

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité
ÉLECTRONIQUE
Présentée

par **Rémi DUBOIS**

pour obtenir le titre de **DOCTEUR de L'UNIVERSITÉ PARIS 6**

Sujet de la thèse :

Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie

Soutenue le 27 janvier 2004

Devant le jury composé de :

Gérard	Dreyfus	Directeur
Pierre	Maison-Blanche	Rapporteur
Manuel	Samuelides	Rapporteur
Yves	Faisandier	Examineur
Patrick	Gallinari	Examineur
Philippe	Mabo	Examineur
Brigitte	Quenet	Examineur

Sommaire

Sommaire.....	1
Table des matières	3
Introduction	7
Chapitre 1 : Le cœur et l'électrocardiographie	17
Chapitre 2 : Les pathologies cardiaques	31
Chapitre 3 : Détection des ondes QRS.....	55
Chapitre 4 : Analyse de la ligne de base	75
Chapitre 5 : Analyse en composantes principales.....	87
Chapitre 6 : Modélisation des battements cardiaques	111
Chapitre 7 : Classification non supervisée des battements	153
Chapitre 8 : Étiquetage des battements et des ondes caractéristiques	175
Conclusion	195
Bibliographie	199
Annexes	205

Table des matières

Introduction	7
Etat de l'art	8
Objectifs	9
Plan du mémoire	11
Validation des algorithmes	14
Chapitre 1 Le cœur et l'électrocardiographie	17
I Le système cardiovasculaire	17
I.1 La circulation artérielle et veineuse	17
I.2 Le cœur	18
II L'électrocardiographie	21
II.1 Principe de l'électrocardiogramme (ECG)	21
II.2 De l'ECG au Holter	22
II.3 Trace électrique du cœur	24
II.4 Les représentations de l'ECG	25
Chapitre 2 Les pathologies cardiaques	31
I Introduction	31
I.1 Le rythme cardiaque	31
I.2 Le battement cardiaque standard et ses caractéristiques	32
I.3 Extrasystole ventriculaire (ESV)	34
II Diagnostic à partir du rythme	35
II.1 Fréquence	35
II.2 Arythmies, ou troubles de la régularité	39
III Diagnostic à partir des ondes	46
III.1 L'onde P	46
III.2 Le complexe QRS	49
III.3 L'onde T	51
Chapitre 3 Détection des ondes QRS	55
I La détection monopiste	55
I.1 L'Algorithme	57
I.2 Résultats	61
I.3 Conclusion	64

II	Analyse multipiste.....	64
II.1	L'algorithme.....	64
II.2	Résultats	71
Chapitre 4	Analyse de la ligne de base	75
I	Détection de la ligne de base	76
I.1	Motivation	76
I.2	Segmentation du signal en zones	76
I.3	Interpolation et extrapolation	78
I.4	Résultats	79
II	Estimation du bruit	82
II.1	Bruit BF.....	82
II.2	Bruit HF.....	83
Chapitre 5	Analyse en composantes principales	87
I	Objectifs	87
II	Principe de l'analyse en composantes principales.....	89
II.1	Matrice de passage et matrice de covariance	90
II.2	Valeurs propres	92
II.3	Voie principale S_1^{ACP}	93
III	Avantage et limites de la méthode	93
III.1	Avantage.....	93
III.2	Limites de la méthode	94
III.3	Résultat de l'analyse en composantes principales.....	98
IV	La respiration.....	105
IV.1	Expériences	105
IV.2	Résultats	106
Chapitre 6	Modélisation des battements cardiaques	111
I	Objectif de la modélisation.....	112
I.1	Présentation	112
I.2	Objectif.....	113
II	Modélisation classique	113
II.1	Décomposition en ondelettes	114
II.2	Modélisation par un réseau de neurones à fonctions dorsales	119
II.3	Modélisation par un réseau de fonctions radiales de base (RBF pour Radial Basis Function)	122
II.4	Régression orthogonale directe généralisée (GOFR : generalized orthogonal forward regression) par fonctions radiales de base	131
III	Régression orthogonale directe généralisée par fonctions bosses	139
III.1	Définition de la bosse.....	139
III.2	L'algorithme GOFR	140

III.3	Résultats de l'algorithme.....	147
Chapitre 7	Classification non supervisée des battements.....	153
I	Motivations et objectifs.....	153
II	Présentation d'algorithmes existants.....	154
III	Principe général de l'algorithme.....	157
III.1	Caractérisation d'une famille.....	157
III.2	Principe de la classification non supervisée.....	163
IV	Résultats.....	171
IV.1	Homogénéité des familles.....	172
IV.2	Nombre final de familles.....	173
IV.3	Perspectives d'amélioration.....	173
Chapitre 8	Étiquetage des battements et des ondes caractéristiques ..	175
I	Étiquetage <i>N</i> ou <i>V</i> des battements.....	175
I.1	Recherche de l'onde R.....	175
I.2	Méthode d'étiquetage des familles.....	181
II	Étiquetage des ondes.....	182
II.1	Zones caractéristiques.....	182
II.2	Repérage des ondes au niveau de chaque battement.....	186
III	Résultats.....	188
III.1	Tracé de la distance PR.....	189
III.2	Tracé de la distance QT.....	190
III.3	Etude d'un foyer ectopique auriculaire.....	191
III.4	Enregistrement annoté.....	193
Conclusion	195
Perspective pour l'analyse ECG.....		196
Perspective de la méthode.....		197
Bibliographie	199
ANNEXE A	Résultats de la détection des complexes QRS	A-1
I	Avec analyse du bruit / Sans analyse du bruit.....	3
II	Faux positifs (FP).....	5
III	Faux négatifs (FN).....	6

ANNEXE B	Algorithmes d'optimisation sous contraintes	B-1
I	Algorithme du gradient projeté	2
II	Algorithme de BFGS projeté.....	4
III	Conclusion.....	5
ANNEXE C	Hypothèse de similitude du signal S^i et E^i	C-1
ANNEXE D	Résultats de la classification non supervisée	D-1
I	Battement de fusion (label F).....	4
II	Erreurs de classification	5
II.1	Analyse sur une voie unique	5
II.2	Erreur en 2 voies	7
ANNEXE E	Réseau de neurones pour la localisation des bosses modélisant l'onde R	E-1
I	Architecture du réseau	1
II	Base d'apprentissage et base de test.....	2
III	Déroulement de l'apprentissage.....	2
IV	Résultats de l'apprentissage.....	3
IV.1	Critère de mesure de la qualité de l'apprentissage.....	3
IV.2	Résultats sur la base d'apprentissage	4
IV.3	Résultats sur la base de validation.....	5
ANNEXE F	Étiquetage des battements N / V	F-1
I	Arbre de décision.....	1
I.1	Mesure de la position RR de la famille (M_{RR}).....	2
I.2	Mesure des différences d'amplitudes (M_A)	2
I.3	Mesure discrète de la corrélation	3
I.4	Décision.....	3
II	Résultats sur les bases MIT et AHA.....	5
ANNEXE G	Comparaison des résultats	G-1
I	Détection des QRS.....	1
II	Étiquetage N / V.....	2

Introduction

Les maladies cardiovasculaires constituent un problème majeur de santé publique : avec un peu plus de 180 000 décès par an, elles sont la première cause de mortalité en France, soit environ 1 décès sur 3. Les facteurs de risques sont multiples : tabac, sédentarité, obésité, hypertension artérielle, diabète, facteurs génétiques aussi parfois.

En raison de l'ampleur du problème, le ministère de la santé a mis en place en 2002 un plan pluriannuel de réduction des risques préconisant notamment l'éducation, la prévention, et le suivi des patients à risques, *ce qui concerne près de 20 millions de personnes par an*.

Le cœur est l'organe central du système cardiovasculaire : il peut être affecté de nombreuses pathologies qui peuvent soit être bénignes, comme certaines tachycardies par exemple, soit s'avérer très sérieuses, comme l'infarctus du myocarde. Avec l'évolution des techniques, les médecins disposent aujourd'hui d'outils performants pour observer le fonctionnement du muscle cardiaque et dresser ainsi leur diagnostic. Parmi les examens cardiologiques possibles, *l'électrocardiogramme (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. Normalement effectué sur un patient au repos en cabinet ou en milieu hospitalier, pendant quelques secondes, l'ECG est très souvent complété par un examen similaire d'une durée de 24 heures appelé « Holter »¹, examen au cours duquel le patient peut vaquer à ses occupations habituelles. Le principal avantage de l'enregistrement Holter par rapport à l'ECG est qu'il permet la détection d'événements sporadiques qui n'interviennent pas nécessairement au cours des quelques secondes de l'enregistrement ECG 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'information enregistrée en 24 heures est très importante : elle correspond à environ *100 000 battements cardiaques* sur 3 voies d'enregistrements, soit environ *52 millions de points* pour un enregistrement échantillonné à 200Hz. Ces outils de lecture doivent permettre le repérage d'informations pertinentes, et le calcul de paramètres caractéristiques comme le rythme cardiaque ou la fréquence d'occurrence des battements anormaux.

¹ Du nom de son inventeur, Norman Holter, qui a mis au point cette méthode d'enregistrement en 1961 [Holter, 1961]

Cette thèse propose une méthode originale d'analyse de ce type d'enregistrements : on présente ici une stratégie de lecture automatique du signal qui, à partir d'un enregistrement brut de 24 heures permet de repérer les zones « pertinentes » qui sont révélatrices du fonctionnement cardiaque et de ses troubles éventuels. A partir d'un tel repérage, il est possible de calculer courbes et tendances dans le but de fournir une vue globale de l'enregistrement et de rendre ainsi son analyse par le cardiologue rapide et efficace.

Ce travail a fait l'objet d'une collaboration avec la société Ela Medical filiale du groupe pharmaceutique italien SNIA. Principalement connu pour la fabrication de stimulateurs cardiaques (pacemakers), elle propose depuis de nombreuses années une gamme complète d'équipements Holter comprenant des appareils enregistreurs et des logiciels d'analyse.

Etat de l'art

Les logiciels d'analyse des enregistrements Holter disponibles aujourd'hui^{II} permettent déjà une analyse performante du *rythme cardiaque* : l'onde R de chaque battement (onde de plus grande amplitude de l'ECG, qui traduit la dépolarisation des cellules des ventricules cardiaques) est précisément repérée, ce qui permet de *détecter les principaux troubles du rythme*.

De plus *l'analyse de la forme de l'onde R* permet de distinguer les battements d'origine sinusale de ceux d'origine ventriculaire.

En revanche, exceptée une étude de l'intervalle S-T (durée séparant le début de l'onde R de l'onde T qui traduit la repolarisation des cellules cardiaques) dans les dernières versions du Synetec de Ela médical, par exemple, les pathologies liées aux autres ondes cardiaques notamment l'onde P (onde de dépolarisation des oreillettes cardiaques) ne sont que mal détectées ou identifiées [Adamec, 2000], [Fillette, 1983].

^{II} Les grandes marques proposant ce type de matériel sont : Ela medical, GE Medical system, Novacor, Oxford Instrument,...

Objectifs

De manière générale, l'analyse d'un signal est habituellement composée de deux étapes : la première consiste à trouver une représentation du signal adaptée à la propriété recherchée : on exprime alors le signal original à l'aide de *descripteurs*, en effectuant une transformation du signal de l'espace d'enregistrement à l'espace des descripteurs. La seconde étape s'effectue dans ce deuxième espace : c'est *l'analyse de la valeurs des descripteurs* pour déduire les propriétés recherchées. 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 simplifiée.

Pour illustrer et expliquer ces propos, prenons l'exemple simple de classification d'images représentant des chiffres manuscrits. La propriété finale souhaitée est la reconnaissance de *la valeur* du chiffre numérique représenté sur l'image (Figure 1).

Une première analyse peut être la suivante : l'image étant composée de N pixels numérotés de 1 à N , la représentation la plus simple de l'image est un vecteur de dimension N dont chaque coordonnée est 1 ou 0 selon l'illumination ou non du pixel correspondant. L'étape de classification étudie alors ce vecteur pour lui associer un chiffre entre 0 et 9. Le descripteur utilisé ici (le vecteur de dimension N) est très simple : aucun travail particulier n'a été effectué au niveau de la recherche de celui-ci, et l'expression du signal dans l'espace des descripteurs est particulièrement rapide. Mais dans ce cas, deux images représentant deux chiffres identiques peuvent avoir des vecteurs descripteurs associés très différents l'un de l'autre, ce qui rend le travail de classification très difficile.

Une seconde méthode d'analyse peut être la suivante : on recherche, dans chaque image, la courbe représentant le dessin du chiffre manuscrit, puis on compte d'une part le nombre de boucles que celle-ci effectue, ainsi que le nombre d'angles droits, le nombre d'angles obtus et le nombre d'angles aigus. On décrit ainsi l'image par un vecteur de 5 paramètres. Le travail effectué pour la construction de ce vecteur est difficile et nécessite de nombreux traitements. Mais, en revanche, la classification qui résulte de l'utilisation de cette description est particulièrement efficace. On peut également noter que l'espace des descripteurs utilisés ici est plus « compréhensible » ou naturel pour un observateur que l'espace des pixels (vecteurs constitué de 0 et de 1).

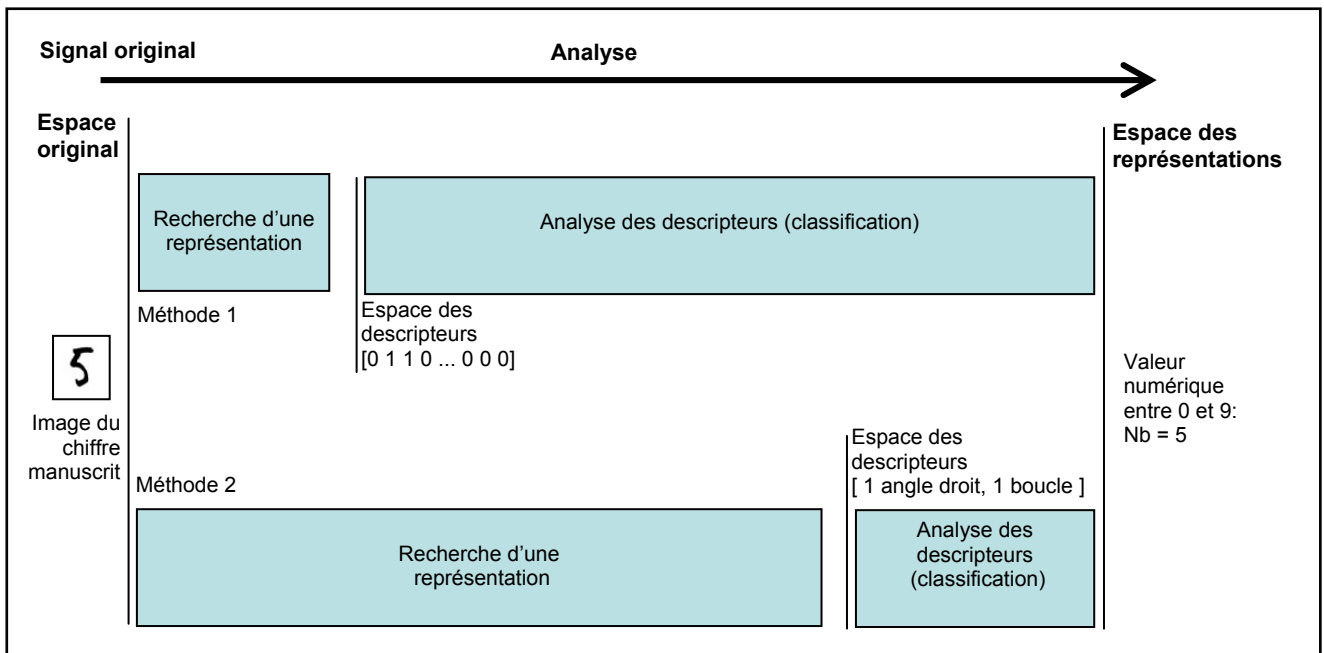


Figure 1 : L'analyse d'un signal se compose habituellement de deux étapes : la première est l'expression de ce signal sous forme de descripteurs, la seconde est l'analyse de ces descripteurs pour obtenir la propriété désirée. Selon la pertinence des descripteurs et le soin apporté à leur recherche, la seconde étape est plus ou moins simplifiée.

Pour revenir au cas de l'analyse du signal ECG, la méthode développée ici est fondée sur deux étapes d'analyses qui correspondent aux deux étapes précédemment décrites : la première est l'expression du signal dans un espace de descripteurs ; la seconde est l'analyse dans cet espace des valeurs des descripteurs pour les associer à des propriétés du signal ECG (Figure 2).

Dans la présente étude, nous nous sommes efforcés de construire un espace de descripteurs qui soit « compréhensible » par les médecins, c'est-à-dire d'exprimer le signal sous forme de descripteurs qui sont ceux habituellement utilisés par le cardiologue pour en faire l'analyse.

L'avantage de cette démarche est double.

- Tout d'abord, la seconde étape - l'analyse des descripteurs - peut se faire par application directe de la connaissance médicale, car celle-ci s'appuie directement sur la valeur de ces descripteurs.
- D'autre part, comme l'outil développé est destiné à constituer un outil d'aide au diagnostic, l'interaction entre le médecin et le programme doit être particulièrement simplifiée, et être aussi proche que possible de la démarche habituelle du médecin.

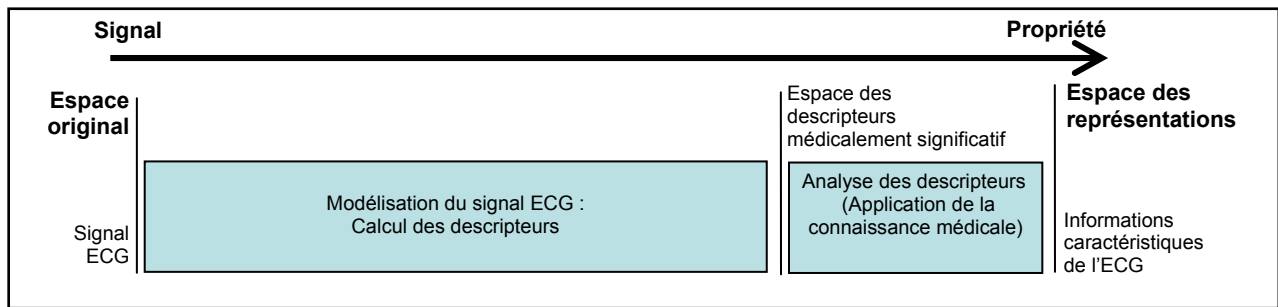


Figure 2 : *Un des objectifs du présent travail était de construire des descripteurs médicalement significatifs, donc directement intelligibles par le médecin. La seconde étape d'analyse se réduit alors à une application de la connaissance experte sur ces descripteurs, pour obtenir les caractéristiques de l'ECG.*

Ainsi, nous avons développé un algorithme capable de détecter pour chaque battement cardiaque les ondes caractéristiques (c'est-à-dire essentiellement les ondes P, R et T) et de les caractériser par leur emplacement temporel, leur largeur et leur amplitude, qui sont les descripteurs de chaque onde caractéristique. Ce sont les mêmes descripteurs que ceux utilisés par le cardiologue pour caractériser ces ondes. Ainsi, une anomalie de dépolarisation des oreillettes ou de transmission de cette dépolarisation aux ventricules se traduit par une modification d'un des descripteurs de l'onde P, par exemple une modification de son amplitude, ou une distance à l'onde R anormalement courte: de telles mesures, qui possèdent des significations physiologiques précises, peuvent facilement être introduites dans la détection automatique. Une telle approche permet donc de détecter les ondes P anormales ou de présenter des courbes de tendances sur toute la durée de l'enregistrement Holter, ce qui était jusqu'ici impossible à réaliser ; parmi de telles courbes, citons l'évolution de la distance entre l'onde P et l'onde R au cours du temps

L'algorithme présenté ici s'articule autour de méthodes mathématiques de traitement du signal développées au laboratoire, méthodes fondées sur les réseaux de neurones et plus généralement sur les algorithmes à apprentissages.

Plan du mémoire

Les deux premiers chapitres présentent sommairement le fonctionnement du système cardiovasculaire, et notamment du cœur. Ils permettent en particulier de comprendre l'origine

des signaux électriques enregistrés par l'électrocardiogramme, et présentent quelques pathologies cardiaques typiques souvent illustrées par un enregistrement ECG correspondant.

Le préalable à toute analyse du signal ECG est la *détection des ondes R* : l'onde ventriculaire qui est l'onde de plus haute amplitude de chaque battement. Le **chapitre 3** présente la méthode développée ici ; elle comporte deux étapes : 1) la détection de ces ondes sur chacune des voies d'enregistrements qui constitue la *détection monopiste* ; 2) la synthèse de ces informations afin d'obtenir une indexation précise de la position des ondes R, qui constitue la *synthèse multipiste*. Une fois ces étapes réalisées, *l'enregistrement est segmenté en fenêtres ne contenant qu'un battement cardiaque*.

Le **chapitre 4** présente la technique utilisée pour compenser les problèmes liés aux *variations de la ligne de base* c'est-à-dire la ligne isoélectrique du cœur. Ce traitement est nécessaire 1) pour la synthèse multipiste des ondes R, car il donne accès à un *indice de fiabilité* des voies d'enregistrement à chaque instant, et 2) pour réaliser une modélisation pertinente du signal comme exposé au chapitre 6.

À ce niveau, l'ECG est donc représenté sous forme de *fenêtres temporelles contenant chacune un battement cardiaque exprimé sur 1, 2 ou 3 voies selon la qualité^{III} de celles-ci*. Afin de travailler avec un nombre constant de paramètres, le **chapitre 5** décrit une application originale à l'ECG d'une méthode d'analyse multivariée permettant de ramener un battement enregistré sur plusieurs voies à un battement équivalent exprimé sur *une voie unique* en gardant un maximum d'information. Cette application est principalement fondée sur la méthode *d'analyse en composantes principales*.

Chaque battement étant maintenant exprimé par *un signal temporel sur une voie unique*, le **chapitre 6** propose une méthode de modélisation de chacun d'eux. On les décompose ici sur une base de fonctions appelées *bosses* particulièrement bien adaptées à la modélisation des ondes d'un battement cardiaque : ces fonctions ont été définies de façon ad hoc pour cette application, mais elles sont d'usage beaucoup plus général ; elles ont fait l'objet d'un dépôt de brevet.

^{III} La qualité d'une voie d'enregistrement est estimée à partir d'indices portant sur l'importance du bruit de basse fréquence, et du bruit haute fréquence également estimé lors de l'application de l'algorithme décrit dans le chapitre 4.