



Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire des Dispositifs de Communication
et de Conversion Photovoltaïque



Mémoire de Master En Electronique

Présenté par :

HADJ-YOUCHEF Mohamed elAmine

Ingénieur d'état en électronique de l'ENP

Intitulé

Outil d'aide au diagnostic basé sur l'analyse multi-résolution pour une classification des signaux EEG

Proposé et dirigé par :

Mr M.ADNAN

Docteur

Ecole Nationale Polytechnique

Mme A.BOUSBIA-SALAH

MAT classe A

USTHB

Octobre 2012

Remerciements

Je remercie DIEU le tout puissant, miséricordieux pour m'avoir donné la foi, le courage, la patience, la volonté et la force nécessaire, pour affronter toutes les difficultés et les obstacles, qui se sont mis au travers de mon chemin durant toutes mes années d'études.

J'exprime mes remerciements à mes encadreurs : Mr M.ADNANE et Mme A.BOUSBIA-SALAH, pour leur assistance, disponibilité et gentillesse, également pour leurs conseils et orientations le long de ce travail, qu'ils trouvent ici l'expression de notre profonde gratitude.

J'exprime mon gratitude à Monsieur le chef de département d'Électronique Mr H.BOUSBIA-SALAH, pour sa disponibilité, son professionnalisme et ces qualités humaines tout au long de ma formation au sein de ce département.

Enfin, je ne pourrai pas terminer ces remerciements sans une pensée à l'ensemble des enseignants, qui sont à l'origine de tout mon savoir.

ملخص

في هذا العمل نقترح نظاما للتصنيف التلقائي لإشارات التخطيط الدماغي (EEG). النظام يحتوي على المراحل الاربع التالية : (أ) تمثيل إشارات التخطيط الدماغي في محور زمن-تواتر باستخدام نوعين من تحويل الموجات (DWT, WPT) (ب) استخراج المكونات الخصوصية (ج) تخفيض البيانات باستعمال التحليل بالمكونات الرئيسية (ACP) (د) التصنيف بطريقة آلة دعم الأشعة (SVM) متبوعة بخوارزمية الحذف التراجعي للمكونات (RFE) لاستنتاج أفضلها.

الهدف من هذا العمل هو تطوير نظام لتصنيف إشارات التخطيط الدماغي (EEG) ضمن صنفيه : العادي والمرضي.

كلمات مفتاحية : إشارات التخطيط الدماغي (EEG), تحويل الموجات (WT), استخراج المكونات الخصوصية, التحليل بالمكونات الرئيسية (ACP), آلة دعم الأشعة (SVM), الحذف التراجعي للمكونات (RFE).

Résumé

Dans ce mémoire, un système de classification automatique des signaux EEG est proposé. Le système contient les quatre étages suivants : (A) Représentation des signaux EEG en temps-fréquence en utilisant deux versions de la transformée en ondelettes (DWT, WPT). (B) Extraction des paramètres caractéristiques (statistique et non-statistique) des coefficients d'ondelette et de paquets d'ondelettes. (C) Réduction des données par l'analyse en composantes principales (ACP). (D) Classification avec la méthode de machine à vecteurs de support (SVM) suivie de l'algorithme d'élimination récursive des paramètres (RFE) pour déterminer les meilleurs paramètres.

Le but de ce travail est de développer un algorithme de classification des signaux EEG en deux classes : Normal, Pathologique (épileptique).

Mots clés : Électroencéphalogramme (EEG), transformée en ondelette (WT), extraction du paramètre caractéristique, Analyse en composantes principales (ACP), machine à vecteurs de support (SVM), élimination récursive des paramètres (RFE).

Abstract

In this work, a system for automatic classification of EEG signals is proposed. The system contains the four following stages: (A) Representation of EEG time-frequency using two versions of the wavelet transform (DWT, WPT). (B) Extraction of the characteristic parameters (statistical and non-statistical) of wavelet coefficients and wavelet packets. (C) Data reduction by principal component analysis (PCA). (D) the method of classification with support vector machine (SVM) followed by recursive feature elimination algorithm (RFE) to determine the best features.

The aim of this work is to develop an algorithm for classifying EEG signals into two classes: Normal, Pathological (epilepsy).

Keywords: electroencephalogram (EEG), wavelet transform (WT), feature extraction, principal component analysis (ACP), Support Vector Machine (SVM), recursive feature elimination (RFE).

Table des matières

Remerciements	i
Table des matières	v
Table des figures	ix
Liste des tableaux	xiii
Introduction	1
1 Contexte médical	3
1.1 Introduction	3
1.2 Le cerveau : bases anatomiques et physiologiques	3
1.2.1 Anatomie du cerveau	3
1.2.2 Activité électrique	4
1.2.3 Mesure de l'activité cérébrale	7
1.3 Électroencéphalogramme (EEG)	8
1.3.1 Histoire	8
1.3.2 Définition	9
1.3.3 Les différents rythmes du signal EEG	10
1.4 La mesure des signaux EEG	12
1.5 L'emplacement d'électrodes : Le système 10–20	14
1.6 Conclusion	16
2 Transformations à base d'ondelettes et de paquets d'ondelettes	17
2.1 Introduction	17

TABLE DES MATIÈRES

2.2	Analyse temps-fréquence	17
2.2.1	La transformée de fourrier (TF)	18
2.2.2	La transformée de fourrier fenêtrée (STFT)	19
2.2.3	La transformée en ondelette WT	21
2.2.3.1	Histoire des ondelettes	21
2.2.3.2	La transformée en ondelette continue (CWT)	22
2.2.3.3	La transformée en paquets d'ondelettes (WPT)	26
2.3	Conclusion	28
3	Extractions des paramètres caractéristiques des signaux EEG et Analyse des données par l'ACP	29
3.1	Introduction	29
3.2	La base de données	29
3.3	Les paramètres caractéristiques extraits	31
3.3.1	Paramètres caractéristiques statistiques	31
3.3.2	Paramètres caractéristiques Non-statistiques	32
3.4	L'analyse en composantes principales (ACP)	32
3.4.1	Tableau des données	32
3.4.1.1	Choix d'une distance	34
3.4.1.2	Choix de l'origine	34
3.4.1.3	Matrices de covariance et de corrélation empiriques	36
3.4.1.4	Moments d'inertie	37
3.4.1.5	Recherche du 1er axe principale	38
3.4.1.6	la Recherche des axes suivants	38
3.4.1.7	Représentation des individus dans les nouveaux axes	39
3.4.1.8	Différentes procédures pour le choix du nombre d'axes	40
3.5	Conclusion	41
4	Méthode de classification : Machine à vecteurs de support (SVM)	43
4.1	Introduction	43
4.2	Différentes Méthodes de classifications	43
4.3	Introduction au Machine à vecteurs de support (SVM)	44
4.4	Les Machines à vecteurs de support	44
4.5	Hyperplan séparable	45

4.6	Marge et hyperplan canonique	47
4.6.1	Trouver l'hyperplan	48
4.6.2	Représentation duale	49
4.7	Vecteur de support	51
4.8	Machines à vecteurs de support pour données non linéairement séparables	53
4.8.1	Astuce du noyau	53
4.9	Marges souples	54
4.9.1	Machines à vecteurs de support et bruit	54
4.9.2	Marge souple	55
4.9.3	Représentation duale	56
4.10	Algorithme d'élimination récursive des paramètres (RFE)	57
4.11	Validation croisée	58
4.12	Les Paramètres de performances	59
4.13	Conclusion	60
5	Implémentation et évaluation des performances	61
5.1	Introduction	61
5.2	Système automatique de classification de signaux EEG	62
5.3	Signaux EEG	63
5.4	La transformée en paquets d'ondelette des signaux EEG	64
5.5	Extraction des paramètres caractéristiques	67
5.5.1	Extraction pour les signaux EEG	67
5.5.2	Extraction des paramètres après la décomposition en paquets d'ondelette (WPT)	71
5.6	Réduction des données par l'Analyse en composantes principales (ACP) .	73
5.7	Classification des signaux EEG par la méthode SVM-RFE	75
5.7.1	Classification des signaux EEG du groupe 1 par la méthode SVM-RFE en utilisant la WPT	76
5.8	Conclusion	83
	Conclusion et perspectives	85
	Bibliographie	87

TABLE DES MATIÈRES

Table des figures

1.1	Le cerveau : représentation de ces deux hémisphères (a) et des différents lobes qui le constituent (b) [1].	4
1.2	Activité électrique du neurone. À gauche, mécanisme de la dépolarisation ; à droite propagation du courant électrique (ou potentiel d'action).	5
1.3	Déclenchement d'un potentiel post-synaptique.	6
1.4	Champs électriques générés par les courants synaptiques.	7
1.5	Schéma comparatif des résolutions temporelles et spatiales et du caractère invasif des différentes techniques d'imagerie cérébrale.	8
1.6	Dispositif pour l'acquisition du signal EEG avec 128 électrodes.[2]	9
1.7	Exemple d'un enregistrement des signaux EEG.	10
1.8	Les différentes ondes constituant le signal EEG.	11
1.9	La chaîne d'acquisition d'un signal EEG.	14
1.10	Système standard international de placement des électrodes. A) vue de la gauche, B) vue de dessus. En C) localisation et nomenclature des électrodes selon la société américaine l'électroencéphalographie	15
2.1	(a) transformée de Fourier à fenêtre (porte étroite) (b) transformée de Fourier à fenêtre (porte large)	20
2.2	La Résolution temps-fréquence avec une TF fenêtrée	21
2.3	Les composants d'ondelette d'un signal transformée en ondelette	23
2.4	Exemple d'une sinusoïde à différentes échelle	24
2.5	Translation d'une ondelette	24
2.6	Différentes résolutions pour différentes analyses d'un signal	26
2.7	Structure de la décomposition en paquets d'ondelettes à 3 niveaux	26

TABLE DES FIGURES

3.1	Schéma de la localisation des électrodes selon le système internationale 10–20 [3]	30
3.2	Exemple des signaux EEG de la base de données [4]	31
3.3	Représentation des nuages de points dans un système d’axes	33
3.4	Le changement d’origine	36
3.5	Projection orthogonale de u_i sur l’axe Δ	37
3.6	Représentation du nuage de points avec deux axes principaux	40
3.7	Histogramme des valeurs propres associées aux vecteurs propres	41
4.1	Des données linéairement séparables	45
4.2	Hyperplan : Il existe une infinité d’hyperplans pouvant séparer les données linéairement séparables.	47
4.3	La marge	48
4.4	Les vecteurs de support[5]	52
4.5	La transformation des données linéaire	54
4.6	La marge souple	56
5.1	Structure du système de classification automatique de signaux EEG. . .	62
5.2	Exemple des signaux EEG de différents Set utilisés dans ce mémoire. . .	63
5.3	Le Taux d’erreur de classification obtenu pour différentes familles d’ondelettes en utilisant la WPT.	65
5.4	Structure de la décomposition en paquet ondelette (3 niveau) utilisés dans ce mémoire.	65
5.5	La décomposition d’un signal EEG du set A en paquets d’ondelettes (WPT), avec l’ondelette mère « db1 » et un niveau de décomposition 3.	66
5.6	Le paramètre Maximum extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	67
5.7	Le paramètre minimum extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	68
5.8	Le paramètre Gamme extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	69
5.9	Le paramètre écart-type extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	70

5.10	Le paramètre Entropie extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	70
5.11	Le paramètre énergie extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.	71
5.12	Histogramme des valeurs propres et variabilités cumulées pour chaque donnée (groupe 1 avec WPT).	74
5.13	Les performances de la classification du groupe 1 par la SVM pour différents paramètres caractéristiques en utilisant la WPT pour $m = 2, 5, 10$	78
5.14	Le Taux d'erreur en fonction des paramètres caractéristiques choisis dans chaque étape de l'algorithme SVM-RFE du groupe 1 en utilisant la WPT pour $m=2, 5, 10$	81
5.15	Système de classification retenu dans cette étude	82

Liste des tableaux

4.1	Les noyaux les plus fréquemment utilisés	53
4.2	Table de diagnostique (confusion)	60
5.1	Les paramètres extraits des signaux EEG du groupe 1 pour chaque paquet d'ondelette.	72
5.2	Résultats de l'application de l'ACP sur les données (WPT) pour deux axes principaux.	75
5.3	Les performances de la classification du groupe 1 par la SVM pour différents paramètres caractéristiques en utilisant la WPT.	77
5.4	Résultat après l'utilisation de la SVM-RFE sur le groupe 1 pour $m=2, 5, 10.$, en utilisant WPT.	80
5.5	Classement en ordre croissant (du premier au dernier) des paramètres caractéristiques par la SVM-RFE du groupe 1 en utilisant la WPT pour $m=2, 5, 10.$	82

LISTE DES TABLEAUX

Liste des symboles

ACP	Analyse en composantes principales
CWT	Transformée en ondelettes continue
Ect	Ecart-type
EEG	Électroencéphalogramme
Eng	Energie
Ent	Entropie
FN	faux négatif signifie que le cas Normal est classé dans la classe des Épileptiques.
FP	faux positif signifie que le cas Épileptique est classé dans la classe des Normaux.
Gam	Gamme
Max	Maximum
Min	Minimum
QMFs	Filtre miroir quadratique
RFE	Elimantion recursive des paramètres caractéristiques.
STFT	Transformée de fourrier fenêtrée
SVM	la Machine à support vecteur.
TF	Transformée de fourrier
TN	vrai négatif signifie que le cas Épileptique est classé dans la classe des Épileptiques.
TP	vrai positif signifie que le cas normal est classé dans la classe des Normaux.
WPT	Transformée en paquet ondelettes
WT	Transformée en ondelette

Introduction

Cadre général et objectifs

Le traitement et l'analyse des signaux biomédicaux à l'aide des logiciels sont entrés en jeu depuis le début des années 1960 offrant aux médecins des moyens rapides et précis pour l'aide au diagnostique.

L'électroencéphalographie (EEG) est l'enregistrement au niveau du scalp de l'activité électrique du cerveau, il permet de mesurer l'activité cérébrale avec une grande précision temporelle. L'épilepsie constitue sans aucun doute la pathologie la mieux et la plus étudiée par électroencéphalographie.

L'objectif de notre travail, est de trouver la méthode d'analyse, les paramètres caractéristiques les plus significatifs et la méthode de classification la plus appropriée pour développer un algorithme de classification efficace et performant des signaux EEG comme une aide au diagnostique pour les praticiens en neurophysiologie, afin de les classifier en signaux EEG normaux ou épileptiques.

Contributions

Depuis les premiers jours du traitement numérique, la transformée de Fourier a été la plus utilisée pour la représentation fréquentielle du signal. Toutefois les signaux biomédicaux sont couramment caractérisés par un comportement non stationnaire et leur traitement avec la transformée de Fourier, ne donnerait pas le meilleur résultat. Par conséquent, pour ces signaux transitoires, une représentation temps-fréquence est très souhaitable dans le but de tirer les caractéristiques significatives. De la variété des approches disponibles, la transformée en ondelettes a été jugé un outil d'analyse temps-fréquence efficace pour analyser les signaux transitoires, car cette méthode unifie les différents outils qui ont été développés pour le traitement des signaux. L'extraction des paramètres caractéristiques est utilisée pour évaluer les différents phénomènes transi-

toires dans les signaux EEG dans un but de classification. Plusieurs familles d'ondelettes sont disponibles pour la caractérisation du signal et la sélection des ondelettes appropriée est très importante pour l'analyse des signaux, car elle a un impact significatif sur la qualité des résultats à l'égard du classificateur. L'ondelette mère est choisie en fonction du signal à traiter.

Le travail présenté dans ce mémoire traite la sélection de la fonction ondelette la plus appropriée pour l'analyse des signaux EEG. Ces signaux réels ont été tirés de la base de données internationale[4] et décomposés pour l'extraction des paramètres caractéristiques et leur réduction par l'analyse en composantes principales (ACP), en vue d'une classification à l'aide de la machine à vecteurs de support (SVM) combinée à l'algorithme d'élimination récursive des paramètres caractéristiques (RFE) , pour pouvoir évaluer la combinaison des paramètres donnant les meilleurs performances avec une précision de diagnostic supérieure à 98%.

Organisation du projet

Pour mener à bien notre travail, nous avons divisé notre étude en cinq chapitres :

1. Le premier chapitre traitera de l'aspect médical de l'encéphalographie et donnera une description de l'acquisition des signaux EEG.
2. Le deuxième chapitre décrira le traitement et l'analyse par la transformée en ondelette.
3. Le troisième chapitre présentera la base de données utilisée, l'extraction des paramètres caractéristiques et leur réduction dimensionnelle.
4. Le quatrième chapitre traitera la méthode de classification et les critères de performances utilisés.
5. Le cinquième chapitre présentera les résultats expérimentaux suivis d'une discussion.

Nous terminerons ce mémoire par une conclusion et quelques perspectives.

Chapitre 1

Contexte médical

1.1 Introduction

Le cerveau est l'organe le plus complexe du corps humain. Il régit notre comportement, nos actions et nos pensées, nos désirs et nos instincts. Grâce à lui, nous pouvons voir, sentir ou entendre, parler et marcher, analyser et comprendre le monde qui nous entoure. Pour cela, il suscite autant d'interrogations que de fascination.

Dans ce chapitre on va voir quelques bases anatomiques du cerveau, les caractéristiques du signal électroencéphalogramme (EEG), et la façon d'effectuer sa mesure.

1.2 Le cerveau : bases anatomiques et physiologiques

1.2.1 Anatomie du cerveau

Le cerveau possède deux moitiés relativement symétriques appelées *des hémisphères*, l'une à droite et l'autre à gauche (Figure 1.1 (a)). Toute la couche externe du cerveau, la substance grise qui présente des circonvolutions complexes en périphérie est appelée *le cortex cérébral*. Le cortex de chaque hémisphère est divisé en quatre lobes (Figure 1.1 (b)) : *le lobe pariétal*, *le lobe occipital*, *le lobe frontal* et *le lobe temporal*, dont l'appellation est directement inspirée du nom des os crâniens qui les recouvrent [6]. Les lobes sont impliqués dans différentes fonctions [7].

- Les lobes occipitaux, localisés à l'arrière du cortex, sont spécialisés dans la vision.
- Les lobes pariétaux sont situés à l'avant des lobes occipitaux. Ils sont impliqués

dans le sens du toucher, dans la détection de mouvements dans l'environnement et la localisation des objets dans l'espace.

- Les lobes frontaux sont impliqués dans le mouvement, l'attention, la planification, les compétences sociales, la pensée abstraite, la mémoire et certains aspects de la personnalité.
- Les lobes temporaux sont importants dans l'audition, le langage et la connaissance visuelle des objets [1].

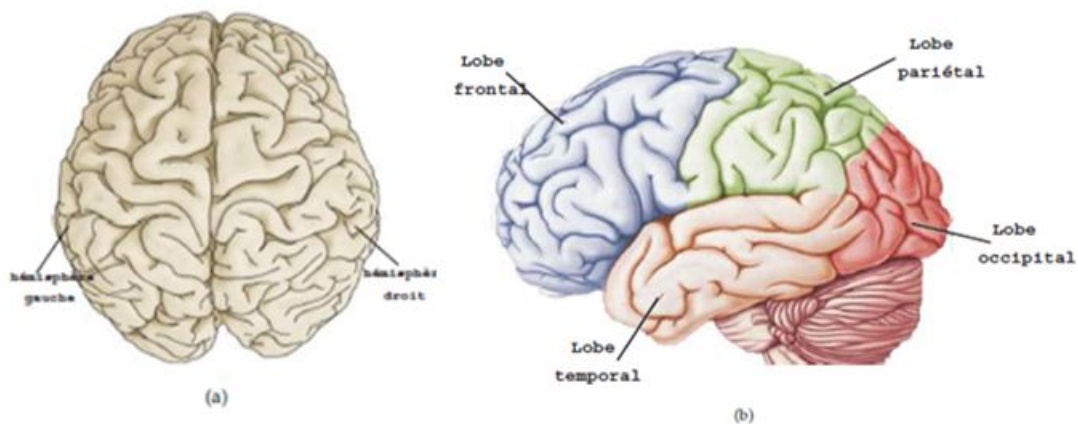


FIGURE 1.1: Le cerveau : représentation de ces deux hémisphères (a) et des différents lobes qui le constituent (b) [1].

1.2.2 Activité électrique

Le cerveau humain est composé de 10 à 100 milliards de neurones dont les corps cellulaires constituent la matière grise. Le rôle principal des neurones est de recevoir, traiter et transmettre l'information. Les échanges d'informations entre cellules sont réalisés au niveau des synapses. La synapse est l'unité structurale et fonctionnelle de connexions entre deux neurones. La membrane des neurones possède la propriété d'être excitable.

Deux mécanismes sont à l'œuvre pour la communication neuronale :

- La transmission de l'influx nerveux le long des fibres nerveuses par le potentiel d'action.
- La transmission synaptique qui fait le lien entre deux neurones.

Le potentiel d'action est une onde bi-phasique (dépolérisation puis hyper polarisation de la membrane) de quelque milliseconde, d'amplitude $1.10^{-13}mA$, qui se propage le long de l'axone.

À l'état de repos, il existe une différence de potentiel entre l'extérieur et l'intérieur d'environ -60 à -70 mv due à une différence de concentration ionique. Les ions potassium ($K+$) sont plus concentrés à l'intérieur de la membrane du neurone, alors que les ions sodium ($Na+$) sont plus concentrés à l'extérieur[8].

Les neurones échangent des influx nerveux véhiculés par la «dépolérisation » de leur membrane (Figure 1.2) : lors d'une stimulation de la membrane, les ions sodium traversent celle-ci à l'emplacement de la stimulation, puis cette réaction se propage le long de la membrane ce qui crée le potentiel d'action du neurone. (Figure 1.3)

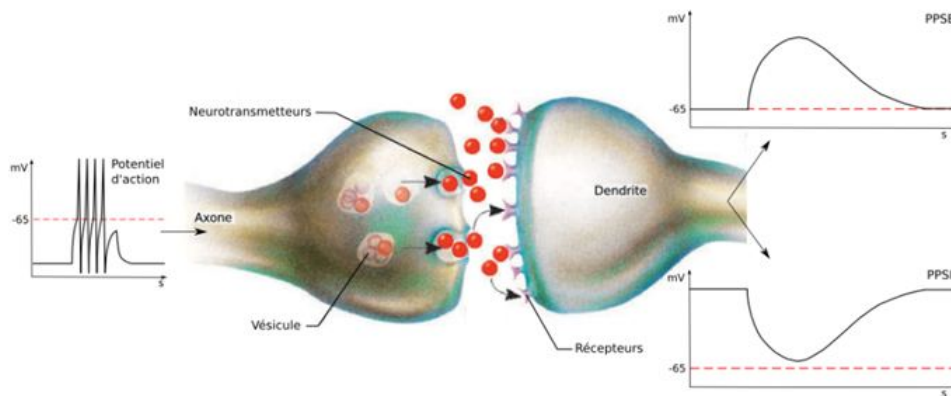


FIGURE 1.2: Activité électrique du neurone. À gauche, mécanisme de la dépolérisation ; à droite propagation du courant électrique (ou potentiel d'action).

Ces échanges d'ions chargés électriquement produisent un courant électrique qui peut être capté si une électrode est placée à l'extérieur du neurone. Si plusieurs neurones sont dans le voisinage de l'électrode, l'électrode enregistre l'ensemble des activités de ces neurones.

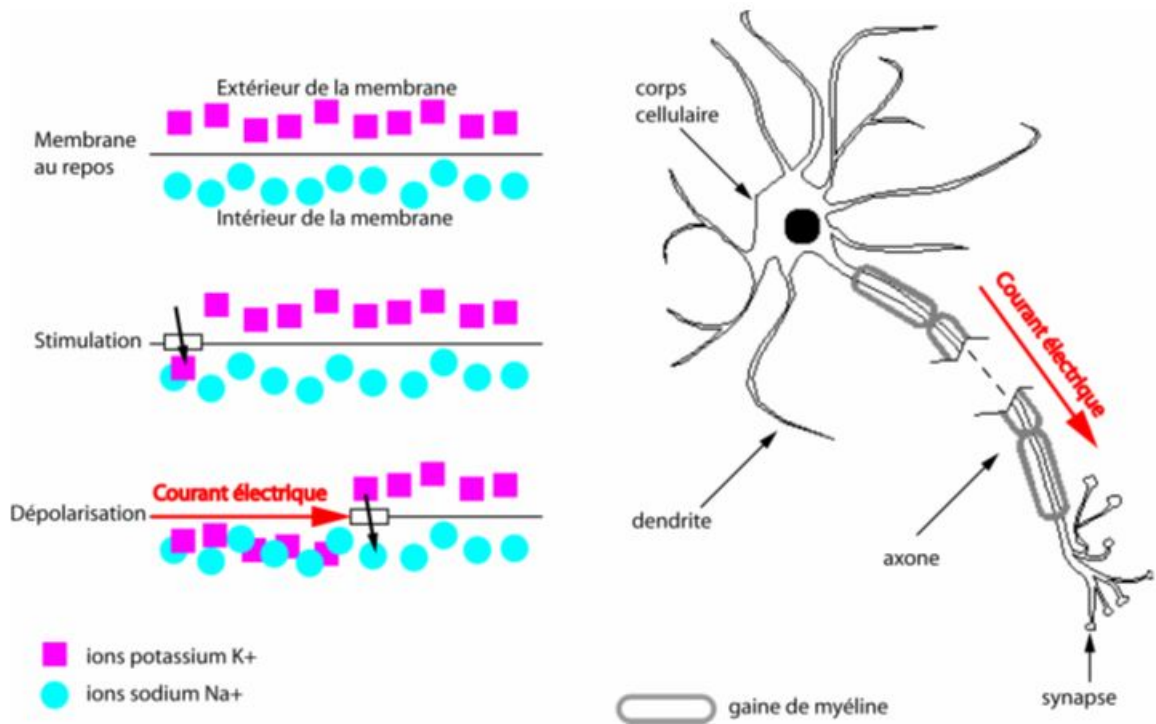


FIGURE 1.3: Déclenchement d'un potentiel post-synaptique.

Dans le neurone, lorsqu'un stimulus provoque l'apport d'ions sodium dans la membrane, une activité électrique se propage le long de l'axone. Les gaines isolantes de myéline permettent d'accélérer la circulation de la dépolarisation, seuls les nœuds entre ces parties isolées se dépolarisent. Après la phase de dépolarisation, suit une phase de repolarisation de la membrane. Enfin, les ions retournent à leurs positions de départ : l'action des « pompes à sodium/potassium » replace les ions potassium à l'extérieur et les ions sodium à l'intérieur de la membrane.

L'activité électrique des neurones est enregistrée par une électrode placée sur le scalp. Si les neurones sont orientés parallèlement, comme c'est le cas pour les cellules pyramidales du cortex représentées ici (en colonnes verticales parallèles), les courants de la population de neurones peuvent se sommer ; sinon les cellules annulent mutuellement leurs signaux (Figure 1.4).

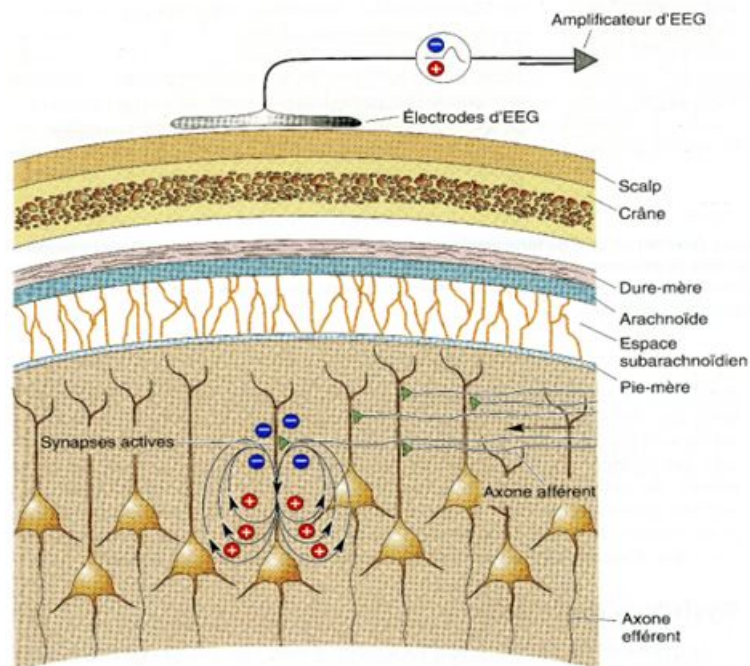


FIGURE 1.4: Champs électriques générés par les courants synaptiques.

1.2.3 Mesure de l'activité cérébrale

Il existe plusieurs types de modalités pour explorer les activités cérébrales, la Figure 1.5 illustre la variabilité des résolutions temporelles et spatiales parmi les méthodes d'exploration cérébrales (techniques d'imagerie et techniques électro-physiologiques). En principe les méthodes peuvent être groupées en deux classes : l'imagerie et l'exploration électro-physiologiques. Nous détaillons ici plus particulièrement une méthode électro-physiologique : l'électroencéphalogramme (EEG), Il existe d'autres méthodes : la magnéto-encéphalographie (MEG), Les EEG invasives (la stéréo-électroencéphalographie (SEEG) et l'électro-corticographie (ECoG), ainsi des méthodes d'imageries. Nous pouvons classer en trois groupes les différentes modalités électro-physiologiques qui permettent d'enregistrer les signaux de l'activité électrique cérébrale : méthodes non invasive, semi-invasive et invasive.

Parmi les méthodes d'analyse physiologiques non invasive, on distingue l'électroencéphalogramme de surface classique et la magnéto-encéphalographie, qui est une technique

relativement récente, nous parlerons plus précisément des premières méthodes (EEG) dans les paragraphes suivants.

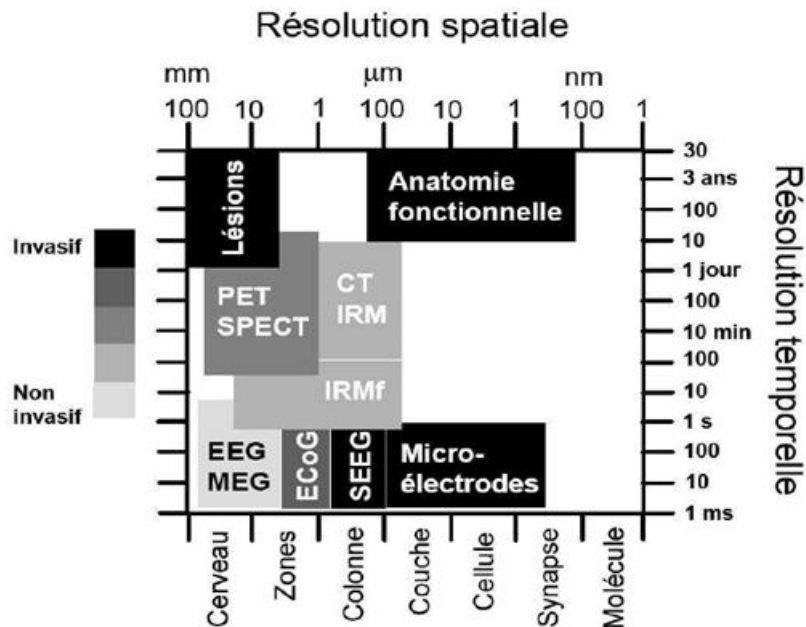


FIGURE 1.5: Schéma comparatif des résolutions temporelles et spatiales et du caractère invasif des différentes techniques d'imagerie cérébrale.

1.3 Électroencéphalogramme (EEG)

1.3.1 Histoire

L'invention de l'électroencéphalographie est généralement attribuée au physiologiste allemand *Hans Berger*, qui commença à l'étudier dans les années 1920 et enregistra le premier signal d'activité cérébrale en 1929. Ses travaux furent repris et complétés par le britannique *Edgar Douglas Adrian*, qui obtint en 1932 le Prix Nobel de physiologie. L'EEG ne s'est vraiment développé et répandu dans la pratique médicale courante, en particulier dans la pratique épileptologique, que dans les années 1950. Historiquement les courbes d'EEG étaient tracées sur des rouleaux de bandes de papier millimétrés

afin d'être relus ensuite par les médecins neurologues pour y déceler d'éventuels signes. Aujourd'hui, le signal est numériquement converti et traité par ordinateur. Le développement de l'EEG en clinique date de l'apparition des amplificateurs différentiels au début des années 1960. Les EEG sont actuellement utilisés afin de diagnostiquer des foyers épileptiques, des tumeurs cérébrales, des lésions. Ils permettent aussi de trouver l'origine de migraines, de problèmes d'étourdissements, ou encore de somnolence.

1.3.2 Définition

L'électroencéphalogramme (EEG) est le moyen le plus utilisé pour mettre en évidence l'activité électrique du cerveau, donc l'activité épileptique proprement dite. Comparé aux autres techniques d'investigation, L'EEG fournit des informations en temps réel et avec une très bonne résolution temporelle, de l'ordre de la milliseconde. Il demeure incontournable pour le diagnostic et la classification des épilepsies. *Berger* a mis cela en évidence chez l'homme en 1924 [6].

L'EEG est le résultat d'une différence de potentiel mesurée entre deux points. L'enregistrement entre deux points se fait grâce à des électrodes placées sur le scalp (Figure 1.6) qui ouvre une fenêtre pour explorer l'activité neuronale et le fonctionnement du cerveau [5], La Figure 1.7 présente un exemple d'un enregistrement EEG. Ces potentiels varient au cours du temps et forment des rythmes représentées dans la Figure 1.8. La morphologie des tracés dépend de l'état d'activation des neurones enregistrés et cet état d'activation va se traduire par des variations de fréquence et d'amplitude des ondes enregistrées.



FIGURE 1.6: Dispositif pour l'acquisition du signal EEG avec 128 électrodes.[2]

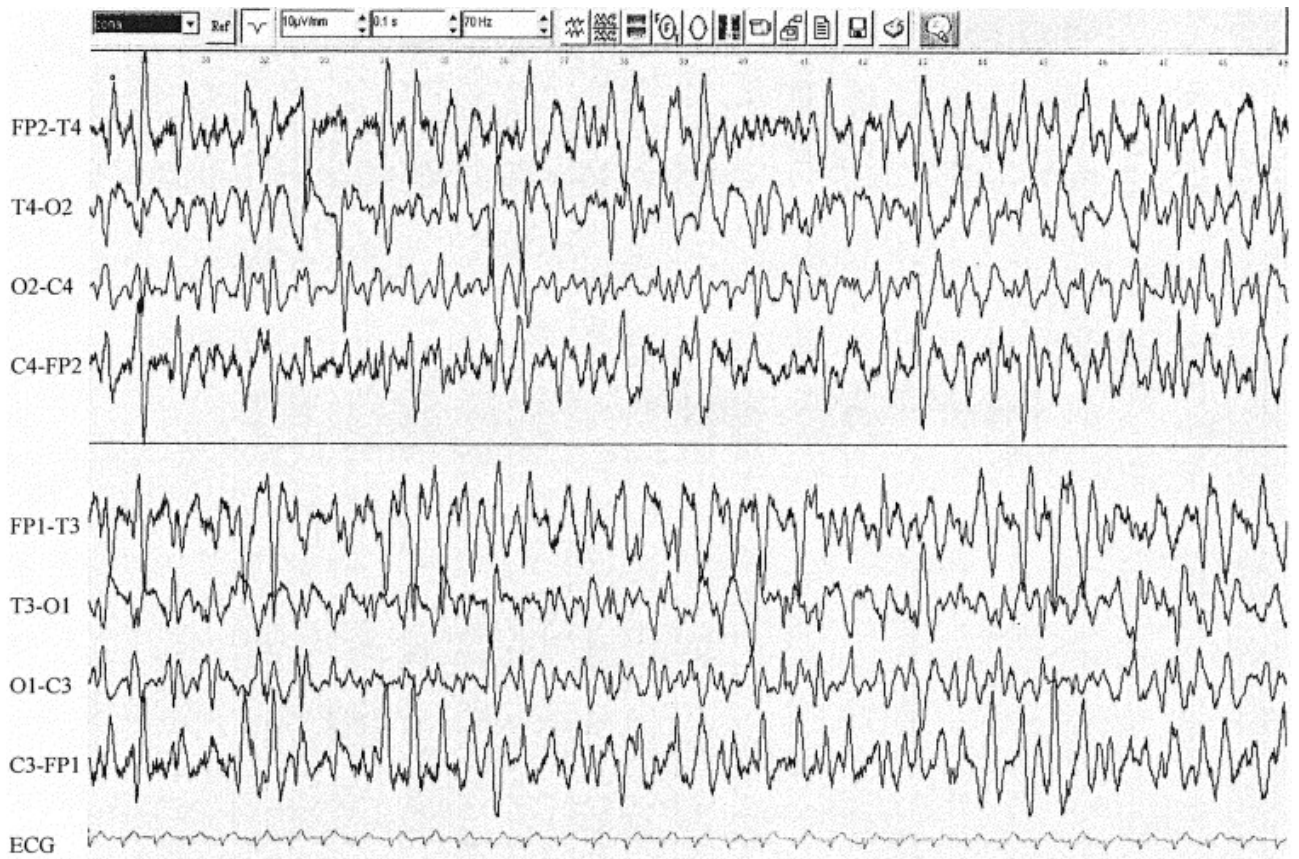


FIGURE 1.7: Exemple d'un enregistrement des signaux EEG.

1.3.3 Les différents rythmes du signal EEG

L'analyse des signaux EEG ou des ondes cérébrales est complexe, en raison de la grande quantité d'informations reçues de chaque électrode. Comme de nombreuses stations de radio, les signaux EEG sont classés par la fréquence de leurs émanations et, dans certains cas, par la forme de leurs d'ondes. Bien qu'aucune de ces ondes n'est jamais émise seule, l'état de conscience des individus peut prendre une gamme de fréquences plus prononcée que d'autres types ; 6 types sont particulièrement importants (Figure 1.8) :

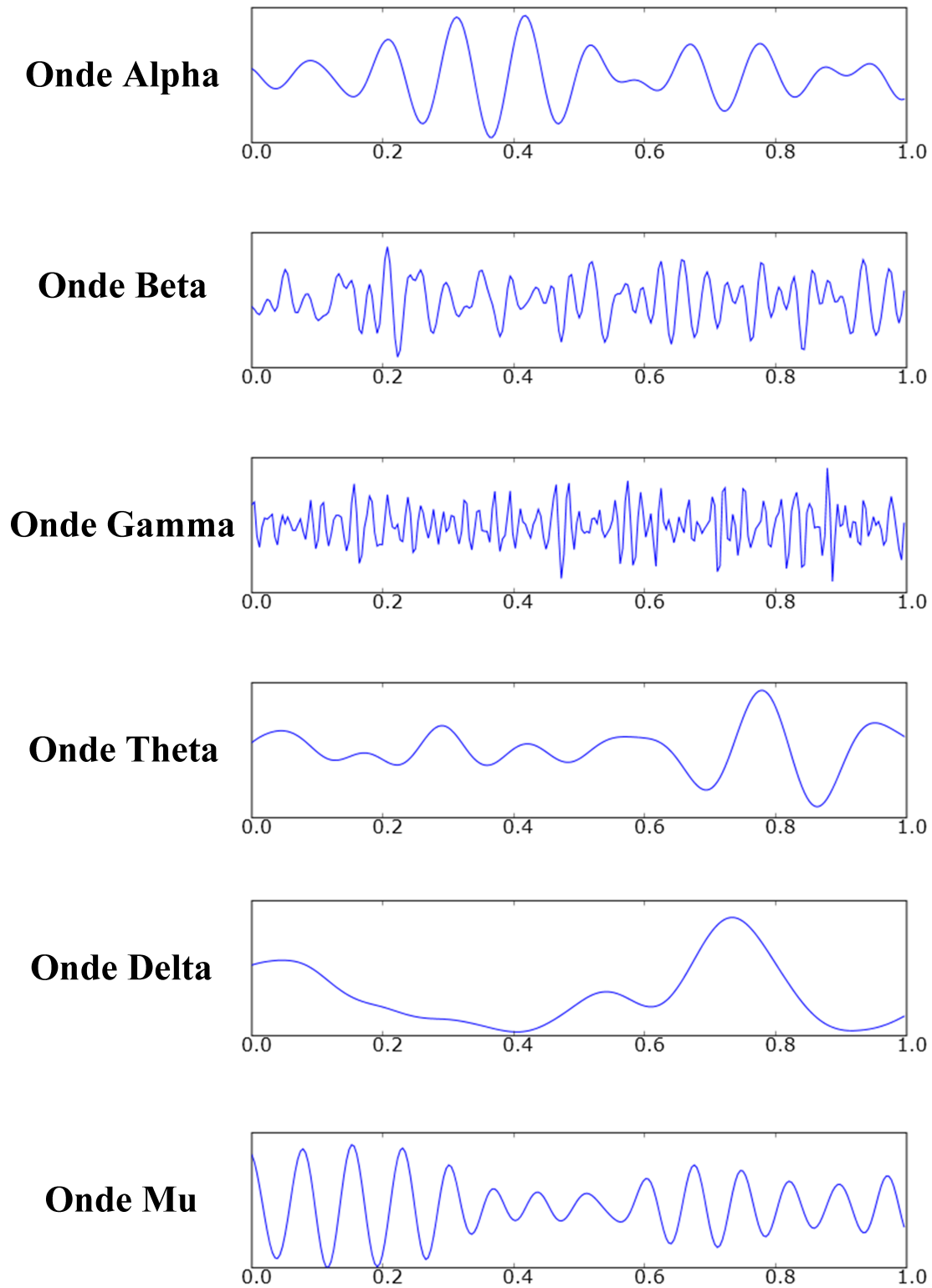


FIGURE 1.8: Les différentes ondes constituant le signal EEG.

- **Ondes ALPHA α** : La fréquence se situe entre 8 à 13 Hz, avec des amplitudes 30-50 μV . Les ondes alpha indiquent à la fois une prise de conscience détendue et aussi l'inattention. Elles ont de fortes amplitudes sur l'occipital (arrière du crâne) du cortex et également sur le cortex frontal.
- **Ondes BÊTA β** : La fréquence se situe entre 14 et 20 Hz, elle a généralement une faible tension entre 5-30 μV . Bêta est une onde du cerveau qui est généralement associée à la pensée active, une attention active. Se concentrer sur le monde extérieur ou de résoudre des problèmes concrets. Elle peut atteindre des fréquences proches de 50 hertz durant l'activité mentale intense.
- **Ondes GAMMA γ** : Les ondes gamma varient entre une fréquence de 20 Hz à 60 Hz ; et caractérisée par une amplitude très faible, elle est associée à des fonctions motrices et cognitives (perception, raisonnement, langage, mémoire).
- **Ondes THÊTA θ** : La fréquence se situe entre 4 à 7 Hz, avec une amplitude généralement supérieure à 20 μV . Thêta provient du stress émotionnel, en particulier de la frustration et la déception. L'onde Thêta a été également associée à l'accès inconscient, à l'inspiration créatrice et à la méditation profonde. Le grand pic dominant de l'onde thêta est d'environ 7 Hz.
- **Ondes DELTA Δ** : Les ondes delta se situent dans la gamme de 0, 5 à 4 Hz, avec des amplitudes variables. Elles sont principalement associées à un sommeil profond, et à l'état de veille.
- **Ondes MU μ** : Il s'agit d'une onde spontanée de 8 à 12 Hz associée à des activités motrices, et l'enregistrement est maximum sur le cortex moteur .Elles diminuent avec le mouvement ou l'intention de se déplacer. L'onde Mu se situe dans la même bande de fréquence que l'onde alpha, mais cette dernière est enregistrée sur le cortex occipital.

1.4 La mesure des signaux EEG

Plusieurs difficultés existent lors de l'acquisition d'un signal EEG avec des électrodes de surface :

- La faible amplitude du signal mesuré (de 10 à 200 μV) causées par la mauvaise conduction du crâne.

- Le bruit de mesure, dû à l'éloignement des sources; la qualité du signal EEG recueilli est testé en mesurant l'impédance de l'ensemble constitué par chaque électrode et son câble, qui doit être inférieure à 10 $K\Omega$, voire à 5 $K\Omega$ pour que les signaux soient interprétables[9].

- La présence d'artéfacts :

- Ils peuvent être d'origine corporelle, c'est-à-dire dus à des mouvements oculaires, aux activités musculaire, respiratoire et cardiaque, au mâchonnement, à la déglutition, à la toux... De plus, il existe des pensées parasites qu'il est difficile de contrôler.

- Les artéfacts peuvent aussi être d'origine extra-corporelle; ils peuvent être induits par l'environnement immédiat, dépendre de l'isolement électrique.

Dans le cas des techniques non invasives, le problème d'une mauvaise résolution du signal à cause de son atténuation par les tissus et os crâniens a souvent été souligné. La forme ronde du crâne vient également entraîner une dispersion et un brouillage des ondes créés par les neurones. Ainsi, même si les ondes peuvent tout de même être détectées, il est plus difficile, par rapport aux méthodes invasives, de déterminer le secteur du cerveau qui les a créés ou les actions générées par différents neurones, d'autant plus que plusieurs régions différentes du cerveau peuvent réagir au même évènement. Par ailleurs, la méthode non invasive nécessite un entraînement intensif du sujet à la maîtrise de cette technologie.

L'acquisition d'un signal EEG comporte plusieurs étapes comme l'indique la Figure 1.9 :

- La mesure de l'activité bioélectrique au moyen d'un casque à électrodes.
- L'amplification du signal (par un facteur 1000 au moins), le signal initial étant de l'ordre du μV , afin d'obtenir des mV .
- Le filtrage du signal (filtrage passe-bas avec une fréquence de coupure à 100 Hz).
- La conversion analogique-numérique dans le but de traiter le signal.

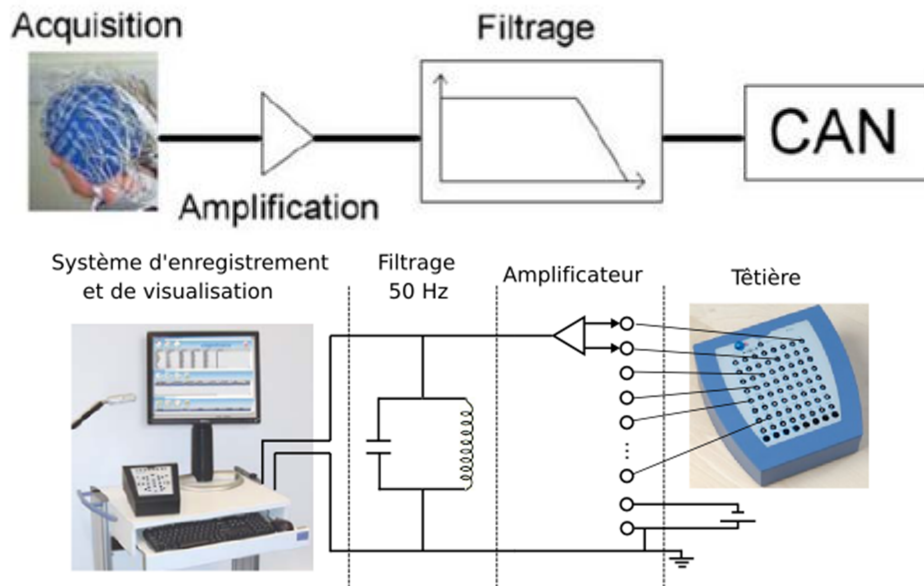


FIGURE 1.9: La chaîne d'acquisition d'un signal EEG.

1.5 L'emplacement d'électrodes : Le système 10–20

L'emplacement des électrodes utilisées en EEG a été standardisé par une nomenclature internationale : le système 10–20. Cette technique est basée sur des repères anatomiques : le nasion, l'inion et les repères pré-auriculaires droit et gauche. La première mesure est faite sur le plan antéro-postérieur : la distance nasion- inion est prise et divisée en cinq segments. Le premier repère est placé à 10% de la distance totale à partir du nasion et est appelé Fp (frontopolaire). Les repères suivants sont placés à 20% de la distance totale et appelés F(Frontale), C(Centrale), P(Pariétale) et O(Occipitale). Une méthode similaire est utilisée pour définir les emplacements dans le plan transversal. Pour distinguer les électrodes, des chiffres sont utilisées : paires pour l'hémisphère droit et impaires pour l'hémisphère gauche (voir Figure 1.10). Le développement des meilleures technologies d'enregistrement ainsi que l'amélioration des électrodes ont permis que ce montage classique puisse être augmenté avec des électrodes supplémentaires placées dans

des positions intermédiaires. Aujourd'hui, il est possible d'enregistrer 64, 128 et jusqu'à 256 voies. Dans ce cas, on parle d'EEG à haute résolution (EEG-HR).

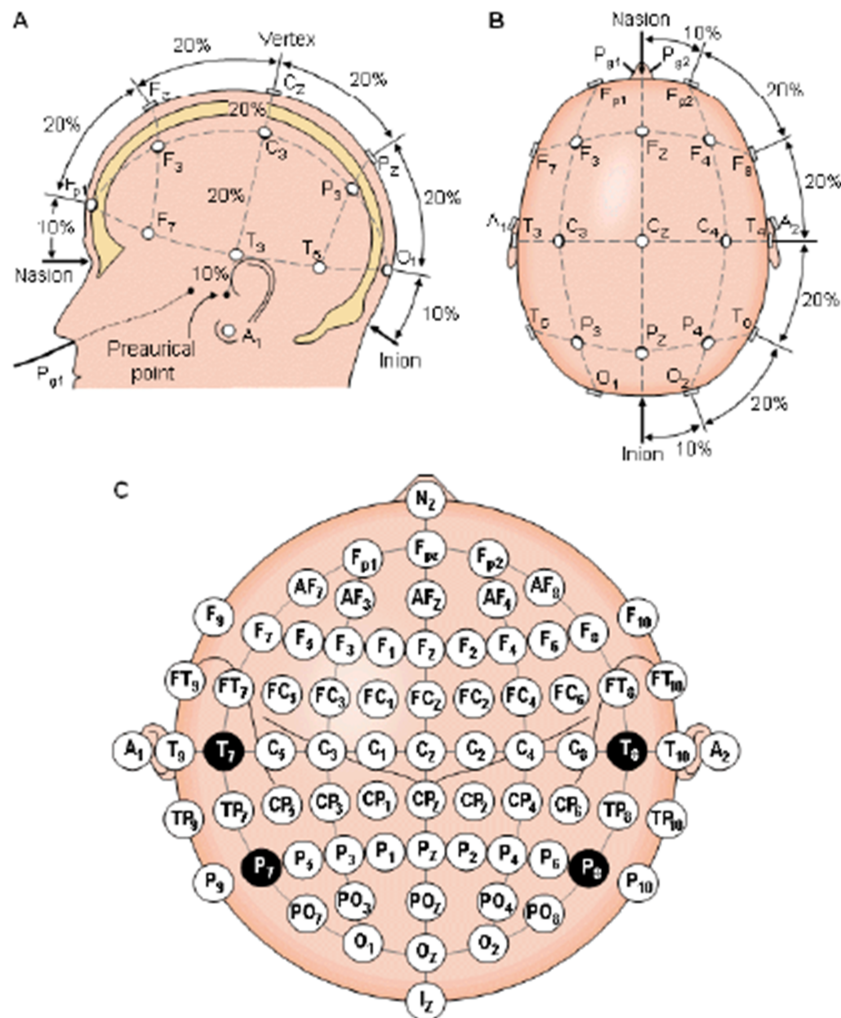


FIGURE 1.10: Système standard international de placement des électrodes. A) vue de la gauche, B) vue de dessus. En C) localisation et nomenclature des électrodes selon la société américaine l'électroencéphalographie

1.6 Conclusion

j'ai présenté, dans ce chapitre, les bases de la neurophysiologie allant du fonctionnement du neurone à celui des différents rythmes du signal EEG ainsi que la méthode d'acquisition de ce dernier. Dans le chapitre suivant, nous allons présenter par la suite les techniques d'analyse des signaux EEG. A partir de cette analyse, nous pourrons extraire des informations, que j'utiliserai pour la classification de ces signaux.

Chapitre 2

Transformations à base d'ondelettes et de paquets d'ondelettes

2.1 Introduction

Les oscillations enregistrées au moyen de électroencéphalogramme sont des évènements transitoires ; l'observation de ce signal au court du temps permet de connaître sa durée et aussi ces variations qualitatives, néanmoins, souvent pour les signaux biomédicaux la détections des pathologies se fait à l'aide des informations de fréquences, il est nécessaire alors d'avoir un outil d'analyse qui permet un passage du domaine *temporel* au domaine *fréquentiel*.

Dans ce chapitre j'étudierai les différentes méthodes de transformations et d'analyse du signal EEG telles que la transformée de fourrier (TF), la transformée de fourrier fenêtrée (STFT) et la transformée en ondelette (WT) qui permettent une représentation adéquate et significative de ce signal.

2.2 Analyse temps-fréquence

En pratique la plupart des signaux sont représentés dans le domaine temporel, la représentation du signal est donc *temps-amplitude*. Dans les traitements des signaux, en particuliers les signaux EEG, c'est l'information de fréquence qui est utilisée pour détecter l'existence d'une pathologie. Un symptôme de maladie peut parfois être mieux diagnostiqué quand on analyse les composantes fréquentielles du signal de plus, l'information qui ne peut pas être distinguée dans le domaine temporel pourrait être facilement visible dans le domaine fréquentiel. La représentation *temps-amplitude* ne permet pas d'analyser cette composante fréquentiel ; d'où la nécessité d'une transformation du signal dans le domaine fréquentiel.

2.2.1 La transformée de fourrier (TF)

La transformée de fourrier (TF) permet de mesurer et de trouver le contenu fréquentiel d'un signal linéaire et stationnaire, si on effectue la TF d'un signal représenté dans le domaine temporel, on obtient la représentation *fréquence-amplitude* de ce signal. Le spectre de fréquence d'un signal est constitué par les composantes de fréquence de ce signal. C'est une transformation réversible entre le signal brut et le signal traité (transformé).

Soit $x(t)$ un signal non-périodique, alors, la transformée de Fourier de $x(t)$, notée $X(f)$ Ou $f\{x(t)\}$, et la transformée de Fourier inverse, notée $f^{-1}\{X(f)\}$ sont données comme suit :

$$X(f) = \int_{-\infty}^{+\infty} x(t). \exp(-2\pi ft). dt \quad (2.1)$$

$$x(t) = \int_{-\infty}^{+\infty} X(f). \exp(2\pi ft). df \quad (2.2)$$

Malgré son avantage pour analyser les signaux linéaires et stationnaires, cette technique a plusieurs limitations en pratique, en particulier son manque évident de localisation temporelle, elle ne permet pas l'étude de signaux dont la fréquence varie dans le temps (signaux non-stationnaire). De tels signaux nécessitent une analyse temps-fréquence qui permettra une localisation et observation des composantes qui constituent le signal dans le temps.

La TF n'est donc pas l'outil adéquat pour les signaux non stationnaires ou transitoire, à une exception : elle peut être utilisée pour les signaux non stationnaires si on ne s'intéresse qu'aux composantes spectrales qui existent dans le signal et non aux instants où elles apparaissent. Si on veut savoir quelles composantes spectrales apparaissent et à quels instants, alors la TF n'est pas la transformée à utiliser.

Dans Les signaux EEG, les instants ou les intervalles du temps auxquels une composante spectrale particulière apparaît sont d'un grand intérêt. Dans ces cas, il est très important de connaître les intervalles de temps d'apparition de ces composantes.

Pour cela, nous pourrons utiliser une parmi les méthodes d'analyse *temps-fréquence* suivantes :

- La Transformée de Fourier fenêtrée (Short Time Fourier Transform (STFT)).
- La Transformée en Ondelettes (Wavelet Transform (WT)).

2.2.2 La transformée de fourrier fenêtrée (STFT)

Définition

Pour donner un pouvoir de localisation aux fonctions analysantes de la transformée de Fourier, qui oscillent avec la même amplitude sur tout l'axe des réels, on pondère ces fonctions par une fonction fenêtrée de manière à sélectionner uniquement la partie utile du signal. La fenêtrée est bien sûr translatée de manière à observer toutes les parties utiles du signal ; la transformée de Fourier à fenêtrée glissante s'exprime par :

$$X(f, \tau) = \int_{-\infty}^{+\infty} x(t) \cdot \exp(-2\pi f t) \cdot W^*(t - \tau) dt \quad (2.3)$$

Où $W(t)$ est la fonction fenêtrée qui est à choisir et τ le paramètre de translation de la fenêtrée ; on notera que la transformée dépend maintenant de deux variables : une variable de fréquence et une variable de localisation temporelle du contenu fréquentiel. Cette transformée nous permet donc bien d'atteindre le but recherché qui était d'avoir des informations sur le signal en temps et en fréquence à partir de la transformation réalisée. Une question qui se pose est de savoir comment choisir cette fonction fenêtrée.

Choix de la fonction fenêtrée

La forme la plus simple de fenêtrée semble être la fonction porte, qui vaut "un" à l'intérieur de la fenêtrée et "zéro" partout ailleurs. En supposant ce choix de forme de fenêtrée pertinent, il nous reste à choisir la largeur de cette fenêtrée[10].

- Fenêtrée étroite \Rightarrow bonne résolution temporelle, mauvaise résolution fréquentielle
- Fenêtrée large \Rightarrow bonne résolution fréquentielle, mauvaise résolution temporelle

Ces observations mettent en évidence un compromis entre résolution fréquentielle et résolution temporelle qui correspond au principe d'incertitude d'*Heisenberg* formulé en mécanique quantique qui s'applique à toute paire de variables dites conjuguées (temps et fréquence).

Le principe d'incertitude peut se quantifier en ayant défini au préalable un paramètre qui caractérise la résolution. En notant $\psi(t)$ une fonction d'énergie finie (dans le cas de la TF à fenêtrée ce serait la fonction $\exp(j2\pi f t) \cdot w * (t)$), nous caractériserons la résolution temporelle et la résolution fréquentielle par :

$$\begin{cases} \sigma_t = \sqrt{\frac{\int_{-\infty}^{+\infty} t^2 \cdot |\psi(t)|^2 \cdot dt}{\int_{-\infty}^{+\infty} |\psi(t)|^2 \cdot dt}} \\ \sigma_f = \sqrt{\frac{\int_{-\infty}^{+\infty} f^2 \cdot |\Psi(f)|^2 \cdot dt}{\int_{-\infty}^{+\infty} |\Psi(f)|^2 \cdot dt}} \end{cases} \quad (2.4)$$

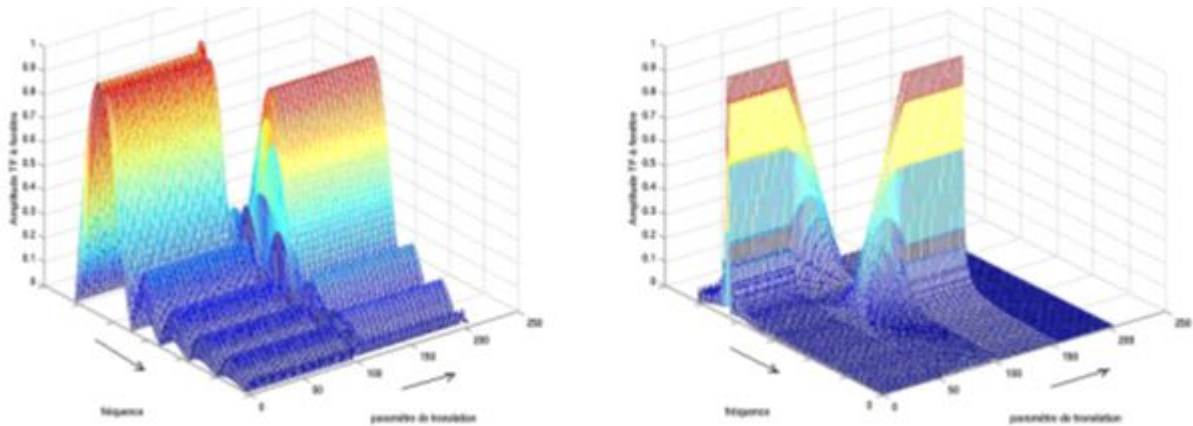


FIGURE 2.1: (a) transformée de Fourier à fenêtre (porte étroite) (b) transformée de Fourier à fenêtre (porte large)

Ces dispersions ou écarts-types caractérisent "l'étendue" spatio-temporelle de la fonction considérée ; plus la dispersion est petite meilleure est la résolution.

Dans le plan temps fréquence, la majeure partie de l'énergie de la fonction considérée est située dans un rectangle de dimension $\sigma_t \times \sigma_f$. Le principe d'incertitude se traduit alors par le théorème suivant :[10]

Relation d'incertitude

Pour toute fonction d'énergie finie $\psi(t)$ telle que les fonctions $t\psi(t)$ et $f\Psi(t)$ soient également d'énergie finie, on a :

$$\sigma_t \times \sigma_f \geq \frac{1}{4\pi} \quad (2.5)$$

σ_t : La Résolution temporelle , σ_f : La Résolution fréquentiel.

On pourrait dire que ' l'image ' que nous fournit la transformée de Fourier à fenêtre du signal ne peut pas être vue avec des pixels dont l'aire serait plus petite que $\frac{1}{4\pi}$ [10].

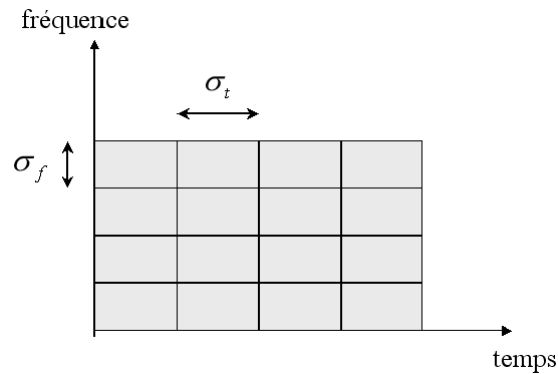


FIGURE 2.2: La Résolution temps-fréquence avec une TF fenêtrée

2.2.3 La transformée en ondelette WT

La transformée en ondelette a été développée pour dépasser les limitations que présentent les transformées (TF, STFT) pour le traitement des signaux non-stationnaires, tels que : la localisation temporelle des composantes fréquentielles de la TF, et le problème de la résolution de la STFT (la STFT donne une résolution fixe pour tous les instants de temps alors que la WT donne une résolution variable).

2.2.3.1 Histoire des ondelettes

D'un point de vue historique, l'analyse en ondelettes est une nouvelle méthode, bien que ces mathématiques remontent aux travaux de *Joseph Fourier* dans le dix-neuvième siècle. Fourier a prévu les fondations avec ses théories de l'analyse de fréquence, qui s'est avéré être extrêmement importante et influente dans le traitement du signal.

L'attention des chercheurs s'est dirigé peu à peu de l'analyse basée sur la fréquence à l'analyse basée sur l'échelle quand il a commencé à devenir clair que l'approche de mesure des fluctuations moyennes à différentes échelles pourrait s'avérer moins sensible au bruit.

La première mention de ce que nous appelons aujourd'hui une «*ondelettes*» semble être en 1909, dans une thèse par *Alfred Haar*. Le concept d'ondelettes dans sa forme théorique actuelle a été proposé par *Jean Morlet* et l'équipe au Centre de Physique Théorique de Marseille. Les méthodes d'analyse en ondelettes ont été développées principalement par *Y. Meyer* et ses collègues, qui ont assuré la diffusion des méthodes. L'algorithme principal remonte à l'œuvre de *Stéphane Mallat* en 1988. Depuis la recherche sur les ondelettes est devenue internationale. Cette recherche est particulièrement active dans les *États-Unis*, où il est dirigé par le travail des scientifiques tels que *Ingrid Daubechies*, *Ronald Coifman*, et *Victor Wickerhauser*. Le domaine des ondelettes grandit très rapidement ;

un grand nombre de documents mathématiques et pratiques essais sont publiés chaque mois.

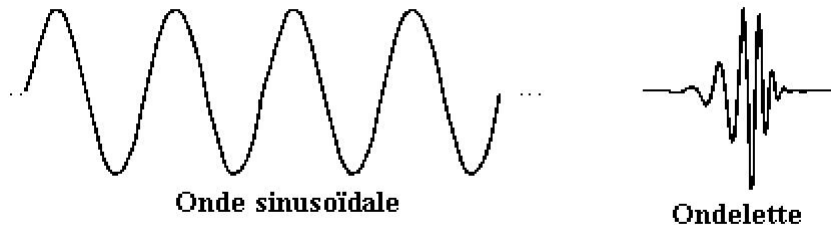
2.2.3.2 La transformée en ondelette continue (CWT)

Définition

Comme la transformée de Fourier, la transformée en ondelettes continue (CWT) utilise des produits internes pour mesurer la similitude entre un signal et une fonction analysante.

Dans la transformée de Fourier, les fonctions d'analyse sont exponentielles et complexes $\exp(j2\pi ft)$. La transformée résultante est une fonction d'une variable unique, f . Dans la transformée de Fourier à court terme (STFT), les fonctions d'analyse sont exponentielles complexes fenêtrées $w(t) \cdot \exp(j2\pi ft)$, et le résultat dans une fonction de deux variables. Les coefficients de STFT $F(f, \tau)$, représentent la correspondance entre le signal et une sinusoïde avec la fréquence f dans un intervalle d'une longueur spécifiée centrée à τ [11].

Une ondelette est une forme d'onde d'une durée limitée qui a une valeur moyenne de zéro contrairement aux sinusoïdes, qui sont la base de l'analyse de Fourier ; les ondelettes ont tendance à être irrégulière et asymétrique.



Dans la CWT, la fonction d'analyse est une ondelette ψ elle compare le signal de versions décalées et compressée (ou dilatées) d'une ondelette ; la compression d'une fonction est dénommé dilatation ou mise à l'échelle et correspond à la notion physique de l'échelle.

En comparant le signal de l'ondelette à différentes échelles et positions, on obtient une fonction de deux variables. La représentation bidimensionnelle d'un signal à une dimension est redondante.

- Si l'ondelette est à valeurs complexes, le CWT est une fonction complexe de l'échelle et de la position.
- Si le signal est à valeurs réelles, la CWT est une fonction réelle de l'échelle et de la position.

Pour un paramètre d'échelle $a > 0$, et de position b , la CWT pour un signal $f(t)$ est définie par :

$$C(a, b) = \int_{-\infty}^{+\infty} f(t) \cdot \frac{1}{\sqrt{a}} \cdot \psi^*\left(\frac{t-b}{a}\right) \cdot dt \quad (2.6)$$

Où $*$ désigne le conjugué complexe. $a \neq 0$

Les valeurs de la CWT sont affectées non seulement par le choix de l'échelle et la position, mais aussi par le choix de l'ondelette mère.

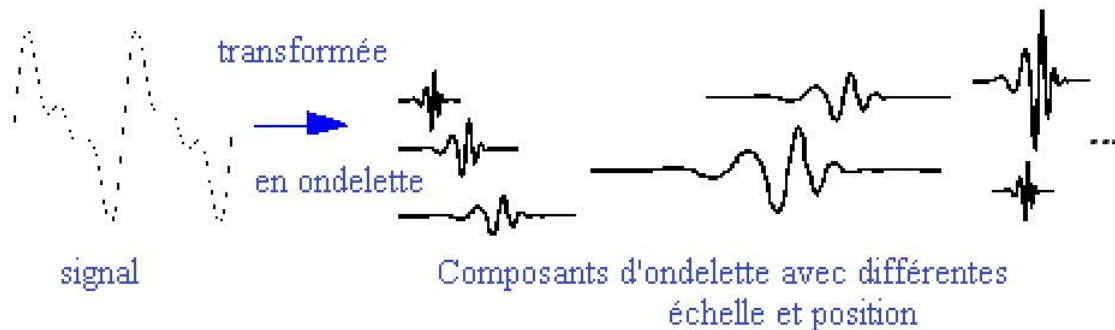


FIGURE 2.3: Les composants d'ondelette d'un signal transformée en ondelette

L'échelle (compression ou dilatation)

Pour aller au-delà des descriptions familières telles que «dilatation» ou «compression», nous introduisons le facteur d'échelle, souvent désigné par la lettre a qui est une quantité positive, $a > 0$. Pour une sinusoïde, l'effet du facteur d'échelle est très facile à voir dans la Figure 2.4 :

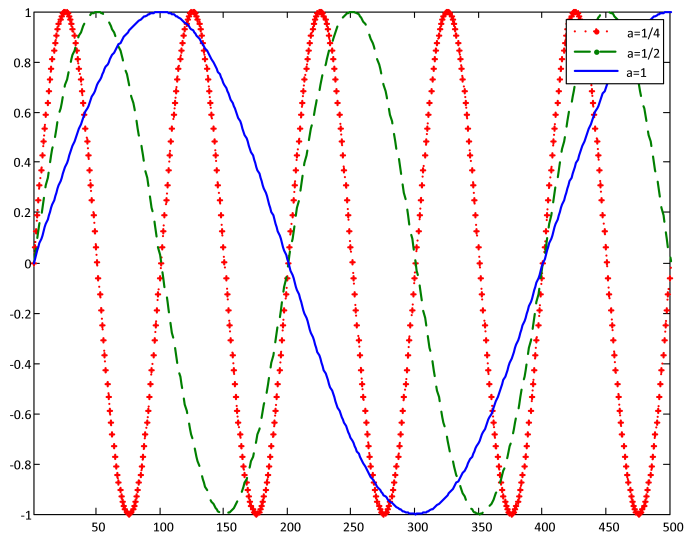


FIGURE 2.4: Exemple d'une sinusoïde a différentes échelle

Translation de l'ondelette

Le décalage d'une ondelette signifie simplement retarder (ou avancer) son apparition ; mathématiquement, ce qui retarde une fonction $f(t)$ par k est représenté par $f(t - k)$

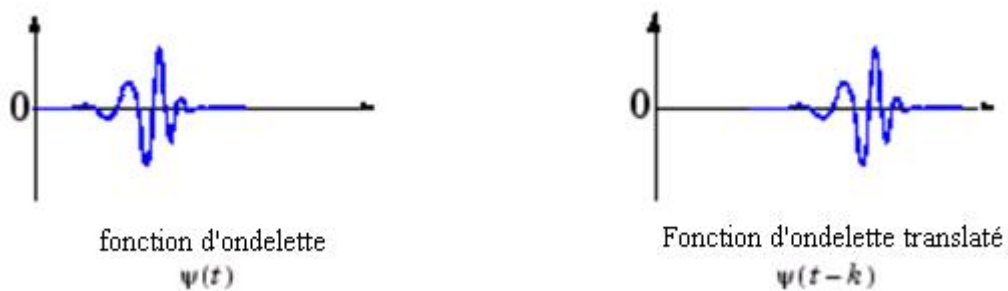


FIGURE 2.5: Translation d'une ondelette

Propriété de la résolution

Vérifions que la transformée en ondelettes a bien les propriétés que nous cherchions, à savoir une résolution temporelle croissante pour des fréquences croissantes et inversement.

Pour la dispersion temporelle, nous avons à l'échelle « a » :

$$\sigma_t(a, b = 0) = \sqrt{\int_{-\infty}^{+\infty} t^2 |\psi_{a,0}(t)|^2 . dt} = a \sqrt{\int_{-\infty}^{+\infty} t^2 |\psi(t)|^2 . dt} = a . \sigma_t \quad (2.7)$$

Pour la dispersion fréquentielle, nous avons à l'échelle « a » :

$$\sigma_f(a, b = 0) = \sqrt{\int_{-\infty}^{+\infty} f^2 |\Psi_{a,0}(f)|^2 . df} = a \sqrt{\int_{-\infty}^{+\infty} f^2 |\Psi(f)|^2 . df} = \frac{\sigma_f}{a} \quad (2.8)$$

On voit que la transformée en ondelettes permet d'avoir un pavage du plan temps-échelle (ou temps-fréquence) qui possède les propriétés recherchées : à une petite échelle (donc à une fréquence élevée) correspond une haute résolution temporelle (petite valeur de la dispersion temporelle) et inversement. L'aire de chaque pavé est constante et sa valeur peut s'approcher de $\frac{1}{4\pi}$ selon le choix de l'ondelette. [10]

La Figure 2.6 présente différentes résolutions pour les analyses en temps, fréquence, temps-fréquence, et échelle-temps.

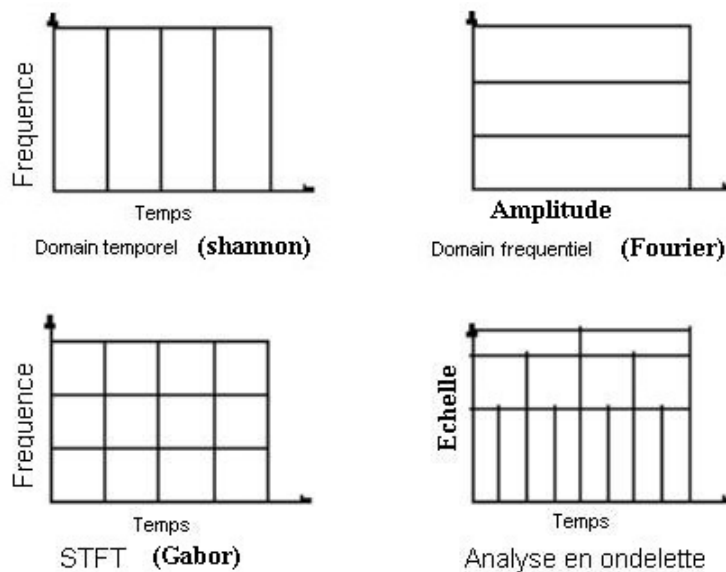


FIGURE 2.6: Différentes résolutions pour différentes analyses d'un signal

2.2.3.3 La transformée en paquets d'ondelettes (WPT)

Le transformée en paquet ondelettes (WPT) peut être considérée comme une généralisation de la transformée en ondelette classique qui offre une analyse plus riche du signal [12]; elle offre une analyse multi-résolution et une analyse temps-fréquence pour les signaux EEG non-stationnaires.

La transformée en paquet d'ondelettes génère l'arbre de décomposition complet, comme représenté dans la Figure 2.7.

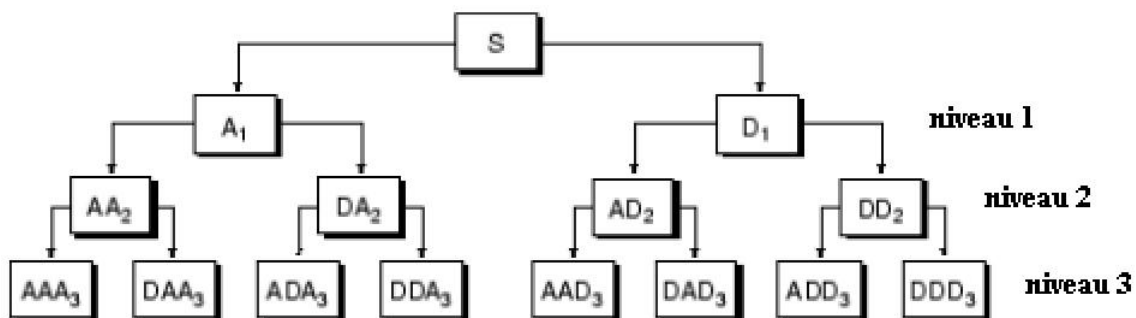


FIGURE 2.7: Structure de la décomposition en paquets d'ondelettes à 3 niveaux

Un paquet d'ondelettes est représenté comme une fonction [13] :

$$\Psi_{(j,k)}^i(t) = 2^{-j/2} \cdot \Psi(2^{-j}t - k) \quad (2.9)$$

Où i est le paramètre de modulation, j est le paramètre de dilatation et k est le paramètre de translation. $i = 1, 2, \dots, j^n$ et n est le degré de décomposition dans l'arbre de paquet d'ondelette.

L'ondelette Ψ^i est obtenue par les relations récursives suivantes :

$$\Psi^{2i} = \frac{1}{\sqrt{2}} \sum_{-\infty}^{\infty} h(k) \cdot \Psi^i\left(\frac{t}{2} - k\right) \quad (2.10)$$

$$\Psi^{2i+1} = \frac{1}{\sqrt{2}} \sum_{-\infty}^{\infty} g(k) \cdot \Psi^i\left(\frac{t}{2} - k\right) \quad (2.11)$$

Ici ψ^i est appelée *ondelette mère* et les filtres discrets $h(k)$ et $g(k)$ sont des filtres miroirs en quadrature (QMFs) associés à la fonction d'échelle et de la fonction ondelette mère [14]. Les coefficients de paquet d'ondelette $C_{j,k}^i$, correspondant au signal $f(t)$, peuvent être obtenus sous forme :

$$C_{j,k}^i = \int_{-\infty}^{\infty} f(t) \cdot \Psi_{j,k}^i(t) dt \quad (2.12)$$

Ce qui fournit les coefficients d'ondelettes satisfaisant la condition d'orthogonalité. La composante de paquet d'ondelettes du signal à un nœud particulier :

$$f_j^i(t) = \sum_{k=-\infty}^{\infty} C_{j,k}^i \Psi_{j,k}^i \quad (2.13)$$

Les coefficients d'ondelettes extraits fournissent une représentation compacte qui présente la distribution de l'énergie du signal dans le temps et fréquence[12].

2.3 Conclusion

Après avoir étudié les différentes méthodes d'analyse *temps-fréquence*, nous avons trouvé que la méthode la plus adaptée pour l'étude des signaux EEG est la transformée en paquets d'ondelettes (WPT), car elle permet le traitement des signaux non-stationnaires et une localisation précise en temps-fréquence des composantes spectrales.

Chapitre 3

Extractions des paramètres caractéristiques des signaux EEG et Analyse des données par l'ACP

3.1 Introduction

Dans ce chapitre je donnerais une description de la base de données utilisée dans ce mémoire ainsi que les paramètres caractéristiques extraits afin de réduire la dimension des données à traiter, suivie d'une analyse en composantes principales (ACP) pour enlever les informations redondantes.

3.2 La base de données

L'étude faite dans ce mémoire traite des signaux EEG disponibles dans une base de données accessible au public, voir [4], cette section va donner une petite description en se référant à [3]. L'ensemble complet des données se compose de cinq ensembles (notée A-B-C-D-E) chacun contient 100 segments mono-canal EEG de durée de 23,6 s. Ces segments ont été sélectionnés et découpés à partir des enregistrements continus et multicanal des signaux EEG. Les ensembles A et B se composent de segments pris à partir d'enregistrements EEG de surface qui ont été effectués sur cinq volontaires saints (Normaux) en plaçant les électrodes suivant le système international 10-20 (voir Figure 3.1). Les volontaires été relaxés dans un état d'éveil avec les yeux ouverts (A) et les

yeux fermés (B), respectivement. Les ensembles C, D, et E proviennent des archives du diagnostic préopératoire EEG.

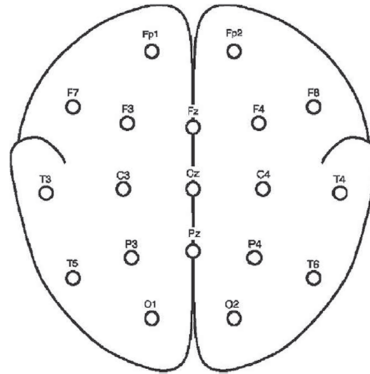


FIGURE 3.1: Schéma de la localisation des électrodes selon le système internationale 10-20 [3]

Les cinq patients avaient tous atteint la maîtrise des crises après résection complète de l'une des formations hippocampique, qui a donc été correctement diagnostiqué à la zone épileptogène. Les segments dans la série D ont été enregistrés au sein de la zone épileptogène, et ceux en C réglée de la formation hippocampique de l'hémisphère opposé du cerveau. Tandis que les ensembles C et D ne contenait que l'activité mesurée pendant des intervalles de saisie en repos, seulement l'ensemble E contient l'activité épileptique.

Tous les signaux EEG ont été enregistrés avec le système d'amplificateur 128-canal, en utilisant un moyen de référence commun. Les données ont été numérisées à 173.61 échantillons par seconde en utilisant une résolution de 12 bits. Les Paramètres filtre passe-bande étaient de 0,53 à 40 Hz (12 dB / oct.) [3].

Dans ce mémoire j'ai traité un groupe de signaux composé des Set A et Set E .

Un exemple des différents signaux constituant la base de données utilisée dans ce mémoire sont représentés dans la Figure 3.2.

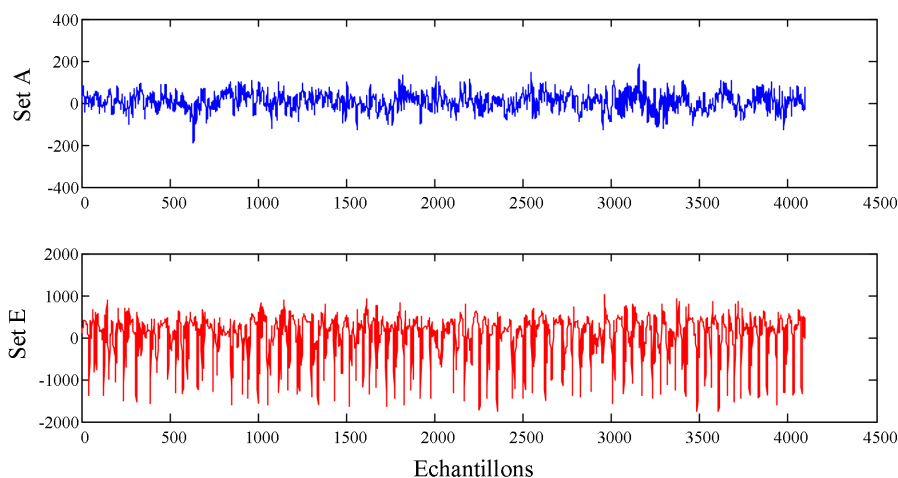


FIGURE 3.2: Exemple des signaux EEG de la base de données [4]

3.3 Les paramètres caractéristiques extraits

Les coefficients d'ondelettes extraits fournissent une représentation compacte qui présente la distribution de l'énergie du signal dans le temps et dans la fréquence. Afin de réduire d'avantage la dimension des vecteurs de caractéristiques extraits, les paramètres statistiques sur l'ensemble des coefficients d'ondelettes ont été utilisés [12].

3.3.1 Paramètres caractéristiques statistiques

Les paramètres statistiques suivants ont été utilisés pour représenter la distribution temps-fréquence des signaux EEG dans cette étude :

- Le maximum : Traduit la plus grande valeur qu'atteint le signal.
- Le minimum : Traduit la plus petite valeur qu'atteint le signal.
- La gamme : Traduit la distance entre la valeur maximale et minimale du signal.
- L'écart-type : (en anglais : standard deviation) est une notion mathématique définie en probabilités et appliquée en statistique. L'écart type est une mesure de la dispersion d'une variable aléatoire réelle [15]

$$Ect = \sigma = \sqrt{\left(\frac{1}{N-1} \cdot \sum_{i=1}^N (x_i - \bar{x})^2\right)}$$

3.3.2 Paramètres caractéristiques Non-statistiques

- L'énergie : C'est la somme quadratique des valeurs du signal ou des coefficients.

$$Eng = \sum_{i=1}^N |x_i|^2$$

- L'entropie : (entropie de Shannon), qu'on doit à *Claude Shannon*, est une fonction mathématique qui, intuitivement, correspond à la quantité d'information contenue ou délivrée par une source d'information (signal électrique). Du point de vue d'un récepteur, plus la source émet d'informations différentes, plus l'entropie (ou incertitude sur ce que la source émet) est grande [16].

$$Ent = - \sum_{i=1}^N x_i^2 \log(x_i^2)$$

Avec la convention $0 \log 0 = 0$

Où les x_i désignent les échantillons constituant le signal ou les coefficients.

3.4 L'analyse en composantes principales (ACP)

L'Analyse en Composantes Principales (ACP) est une méthode de la famille de l'analyse des données et plus généralement de la statistique multivariée, qui consiste à transformer des variables liées entre elles (dites "corrélées" en statistique) en nouvelles variables décorrélées les unes des autres qui sont une combinaison linéaire de ces variables corrélées. Ces nouvelles variables sont nommées "composantes principales", ou axes. Elle permet une meilleure représentation des données et permet de réduire l'information en un nombre de composantes plus limité que le nombre initial de variables avec une perte minimum d'informations [17].

3.4.1 Tableau des données

Les données sont les mesures effectuées sur n unités $u_1, u_2, \dots, u_i, \dots, u_n$, Les p variables quantitatives qui représentent ces mesures sont $v_1, v_2, \dots, v_j, \dots, v_p$

Le tableau des données brutes à partir duquel on va faire l'analyse est noté X et à la forme suivante :

$$X = \begin{matrix} & \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_j & \dots & \mathbf{v}_p \\ \mathbf{u}_1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ \mathbf{u}_2 & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{u}_i & x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{u}_n & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{matrix}$$

On peut représenter chaque unité par le vecteur de ses mesures sur les p variables

$${}^tU_i = [x_{i1} \ x_{i2} \ \dots \ x_{ij} \ \dots \ x_{ip}] \text{ Ce qui donne } U_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{ip} \end{bmatrix}$$

L'idée générale des méthodes factorielles est de trouver un système d'axes et de plans tels que les projections de ces nuages de points sur ces axes et ces plans permettent de reconstituer les positions des points les uns par rapport aux autres, c'est à dire avoir des images les moins déformées possibles.

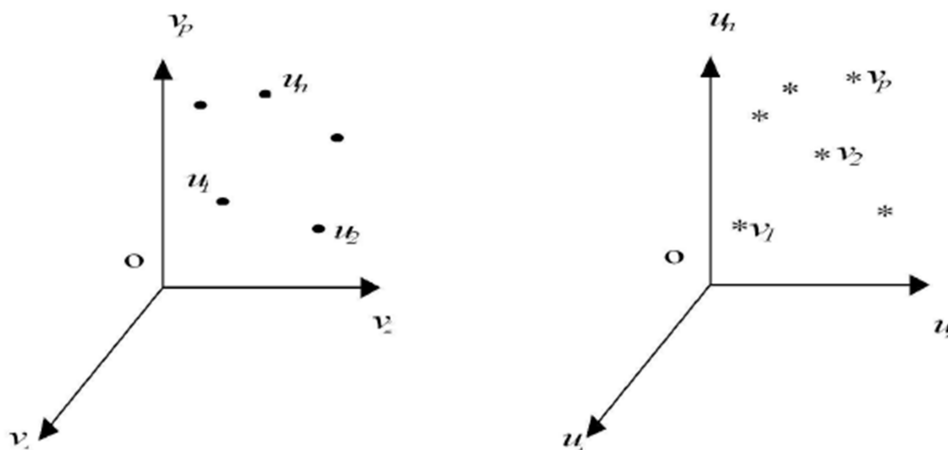


FIGURE 3.3: Représentation des nuages de points dans un système d'axes

3.4.1.1 Choix d'une distance

À propos de deux individus, on essaye d'évaluer leur ressemblance : deux individus se ressemblent d'autant plus qu'ils possèdent des valeurs proches pour l'ensemble des variables. En ACP, la distance $d(u_i, u'_i)$ entre deux individus u_i et u'_i est la distance euclidienne classique définie par :

$$d^2(u_i, u'_i) = \sum_{j=1}^p (X_{ij} - X'_{ij})^2 \quad (3.1)$$

3.4.1.2 Choix de l'origine

Le point o correspondant au vecteur de coordonnées toutes nulles n'est pas forcément une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine. Il apparaît plus judicieux de choisir une origine liée au nuage lui-même « le centre de gravité du nuage ».

Pour définir ce centre de gravité, il faut choisir un système de pondération des unités :

$\forall i = 1, \dots, n$ p_i = poids de l'unité u_i tel que $\sum_{i=1}^n p_i = 1$. Par définition le centre de gravité est défini comme le point tel que :

$$\sum_{i=1}^n p_i \vec{G}u_i = \vec{0} \quad (3.2)$$

Pour l'ACP on choisit de donner le même poids $\frac{1}{n}$ à tous les individus.

Le centre de gravité G du nuage des individus est alors le point dont les coordonnées sont les valeurs moyennes des variables :

$$G = \begin{bmatrix} \frac{1}{2} \sum_{i=1}^n x_{i1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{2} \sum_{i=1}^n x_{ij} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{2} \sum_{i=1}^n x_{ip} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}_j \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}_p \end{bmatrix}$$

Prendre G comme origine, conformément à la Figure suivante, revient alors à travailler sur le tableau des données centrées :

$$X_c = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1j} - \bar{x}_j & \dots & x_{1p} - \bar{x}_p \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ x_{i1} - \bar{x}_1 & \dots & x_{ij} - \bar{x}_j & \dots & x_{ip} - \bar{x}_p \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ x_{ni} - \bar{x}_1 & \dots & x_{nj} - \bar{x}_j & \dots & x_{np} - \bar{x}_p \end{bmatrix}$$

et le vecteur des coordonnées centrées de l'unité u_i est :

$$U_{ci} = \begin{bmatrix} x_{i1} - \bar{x}_1 \\ x_{i2} - \bar{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{ij} - \bar{x}_j \\ \cdot \\ \cdot \\ \cdot \\ x_{ip} - \bar{x}_p \end{bmatrix}$$

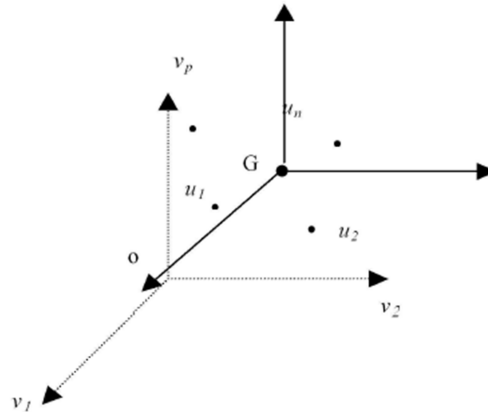


FIGURE 3.4: Le changement d'origine

3.4.1.3 Matrices de covariance et de corrélation empiriques

On appelle matrice de covariance empirique de p variables quantitatives $(v_1, v_2, \dots, v_j, \dots, v_p)$ mesurées sur un ensemble de n unités, la matrice à p lignes et p colonnes contenant sur sa diagonale principale les variances empiriques des p variables, et ailleurs, les covariances empiriques de ces variables deux à deux :

$$\Sigma = \begin{bmatrix} \text{Var}(v_1) & \text{Cov}(v_1, v_2) & \dots & \text{Cov}(v_1, v_j) & \dots & \text{Cov}(v_1, v_p) \\ \text{Cov}(v_2, v_1) & \text{Var}(v_2) & \dots & \text{Cov}(v_2, v_j) & \dots & \text{Cov}(v_2, v_p) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \text{Cov}(v_j, v_1) & \text{Cov}(v_j, v_2) & \dots & \text{Var}(v_j) & \dots & \text{Cov}(v_j, v_p) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(v_p, v_1) & \text{Cov}(v_p, v_2) & \dots & \text{Cov}(v_p, v_j) & \dots & \text{Var}(v_p) \end{bmatrix}$$

Avec :

$$\text{Var}(v_j) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

$$\text{Cov}(v_j, v_k) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

3.4.1.4 Moments d'inertie

On note I_G le moment d'inertie du nuage des individus par rapport au centre de gravité G :

$$I_G = \sum_{j=1}^p \left[\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right] = \sum_{j=1}^p \text{Var}(V_j)$$

ou $\text{Var}(v_j)$ est la variance empirique de la variable v_j . Sous cette forme, on constate que l'inertie totale est égale à la trace de la matrice de covariance Σ des p variables v_j :

$$I_G = \text{trace}(\Sigma)$$

L'inertie du nuage des individus par rapport à un axe Δ passant par G est égale par définition à :

$$I_\Delta = \frac{1}{n} \sum_{i=1}^n d^2(h_{\Delta_i}, u_i)$$

où h_{Δ_i} est la projection orthogonale de u_i sur l'axe Δ (Cette inertie mesure la proximité à²²² l'axe Δ du nuage des individus.).

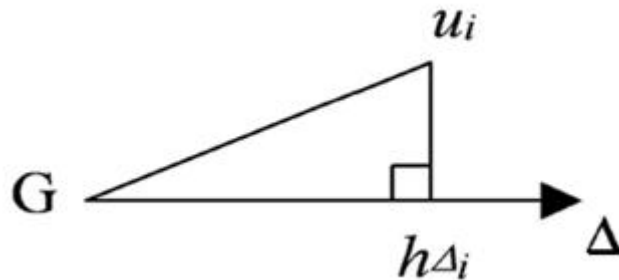


FIGURE 3.5: Projection orthogonale de u_i sur l'axe Δ

3.4.1.5 Recherche du 1er axe principale

On cherche un axe Δ_1 passant par G d'inertie I_{Δ_1} minimum car c'est l'axe le plus proche de l'ensemble des points du nuage des individus, et donc, si l'on doit projeter ce nuage sur cet axe, c'est lui qui donnera l'image la moins déformée du nuage.

I_{Δ_1} est minimum $\iff I_{\Delta_1^*}$ est maximum (tel que $I_{\Delta_1^*}$ est le complémentaire de I_{Δ_1})

On définit l'axe Δ_1 par son vecteur directeur unitaire $\overrightarrow{Ga_1}$

Il faut donc trouver $\overrightarrow{Ga_1}$ tel que $I_{\Delta_1^*}$ est maximum sous la contrainte que $\|\overrightarrow{Ga_1}\|^2=1$

Les expressions algébriques de $I_{\Delta_1^*}$ et $\|\overrightarrow{Ga_1}\|^2$ sont donnés par :

$$I_{\Delta_1^*} = {}^t a_1 \Sigma a_1$$

$$\|\overrightarrow{Ga_1}\|^2 = {}^t a_1 a_1$$

Le problème à résoudre est de trouver a_1 tel que ${}^t a_1 \Sigma a_1$ soit maximum avec la contrainte ${}^t a_1 a_1 = 1$

La méthode des multiplicateurs de Lagrange montre que ${}^t a_1 \Sigma a_1 = \lambda_1$

Tel que a_1 est le vecteur propre de la matrice de covariance Σ qui a comme valeur propre λ_1

Ce qui signifie que L'axe Δ_1 pour lequel le nuage des individus à une inertie minimum a comme vecteur directeur unitaire le premier vecteur propre associé à la plus grande valeur propre de la matrice de covariance Σ .

3.4.1.6 la Recherche des axes suivants

On recherche ensuite un deuxième axe Δ_2 orthogonal au premier et d'inertie minimum. On peut, comme dans le paragraphe précédent, définir l'axe Δ_2 passant par G par son vecteur directeur unitaire a_2 .

L'inertie du nuage des individus par rapport à son complémentaire orthogonal est égale à :

$$I_{\Delta_1^*} = {}^t a_2 \Sigma a_2 = \lambda_2$$

Et elle doit être maximum avec les deux contraintes suivantes :

$${}^t a_1 a_1 = 1 \text{ et } {}^t a_2 a_1 = 0$$

La deuxième contrainte exprime que le deuxième axe doit être orthogonal au premier et donc que le produit scalaire des deux vecteurs directeurs est nul.

On peut rechercher de nouveaux axes en suivant la même procédure.

Les nouveaux axes sont tous vecteurs propres de Σ correspondant aux valeurs propres ordonnées. Ce qui nous donne le schéma suivant :

$$\left\{ \begin{array}{l} \Delta_1 \perp \Delta_2 \perp \dots \perp \Delta_p \\ a_1 \perp a_2 \perp \dots \perp a_p \\ \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \\ I_{\Delta_1^*} \geq I_{\Delta_2^*} \geq \dots \geq I_{\Delta_p^*} \end{array} \right.$$

On passera ainsi de la base orthogonale initiale des variables centrées à la nouvelle base orthogonale des vecteurs propres de Σ . On appelle les nouveaux axes « axes principaux ».

3.4.1.7 Représentation des individus dans les nouveaux axes

Pour faire la représentation des individus dans les plans définis par les nouveaux axes, il suffit de calculer les coordonnées des individus dans les nouveaux axes. Pour obtenir y_{ik} , coordonnée de l'unité u_i sur l'axe Δ_k , on projette orthogonalement le vecteur $\overrightarrow{Gu_i}$ sur cet axe et on obtient :

$$y_{ik} = {}^t a_k u_{ci}$$

Et

$$Y_i = {}^t A u_{ci}$$

Où Y_i est le vecteur des coordonnées de l'unité u_i et A est la matrice du changement de base (A matrice des vecteurs propres orthogonaux et de norme 1 est une matrice orthogonale, son inverse est égale à sa transposée).

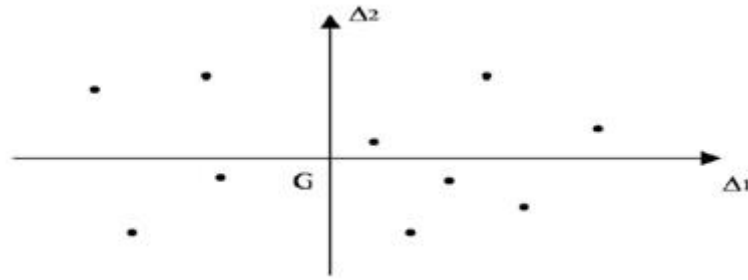


FIGURE 3.6: Représentation du nuage de points avec deux axes principaux

3.4.1.8 Différentes procédures pour le choix du nombre d'axes

- Pourcentage d'inertie souhaité : $\frac{\lambda_1}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$ Mesure la part d'inertie expliquée par l'axe Δ_1 . Ce critère (souvent exprimé en pourcentage) mesure le degré de reconstitution des carrés des distances.
- Diviser l'inertie totale par le nombre de variables initiales \Rightarrow inertie moyenne par variable : I.M. Conserver tous les axes apportant une inertie supérieure à cette valeur I.M. (inertie > 1 si variables centrées réduites).
- Histogramme : On trace un histogramme des valeurs propres associées aux vecteurs propres de la matrice de covariance Σ et on choisit un certain seuil (cassure) au dessus duquel on néglige les autres axes principaux et on ne prendra que ceux qui sont au dessous de ce seuil comme le montre la Figure 3.7 [18].

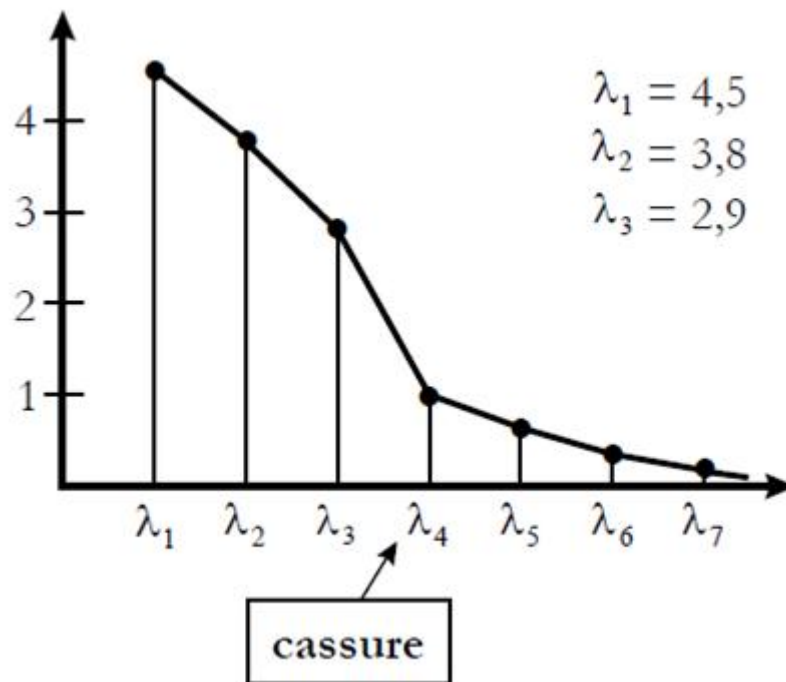


FIGURE 3.7: Histogramme des valeurs propres associées aux vecteurs propres

3.5 Conclusion

L'extraction des paramètres caractéristiques des données à pour but de réduire leur dimensions, et l'analyse en composantes principales permet de ne garder que les informations utiles, ce qui rend leur manipulation plus flexible, afin de gagner en temps et en espace mémoire lors du développement de système de classification des signaux EEG.

Chapitre 4

Méthode de classification : Machine à vecteurs de support (SVM)

4.1 Introduction

Dans le présent chapitre nous donnerons un aperçu sur les méthodes de classification utilisées dans le domaine du traitement des signaux EEG, et nous développerons la méthode choisie qui est la classification par (SVM), nous parlerons de l'algorithme d'élimination récursive des paramètres caractéristiques (RFE) et les critères de performances sur lesquels nous nous sommes basé dans notre chapitre final pour évaluer notre classificateur et préciser les paramètres à choisir pour une bonne classification.

4.2 Différentes Méthodes de classifications

Il existe une multitude de méthodes de classification qui sont utilisées pour classer les signaux EEG parmi ces méthodes on cite :

- ANN : Artificial Neural Network [19]
- ANFIS : adaptive neuro-fuzzy inference system [20]
- kNN : k nearest neighbors classification algorithm [21]
- RBFNN : radial basis function neural network[22]
- RNN : recurrent neural network[23]
- SVM : Machine à vecteurs de support[24]

Dans notre travail on cherche à identifier les caractéristiques du signal EEG qui permettent de réaliser la meilleure classification possible (EEG normal ou pathologique) du signal EEG. . Pour cela, j'ai choisi la méthode SVM qui s'adapte le mieux à notre besoin car elle permet d'avoir une bonne précision lors de la classification en utilisant des données de grande dimension en prenant une partie pour l'apprentissage et une autre pour le test ; en combinant cette méthode à l'algorithme RFE on peut classer les paramètres caractéristiques selon leurs performances.

4.3 Introduction au Machine à vecteurs de support (SVM)

La classification consiste à classer des individus en fonction de certaines de leurs caractéristiques. L'idée de la classification supervisée est d'apprendre une règle de classification à partir d'un ensemble de données dont le classement est déjà connu. Une fois la règle apprise, il est possible de l'appliquer pour catégoriser de nouvelles données, dont le classement est inconnu. Les machines à vecteurs de support, ou SVM (Support Vector Machines), sont une technique relativement récente de classification supervisée qui suscite beaucoup d'intérêt pour ses bonnes performances dans un large éventail d'applications pratiques (elles ont été introduites en 1992 par *Vladimir Vapnik, Bernhard Boser et Isabelle Guyon*). Ce chapitre constitue une introduction aux machines à vecteurs de support. Il explique leur fonctionnement d'abord dans le cas simple où les données sont linéairement séparables, puis dans le cas général où elles ne le sont pas nécessairement (non linéaire).

4.4 Les Machines à vecteurs de support

Les machines à vecteurs de support (SVM) sont un algorithme dont le but est de résoudre les problèmes de discrimination à deux classes. On appelle problème de discrimination à deux classes un problème dans lequel on tente de déterminer la classe à laquelle appartient un individu.

Pour ce faire, on utilise les caractéristiques connues de cet individu. Ces n caractéristiques sont représentées par un vecteur $x \in R^n$. La classe à laquelle appartient l'individu

est représentée par $y \in \{-1, 1\}$, où une des classes possible est représentée par -1 et l'autre par 1. Par conséquent, avec cette notation, le problème est de déterminer la valeur de y en se servant de x . Pour y parvenir, les machines à vecteurs de support utilisent un ensemble de données pour lesquelles le classement est déjà connu et s'en servent pour construire une règle qui permet d'effectuer une bonne classification. Cet ensemble de données est appelé l'ensemble d'apprentissage. La règle trouvée avec l'ensemble d'apprentissage doit être la plus générale possible, puisqu'il faut aussi qu'elle soit bonne pour de nouvelles données qui n'étaient pas dans l'ensemble d'apprentissage.

4.5 Hyperplan séparable

Supposons que nous disposons d'un ensemble d'apprentissage de l données de la forme $(x_i, y_i) \in R^n \times \{-1, 1\}$ ($i = 1, 2, \dots, l$) dont nous voulons nous servir pour déterminer une règle permettant de classer les données. Supposons aussi que ces données sont linéairement séparables, c'est-à-dire qu'il existe un hyperplan dans R^n tel que toutes les données appartenant à la classe 1 se retrouvent d'un côté de l'hyperplan alors que celles de la classe -1 se situent de l'autre côté.

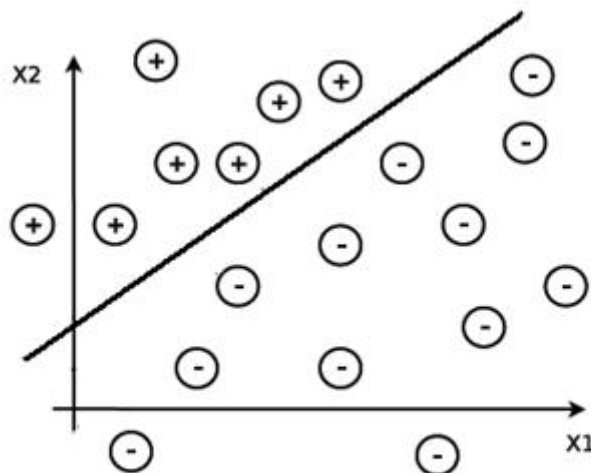


FIGURE 4.1: Des données linéairement séparables

Plus formellement, les données sont dites linéairement séparables s'il existe un hyperplan $w.x + b = 0$ tel que

- $w.x + b > 0$ pour tout x appartenant à la classe 1,
- $w.x + b < 0$ pour tout x appartenant à la classe -1 ,

avec $w = (w_1; \dots; w_n) \in R^n$ le vecteur des coefficients de l'hyperplan et $b \in R$ un scalaire appelé le biais (remarquons que tout hyperplan peut s'écrire sous cette forme). Nous dirons d'un tel hyperplan qu'il sépare et classe les données sous l'hypothèse que les données sont linéairement séparables car il suffit d'utiliser la fonction suivante pour effectuer la classification :

$$\text{Classe}(x) = \text{signe}(w.x + b) \tag{4.1}$$

où

$$\begin{aligned} \text{signe}(w.x + b) &= -1 \quad \text{si } w.x + b < 0 \\ &= 0 \quad \text{si } w.x + b = 0 \\ &= 1 \quad \text{si } w.x + b > 0 \end{aligned}$$

Notons que si une donnée est directement sur l'hyperplan, elle sera assignée à la classe 0. Dans ce cas, il est possible de la laisser inclassée, d'utiliser une autre règle ou de l'assigner aléatoirement à l'une des deux classes.

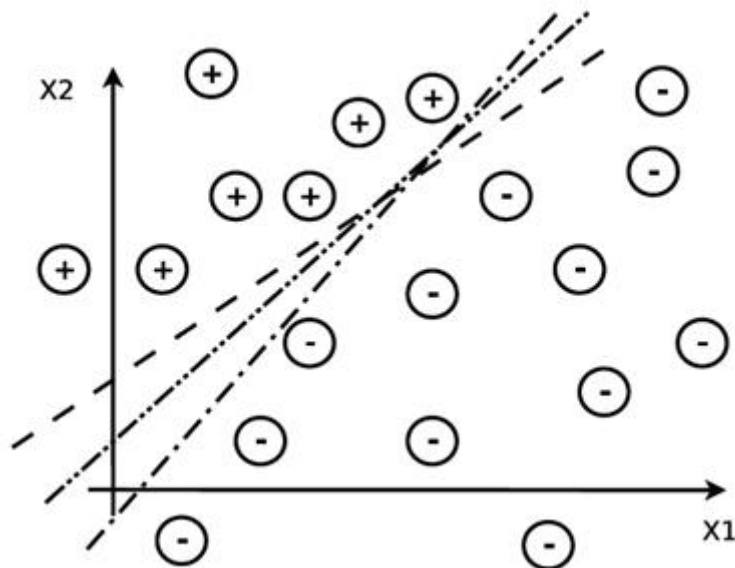


FIGURE 4.2: Hyperplan : Il existe une infinité d'hyperplans pouvant séparer les données linéairement séparables.

Si les données sont linéairement séparables, il existe une infinité d'hyperplans qui peuvent servir de séparateurs. L'idée des SVM est de choisir le meilleur hyperplan, c'est-à-dire celui qui donnera la règle qui se généralisera le mieux à d'autres données que celles de l'ensemble d'apprentissage. Afin de déterminer ce qui caractérise le meilleur hyperplan, introduisons le concept de marge.

4.6 Marge et hyperplan canonique

La marge d'un hyperplan est la distance entre l'hyperplan et la donnée la plus proche. Si $dist(x, w, b)$ représente la distance euclidienne entre le point x et l'hyperplan $w \cdot x + b = 0$, alors la marge M est définie ainsi :

$$M = \min\{dist(x_i, w, b) : i = 1; \dots; l\} \quad (4.2)$$

où les x_i sont les données de l'ensemble d'apprentissage.

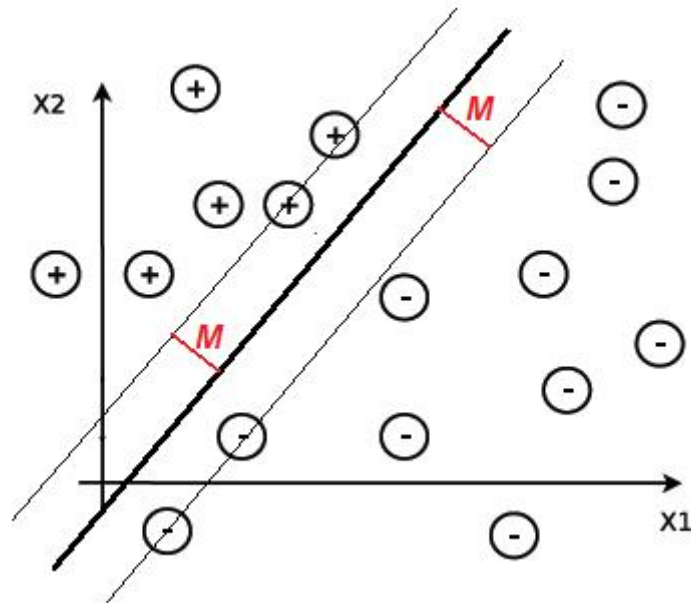


FIGURE 4.3: La marge

D'après un résultat de la théorie de l'apprentissage statistique, l'hyperplan qui aura la meilleure généralisation est celui qui possède la plus grande marge. Toutefois il existe une infinité de manières différentes d'écrire le même hyperplan avec une marge maximale.

En effet, supposons que l'hyperplan $w \cdot x + b = 0$ soit un hyperplan dont la marge est maximale, et soit $\lambda \in \mathbb{R}^+ \setminus \{0\}$. Alors, l'hyperplan correspond aussi à l'hyperplan dont la marge est maximale, mais possède un vecteur des coefficients et un biais différents (si $\lambda \neq 1$). Afin d'éviter ce problème nous introduisons le concept d'hyperplan canonique.

Un hyperplan $w \cdot x + b = 0$ est dit canonique si $\min\{|w \cdot x_i + b| : i = 1; \dots; l\} = 1$.

où les x_i sont les données d'apprentissage. On peut montrer que ce minimum correspond aux données qui sont directement sur la marge. On peut aussi montrer que tout hyperplan qui sépare les données peut s'écrire sous forme canonique et qu'il n'existe qu'une seule façon d'écrire un hyperplan pour qu'il soit canonique.

4.6.1 Trouver l'hyperplan

On peut montrer que pour un hyperplan canonique $w \cdot x + b = 0$ la marge M est donnée par l'expression

$$M = \frac{1}{\|w\|} \quad (4.3)$$

Où $\|w\| = \sqrt{(w_1^2 + \dots + w_n^2)}$ On voit donc que plus $\|w\|$ est petite, plus la marge de l'hyperplan canonique correspondant est grande. La recherche du meilleur hyperplan peut donc s'écrire sous la forme du problème d'optimisation suivant :

$$\text{minimiser } \|w\| \text{ avec } y_i(w \cdot x_i + b) \geq 1, i = 1; \dots; l. \quad (4.4)$$

Les contraintes $y_i(w \cdot x_i + b) \geq 1, i = 1; \dots; l$ assurent d'une part que l'hyperplan sépare les données correctement, et d'autre part qu'il est canonique. Cependant, il est possible de formuler un problème équivalent, mais avec une fonction objectif plus simple. En effet, comme

$$\|w\| = \sqrt{(w \cdot w)}$$

minimiser $\|w\|$ est équivalent à minimiser $w \cdot w$ évidemment, minimiser $w \cdot w$ est équivalent à minimiser $\frac{1}{2}w \cdot w$ (cette petite modification permet d'éviter d'avoir une constante dans la représentation duale du problème, comme nous le verrons un peu plus loin). Par conséquent, afin de trouver l'hyperplan canonique qui sépare les données avec la plus grande marge possible, il suffit de résoudre le problème d'optimisation suivant :

$$\text{minimiser } \frac{1}{2}w \cdot w \text{ sujet à } y_i(w \cdot x_i + b) \geq 1; i = 1; \dots; l.$$

Une propriété très intéressante de ce problème est que $f(w) = w \cdot w = w_1^2 + \dots + w_n^2$ est une fonction strictement convexe. Ceci assure qu'il n'y a pas de minimum relatif et qu'il n'existe qu'une unique solution optimale.

4.6.2 Représentation duale

Il serait possible de résoudre le problème d'optimisation ci-dessus directement. Toutefois, sa représentation duale possède des propriétés très intéressantes qui auront des répercussions majeures lorsque nous considérerons les machines à vecteurs de supports pour le cas où les données ne sont pas linéairement séparables. Commençons tout d'abord par écrire le Lagrangien. Pour ce faire, il est nécessaire de réécrire les contraintes ainsi :

$$-(y_i(w \cdot x_i + b) - 1) \leq 0; i = 1; \dots; l.$$

La formulation de Lagrange permet de trouver les extrémums en annulant les dérivées partielles de la fonction $L(w, b, \alpha) = \frac{1}{2}w.w - \sum_{i=1}^l \alpha_i(y_i(w.x_i + b) - 1)$. Le lagrangien L doit être minimisé par rapport à w et b et maximisé par rapport à α . On résout ce nouveau problème en calculant les dérivées partielles :

$$\frac{\partial L(w,b,\alpha)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0,$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = -\sum_{i=1}^l \alpha_i y_i = 0,$$

ce qu'il est possible de réécrire de cette manière :

$$w = \sum_{i=1}^l \alpha_i y_i x_i \text{ et } \sum_{i=1}^l \alpha_i y_i = 0$$

Utilisons ces équations pour réécrire le Lagrangien minimal uniquement en fonction des variables duales :

$$\min_w L(w, b, \alpha) = \frac{1}{2}w.w - \sum_{i=1}^l \alpha_i(y_i(w.x_i + b) - 1) = \frac{1}{2}\sum_{i=1}^l \alpha_i y_i \sum_{i=1}^l \alpha_j y_j x_i . x_j - \sum_{i=1}^l \alpha_i y_i \sum_{i=1}^l \alpha_j y_j x_i . x_j - \sum_{i=1}^l \alpha_i y_i b + \sum_{i=1}^l \alpha_i = \sum_{i=1}^l \alpha_i - \frac{1}{2}\sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j x_i . x_j) \quad [5]$$

Ainsi, nous avons le problème dual suivant :

$$\text{maximiser } \sum_{i=1}^l \alpha_i - \frac{1}{2}\sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j x_i . x_j)$$

Avec $\sum_{i=1}^l \alpha_i y_i = 0$ et $\alpha_i \geq 0 \quad i = 1, \dots, l$.

On remarque qu'il est nécessaire d'ajouter la contrainte $\sum_{i=1}^l \alpha_i y_i = 0$ pour s'assurer que $\sum_{i=1}^l \alpha_i - \frac{1}{2}\sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j x_i . x_j)$ Correspond bien au minimum du Lagrangien. En effet, alors que la contrainte $w = \sum_{i=1}^l \alpha_i y_i x_i$ est automatiquement satisfaite par la construction de la fonction, rien n'assure que $\sum_{i=1}^l \alpha_i y_i = 0$ est elle aussi respectée.

La solution de ce problème d'optimisation sera bien sûr un vecteur $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$, alors que c'est l'équation d'un hyperplan qu'il faut pour classer les données à l'aide de la fonction indicatrice

$$\text{Classe}(x) = \text{signe}(w.x + b).$$

Il est toutefois possible de réécrire la fonction indicatrice ainsi, puisque $w = \sum_{i=1}^l \alpha_i y_i x_i$

$$\text{Classe}(x) = \text{signe}\sum_{i=1}^l (\alpha_i y_i x_i . x) + b).$$

Tout ce qu'il manque pour pouvoir utiliser cette fonction est la valeur de b . Or, comme l'hyperplan est canonique, et d'après les contraintes du problème primal, si une donnée x_m se trouve sur la marge, alors on sait que $y_m(w.x_m + b) = 1$. Donc,

$$b = \frac{1}{y_m} - w \cdot x_m = y_m - \sum_{i=1}^l \alpha_i y_i x_i \cdot x_m \text{ puisque } y_m \in \{-1, 1\}$$

Ainsi, la résolution du problème dual permet de construire l'hyperplan canonique séparant les données avec la plus grande marge et de l'utiliser pour classer des données, tout comme la résolution du problème primal.

4.7 Vecteur de support

Comme la fonction $w \cdot w$ est une fonction convexe continue et dérivable, que les contraintes sont des fonctions affines $(- (y_i(w \cdot x_i + b) - 1) \leq 0)$ et que le domaine du problème est R^n , la solution optimale trouvée respecte nécessairement les conditions de *Karush-Kuhn-Tucker*. En particulier, elle respecte la condition complémentaire de *Karush-Kuhn-Tucker*, c'est-à-dire que

$$\alpha_i^* (y_i(w^* \cdot x_i + b^*) - 1) = 0 \quad i = 1, \dots, l.$$

où α^* représente la solution optimale du problème dual et (w^*, b^*) représente celle du problème primal.

Cette condition implique que si $y_i(w^* \cdot x_i + b^*) - 1 \neq 0$ alors $\alpha_i = 0$. Par conséquent, les seuls cas où α_i peut ne pas être nul sont ceux où $y_i(w^* \cdot x_i + b^*) - 1 = 0$, c'est-à-dire ceux où

$$y_i(w^* \cdot x_i + b^*) = 1.$$

Or, les seuls points où $y_i(w^* \cdot x_i + b^*) = 1$ sont ceux qui sont sur la marge. Par conséquent, seuls les points sur la marge peuvent avoir des α_i non nuls. Ces points sont appelés les vecteurs de support.

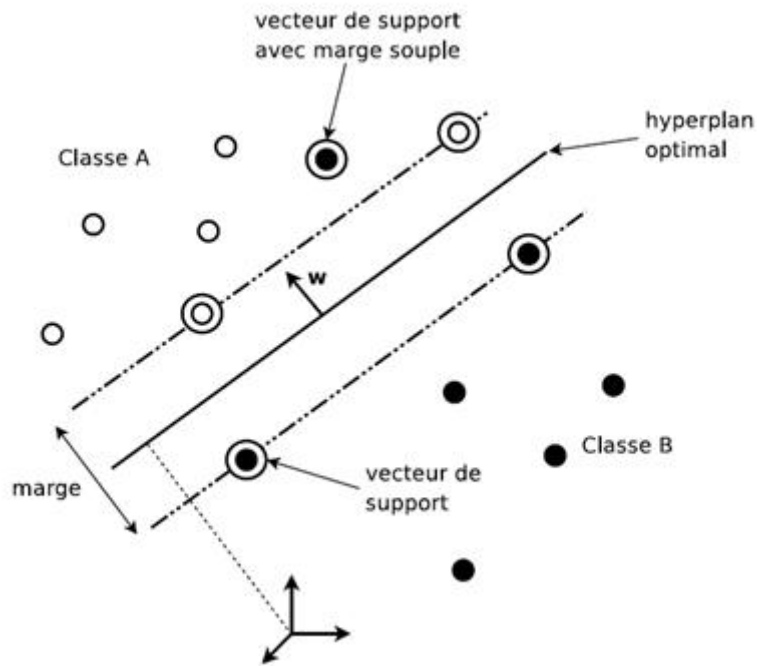


FIGURE 4.4: Les vecteurs de support[5]

La raison de ce nom est que ce sont les seuls points utiles pour déterminer l'hyperplan. En effet, rappelons que le vecteur des coefficients de l'hyperplan est donné par

$$w = \sum_{i=1}^l \alpha_i y_i x_i.$$

Ainsi, tout point qui n'est pas sur la marge n'apporte aucune contribution, puisque α_i est alors nul. Si tout les points sauf les vecteurs de support étaient retirés de l'ensemble d'apprentissage, on retrouverait le même hyperplan. Les vecteurs de support peuvent donc être vu comme les points contenant toute l'information essentielle du problème.

4.8 Machines à vecteurs de support pour données non linéairement séparables

4.8.1 Astuce du noyau

Le cas linéairement séparable vu jusqu'à présent est peu intéressant, car les problèmes de classification sont souvent non linéaires. Pour résoudre ce point la méthode classique est de projeter les données dans un espace de dimension supérieur appelé espace de redescription. L'idée étant qu'en augmentant la dimensionnalité du problème on se retrouve dans le cas linéaire vu précédemment. Nous allons donc appliquer une transformation non linéaire $\phi(\bullet)$ aux vecteurs d'entrée x_i tel que, $x_i \in R$ et $\phi(x_i) \in R$. Ce changement va conduire à passer d'un produit scalaire dans l'espace d'origine $x_i \cdot x_j$ à un produit scalaire $\phi(x_i) \cdot \phi(x_j)$ dans l'espace de redescription (voir la figure 4.5). L'astuce est d'utiliser une fonction noyau notée K qui évite le calcul explicite du produit scalaire dans l'espace de redescription. Les fonctions noyaux doivent satisfaire le théorème de Mercer (une fonction est un noyau si et seulement si elle est symétrique et semi-définie positive). Nous avons alors l'égalité suivante[5] :

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j).$$

Il existe de nombreuses fonctions noyau prédéfinies dans le Tableau 4.1 :

TABLE 4.1: Les noyaux les plus fréquemment utilisés

Nom	Noyau
Linéaire	$K(x_i, x_j) = (x_i, x_j)$
Polynomial de degré d	$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Gaussien (RBF)	$K(x_i, x_j) = e^{-\frac{(x_i - x_j) \cdot (x_i - x_j)}{2\sigma^2}}$
Multiquadratique inverse	$K(x_i, x_j) = \frac{1}{\sqrt{(x_i - x_j) \cdot (x_i - x_j) + \beta}}$

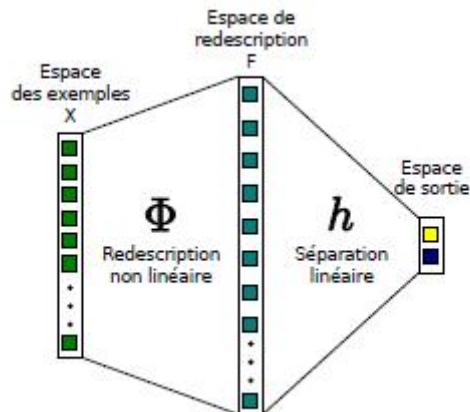


FIGURE 4.5: La transformation des données linéaire

Grâce au concept de noyau, il est possible de réécrire le problème dual de cette manière :

$$\text{maximiser } \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j K(x_i, x_j))$$

$$\text{Avec } \sum_{i=1}^l \alpha_i y_i = 0 \text{ et } \alpha_i \geq 0 \text{ } i = 1, \dots, l.$$

La fonction indicatrice peut elle aussi être réécrite :

$$\text{Classe}(x) = \text{signe}(\sum_{i=1}^l (\alpha_i y_i K(x_i, x_j)) + b),$$

où $b = y_m - \sum_{i=1}^l \alpha_i y_i K(x_i, x_j)$ avec x_m un vecteur de support.

On remarque que de cette manière, lorsque la fonction noyau est connue, la transformation $\phi(x)$ n'apparaît nulle part, ni dans le problème, ni dans l'application de la solution. Par conséquent, grâce à la fonction noyau, il n'est pas nécessaire d'effectuer la transformation sur les données. Cette fonction permet donc de faire tous les calculs nécessaires sans avoir à se préoccuper de la dimension de l'espace de redescription.

4.9 Marges souples

4.9.1 Machines à vecteurs de support et bruit

En pratique, les données sont rarement parfaites. Il y a souvent du « bruit », c'est-à-dire des données qui sont mal classées par un modèle qui est toutefois excellent en

général. Il s'agit donc d'erreurs qui sont inévitables, même pour les meilleurs modèles. Toutefois, les machines à vecteurs de support ne permettent pas de tenir compte de ce phénomène, puisque dans les contraintes, toutes les données doivent être correctement classées. Supposons par exemple qu'un ensemble de données serait très bien séparé par un hyperplan, mais qu'il n'est pas linéairement séparable dû à la présence d'un certain bruit dans les données. Dans un tel cas, il serait impossible de construire un SVM linéaire, car il est impossible que toutes les contraintes soient respectées. Afin de contourner ce problème, il peut être tentant d'utiliser un noyau quelconque afin de rendre les données linéairement séparables. Ceci est en effet toujours possible en utilisant un noyau polynomial avec un degré suffisamment élevé. Toutefois, bien que les données de l'ensemble d'apprentissage seront parfaitement séparées, la règle trouvée risque de très mal se généraliser, puisqu'elle va tenir compte de toutes les petites variations et ainsi généraliser des phénomènes qui sont en réalité bien spécifiques à l'ensemble de données actuel.

4.9.2 Marge souple

Un meilleur moyen serait de permettre à quelques données d'être à l'intérieur de la marge ou du mauvais côté de l'hyperplan. Il s'agit du concept de marge souple (soft margin). Une première idée serait de tenter de maximiser la marge tout en minimisant le nombre de données mal classées. Toutefois, le nombre de données mal classées peut être trompeur, puisqu'il ne permet pas de déterminer si une donnée était presque correctement classée ou si elle était en réalité très loin de l'hyperplan. Une meilleure idée est d'attribuer à chaque donnée x_i une valeur ξ_i qui représente à quel point la donnée est éloignée d'un bon classement, puis de tenter de minimiser la somme des ξ_i . Plus formellement, au lieu d'imposer

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1; \dots; l.$$

ce qui oblige les données à être bien classées, les contraintes seront plutôt

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1; \dots; l \text{ avec } \xi_i \geq 0$$

Par conséquent, il est possible pour une donnée d'être du mauvais côté de la marge, si, ξ_i est non nul. On dira d'une donnée qu'elle est du mauvais côté de la marge si elle est mal classée ou si sa distance par rapport à l'hyperplan séparateur est plus petite que la marge (remarquons que les points pour lesquels $\xi_i \neq 0$ ne sont pas considérés dans

le calcul de la marge). L'objectif est ainsi de maximiser la marge tout en minimisant la somme des ξ_i . Le problème d'optimisation devient alors

$$\text{minimiser } \frac{1}{2}w \cdot w + C \sum_{i=1}^l \xi_i \text{ avec } y_i(w \cdot x_i + b) \geq 1 - \xi_i \text{ et } \xi_i \geq 0 \text{ } i = 1; \dots; l.$$

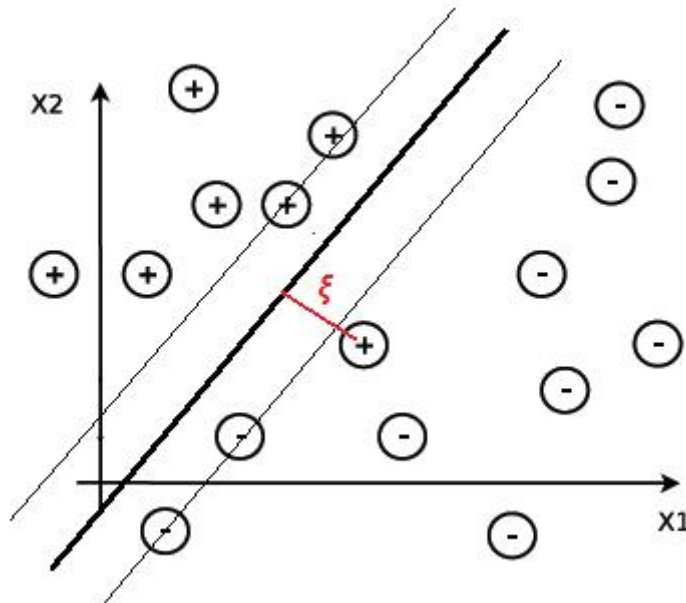


FIGURE 4.6: La marge souple

où $C > 0$ est une constante qui représente la pénalité d'avoir des données mal classées. Lorsque C est très élevé, il y aura très peu de données mal classées, alors qu'il y en aura plus pour une valeur plus faible de cette constante. Le choix de C a une grande influence sur le modèle. En pratique, plusieurs modèles sont souvent construits, avec différentes valeurs de C , puis le meilleur est choisi.

4.9.3 Représentation duale

Il est possible de construire le dual de ce problème de la même manière que précédemment. Le Lagrangien est

$$L(w, b, \xi, \alpha) = \frac{1}{2}w \cdot w + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i (y_i (w \cdot x_i + b) - 1 + \xi_i)$$

Afin de trouver le Lagrangien minimal pour un α donné, il faut le dériver par rapport aux variables primales. On obtient alors

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0,$$

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial b} = -\sum_{i=1}^l \alpha_i y_i = 0,$$

$$\frac{\partial L(w,b,\xi,\alpha)}{\partial \xi_i} = C - \alpha_i = 0.$$

De ceci, on obtient

$$w = \sum_{i=1}^l \alpha_i y_i x_i, \quad \sum_{i=1}^l \alpha_i y_i = 0 \text{ et } C = \alpha_i \text{ pour } i \in \{1, \dots, l\}$$

Utilisons ces expressions pour réécrire le Lagrangien uniquement en fonction des variables duales :

$$L(w, b, \xi, \alpha) = \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i (y_i (w \cdot x_i + b) - 1 + \xi_i) = \\ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j x_i \cdot x_j)$$

Il s'agit de la fonction objectif du problème dual pour la marge souple. On remarque que cette fonction est exactement la même que celle obtenue précédemment. La différence se situe au niveau des contraintes. En effet, rappelons que dans le problème dual, les multiplicateurs de Lagrange qui sont associés à des contraintes d'inégalités doivent être supérieurs ou égaux à zéro. Par conséquent, $\alpha_i \geq 0$. Toutefois $C - \alpha_i \geq 0$. Ceci implique que $\alpha_i \leq C$. Le problème dual est ainsi

$$\text{maximiser } \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i \alpha_j y_i y_j x_i \cdot x_j) \text{ avec } \sum_{i=1}^l \alpha_i y_i = 0 \text{ et } 0 \leq \alpha_i \leq C \text{ } i = \\ 1, \dots, l.$$

4.10 Algorithme d'élimination récursive des paramètres (RFE)

L'avantage de cette méthode réside dans le fait qu'elle fasse apparaître la meilleure précision lors de la classification en utilisant un paramètre caractéristique ou une combinaison de plusieurs paramètres et qui seront utiles par la suite pour le choix de la

meilleure méthode à utiliser pour la classification des signaux EEG.

Dans le procédé SVM-RFE qui a été développé par *Guyon, Weston, Barnhill, and Vapnick* (2002) et utilisé pour la classification de malades atteints du cancer par sélection de gènes, l'effet de supprimer une caractéristique et observer la précision de la classification est utilisé en tant que critère de classement pour les paramètres caractéristiques. Les étapes de l'algorithme SVM-RFE sont les suivants :

- (1) Supprimer un paramètre caractéristique hors du nombre de paramètres N_F (initialement $N_F = 6$) et calculer la précision de la classification.
- (2) Remettre ce paramètre et supprimer un autre paramètre et refaire l'opération (1).
- (3) Classer les précisions par ordre croissant et supprimer le paramètre caractéristique correspondant à la plus grande précision.
- (4) Refaire les opérations (1), (2) et (3) jusqu'à avoir un seul paramètre caractéristique.
- (5) Faire un classement final tel que le paramètre éliminé en premier est classé en dernier, le paramètre supprimé en deuxième est classé en avant dernier et ainsi de suite.

A la fin de cet algorithme nous aurons un classement des meilleurs paramètres, aussi les groupes de paramètres qui mènent à un résultat performant.

4.11 Validation croisée

La validation croisée est une technique qui permet de tester un modèle d'apprentissage. La validation croisée se décline en plusieurs sous-méthodes. La plus répandue est la méthode « m-fold » avec typiquement $m \in [4 : 10]$. Si l'on a une base d'apprentissage A_p contenant p éléments : $A_p = \{x_1; \dots; x_p\}$ la validation croisée consiste à appliquer les cinq étapes suivantes :

1. Découper l'ensemble des exemples en m sous-ensembles disjoints de taille $\frac{p}{m}$.
2. Apprendre sur les $m - 1$ sous-ensembles.
3. Calculer l'erreur sur la m^{ieme} partie.
4. Répéter le processus p fois.
5. Obtenir l'erreur finale en calculant la moyenne des m erreurs précédentes.

La validation croisée est simple à mettre en œuvre et utilise toutes les données. Elle permet d'obtenir une estimation de l'erreur de généralisation. Cela permet d'éviter le

sur-apprentissage[25].

4.12 Les Paramètres de performances

Afin d'évaluer les performances de notre classificateur on s'est intéressé à un ensemble de paramètres nécessaires qui sont les suivants :

- Précision : rapport des vrais positifs (TP) et vrais négatifs (TN). Elle mesure le taux de classification correcte des cas Normaux et Épileptiques.

Elle est donnée par la formule suivante :

$$Précision = \frac{TP+TN}{TP+FN+FP+TN} \times 100\%$$

- Sensibilité : rapport des vrais positifs (TP), elle mesure le taux de classification correcte des cas Normaux. Elle est donnée par la formule suivante :

$$Sensibilité = \frac{TP}{TP+FN} \times 100\%$$

- Spécificité : rapport des vrais négatifs (TN), elle mesure le taux de classification correcte des cas Épileptiques. Elle est donnée par la formule suivante :

$$Spécificité = \frac{TN}{TN+FP} \times 100\%$$

- Taux d'erreur : mesure l'erreur commise lors de la classification des cas Normaux et Épileptiques. Il est donné par la formule suivante :

$$Taux\ d'erreur = \left(1 - \frac{TP+TN}{TP+FN+FP+TN}\right) \times 100\%$$

- TP (true positive) : vrai positif signifie que le cas normal est classé dans la classe des Normaux.
- TN(true negative) : vrai négatif signifie que le cas Épileptique est classé dans la classe des Épileptiques.
- FP (false positive) : faux positif signifie que le cas Épileptique est classé dans la classe des Normaux.
- FN (false negative) : faux négatif signifie que le cas Normal est classé dans la classe des Épileptiques.

TABLE 4.2: Table de diagnostique (confusion)

		classe réelle	
		classe +	classe -
classe prédite	classe +	TP(vrai positifs)	FP(faux positifs)
	classe -	FN(faux négatifs)	TN(vrais négatifs)

4.13 Conclusion

Dans ce chapitre, nous avons présenté de manière simple et complète le concept du système d'apprentissage et classification introduit par *Vladimir Vapnik*, les « Support Vector Machine » avec une vision purement mathématique des SVM. Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données. Nous avons exposé le cas des données linéairement séparables et le cas des données non linéairement séparables qui nécessitent l'utilisation de la fonction noyau (kernel) pour changer d'espace. Cette méthode est applicable pour des tâches de classification à deux classes, l'avantage des SVM est la possibilité de les utiliser pour des données de grandes taille et aussi dans des dimensions supérieures à 2 ; on appliquera cette propriété dans notre dernier chapitre qu'on associera à l'algorithme RFE afin d'obtenir un classificateur complet et de bonne précision.

Chapitre 5

Implémentation et évaluation des performances

5.1 Introduction

Nous avons vu dans les chapitres précédents des notions indispensables pour réaliser notre classificateur ; dans ce présent chapitre nous allons mettre en évidence et en pratique ce qui a été cité précédemment à savoir : la transformée en paquets d'ondelettes (WPT) pour avoir les différents coefficients constituant nos signaux (Set A, Set E). Ensuite l'extraction des différents paramètres caractéristiques de ces coefficients (Max, Min, Gam, Ect, Ent, Eng), sur lesquels j'appliquerai une réduction de données par l'ACP pour éviter la redondances et les informations inutiles, une fois cette opération réalisée j'utiliserai ces données pour faire une classification par la méthode SVM-RFE sur le groupe1 des signaux EEG composé de Set A et Set E, nous discuterons par la suite les performances de classificateur en se basant sur trois paramètres qui sont : la précision, la sensibilité et la spécificité.

Pour cela j'ai exploité la plateforme Matlab (R2012a) avec ces différentes fonctions incluant d'autres fonctions disponibles dans la Wavelet Toolbox, Bioinformatics Toolbox, et un logiciel XLSTAT (2009) qui est un module complémentaire sous MS Excel pour l'analyse des données. Au final j'afficherai les différentes simulations effectuées .

5.2 Système automatique de classification de signaux EEG

Pour le développement d'un système automatique pour la classification des signaux EEG, nous proposons la structure montrée dans la Figure 5.1 :

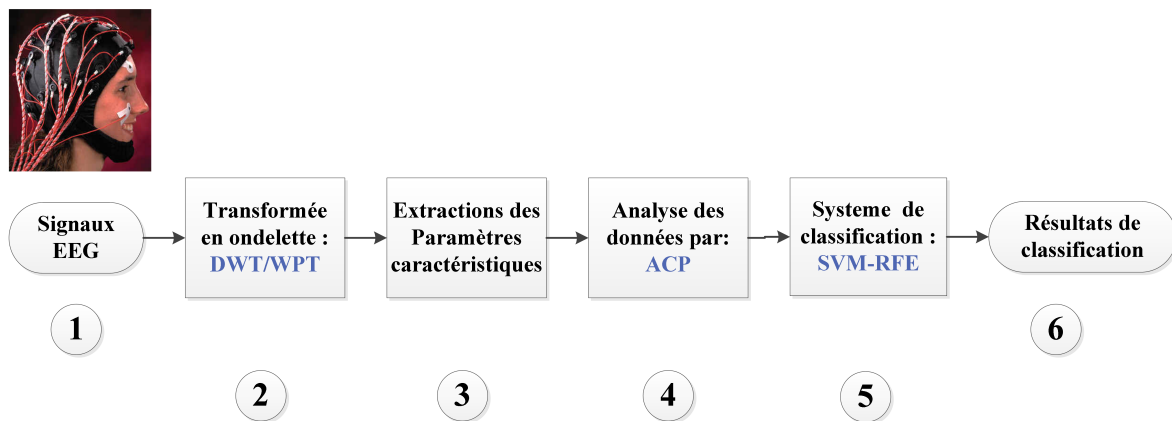


FIGURE 5.1: Structure du système de classification automatique de signaux EEG.

1. j'ai utilisé des signaux EEG issus de la base de données [4], qui sont composés du set A (personnes normales : yeux ouverts), et set E (personnes épileptiques pendant la crise).
2. Pour l'analyse des signaux EEG, j'ai appliqué la transformée en paquets d'ondelettes (WPT).
3. Afin de réduire la dimension des données traitées dans cette étude, six paramètres statistiques et non-statistiques ont été utilisées dans ce projet, à savoir : le Maximum, le Minimum, la Gamme, l'Écart-type, l'Entropie et l'Énergie des coefficients (approximation et détails) de chaque niveau de décomposition.
4. Les paramètres extraits contiennent une redondance qui risque de réduire les performances de notre classificateur, pour cela j'utilise l'analyse en composante principale (ACP) afin de remédier à ce problème, et de ne garder que les données utiles.

5. La méthode de classification utilisée est la machine à support de vecteurs (SVM), qui permet de faire une séparation entre les signaux EEG normaux et les signaux EEG épileptiques, suivi d'un traitement par l'algorithme d'élimination récursive des paramètres caractéristiques (RFE) permettant de trouver le ou les paramètres qui donne(ent) la meilleure précision de classification.
6. Afin d'évaluer les performances de notre classificateur, trois paramètres ont été utilisés : précision, sensibilité, et la spécificité.

5.3 Signaux EEG

Un exemple des signaux EEG appartenant aux Set A, E utilisés dans ce mémoire est illustré sur la Figure 5.2.

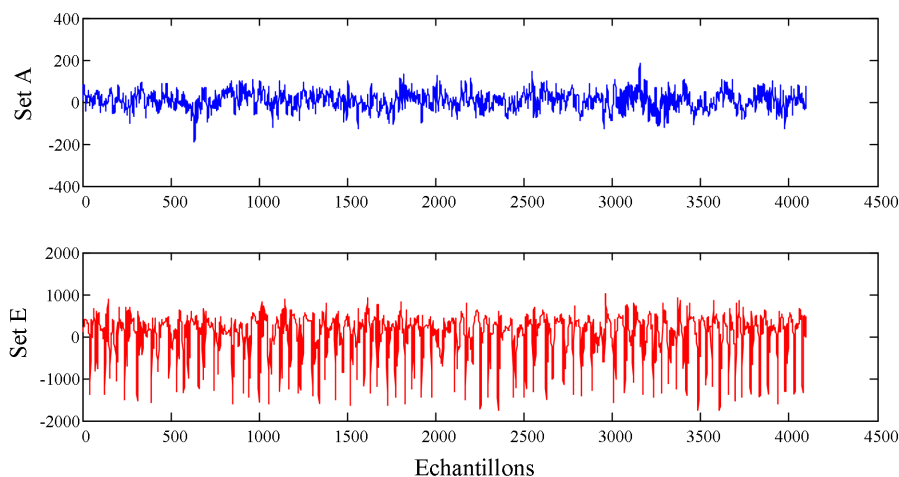


FIGURE 5.2: Exemple des signaux EEG de différents Set utilisés dans ce mémoire.

Nous remarquons une différence de forme et d'amplitude entre les différents signaux, en effet les variations du signal issu du set A sont moins importantes que celles du signal issu du Set E. Les amplitudes qu'atteint le signal du Set E sont très grandes par rapport à celles du Set A ce qui ma poussé à utiliser ces différences comme critères de séparation entre ces classes. Pour cela j'ai choisi d'utiliser les paramètres statistiques suivants : le Maximum (Max), le Minimum (Min), la Gamme (Gam), l'Écart-type (Ect), d'autres

paramètre ont été utilisés tel que : l'Énergie (Eng), et l'Entropie (Ent) pour évaluer le désordre et la complexité du signal.

5.4 La transformée en paquets d'ondelette des signaux EEG

Dans cette section nous allons mettre en évidence l'application de la transformée en paquets d'ondelette sur nos signaux EEG

Nous avons besoin de choisir une ondelette mère et un niveau de décomposition optimaux pour l'analyse des signaux EEG en paquets d'ondelettes (WPT), en suivant la même démarche que pour l'analyse en ondelette discrète nous trouvons que l'ondelette mère « db1 » avec un niveau de décomposition égale à 3 mène à une analyse optimale en paquets d'ondelette pour les signaux EEG.

La Figure 5.3 donne un aperçu sur le choix de l'ondelette mère pour l'analyse en paquets d'ondelettes.

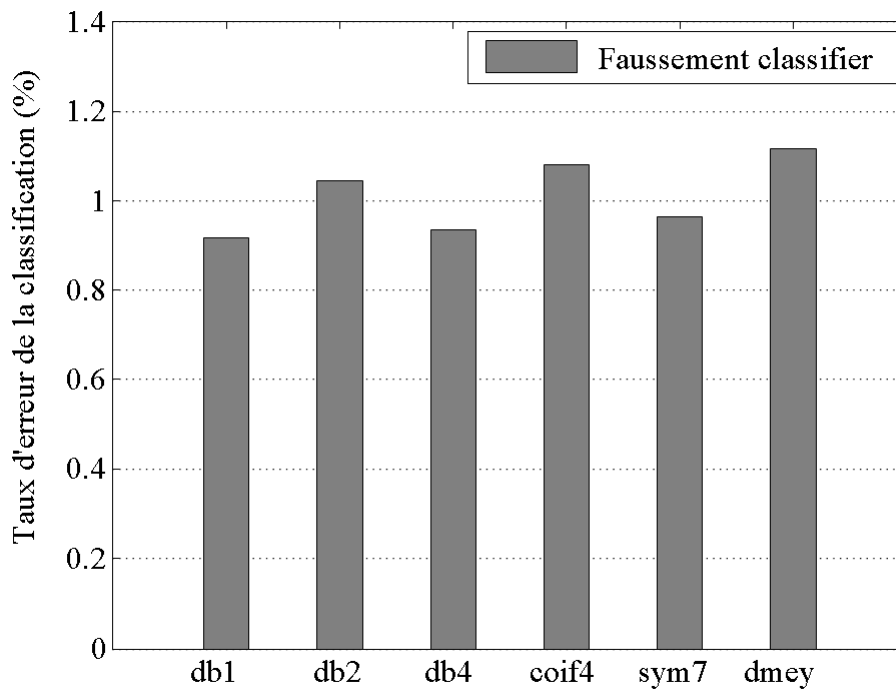


FIGURE 5.3: Le Taux d'erreur de classification obtenu pour différentes familles d'ondelettes en utilisant la WPT.

La structure de l'arbre en WPT est affiché sur la Figure 5.4 :

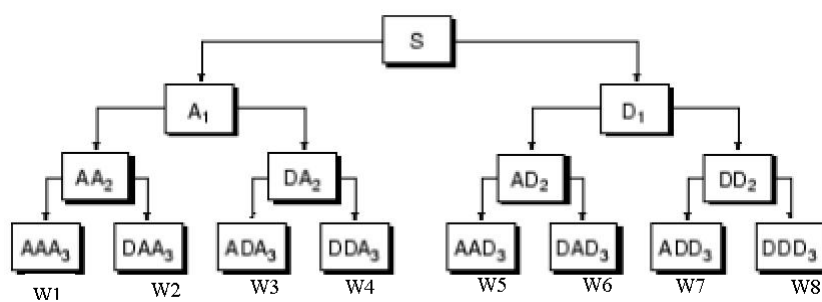


FIGURE 5.4: Structure de la décomposition en paquet ondelette (3 niveau) utilisés dans ce mémoire.

Pour l'analyse en paquets d'ondelette, l'ondelette mère choisi est la db1 avec un niveau de décomposition égale à 3.

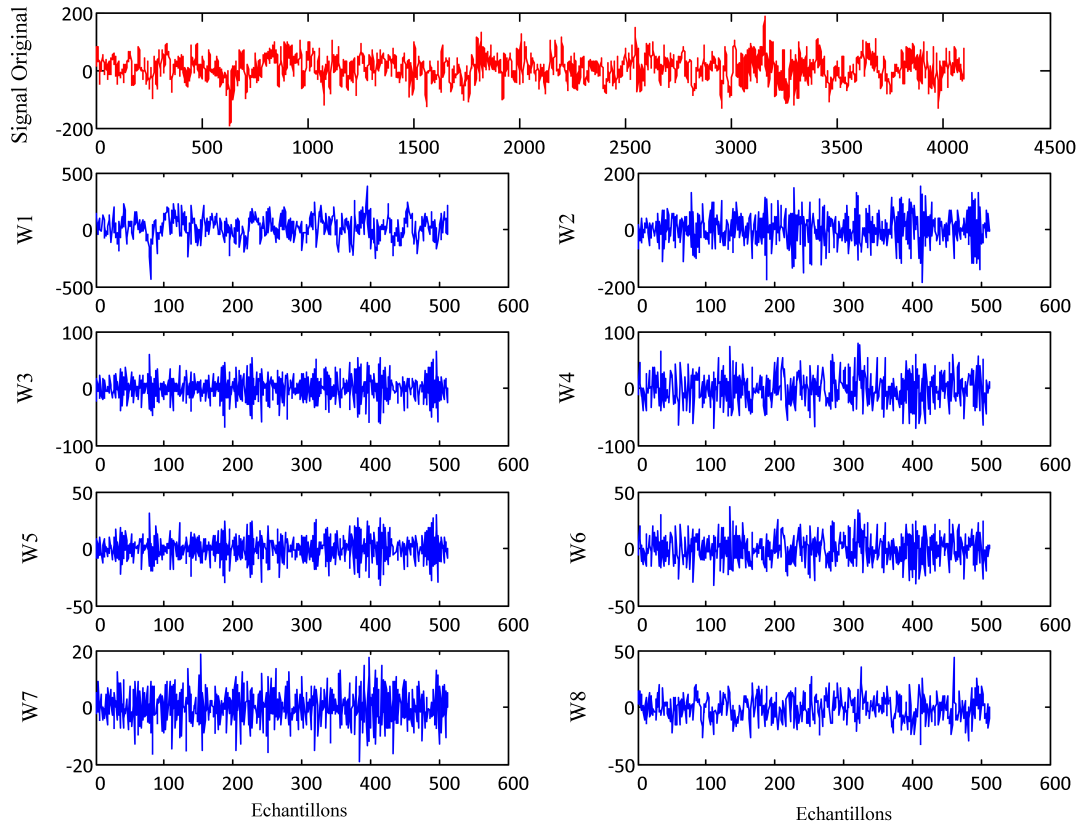


FIGURE 5.5: La décomposition d'un signal EEG du set A en paquets d'ondelettes (WPT), avec l'ondelette mère « db1 » et un niveau de décomposition 3.

le signal original sera représenté par les huit coefficients de paquets d'ondelettes
les paquets d'ondelettes obtenus (W1, ..., W8) avec la décomposition représentent le signal original; dans la suite de l'étude on va utiliser ces coefficients pour extraire de l'information en utilisant quelques paramètres caractéristiques.

5.5 Extraction des paramètres caractéristiques

Dans cette section nous allons extraire des paramètres caractéristiques à savoir le Maximum (Max), le Minimum (Min), la Gamme (Gam), l'Écart-type (Ect), l'Entropie (Ent), et l'Énergie (Eng) avec lesquels nous allons manipuler nos signaux pour pouvoir les classer par la suite ; pour cela notre étude portera le groupe 1 (Set A-Set E) de signaux EEG .

5.5.1 Extraction pour les signaux EEG

Pour justifier le choix de nos paramètres caractéristiques, j'ai préféré tracer des courbes pour paramètre du groupe 1 dans les Figures 5.6, 5.7, 5.8, 5.9, 5.10, 5.11.

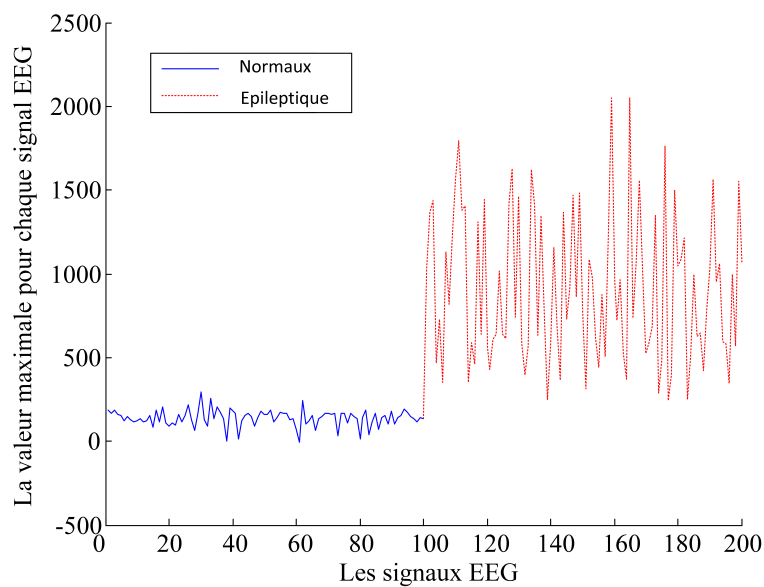


FIGURE 5.6: Le paramètre Maximum extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

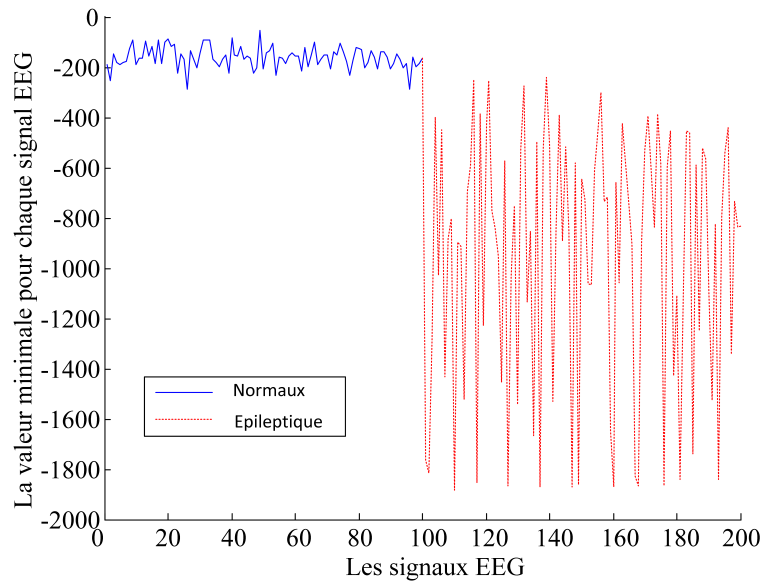


FIGURE 5.7: Le paramètre minimum extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

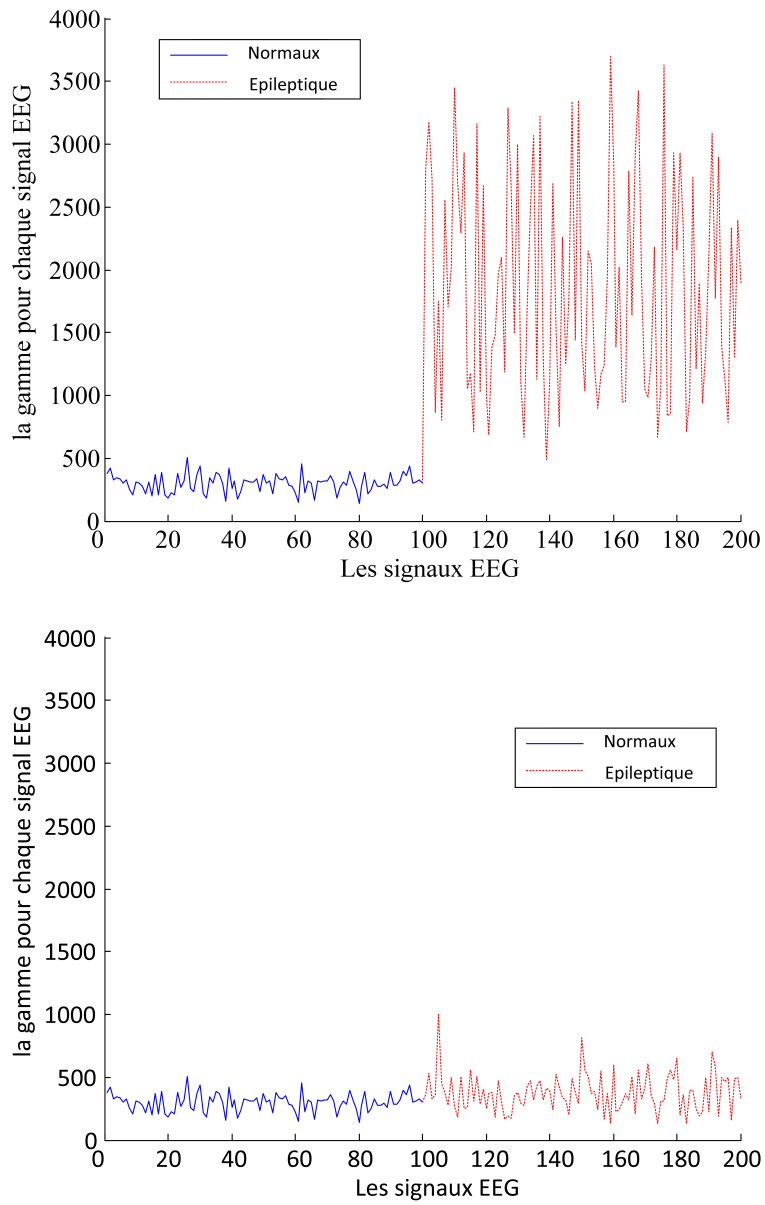


FIGURE 5.8: Le paramètre Gamme extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

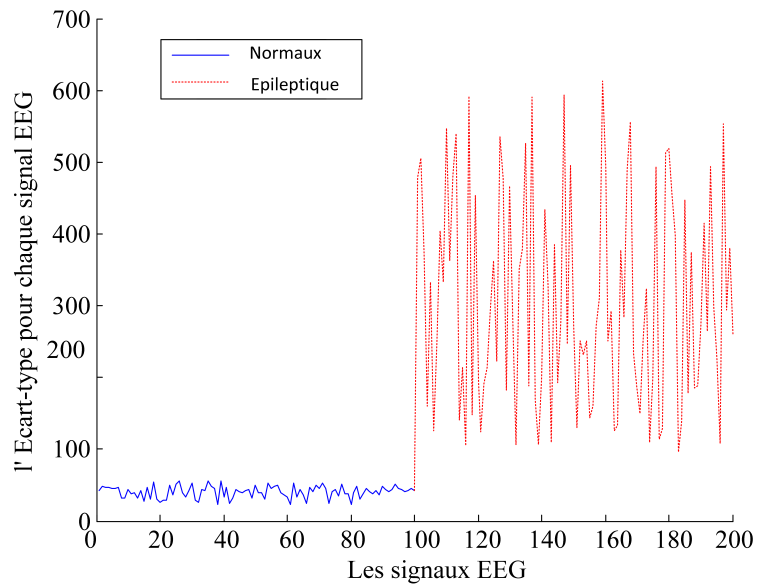


FIGURE 5.9: Le paramètre écart-type extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

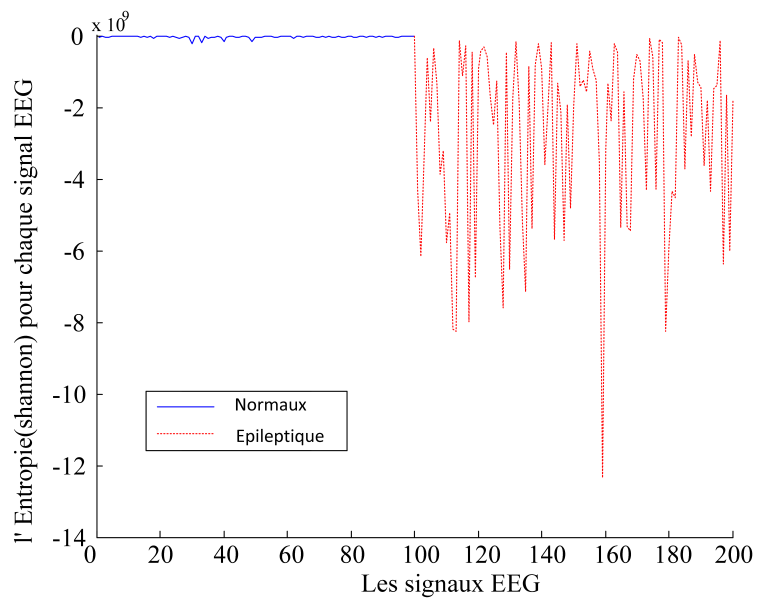


FIGURE 5.10: Le paramètre Entropie extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

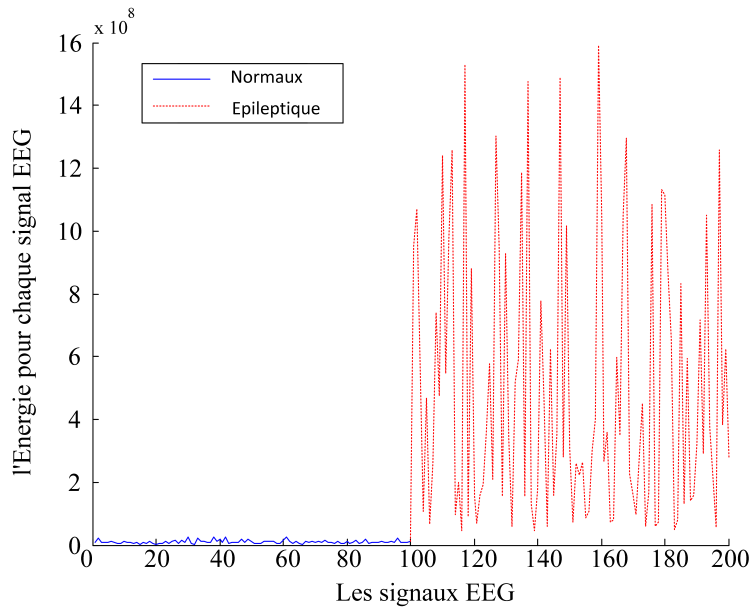


FIGURE 5.11: Le paramètre énergie extrait des signaux EEG avec les 100 signaux normaux à gauche et les 100 signaux épileptiques à droite.

D'après les figures obtenues, la différence visible entre les signaux normaux et les signaux épileptique explique le choix de nos paramètres.

5.5.2 Extraction des paramètres après la décomposition en paquets d'ondelette (WPT)

L'extraction des paramètres est effectuée sur les coefficients de paquets d'ondelette (W_1, W_2, \dots, W_8) est affiché dans le Tableau 5.1.

Remarque :

Pour pouvoir afficher ces résultats et avoir l'ordre de grandeur pour chaque paramètre, nous avons calculé la moyenne des résultats obtenus des 100 signaux EEG du Set A, Set E .

TABLE 5.1: Les paramètres extraits des signaux EEG du groupe 1 pour chaque paquet d'ondelette.

Paramètre	W1	W2	W3	W4	W5	W6	W7	W8	
Set A	Max	282.65	168.41	66.08	98.33	32.88	44.36	23.00	36.77
	Min	-331.35	-168.66	-67.44	-96.33	-33.10	-42.48	-22.41	-38.15
	Gam	613.99	337.06	133.51	194.66	65.98	86.84	45.41	74.92
Set B	Ect	93.83	50.59	20.46	30.26	10.05	13.55	7.19	11.79
	Ent	$-2.9 \cdot 10^7$	$-5.97 \cdot 10^6$	$-7.79 \cdot 10^5$	$-1.94 \cdot 10^6$	$-1.50 \cdot 10^5$	$-3.08 \cdot 10^5$	$-7.3 \cdot 10^5$	$-2.4 \cdot 10^5$
	Eng	$7.57 \cdot 10^8$	$1.14 \cdot 10^8$	$2.30 \cdot 10^7$	$5.16 \cdot 10^7$	$5.57 \cdot 10^6$	$1.03 \cdot 10^7$	$2.98 \cdot 10^6$	$8.14 \cdot 10^6$
Set C	Max	1759.07	1249.21	472.51	776.08	223.69	341.69	138.91	270.11
	Min	-1822.91	-1306.74	-489.04	-755.78	-231.03	-335.38	-144.68	-272.29
	Gam	3581.98	2555.95	961.55	1531.85	454.72	677.07	283.59	542.40
Set D	Ect	683.71	416.88	165.66	214.82	79.34	96.84	38.55	69.76
	Ent	$-2 \cdot 10^9$	$-8.2 \cdot 10^8$	$-1 \cdot 10^8$	$-2.4 \cdot 10^8$	$-2.1 \cdot 10^7$	$-4.2 \cdot 10^7$	$-4.8 \cdot 10^6$	$-2.1 \cdot 10^7$
	Eng	$2.99 \cdot 10^{10}$	$1.22 \cdot 10^{10}$	$1.85 \cdot 10^9$	$3.6 \cdot 10^9$	$4.24 \cdot 10^8$	$7.28 \cdot 10^8$	$1.09 \cdot 10^8$	$3.93 \cdot 10^8$

5.6 Réduction des données par l'Analyse en composantes principales (ACP)

L'application de l'ACP sur les données du groupe 1 (Max, Min, Gam, Ect, Ent, Eng) donne les résultats illustrés dans la Figure 5.12 :

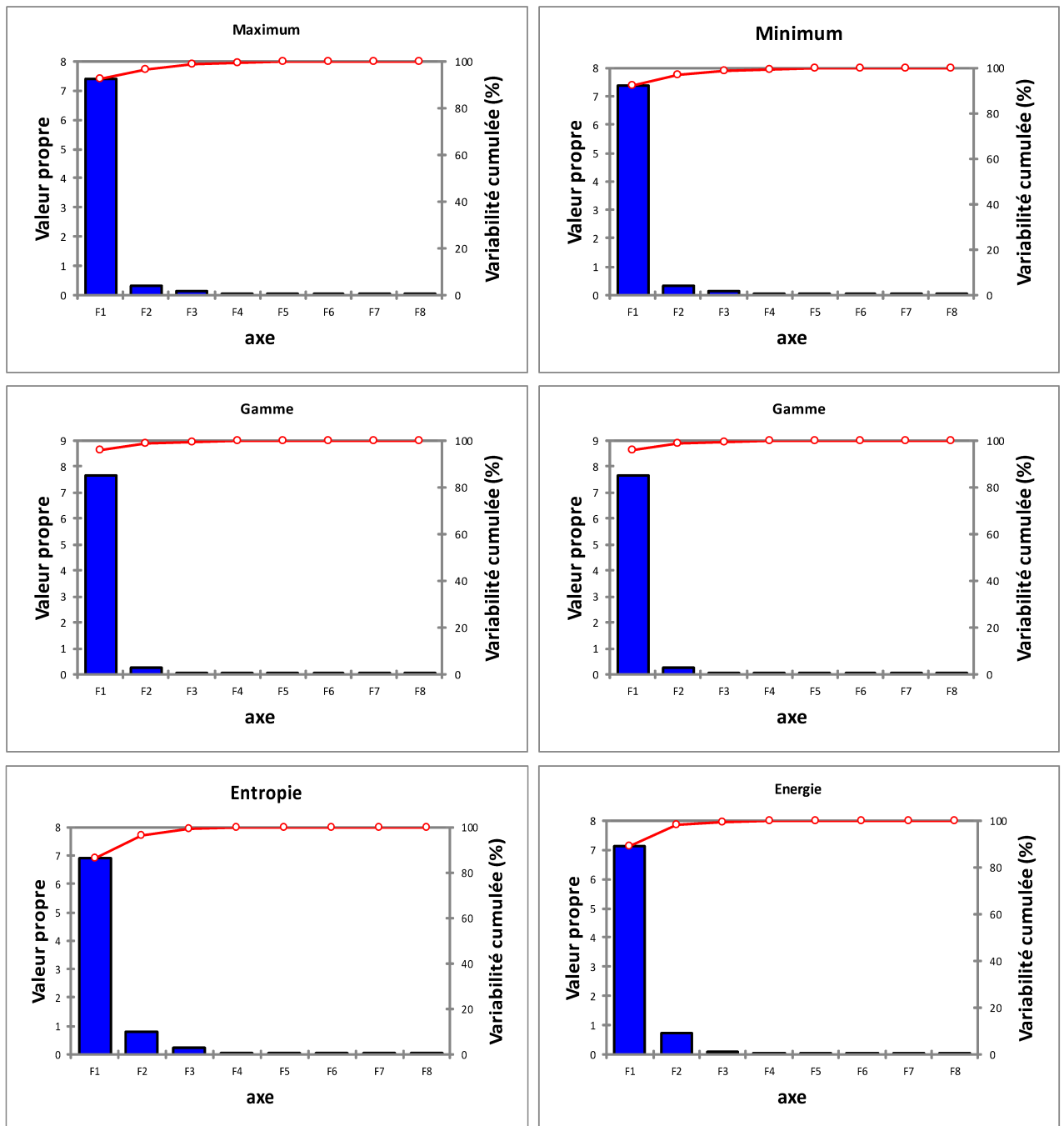


FIGURE 5.12: Histogramme des valeurs propres et variabilités cumulées pour chaque donnée (groupe 1 avec WPT).

TABLE 5.2: Résultats de l'application de l'ACP sur les données (WPT) pour deux axes principaux.

données	variabilité cumulée (%)	pertes (%)
Max	96.54	3.46
Min	96.63	3.37
Gam	98.77	1.23
Ect	99.19	0.81
Ent	96.4	3.6
Eng	98.45	1.55

La Figure 5.12 représente l'histogramme des valeurs propres pour chaque donnée, nous pouvons voir que la cassure au niveau des valeurs propres apparaît au-delà du deuxième axe principale (F2) ce qui signifie que nous pouvons prendre uniquement ces deux derniers pour représenter l'ensemble des données sans perte significative d'informations. Le Tableau 5.2 représente la quantité d'information utile (Variabilité cumulée) et les pertes (redondance et informations) résultantes de l'ACP en utilisant seulement deux axes principaux, ce qui justifie notre passage de **huit** axes (nombres de coefficients) à **deux** axes; cela réduit la quantité de données (dimensions), en gardant un maximum d'informations utiles, et un minimum de pertes, permettant ainsi un gain considérable en temps de calcul et en espace mémoire utilisé par notre programme.

5.7 Classification des signaux EEG par la méthode SVM-RFE

Dans cette partie nous ferons une classification des signaux EEG (séparer les signaux normaux des signaux épileptiques) pour le groupe 1 par la méthode SVM combiner avec l'algorithme RFE, et pour cela j'ai choisi de faire l'apprentissage et le test en utilisant la technique de la validation croisée m-fold (m=2, 5, 10) . j'exploiterai les paramètres de performance à la fin de la classification pour évaluer le système de classification.

Remarque :

- Le noyau utilisé dans la méthode de classification SVM pour la séparation des classes est un noyau Gaussien « RBF » (radial basis function).
- Les pourcentages 50%-50%, 80%-20% et 90%-10% des couples *apprentissage-test* correspondent respectivement aux $m=2$, $m=5$ et $m=10$.

5.7.1 Classification des signaux EEG du groupe 1 par la méthode SVM-RFE en utilisant la WPT

Après l'application de notre programme de classification SVM pour chaque paramètre nous obtenons le Tableau 5.3.

- Nous remarquons que l'augmentation des données utilisées pour l'apprentissage (m) donne une amélioration des résultats (Précision, Spécificité), on voit bien que les résultats obtenus avec les paramètres statistiques sont meilleurs que ceux obtenus avec les paramètres non-statistiques en raisons des différences d'amplitudes des signaux EEG du groupe 1.

La Figure 5.13 présente une illustration graphique des résultats obtenus dans le Tableau 5.3.

TABLE 5.3: Les performances de la classification du groupe 1 par la SVM pour différents paramètres caractéristiques en utilisant la WPT.

m-fold	Performances	Maximum	Minimum	Gamme	Écart-type	Entropie	Énergie
m=2	Précision(%)	98, 11	98, 58	98, 96	98, 99	88, 89	90, 43
	Sensibilité(%)	99, 99	100	100	100	100	100
	Spécificité(%)	96, 23	97, 16	97, 92	97, 98	77, 78	80, 85
m=5	Précision(%)	98, 89	98, 97	99, 13	99, 07	89, 94	91, 89
	Sensibilité(%)	100	100	100	100	100	100
	Spécificité(%)	97, 78	97, 95	98, 25	98, 14	79, 88	83, 78
m=10	Précision(%)	98, 98	99, 17	99, 10	99, 18	90, 25	92, 58
	Sensibilité(%)	100	100	100	100	100	100
	Spécificité(%)	97, 96	98, 35	98, 20	98, 36	80, 49	85, 16

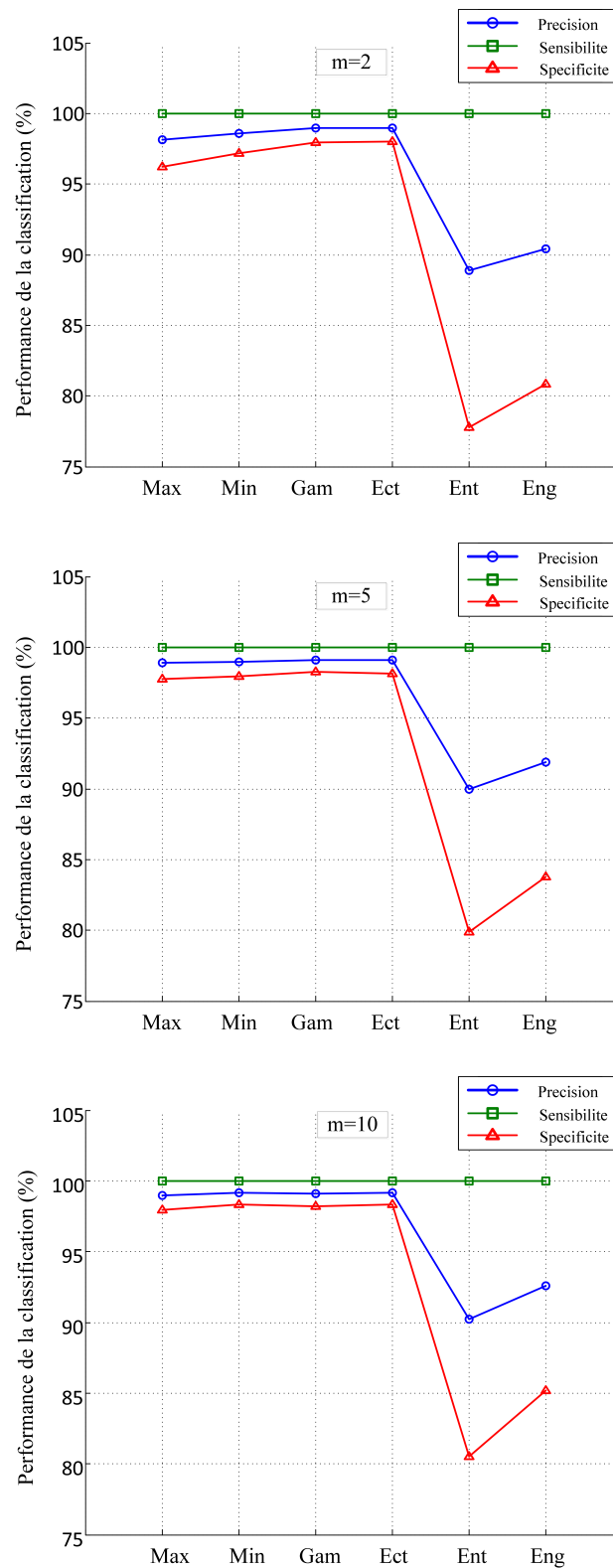


FIGURE 5.13: Les performances de la classification du groupe 1 par la SVM pour différents paramètres caractéristiques en utilisant la WPT pour $m = 2, 5, 10$.

Après simulation du programme SVM combiner avec l'algorithme RFE sur le groupe 1 pour des « m » différents, nous obtenons le Tableau 5.4.

Nous remarquons que nous pouvons avoir de très bons résultats avec une précision de 99,89 %, en utilisant seulement la combinaison de deux paramètres qui sont le Maximum et la Gamme avec $m=10$.

- La Figure 5.14 représente les variations de taux d'erreurs en réduisant à chaque fois le nombre de paramètres caractéristiques.

TABLE 5.4: Résultat après l'utilisation de la SVM-RFE sur le groupe 1 pour $m=2, 5, 10$, en utilisant WPT.

Étape	Nombre de caractéristiques	Précision (%)	Sensibilité (%)	Spécificité (%)	Taux d'erreur (%)
m=2					
1	6 [Max Min Gam Ect Ent Eng]	99, 53	100	99, 07	0.47
2	5 [Max Gam Ect Ent Eng]	99, 55	100	99, 10	0.45
3	4 [Max Gam Ect Eng]	99, 53	100	99, 07	0.47
4	3 [Max Ect Eng]	99, 59	100	99, 18	0.41
5	2 [Max Ect]	99, 56	100	99, 12	0.44
6	1 [Ect]	98, 99	100	97, 98	1.01
m=5					
1	6 [Max Min Gam Ect Ent Eng]	99, 50	100	99	0.50
2	5 [Max Min Gam Ent Eng]	99, 63	100	99, 26	0.37
3	4 [Max Gam Ent Eng]	99, 73	100	99, 47	0.27
4	3 [Max Gam Eng]	99, 84	100	99, 68	0.16
5	2 [Max Gam]	99, 81	100	99, 61	0.19
6	1 [Gam]	99, 13	100	98, 25	0.87
m=10					
1	6 [Max Min Gam Ect Ent Eng]	99, 50	100	99	0.50
2	5 [Max Min Gam Ent Eng]	99, 66	100	99, 32	0.34
3	4 [Max Gam Ent Eng]	99, 91	100	99, 81	0.09
4	3 [Max Gam Eng]	99, 91	100	99, 81	0.09
5	2 [Max Gam]	99, 89	100	99, 79	0.11
6	1 [Gam]	99, 10	100	98, 20	0.90

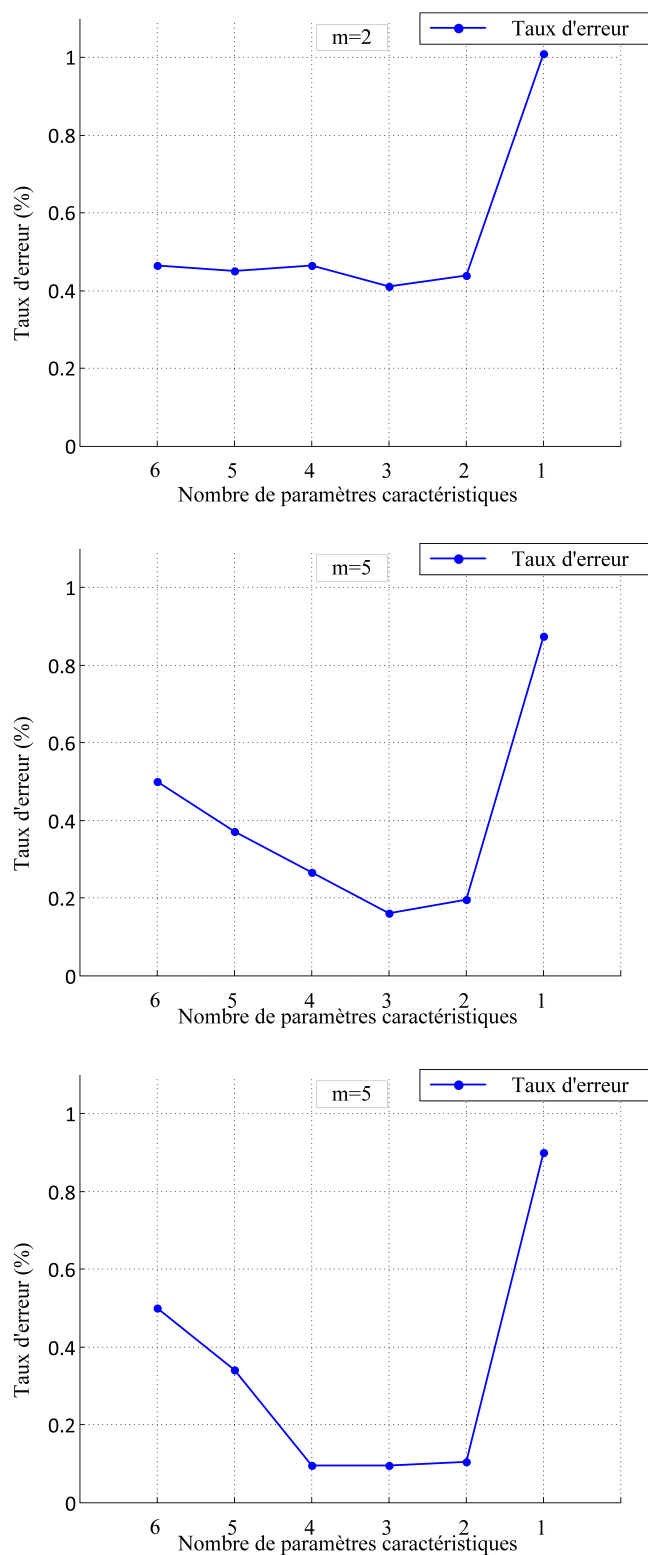


FIGURE 5.14: Le Taux d'erreur en fonction des paramètres caractéristiques choisis dans chaque étape de l'algorithme SVM-RFE du groupe 1 en utilisant la WPT pour $m=2, 5, 10$.

Le classement des meilleurs paramètres caractéristiques par ordre décroissant d'après l'algorithme RFE (en se basant sur le taux d'erreur de classification) est donné dans le Tableau 5.5.

TABLE 5.5: Classement en ordre croissant (du premier au dernier) des paramètres caractéristiques par la SVM-RFE du groupe 1 en utilisant la WPT pour $m=2$, 5, 10.

Classement	Paramètres caractéristique (m=2)	Paramètres caractéristique (m=5)	Paramètres caractéristique (m=10)
1	Écart-type	Gamme	Gamme
2	Maximum	Maximum	Maximum
3	Énergie	Énergie	Énergie
4	Gamme	Entropie	Entropie
5	Entropie	Minimum	Minimum
6	Minimum	Écart-type	Écart-type

Cette étude ma mené à un système développé pour la classification automatique des signaux EEG (Saints, Épileptiques) schématisé par la Figure 5.15.

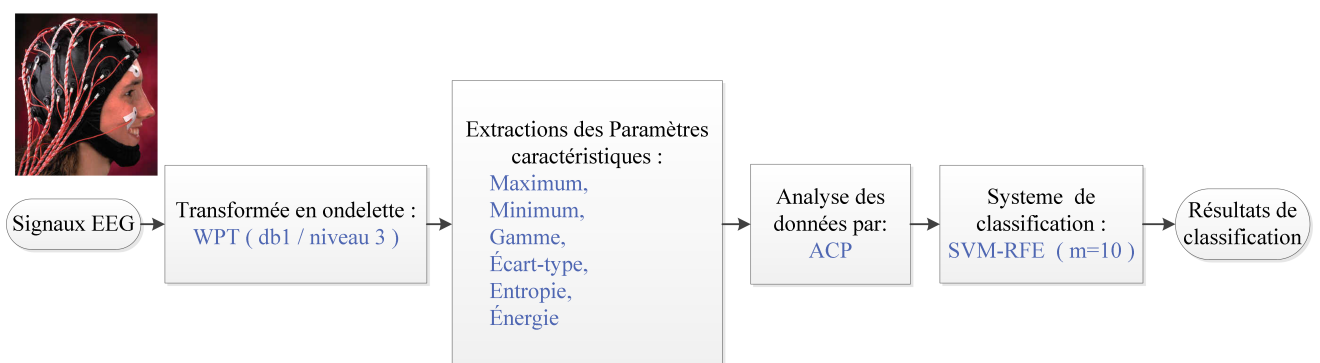


FIGURE 5.15: Système de classification retenu dans cette étude .

5.8 Conclusion

Après avoir simulé les programmes et évalué les performances de notre système de classification pour différents paramètres caractéristiques, et test-apprentissage. nous concluons que l'analyse en WPT donne des résultats performants avec une précision de classification de 99.89% (groupe 1), et ceci a été obtenu avec $m=10$ (apprentissage = 90%, test = 10%).

Conclusion et perspectives

Bilan des travaux effectués

Dans le but de chercher un système performant pour la classification des signaux EEG, J'ai utilisé la méthode d'analyse en paquet d'ondelettes (WPT). Afin d'extraire des informations significatives de ces signaux, l'étude m'a mené à choisir six paramètres caractéristiques (statistiques et non-statistiques). Pour enlever les redondances et réduire la dimension des données traitées, une analyse en composantes principales (ACP) a été utilisée. Pour des raisons de performances et de rapidité la méthode de classification machine à vecteurs de support (SVM) a été utilisée combinée à l'algorithme d'élimination récursive des paramètres caractéristiques (RFE), pour pouvoir évaluer la combinaison des paramètres donnant les meilleures performances. J'ai trouvé que les résultats de classification des signaux EEG avec les paramètres caractéristiques suivants : Maximum et Gamme pour classifier les signaux EEG du groupe 1 avec un pourcentage apprentissage-test de 90%-10%.

Apports, limites et perspectives

- J'envisage comme continuité à mon travail, expérimenter et évaluer d'autres méthodes de réduction et de classification, afin de sélectionner les plus performantes et les plus rapides pour améliorer mon algorithme et l'implémenter sur une carte DSP pour en faire un outil utilisable par les médecins .
- étendre l'étude pour d'autre type de pathologie et d'autres types de signaux biomédicaux

Bibliographie

- [1] ROMO VAZQUER. Contribution à la detection et l'analyse des signaux eeg épileptique :débruitage et séparation de sources. *Universite de Nancy (INPL)*, 2010.
- [2] Elana Zion-Golumbic. What is eeg? *The Department of Psychology and the Department of Cognitive Science*.
- [3] Lehnertz K Mormann F. Rieke C David P Elger C. E. Andrzejak, R. G. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity : Dependence on recording region and brain state. *Physical Review E*, vol. 64, no. 6, pp. 061907, 2001.
- [4] data base. Eeg time series are available under. [http ://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html](http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html).
- [5] Ludovic Mercier. Les machines à vecteurs support pour la classification en imagerie hyperspectrale-inp grenoble implémentation et mise en oeuvre. *CENTRE D'ENSEIGNEMENT DE GRENOBLE*, www.coriolys.org/images/Sciences/rapportfinal.pdf, 2010.
- [6] WHISHAW KOLB, B Et I. Cerveau et comportement. *Deboeck Universite*, 2002.
- [7] D. E WESTEN. Psychologie : Pensée, cerveau et culture. *DeBoeck Université*, 2000.
- [8] Montain BERNARD. Méthodologie d'analyse des synchronisations neuronales dans les signaux eeg à l'aide de graphes d'informations temps-fréquence. *UNIVERSITÉ DE POITIERS*, 2002.
- [9] Victor BRAZ. Interface cerveau-machine. *ensem*, 2008.
- [10] S. Lasaulce. Module ondelette du dea tis. [http ://fr.scribd.com/doc/95560884/Ondelettes-Lecture-Notes-Wave](http://fr.scribd.com/doc/95560884/Ondelettes-Lecture-Notes-Wave), 2010.
- [11] mathworks. wavelet toolbox,user's guide,continuous wavelet transform. [http ://www.mathworks.com/help/toolbox/wavelet/g3-1000759.html](http://www.mathworks.com/help/toolbox/wavelet/g3-1000759.html), 2012.

- [12] Wang. D ; Miao. D ; Xie. Cl. Best basis-based wavelet packet entropy feature extraction and hierarchical eeg classification for epileptic detection. *Expert Systems with Applications*, vol. 38, no. 11, pp. 14314-14320, 2011.
- [13] Z. K Shinde, A ; Hou. A wavelet packet based sifting process and its application for structural health monitoring. *Master thesis, Faculty of Worcester Polytechnic Institute*, 2004.
- [14] I Daubechies. lectures on wavelets. philadelphia. *Capital City Press*, 10nd edition, 1992.
- [15] Wikipedia. <http://fr.wikipedia.org/wiki/Ecart-type>.
- [16] Wikipedia. <http://fr.wikipedia.org/wiki/EntropiedeShannon>.
- [17] S. Robin C. Duby. Analyse en composante principales. www.agroparistech.fr/IMG/pdf/ACP2006.pdf, 2006.
- [18] Pierre-Louis Gonzalez. L'analyse en composantes principales. <http://maths.cnam.fr/IMG/pdf/A-C-P-.pdf>.
- [19] Subasi Alkan, Koklukaya. Automatic seizure detection in eeg using logistic regression and artificial neural network. *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167-76, 2005.
- [20] Ubeyli Guler. Adaptive neuro-fuzzy inference system for classification. *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113-121, 2005.
- [21] Bishop. 2006.
- [22] Sahin Ogulata Erol Aslan, Bozdemir. A radial basis function neural network model for classification of epilepsy using eeg signals. *The Journal of Medical Systems*, vol. 32, no. 5, 2008.
- [23] Homan Dashei Wunsch Petrosian, Prokhorov. Recurrent neural network based prediction of epileptic seizures in intra and extracranial eeg. *Neurocomputing*, vol. 30, no. 1-4, pp. 201-218, 2000.
- [24] Mourad Adnane ; Zhongwei Jiang ; Zhonghong Yan. Sleep-wake stages classification and sleep efficiency estimation using single-lead electrocardiogram. *Expert Systems with Applications*, vol. 39, pp. 1401-1413, 2012.
- [25] C.-C. et LIN C.-J HSU, C.-W ; CHANG. A practical guide to support vector classification. *Rapport technique, National Taiwan University*, www.csie.ntu.edu.tw/~cj-lin/papers/guide/guide.pdf, 2009.