si et seulement si :

$$\frac{r_+}{c_+} + \frac{r_-}{c_-} < 1 \quad (6.2)$$

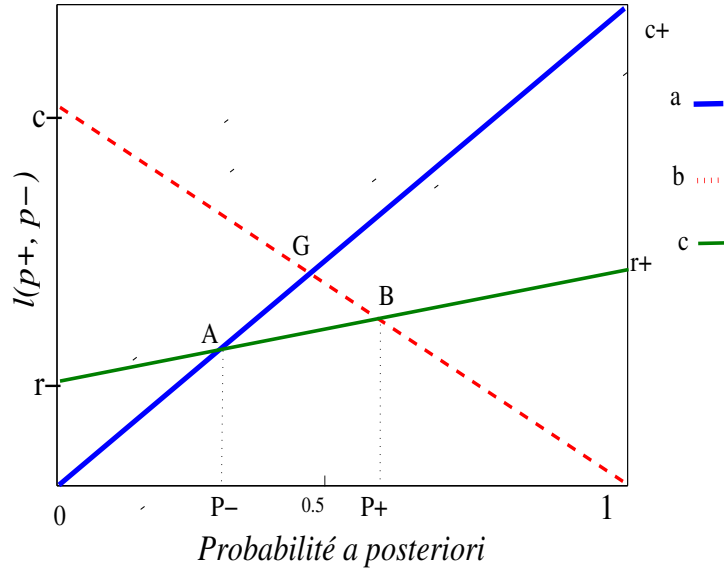


FIG. 6.1 – Coût de chaque classification en fonction de la probabilité a posteriori P . Les courbes a , b et c représentent respectivement : $E_y = c_+ P$, $E_y = c_- (1 - P)$ et $E_y = r_+ P + r_- (1 - P)$.

La règle de décision de Bayes peut être tout simplement exprimée en utilisant les deux seuils :

$$p_+ = \frac{c_- - r_-}{c_- - r_- + r_+} , \quad (6.3)$$

$$p_- = \frac{r_-}{c_+ - r_+ + r_-} , \quad (6.4)$$

Si les inégalités (6.1) et (6.2) sont satisfaites, nous avons $0 < p_- < p_+ < 1$. la règle de Bayes peut alors être énoncée comme suit :

$$d^*(\mathbf{x}) = \begin{cases} +1 & \text{si } P(Y = 1|X = \mathbf{x}) > p_+ , \\ -1 & \text{si } P(Y = 1|X = \mathbf{x}) < p_- , \\ 0 & \text{sinon .} \end{cases} \quad (6.5)$$

Cette règle est illustrée par la figure 6.2. Sa construction tend à minimiser la fonction de perte : (6.6) :

$$L(d) = c_+ P(Y = 1, d(X) = -1) + c_- P(Y = -1, d(X) = 1) + r_+ P(Y = 1, d(X) = 0) + r_- P(Y = -1, d(X) = 0) . \quad (6.6)$$

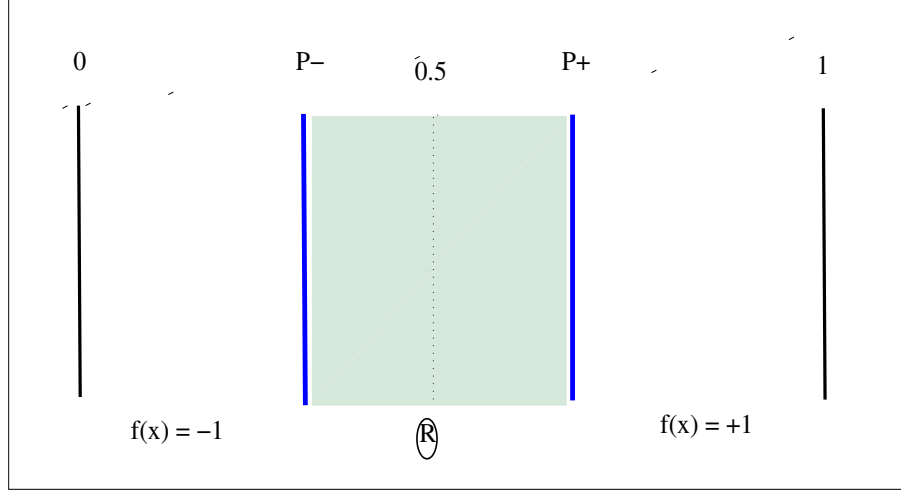


FIG. 6.2 – Illustration de la règle de Chow.

6.3 Formulation du problème

Le problème est basé sur la minimisation du risque structurel (6.6). En classification, ce principe aboutit souvent à un problème d'optimisation de type NP-difficile¹. Puisque la règle de Bayes est donnée par des probabilités conditionnelles, plusieurs classifieurs commencent par estimer la probabilité conditionnelle $\hat{P}(Y = 1|X = \mathbf{x})$, pour l'exploiter dans (6.5) et poser la règle de décision :

$$d^*(\mathbf{x}) = \begin{cases} \text{classer } \mathbf{x} \text{ comme } +1 & \text{si } \hat{P}(Y = 1|X = \mathbf{x}) > p_+ , \\ \text{classer } \mathbf{x} \text{ comme } -1 & \text{si } \hat{P}(Y = 1|X = \mathbf{x}) < p_- , \\ 0 & \text{sinon} . \end{cases} \quad (6.7)$$

6.3.1 Fonction de coût du SVM avec rejet

Pour construire la fonction de perte du SVM avec rejet, nous nous basons sur deux fonctions de pertes standards : la fonction *hinge loss*, très recommandée par V.Vapnik car elle conduit à une solution consistante [Vapnik 1995, Zhang 2004, Grandvalet 2009] et la fonction logistique car elle permet une bonne estimation des probabilités *a posteriori* $P(Y = 1|X = \mathbf{x})$ [Bartlett 2007, Herbei 2006, Grandvalet 2009]. Généralement, une meilleure estimation de cette probabilité est obtenue lorsque la fonction *hinge loss* est tangente à la fonction logistique au point $f = \log(p/(1 - p))$.

Dans une première étape, nous considérons le modèle de régression logistique

¹Problème d'optimisation difficile dont le temps de résolution est exponentiel.

pour la classification binaire où :

$$\widehat{P}(Y = y|X = \mathbf{x}) = \frac{1}{1 + \exp(-yf(\mathbf{x}))} , \quad (6.8)$$

la fonction f est estimée par la minimisation du risque empirique régularisé sur un ensemble d'apprentissage $\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$:

$$\sum_{i=1}^n \ell(y_i, f(\mathbf{x}_i)) + \lambda \Omega(f) \quad (6.9)$$

où ℓ est la fonction de perte et $\Omega(\cdot)$ est le fonctionnelle de régularisation définie comme le carré de la norme de f dans l'espace de Hilbert, $\Omega(f) = \|f\|_{\mathcal{H}}^2$. Pour la regression logistique standard, ℓ est le logarithme négatif du maximum de vraisemblance :

$$\ell(y, f(\mathbf{x})) = \log(1 + \exp(-yf(\mathbf{x}))) . \quad (6.10)$$

La règle de décision (6.5) montre que nous n'avons besoins que des probabilités *a posteriori* $P(Y = 1|X = \mathbf{x})$ aux points de p_- and p_+ . L'algorithme d'apprentissage proposé dans [Grandvalet 2009] poursuit cet objectif sans perdre la consistance de la solution et de la règle de décision. La fonction de coût doit être alors tangente à la fonction de régression logistique en deux points :

$$f_+ = \log \frac{p_+}{1 - p_+} , \quad (6.11)$$

$$f_- = \log \frac{p_-}{1 - p_-} , \quad (6.12)$$

Cette fonction de coût est alors à double coude (voir figure 6.3). Le coût des points correctement classés est remis à zéro.

Double hinge : La fonction de coût satisfaisant les conditions précédentes est donnée par les équations suivantes respectivement, pour les exemples positifs et pour les exemples négatifs.

$$\ell_{p_-, p_+}(+1, f(\mathbf{x})) = \max(- (1 - p_-)f(\mathbf{x}) + H(p_-), - (1 - p_+)f(\mathbf{x}) + H(p_+), 0) ,$$

$$\ell_{p_-, p_+}(-1, f(\mathbf{x})) = \max(p_+f(\mathbf{x}) + H(p_+), p_-f(\mathbf{x}) + H(p_-), 0) .$$

Ces expressions peuvent être reformulées comme suit :

$$\ell_{p_-, p_+}(+1, f(\mathbf{x})) = \begin{cases} - (1 - p_-)f(\mathbf{x}) + H(p_-) & \text{si } f(\mathbf{x}) < \frac{H(p_-) - H(p_+)}{p_+ - p_-} , \\ - (1 - p_+)f(\mathbf{x}) + H(p_+) & \text{si } \frac{H(p_-) - H(p_+)}{p_+ - p_-} \leq f(\mathbf{x}) < \frac{H(p_+)}{1 - p_+} , \\ 0 & \text{sinon} . \end{cases} \quad (6.13)$$

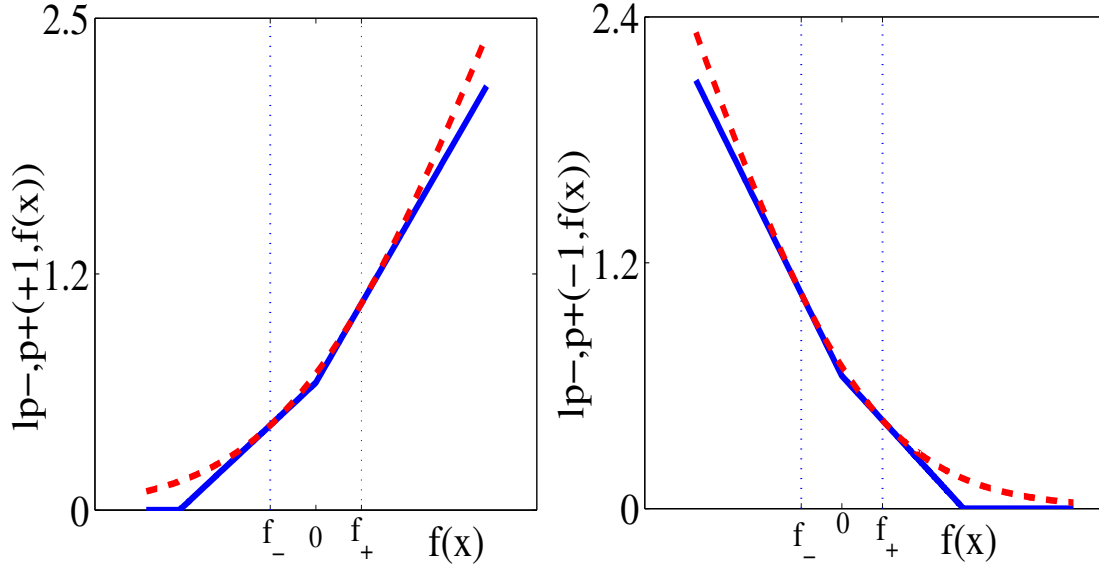


FIG. 6.3 – Fonction de coût pour les exemples positifs (gauche) et pour les exemples négatifs (droite) avec $p_- = 0.35$ and $p_+ = 0.6$ (trait continu : double hinge, trait interrompu : logistique).

et

$$\ell_{p_-, p_+}(-1, f(\mathbf{x})) = \begin{cases} p_+ f(\mathbf{x}) + H(p_+) & \text{si } f(\mathbf{x}) > \frac{H(p_-) - H(p_+)}{p_+ - p_-}, \\ p_- f(\mathbf{x}) + H(p_-) & \text{si } \frac{-H(p_-)}{p_-} \geq f(\mathbf{x}) > \frac{H(p_-) - H(p_+)}{p_+ - p_-}, \\ 0 & \text{sinon.} \end{cases} \quad (6.14)$$

où $H(p) = -p \log(p) - (1 - p) \log(1 - p)$ et $f(x) = w^T x$. Cette fonction de perte est convexe et linéaire par morceau.

Après la phase d'apprentissage, la règle de décision (6.5) peut alors être reformulée comme suit :

$$d_{p_-, p_+}(\mathbf{x}) = \begin{cases} +1 & \text{si } f(\mathbf{x}) > \log\left(\frac{p_+}{1 - p_+}\right), \\ -1 & \text{si } f(\mathbf{x}) < \log\left(\frac{p_-}{1 - p_-}\right), \\ 0 & \text{sinon.} \end{cases} \quad (6.15)$$

6.3.2 Formulation du problème primal

Dans cette section, nous allons montrer comment le problème d'optimisation d'un SVM standard a été modifié en remplaçant la fonction de perte par une fonction de perte à double coude (*double hinge*). Le problème d'optimisation est d'abord écrit sous une forme compactée, après quoi, le problème dual est déduit.

Soit C une constante de pénalisation dont la valeur doit être calculée pendant la phase d'apprentissage, nous définissons $B = C(p_+ - p_-)$, $C_i = C(1 - p_+)$ pour les exemples positifs et $C_i = Cp_-$ pour les exemples négatifs ; le problème d'optimisation est posé comme suit :

$$\min_{f,b} \frac{1}{2} \|f\|_{\mathcal{H}}^2 + \sum_{i=1}^n C_i |t_i - y_i(f(x_i) + b)|_+ + B \sum_{i=1}^n |\rho_i - y_i(f(x_i) + b)|_+ , \quad (6.16)$$

où $|\cdot|_+ = \max(\cdot, 0)$. Pour les exemples positifs, $t_i = \frac{H(P_+)}{1-P_+}$ et $\rho_i = \frac{H(P_-)-H(P_+)}{P_+-P_-}$. Pour les exemples négatifs, $t_i = \frac{H(P_-)}{P_-}$ et $\rho_i = \frac{H(P_-)-H(P_+)}{P_+-P_-}$.

Le problème quadratique (6.16) devient plus lisible lorsque nous introduisons les variables d'écart ξ et η :

$$\begin{cases} \min_{f,b,\xi,\eta} & \frac{1}{2} \|f\|_{\mathcal{H}}^2 + \sum_{i=1}^n C_i \xi_i + B \sum_{i=1}^n \eta_i , \\ \text{S.c} & y_i(f(x_i) + b) \geq t_i - \xi_i \quad i = 1, \dots, n , \\ & y_i(f(x_i) + b) \geq \rho_i - \eta_i \quad i = 1, \dots, n , \\ & \xi_i \geq 0 , \quad \eta_i \geq 0 \quad i = 1, \dots, n . \end{cases} \quad (6.17)$$

6.3.3 Formulation du problème dual

La solution du problème (6.17) est aussi le point selle de son Lagrangien, ce qui nous conduit au problème dual (6.18) :

$$\begin{cases} \max_{\alpha,\gamma} & -\frac{1}{2} \gamma^T G \gamma + (\mathbf{t} - \boldsymbol{\rho})^T \boldsymbol{\alpha} + \boldsymbol{\rho}^T \boldsymbol{\gamma} , \\ \text{S.c} & y^T \gamma = 0 , \\ & 0 \leq \alpha_i \leq C_i \quad i = 1, \dots, n , \\ & 0 \leq \gamma_i - \alpha_i \leq B \quad i = 1, \dots, n , \end{cases} \quad (6.18)$$

où $\mathbf{t} = (t_1, \dots, t_n)^T$ et $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)^T$ sont les vecteurs seuils dans $\mathbb{R}^n$ et G , la matrice d'influence de dimension $n \times n$ et de terme général $G_{ij} = y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$ et $k(x_i, x_j)$, le noyau choisi.

Le problème (6.18) est un problème d'optimisation quadratique sous contraintes de boîte ². Comparé au SVM standard, le dual (6.18) a deux vecteurs à optimiser ($\boldsymbol{\alpha}$ et $\boldsymbol{\gamma}$) mais nous verrons plus loin que le vecteur $\boldsymbol{\alpha}$ peut être facilement recouvert par le vecteur $\boldsymbol{\gamma}$.

6.4 Résolution du problème

L'apprentissage du SVM avec rejet consiste à résoudre le problème d'optimisation (6.18). Pour ce faire, nous utilisons un algorithme à contraintes actives selon

²Les contraintes d'inégalité sont comprises entre une borne inférieure et une borne supérieure, appelées frontières de la boîte.

la stratégie utilisée dans [Loosli 2005] et dans [Vishwanathan 2003] pour l'apprentissage d'un SVM standard. Nous avons adapté cet algorithme au SVM avec rejet en optimisant deux vecteurs. La première étape de la méthode consiste à partitionner l'ensemble d'apprentissage en supports vecteurs et non supports vecteurs et de calculer la fonction de coût correspondante. La deuxième étape évalue la fonction de coût et engendre une nouvelle partition de points. Ces deux étapes sont reproduites d'une manière itérative jusqu'à ce que la fonction de coût soit minimale.

6.4.1 Partitions de l'ensemble d'apprentissage

L'ensemble d'apprentissage est partitionné selon les contraintes du problème (6.17), en 5 sous ensembles (voir table 6.2). La fonction de coût correspondante est indiquée par la figure (6.4).

TAB. 6.2 – *Partition de l'ensemble d'apprentissage*

I_0	partie saturée de la fonction de coût	$I_0 = \{i y_i(f(x_i) + b) > t_i\}$
I_t	première charnière de la fonction de coût	$I_t = \{i y_i(f(x_i) + b) = t_i\}$
I_C	première pente de la fonction de coût	$I_C = \{i \rho_i < y_i(f(x_i) + b) < t_i\}$
I_ρ	deuxième charnière de la fonction de coût	$I_\rho = \{i y_i(f(x_i) + b) = \rho_i\}$
I_B	deuxième pente de la fonction de coût	$I_B = \{i y_i(f(x_i) + b) < \rho_i\}$

Cette partition engendre des conséquences sur les variables duales (α, γ) :

$$\left\{ \begin{array}{ll} i \in I_0 \Rightarrow \alpha_i = 0 & , \gamma_i = 0 \\ i \in I_t \Rightarrow 0 \leq \alpha_i \leq C_i & , \gamma_i = \alpha_i \\ i \in I_C \Rightarrow \alpha_i = C_i & , \gamma_i = C_i \\ i \in I_\rho \Rightarrow \alpha_i = C_i & , C_i < \gamma_i < C_i + B \\ i \in I_B \Rightarrow \alpha_i = C_i & , \gamma_i = C_i + B \end{array} \right. \quad (6.19)$$

En considérant la répartition précédente, les valeurs de γ_i doivent être calculées uniquement pour les indices $i \in I_t \cup I_\rho$. cependant, sur cet ensemble, la variable α_i est constante ou égale à γ_i .

6.4.2 Calcul des variables duales

D'après la répartition de la table 6.2, le problème (6.18) peut être réduit pour ne calculer que les γ_i pour $i \in I_t \cup I_\rho$. Soit $I_D = \{I_C, I_B\}$ et $I_h = \{I_t, I_\rho\}$, le

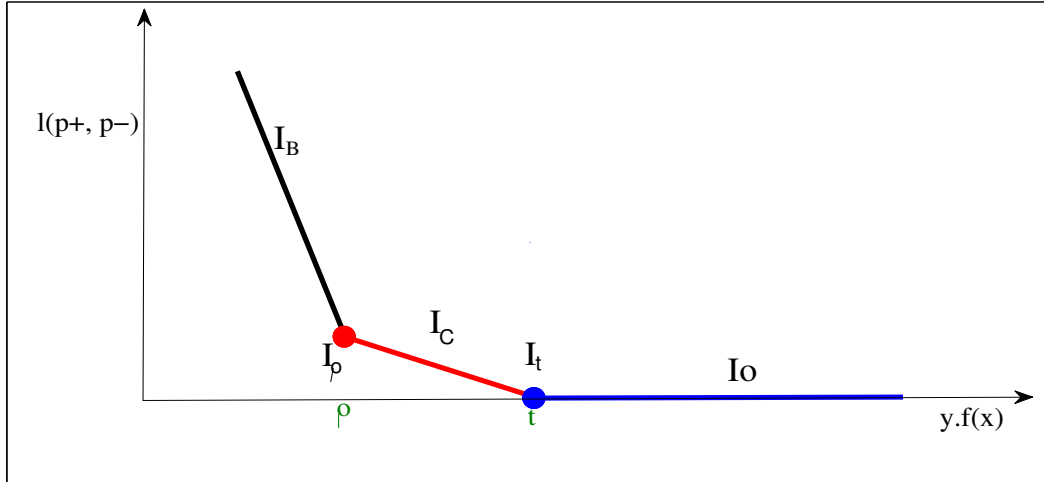


FIG. 6.4 – Correspondance entre les vecteurs d'apprentissage et les partitions de la fonction de coûts (Double hinge loss).

problème (6.18) vu comme un problème de minimisation devient :

$$\begin{cases} \min_{\gamma} & \frac{1}{2} \gamma^T G \gamma - S^T \gamma \\ \text{S.c} & y^T \gamma = 0 \\ & 0 \leq \gamma_i \leq D_i, \quad i = 1, \dots, n \text{ et } D = C_i + B. \end{cases} \quad (6.20)$$

où S regroupe les deux seuils t_i et ρ_i . Pour résoudre le problème dual (6.20), le Lagrangien est encore une fois utilisé.

$$\mathcal{L}(\gamma, \lambda, \mu, \nu) = \frac{1}{2} \gamma^T G \gamma - S^T \gamma - \lambda \gamma^T y - \nu^T \gamma + \mu^T (\gamma - D \mathbb{I}_n) \quad (6.21)$$

avec les multiplicateurs de Lagrange λ , μ et ν positifs ou nuls et $\mathbb{I}_n$ un vecteur unitaire de dimension n . Ce Lagrangien ramène à celui du problème primal (6.17) en exprimant la variable f par γ , le problème (??) devient :

$$\mathcal{L} = \frac{1}{2} \gamma^T G \gamma - b \gamma^T y - (\mathbf{t} - \boldsymbol{\rho})^T \alpha - \boldsymbol{\rho}^T \gamma + \boldsymbol{\xi}^T (\alpha - \nu + C_i \mathbb{I}_n) + \boldsymbol{\eta}^T (\gamma - \omega + B \mathbb{I}_n) \quad (6.22)$$

Pour énoncer les relations entre les variables primales et duales, nous devons considérer les conditions de KKT qui consistent à annuler le gradient du Lagrangien (6.22) par rapport à la variable primale γ dans les ensembles de points I_t, I_ρ, I_C, I_B et I_0 .

La table 6.3 indique les propriétés de chaque ensemble de points de l'ensemble d'apprentissage en considérant les variables primales et duales.

6.4.3 Algorithme d'apprentissage

Nous considérons la répartition des points (I_0, I_h et I_D). Les valeurs de γ non connues sont uniquement celles qui appartiennent à l'ensemble I_h . Ces valeurs

TAB. 6.3 – Situation des contraintes pour les cinq types d'exemples

ens	Contraintes initiales	Contraintes primales	Contraintes duales
I_0	$y_i[f(\mathbf{x}_i) + b] > t_i$	$\xi_i = \eta_i = 0$	$\mu = 0, \nu = G\gamma + by - \mathbf{t} \neq 0$
I_t	$y_i[f(\mathbf{x}_i) + b] = t_i$	$\xi_i = \eta_i = 0$	$\mu = 0, \nu = G\gamma + by - \mathbf{t} = 0$
I_C	$\rho_i < y_i[f(\mathbf{x}_i) + b] < t_i$	$\xi_i \neq 0, \eta_i = 0$	$\mu = 0, \nu = -G\gamma - by + \mathbf{t} = \xi$
I_ρ	$y_i[f(\mathbf{x}_i) + b] = \rho_i$	$\xi_i \neq 0, \eta_i = 0$	$\nu = 0, \mu = G\gamma + by - \boldsymbol{\rho} = 0$
I_B	$y_i[f(\mathbf{x}_i) + b] < \rho_i$	$\xi_i \neq 0, \eta_i \neq 0$	$\nu = 0, \mu = -G\gamma - by + \boldsymbol{\rho} = \eta_i$

doivent être données par la solution du problème d'optimisation suivant dont la dimension est inférieure à celle du problème initial.

$$\begin{cases} \min_{\gamma_h} & \frac{1}{2} \gamma_h^T G_{hh} \gamma_h - S_h^T \gamma_h \\ \text{S.c} & y_h^T \gamma_h + c_D = 0 \end{cases}, \quad (6.23)$$

Avec :

$$\begin{aligned} \gamma_h &= \gamma(I_h), \quad y_h = y(I_h), \quad G_{hh} = G(I_h, I_h), \\ c_D &= \sum_{i \in I_C} C_i y_i + \sum_{i \in I_B} D_i y_i, \\ S_h &= \mathbf{t} - \sum_{j \in I_C} C_j G_{ji} - \sum_{j \in I_B} D_j G_{ji} \text{ si } i \in I_t \text{ et} \\ S_h &= \boldsymbol{\rho} - \sum_{j \in I_C} C_j G_{ji} - \sum_{j \in I_B} D_j G_{ji} \text{ si } i \in I_\rho. \end{aligned}$$

Les conditions de Kuhn-Tucker nous donnent le système à résoudre pour trouver les valeurs de γ non encore connues.

$$\begin{cases} G_{hh} \gamma_h = S_h - y_h^T \lambda \\ y_h^T \gamma_h = -c_D \end{cases}, \quad (6.24)$$

Après la résolution de ce système, les composantes de γ qui violeraient les contraintes primales ou duales doivent être déplacées vers un ensemble approprié. Cette itération doit être reproduite jusqu'à la satisfaction des contraintes de boîte, ce qui correspond alors à la minimisation du problème (6.23). L'algorithme de résolution est résumé comme suit :

Algorithme

1. Introduire l'ensemble $\{x_i, y_i\}$ et les paramètres C , p^+ et p^- .
2. Initialiser les ensembles (I_0 , I_h et I_D).
3. Résoudre le problème (6.24) $\rightarrow \gamma^{r+1}$ et $b = \lambda$.
4. Pour tout point γ_i^{r+1} ne satisfaisant pas les contraintes de boîte (6.18), on calcule le plus grand pas $\Delta\gamma_i$ qui nous permet de le projeter sur les bords de la boîte. Ce point est déplacé de I_h vers I_D ou I_0 . $\gamma^r = \gamma^{new}$. $\rightarrow$ étape 3.
Soit $i \in \{I_0 \cup I_D\}$,
 - si $i \in I_0$ et $\nu_i \leq 0$, i est déplacé vers I_t , $\gamma^r = \gamma^{new}$, $\rightarrow$ étape 3
sinon
 - si $i \in I_C$ et $\xi_i \leq 0$, i est déplacé vers I_t , $\gamma^r = \gamma^{new}$, $\rightarrow$ étape 3
sinon
 - si $i \in I_B$ et $\eta_i \leq 0$, alors :
 - si $\xi_i \leq 0$, i est déplacé vers I_t , $\gamma^r = \gamma^{new}$, $\rightarrow$ étape 3
sinon
 - si $\xi_i > 0$, i est déplacé vers I_ρ , $\gamma^r = \gamma^{new}$, $\rightarrow$ étape 3
sinon le coût minimal est atteint.
5. fin $\mapsto \gamma$, b .

6.5 Validation de l'algorithme

Pour tester l'algorithme précédent, nous l'avons appliqué à une base de données synthétiques générée sous Matlab par la fonction "datasets". Cette base contient deux classes de données non balancées réparties selon une distribution choisie dans la fonction "datasets". Pour l'apprentissage, nous avons utilisé un noyau gaussien et un paramètre de pénalisation $C = 10$ et un rapport $\frac{c_+}{c_-} = 1$ (coûts symétriques).

Le résultat obtenu en faisant une classification avec rejet est donnée en figure 6.5 en projetant l'espace des caractéristiques sur un espace à deux dimensions, plus explicite. Nous constatons sur cette figure que les frontières de décision sont symétriques par rapport à l'hyperplan optimal tels que $f_+ = 0.40547$ et $f_- = -0.40547$.

La figure 6.6 indique un exemple où nous avons affecté des coûts asymétriques aux deux classes en considérant $c_+ > c_-$. Dans ce cas, nous avons obtenu $f_+ = 32277$ et $f_- = -0.8473$. Nous rejetons ainsi plus de faux négatifs car leur coût est considéré plus élevé.

6.6 Conclusion

A travers ce chapitre, nous avons exposé une extension aux Support Vector Machines en introduisant une option de rejet d'ambiguïtés afin d'optimiser le coût de la classification. Le rejet est introduit pendant la phase d'apprentissage en faisant

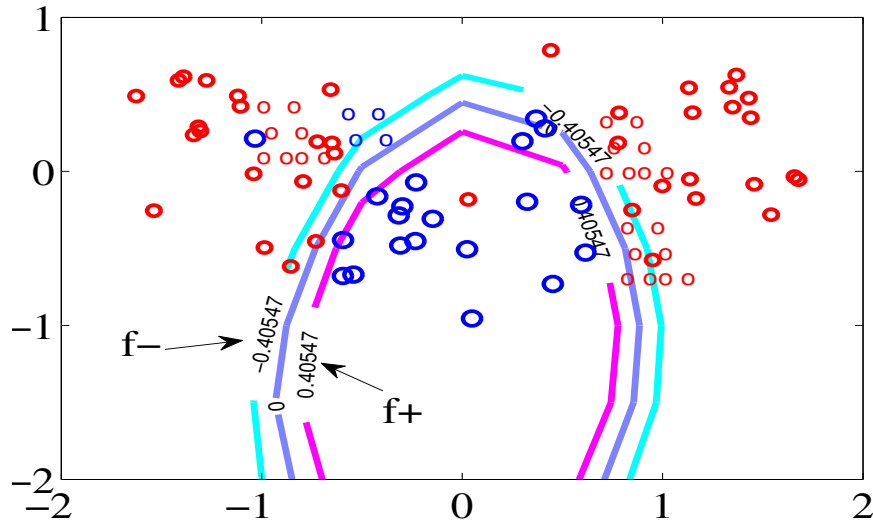


FIG. 6.5 – Résultat obtenu en appliquant l'algorithme de classification symétrique avec rejet sur une base de données générée sous Matlab.

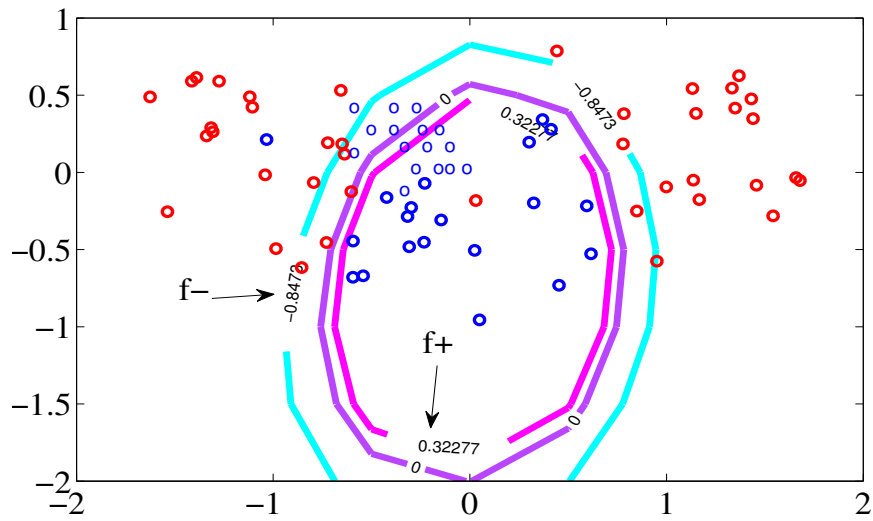


FIG. 6.6 – Résultat obtenu en appliquant l'algorithme de classification asymétrique avec rejet sur une base de données générée sous Matlab.

appel aux deux principales fonctions de coûts des SVM, la fonction "hinge loss" offrant une solution consistante et la fonction "*logistic*" qui permet une meilleure estimation des probabilités *a posteriori* p_+ et p_- [Herbei 2006] ce qui rend la méthode plus robuste.

L'algorithme d'apprentissage proposé repose donc sur une fonction de pénalisation convexe et linéaire par morceau (double hinge loss). La résolution de ce nouveau problème quadratique nous a permis d'effectuer une classification avec des frontières de décision f_+ et f_- flexibles suivant les coûts respectifs des faux négatifs, des faux positifs et des points rejetés. La validation de cet algorithme est clairement indiquée par les figures 6.5 et 6.6.

Validation expérimentale

7.1 Introduction

Dans ce chapitre, nous décrivons notre base d'étude ainsi que les enregistrements choisis puis les résultats obtenus avec les algorithmes que nous avons développés dans les chapitres précédents. L'algorithme de détection des complexes QRS est appliqué à la totalité de la base MITDB et à quelques signaux de la base AHA afin de localiser les battements. Les vecteurs caractéristiques sont ensuite calculés pour chaque battement segmenté. L'algorithme de classification par SVM avec rejet décrit au sixième chapitre est appliqué à la base de test. Les résultats obtenus sont commentés et comparés à ceux obtenus dans des travaux existant dans la littérature.

7.2 Base de données

Les enregistrements ECG que nous avons utilisés dans cette étude proviennent principalement de la base de données internationale MIT-BIH (Massachusetts Institute of Technology/Beth Israel Hospital) arrhythmia data base [Mark 1988] disponible sur le lien (www.physionet.org/physiobank/mitdb).

Cette base est un ensemble de 48 enregistrements de 30 minutes chacun provenant d'un moniteur cardiaque Holter. Elle contient 23 enregistrements numérotés entre 100 et 124 pour le premier groupe ; et de 25 enregistrements numérotés entre 200 et 234 pour le deuxième groupe. Le premier groupe est prévu pour servir d'échantillon représentatif de variété de formes d'ondes qu'un détecteur d'arythmie pourrait rencontrer dans l'utilisation clinique courante ; tandis que le deuxième groupe est choisi pour inclure une variété de cas pathologiques. Les sujets étaient 25 hommes âgés de 32 à 89 ans, et 22 femmes âgées de 23 à 89 ans.

La base MITDB contient 116.137 battements annotés : Chaque battement (complexe QRS) est décrit par une étiquette (annotation) indiquant la position temporelle du pic R ainsi que le type de battement.

Chaque enregistrement est recueilli sur deux pistes correspondant aux dérivations D_2 et V_5 . Nous avons utilisé la dérivation D_2 dans toutes nos analyses vues ses caractéristiques représentatives pour identifier les battements du cœur d'une part et d'autre part, l'utilisation d'un canal unique nécessite moins d'espace mémoire.

Les signaux sont échantillonnés avec une fréquence de 360 Hz et quantifiés à une résolution de 12 bits. Les enregistrements contiennent plusieurs bruits, artefacts, différents types de battements : ventriculaires, supra-ventriculaires, jonctionnels et plusieurs autres anomalies de conduction, c'est pourquoi, elle est très souvent utilisées pour valider tout algorithme lié aux arythmies cardiaques et aux artefacts.

Parce qu'il est un système interactif et convivial de calcul numérique et de visualisation graphique, Matlab est l'environnement sous lequel, nous avons développé nos traitements. Le chargement du signal ECG sous Matlab constitue l'étape initiale dans notre algorithme. Il s'agit de convertir les données codées sous la forme initiale en un autre format compréhensible par Matlab. En effet, chaque enregistrement MITDB est recueilli à l'aide de trois fichiers (*.dat, *.hea et *.anot). Après conversion, un enregistrement est totalement décrit par un simple fichier (*.mat).

7.3 Résultats de la détection des complexes QRS

L'algorithme de détection des complexes QRS que nous avons développé est désigné pour opérer avec des fréquences d'échantillonnage de 360Hz mais aussi avec 250 Hz. Il est alors testé sur quelques enregistrements de la base AHA disponibles sur (www.physionet.org) et sur l'ensemble des enregistrements de la base MITDB à l'exception de quelques intervalles de l'enregistrement " MITDB 207 " présentant des Fibrillations ventriculaires (VF), soit au total 142.5 s d'enregistrement sont exclues. Notons que ces intervalles ne sont pas annotés.

Les performances d'un algorithme sur une base de données standard ne sont pas uniquement une réponse à une utilité clinique mais permettent également la comparaison entre la robustesse de différents algorithmes. Un autre avantage de ces bases est qu'elles couvrent un grand nombre de pathologies, d'artefacts ainsi qu'un grand éventail de types de battements ce qui nous permet de valider la détection des ondes R pour un grand nombre de cas réels.

7.3.1 Détection de battements atypiques avec artefacts

Pour montrer la robustesse de notre détecteur, nous avons opté pour l'illustration de quelques segments d'ECG bruts contenant plusieurs artefacts avec des battements QRS atypiques. Ces résultats peuvent être retrouvés dans [Zidelmal 2012a]. La figure 7.1 indique un segment de l'enregistrement 107 contenant des battements extrasystoliques (PVC) suivis d'ondes T très élevées, parfois de même amplitude que l'onde R. L'effet de l'onde T est nettement atténué sur le signal de localisation h et les pics R sont correctement détectés.

Les fluctuations de la ligne de base couvrent une bande de fréquences allant de 0 Hz à 2.5 Hz [Thakor 1984]. Cette bande est complètement éliminée lorsque nous avons considéré les coefficients d_n^4 et d_n^5 couvrant la bande [5.62 Hz, 22.5 Hz].

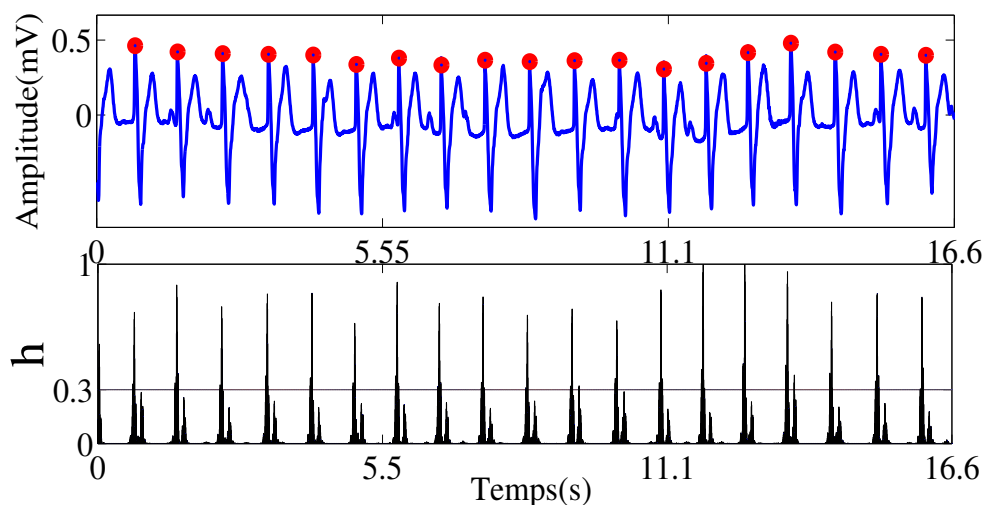


FIG. 7.1 – Segment de l'enregistrement MITDB : 107 $[0\text{ s}, 16.6\text{ s}]$ avec les points rouges représentant les positions des pics R détectés (haut); produit h des coefficients détails avec son seuillage (bas).

Ceci est très visible sur la figure 7.2 lorsque l'algorithme de détection est appliqué à l'enregistrement MITDB : 222.

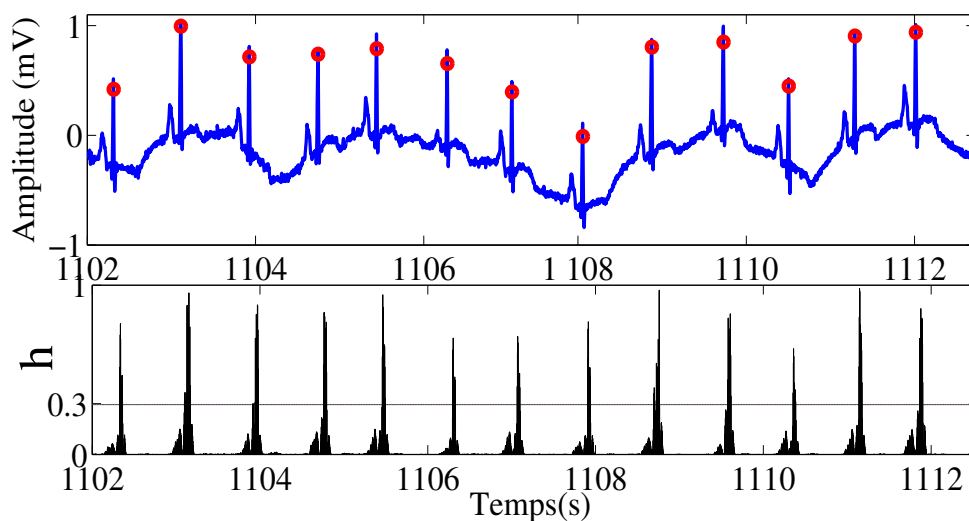


FIG. 7.2 – Détection des pics R sur un segment de l'enregistrement MITDB : 222 $[1102\text{ s}, 1112\text{ s}]$.

La figure 7.3, indique un segment de l'enregistrement 203 prélevé sur un homme âgé de 43 ans. Ce segment présente des changements brusques de l'amplitude de l'onde R, des fluctuations de la ligne de base, des extrasystoles suivies d'ondes T

très importantes (battements encerclés). Sur ce segment, les ondes R sont détectées avec le seuil primaire $\lambda = 0.3 \cdot \max(h)$.

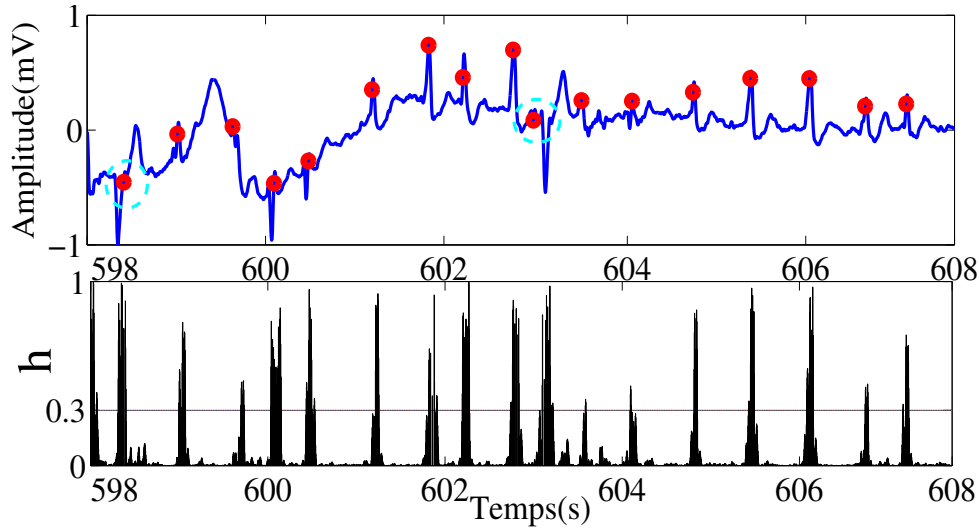


FIG. 7.3 – *détection des ondes R (points rouge sur un segment de l'enregistrement MITDB : 203 [598 s, 608 s] présentant des battements et des inter-battements atypiques.*

Pour montrer l'efficacité de notre algorithme lors de la présence de bruits musculaires, nous avons considéré un segment de l'enregistrement MITDB : 104 (ECG d'une patiente de 66 ans), pris entre 310 s et 320 s (voir figure 7.4). Sur ce segment, nous constatons un sévère bruit musculaire, parfois de même amplitude que l'onde R notamment entre 318 s et 320 s. L'effet de ces bruits est nettement supprimé sur le paramètre de localisation h et visiblement, tous les pics R sont détectés avec le premier seuil λ .

Lors des enregistrements, la ligne de base peut chuter brusquement dans certains cas. Ce bruit est très difficile à supprimer car il appartient à la même bande de fréquence que les complexes QRS. Un tel artefact est présent sur un segment de l'enregistrement 101 entre 174 s et 180 s. Nous constatons alors sur la figure 7.5 que cette chute de tension brusque engendre des coefficients de détails très élevés, ce qui induit des erreurs de détection, deux Faux Positifs et un Faux négatif.

La base AHA n'étant pas aussi bruitées et ne contient pas autant de battements atypiques que la base MITDB, les chercheurs ne l'ont donc pas beaucoup utilisée pour évaluer leurs algorithmes. Nous nous contentons d'illustrer un exemple de détection sur des segments contenant des battements ectopiques suivis d'ondes T très pertinentes (figures 7.6 et 7.7). Sur les deux enregistrements présentés, nous avons obtenu 100% de bonne détection.

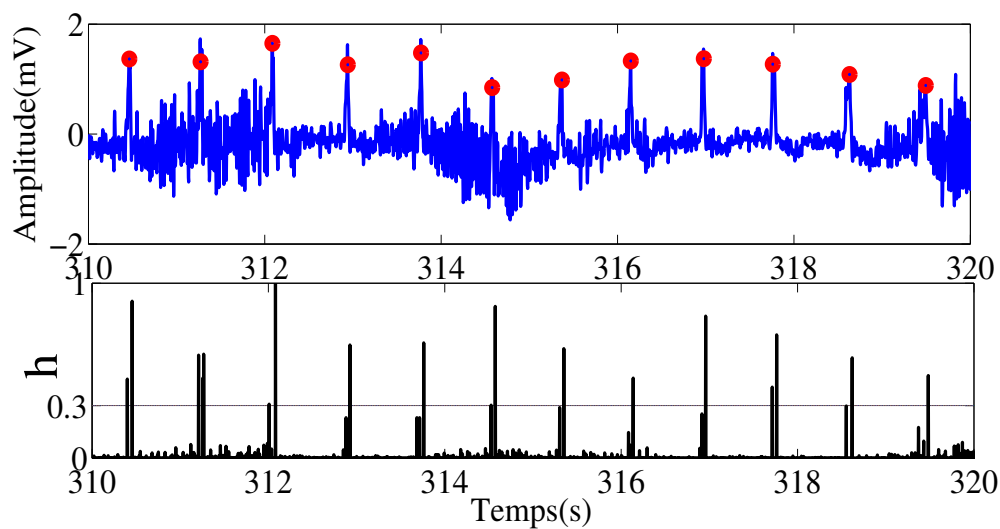


FIG. 7.4 – Résultat de la détection des pics R sur le segment MITDB : 104 [310s, 320s].

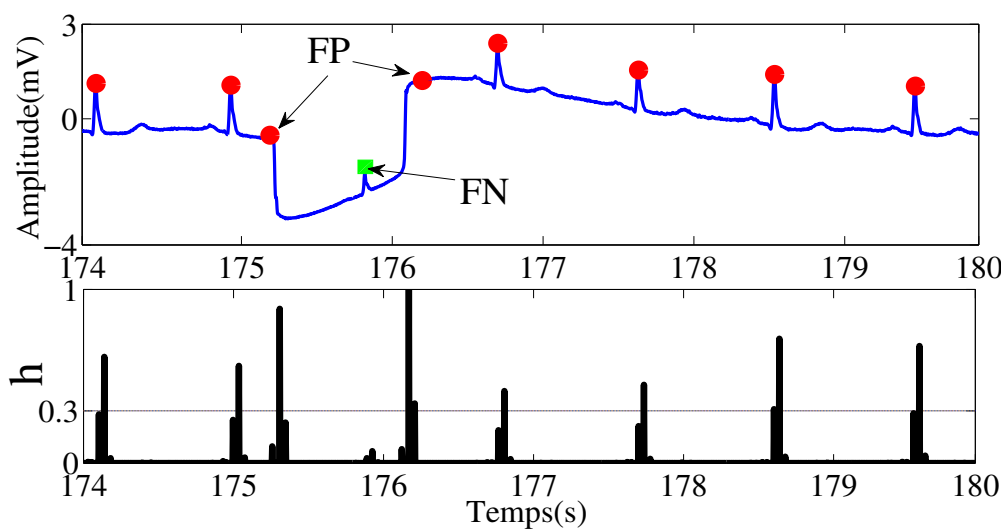


FIG. 7.5 – Segment de l'enregistrement MITDB : 101 [174s, 180s] indiquant un cas de chute brusque de la ligne de base.

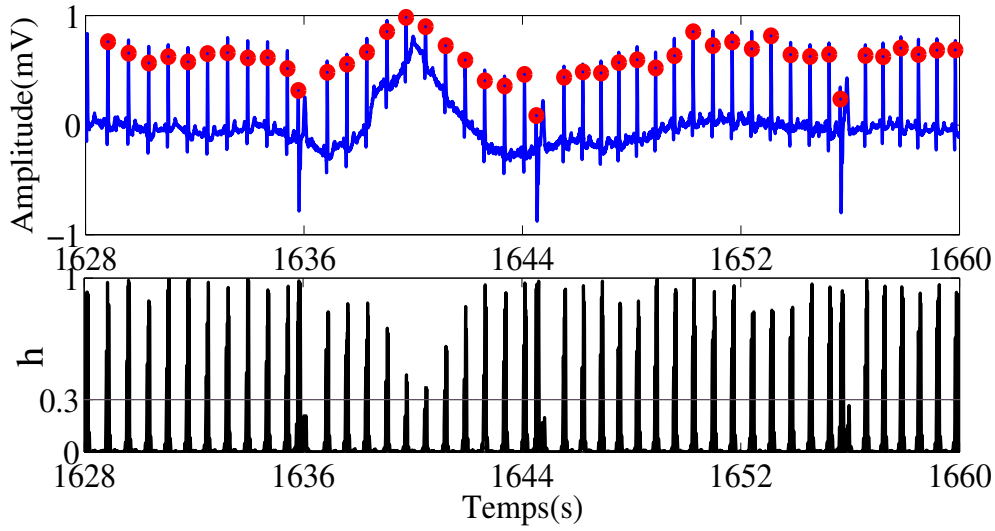


FIG. 7.6 – Résultats de la détection des pics R sur le segment AHA : 0001 [1628 s, 1660 s]. Nous remarquons que les trois battements PVC sont correctement détectés malgré qu'ils sont suivis chacun d'une onde T plus importante car l'effet de cette dernière est très amorti sur le signal de localisation h .

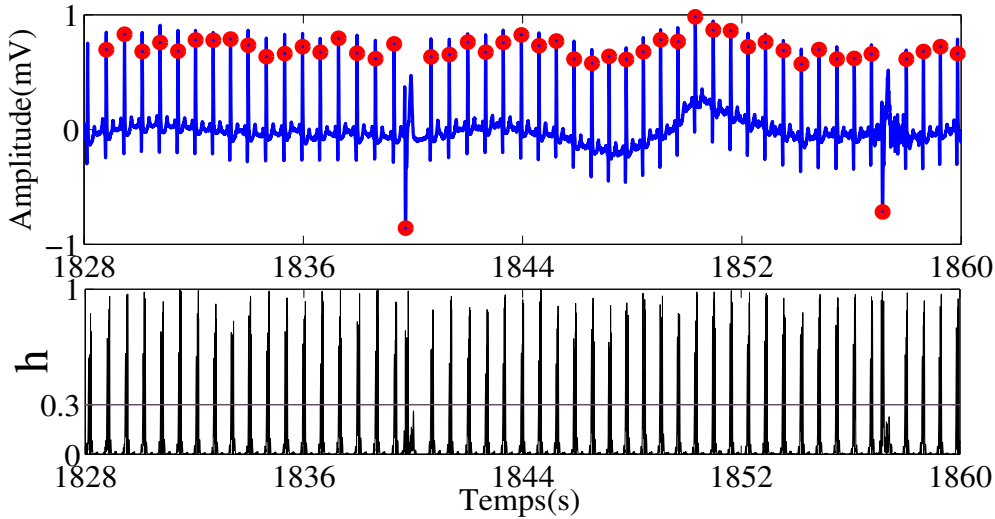


FIG. 7.7 – Résultats de la détection des pics R sur le segment AHA : 0201 [1828 s, 1860 s]. Nous remarquons que les deux battements PVC sont inversés mais correctement détectés.

7.3.2 Résultats statistiques de la détection

Pour juger la qualité d'un algorithme de ce type et pouvoir comparer les résultats à ceux obtenus par d'autres chercheurs, quatre grandeurs sont habituellement mesurées pour chaque enregistrement :

- NTA (Nombre Total Analysé), qui est le nombre de battements analysés par l'algorithme
- FP, qui est le nombre de Faux Positifs : c'est le nombre d'ondes R qui ont été détectées par l'algorithme alors qu'elles ne font pas partie de cette catégorie : ces erreurs peuvent correspondre à des emplacements repérés par l'algorithme alors qu'il n'y avait aucune onde caractéristique, ou encore à une onde repérée comme R alors qu'il s'agit d'une autre onde caractéristique ou d'un artefact.
- FN, qui est le nombre de Faux Négatifs : c'est le nombre d'ondes étiquetées R que l'algorithme n'a pas détectées.
- TP, qui est le nombre de pics R correctement détectés.

Ces grandeurs nous permettent le calcul de certains paramètres statistiques qui sont la sensibilité S_e , la prédictivité positive P_p et le taux d'erreurs E_r respectivement donnés par les équations suivantes et récapitulés par la table 7.2.

$$S_e = \frac{TP}{TP + FN} \quad (7.1)$$

$$P_p = \frac{TP}{TP + FP} \quad (7.2)$$

$$E_r = \frac{FN + FP}{NTA} \quad (7.3)$$

La sensibilité est utilisée pour évaluer la capacité de l'algorithme à détecter correctement les pics R, la prédictivité positive est utilisée pour évaluer la capacité de l'algorithme à séparer les vrais positifs (TP) des faux positifs (FP) et le taux d'erreurs est utilisé pour évaluer le score de l'algorithme. Généralement, un battement est considéré comme correctement détecté (TP) si la position détectée est décalée à moins de 100mS de la position annotée [Meyer 2006]. Un exemple de positions décalées a été indiqué sur la figure 7.3, ces décalages sont nettement inférieurs à l'intervalle cité dans [Meyer 2006]. La table (7.2) montre que sur la totalité des enregistrements de la base MITDB, notre algorithme assure un score de 99,46%, une sensibilité de 99,64% et une prédictivité positive de 99,82%. Ces résultats sont comparables ou meilleurs que ceux publiés dans la littérature (voir Table 7.1).

7.4 Résultats de la classification

Notre objectif étant d'évaluer les performances d'un classifieur binaire avec rejet, nous avons d'abord retenu les annotations de " American Hearth Association

TAB. 7.1 – *Comparaison avec d'autres algorithmes*

Méthode	$S_e(\%)$	$P_p(\%)$	$Er(\%)$
Notre algorithme	99,64	99,82	0,54
[Pan 1985]	99,75	99,54	0,71
[Adnane 2010]	99,66	99,80	0,54
[Chen 2006]	99,47	99,54	0,98

(AHA) " [Mark 1988] et les recommandations pratiques de la norme AAMI pour former deux classes de battements (battements normaux et battements ectopiques)

- La classe positive (P) représente les battements ectopiques, les battements prématurés et quelques battements inconnus.
- La classe négative (N) représente les battements normaux (environ 70 % de la base d'étude) et quelques battements anormaux LBBB (Left Bundle Branch Block), RBBB (Right Bundle Branch Block).

D'après les recommandations pratiques de la norme AAMI (American Association for Medical Instrumentation), les enregistrements (102, 104, 107, 217) contenant des battements provenant de pacemakers sont exclus de cette étude. Les enregistrements ne contenant pas de battements PVC (11 enregistrements) sont également exclus [Mark 1987]. Il nous reste alors 33 enregistrements d'intérêt. Notons qu'aucune sélection de signaux n'est basée sur leur qualité.

7.4.1 Base d'apprentissage

Pour constituer une bonne base d'apprentissage qui nous permettrait la généralisation de notre classifieur et d'obtenir ainsi un classifieur global, nous avons pris d'une manière aléatoire 10 enregistrements desquels nous avons utilisé les battements présents dans les 5 premières minutes, soit 1500 battements étiquetés par des cardiologues.

Les battements cardiaques étant segmentés et quantifiés par des paramètres discriminant, sont représentés chacun par un vecteur de caractéristique déjà décrit au quatrième chapitre.

7.4.2 Base de test

De chaque enregistrement, nous utilisons les 25 dernières minutes pour la phase de test. Ainsi, la base d'apprentissage est complètement dissociée de la base de test. Ceci nous permet d'évaluer la capacité de généralisation de notre algorithme de classification.

7.4.3 Protocole expérimental

Les conditions sous lesquelles doivent se dérouler les tests de classification doivent répondre à certains critères. Ainsi tous les test doivent être effectués de la même manière.

7.4.3.1 Normalisation des données

Les données d'apprentissage et de test sont structurées sous forme de matrices. Le nombre de lignes représente le nombre de battements (ou d'exemples x_i) et le nombre de colonnes représente le nombre de caractéristiques y_j quantifiant chaque battement. Chaque ligne représente donc un vecteur caractérisant un battement.

Ces caractéristiques peuvent être très différentes d'un exemple à un autre. Leur normalisation est par conséquent nécessaire. Cette opération consiste à ramener chaque caractéristique (chaque colonne de la matrice de données) à suivre une distribution gaussienne réduite c'est-à-dire de moyenne nulle et d'écart type égal à l'unité. La formule de normalisation est donnée par :

$$\tilde{y}_{ij} = \text{tansig}\left(\frac{y_{ij} - \bar{y}_j}{\sigma_{yj}}\right) \quad (7.4)$$

Où y_{ij} est la i^{me} composante de la j^{me} caractéristique. $\bar{y}_j$ et σ_{yj} sont la moyenne et l'écart type de la j^{me} composante du vecteur de caractéristiques et $\text{tansig}(n) = 2/(1 + \exp(-2n)) - 1$.

7.4.3.2 Choix du noyau

Dans les méthodes à noyau, comme les SVMs, les observations x_i sont implicitement projetées dans un espace de caractéristiques de plus grande dimension dans lequel le problème devient linéairement séparable. Ainsi, chaque produit scalaire $\langle x_i, x_j \rangle$ est remplacé par son correspondant $k(x_i, x_j)$ dans le nouvel espace, k est la fonction noyau (voir Chapitre 5).

Cette propriété montre à la fois l'importance et la difficulté du choix du noyau dans les SVMs car il n'existe pas de méthodes permettant un choix *a priori*. Généralement, plusieurs noyaux sont testés afin de sélectionner celui qui minimiserait le coût de la classification. Pour la présente application, nous avons retenu le noyau gaussien.

7.4.3.3 Sélection des hyper-paramètres par validation croisée

La précision d'un modèle SVM est largement tributaire de la sélection des paramètres du noyau et des paramètres de pénalisation. Les paramètres optimaux sont appelés hyper-paramètres. Dans notre cas, les paramètres à optimiser sont : le paramètre de pénalisation C (voir paragraphe 6.3.2) et la largeur du Kernel exprimée par σ . Ils sont déterminés par une technique de validation croisée. La

validation croisée se décline en plusieurs sous-méthodes. La plus répandue est la méthode " k -Fold " avec typiquement $k \in [4, 10]$. Elle consiste à partager un ensemble de validation composé de n exemples en k parties égales. On procède alors à k expériences d'apprentissage à partir de $\frac{k-1}{k}n$ exemples. Les $\frac{n}{k}$ exemples restants servent au test des règles apprises et à l'évaluation du résultat. Une Cross-Validation avec $k = 3$ est illustrée par la figure 7.8. Cette technique est appliquée

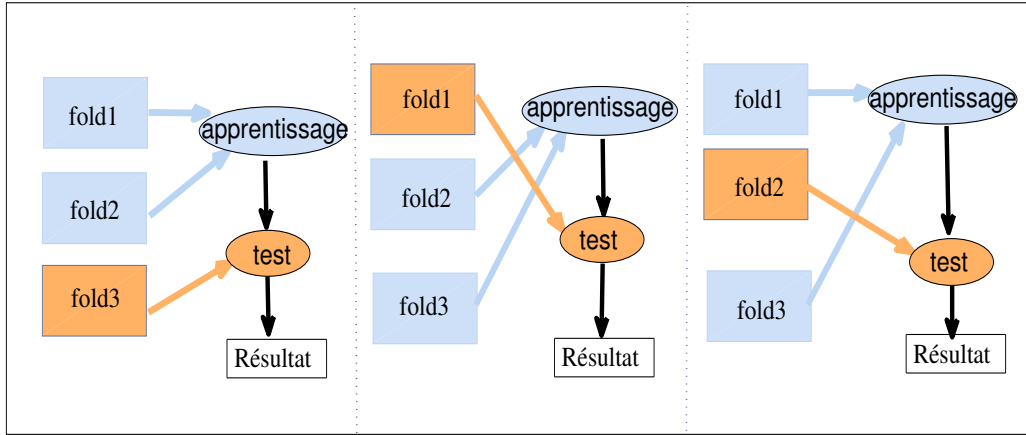


FIG. 7.8 – Procédure de validation croisée avec $k = 3$.

selon l'algorithme suivant :

1. diviser l'ensemble de validation (EV) en k sous ensembles équilibrés (même nombre d'exemples positifs et même nombre d'exemples négatifs).
2. Résoudre le problème d'optimisation (Apprentissage) sur l'union des $k - 1$ sous ensembles.
3. Tester sur le k^{eme} sous ensemble et calculer l'erreur quadratique moyenne.
4. Refaire k fois les étapes 3 et 4.
5. La division de EV engendrant l'erreur minimal doit être testée sur une grille de valeurs C et σ .

Le couple (C, σ) correspondant à l'erreur minimale est retenu comme couple d'hyper-paramètres.

7.4.4 Classification à coûts symétriques

Comme la base de données n'est accompagnée d'aucune matrice de coûts, nous supposons alors que le rejet d'un exemple positif et le rejet d'un exemple négatif ont des coûts identiques $r_+ = r_- = r$ (voir table 1, Chapitre 6). Supposons aussi

dans un premier test qu'un exemple positif mal classé (FN) et un exemple négatif mal classé (FP) sont pénalisés de la même façon c'est-à-dire : $c_+ = c_- = c$ ou encore $\frac{c_+}{c_-} = \beta = 1$. Dans ce cas là, $p_+ = 1 - p_-$.

Les SVMs sont souvent non linéaires dans l'espace d'entrée. Ils sont vus comme une "boîte noire" par l'utilisateur vue la grande dimension de l'espace des représentations. Pour visualiser la classification effective et montrer la région de rejet, nous avons projeté l'espace des représentations sur un espace à deux dimensions. Nous pouvons alors voir sur la figure 7.9 la région de rejet induite par notre classifieur lorsque nous avons effectué un test sur l'enregistrement MITDB : 106 ne contenant que deux types de battements (Normaux et PVC). Cette figure montre que les deux types de battements sont nettement séparés et indique clairement les seuils de rejet évoqués dans l'équation (6.13). Ses seuils sont : $f_+ = 0.75377$ et $f_- = -0.75377$. L'hyperplan optimal est $f = 0$.

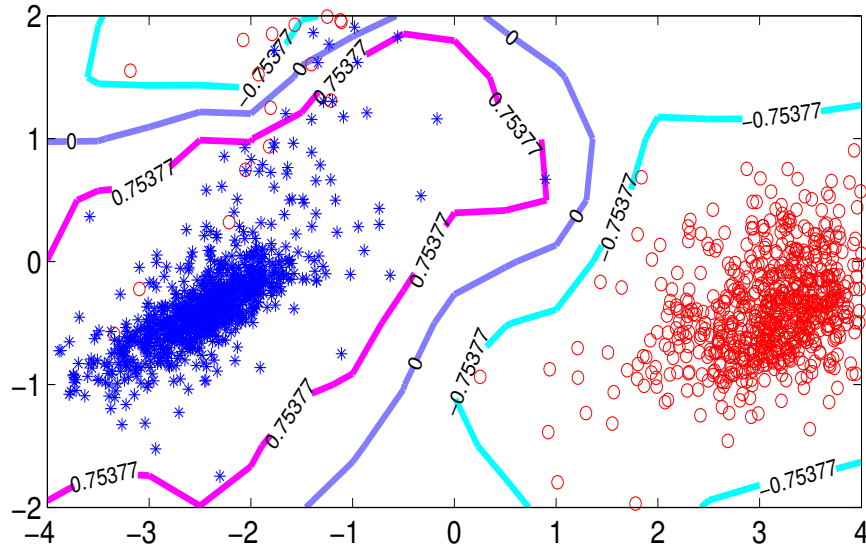


FIG. 7.9 – Région de rejet induite par les seuils de rejet fixés avec les coûts des rejets ($r = 0.35$ et le rapport des coûts des fausses classifications $\beta = 1$). Les exemples représentés par (o rouge) représentent la classe négative. Les exemples représentés par (* bleue) représentent la classe positive.

7.4.5 Classification à coûts asymétriques

Souvent, en pratique, un exemple FN est plus pénalisé qu'un exemple FP. En particulier lorsqu'il s'agit d'un diagnostic médical où nous devons encourager le rejet de FN lorsque la probabilité de leur appartenance à une classe est proche de 0.5. *Ceci est l'objectif même de notre approche.*

La figure 7.10 indique la région de rejet induite par les seuils obtenus avec des coûts asymétriques. Nous constatons sur cette figure que la règle de décision de Bayes a été appliquée avec des seuils de décision $f_+ = 0.4726$ et $f_- = -0.75377$ et que la région de rejet est plus large du côté négatif de l'espace projeté. Nous rejetons alors plus de FN que de FP.

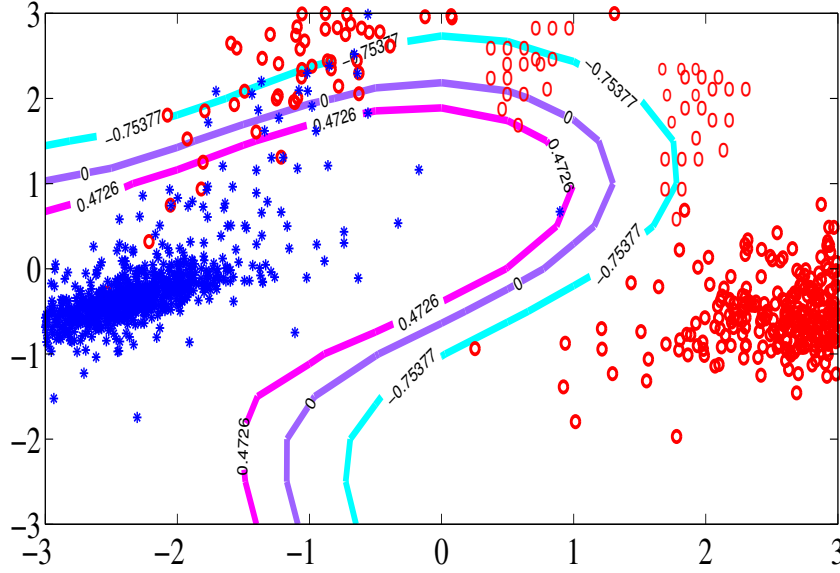


FIG. 7.10 – région de rejet induite par les seuils de rejet fixés avec les coûts asymétriques ($r = 0.35$ et $\beta = 1.2$). Les exemples représentés par $\circ$ appartiennent à la classe négative. Les exemples représentés par $*$ appartiennent à la classe positive.

Dans tous les travaux introduisant du rejet dans les classifieurs [Chow 1970, Chow 1957, Herbei 2006], les auteurs ont considéré que pour une description efficace des performances d'un système de reconnaissance, il faut considérer le compromis erreur-rejet car rejeter trop d'exemples n'est tout de même pas intéressant.

Soit E et R respectivement le taux d'erreurs et le taux de rejet. Ces fonctions monotones de r nous permettent de calculer le compromis erreur-rejet à partir de $E(r)$ et $R(r)$ en faisant varier le coût de rejet r entre 0.5 et 0.1 comme dans [Bartlett 2008]. Le résultat moyen obtenu avec les enregistrements sélectionnés, est représenté par la figure 7.11. Le taux d'erreur diminue jusqu'à un taux pratiquement constant. Une autre courbe intéressante sur la même figure représente le rapport entre la variation de l'erreur et la variation du rejet en fonction du coût de rejet r . Cette fonction a été définie dans [Chow 1970] par :

$$V_{E/R} = \frac{E(r - \Delta r) - E(r)}{R(r - \Delta r) - R(r)}$$

Cette dernière représente la pente de la courbe de compromis erreur/rejet. Cette pente négative croît jusqu'à une valeur presque nulle correspondant à un taux d'erreur constant. Le point d'inflexion sur cette courbe indique la plus grande diminution de l'erreur.

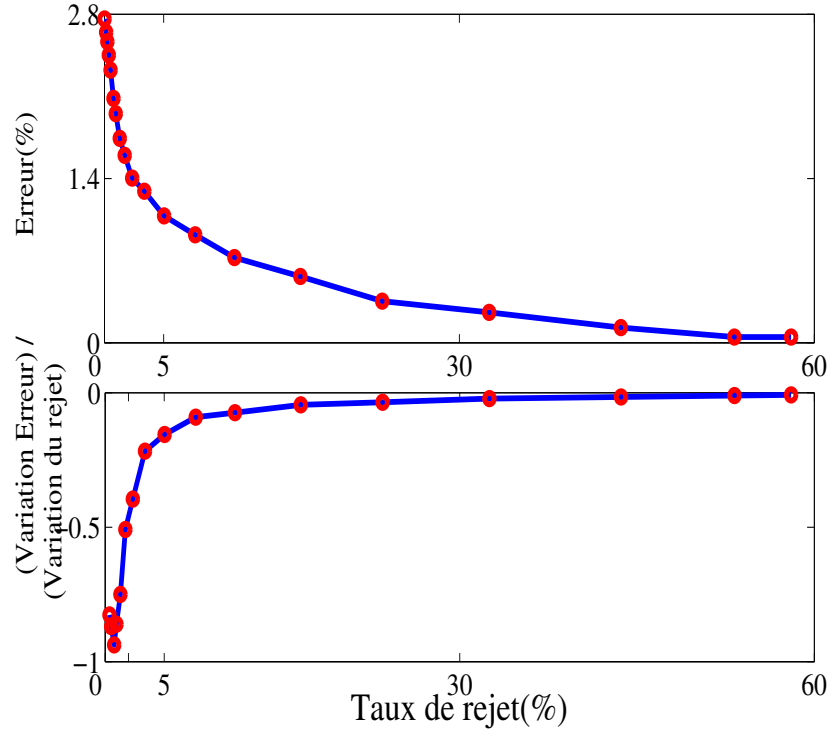


FIG. 7.11 – *Courbe de compromis erreur-rejet (haut); Pente de la courbe de compromis (bas).*

Pour comparer nos résultats avec ceux exposés dans la littérature, nous avons utilisé deux paramètres statistiques : La sensibilité et la prédictivité positive du classifieur. Ces deux paramètres sont présentés en figure 7.12. Pour mieux illustrer l'objectif de approche qui consiste à optimiser le coût de la classification, nous avons représenté sur la même figure la variation du coût global de la classification donné par :

$$C_c = c_+FN + c_-FP + rR$$

Cette figure montre une bonne sensibilité du classifieur. Nous avons obtenu plus de 98% de sensibilité pour une classification sans rejet (0%) et plus de 99% de sensibilité pour le coût optimal de classification correspondant à un rejet de 1.8%. Notons que les seuils de rejet sont flexibles et définis automatiquement selon les coûts des Faux Positifs, des Faux Négatifs et des

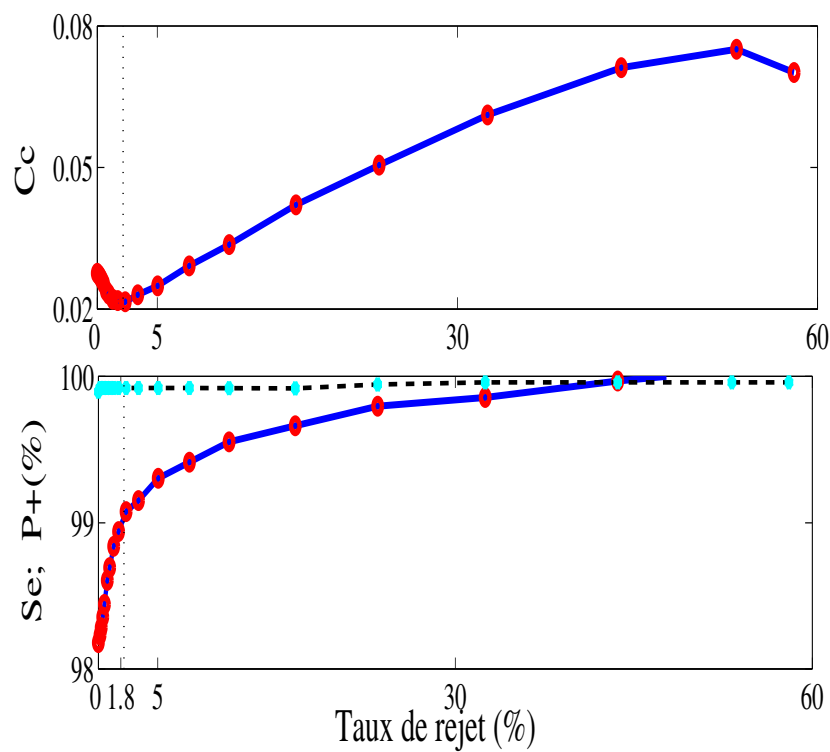


FIG. 7.12 – Coût global de la classification en fonction du taux de rejet (haut); Sensitivité et Prédicativité positive du classifieur en fonction du taux de rejet (bas).

exemples rejetés. Il est clair que pour la base d'étude considérée, le meilleur taux de rejet est de 1.8% correspondant à un coût global minimal. Les résultats obtenus sont comparables à ceux publiés par certains auteurs : *eg.* [Afonso 1999, Gomez 2006, Krasteva 2007, Christov 2005] obtenant respectivement 85.3%, 97.7%, 98.4% et 96.7% de sensibilité.

7.5 Conclusion

Dans ce chapitre nous avons présenté et discuté les résultats obtenus par les méthodes proposées : la détection des complexes QRS puis la classification avec rejet d'ambiguïtés.

La détection des ondes R est basée sur la densité spectrale de puissance des complexes QRS à différents niveaux de résolution. Une telle méthode est testée sur la totalité de la base MITDB ainsi que sur quelques enregistrements de la base AHA. La possibilité de détection d'un grand éventail de battements atypiques sévèrement bruités est mise en évidence. Notons que contrairement à d'autres méthodes exposées dans la littérature, *eg.* [Pan 1985], notre algorithme détecte la majorité des battements avec le seuil primaire, ce qui le rend plus simple et plus rapide.

L'algorithme de classification avec rejet est testé sur les enregistrements sélectionnés. Les figures illustrant les résultats obtenus montrent que la règle de décision utilisée est sensible au coût des erreurs qui peuvent être symétriques ou asymétriques. Les seuils de rejet sont alors flexibles et suivent le coût d'un exemple rejeté et le coût d'un exemple mal classé. L'objectif d'une telle approche est l'optimisation du coût global de la classification ce qui est intéressant pour un diagnostic médical.

TAB. 7.2 – *Résultats de la détection par notre méthode, testée sur la base MITDB*

Rec	NTA	TP	FP	FN	$S_e(\%)$	$P_p(\%)$	$E_r(\%)$
100	2273	2273	0	0	100,00	100,00	0,00
101	1865	1864	2	1	99,95	99,89	0,16
102	2187	2185	0	2	99,91	100,00	0,09
103	2084	2084	0	0	100,00	100,00	0,00
104	2229	2211	12	18	99,19	99,46	1,35
105	2572	2528	15	44	98,29	99,41	2,29
106	2027	2027	8	0	100,00	99,61	0,39
107	2137	2134	4	3	99,86	99,81	0,33
108	1763	1728	25	35	98,01	98,57	3,40
109	2532	2532	0	0	100,00	100,00	0,00
111	2124	2122	0	2	99,91	100,00	0,09
112	2539	2537	0	2	99,92	100,00	0,08
113	1795	1793	1	2	99,89	99,94	0,17
114	1879	1875	1	4	99,79	99,95	0,27
115	1953	1953	0	0	100,00	100,00	0,00
116	2412	2397	4	15	99,38	99,83	0,79
117	1535	1535	0	0	100,00	100,00	0,00
118	2278	2273	0	5	99,78	100,00	0,22
119	1987	1987	0	0	100,00	100,00	0,00
121	1863	1861	1	2	99,89	99,95	0,16
122	2476	2476	0	0	100,00	100,00	0,00
123	1518	1518	0	0	100,00	100,00	0,00
124	1619	1619	1	0	100,00	99,94	0,06
200	2601	2586	11	15	99,42	99,58	1,00
201	1963	1938	7	25	98,73	99,64	1,63
202	2136	2135	1	1	99,95	99,95	0,09
203	2980	2944	25	36	98,79	99,16	2,05
205	2656	2643	0	13	99,51	100,00	0,49
207	1860	1848	8	12	99,35	99,57	1,08
208	2955	2930	5	25	99,15	99,83	1,02
209	3005	3004	1	1	99,97	99,97	0,07
210	2650	2627	18	23	99,13	99,32	1,55
212	2748	2748	0	0	100,00	100,00	0,00
213	3251	3250	0	1	99,97	100,00	0,03
214	2262	2257	2	5	99,78	99,91	0,31
215	3363	3353	4	10	99,70	99,88	0,42
217	2208	2203	2	5	99,77	99,91	0,32
219	2154	2154	0	0	100,00	100,00	0,00
220	2048	2048	1	0	100,00	99,95	0,05
221	2427	2424	1	3	99,88	99,96	0,16
222	2483	2481	2	2	99,92	99,92	0,16
223	2605	2604	0	1	99,96	100,00	0,04
228	2053	1996	27	57	97,22	98,67	4,09
230	2256	2255	3	1	99,96	99,87	0,18
231	1571	1571	0	0	100,00	100,00	0,00
232	1780	1779	0	1	99,94	100,00	0,06
233	3079	3066	0	13	99,58	100,00	0,42
234	2753	2745	1	8	99,71	99,96	0,33
all	109494	109101	193	393	99,64	99,82	0,54

Conclusion et perspectives

Les travaux présentés dans cette thèse rassemblent principalement deux domaines : traitement du signal et apprentissage statistique. Les techniques du traitement du signal nous ont permis de segmenter et de représenter chaque battement cardiaque par un vecteur de caractéristiques. Autour de ces vecteurs, nous avons construit une règle de décision qui nous permet de distinguer un battement ectopique d'un battement normal. Pour ce faire, nous avons choisi de travailler avec la méthode des SVMs, algorithme ayant fait ses preuves en termes de performances en discrimination et en pouvoir de généralisation.

Les algorithmes développés ont été validés sur la base AHA (*American Heart Association*) et la base MITDB sévèrement bruitée. Une étape de prétraitement était alors incontournable. Ces prétraitements ont été réalisés dans le domaine temps-échelle car les signaux souffrent de non stationnarité, notamment pendant les phases d'arythmies.

Lors de la segmentation des battements cardiaques, la présence d'arythmies et d'une variété de battements atypiques était contraignante pour la détection des ondes R et la localisation des complexes QRS. Les méthodes de détection basées sur les dérivées ont tendance à amplifier les battements normaux et atténuer les battements ectopiques caractérisés par une onde large nécessitant ainsi des seuils adaptatifs [Pan 1985] et un temps de calcul élevé.

Une première contribution de ce travail a été de proposer une méthode de localisation de battements cardiaques en utilisant les coefficients de détail issus d'une analyse multi-résolution et en considérant les Densités Spectrales de Puissance correspondant aux battements normaux et aux battements pathologiques [Zidelmal 2012a]. Cet algorithme permet la détection d'une grande variété de battements malgré qu'ils sont parfois noyés dans différents types de bruits. Pour montrer la robustesse d'un tel algorithme, nous avons présenté à travers la section 7, des exemples de détection sur des segments ECG non filtrés. Les résultats statistiques auxquels nous avons abouti sont de 99,64% de sensibilité pour la base MITDB et de 100% de sensibilité pour les enregistrements AHA.

Les battements ainsi localisés ont été pondérés avec une fenêtre rectangulaire de 72 échantillons, soit un segment de 180 ms (environ le double de la largeur d'un battement moyen). Chaque battement i a été ensuite quantifié par un vecteur x_i formé avec les caractéristiques soigneusement choisies. Les descripteurs que nous avons retenus sont ceux habituellement utilisés par les médecins à savoir : la largeur du battement, les inter-battements RR, la morphologie et le contenu fréquentiel

du battement. Ces grandeurs ont été obtenues dans les domaines temporels et fréquentiels.

Dans la section 2.7, nous avons récapitulé quelques problèmes auxquels est souvent confrontée la reconnaissance automatique de battements cardiaques. Ces problèmes sont liés à la grande variabilité de la morphologie et de la synchronisation de battements. Certains battements sont tout simplement ambigus. Dans ce contexte là, nous avons introduit une nouvelle méthode de classification de battements cardiaque en introduisant l'option de rejet dans l'algorithme SVM. La règle de décision posée est sensible au coût des erreurs qui peuvent être symétriques ou asymétriques. Les seuils de rejet sont alors flexibles et suivent le coût d'un exemple rejeté et le coût d'un exemple mal classé [Zidelmal 2012b]. L'objectif d'une telle approche notamment en diagnostic médical est l'optimisation du coût global de la classification.

Comme tous les classifieurs d'arythmies cardiaques exposés dans la littérature ont travaillé avec deux décisions (sans rejet), nous avons comparé les résultats que nous avons obtenus avec 0% de rejet à ceux publiés dans quelques articles. Cette comparaison est très satisfaisante. Notons que le point le plus satisfaisant dans notre approche reste la possibilité de minimiser le coût de la classification, point sur lequel les autres auteurs ne se sont pas penchés.

Perspectives : Les expériences menées sur certains enregistrements nous ont permis d'ouvrir ce travail sur quelques perspectives. Concernant la source ECG, les exemples d'apprentissage pour ces arythmies contiennent en effet un grand nombre de cycles cardiaques normaux, ce qui rend difficile la discrimination des arythmies dont la manifestation est souvent ponctuelle, face au rythme sinusal. Dans la base d'apprentissage, les arythmies doivent être présente avec une bonne probabilité *a priori*.

Travailler sur des données multisources reflétant un même phénomène telles que les différentes dérivations d'un électrocardiogramme et le signal de pression artérielle pourrait être également une piste de recherche intéressante.

La détection des complexes QRS sur certains segments où les positions détectées sont légèrement décalée de la position annotée pourrait être améliorée par une méthode itérative basée sur les enveloppes des minimas et des maximas.

En appliquant l'approche un contre tous, la classification binaire avec rejet pourrait facilement être étendue à la classification multi-classes afin de reconnaître les différentes arythmies. Cette approche consisterait la création d'un classifieur par classe. Chaque classifieur serait entraîné à différencier les battements qui appartiennent à une classe et ceux qui n'y appartiennent pas.

Bilan scientifique relatif à la thèse

1. Z.Zidelmal, A.Amirou, M.Adnane and A.Belouchrani, "*QRS detection based on wavelet coefficients*", ELSEVIER, Computer Methods and Programs in Biomedicine (CMPB), 2012.
2. Z.Zidelmal, A.Amirou and A.Belouchrani, "*Heartbeat classification using Support Vector Machines (SVMs) with an embedded reject option*", International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), DOI :10.1142/S0218001412500012, 2012.
3. Z.Zidelmal, A.Amirou and A.Belouchrani, "*Using Support Vector Machines (SVMs) with reject option for heartbeat classification*", In proceeding of the 2nd International Conference on Bio-inspired Systems and Signal Processing-BIOSIGNALS, 14-17 January, Porto (Portugal), 2009.
4. Z.Zidelmal, A.Amirou and N.Djouaher, "*Application des SVMs basés sur l'algorithme SMO pour la detection d'anomalies cardiaques*", In proceeding of the 4th International Conference : Sciences of Electronic, Technologies of Information and Telecommunication, March 25-29, Tunisia, 2007.
5. Z.Zidelmal and N.Djouaher, "*Identification d'une extrasystole ventriculaire par Support Vector Machines (SVMs)*", ISSPA'07, Boumerdes (Algeria), July, 2007.
6. N.Djouaher and Z.Zidelmal, "*Détection Automatique des battements PVC par la spline quadratique*", ISSPA'07, Boumerdes (Algeria), July, 2007.
7. Z.Zidelmal, N.Djouaher and M.Djeddi, "*Classification of patients prone to Atrial Fibrillation using SVMs*", International conference on Computer System and Information Technology, (ICESIT), Algiers (Algerie), July, 2005.

Bibliographie

- [Adnane 2010] M. Adnane, G.J. Lee, H. Jang, Z. Jiang et H.K. Park. *Neural network based adaptive matched filtering for QRS detection*. EXPERT Syst. Appl., pages 5208–5218, 2010. (Cité en pages [45](#) et [110](#).)
- [Afonso 1999] O. Afonso et W.J. Tompkins. *Classification of premature ventricular complexes using filter bank features, induction of decision trees and a fuzzy rule-based system*. Trans. on Biomedical Engineering, vol. 46, no. 2, pages 192–202, 1999. (Cité en pages [23](#), [68](#) et [117](#).)
- [Arzeno 2008] N. Arzeno, Deng. Zhi et C. Poon. *Analysis of First Derivative Based QRS Detection Algorithms*. IEEE Trans. Biomed. eng., vol. 55, no. 2, pages 478–484, 2008. (Cité en pages [40](#) et [45](#).)
- [Bartlett 2007] P. L. Bartlett et A. Tewari. *Sparseness vs Estimating Conditional Probabilities : Some Asymptotic Results*. Journal of Machine Learning Research, vol. 8, pages 775–790, 2007. (Cité en page [92](#).)
- [Bartlett 2008] P. L. Bartlett et M. H. Wegkamp. *Classification with a reject option using a hinge loss*. Department of Statistics, Florida State University, no. M980, pages 709–721, 2008. (Cité en pages [89](#) et [114](#).)
- [Beasley 2003] B.M. Beasley. *Understanding EKGs, A practical approach*. Prentice Hall, 2003. (Cité en page [18](#).)
- [Borries 2005] R.F. Borries, H.J. Pierluissi et H. Nazeran. *Wavelet Transform-Based ECG Baseline Drift Removal for Body Surface Potential Mapping*. In proc of the 27th Annual Conference on Engineering in Medicine and Biology, Shanghai, pages 3891–3894, 2005. (Cité en pages [14](#), [34](#), [35](#) et [40](#).)
- [Bossert 1992] B.E. Bossert, M. Guymon et V.N. Vapnik. *A Training Algorithm for Optimal Margin Classifiers*. In Proc. Of the Fifth Annual ACM Conference on Computational Learning Theory, 1992. (Cité en page [73](#).)
- [Chen 2006] S.W. Chen, C.H. Chen et H.L. Chan. *A real-time QRS detection-method based on moving-averaging incorporating with wavelet denoising*. Comput. Method. Program. Biomed, vol. 82, no. 3, pages 187–195, 2006. (Cité en pages [14](#), [45](#) et [110](#).)
- [Chow 1957] C. K. Chow. *An Optimum Character Recognition System Using Decision Function*. IRE Trans. Electronic Computers, vol. EC-6, no. 4, pages 247–254, 1957. (Cité en pages [89](#) et [114](#).)
- [Chow 1970] C. K. Chow. *On optimum recognition error and reject tradeoff*. IEEE Trans. on Information Theory, vol. 16, no. 41, pages 41–46, 1970. (Cité en pages [89](#) et [114](#).)
- [Christopher 1998] J. Christopher et C. Burges. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, pages 121–167, 1998. (Cité en page [81](#).)

- [Christov 2005] I. Christov, I. Jecova et G. Botolan. *Premature ventricular contraction classification by the K^{th} nearest-neighbours rule*. *Physiol. Meas.*, vol. 6, pages 123–130, 2005. (Cité en pages 24 et 117.)
- [Christov 2006] I. Christov, G. Herrero et V. Krasteva. *Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification*, 2006. (Cité en page 23.)
- [De-Chazal 2004] P. De-Chazal, M. Dwyer et R. Reilly. *Automatic classification of hearthbeats using ECG morphology and hearthbeat interval features*. *IEEE.Trans. on Biomedical Engineering*, vol. 51, no. 7, pages 1196–1206, 2004. (Cité en pages 23, 24 et 66.)
- [De-Chazal 2006] P. De-Chazal et R. Reilly. *A Patient Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features*. *IEEE.Trans. on Biomedical Engineering*, vol. 53, no. 12, pages 2535–2543, 2006. (Cité en pages 23, 24, 62, 64 et 66.)
- [Dinh 2002] A.N. Dinh, D.K. Kumar et P. Burton. *Wavelet for QRS detection*. *Engineering in medicine and Biology society, 23rd Conference of IEEE*, no. 23, pages 7803–7211, 2002. (Cité en pages 45 et 48.)
- [Donoho 1995] D.L. Donoho. *Denoising by soft thresholding*. *IEEE, Trans. Information Theory*, vol. 41, no. 3, pages 613–627, 1995. (Cité en page 35.)
- [Evanich 1972] M.J. Evanich, O. Newberry et L.D. Patridge. *Some limitations of periodic noise removal by averaging techniques*. *Journal. Appl. Physiol.*, vol. 33, pages 536–541, 1972. (Cité en page 35.)
- [Focapu 1993] O. Focapu et J.P. Girard. *A new approach for P wave detection using analytic signal*. *IEEE Engineering in Medecine and Biology Conference*, pages 400–401, 1993. (Cité en page 46.)
- [Fumera 2002] G. Fumera et F. Roli. *Support Vector Machines with Embedded Reject Option*. *Pattern Recognition with Support Vector Machines : First International Workshop*, pages 68–82, 2002. (Cité en page 89.)
- [Gomez 2006] H.G. Gomez, I. Jecova, V Krasteva et I. Christov. *Relative Estimation of the Karhunen Loeve Transform Basis Functions for Detection of Ventricular Ectopic Beats*. *Computers in Cardiology*, vol. 33, pages 569–572, 2006. (Cité en pages 23, 62, 64 et 117.)
- [Grandvalet 2006] Y. Grandvalet, J. Mariéthoz et S. Bengio. *A probabilistic interpretation of SVMs with an application to unbalanced classification*. *Advances in Neural Information Processing Systems 18*, pages 467–474, 2006. (Cité en page 89.)
- [Grandvalet 2009] Y. Grandvalet, A. Rakotomamonjy, J. Keshet et S. Canu. *Support Vector Machines With a Reject Option*. *Advances in Neural Information Processing Systems*, no. 21, pages 537–544, 2009. (Cité en pages 89, 92 et 93.)

- [Hamilton 1986] P.S. Hamilton et W.J. Tompkins. *Quantitative Investigation of QRS Detection Rule Using the MIT-BIH Arrhythmia Database*. IEEE Trans. Biomed. eng., vol. 12, no. 33, pages 479–482, 1986. (Cité en page 56.)
- [Herbei 2006] R. Herbei et M. H. Wegkamp. *Classification with reject option*. The Canadian Journal of Statistics, vol. 34, no. 4, pages 709–721, 2006. (Cité en pages 89, 92, 101 et 114.)
- [Hu 1997] Y.H. Hu, S.H. Paleady et W.J. Tompkins. *A Patient Adaptable ECG Beat Classifier Using a Mixture of Experts Approach*. Trans. on Biomedical Engineering, vol. 44, no. 9, pages 891–900, 1997. (Cité en page 24.)
- [Joachim 2002] T. Joachim. *Apprendre à classifier texte à l'aide Support Vector Machines*. Dissertation, Kluwer, 2002. (Cité en page 88.)
- [Johnstone 1996] M.I. Johnstone et B.W. Silverman. *Wavelet threshold estimators for data with correlated noise*. Technical report Dept. of Statistics, Stanford University, 1996. (Cité en page 35.)
- [Kohler 2002] B. Kohler, C. Henning et Orglmeister. *The principles of software QRS detection*. IEEE Eng. Med. Biol. Mag., vol. 3, no. 21, pages 42–57, 2002. (Cité en pages 23 et 45.)
- [Krasteva 2007] V. Krasteva et I. Jecova. *QRS Template Matching for Recognition of Ventricular Ectopic Beats*. Annals of Biomedical Engineering, vol. 55, no. 12, pages 2065–2076, 2007. (Cité en pages 23, 24, 62, 64, 66, 68 et 117.)
- [Kunt 1999] M. Kunt. *Traitement numérique des signaux*. Ed, Press polytechnique et universitaire omande, no. 3, 1999. (Cité en page 51.)
- [Kwok 1999] J.T. Kwok. *Moderating the outputs of support vector machine classifiers*. IEEE Trans. on Neural Networks, vol. 10, no. 5, pages 1018–1031, 1999. (Cité en page 89.)
- [Lagerholm 2000] M. Lagerholm, G. Peterson et L. Edenbrandt. *Clustering ECG complex using Hermite Functions and self-organizing maps*. IEEE Trans.Biomed.Eng, vol. 47, no. 7, pages 838–848, 2000. (Cité en page 23.)
- [Lake 1990] C. Lake. *Clinical Monitoring*. Saunders, Philadelphia, 1990. (Cité en page 18.)
- [Lepage 2003] R. Lepage. *Detection et analyse de l'onde P d'un electrocardiogramme : Application au dépistage de la fibrillation auriculaire*. These de Doctorat de l'universite de Bretagne, 2003. (Cité en pages 34, 35, 45 et 50.)
- [Loosli 2005] G. Loosli, S. Canu, S. Vishwanathan et M. Chattopadhyay. *Boite à outils SVM simple et rapide*. RIA - Revue d'Intelligence Artificielle, vol. 19, no. 4, pages 741–767, 2005. (Cité en pages 81, 88 et 96.)
- [Mallat 2009] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, no. Ed, 3, 2009. (Cité en page 28.)

- [Mark 1987] R. Mark et R. Wallen. *AAMI-recommended practice : Testing and reporting performance results of ventricular arrhythmia detection algorithms*. AAMI, Tech. Rep. ECAR, 1987. (Cité en page 110.)
- [Mark 1988] R. Mark et G. Moody. *MIT-BIH Arrhythmia database directory*. Massachusetts Inst. Technol. (MIT), 1988. (Cité en pages 4, 14, 59, 103 et 110.)
- [Mercer 1909] J. Mercer. *Functions of positive and negative type and their connection with the theory of integral equations*. Philos. Trans. Roy. Soc, 1909. (Cité en page 73.)
- [Meyer 2006] C. Meyer, J.F. Gavela et M. Harris. *Combining algorithms in automatic detection of QRS complexes in ECG signals*. IEEE.Trans.Inf.Technol, vol. 10, no. 3, pages 468–475, 2006. (Cité en page 109.)
- [Minami 1999] K. Minami, H. Nakajima et T. Toyoshima. *Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network*. IEEE. Trans. on Biomed.Eng, vol. 46, no. 2, pages 179–185, 1999. (Cité en pages 23 et 68.)
- [Moody 1984] G. Moody, W. Muldrow et R. Mark. *A noise stress test for arrhythmia detectors*. IEEE Computers in Cardiology, no. 11, pages 381–384, 1984. (Cité en page 14.)
- [Moody 1989] G. Moody et R. Mark. *QRS morphology representation and noise estimation using the Karhunen-Loève transform*. Computers in Cardiology, vol. 16, pages 269–272, 1989. (Cité en pages 23, 58, 62 et 64.)
- [Osowski 1997] S. Osowski, L.T. Hoai et T. Markiewicz. *Support Vector Machine-Based Expert System for Reliable Heartbeat Recognition*. Trans. on Biomedical Engineering, vol. 51, no. 4, pages 582–589, 1997. (Cité en page 24.)
- [Pan 1985] J. Pan et W.J. Tompkins. *A real-time QRS detection algorithm*. IEEE Trans. Biomed. ENG., vol. 3, no. 32, pages 479–482, 1985. (Cité en pages 3, 34, 35, 40, 45, 56, 110, 117 et 119.)
- [Patra 2010] P.D. Patra. *Integration of FCM, PCA and Neural Networks for Classification of ECG Arrhythmias*. IAENG International Journal of Computer Science, vol. 36, no. 3, 2010. (Cité en pages 24, 62 et 64.)
- [Ramaswamy 2004] P. Ramaswamy, N.G. Cota, K.L. Chan et M.K. Shankar. *MultiParameter Detection of Ectopic Heartbeats*. IEEE, Int.Workshop.BioCAS, 2004. (Cité en page 70.)
- [Richard 1973] O. Richard et E. Petre. *Pattern classification and Scene Analysis*. Wiley, 1973. (Cité en page 73.)
- [Ruchita 2010] G. Ruchita et A.K. Sharma. *Detection of QRS complexes of ECG recording based on wavelet transform using Matlab*. Int. J. Eng. SCI, vol. 2, no. 7, pages 3038–3034, 2010. (Cité en pages 45 et 48.)

- [Sornmo 2006] L. Sornmo et P. Laguna. *Electrocardiogram ECG signal processing*. Wiley Encyclopedia of Biomedical Engineering, 2006. (Cité en pages 34 et 35.)
- [Sung-Nien 2007] Y. Sung-Nien et C. Kuan. *A switchable scheme for ECG beat classification based on Independent Component Analysis*. Elsevier, Expert Systems with Applications, vol. 33, no. 1, pages 824–829, 2007. (Cité en page 64.)
- [Tawara 2000] S. Tawara. *The conduction System of the Mammalian Heart*. Imperial Press, 2000. (Cité en page 8.)
- [Thakor 1984] N.V. Thakor, J.G Wbstor et W.J. Tompkins. *Estimation of the QRS Complex Power Spectra for Design of a QRS Filter*. IEEE.Trans. on Biomed. Eng, no. 11, pages 702–706, 1984. (Cité en pages 35, 40, 53, 68 et 104.)
- [Thakor 1993] N.V. Thakor, Y. Sun et P. Caminal. *A Wavelet- Based ECG Data Compression Algorithm*. IEEE Transactions on Information and system, vol. 76, no. 12, pages 1462–1469, 1993. (Cité en page 64.)
- [Tortorella 2004] F. Tortorella. *Reducing the classification cost of support vector classifiers through an ROC-based reject rule*. Pattern Analysis & Applications, vol. 7, no. 2, pages 128–143, 2004. (Cité en page 89.)
- [Truchetet 1998] F Truchetet. *Ondelettes pour le signal numérique*. Edition, Hermes, 1998. (Cité en page 28.)
- [Vapnik 1963] V.N. Vapnik et A. Lerner. *Pattern Recognition using Generalized Portrait Method*. Automation and Remote Control, 1963. (Cité en page 73.)
- [Vapnik 1995] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer Series in Statistics, 1995. (Cité en pages 73, 76, 81, 85 et 92.)
- [Vishwanathan 2003] S.V.N. Vishwanathan, A. Smola et N. Murty. *SimpleSVM*. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), pages 68–82, 2003. (Cité en pages 88 et 96.)
- [Wang 1994] L. Wang, J. Belina et M. Berner. *Compression of ECG using a code excited linear prediction (CELP)*. IEEE, engineering in medecine and biomedical, vol. 2, pages 1264–1265, 1994. (Cité en page 64.)
- [Wilson 1944] F. Wilson, F. Johnston et H. Erlanger. *The precordial electrocardiogram*. American Heart Journal, no. 27, pages 19–85, 1944. (Cité en page 13.)
- [Yeap 1990] T.H. Yeap, F. Johson et M. Rachniowski. *ECG beat classification by a neural network*. Annual International Conf of IEEE Eng in Med and Biol, vol. 12, pages 1457–1458, 1990. (Cité en page 24.)
- [Zhang 2004] T. Zhang. *Statistical behavior and consistency of classification methods based on convex risk minimization*. The Annals of Statistics, vol. 32, no. 1, pages 56–85, 2004. (Cité en page 92.)

- [Zhao 2008] Y. Zhao et H. Wenxue. *ECG Beats Classification Based on Ensemble Feature Composed of Independent Components and QRS Complex Width*. IEEE Xplore, no. 1, 2008. (Cité en pages [64](#) et [66](#).)
- [Zidelmal 2009] Z. Zidelmal, A. Amirou et A. Belouchrani. *Using Support Vector Machine SVMs with reject option for heartbeat classification*. 2nd International Conference on Bio-inspired Systems and Signal Processing, no. 2, pages 204–210, 2009. (Cité en pages [3](#) et [24](#).)
- [Zidelmal 2012a] Z. Zidelmal, A. Amirou, M. Adnane et A. Belouchrani. *QRS detection using wavelet coefficients*. Computer Methods and Programs in Biomedicine, 2012. (Cité en pages [3](#), [50](#), [104](#) et [119](#).)
- [Zidelmal 2012b] Z. Zidelmal, A. Amirou et A. Belouchrani. *Heartbeat classification using Support Vector Machines (SVMs) with an embedded reject option*. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), no. DOI : 10.1142/S0218001412500012, 2012. (Cité en pages [3](#) et [120](#).)