



THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

Diplôme National - Arrêté du 7 août 2006

ÉCOLE DOCTORALE MATHÉMATIQUES ET INFORMATIQUE

Spécialité : Mathématiques appliquées

Présentée par

Laurent Vézard

Réduction de dimension en apprentissage supervisé. *Applications à l'étude de l'activité cérébrale*

Directrice de thèse : **Marie Chavent**

Co-direction : **Frédérique Faïta-Aïnseba, Pierrick Legrand**

Soutenue le 13 décembre 2013
Devant la Commission d'Examen

JURY

M. Saracco Jérôme	Professeur d'Université	Président
M. Biernacki Christophe	Professeur d'Université	Rapporteur
Mme Lutton Evelyne	Directeur de Recherche INRA	Rapporteur
Mme Chavent Marie	Maître de conférences, HDR	Examineur
Mme Faïta-Aïnseba Frédérique	Maître de conférences	Examineur
M. Legrand Pierrick	Maître de conférences	Examineur

Thèse préparée à l'Université Bordeaux 1
351 cours de la libération - 33405 Talence Cedex

Laboratoires d'accueil :
INRIA Bordeaux Sud Ouest
200 Rue Vieille Tour - 33400 Talence

Institut de Mathématiques de Bordeaux (IMB)
Université Bordeaux 1
351, cours de la Libération - F 33405 TALENCE cedex

Remerciements

Mes premiers remerciements vont naturellement à mes directeurs de thèse, Marie Chavent, Frédérique Faïta-Aïnseba et Pierrick Legrand pour la qualité de leur encadrement et leur gentillesse. Je tiens à les remercier pour leur patience et la qualité de leurs explications. Leurs disponibilités et leurs qualités pédagogiques m'ont permis d'avancer et de mener à bien cette thèse. Je vous remercie donc tous les trois du fond du cœur, vous m'avez beaucoup appris, tant scientifiquement qu'humainement.

Je remercie sincèrement les membres du jury qui ont accepté d'assister à ma soutenance de thèse. Je remercie particulièrement mes rapporteurs, Evelyne Lutton et Christophe Biernacki, pour avoir évalué ce travail. Je remercie également Jérôme Saracco pour m'avoir fait l'honneur de faire partie du jury. Merci à vous pour vos commentaires pertinents et vos conseils.

J'ai également une pensée pour les chercheurs avec lesquels j'ai eu la chance d'échanger durant cette thèse. Ainsi, je remercie chaleureusement Leonardo Trujillo pour qui l'océan atlantique ne fut pas un obstacle pour suivre l'évolution de ces travaux et contribuer très activement à ces derniers. Tu as, par ta disponibilité, ta patience et tes nombreuses remarques enrichissantes, éclairé à de nombreuses reprises ces trois années de thèse. Je te remercie pour tes nombreux apports sur nos papiers et notamment pour avoir corrigé l'ensemble de nos papiers en anglais. Merci également pour ta gentillesse et ton accueil lors de la conférence à Cancun. Je remercie aussi Fabien Lotte qui, par son exposé sur le CSP, nous a donné de nouvelles pistes de recherche. Merci également d'avoir partagé tes codes Matlab avec nous. Enfin, je remercie Robin Génuer pour sa gentillesse et pour m'avoir fait profiter de son expertise sur les forêts aléatoires.

Je tiens également à remercier Ingrid Rochel, assistante de recherche à l'Université de Bordeaux Segalen, pour avoir facilité ma venue sur Bordeaux et réglé les petits problèmes administratifs que nous avons rencontrés au début de cette thèse. Je remercie aussi Nicolas Jahier et Catherine Cattaert Mégrat, tour à tour assistant de l'équipe CQFD, pour avoir facilité toutes les démarches administratives au cours de cette thèse.

Je remercie les étudiants de Bordeaux 2 qui, en acceptant d'être les sujets de ce travail, ont permis à celui-ci d'exister. Je garde de très bons souvenirs de ces expériences et de ces moments de partage. J'ai également une pensée pour Julien Clauzel, Vérane Faure, Émilie Drouineau et Nidal El Yacoubi pour leur aide durant les passations. Leurs présences ont permis d'assurer le recueil des données EEG nécessaires à ces travaux. Merci à vous d'avoir participé à cette aventure. Julien, je garde un très bon souvenir de ces quelques mois où l'on a travaillé ensemble. Tes choix de noms de fonctions Matlab et tes commentaires hilarants dans celles-ci ont continué à me faire rire durant l'ensemble de cette thèse. Merci à toi pour ta bonne humeur et ton énergie.

Je remercie Benoite de Saporta de m'avoir proposé d'enseigner les travaux dirigés à Bordeaux IV et de m'avoir soutenu durant la préparation de ces derniers.

J'ai également une pensée pour mes camarades de thèse, Camille, Isabelle, Adrien, Christophe, François, Fred, Raphi, Roro et Philou. Vous avez fait du bureau 35 puis de l'open space CQFD-ALEA des endroits sympathiques de travail et de partage. Cela fut un plaisir de passer ces trois années en vos compagnies.

Pour finir, je souhaite remercier ma famille pour son soutien tout au long de mes études. Je remercie mes parents et Priscilla pour avoir assisté à ma soutenance et pour m'avoir aidé lors de la préparation du pot de thèse. Je remercie très sincèrement Papa et Déborah pour leurs nombreuses relectures de ce manuscrit et pour leur remarques pertinentes qui ont permis de sensiblement améliorer la qualité de ce document.

Enfin, je remercie Marie-Sara qui m'a accompagné, encouragé et soutenu tout au long de cette thèse. Sans elle, tout aurait été beaucoup plus difficile.

Table des matières

Acronymes et abréviations	ix
Table des figures	xi
Liste des tableaux	xxi
Introduction générale	1
I L'activité cérébrale humaine : genèse, acquisition et utilisation	5
1 Structure cérébrale	7
2 Genèse de l'activité électromagnétique cérébrale	8
2.1 Les neurones	8
2.2 Structure du neurone	8
2.3 Origine de l'influx nerveux	9
2.4 Transmission de l'influx nerveux	9
3 Les différentes méthodes d'acquisition de l'activité cérébrale	11
3.1 Les méthodes non invasives	11
3.1.1 L'électro-encéphalographie (EEG)	11
3.1.1.1 Positionnement des électrodes	13
3.1.1.2 Artéfacts	14
3.1.1.3 Montage	15
3.1.1.4 Matériel de recueil de l'EEG	17
3.1.1.5 Les rythmes cérébraux	18
3.1.1.6 Les états non altérés de conscience	21

3.1.1.7	Classification des états de conscience	22
3.1.1.8	Intérêt de l'étude des rythmes cérébraux	23
3.1.1.9	Les potentiels évoqués	23
3.1.2	La magnétoencéphalographie (MEG)	28
3.1.3	L'imagerie par résonance magnétique (IRM)	28
3.1.4	Tomographie par émission de positons (TEP)	30
3.2	Les méthodes invasives	31
3.2.1	La stéréo-électro-encéphalographie (SEEG)	31
3.3	Classification des méthodes d'acquisition	33
4	Interface Cerveau-Machine	34
4.1	Communication	35
4.2	Locomotion	37
4.3	Restauration de mouvements	37
4.4	Applications ludiques	38
4.4.1	Contrôle de jeux vidéos	38
4.4.2	Visite virtuelle de musée	40
4.5	Jeux de données des compétitions BCI	40
II Acquisition de données d'Electro-EncéphaloGraphie		43
1	Recrutement des sujets et conditions expérimentales	44
1.1	Recrutement des sujets	44
1.2	Environnement de l'acquisition	45
2	Préparation du sujet	46
3	Pose du casque	47
4	Enregistrement de signaux EEG	48
5	Séance de relaxation	50
6	Recueil de la variation contingente négative (VCN)	50
7	Remarques sur le protocole expérimental	51
8	Validation des données	52
8.1	Pré-traitement des signaux	52
8.2	Extraction de la VCN et règle de décision	53
8.2.1	Sélection de la période pour le calcul de l'aire	53
8.2.2	Sélection des électrodes par test statistique	54
8.2.3	Sélection des sujets par test statistique	57
8.3	Inspection visuelle de la VCN	59

8.4	Résultats de l'étude de la VCN	60
9	Les données	64
III État de l'art sur la classification des signaux EEG		65
1	Extraction de caractéristiques	68
1.1	Extraction de caractéristiques fondées sur l'analyse du contenu fréquentiel	68
1.1.1	La transformée de Fourier	68
1.1.2	La décomposition en ondelettes	69
1.1.3	Densité spectrale de puissance	73
1.1.4	Extraction de caractéristiques fondées sur le contenu fréquentiel	73
1.2	Extraction de caractéristiques fondées sur le filtrage spatial	74
1.2.1	Le Common Spacial Pattern (CSP)	74
1.2.2	Extraction de caractéristiques par transformation log-var	81
1.2.3	Applications en BCI	82
2	Classification de signaux EEG	82
2.1	Estimation du taux de bons classements	83
2.2	Méthode de classification	85
3	Sélection de variables en BCI	87
3.1	Sélection à l'aide d'algorithmes génétiques	87
3.1.1	Principe général d'un algorithme génétique	87
3.1.2	Utilisation en BCI	89
3.2	Sélection à l'aide d'algorithmes de recherche séquentielle	89
3.2.1	Principe général	89
3.2.2	Utilisation en BCI	90
3.3	Le CSP parcimonieux	91
4	Conclusion	92
IV Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique		93
1	Introduction	94
2	Les données utilisées	95
3	Extraction de caractéristiques	96
3.1	Énergie d'un signal	96
3.2	Le critère des pentes	97
3.3	Le critère de l'exposant de Hölder et le critère des ondes α	99

3.4	Résultats préliminaires	100
4	Sélection de variable et extraction de caractéristiques par algorithme génétique	103
4.1	Choix algorithmiques	103
4.1.1	Opérateurs génétiques	104
4.1.2	Fonctions d'évaluations	106
4.1.2.1	Le SVC	107
4.1.2.2	CART	107
4.1.3	Critère d'arrêt	109
4.2	Résultats	111
5	Conclusions	115
V	Algorithme génétique pour la sélection de variables dans les filtres CSP.	117
1	Introduction	119
2	Présentation des données utilisées	120
3	Classification de signaux EEG avec filtre spatial	121
3.1	Adaptation de la méthode CSP au problème multi-sujets	121
3.2	Pré-traitement des données	122
3.2.1	Filtrage passe-bande des données	122
3.2.2	Normalisation des données	123
3.2.3	Projection des données	125
3.2.4	Nombre de paires de filtres spatiaux	125
3.3	Optimisation des paramètres	125
3.3.1	Estimation du TBC par Leave One Out (LOO)	126
3.3.2	Résultats de l'optimisation des paramètres	128
4	Sélection d'un sous-ensemble d'électrodes par algorithmes génétique	133
4.1	Choix algorithmiques	133
4.1.1	Fonction d'évaluation	134
4.1.2	Opérateurs génétiques	136
4.1.3	Critère d'arrêt	138
4.2	Application expérimentale	138
4.2.1	Stratégie expérimentale	138
4.2.2	Résultats	139
4.2.2.1	Performances de la méthode	139
4.2.2.2	Choix de la valeur de p'	142
4.2.2.3	Électrodes sélectionnées	143

5	Sélection de variables par méthode séquentielle	147
5.1	Le forward et le backward CSP	147
5.2	Résultats	148
6	Prédiction de l'état de vigilance d'un nouvel individu	156
7	Conclusion	157
VI Le CSP parcimonieux		159
1	Introduction	160
2	Le CSP : une double ACP avec métrique	161
2.1	ACP de \mathbf{X}_1 avec métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p	161
2.2	ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p	162
3	L'ACP parcimonieuse	162
3.1	L'approche de Zou, Hastie et Tibschirani (2006)	163
3.1.1	Le critère pénalisé	163
3.1.2	La stratégie d'optimisation	164
3.1.3	L'algorithme et son implémentation	166
3.2	Grille matricielle de paramètres de régularisation	167
4	Le CSP parcimonieux	171
4.1	Le CSP : une double ACP avec métrique identité	171
4.2	L'algorithme de CSP parcimonieux	172
4.3	Grille matricielle de paramètres de régularisation	174
4.4	Un exemple	174
5	Applications	178
5.1	Comparaison avec les résultats de l'algorithme génétique proposé en Chapitre V	178
5.2	Evaluation du sCSP sur 35 découpages différents	181
6	Conclusion	184
Conclusion		187
Références bibliographiques		197

Acronymes et abréviations

ACP : Analyse en Composantes Principales

BCI : Interface Cerveau-Ordinateur (*Brain Computer Interface* en anglais)

CART : *Classification And Regression Trees*

CSP : *Common Spatial Pattern*

ddp : Différence De Potentiel

ECOG : ÉlectroCOrticoGraphie

EEG : ÉlectroEncéphaloGraphie

EMG : Électromyographie

EOG : ÉlectroOculoGramme

FDR : *False Discovery Rate*

FFT : *Fast Fourier Transform*

ICM : Interface Cerveau-Machine

INRIA : Institut National de Recherche en Informatique et Automatique

IRM : Imagerie par Résonance Magnétique

IRMa : Imagerie par Résonance Magnétique anatomique

IRMd : Imagerie par Résonance Magnétique de diffusion

IRMf : Imagerie par Résonance Magnétique fonctionnelle

KNN : K plus proches voisins (K -nearest neighbors)

LASSO : *Least Absolute Shrinkage and Selection Operator*

LDA : Analyse Linéaire Discriminante (*Linear Discriminant Analysis*)

LOO : *Leave One Out*

MEG : MagnétoEncéphaloGraphie

PLS : *Partial Least Squares*

PSD : *Power Spectral Density*

sACP : *Sparse Principal Component Analysis* (ACP parcimonieuse)

SBS : *Sequential Backward Selection*

sCSP : *Sparse Common Spatial Pattern* (CSP parcimonieux)

SEEG : Stéréo-Électro-EncéphaloGraphie

SFS : *Sequential Forward Selection*

SLA : Sclérose latérale amyotrophique

SQUID : *Superconducting Quantum Interference Device*

SSVEP : *Steady State Visually Evoked Potential*

SVC : *Single Variable Classifier*

SVM : Supports Vecteurs Machine

TBC : Taux de Bons Classements

TEP : Tomographie par Émission de Positons

VCN : Variation Contingente Négative

Table des figures

I.1	Représentation des deux hémisphères du cerveau et des différents lobes qui les composent. D'après [1].	8
I.2	Schéma représentant la structure d'un neurone. <i>Crédits : Chantal Proulx</i> . . .	9
I.3	Représentation du fonctionnement d'une synapse chimique.	10
I.4	Electroencéphalogramme d'un sujet sain, acquis par 58 électrodes positionnées selon le système international 10/10 [2].	12
I.5	Représentation des systèmes internationaux de positionnements d'électrodes pour la capture de signaux EEG. Positions des électrodes dans les systèmes respectifs 10/20 (cercles noirs) [3]; 10/10 (ajout des cercles gris) [2] et 10/5 (ajout des cercles blancs) [4]. D'après [4].	13
I.6	Représentation d'un artéfact musculaire contaminant l'électroencéphalogramme d'un sujet sain enregistré à l'aide de 58 électrodes placées selon le système international 10/10. La partie encadrée en bleu correspond au bruit généré par un artéfact musculaire.	15
I.7	Représentation de l'électroencéphalogramme d'un sujet sain enregistré à l'aide de 58 électrodes placées selon le système international 10/10 lors de l'apparition d'un clignement (entouré en trait plein bleu) et d'une saccade oculaire (entourée en trait pointillé rouge). Le troisième tracé en partant du bas correspond au tracé d'EOG.	16
I.8	Représentation de l'équipement usuel pour l'enregistrement de signaux en EEG.	19
I.9	Principaux rythmes oscillatoires de l'EEG humain. D'après H. Gamboa.	20

I.10	Électroencéphalogramme d'un sujet sain dans un bas niveau de vigilance enregistré à l'aide de 58 électrodes placées selon le système international 10/10. La partie encadrée correspond à une bouffée d'ondes α	20
I.11	Classification des différents états de conscience en fonction du niveau de vigilance. Les états de conscience normaux apparaissent en bleu tandis que les états de conscience altérés apparaissent en rouge. Le rouge foncé correspond à l'état de mort cérébrale (noté M). Ce schéma est adapté à partir de celui proposé dans [5].	22
I.12	Représentation de l'apparition des potentiels évoqués par la technique du moyennage.	26
I.13	Représentation de la variation d'amplitude de la VCN en fonction de l'état de vigilance du sujet.	27
I.14	Représentation de l'appareil de mesure utilisé en MEG ainsi que des capteurs. <i>Credits : Inserm-CERMEP</i>	29
I.15	Représentation de l'appareil de mesure utilisé en IRM ainsi que de différentes images provenant de l'IRMa, l'IRMf et de l'IRMd.	31
I.16	Représentation de l'appareil de mesure utilisé en TEP ainsi que d'un exemple d'image cérébrale obtenue par TEP. Les images proviennent de la thèse de J. Langner [6].	32
I.17	Représentation de capteurs utilisés en SEEG ainsi que d'une implantation d'électrodes.	32
I.18	Illustration schématique des différentes méthodes d'acquisitions de données cérébrales présentées en fonction de leur résolution spatiale, temporelle et de leur degré d'invasivité. Ce schéma est adapté à partir de celui de [7].	33
I.19	Représentation du fonctionnement d'une interface cerveau-machine. Schéma issu de [8].	34
I.20	Représentation du clavier virtuel proposé dans [9] et [10]. Les images sont issues de [10].	36
I.21	Représentation du jeu MindBalance. Le but de ce jeu est de maintenir l'équilibre du personnage qui se déplace le long d'un fil. Cette image provient de [11].	39
I.22	ICM permettant de réaliser une visite virtuelle de musée. Image provenant de https://team.inria.fr/potioc/fr/scientific-subjects/bci-2/	40

II.1	Diagramme représentant la procédure d'acquisition des données EEG. Les cinq étapes clés de la procédure sont encadrées.	44
II.2	Schéma représentant les deux pièces utilisées lors des acquisitions.	46
II.3	Représentation de la répartition des électrodes lors des acquisitions.	47
II.4	Photographie représentant les conditions d'enregistrement des signaux EEG.	49
II.5	Schéma représentant le protocole de recueil de la VCN. La tâche est composée de 50 paires de stimuli « S1 : bip - S2 : carré » durant lesquels le signal EEG du sujet est enregistré.	51
II.6	Représentation de la gestion des artéfacts oculaires.	53
II.7	Représentation de tracés EEG moyens obtenus dans le protocole de recueil de la VCN dans les deux états de vigilance (courbe bleue pour l'état avant relaxation et rouge pour l'état après relaxation). Ces tracés sont ceux de l'électrode FCz.	54
II.8	Liste des électrodes classées par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs inférieurs au seuil de 5% ainsi que les électrodes associées apparaissent en bleu.	55
II.9	Liste des électrodes classées par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs sont corrigées à l'aide de la méthode de Benjamini et Hochberg.	57
II.10	Liste des sujets classés par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs inférieurs au seuil de 5% ainsi que les sujets associés apparaissent en bleu.	58
II.11	Liste des sujets classés par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs sont corrigées à l'aide de la méthode de Benjamini et Hochberg. Les p -valeurs inférieurs au seuil de 5% ainsi que les sujets associés apparaissent en bleu.	58
II.12	Interface graphique réalisée pour l'inspection visuelle des tracés VCN.	62
II.13	Représentation de la VCN enregistrée chez le sujet 4 de l'étude avant (ligne en pointillé) et après (ligne pleine) la séance de relaxation. Les lignes verticales pleines correspondent aux signaux avertisseurs (S1 : signal sonore, S2 : carré). Ce sujet est conservé car la courbe pleine est majoritairement au dessus de la courbe en pointillé entre T1 et T2 (période d'étude de la VCN encadrée par des lignes verticales en pointillé).	63

II.14	Représentation de la VCN enregistrée chez le sujet 9 de l'étude avant (ligne en pointillé) et après (ligne pleine) la séance de relaxation. Les lignes verticales pleines correspondent aux signaux avertisseurs (S1 : signal sonore, S2 : carré). Ce sujet est rejeté car la courbe pleine est majoritairement au dessous de la courbe en pointillé entre T1 et T2 (période d'étude de la VCN encadrée par des lignes verticales en pointillé).	63
II.15	Représentation de la matrice de données en trois dimensions : une pour les sujets, une pour le temps (46000 points correspondant au nombre de points dans un signal EEG de 3 minutes mesuré en utilisant une fréquence d'échantillonnage de 256 Hz) et une pour les électrodes.	64
III.1	Représentation de quelques ondelettes.	70
III.2	Représentation de la grille dyadique pour 4 niveaux de décomposition (4 échelles) correspondant à la décomposition en ondelettes discrète d'un signal x de 2^4 points.	71
III.3	Représentation d'un signal (a) et de la grille dyadique de sa décomposition en ondelettes discrète (b). Les figures sont obtenues à l'aide de la toolbox Fraclab.	72
III.4	L'interface graphique pour l'utilisateur de la toolbox Fraclab.	73
III.5	Représentation des signaux EEG bruts recueillis par les électrodes F6 et Pz. L'utilisation du CSP permet de représenter les signaux projetés à l'aide du premier et du dernier filtre CSP obtenu. Les courbes en rouge et en noir correspondent respectivement à l'essai enregistré durant le mouvement imaginé de la main droite et de la main gauche. Ces graphiques sont obtenus à partir du jeu de données IIIa [12] de la compétition BCI III [13].	81
III.6	Boucle évolutionnaire d'un algorithme génétique classique.	88
IV.1	Représentation de la matrice de données contenant uniquement les données EEG des sujets de la première campagne d'acquisition de données. Il y a trois dimensions : une pour les sujets, une pour le temps (46000 points correspondant au nombre de points dans un signal EEG de 3 minutes mesuré en utilisant une fréquence d'échantillonnage de 256 Hz) et une pour les électrodes.	96
IV.2	Représentation du calcul de l'énergie du signal X aux échelles 1, 2, 3 et 4 grâce à une décomposition en ondelettes discrète d'un X de 2^4 points.	97

IV.3	Représentation des énergies du signal X^p , obtenues à l'aide d'une décomposition en ondelettes discrète dyadique de X^p , en fonction de la fréquence. Pour calculer le critère des pentes, une régression linéaire est effectuée (droite en pointillé) entre 4, 8 et 16 Hz (cercles).	98
IV.4	Représentation de la matrice de données après l'étape d'extraction de caractéristiques. Sur la gauche, les données obtenues après la décomposition en ondelettes discrète. Il y a toujours trois dimensions : une pour les 15 niveaux de décomposition, une pour les sujets et une pour les électrodes. Sur la droite, la matrice de données obtenue après le calcul du critère des pentes dans laquelle seules deux dimensions demeurent : une pour les sujets et une pour les électrodes.	99
IV.5	Critère des pentes sommé sur toutes les électrodes pour chacun des 13 individus (à gauche) et sur tous les sujets pour chaque électrode (à droite)	100
IV.6	Taux de bons classements pour différentes méthodes de classification appliquées sur le critère des pentes.	102
IV.7	Exemple d'un génome dans l'algorithme génétique	104
IV.8	Relation entre le génome et le calcul du critère des pentes.	105
IV.9	Représentation du croisement des électrodes. L'enfant hérite de l'électrode qui se trouve à mi-chemin entre les électrodes de ses deux parents.	106
IV.10	Représentation graphique de la méthode du Single Variable Classifier. Les sujets de l'échantillon d'apprentissage dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. Les moyennes du critère des pentes sont représentées par un triangle bleu (état normal) ou rouge (état relaxé). Un individu de l'échantillon test (rond gris) est affecté à la classe correspondant à la moyenne la plus proche.	108
IV.11	Représentation d'un arbre de décision binaire élagué. Les sujets dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. Chaque feuille de l'arbre est associée à une modalité (normal ou relaxé) représentée par un triangle (bleu ou rouge). La prédiction de la classe d'un sujet de l'échantillon test (rond gris) s'obtient en lui faisant parcourir l'arbre et en l'affectant à la modalité associée à la feuille atteinte par celui-ci (classe « relaxé » dans ce cas).	110
IV.12	Taux de bons classements calculé par validation croisée de type 5 folds avec CART (étoiles) et SVC (cercles) pour chaque lancement de l'algorithme génétique avec 300 parents et 150 enfants.	112

IV.13 Occurrence des électrodes dans les meilleurs génomes trouvés par l’algorithme génétique durant les 100 lancements avec 300 parents, 150 enfants et CART (courbe en tiret pointillé) ou SVC (courbe pleine). 113

IV.14 Nombre de différences au sein des parents en fonction du numéro de l’itération lors d’un lancement de l’algorithme génétique avec 300 parents, 150 enfants et la méthode d’évaluation SVC (a) ou la méthode CART (b). 114

V.1 Représentation des matrices de données correspondant aux signaux EEG obtenus pendant la première campagne d’acquisition (13 sujets) et la seconde (6 sujets). Les matrices sont en 3 dimensions : une pour les sujets (2 enregistrements par sujets), une pour les électrodes et une autre pour les points de mesure (46000 points correspondant au signal EEG recueilli durant 3 minutes avec une fréquence d’échantillonnage de 256 Hz). 121

V.2 Diagramme représentant l’étape de pré-traitement des données. Les données brutes sont filtrées entre a et b Hz et découpées en essais de longueur T . Une normalisation est alors appliquée aux données pour obtenir $\tilde{\mathbf{X}}_1$ et $\tilde{\mathbf{X}}_2$ 126

V.3 Création de l’ensemble d’apprentissage et de l’ensemble test lors de la $m^{\text{ème}}$ itération de l’estimation par LOO du TBC (avec $m = 1, \dots, q$). 127

V.4 TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). **Les données ne sont pas normalisées** mais sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L’extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP. 129

V.5 TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). **Les données sont normalisées en utilisant l’inertie** puis sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L’extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP. 130

V.6	TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). Les données sont normalisées en utilisant les écarts types puis sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L'extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP.	131
V.7	TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre 7 et b Hz avec b qui est un entier entre 10 et 40). Les données sont normalisées en utilisant les écarts types puis sont découpées en essais de 8 secondes (correspondant à $T = 2048$ points). L'extraction de caractéristiques est réalisée en utilisant 3 paires de filtres CSP.	132
V.8	Exemple d'un génome dans l'algorithme génétique. Les composantes égales à 1 conduisent à la conservation du signal des électrodes associées (traits bleus) tandis que les composantes égales à 0 conduisent au rejet du signal des électrodes concernées (traits rouges).	134
V.9	Diagramme représentant l'étape d'évaluation d'un génome lors de l'algorithme génétique. Pour un génome donné, un sous-ensemble d'électrode est sélectionné et un CSP est utilisé afin de construire la matrice des filtres W . Les essais sont alors projetés et une transformation log-var est appliquée pour obtenir Z . Une règle de décision est apprise sur l'ensemble d'apprentissage à l'aide de la méthode LDA. Un TBC est obtenu en utilisant le modèle construit (filtres CSP et règle de décision) sur les essais de l'ensemble test.	136
V.10	Représentation schématique de la méthode de croisement utilisée dans l'algorithme génétique.	137
V.11	Taux de bons classements obtenus en LOO sur les sujets de la première campagne lors de l'algorithme génétique pour différentes valeurs de p' . La ligne en pointillé représente le taux de bons classements de 75.35% obtenu en utilisant les 58 variables initiales (pas de sélection de variable).	139

V.12	Les figures a, b, c, g et h représentent le nombre de sélection de chaque électrode dans le meilleur génome (1 par lancement) trouvé par l’algorithme génétique lors des 15 lancements respectivement pour $p' = 3, 6, 9, 12$ et 15. Les figures d, e, f, i et j sont des représentations en 3D qui permettent de visualiser les meilleurs génomes trouvés par l’algorithme génétique lors des 15 lancements pour respectivement $p' = 3, 6, 9, 12$ et 15.	145
V.13	Représentation 3D permettant de visualiser le nombre de sélections de chaque électrode dans les meilleurs sous-ensemble d’électrodes trouvés par l’algorithme génétique au cours des 15 lancements.	146
V.14	Meilleure combinaison de 9 électrodes obtenue pour le meilleur des 15 lancements de l’algorithme génétique avec $p' = 9$. Les électrodes sélectionnées apparaissent en noirs. Cette figure est obtenue en utilisant la Visual RCSP toolbox [14].	146
V.15	Représentation des TBC obtenus en utilisant le forward CSP (courbe bleue), le backward CSP (courbe noire) et l’algorithme génétique (courbe rouge) en fonction du nombre d’électrodes sélectionnées.	151
V.16	Taux de bons classements obtenus en fonction du nombre de variables sélectionnées par forward CSP (a) et par backward CSP (b). Les TBC sont estimés par LOO sur les sujets de la première campagne. Les droites en traits pointillé correspondent à la valeur de p' choisie pour chacune des deux approches. . . .	152
V.17	Liste d’ajout des électrodes lors des itérations du forward CSP avec les taux de bons classements correspondants.	153
V.18	Liste de retrait des électrodes lors des itérations du backward CSP avec les taux de bons classements correspondants.	153
V.19	Représentation de la localisation des électrodes sélectionnées par forward CSP (a) et par backward CSP (b). Les électrodes sélectionnées apparaissent en noir.	154
V.20	Diagramme de Venn présentant les similitudes entre les sous-ensembles de variables sélectionnées par algorithme génétique (9 électrodes), forward (16) et backward CSP (18).	155
VI.1	Chemin de régularisation produit par la fonction <i>lars()</i> du package <i>R</i> portant le même nom. Les lignes verticales marquent l’introduction d’une nouvelle variable dans le modèle. L’axe supérieur de la figure indique le nombre de variables incluses dans le modèle pour la valeur considérée du paramètre de régularisation.	169

VI.2	Chemin de régularisation produit par la fonction <i>glmnet()</i> du package <i>R</i> portant le même nom. L'axe supérieur de la figure indique le nombre de variables incluses dans le modèle pour la valeur considérée du paramètre de régularisation.	169
VI.3	Représentation des cercles de corrélations pour les deux premières composantes principales des ACP respectives de \mathbf{X}_1 , Z_1 et Z_2	176
VI.4	Première paire de composantes CSP (a) et sCSP (c) et deuxième paire de composantes CSP (b) et sCSP (d).	177
VI.5	Taux de bons classements obtenus par LOO sur les individus de la campagne 1 pour les différents pas de la grille matricielle de paramètres de régularisation. L'axe supérieur de la figure indique le nombre d'électrodes sélectionnées pour le pas correspondant dans la grille.	179
VI.6	Représentation de la localisation des 13 électrodes sélectionnées par sCSP. Les électrodes sélectionnées apparaissent en noir.	181
VI.7	Diagramme de Venn comparant le sous-ensemble de 9 électrodes sélectionnées en utilisant l'algorithme génétique (voir Section 4.2.2.3 du Chapitre V) et celui de 13 électrodes obtenu par sCSP.	182
VI.8	Boîtes à moustaches des TBC (a) et des nombres d'électrodes sélectionnées (b) lors du lancement du sCSP sur 35 découpages différents (13 sujets en échantillon d'apprentissage et 6 en test). Les résultats sont obtenus sur l'échantillon test.	183
VI.9	Représentation de la localisation des électrodes sélectionnées par sCSP lors des 35 répétitions de découpage.	183
VI.10	Représentation du protocole expérimentale permettant d'amener un sujet vers un état de vigilance cible. La partie encadrée en pointillées correspond à l'interface Cerveau-Ordinateur (ICM). Les parties représentées en rouge et en bleu correspondent aux apports principaux de cette thèse. La partie grisée correspond à l'algorithme génétique réalisé par Mr Legrand pour l'optimisation des paramètres musicaux.	192

Liste des tableaux

IV.1	Moyennes et écarts types des taux de bons classements pour différentes méthodes de classification utilisées sur le critère des pentes.	102
IV.2	Opérateur logique utilisé pour le croisement des fréquences. Pour une composante binaire donnée, lorsque les deux parents ont des valeurs différentes, une loi de Bernoulli de probabilité $\frac{1}{2}$ est utilisée avec afin de déterminer la composante de l'enfant.	106
IV.3	Moyennes et écarts types des taux de bons classements obtenus pour les 100 lancements de l'algorithme génétique et pour les deux méthodes d'évaluation.	114
IV.4	Tableau résumant les deux meilleurs génomes trouvés lors des 100 lancements de l'algorithme génétique avec les deux méthodes d'évaluation.	114
IV.5	Comparaison entre les taux de bons classements obtenus lors de l'étude préliminaire (première ligne) et les taux de bons classements obtenus avec l'algorithme génétique (seconde ligne).	115
V.1	Résultats obtenus en LOO dans l'algorithme génétique pour différentes valeurs de p' . Les résultats entre parenthèses sont obtenus en considérant les sujets au lieu de considérer les essais. La dernière ligne correspond au taux de bons classements référence obtenu en conservant les 58 variables initiales.	141
VI.1	Comparaison entre les taux de bons classements obtenus avec l'algorithme génétique dans le Chapitre V (première ligne) et ceux obtenus avec le sCSP (seconde ligne).	180

Introduction générale

L'époque où seuls une souris et un clavier nous permettaient d'interagir avec un ordinateur semble aujourd'hui révolue. De nombreux et innovants modes d'interactions entre l'Homme et l'ordinateur se sont rapidement développés, se fondant ainsi dans notre vie de tous les jours (sélection tactile d'item, reconnaissance vocale, ...). Au cours des dernières décennies, le développement d'interfaces Homme-Ordinateur (IHO) s'est imposé comme un domaine de recherche porteur, attirant une communauté de chercheurs toujours plus nombreux. Parmi les interfaces Homme-Ordinateur, les interfaces les plus ambitieuses sont les Interfaces Cerveau-Ordinateur (BCI pour *Brain Computer Interface*). L'objectif de ces interfaces est de permettre à un individu d'interagir avec un système artificiel par le seul biais de son activité électrique cérébrale. D'abord conçues avec une visée médicale (restauration de fonctions motrices [15] [16] ou de fonctions de communication [9] par exemple), ces interfaces ont été adaptées au grand public au moyen d'applications ludiques (visite virtuelle de musée [17], contrôle de jeux vidéos [11] [18], ...). L'approche la plus courante pour concevoir une interface Cerveau-Ordinateur est d'analyser, classer et interpréter des signaux recueillis par électroencéphalographie (EEG) afin d'agir sur un ordinateur. Ces interfaces profitent ainsi du caractère non invasif et portatif de l'électroencéphalographie, de son coût relativement faible et de sa résolution temporelle extrêmement précise.

L'électroencéphalographie désigne la mesure de l'activité cérébrale à l'aide de capteurs (électrodes) positionnés sur le scalp. Depuis sa première utilisation en 1929, l'usage de l'électroencéphalographie s'est largement répandu afin d'établir des relations entre les enregistrements et le fonctionnement cérébral et/ou le comportement des individus. Initialement, l'un des objectifs de ces recherches consistait à caractériser les enregistrements cérébraux propres à certaines

pathologies. C'est ainsi qu'à partir de 1950, l'électroencéphalographie s'est imposée comme un examen incontournable pour le diagnostic de l'épilepsie.

Depuis le début des années 1990, l'analyse des rythmes oscillatoires de l'électrogénèse corticale, un temps délaissée, connaît un regain d'intérêt pour appréhender les états mentaux des individus et, notamment, pour mieux cerner l'état fonctionnel du cerveau et ses corrélations avec le niveau de vigilance ainsi que l'état émotionnel de l'individu dans un contexte donné. Dans ce cadre, de nombreux chercheurs se sont attachés à établir des liens entre les enregistrements électroencéphalographiques et l'état d'un sujet ([19], [20], [21]).

Contexte général du travail

Les travaux réalisés au cours de cette thèse s'inscrivent dans le cadre d'un projet plus large. Ce dernier vise à créer une interface Cerveau-Ordinateur permettant de modifier l'état de vigilance d'un individu à l'aide de stimuli musicaux générés en temps réel. Ces stimuli sont adaptés à l'état de vigilance de l'individu qui doit être déduit automatiquement et en temps réel à partir de l'étude du signal électroencéphalographique recueilli. L'objectif est ainsi de pouvoir amener une personne vers un état psycho-physiologique cible (état de vigilance haute ou basse en fonction de l'objectif fixé) à partir de musiques synthétisées en temps réel et personnalisées.

La partie centrale de cette interface consiste en l'obtention de la mesure de l'état de vigilance. C'est autour de la résolution de ce problème que vont s'articuler les travaux réalisés dans le cadre de cette thèse. Dans ce but, une expérience a été réalisée pour recueillir des données électroencéphalographiques permettant de spécifier ces deux états de vigilance. L'analyse de ce jeu de données va permettre d'en apprécier la qualité prédictive et de construire un modèle de prédiction de l'état de vigilance.

Contributions

Les principales contributions proposées dans ces travaux s'articulent autour de deux axes majeurs : l'extraction de caractéristiques dans les signaux électroencéphalographiques et la sélection d'électrodes pour la construction des modèles de classification.

Lorsque le signal électroencéphalographique est utilisé pour concevoir une interface Cerveau-Ordinateur, une étape consiste à extraire des caractéristiques, à partir de ce signal, afin de simplifier la tâche de classification. Dans ce travail, une approche fondée sur le calcul par décomposition en ondelettes discrète des énergies du signal électroencéphalographique dans certaines fréquences est proposée. Dans un premier temps, une régression linéaire est calculée sur les énergies de la bande de fréquences associée à l'apparition des ondes α , connues dans la littérature du domaine pour être spécifiques à l'état de veille passive. Dans un second temps, la plage de fréquences, sur laquelle la régression est calculée, est optimisée par un algorithme génétique. Celui-ci est également utilisé afin de sélectionner une unique électrode. Le couple électrode/plage de fréquences permet ainsi d'obtenir une caractéristique. Le critère optimisé dans l'algorithme génétique est fondé sur une mesure de la qualité de la prédiction obtenue en utilisant une méthode de classification sur les caractéristiques extraites.

Dans la suite des travaux réalisés, la méthode du Common Spatial Pattern (CSP) a été utilisée. Cette méthode permet de créer des signaux synthétiques par combinaison linéaire des variables initiales. L'extraction de caractéristiques est alors réalisée en considérant la variance de ces signaux synthétiques (transformation « log-var »). Dans une première approche, cette méthode d'extraction de caractéristiques est couplée à un algorithme génétique afin de sélectionner un sous-ensemble de variables. Des méthodes de recherche séquentielle sont également proposées afin de sélectionner un groupe de variables.

Une seconde approche est ensuite développée consistant à éliminer directement des variables lors du calcul des filtres spatiaux par la méthode du common spatial pattern en introduisant une contrainte dans son problème d'optimisation. Cette approche a abouti à l'obtention d'une méthode du common spatial pattern dite parcimonieuse.

Plan de la thèse

Le Chapitre I retrace le parcours de l'influx nerveux, de sa naissance à son recueil. Il contient une synthèse bibliographique concernant les méthodes de recueil de l'activité cérébrale. Une attention particulière est portée sur l'électroencéphalographie, méthode centrale de ces travaux de thèse. Ce chapitre introduit également les interfaces Homme-Ordinateur fondée sur l'utilisation du signal électroencéphalographique en présentant une synthèse bibliographique de leurs principales applications.

Le Chapitre II est consacré à la présentation du protocole expérimental utilisé au cours de cette étude afin de recueillir les signaux d'électroencéphalographie. Les principales étapes de

ce protocole ainsi que la procédure de validation des données feront l'objet d'une description détaillée.

Le Chapitre III va permettre de se familiariser avec les concepts d'extraction de caractéristiques et de classification des signaux électroencéphalographiques. Dans une dernière partie, les méthodes de sélection de caractéristiques et/ou d'électrodes ayant été utilisées dans des interfaces Cerveau-Ordinateur fondées sur l'électroencéphalographie sont présentées. Ce chapitre permet de poser les bases des principales méthodes qui seront exploitées au cours de ces travaux. À l'aide de cette synthèse des recherches dans le domaine, les principaux apports de ces travaux, comparativement à ceux qui ont déjà été menés, sont soulignés.

Le Chapitre IV présente une approche d'extraction de caractéristiques qui s'appuie sur une décomposition en ondelettes discrète du signal électroencéphalographique et sur le calcul des énergies. Une régression linéaire des énergies calculées pour les fréquences correspondant aux ondes α est réalisée. Un algorithme génétique est également utilisé afin d'optimiser la sélection d'une électrode et d'une bande de fréquences sur laquelle la régression est réalisée.

Le Chapitre V propose une approche qui vise à sélectionner un sous-ensemble d'électrodes par algorithme génétique. La méthode d'évaluation de cet algorithme génétique est fondée sur l'obtention de signaux synthétiques par common spatial pattern et transformation « log-var » et sur la classification de ces derniers par une analyse linéaire discriminante. Dans ce même chapitre, une autre approche fondée sur la combinaison d'une méthode de recherche séquentielle (de type *sequential forward search* ou *sequential backward search*) avec une évaluation identique à celle utilisée dans l'algorithme génétique est présentée. Les différentes approches développées sont ensuite comparées sur la base de leurs résultats de classification et sur leurs coûts computationnels respectifs.

Le Chapitre VI propose un algorithme de common spatial pattern parcimonieux fondé sur l'utilisation des travaux existants portant sur l'analyse en composantes principales parcimonieuse. Cet algorithme, par le biais d'une pénalisation introduite dans le problème d'optimisation, permet d'annuler le poids de certaines variables et ainsi éliminer ces dernières.

Chapitre I

L'activité cérébrale humaine : genèse, acquisition et utilisation

Sommaire

1	Structure cérébrale	7
2	Genèse de l'activité électromagnétique cérébrale	8
2.1	Les neurones	8
2.2	Structure du neurone	8
2.3	Origine de l'influx nerveux	9
2.4	Transmission de l'influx nerveux	9
3	Les différentes méthodes d'acquisition de l'activité cérébrale . . .	11
3.1	Les méthodes non invasives	11
3.1.1	L'électro-encéphalographie (EEG)	11
3.1.2	La magnétoencéphalographie (MEG)	28
3.1.3	L'imagerie par résonance magnétique (IRM)	28
3.1.4	Tomographie par émission de positons (TEP)	30
3.2	Les méthodes invasives	31
3.2.1	La stéréo-électro-encéphalographie (SEEG)	31
3.3	Classification des méthodes d'acquisition	33
4	Interface Cerveau-Machine	34
4.1	Communication	35
4.2	Locomotion	37
4.3	Restauration de mouvements	37
4.4	Applications ludiques	38
4.4.1	Contrôle de jeux vidéos	38

Chapitre I. L'activité cérébrale humaine : genèse, acquisition et utilisation

4.4.2	Visite virtuelle de musée	40
4.5	Jeux de données des compétitions BCI	40

Le cerveau est l'organe central de notre système nerveux. Il est sans doute l'organe le plus complexe et le plus étudié du corps humain. Il est composé de cellules nerveuses qui communiquent entre elles : les neurones. L'étude de son fonctionnement a donné naissance aux Neurosciences.

1 Structure cérébrale

Le cerveau est composé de deux hémisphères quasi symétriques (droit et gauche) séparés par la fissure inter-hémisphérique (voir Figure I.1(a)). Globalement, l'hémisphère droit contrôle le côté gauche du corps et inversement. Le cortex cérébral est composé, tout comme la moelle épinière ou les noyaux sous corticaux, d'une matière organique appelée substance grise qui constitue la couche externe du cerveau, recouvrant les deux hémisphères cérébraux. Chaque hémisphère est lui-même divisé en quatre parties (voir Figure I.1(b)), appelées lobes cérébraux, impliqués dans différentes fonctions :

- Les lobes frontaux : situés à l'avant du cerveau, ils sont impliqués dans la motricité, la production du langage, le raisonnement (résolution de problèmes, organisation), certains aspects de la personnalité et des fonctions dites « exécutives » (prise de décision, jugement, ...).
- Les lobes pariétaux : localisés derrière les lobes frontaux, ils sont dévolus au contrôle du langage (lecture, écriture, compréhension), au calcul et au traitement des informations sensorielles, notamment tactiles (reconnaissance de textures, formes, tailles, ...).
- Les lobes temporaux : situés sous les lobes frontaux et pariétaux, ils interviennent dans la mémoire (visuelle pour le lobe droit et verbale pour le lobe gauche), le repérage dans l'espace, et les émotions.
- Les lobes occipitaux : localisés à l'arrière du cerveau, ils supportent essentiellement les fonctions visuelles (reconnaissance des formes, couleurs et autres signaux visuels).

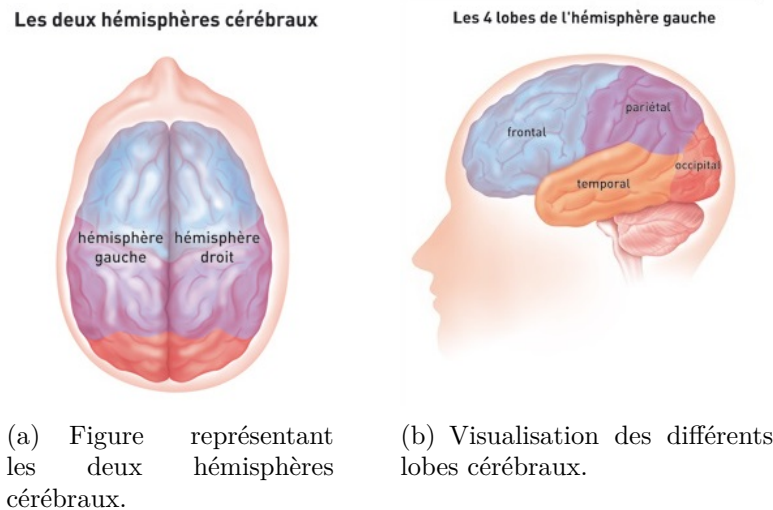


Figure I.1 – Représentation des deux hémisphères du cerveau et des différents lobes qui les composent. D'après [1].

2 Genèse de l'activité électromagnétique cérébrale

2.1 Les neurones

Le cerveau humain est composé d'environ 100 milliards de petites cellules excitables appelées neurones. Le neurone constitue l'unité fonctionnelle élémentaire du système nerveux. Il est en charge du traitement et de la diffusion de l'information. Sa structure particulière lui permet d'acheminer un signal bioélectrique appelé influx nerveux et d'établir des connexions avec d'autres neurones. Ces connexions permettent la transmission du message d'un neurone à l'autre grâce à des messagers chimiques : les neurotransmetteurs.

2.2 Structure du neurone

Le neurone est une cellule nerveuse dont la structure, représentée en Figure I.2, est composée :

- du soma qui est le corps du neurone, contenant le noyau. Il a pour rôle de synthétiser les composants chimiques nécessaires à son fonctionnement ;
- d'un axone qui est un prolongement du soma et qui se termine par des ramifications permettant de lier le neurone à d'autres neurones au moyen d'une jonction particulière appelée synapse ;
- des dendrites qui sont des prolongements généralement plus courts du soma. Ils sont

souvent très nombreux et permettent de récolter un nombre important d'informations provenant d'une multitude d'autres neurones.

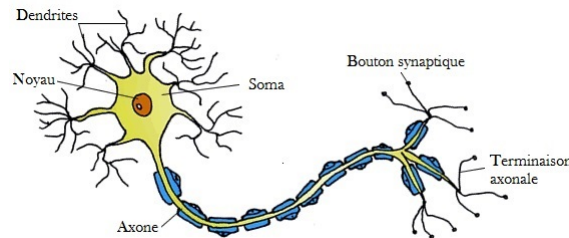


Figure I.2 – Schéma représentant la structure d'un neurone. *Crédits : Chantal Proulx*

2.3 Origine de l'influx nerveux

Lorsqu'une stimulation d'une intensité suffisamment importante intervient à l'intérieur ou à l'extérieur du corps humain, celle-ci est détectée à l'aide de capteurs, constitués de neurones sensitifs, appelés « récepteurs ». Cette stimulation peut être de nature mécanique, chimique ou encore thermique. Le récepteur va alors coder le stimulus détecté sous forme d'un train d'impulsions électriques appelées potentiels d'actions (ou influx nerveux). L'intensité de la stimulation du neurone est codée en fréquence de potentiels d'actions.

2.4 Transmission de l'influx nerveux

Une fois créé, l'influx nerveux est transmis de neurone en neurone. L'influx nerveux a un sens de parcours prédéfini au sein du neurone. Un neurone reçoit les messages en provenance d'autres neurones au niveau de ses dendrites. Un signal généré dans le soma parcourt ensuite l'axone avant d'être transmis à d'autres neurones par le biais des connexions synaptiques. Ainsi, le fonctionnement cérébral repose sur des transmissions d'impulsions électriques qui permettent de véhiculer l'information. Les synapses et leurs messagers biochimiques jouent aussi un rôle essentiel dans la transmission du message nerveux.

La synapse est la zone d'interface entre deux neurones. Chaque neurone possède de 1 à 100 000 (en moyenne 10 000) synapses, lui permettant autant de connexions avec d'autres neurones. Ces zones d'interface sont composées d'un élément pré-synaptique localisé sur l'axone du neurone parcouru par le train de potentiels d'actions (neurone 1), d'une fente synaptique séparant les deux neurones et d'un élément post-synaptique positionné sur un dendrite du

neurone qui va recevoir le message nerveux (neurone 2). Il existe deux types de synapses : les synapses électriques (peu courantes chez l'être humain) et les synapses chimiques. Seul le fonctionnement des synapses chimiques est détaillé ici. Le fonctionnement d'une synapse chimique est représenté en Figure I.3.

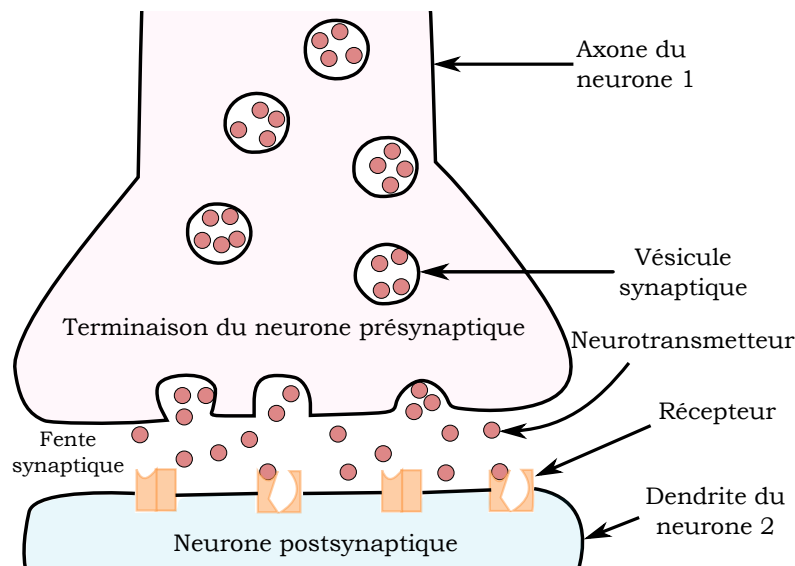


Figure I.3 – Représentation du fonctionnement d'une synapse chimique.

Les éléments pré et post-synaptiques ont la forme de boutons. Lorsque l'influx nerveux arrive au bouton pré-synaptique, il déclenche la libération de molécules chimiques dans la fente synaptique : les neurotransmetteurs. La fréquence des potentiels d'action est alors codée en concentration de neurotransmetteurs. Ce neurotransmetteur traverse la fente synaptique et est alors détecté par l'élément post-synaptique à l'aide de récepteurs membranaires. Selon la nature du neurotransmetteur pré-synaptique, le neurone 2 va alors être soit excité, soit inhibé. S'il est excité, il va déclencher à son tour un train de potentiels d'actions identique à celui provenant du neurone 1. S'il est inhibé, il va produire un train de potentiels d'actions identique d'intensité moins importante que celui du neurone 1.

Ainsi, l'activité électromagnétique cérébrale enregistrée en surface est la résultante des communications entre neurones. Afin que celle-ci puisse être décelée au niveau du scalp d'un individu, il faut qu'un grand nombre de neurones soient excités simultanément.

3 Les différentes méthodes d'acquisition de l'activité cérébrale

Il existe de nombreuses méthodes permettant le recueil de l'activité cérébrale qui diffèrent en fonction de leur résolution spatiale, temporelle, de leur coût ou encore de leur degré d'invasivité.

Les principales méthodes d'acquisition qui vont être présentées seront classées en méthodes non invasives ou méthodes invasives.

3.1 Les méthodes non invasives

Ces méthodes d'acquisition sont privilégiées pour le recueil de l'activité cérébrale. Elles sont indolores et n'entraînent tout au plus qu'une gêne provisoire du sujet durant le recueil. Dans les prochaines sections, l'électro-encéphalographie, la magnéto-encéphalographie, l'imagerie par résonance magnétique et la tomographie par émission de positons vont être présentées. Parmi ces méthodes, que nous allons présenter ci-après, l'électro-encéphalographie fera l'objet d'un développement plus important car c'est la méthode d'acquisition qui a été utilisée pour recueillir l'activité cérébrale dans le cadre de ce travail.

3.1.1 L'électro-encéphalographie (EEG)

L'électro-encéphalographie est la méthode la plus courante pour le recueil de l'activité électrique cérébrale est l'électro-encéphalographie. En 1929, le physiologiste allemand Hans Berger a été le premier à enregistrer un signal électrique provenant de l'activité cérébrale à partir d'électrodes placées sur le scalp. L'activité électrique cérébrale capturée par EEG est représentée sous forme d'un tracé appelé électroencéphalogramme. Un exemple de tracé, recueilli au cours de ces travaux chez un individu sain, est montré en Figure I.4.

Depuis, l'électroencéphalogramme (EEG) a été largement étudié afin d'établir des liens entre ces enregistrements et le fonctionnement cérébral et/ou le comportement des individus. Un intérêt particulier a rapidement porté sur le fait de pouvoir caractériser les enregistrements cérébraux propres à certaines pathologies. C'est ainsi qu'à partir de 1950, l'électroencéphalographie est devenue un examen incontournable pour le diagnostic de l'épilepsie.

L'EEG permet de capter le champ électrique résultant de l'activité neuronale globale. La principale limite de l'EEG réside dans le fait que des électrodes placées sur le scalp ne permettent pas d'obtenir une très bonne résolution spatiale. En effet, elles ne donnent que peu

3.1.1.1 Positionnement des électrodes

Le signal EEG est recueilli à l'aide d'électrodes positionnées sur le scalp. Suivant les besoins et la précision que l'on souhaite obtenir, le nombre d'électrodes peut varier d'une dizaine d'électrodes à plusieurs centaines. Elles sont placées sur le scalp selon des règles de positionnement répondant à une nomenclature standard. Le système international 10/20 a été proposé par Jasper en 1958 [3] et a couramment été adopté dans les travaux qui ont suivis. Ce système définit les positions de 21 électrodes, standardisation permettant la comparaison simple des travaux fondés sur le recueil de signaux EEG dans différents laboratoires. Au cours du temps, les méthodes d'analyse des signaux EEG se perfectionnant, le nombre d'électrodes utilisées a augmenté et des systèmes de positionnements et de nomenclatures étendus ont vu le jour (système 10/10 [2] pour un maximum de 81 électrodes, système 10/5 [4] pouvant aller jusqu'à 345 électrodes). Une représentation des trois systèmes de positionnements est proposée en Figure I.5.

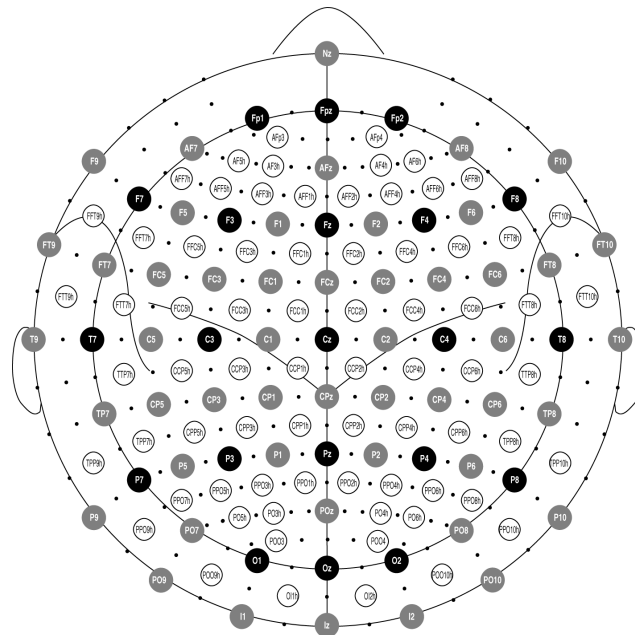


Figure I.5 – Représentation des systèmes internationaux de positionnements d'électrodes pour la capture de signaux EEG. Positions des électrodes dans les systèmes respectifs 10/20 (cercles noirs) [3] ; 10/10 (ajout des cercles gris) [2] et 10/5 (ajout des cercles blancs) [4]. D'après [4].

Dans ces systèmes, la nomenclature des électrodes est composée d'une ou plusieurs lettres et d'un nombre. Les lettres correspondent à la zone corticale au dessus de laquelle l'électrode est positionnée (F, T, C, P et O désignant respectivement les zones Frontale, Temporale, Centrale, Pariétale et Occipitale) et les chiffres en indice caractérisent plus finement le positionnement

du centre vers la périphérie (les chiffres pairs désignant les électrodes positionnées au-dessus de l'hémisphère droit et les impairs celles au-dessus de l'hémisphère gauche). Les électrodes portant l'indice 'z' sont positionnées sur la ligne médiane « inion-nasion », du nez à l'arrière de la tête.

3.1.1.2 Artéfacts

Les signaux EEG qui sont enregistrés à la surface du scalp sont relativement faibles (de l'ordre de quelques μV) et sont donc sensibles à des perturbations qui viennent se superposer au signal d'origine. Ces perturbations, appelées artéfacts, peuvent être de deux types : induites par le sujet (physiologiques) ou par l'environnement (extra-physiologiques). Les principaux artéfacts de l'EEG vont être détaillés dans les prochains paragraphes.

Les artéfacts musculaires : Les neurones cérébraux ne sont pas les seules cellules du corps à avoir une activité bioélectrique. En effet, les muscles du corps, lors de leurs contractions, produisent également une activité électrique. La technique permettant d'enregistrer l'activité électrique musculaire est l'électromyographie (EMG). Le type d'électromyogramme le plus connu est l'électrocardiogramme, décrivant l'activité du muscle cardiaque. Les artéfacts musculaires de l'EEG sont d'autant plus importants que l'amplitude du mouvement est grande ou que les muscles impliqués sont proches du scalp. Le signal électrique musculaire étant d'amplitude bien plus importante que le signal électrique cérébral, ces artéfacts couvrent totalement le signal provenant de l'activité neuronale. C'est pourquoi de nombreux travaux ont été consacrés à leur détection et leur élimination automatique du signal EEG impacté [25], [26], [27]. Un exemple d'artéfact musculaire contaminant un signal EEG est représenté en Figure I.6

Les artéfacts oculaires : Il s'agit d'artéfacts EMG particulièrement gênants lors d'acquisition EEG. Ils ont deux causes distinctes : les clignements des paupières et les saccades (mouvements) oculaires. Les clignements perturbent localement le signal par l'apparition d'un pic, surtout visible sur les électrodes préfrontales et frontales. Les saccades oculaires apparaissent quant à elles lorsque le sujet change la position de son regard. Ces mouvements sont plus compliqués à détecter et à éliminer de l'EEG car ils forment des perturbations plus aléatoires (signaux carrés, dérives lentes du signal,...). Afin de se prémunir contre ces artéfacts, il est d'usage d'enregistrer, en complément de l'électroencéphalogramme, l'EMG oculaire, que l'on nomme Électrooculogramme (EOG). Les signaux EOG sont recueillis à l'aide d'électrodes placées sur les orbites, en général en dessous (pour les mouvements verticaux, dont les cligne-

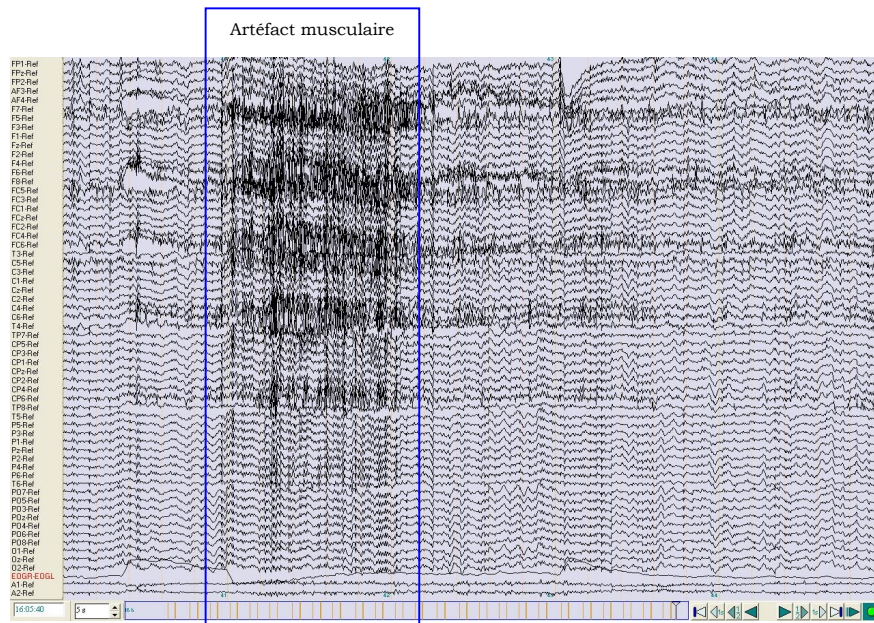


Figure I.6 – Représentation d'un artéfact musculaire contaminant l'électroencéphalogramme d'un sujet sain enregistré à l'aide de 58 électrodes placées selon le système international 10/10. La partie encadrée en bleu correspond au bruit généré par un artéfact musculaire.

ments) et sur le côté extérieur (pour les mouvements horizontaux comme les saccades) de l'œil. On calcule alors la différence des deux signaux obtenus. Lorsqu'aucun artéfact n'intervient, le tracé obtenu est plat. Lorsqu'un clignement intervient, il se traduit par un pic de forte amplitude sur le tracé. Lorsqu'il s'agit d'une saccade oculaire, un signal créneau est obtenu. Un exemple des deux types d'artéfacts oculaires est donné en Figure I.7.

Artéfact secteur : Il s'agit d'un artéfact lié à l'environnement. Il est généré par le courant secteur qui alimente les ordinateurs, haut parleurs, ou autres appareils électriques branchés dans la salle d'enregistrement. Ce courant parasite les enregistrements EEG en introduisant des signaux d'une fréquence fixe (généralement 50 Hz). Il est donc important de protéger les connectiques permettant l'acquisition de données EEG afin d'éviter de récupérer cet artéfact dans les enregistrements EEG. Dans ce but, l'enregistrement se déroule parfois dans un local expérimental où matériel d'acquisition et participant sont placés dans une cage de Faraday [28]. Lorsque les fréquences supérieures à 50 Hz ne présentent pas d'intérêt pour l'étude, ces artéfacts sont facilement éliminés à l'aide d'un filtrage passe bas.

3.1.1.3 Montage

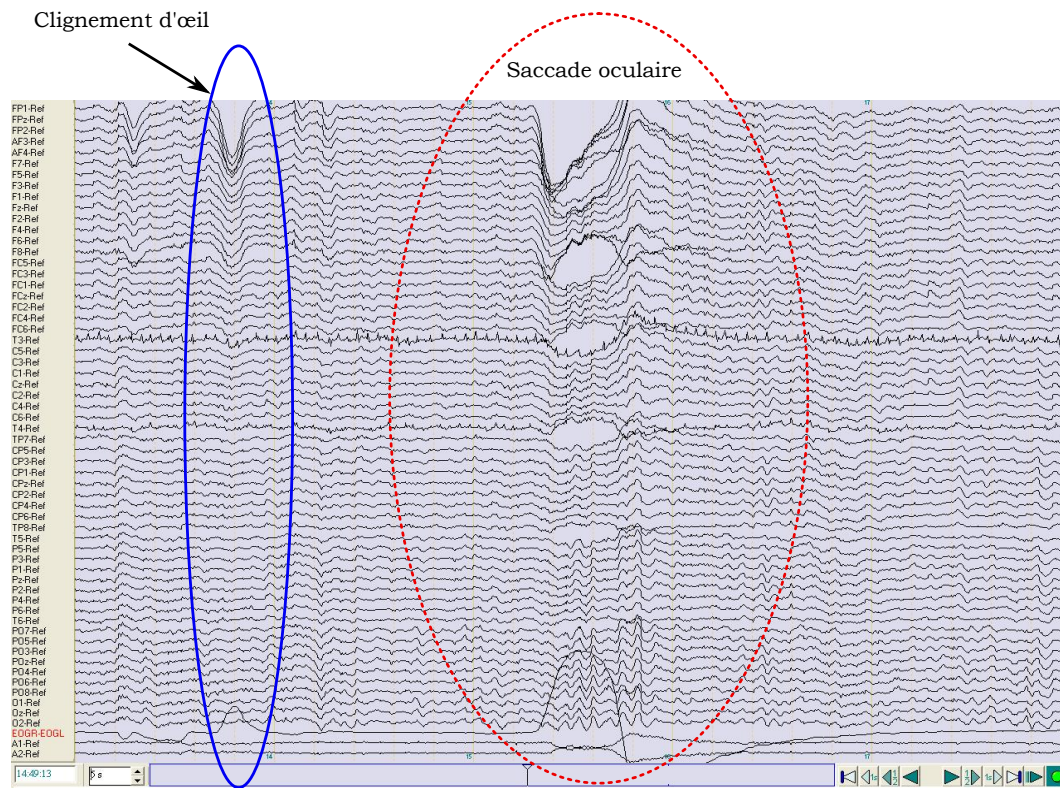


Figure I.7 – Représentation de l'électroencéphalogramme d'un sujet sain enregistré à l'aide de 58 électrodes placées selon le système international 10/10 lors de l'apparition d'un clignement (entouré en trait plein bleu) et d'une saccade oculaire (entourée en trait pointillé rouge). Le troisième tracé en partant du bas correspond au tracé d'EOG.

Le signal EEG est une différence de potentiel électrique (ddp). Une électrode seule ne donnant qu'un potentiel électrique, il faut coupler plusieurs électrodes afin d'obtenir une ddp. La manière de coupler les électrodes est appelée montage. Il existe de nombreux montages tels que le montage référentiel, le bipolaire ou encore le référence moyen. Il existe d'autres types de montage, tels que le montage Laplacien, qui ne seront pas discutés ici car ils sont beaucoup moins couramment utilisés.

Le montage référentiel : Dans ce montage, chaque électrode placée sur le scalp est couplée avec une électrode (ou plusieurs) dite(s) de référence afin d'obtenir une ddp. Les électrodes du scalp sont alors dites actives. Il existe de nombreuses façons de choisir l'électrode de référence. L'idéal est d'en choisir une dont le potentiel est le plus proche de 0 possible afin que la ddp obtenue corresponde uniquement au signal enregistré sous l'électrode active. Ainsi, il faut absolument éviter de choisir une électrode du scalp comme référence car celle-ci recueille un potentiel électrique correspondant à l'activité neuronale. De fait, il est courant de choisir

une électrode placée sur le corps à bonne distance du scalp (appelée alors extra-céphalique) mais cette option présente l'inconvénient de la sensibilité aux émissions électriques issues de l'activité du muscle cardiaque.

Lorsque l'électrode de référence est placée sur la tête du sujet (alors dite céphalique), les emplacements privilégiés sont le bout du menton, le nez, les lobes des deux oreilles ou encore les os mastoïdes. Dans les deux derniers cas, deux électrodes sont utilisées. Le signal de référence est obtenu en réalisant la moyenne des signaux des deux électrodes. L'avantage de ces deux derniers emplacements est que la valeur de référence ainsi calculée simule l'activité d'une électrode de référence virtuelle qui se situerait au centre de la boîte crânienne, donc à équidistance de toutes les autres électrodes du scalp. Ainsi, aucune zone cérébrale n'est privilégiée par rapport aux autres. De plus, les os mastoïdes, situés derrière les oreilles, sont des emplacements pertinents d'une électrode de référence en raison de leur épaisseur importante qui permet de stopper la grande majorité de l'activité électrique cérébrale.

Le montage référentiel moyen : Ce montage est un dérivé du montage référentiel précédemment présenté. Dans celui-ci, la référence correspond à la moyenne des potentiels enregistrés par toutes les électrodes actives du scalp.

Le montage bipolaire : Le montage bipolaire est un montage dans lequel chaque électrode est couplée à une électrode adjacente, permettant ainsi de calculer une ddp. Plusieurs montages bipolaires existent (longitudinal, transverse, circonférentiel) ayant chacun son propre sens de parcours du scalp. Par exemple, le montage longitudinal couple les électrodes deux à deux, d'avant en arrière, tandis que le montage transverse couple les électrodes de droite à gauche.

Enfin, quel que soit le montage choisi, le système d'acquisition nécessite l'utilisation d'une électrode de masse (terre). Il est courant d'utiliser une électrode du scalp afin de remplir ce rôle (généralement FCz).

3.1.1.4 Matériel de recueil de l'EEG

Le système d'acquisition de données EEG suppose le recours à des électrodes, un boîtier de connexion appelé boîte têtère, un amplificateur, un convertisseur analogique/numérique et un ordinateur. Les électrodes sont généralement placées sur un bonnet élastique permettant de garantir le respect du système de positionnement des électrodes quelles que soient la conforma-

tion et la taille du crâne de l'individu enregistré. Les électrodes sont le plus souvent en argent chloruré (AgCl). Les électrodes complémentaires de celles placées sur le scalp (référence, EOG) sont généralement des électrodes cupules maintenues sur la peau à l'aide de sparadrap (voir Figure I.8(a)). Il convient d'avoir le meilleur contact possible entre l'électrode et l'épiderme. Dans le cas contraire, l'électrode va enregistrer le signal électrique ambiant et non plus l'activité cérébrale générée par les neurones situés sous elle. Afin de faciliter la capture des signaux électriques cérébraux, un gel électrolytique très conducteur est utilisé. Le gel facilite le contact entre l'électrode et la peau. La qualité de ce contact est caractérisée par son impédance, c'est-à-dire sa résistance (l'impédance est une résistance qui, de surcroît, déphase le signal). La mesure de l'impédance est réalisée par rapport à l'électrode de terre. On considère qu'en deçà de 5 kOhms, la valeur d'impédance est acceptable. Le câblage des électrodes sort du bonnet par le biais de deux nappes terminées par des connecteurs à broches permettant de les relier à la boîte têtère. Une photographie d'un casque à électrode branché à la boîte têtère est proposée en Figure I.8(b). La boîte têtère fait office de boîtier d'amplification (amplificateurs miniaturisés *Brainbox* dans notre système d'acquisition) qui permet d'augmenter l'amplitude des signaux recueillis.

Anciennement, chaque électrode était connectée à un amplificateur propre, lui-même relié à une plume, dont le mouvement était proportionnel à l'amplitude du signal reçu. Ce système permettait d'inscrire le signal EEG recueilli sur des feuilles enroulées sur un rouleau en rotatif. Ces systèmes à plume ont été progressivement remplacés par des systèmes informatiques. Aujourd'hui, l'amplificateur est relié à un convertisseur analogique/numérique qui traduit le signal électrique en un signal numérisé (série de bits). Une photographie d'un convertisseur analogique est donnée en Figure I.8(c). Enfin, le convertisseur est relié à un ordinateur qui, au moyen d'un logiciel dédié, permet la visualisation des tracés et le stockage du signal. Enfin, il faut noter que pour garantir qu'aucun courant ne revienne vers le scalp du sujet, un boîtier jouant le rôle de disjoncteur est utilisé (voir Figure I.8(d)).

3.1.1.5 Les rythmes cérébraux

L'activité synchrone de neurones, à des échelles spatiales et temporelles variées, a pour effet l'émergence de rythmes cérébraux couvrant un large spectre de fréquence. Ces rythmes cérébraux correspondent à des oscillations électromagnétiques dans des bandes de fréquences données. Chaque bande de fréquence peut caractériser l'état du sujet. On distingue quatre principaux rythmes cérébraux :

- Les ondes δ (delta) incluant les fréquences entre 1 et $3.5Hz$.

I.3 Les différentes méthodes d'acquisition de l'activité cérébrale



(a) Photographie d'une électrode cupule.



(b) Photographie d'un casque à 60 électrodes. Chaque plot blanc abrite une électrode. Le casque est relié à la boîte tête.



(c) Photographie d'un convertisseur analogique. Son rôle est de convertir le signal électrique enregistré en un signal numérique.



(d) Représentation d'un boîtier disjoncteur.

Figure I.8 – Représentation de l'équipement usuel pour l'enregistrement de signaux en EEG.

- Les ondes θ (thêta) ayant des fréquences comprises entre 4 et $8Hz$.
- Les ondes α (alpha), dont les fréquences sont comprises entre 8 et $12Hz$.
- Les ondes β (bêta) qui ont une très basse amplitude et une haute fréquence. Elles varient entre 19 et $26Hz$.

Une représentation de ces rythmes cérébraux est donnée en Figure I.9. À noter que l'on parle

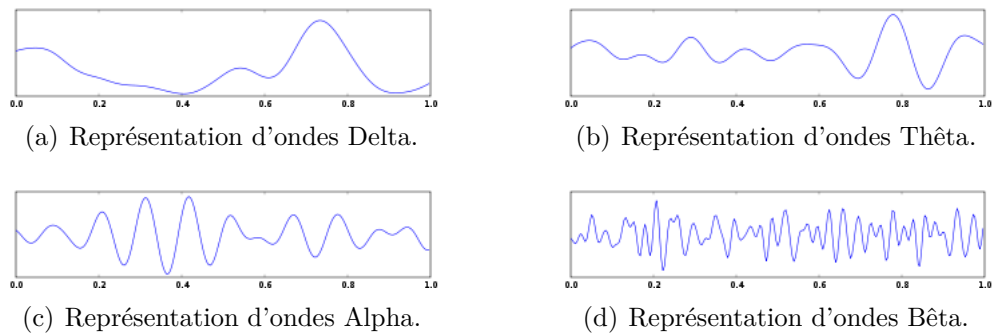


Figure I.9 – Principaux rythmes oscillatoires de l'EEG humain. D'après H. Gamboa.

également parfois de rythmes gamma. Ils ont une fréquence comprise entre $30Hz$ et $40Hz$. Ils accompagneraient les activités cognitives intenses et seraient notamment caractéristiques du sommeil dit paradoxal. Il y a toutefois une polémique au sujet de la pertinence de la distinction entre oscillations de haute fréquence de type bêta et gamma.

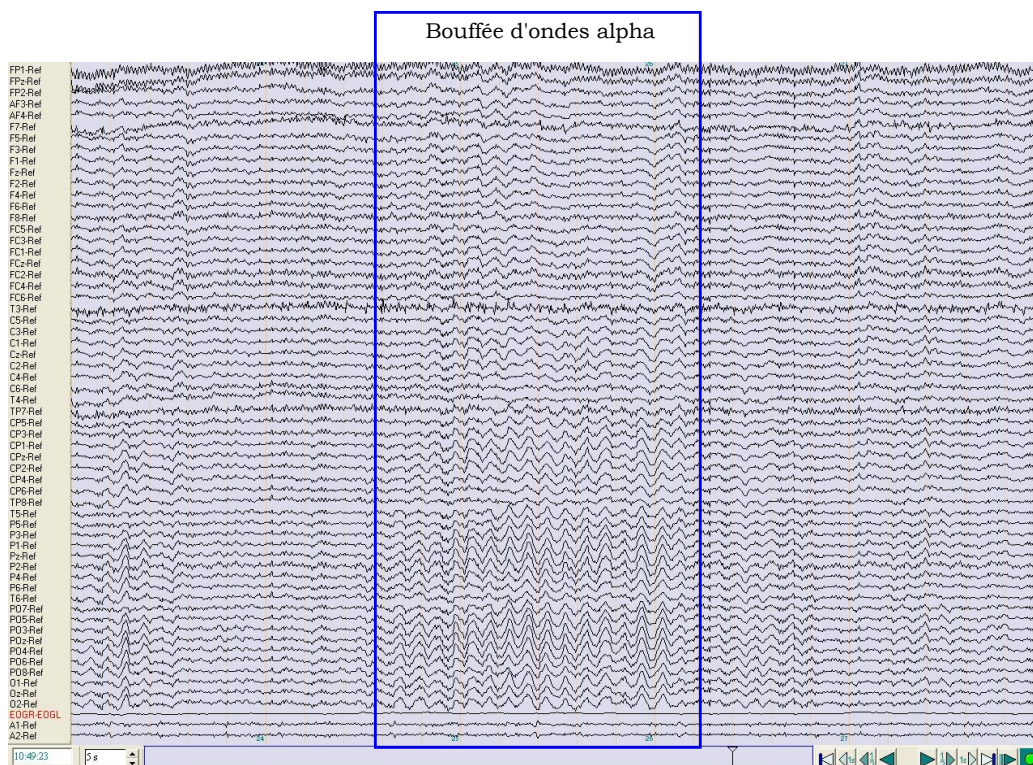


Figure I.10 – Électroencéphalogramme d'un sujet sain dans un bas niveau de vigilance enregistré à l'aide de 58 électrodes placées selon le système international 10/10. La partie encadrée correspond à une bouffée d'ondes α .

3.1.1.6 Les états non altérés de conscience

La nature des rythmes cérébraux est intimement liée à l'état de conscience du sujet. Cette partie va permettre de définir les principaux niveaux de conscience et les rythmes cérébraux correspondants. Dans ce travail, seuls les états de conscience d'un sujet sain seront présentés.

La conscience peut être définie comme la capacité d'un être humain à s'auto-percevoir et à percevoir son environnement. Les changements d'état de conscience rythment la vie de l'être humain. Les trois principaux états de conscience sont l'état d'éveil, le sommeil lent et le sommeil paradoxal. Chacun de ces états est lié à l'apparition de rythmes cérébraux qui lui sont propres.

L'état d'éveil : Il est caractérisé par un maintien de l'activation cérébrale permettant un traitement rapide et correct de l'information. On distingue deux types d'éveil : l'éveil actif et l'éveil passif. Lors d'une période d'éveil, l'être humain oscille de manière plus ou moins rapide entre ces deux types d'état. L'éveil actif correspond à un état de vigilance élevé. Au cours de cet état, l'émission d'ondes β est importante. L'état passif correspond à un état de vigilance plus bas (sujet inactif, détendu, relaxé). Il s'accompagne d'une émission d'ondes α . L'apparition de ces ondes a tendance à être plus marquée lorsque le sujet a les yeux fermés. Leur grande amplitude, comparée au tracé EEG usuel, permet de les repérer facilement à l'œil nu sur les tracés EEG. Elles apparaissent sous forme de bouffées. Un exemple de bouffée d'ondes α est donné en Figure I.10. Le rythme α est surtout retrouvé dans les régions occipitales du cerveau, et à moindre échelle dans les zones antérieures.

Le sommeil : Cet état se traduit par une conscience réduite, une relative immobilité corporelle et une perception très amoindrie des stimuli usuels. Cet état diffère du coma notamment par son côté réversible. On distingue plusieurs phases dans le sommeil : le sommeil lent (stades I,II,III et IV) et le sommeil paradoxal.

Le sommeil lent : Il doit son nom à un ralentissement généralisé des ondes cérébrales. Il est caractérisé par 4 stades distincts allant du sommeil le plus léger (stade I) au sommeil le plus profond (stade IV). Le stade I correspond à l'état de somnolence. Ce dernier est l'étape de transition entre l'éveil et le sommeil, soit l'endormissement. Cet état est caractérisé par une forte baisse de la vigilance. Le sommeil lent de stade II correspond au sommeil léger. Durant ce stade, le sujet peut encore être réveillé par des stimuli d'intensité modérée. Au cours de ces

deux premiers stades apparaissent les ondes θ .

Les deux derniers stades du sommeil lent (III et IV) correspondent au sommeil profond. Ils sont caractérisés par l'apparition d'ondes δ .

Le sommeil paradoxal : C'est durant cette période qu'apparaissent les rêves (on parle de sommeil du rêve mais aussi de sommeil REM pour *Rapid Eye Movements*). Durant le sommeil paradoxal, l'EEG d'un sujet endormi est presque indistinguishable de l'EEG d'un sujet éveillé et actif.

3.1.1.7 Classification des états de conscience

Une classification des différents états de conscience en fonction de l'état de vigilance a été proposée dans [5]. Présentée sous forme de graphique, elle permet de résumer les précédents paragraphes. Cette classification est reprise et adaptée en Figure I.11.

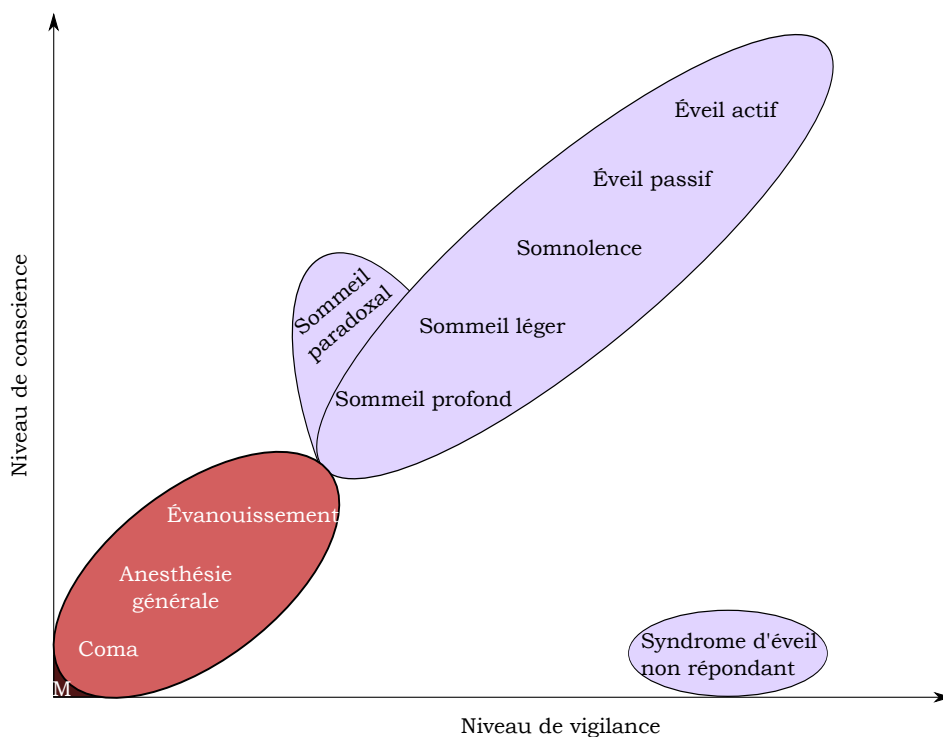


Figure I.11 – Classification des différents états de conscience en fonction du niveau de vigilance. Les états de conscience normaux apparaissent en bleu tandis que les états de conscience altérés apparaissent en rouge. Le rouge foncé correspond à l'état de mort cérébrale (noté M). Ce schéma est adapté à partir de celui proposé dans [5].

3.1.1.8 Intérêt de l'étude des rythmes cérébraux

À la lecture de ce qui vient d'être présenté, l'intérêt de l'étude des rythmes cérébraux paraît évidente. Il s'agit de distinguer dans quel état de conscience se trouve un sujet. Cela peut permettre de repérer et/ou d'identifier certains troubles du sommeil (hypersomnie, insomnie, apnée du sommeil) et donc de les traiter.

De plus, la distinction entre certains états altérés de conscience peut être délicate à établir lors d'une simple consultation du sujet. L'EEG apparaît donc, au travers de l'étude des rythmes cérébraux, comme une méthode complémentaire pour diagnostiquer un état de conscience altéré chez un sujet.

3.1.1.9 Les potentiels évoqués

L'étude des potentiels évoqués est une méthode fondée sur l'analyse, au moyen de l'EEG, de la réponse cérébrale d'un sujet à un stimulus qui lui a été présenté. Un potentiel évoqué est un changement bref de l'activité électrique cérébrale qui apparaît en réponse à un stimulus extérieur (stimulus visuel, olfactif, sonore,...) ou interne (préparation d'une action cognitive ou motrice). Les potentiels évoqués ont la particularité d'être synchronisés sur l'apparition du stimulus déclencheur. Ils apparaissent sous forme d'un pic positif ou négatif, dans une fenêtre de temps couvrant plusieurs centaines de millisecondes (ms) après le stimulus. On distingue plusieurs types de potentiels évoqués, identifiés grâce à leur temps de latence ainsi qu'au type de stimulus qui les suscitent. La nomenclature des composantes constitutives des potentiels évoqués est généralement la suivante : une lettre (P pour les composantes positives et N pour les négatives) suivie d'un nombre correspondant à la latence approximative (exprimée en ms). Quelques exemples de composantes couramment observées sont donnés dans les prochains paragraphes. La variation contingente négative, composante exploitée au cours de cette thèse, sera détaillée de manière plus approfondie.

La composante N100 : La N100 est un pic négatif qui apparaît lorsqu'un stimulus (visuel, sonore, olfactif,...) est présenté au sujet sans qu'une réponse à ce stimulus ne lui soit demandée. Comme son nom l'indique, la N100 apparaît environ 100 ms après la présentation du stimulus. La N100 est un potentiel exogène, c'est-à-dire très précoce et fortement lié à la nature physique du stimulus (son intensité, sa modalité sensorielle). On considère généralement qu'elle correspond au traitement de l'information par les aires sensorielles. L'apparition de la

N100 est couplée à celle de la P200 dans le cas d'un stimulus auditif. On parle alors de complexe N1-P2 (où N1 et P2 sont les raccourcis pour N100 et P200).

La composante P200 : Cette composante, lorsqu'elle est provoquée par un stimulus auditif, traduit la réaction cérébrale de traitement du stimulus, le cerveau cherchant à associer le stimulus qui vient d'être perçu à un stimulus qu'il a déjà traité par le passé. Il semble donc que la P200 soit liée à l'accès à la mémoire. L'amplitude maximale de la P200 est généralement enregistrée au dessus de la zone centrale (électrodes commençant par C sur la Figure I.5).

La composante P300 : Elle apparaît lorsque le sujet a détecté un stimulus inattendu ou inhabituel. Le procédé le plus courant pour la faire apparaître est appelé le paradigme du stimulus discordant [29] (*Oddball* en anglais). Cela consiste à présenter au sujet une série de stimuli composée de deux types de stimuli différents. L'un est le stimulus cible que le sujet doit détecter (appui sur un bouton, comptage, ...) et l'autre joue le rôle de stimulus distracteur. Dans la série présentée, le stimulus distracteur apparaît de manière très fréquente tandis que le stimulus cible n'apparaît que rarement. Cette composante est donc liée à la mémoire, l'évaluation, la catégorisation et la prise de décision.

La composante N400 : La N400 est liée au traitement linguistique et sémantique. Le procédé le plus courant pour la susciter est de faire lire au sujet des phrases contenant un mot sémantiquement incongru (par exemple : « tous les matins, il boit un *réverbère* »). La N400 est principalement visible sur les électrodes centro-pariétales (électrodes commençant par CP sur la Figure I.5).

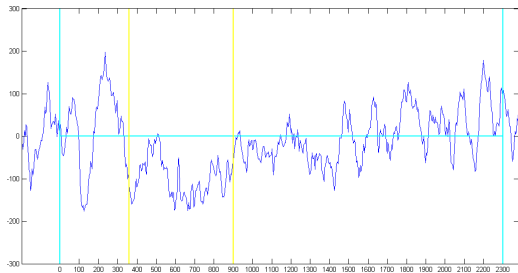
La Variation Contingente Négative (VCN) : La VCN [30] est une composante du potentiel évoqué un peu différente de celles qui viennent d'être présentées, d'abord par sa nomenclature puis par son paradigme expérimental. Son appellation est liée à la forme de ce potentiel, apparaissant comme un grand plateau négatif. Le procédé expérimental pour son recueil nécessite non pas un stimulus comme pour les précédents, mais deux. Le premier stimulus présenté au sujet est nommé « signal avertisseur » et noté S1. Il peut s'agir d'un stimulus visuel ou sonore. Ce stimulus prévient le sujet de l'imminence d'un second stimulus, dit « signal impératif » et noté S2. À l'apparition de S2, le sujet a pour consigne de fournir une réponse le plus vite possible (en général, il s'agit d'appuyer sur un bouton). La VCN traduit donc la façon dont le cerveau se prépare à effectuer la tâche. La latence de ce potentiel varie entre 250 et 450 ms selon les conditions expérimentales. La VCN est composée d'au moins 2

I.3 Les différentes méthodes d'acquisition de l'activité cérébrale

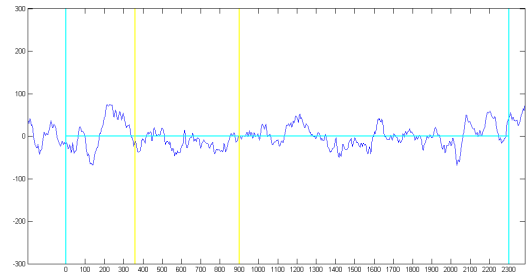
sous-composantes : la VCN précoce et la VCN tardive. La VCN précoce, plutôt fronto-centrale (FC sur le graphique II.3), est associée à un processus d'orientation par rapport au stimulus. La VCN tardive, centro-pariétale (CP sur le graphique II.3), semble, elle, traduire l'attente et la préparation de la réponse à S2.

Recueil et mise en évidence des potentiels évoqués : Malgré la synchronie des potentiels évoqués avec le stimulus déclencheur, ces derniers ne sont pas simples à détecter. En effet, les potentiels évoqués ne sont pas visibles à l'échelle d'un seul tracé EEG. Ces derniers sont fondus au milieu des autres activités électriques cérébrales concomitantes. Ainsi, pour les isoler du bruit de fond du cerveau, la méthode consiste à réaliser une moyenne sur plusieurs dizaines d'essais [31]. Il est donc nécessaire de répéter l'enregistrement du signal EEG après la présentation répétée du stimulus ou de stimuli jugés équivalents au regard des processus de traitement étudiés, afin de pouvoir extraire ensuite le potentiel évoqué recherché. Pour cela, les signaux EEG recueillis vont être superposés en prenant le stimulus déclencheur comme instant initial ($t = 0$) et une courbe moyenne de tous ces tracés va être calculée. Les signaux cérébraux reflétant l'ensemble du fonctionnement cérébral hormis celui consacré au traitement des stimuli qui nous intéressent vont être aléatoires car ils ne sont pas seront pas « déclenchés » par ces stimuli, donc pas synchrones sur leurs instants d'apparition. Leur moyenne sur l'ensemble des époques d'acquisitions toutes synchronisées sur les instants de présentation des stimuli tendra donc vers 0. Cette activité cérébrale non liée au traitement des stimuli à l'étude va donc disparaître du tracé moyen. À l'inverse, les potentiels évoqués, synchronisés sur l'apparition du stimulus déclencheur, vont se sommer et croître en amplitude au fur et à mesure des répétitions, jusqu'à ce qu'il ne reste plus qu'eux. On considère qu'il faut au minimum une trentaine d'essais pour arriver à un « filtrage » suffisant du bruit constitué par l'activité générale corticale pour pouvoir voir apparaître les composantes spécifiques des potentiels évoqués par les stimuli à l'étude. Une représentation de l'influence du nombre d'essais sur l'émergence des potentiels évoqués est donnée en Figure I.12. Le protocole expérimental utilisé pour le recueil du tracé EEG moyen présenté en Figure I.12 correspond à celui décrit en Section 3.1.1.9 pour l'obtention de la VCN. Le signal avertisseur est un stimulus sonore et le signal impératif est un stimulus visuel. L'apparition des deux stimuli est représentée par des lignes verticales bleues. Les Figures I.12(a), I.12(b), I.12(c), I.12(d), I.12(e) et I.12(f) représentent les tracés moyens obtenus respectivement à partir de 1, 5, 10, 20, 30 et 50 essais. Il apparaît que les tracés deviennent de plus en plus « lisses » au fur et à mesure que le nombre d'essais utilisés pour le calcul du tracé moyen augmente. Grâce à ce lissage, trois potentiels évoqués émergent nettement : la N100, la P200 et la VCN (encadrée par les lignes verticales jaunes). Il est important que les tracés

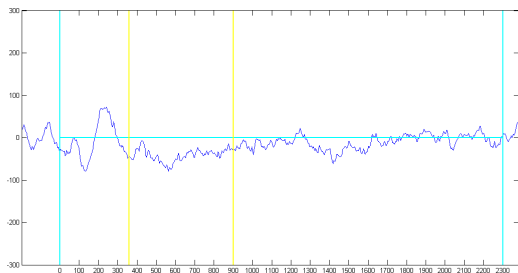
EEG impliqués dans le calcul du tracé moyen ne présentent aucun artefact (oculaire, musculaire ou secteur). En effet, un artefact, du fait de son amplitude, pourrait masquer l'émergence des potentiels évoqués.



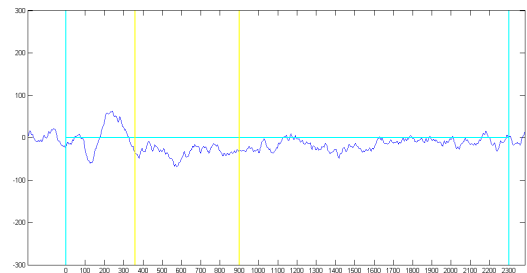
(a) Tracé d'un essai VCN.



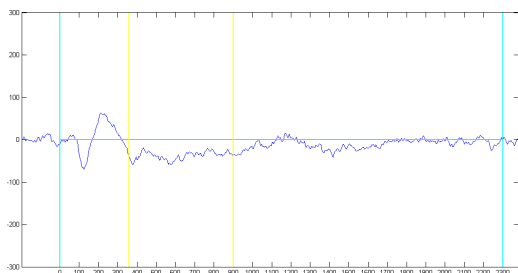
(b) Tracé de la moyenne de 5 essais VCN.



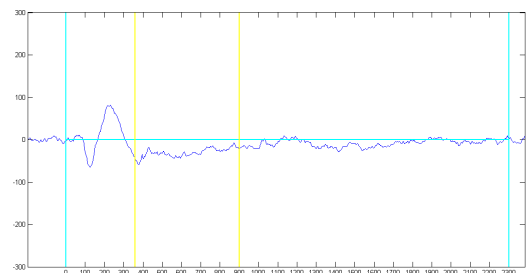
(c) Tracé de la moyenne de 10 essais VCN.



(d) Tracé de la moyenne de 20 essais VCN.



(e) Tracé de la moyenne de 30 essais VCN.



(f) Tracé de la moyenne de 50 essais VCN.

Figure I.12 – Représentation de l'apparition des potentiels évoqués par la technique du moyennage.

La VCN comme indicateur de l'état de vigilance : La VCN a la propriété de diminuer en amplitude quand le sujet est moins vigilant, que cela soit parce qu'il est distrait [32], privé de sommeil [33] ou en train de s'endormir [34]. Ce résultat fondamental est illustré par la figure I.13. Sur cette figure, la VCN d'un sujet dans un état de vigilance normale (état

I.3 Les différentes méthodes d'acquisition de l'activité cérébrale

de veille active) est représentée par une ligne en pointillé et celle d'un sujet dans un bas niveau de vigilance (veille passive ou somnolence) est en trait plein. On peut ainsi constater que l'amplitude de la VCN est proportionnelle à la vigilance du sujet. C'est pourquoi nous avons choisi cette VCN comme marqueur électrophysiologique pertinent de l'état de vigilance d'un sujet. En comparant les amplitudes des VCN d'un même sujet placé dans différentes conditions expérimentales, il devrait en effet être possible de déduire l'effet de ces conditions sur le niveau de vigilance du sujet. L'utilisation précise de ce marqueur de la vigilance qu'est la VCN dans le cadre de notre étude sera discutée dans le Chapitre II.

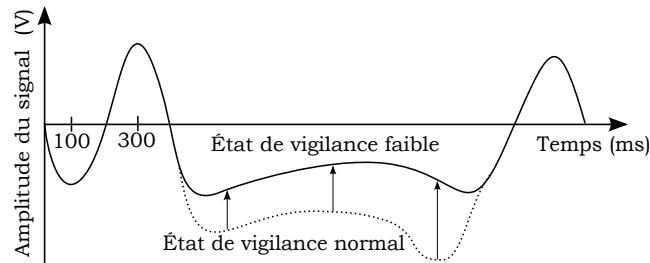


Figure I.13 – Représentation de la variation d'amplitude de la VCN en fonction de l'état de vigilance du sujet.

Remarque : le paradigme expérimental utilisé pour recueillir la composante VCN peut entraîner un léger changement de l'état de vigilance du sujet (diffusion de signaux sonores ou visuels, sollicitation de mouvements). Cependant, ce protocole d'acquisition, couramment utilisé dans la littérature ([32], [33] ou encore [34]), a permis de mettre en évidence un lien entre une baisse de la vigilance et une diminution en amplitude de la VCN. Ainsi, ces perturbations ne semblent pas être trop préjudiciables.

Intérêt des potentiels évoqués : Comme cela vient d'être évoqué avec la VCN, l'étude des potentiels évoqués, à l'instar de l'étude des rythmes cérébraux, peut permettre de déterminer l'état de conscience d'un sujet. En particulier, elle peut être utile pour établir un pronostic de réveil chez des sujets comateux. Les potentiels évoqués sont également très utilisés pour vérifier l'intégrité du système nerveux d'un sujet. En particulier, ils permettent de détecter la présence de lésions cérébrales chez le patient, d'en préciser l'ampleur ainsi que d'en apprécier les répercussions sur le fonctionnement cognitif du patient. Cette méthode est complémentaire des méthodes d'imagerie cérébrale qui vont être présentées.

3.1.2 La magnétoencéphalographie (MEG)

La MEG est une méthode d'imagerie cérébrale, initiée par Cohen en 1958 [35], visant à enregistrer les champs magnétiques émis par l'activité cérébrale. Ces champs cérébraux naissent des déplacements d'ions générés par l'activité électrique neuronale. La complexité de l'enregistrement vient du fait que les champs magnétiques, que l'on cherche à mesurer, sont extrêmement faibles en comparaison du champ électrique terrestre (environ 10 milliards de fois plus faibles). Ainsi, durant l'enregistrement, la tête du patient est placée à l'entrée d'un cylindre imposant (voir Figure I.14(a)). L'entrée de ce cylindre est équipée de capteurs (de 50 à 306 capteurs pour les machines les plus élaborées) spéciaux appelés SQUIDS (Superconducting Quantum Interference Devices) permettant d'enregistrer le champ magnétique cérébral (voir Figure I.14(b)). Le cylindre est rempli d'hélium afin de conserver en permanence les capteurs immergés dans le liquide. Cette température permet aux capteurs SQUIDS de préserver leur caractère supraconducteur et de maintenir un très faible bruit électronique. Contrairement à l'EEG, les capteurs ne sont pas en contact avec le sujet mais seulement disposés autour de sa tête. Cela fait de la MEG la moins invasive des techniques d'acquisition de l'activité cérébrale. Les champs magnétiques n'étant quasiment pas déformés par leur passage au travers des tissus organiques, la MEG permet d'obtenir des signaux moins bruités que ceux issus de l'EEG. La MEG dispose également d'une très bonne résolution temporelle. Malgré tous ces points positifs, cette méthode reste peu répandue en raison de son coût très élevé (environ 2 millions d'euros pour l'achat du système d'acquisition et environ 80 000 euros par an pour l'entretien).

3.1.3 L'imagerie par résonance magnétique (IRM)

Il existe trois types d'imageries par résonance magnétique : l'IRM anatomique (IRMa), l'IRM fonctionnelle (IRMf) et l'IRM de diffusion (IRMd). Ces trois techniques vont être détaillées dans les prochains paragraphes.

L'IRM anatomique : Il s'agit d'une méthode d'imagerie médicale permettant de visualiser à l'aide d'images en 2D ou 3D des parties du corps d'un patient. L'IRM exploite une caractéristique structurale du corps humain : 63% des atomes qui le composent sont des atomes d'hydrogène [36]. L'hydrogène présente l'intérêt d'avoir des propriétés magnétiques. Ainsi, le principe de l'IRM est le suivant : le patient est placé dans un cylindre ayant la forme d'un tunnel dans lequel est généré un champ magnétique. Sous l'action de ce champ magnétique, tous les atomes d'hydrogène vont s'orienter dans la même direction. Des ondes radios sont alors émises par une antenne afin de perturber l'orientation des atomes d'hydrogène et ainsi

I.3 Les différentes méthodes d'acquisition de l'activité cérébrale



(a) Représentation d'un appareil de mesure utilisé en MEG. *Crédits : NIMH Image library.*



(b) Représentation des capteurs installés à l'intérieur du cylindre durant les enregistrements de MEG.

Figure I.14 – Représentation de l'appareil de mesure utilisé en MEG ainsi que des capteurs. *Crédits : Inserm-CERMEP.*

les exciter. Lorsque cette stimulation cesse, les atomes d'hydrogène vont retrouver leur position d'origine tout en restituant l'énergie emmagasinée. Ce processus va entraîner l'émission de signaux électromagnétiques qui vont permettre d'obtenir une image. L'hydrogène étant principalement présent dans l'eau et les graisses du corps humain, l'image va en quelque sorte donner un aperçu de la distribution anatomique d'eau et de graisse dans l'organe observé.

L'IRM est notamment utilisée lors du diagnostic des maladies de Parkinson et d'Alzheimer. Même si le diagnostic de ces maladies est avant tout clinique, il existe plusieurs pathologies qui peuvent avoir des symptômes similaires à chacune de ces deux maladies. L'IRM est donc utilisée comme un examen complémentaire afin d'éliminer une à une ces pathologies et donc confirmer la présence de la maladie de Parkinson ou de la maladie d'Alzheimer.

L'IRM fonctionnelle : L'IRM fonctionnelle (IRMf) est une méthode dérivée de l'IRM anatomique. Elle est dite fonctionnelle car elle permet d'obtenir des images qui mettent en évidence le fonctionnement du cerveau. L'IRMf est fondée sur le fait que lorsque les neurones d'une région cérébrale sont activés, ils consomment davantage d'oxygène. Le débit sanguin va alors augmenter dans ces régions afin de fournir l'oxygène nécessaire. Contrairement à l'IRM anatomique, qui utilise les propriétés de l'hydrogène, le fonctionnement de l'IRMf est fondé sur les propriétés de l'hémoglobine. L'hémoglobine est une protéine, située à l'intérieur des globules

rouges sanguins, capable de se lier aux atomes d'oxygène. De cette façon, l'hémoglobine est impliquée dans le transport de l'oxygène au sein du corps. Elle a des propriétés magnétiques légèrement différentes selon qu'elle est liée ou non à l'oxygène. Ainsi, l'IRMf permet d'obtenir une cartographie 3D des zones irriguées (activées) dans le cerveau.

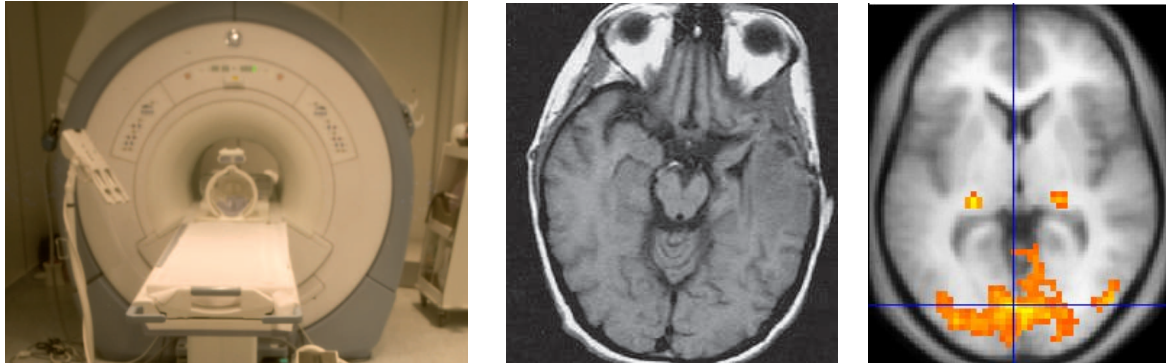
L'IRM de diffusion : L'IRM de diffusion (IRMd) est une autre méthode dérivée de l'IRMa. Elle s'appuie sur la détection des mouvements (considérés comme aléatoires et modélisés par un mouvement brownien) des molécules d'eau. Ces molécules d'eau ont des mouvements perturbés par les obstacles qu'elles rencontrent. Ainsi, les images recueillies permettent de mettre en évidence la position et la nature des obstacles rencontrés (fibres, membranes cellulaires, macromolécules, ...).

Alors que la MEG est parfois combinée avec l'EEG, l'IRM ne peut pas l'être du fait de l'incompatibilité de l'utilisation concomitante des différents capteurs. Il existe cependant des méthodes tentant de fusionner des images IRMf et des données EEG recueillies séparément, mais leur recours est limité du fait d'un travail long, coûteux et malgré tout entaché de faiblesses méthodologiques.

3.1.4 Tomographie par émission de positons (TEP)

La tomographie par émission de positons est une technique d'imagerie médicale qui permet de visualiser le fonctionnement d'un organe. Le principe est le suivant : un traceur, marqué par un élément radioactif à courte durée de vie, est injecté par intraveineuse au patient. Le traceur est choisi en fonction de sa capacité à se fixer sur les récepteurs de l'organe que l'on cherche à visualiser. Le sujet est alors placé sous une caméra TEP, ressemblant à un anneau. L'élément radioactif, en se désintégrant, va émettre un positron. Ce positron va alors entrer en collision avec un électron du milieu. Ce contact va produire deux photons. La caméra TEP va permettre de détecter la position de ces photons et donc de mettre en évidence la localisation du traceur.

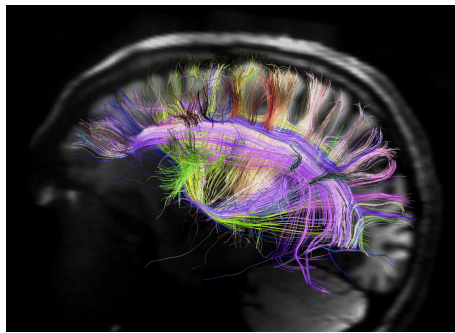
Lorsque la TEP est utilisée en imagerie cérébrale, le traceur utilisé est un dérivé du glucose. Il va venir se fixer au niveau des tissus cérébraux qui, en passant d'un état de repos à un état activé, consomment davantage de glucose. Ainsi, lorsque le composé radioactif marquant le traceur va se désintégrer, il va indirectement révéler la position des cellules cérébrales actives. Un exemple d'acquisition d'une TEP cérébrale est donné en Figure [I.16](#).



(a) Représentation d'un appareil de mesure utilisé en IRM. Image issue de [1].

(b) Représentation d'un enregistrement IRMa sur un sujet sain. Image issue de [37].

(c) Représentation d'un enregistrement IRMf sur un sujet sain. *Crédits : Washington Irving.*



(d) Représentation d'un enregistrement IRMd sur un sujet sain. Image issue de [38].

Figure I.15 – Représentation de l'appareil de mesure utilisé en IRM ainsi que de différentes images provenant de l'IRMa, l'IRMf et de l'IRMd.

3.2 Les méthodes invasives

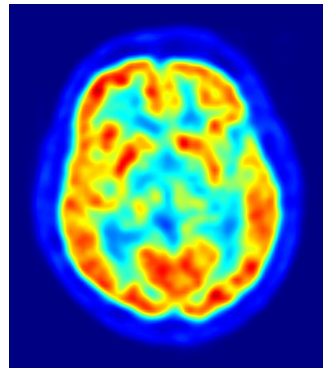
Ces méthodes sont exclusivement mises en œuvre lorsque le sujet est atteint d'une pathologie cérébrale (épilepsie par exemple). Elles nécessitent l'insertion d'un ou plusieurs capteurs dans la boîte crânienne du sujet. Dans le cadre de ce travail, seule la stéréo-électro-encéphalographie va être décrite.

3.2.1 La stéréo-électro-encéphalographie (SEEG)

Cette méthode, développée par Talairach et Bancaud à l'hôpital Sainte-Anne de Paris, consiste à enregistrer l'activité cérébrale à l'aide d'électrodes dites intracrâniennes, c'est-à-dire implantées, par acte chirurgical, en profondeur dans les structures cérébrales du patient (voir

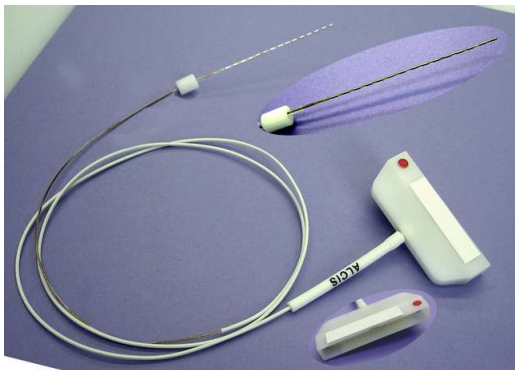


(a) Représentation d'un appareil de mesure utilisé en TEP.

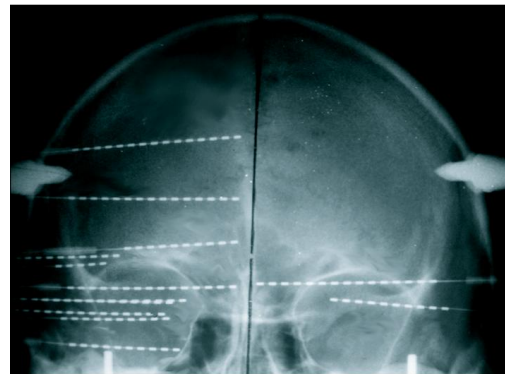


(b) Image cérébrale obtenue par TEP chez un sujet sain.

Figure I.16 – Représentation de l'appareil de mesure utilisé en TEP ainsi que d'un exemple d'image cérébrale obtenue par TEP. Les images proviennent de la thèse de J. Langner [6].



(a) Exemple d'électrode utilisée en stéréo-électro-encéphalographie. Cette image provient du site <http://www.alcis.net>



(b) Radiographie de face montrant l'implantation des électrodes dans le cerveau pour l'acquisition de données de stéréo-électro-encéphalographie. Les points qui apparaissent en blanc sont les capteurs. Image issue de [39]

Figure I.17 – Représentation de capteurs utilisés en SEEG ainsi que d'une implantation d'électrodes.

Figure I.17(b)). Une électrode présente la forme d'une tige sur laquelle 5 à 18 capteurs sont placés en ligne (voir Figure I.17(a)). En général, 10 à 15 électrodes sont utilisées par patient. Un exemple d'électrodes implantées afin d'enregistrer des signaux de SEEG est donné en Figure I.17.

Cette pratique est principalement adaptée à des patients atteints d'épilepsies résistantes aux traitements (pharmaco-résistantes) afin de localiser les tissus cérébraux contenant les foyers de la survenue des crises épileptiques. Une fois ces foyers localisés, une intervention chirurgicale

I.3 Les différentes méthodes d'acquisition de l'activité cérébrale

est réalisée afin de les retirer.

Les enregistrements issus de la SEEG sont très similaires à ceux issus de l'EEG. Il s'agit de potentiels électriques induits par l'activité neuronale. Cependant, le principal avantage de la SEEG réside dans la proximité des électrodes avec les sources de l'activité électrique. Ceci résout de fait les problèmes de résolution spatiale et donc d'identification de sources rencontrés en enregistrements EEG de surface, donc à distance de ces sources. De plus, contrairement à l'EEG, les électrodes étant à l'intérieur de la boîte crânienne, le signal recueilli par la SEEG n'est pas perturbé par le passage à travers différents milieux susceptibles d'altérer le signal, comme l'os crânien par exemple.

3.3 Classification des méthodes d'acquisition

À partir des descriptions précédentes, il est possible de classer les différentes méthodes d'acquisition en fonction de leur degré d'invasivité et de leur résolutions spatiale et temporelle. L'idée d'une telle classification vient de [7]. Cette classification est adaptée sur la Figure I.18.

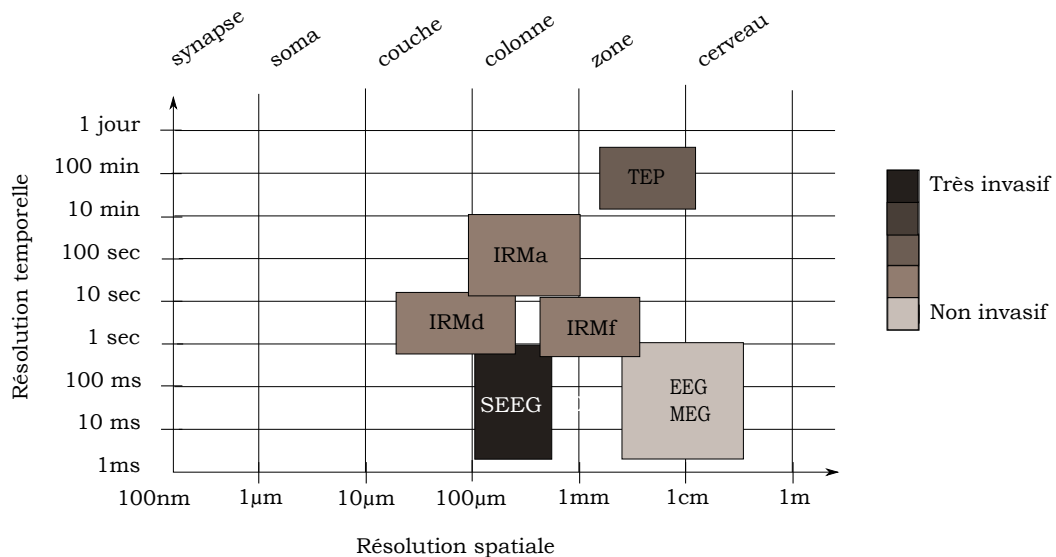


Figure I.18 – Illustration schématique des différentes méthodes d'acquisitions de données cérébrales présentées en fonction de leur résolution spatiale, temporelle et de leur degré d'invasivité. Ce schéma est adapté à partir de celui de [7].

4 Interface Cerveau-Machine

Une interface Cerveau-Machine est un système qui permet à une personne de contrôler une machine (ordinateur, appareil électronique, ...) à l'aide de signaux qui proviennent de son activité cérébrale. Le fonctionnement général d'une ICM est représenté par la Figure I.19 provenant de [8].

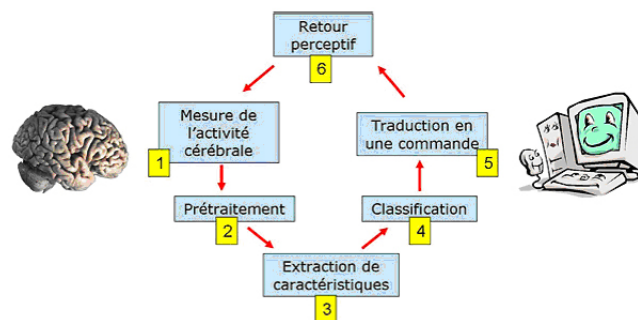


Figure I.19 – Représentation du fonctionnement d'une interface cerveau-machine. Schéma issu de [8].

Une ICM relie donc un sujet à une machine. Le but est que ce sujet puisse envoyer des ordres (déplacement de curseur, sélection d'items, ...) à cette machine par le biais de son activité cérébrale. Ainsi, l'activité cérébrale du sujet va être recueillie afin de pouvoir en extraire un « ordre ». Après ce recueil, les données sont pré-traitées et des caractéristiques sont extraites des signaux sur lesquelles reposera la tâche de classification. Cette dernière va permettre de déterminer l'ordre émis par le sujet. Cet ordre va alors être traduit en une commande pour l'appareil. Les conséquences de l'ordre émis sur sa réalisation par la machine vont alors pouvoir être perçues et analysées par le sujet qui va pouvoir ensuite émettre un nouvel ordre, éventuellement « recalibré ». Lorsque la machine qui est pilotée par l'activité cérébrale est un ordinateur, on parle d'interface cerveau ordinateur (BCI en anglais pour *Brain Computer Interface*).

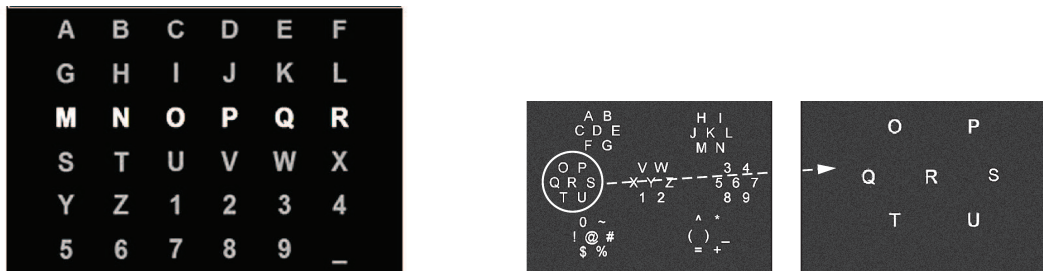
Parmi les différentes méthodes permettant le recueil de l'activité cérébrale, l'EEG est la méthode la plus couramment utilisée en BCI du fait de sa nature peu invasive, de son coût faible, de sa très bonne résolution temporelle et de son transport facile. Compte-tenu, de plus, de l'utilisation de signaux EEG dans le cadre de cette thèse, nous nous limiterons aux ICM fondées sur les signaux EEG dans le reste de ce chapitre. Les méthodes permettant l'extraction de caractéristiques et la classification seront présentées dans le Chapitre III. Nous allons ici nous intéresser aux domaines d'applications des ICM. Le principal domaine d'application des ICM est le domaine médical. Beaucoup de travaux visent à créer des outils permettant à des personnes

souffrant de graves atteintes cérébrales (dus par exemple à une maladie ou à un accident vasculaire cérébral sévère) de retrouver la possibilité de communiquer avec leur entourage ou de se déplacer. Par exemple, la Sclérose latérale amyotrophique (SLA) est une maladie neuro-dégénérative qui entraîne une paralysie totale du patient sans que le fonctionnement du système cérébral gérant les autres fonctions du patient ne soit atteint. Cette maladie mène le patient dans un état dit « d'enfermement » caractérisé par un éveil et une conscience non altérés mais également par une incapacité de mouvement ou de communication (seuls les mouvements des paupières et des yeux sont parfois possibles). Les ICM offrent donc un moyen d'améliorer la qualité de vie de ces patients dont les facultés de déplacement ou de communications ont été détériorées. Quelques applications des ICM vont être détaillées dans les prochains paragraphes, allant du domaine médical au contrôle de jeux vidéos.

4.1 Communication

La restauration de la capacité de communication par le biais des ICM a fait l'objet d'une littérature abondante. Dans la plupart des cas, il s'agit de présenter au sujet un écran sur lequel apparaît un clavier virtuel. L'analyse de son activité cérébrale va permettre de déterminer le caractère qu'il veut sélectionner. Au cours du temps, plusieurs types de claviers virtuels et plusieurs moyens pour le sujet d'envoyer l'ordre de sélection du caractère ont été développés.

Dans [40], le caractère sélectionné par le sujet est déterminé par choix dichotomiques successifs. Pour chacun des trois sujets de l'étude, une première étape consiste à déterminer quelles sont les deux tâches les plus faciles à différencier parmi les 5 suivantes : mouvement imaginé de la main droite, de la main gauche, des deux jambes en même temps, de la langue et une tâche de calcul arithmétique. Une fois que ces deux tâches sont sélectionnées, l'expérience du clavier virtuel peut commencer. Lors de la première étape, 32 caractères, répartis en 2 groupes, sont présentés au sujet sur un écran. Une tâche est associée à chacun des deux groupes. Le sujet peut choisir le groupe contenant la lettre qu'il souhaite sélectionner en réalisant la tâche qui lui est associée. Les 26 caractères du groupe sélectionné vont alors se scinder en deux sous groupes. Le sujet va une nouvelle fois choisir le groupe qu'il souhaite en réalisant la tâche correspondante et ainsi de suite jusqu'à ce qu'il ne reste qu'une seule lettre. Le processus se répète alors pour la lettre suivante. Ainsi, 5 étapes sont nécessaires pour sélectionner un caractère. D'autres type de claviers sont proposés par exemple dans [41] (clavier déroulant de 5 lettres par 5 lettres choisies par des mouvements imaginé des deux mains) et dans [42] (clavier virtuel composé de caractères répartis sur 3 lignes et 9 colonnes). Dans ces deux travaux, les caractères sont,



(a) Représentation du clavier virtuel proposé dans [9].

(b) Représentation du clavier virtuel proposé dans [10].

Figure I.20 – Représentation du clavier virtuel proposé dans [9] et [10]. Les images sont issues de [10].

une fois encore, sélectionnés par la réalisation de tâches mentales (mouvements imaginés des membres, calcul mental, ...).

Enfin, il est courant d'utiliser la composante de potentiel évoqué $P300$ en ICM. En effet, [9] propose l'expérience suivante : un clavier composé de 6 lignes et 6 colonnes est présenté au sujet. Aléatoirement, une colonne ou une ligne de ce clavier virtuel est « flashée » (les caractères de cette ligne ou colonne apparaissant dans une couleur plus vive). Lorsque la ligne ou la colonne contenant la lettre que le sujet souhaite sélectionner est flashée, cela génère une composante $P300$ sur le tracé EEG du sujet. Lorsque la ligne et la colonne sont identifiées (2 enregistrements de $P300$ sont donc nécessaires par caractère), le caractère qui se trouve à la croisée des deux est sélectionné. Il est à noter que la composante $P300$ est ici obtenue à l'aide d'un seul enregistrement EEG contrairement à ce qui a été présenté en Section 3.1.1.9. On parle alors de *Single-trial event-related potentials*. L'utilisation de cette méthode est uniquement possible lorsque l'on sait à l'avance quelle composante du potentiel évoqué on souhaite étudier et que la latence de celle-ci est très peu variable et nécessite l'utilisation de méthodes de traitement du signal particulières qui ne seront pas détaillées ici. L'avantage de ce type d'ICM est que l'apparition du potentiel $P300$ ne nécessite pas d'entraînement. Le principal désavantage du protocole décrit dans [9] est que des $P300$ peuvent apparaître lorsqu'une colonne ou une ligne proche de celle contenant le caractère souhaité est flashée. Afin de palier ce risque d'erreur, [10] propose de séparer l'écran en 7 régions. Chacune des régions contient 7 caractères. Les régions sont flashées aléatoirement comme les lignes et les colonnes dans [9], provoquant l'apparition d'une $P300$ lorsque la bonne région est flashée. Celle-ci est alors sélectionnée et les 7 caractères correspondants sont flashés aléatoirement jusqu'à ce qu'une nouvelle $P300$ soit déclenchée. Les claviers virtuels proposés dans [9] et dans [10] sont respectivement représentés en Figure I.20(a) et I.20(b).

Enfin, [43] propose une amélioration permettant de réduire le taux d'erreur sur les mots obtenus par une ICM fondée sur les travaux de [9]. Cette amélioration consiste à inclure un dictionnaire qui va proposer des corrections automatiques au sujet lorsque le mot obtenu est mal orthographié (à cause de lettres mal sélectionnées).

4.2 Locomotion

En ICM, les travaux visant à améliorer les capacités de locomotion de patients paralysés sont principalement axés sur le contrôle de chaises roulantes par le biais de leurs enregistrements EEG. Par exemple, [44] propose de discrétiser une pièce en 28 rectangles de 60 centimètres par 90 centimètres, organisés en 4 lignes et 7 colonnes. Le sujet peut alors se déplacer sur ce damier à l'aide de mouvements imaginés des membres droits et gauches. Dans un autre travail, [45] propose une approche fondée sur l'utilisation de la composante P300 des potentiels évoqués. Dans cette étude, des chemins prédéfinis menant à des pièces distinctes d'un appartement ou à des points dans l'une de ces pièces sont préalablement enregistrés dans le système. Les différents chemins sont alors proposés aléatoirement au sujet. Lorsque le chemin désiré est proposé, cela fait naître une composante P300 qui peut être détectée. La chaise se met alors en mouvement en suivant ce trajet ainsi sélectionné. Le sujet peut, à tout moment, arrêter la progression de la chaise.

La limite de ces approches repose sur le manque d'adaptabilité à un environnement inconnu. Une solution à ce problème est proposée par [15]. Dans ce travail, les lieux sont modélisés à l'aide d'un scanner laser. Un maillage de la pièce (4 lignes et 5 colonnes) est alors représenté au sujet avec des points accessibles (intersections entre les lignes et les colonnes). Le système est fondé sur l'analyse des composantes P300. Au lieu d'afficher les points un par un (problème de classification à 20 essais), les lignes et les colonnes sont présentées de la même façon que le clavier virtuel proposé par [9] et présenté précédemment. Ce procédé limite alors le problème à deux classifications de 4 et 5 essais. Ce travail a abouti à de très bons résultats (aucune collision et une vitesse de parcours de 0.16 mètre par seconde).

4.3 Restauration de mouvements

La paralysie partielle ou totale entraîne une forte dégradation des conditions de vie d'un patient et le rend dépendant de son entourage. La restauration de la capacité de mouvement est donc un axe de recherche très important. La Stimulation Électrique Fonctionnelle (SEF) est une technique qui consiste à activer artificiellement un muscle par le biais de signaux électriques

qui se substituent à ceux qui devraient être émis par les neurones du patient. Depuis la première utilisation de la SEF en 1961 [46], des développements visant à la coupler avec une ICM ont été réalisés. Par exemple, [47] a réalisé des travaux permettant à des patients tétraplégiques de saisir un cylindre à l'aide d'un bras paralysé. Cette approche est fondée sur la détection d'ondes β dans l'électroencéphalogramme provoquées par des mouvements imaginés des pieds.

L'utilisation de la SEF est concevable lorsque l'intégrité des membres que l'on cherche à contrôler est préservée. Lorsque ce n'est pas le cas, des ICM pour contrôler des prothèses remplaçant un membre perdu ont été élaborées. Ainsi, une approche fondée sur l'utilisation d'une prothèse de bras (de l'avant bras aux doigts) contrôlé par une ICM est proposée dans [16]. Cette ICM s'appuie sur l'utilisation d'un potentiel évoqué particulier, appelé *Steady State Visually Evoked Potential* (SSVEP). Ce potentiel est évoqué par des stimuli visuels présentés à une fréquence fixe (entre 3.5 Hz et 75Hz). Lorsqu'il se concentre sur ces stimuli, le sujet va générer une activité électrique cérébrale caractéristique de la fréquence des stimuli présentés. Lorsque plusieurs stimuli sont présentés à différentes fréquences, il est possible de distinguer celui sur lequel le sujet a porté son attention en analysant le SSVEP qui a été généré par ce stimulus. Ainsi, dans [16], quatre signaux lumineux clignotant à des fréquences distinctes sont présentés au sujet. Chacun de ces quatre stimuli est associé à une commande spécifique. En effet, les lumières clignotant à la fréquence de 6 et 7 Hz sont associées respectivement à un mouvement pivotant du poignet vers la gauche ou vers la droite tandis que les lumières clignotant à la fréquence de 8 et 13 Hz sont associées respectivement à l'ouverture et la fermeture de la main artificielle. Pour choisir l'un de ces mouvements, le sujet doit focaliser son attention sur la lumière correspondante.

4.4 Applications ludiques

4.4.1 Contrôle de jeux vidéos

Les premiers jeux vidéos ont vu le jour dans les années 1950. La qualité de ces jeux n'a cessé de s'améliorer permettant une immersion de plus en plus importante du joueur. Le contrôle du jeu vidéo a évolué de façon continue. Il s'est d'abord fait par des manettes contenant de plus en plus de boutons, permettant donc de plus en plus de possibilités d'interactions. En 2006, un nouveau système fondé sur l'utilisation d'une caméra a permis une avancée majeure dans le domaine du contrôle des jeux. Avec ce système, le joueur peut commander un personnage virtuel à l'aide de mouvements qui sont captés par la caméra et traduits en mouvements virtuels. Les ICM en jeu vidéo s'inscrivent dans la lignée de ces améliorations. Elles permettent de contrôler

un personnage virtuel sans manette, à l'aide de son activité cérébrale.

Un exemple de jeu vidéo contrôlé par ICM est donné dans [11]. Le jeu, appelé « Mind-Balance », consiste à faire avancer un personnage funambule et de parvenir à lui faire garder son équilibre. Le personnage porte un balancier avec un drapeau à damier de chaque côté. Le personnage avance le long du fil et perd l'équilibre au bout d'un temps aléatoire (entre 1.5 et 5 secondes) en penchant vers un côté choisi aléatoirement. Une image issue du jeu présentée dans [11] est montrée en Figure I.21. Le joueur doit donc tenter de maintenir l'équilibre du personnage en se concentrant sur le drapeau positionné à l'extrémité du balancier du côté opposé à celui vers lequel penche le personnage. Ce système de contrôle est fondé sur le SSVEP. En effet, les deux drapeaux du personnage sont présentés avec une fréquence de scintillement de l'image sur l'écran différente (17 Hz pour l'un et 20 Hz pour l'autre). La focalisation de l'attention du joueur sur l'un des deux drapeaux va entraîner l'émergence d'un SSVEP. En analysant les caractéristiques de ce SSVEP, il est en effet possible de déterminer sur lequel des deux drapeaux le sujet se concentre et de modifier en conséquence l'équilibre du personnage virtuel représenté à l'écran .



Figure I.21 – Représentation du jeu MindBalance. Le but de ce jeu est de maintenir l'équilibre du personnage qui se déplace le long d'un fil. Cette image provient de [11].

D'autres ICM appliquées au contrôle de jeux vidéos ont été développées. On peut par exemple citer les travaux de [48], également fondés sur les SSVEP, dans lesquels le joueur de contrôle les mouvements vers la droite et la gauche d'un avion ou les travaux de [18] qui

proposent de contrôler un vaisseau spatial de l'univers de Star Wars. Dans cette application, appelée « Use-the-force » et utilisant la plateforme OpenVibe¹, le contrôle du vaisseau se fait à l'aide de mouvements imaginés des pieds.

4.4.2 Visite virtuelle de musée

Une ICM appliquée à la visite virtuelle d'un musée a été développée par [17]. Dans celle-ci, deux modes de navigation sont proposés : une navigation assistée et une navigation non assistée. Dans le mode de navigation assistée, des points d'intérêts (tableaux, sculptures, points dans une pièce afin de changer de position,...) sont proposés au « visiteur » sur sa droite ou sa gauche, qu'il peut choisir à l'aide de mouvements imaginés de la main droite ou de la main gauche. Le mode de navigation non assistée (sans proposition de points d'intérêts) est accessible, à partir du mode de navigation assistée, en réalisant un mouvement imaginé des deux pieds. Dans ce mode, le sujet peut changer l'orientation de la caméra vers la droite ou la gauche, par un mouvement imaginé de la main correspondante, afin de localiser les positions des points d'intérêts qu'il souhaite rejoindre et d'adapter ainsi son déplacement à l'intérieur du musée. Une illustration de cette application est représentée en figure I.22.



Figure I.22 – ICM permettant de réaliser une visite virtuelle de musée. Image provenant de <https://team.inria.fr/potioc/fr/scientific-subjects/bci-2/>.

4.5 Jeux de données des compétitions BCI

Les recherches sur les méthodes utilisant le traitement des données cérébrales dans le but de créer une BCI sont très actuelles. Afin de susciter des avancées rapides de la communauté

1. Pour plus de renseignements, voir <http://openvibe.inria.fr/>

de chercheurs intéressés par ces interfaces et les problèmes concrets qui se posent dans leur développement, des compétitions de BCI ont été organisées. Des bases de données sont également mises à disposition permettant de tester de nouveaux outils en BCI et de comparer une nouvelle idée à des approches déjà existantes (voir par exemple [14], [49] ou encore [50]).

Quatre compétitions, avec plusieurs jeux de données correspondant à des problèmes différents, ont été organisées. Les jeux de données sont accessibles à l'adresse : <http://www.bbci.de/activities>. Du fait de la diversité des problèmes posés et du nombre de travaux proposés en réponse, les bases de données des compétitions BCI constituent un atout majeur pour le développement de nouveaux outils en BCI.

Chapitre I. L'activité cérébrale humaine : genèse, acquisition et utilisation

Chapitre II

Acquisition de données d'Electro-EncéphaloGraphie

Sommaire

1	Recrutement des sujets et conditions expérimentales	44
1.1	Recrutement des sujets	44
1.2	Environnement de l'acquisition	45
2	Préparation du sujet	46
3	Pose du casque	47
4	Enregistrement de signaux EEG	48
5	Séance de relaxation	50
6	Recueil de la variation contingente négative (VCN)	50
7	Remarques sur le protocole expérimental	51
8	Validation des données	52
8.1	Pré-traitement des signaux	52
8.2	Extraction de la VCN et règle de décision	53
8.2.1	Sélection de la période pour le calcul de l'aire	53
8.2.2	Sélection des électrodes par test statistique	54
8.2.3	Sélection des sujets par test statistique	57
8.3	Inspection visuelle de la VCN	59
8.4	Résultats de l'étude de la VCN	60
9	Les données	64

La recherche d'un critère permettant de déterminer l'état de vigilance d'une personne a donné lieu à une expérience. Le but de cette expérience était d'obtenir un jeu de données réelles afin de rechercher ce critère. Le protocole expérimental est composé de 6 étapes qui sont représentées sur la Figure II.1 :

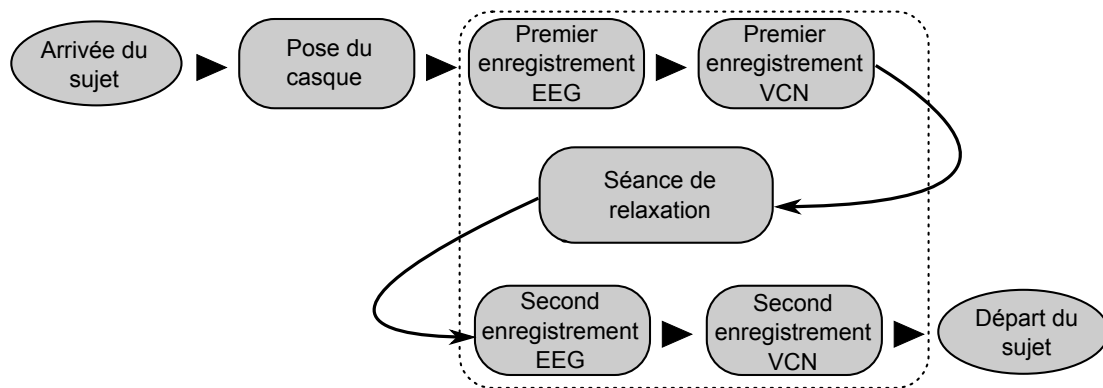


Figure II.1 – Diagramme représentant la procédure d'acquisition des données EEG. Les cinq étapes clés de la procédure sont encadrées.

Environ deux heures trente sont nécessaires pour installer le casque à électrodes, réaliser l'expérience et avoir un bref entretien avec le sujet. Cet entretien, réalisé à la fin de l'expérience pour ne pas affecter les résultats, permet de recueillir les impressions du sujet et de lui expliquer a posteriori les attendus de l'expérience.

Julien Clauzel, Nidal El Yacoubi, Vérane Faure, Émilie Drouineau, Frédérique Faïta-Aïnseba, Pierrick Legrand et moi-même, avons administré en binôme le protocole expérimental. Les quatre premières personnes sont des étudiants stagiaires de niveau Master encadrés par Frédérique Faïta-Aïnseba.

1 Recrutement des sujets et conditions expérimentales

1.1 Recrutement des sujets

Les sujets ont été recrutés en respectant certains critères d'inclusion, permettant d'éviter certains biais expérimentaux. Ainsi, ils ont tous moins de trente-cinq ans, afin d'éviter les modifications des caractéristiques de l'EEG liées à l'âge, et sont tous droitiers, afin de contrôler la variabilité de résultats due à des différences de latéralisation hémisphérique (« inversion » des

II.1 Recrutement des sujets et conditions expérimentales

aires cérébrales droites et gauches entre cerveaux de droitiers et de gauchers).

De cette façon, la localisation des électrodes montrant les activités les plus significatives pourra se faire en limitant les risques d'hétérogénéité spatiale due aux variabilités inter-individuelles de conformation cérébrale.

Enfin, les sujets étaient non-fumeurs afin de limiter le stress pouvant être généré par une éventuelle envie de fumer au cours de l'expérience. Avant de procéder à tout recueil de signaux EEG, le sujet était invité à lire et signer une feuille de consentement éclairé pour s'assurer de sa compréhension de la nature de l'expérience et de son accord pour y participer. Il y était notamment précisé que l'expérience pouvait être interrompue à tout moment si le participant en ressentait la nécessité.

Au cours de ces travaux, deux campagnes d'acquisitions ont été menées. La première, en 2011, a permis de recueillir les signaux EEG de 44 sujets. En 2012, la seconde campagne d'acquisition permis de réaliser des enregistrements chez 14 sujets supplémentaires. Au total, 58 sujets (35 femmes et 23 hommes) ont participé à cette étude.

1.2 Environnement de l'acquisition

Deux pièces, représentées en Figure II.2, sont utilisées lors des acquisitions. Les participants viennent en binôme assister à l'expérience. L'un des deux est le sujet et l'autre assiste les expérimentateurs, dans le cadre d'une participation volontaire à une expérience venant illustrer un cours de « bases biologiques du comportement » dispensé en Licence MASS à l'Université Bordeaux Segalen. Deux membres de l'équipe sont également présents en tant qu'expérimentateurs. Lorsque la procédure d'enregistrement des signaux EEG débute, les expérimentateurs se situent dans la pièce A tandis que le sujet est seul dans la pièce B. De cette manière, le sujet n'est pas perturbé par la présence d'autres personnes dans la pièce. Une caméra avec un micro intégré (5), reliée en wifi à l'ordinateur 3, permet de visualiser la pièce B afin de s'assurer du bon déroulement de l'expérience et de pouvoir l'interrompre en cas de demande du sujet ou de problème majeur. L'ordinateur 1 pilote les enregistrements des signaux EEG. L'ordinateur 2 gère l'interface avec le participant (affichage sur l'écran placé devant le participant (4), lancement du protocole géré sous Matlab, etc). Dans la pièce B, des hauts parleurs permettent de diffuser une bande sonore au sujet (lors de la séance de relaxation). L'écran 4 permet quant à lui d'afficher les consignes visuelles, lancées par les expérimentateurs depuis l'ordinateur 2.

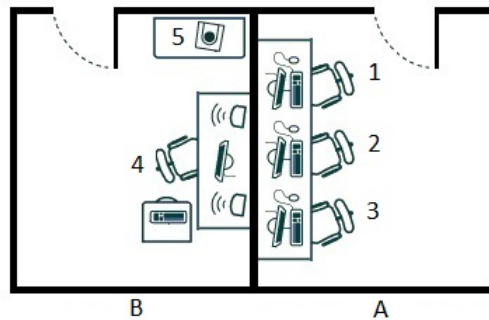


Figure II.2 – Schéma représentant les deux pièces utilisées lors des acquisitions.

2 Préparation du sujet

Lors de son arrivée, le participant est invité à retirer tout objet qui pourrait interférer avec le champ électrique environnant (bijoux, barrettes métalliques, téléphone portable, etc). Si le participant porte des lunettes, il peut les conserver une fois le casque posé.

La première étape consiste à poser des électrodes cupules (voir Figure I.8(a) dans le Chapitre I) au niveau des yeux et derrière les oreilles. Pour ce faire, la peau du sujet doit être nettoyée et dégraissée à l'alcool afin d'améliorer le contact entre la peau et les électrodes et diminuer ainsi leur impédance.

Une fois que la peau a été préparée, un gel électrolytique est disposé dans chaque électrode-cupule pour favoriser son contact. L'électrode est alors posée directement sur la peau du sujet où elle est maintenue à l'aide de sparadrap.

Les électrodes oculaires permettent d'enregistrer un Electro-OculoGramme (EOG). L'analyse de l'EOG permet de détecter les mouvements oculaires ainsi que les clignements (voir Section 3.1.1.2 dans le Chapitre I). Ces artéfacts nettement visibles sur l'électroencéphalogramme recueilli perturbent la future exploitation de celui-ci (voir la Figure I.7 du Chapitre I). Pour certains signaux, il peut donc être intéressant de savoir à quel moment un artéfact oculaire est apparu afin de pouvoir retirer la portion de signal EEG impactée. Les électrodes oculaires permettent donc d'obtenir cette information. Afin de capter à la fois les clignements et les saccades oculaires, une électrode est placée en dessous de l'œil gauche du sujet et une autre électrode est placée à droite de son œil droit.

Dans la présente expérience, un montage référentiel est utilisé (voir section 3.1.1.3 dans le Chapitre I). Dans le montage utilisé, la référence est obtenue en moyennant les signaux enregistrés par les deux électrodes placées sur les os mastoïdes (situés derrière l'oreille).

3 Pose du casque

Une fois les quatre électrodes servant de référence et au recueil de l'EOG placées, un casque à 58 électrodes (Electrocap) est placé sur le crâne du sujet. Le casque est connecté au système d'acquisition via la boîte têtère (voir Section 3.1.1.4 du Chapitre I). Afin de limiter la captation de signaux électriques ambiants, l'ensemble des branchements des matériels électriques du box expérimental où se trouve le sujet (écrans, haut-parleurs, caméra, etc) se fait sur des prises situées dans la pièce de contrôle adjacente (grâce au passage des câbles par une ouverture pratiquée dans le mur).

Ainsi, dans ce travail, 58 électrodes actives sont utilisées pour le recueil du signal EEG. Elles sont fixées au casque et ainsi positionnées sur le scalp selon le système international 10/10, qui répond à des règles de placement et une nomenclature standard. La répartition des électrodes dans ce travail est illustrée sur la Figure II.3(a).

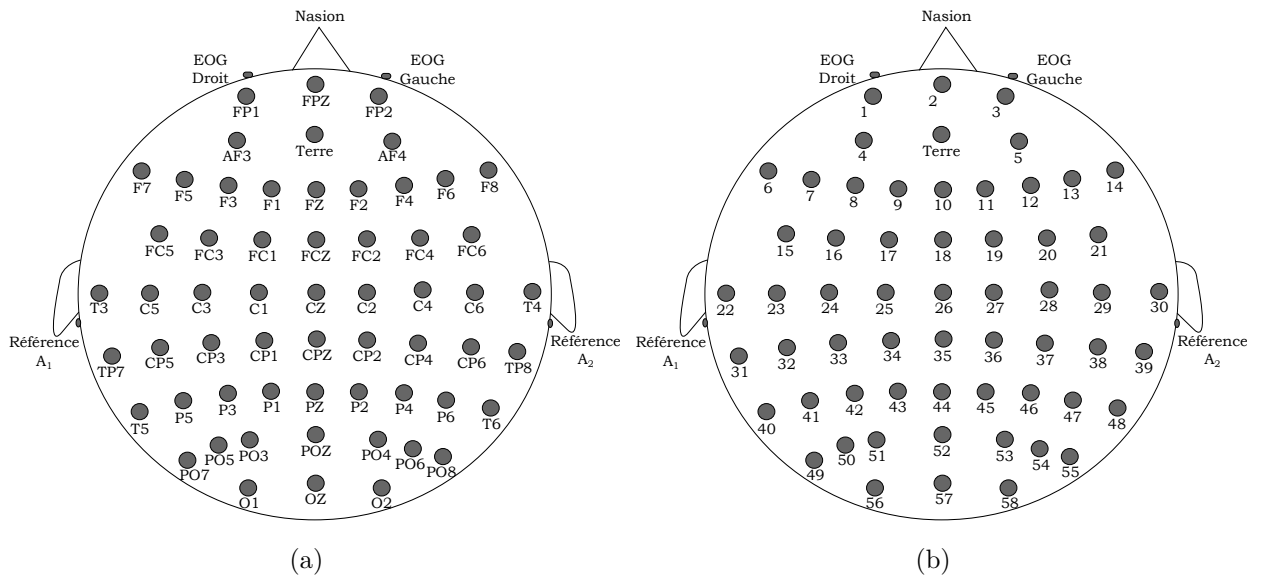


Figure II.3 – Représentation de la répartition des électrodes lors des acquisitions.

La nomenclature associée à ce système a été détaillée en Section 3.1.1.1 dans le Chapitre I. La Figure II.3(b) montre la numérotation des électrodes utilisée dans cette étude. Cette numé-

rotation va permettre de simplifier les représentations des résultats dans la suite de ce travail. Elle se fait dans la direction antéro-postérieure, en commençant par les électrodes placées en région pré-frontale (notée FP) et en augmentant jusqu'aux électrodes placées à l'arrière de la tête (zone occipitale notée O). L'acquisition nécessite l'utilisation d'une électrode de masse (terre), constituée ici de l'électrode AFz (voir Figures II.3 et I.5).

Le logiciel Cohérence 3NT (Deltamed, <http://www.natus.com/>) est utilisé dans ce travail pour le recueil des signaux EEG. Il a une fonctionnalité qui permet en cours de pose de visualiser l'impédance des électrodes, et de s'assurer ainsi du bon contact électrique de celles-ci. L'expérimentateur va donc, grâce à une seringue munie d'une canule pouvant s'introduire dans les plots enchâssant les électrodes, écarter les cheveux du sujet et ajuster la quantité de gel et sa répartition dans la cupule, jusqu'à obtenir la meilleure impédance possible (idéalement inférieure à 5KOhms). La pose des 58 électrodes nécessite environ 45 minutes.

Ces précisions sont importantes pour expliquer un des intérêts de ce travail de thèse. En effet, on comprend bien que la pose des électrodes, du fait qu'elle nécessite beaucoup de temps, peut réduire le champ des applications possibles de l'EEG en BCI. Il s'avère donc utile, dans des perspectives d'applications ultérieures (cliniques, commerciales ...) d'effectuer une réduction de variables en minimisant le nombre d'électrodes nécessaires ce qui réduira de la pose d'un casque ou d'éventuelles électrodes libres en nombre restreint.

Après la pose du casque, le sujet est renseigné sur les contraintes imposées pour garantir la qualité du signal EEG. Les consignes sont de ne pas bouger, d'être le plus décontracté possible, de respirer par la bouche pour éviter toute crispation des maxillaires, et de garder les bras et les jambes décroisées afin d'éviter au maximum les artefacts musculaires (voir la Figure I.6 du Chapitre I). Une fois les consignes données, le sujet est laissé seul dans le box expérimental, les expérimentateurs se placent dans la pièce de contrôle voisine et l'expérience peut débuter.

4 Enregistrement de signaux EEG

Durant l'expérience, il y a deux enregistrements EEG. Le premier enregistrement intervient au début de l'expérience avant toute activité préalable du sujet. Il vise à recueillir l'EEG du sujet considéré comme étant dans un état de vigilance de « référence », dit état « normal ».

II.4 Enregistrement de signaux EEG

Le second enregistrement, intervenant après une séance de relaxation, vise à recueillir l'activité cérébrale du sujet dans un état de basse vigilance, dit état « relaxé ».

Chacune de ces périodes d'enregistrement dure trois minutes durant lesquelles le sujet doit fixer une croix blanche, située au centre d'un écran, pour éviter toute saccade oculaire. Le sujet garde donc les yeux ouverts et peut cligner des yeux sans contrainte particulière. Une photographie d'un membre de notre équipe, prise afin d'illustrer les conditions d'enregistrements des signaux EEG, est donnée en figure II.4.

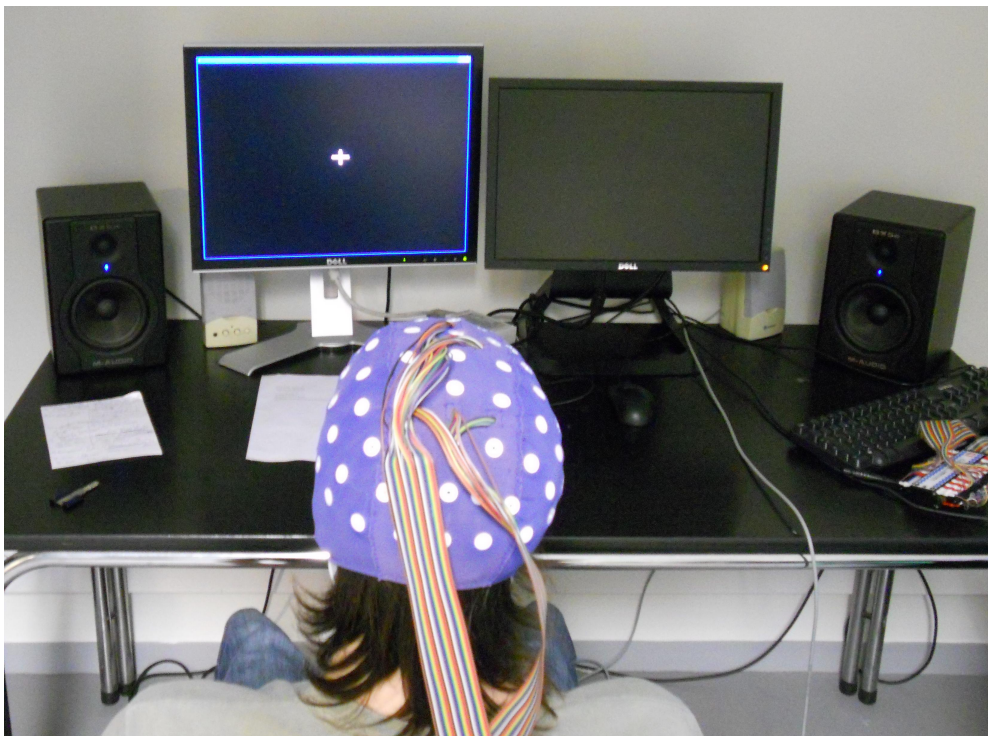


Figure II.4 – Photographie représentant les conditions d'enregistrement des signaux EEG.

Les enregistrements EEG ont été réalisés à une fréquence d'échantillonnage de 256 Hz. Ainsi, chaque tracé obtenu pour une électrode dans chaque épisode enregistré contient plus de 46000 points.

Remarque : Le fait que le sujet garde les yeux ouverts peut être un point discutable car les ondes alpha, caractéristiques d'un état de veille passif, apparaissent principalement lorsque le cerveau du sujet n'est pas sollicité pour analyser un contenu sensoriel. Elles sont ainsi plus marquées lorsque le sujet a les yeux fermés. Cependant, le but de notre étude est de pouvoir

prédire l'état de vigilance d'une personne sans que cette dernière n'ait forcément les yeux fermés. C'est pourquoi, afin de ne pas restreindre les domaines d'applications de ce travail, il a été choisi que les sujets aient les yeux ouverts durant les acquisitions.

5 Séance de relaxation

Le but de la séance de relaxation est d'abaisser le niveau de vigilance du sujet afin de l'amener dans un état dit « relaxé ». Durant la séance, le sujet est guidé par une bande son, diffusée par les haut-parleurs de la salle B (voir Figure II.2). La séance comprend trois exercices d'auto-relaxation fondés sur de la relaxation musculaire et de la visualisation mentale¹. Le premier exercice est issu de la méthode du training autogène [51]. Durant celui-ci, le sujet doit se répéter mentalement des phrases qui favorisent l'auto-hypnose. Le deuxième exercice s'apparente à la méthode de la relaxation progressive [52]. Le sujet doit successivement contracter et décontracter certains muscles, en particulier ceux des bras et des jambes. Enfin, un exercice de visualisation mentale clôture la séance. Le sujet doit alors imaginer qu'il se trouve dans un lieu qui lui est familier et qu'il affectionne et se concentrer sur les sensations éprouvées (odeurs, bruits, couleurs). Durant l'intégralité de la séance de relaxation, aucun enregistrement EEG n'est réalisé.

6 Recueil de la variation contingente négative (VCN)

Si nous nous étions contentés de recueillir l'EEG du sujet avant et après la séance de relaxation, il aurait été impossible de savoir si cette séance a permis d'atteindre le but recherché, c'est-à-dire a effectivement abaissé le niveau de vigilance du sujet. C'est donc dans le but de vérifier l'efficacité de la relaxation sujet par sujet qu'une tâche permettant de recueillir la VCN a été introduite dans le protocole expérimental. Cette tâche a été effectuée avant et après la séance de relaxation.

La VCN est une composante des potentiels évoqués qui se développe lorsqu'un sujet se prépare à réaliser une tâche. Cette composante a la particularité d'avoir une amplitude qui varie en fonction de l'état de vigilance du sujet (voir Section 3.1.1.9 dans le Chapitre I). L'analyse des amplitudes des VCN obtenues avant et après la séance de relaxation va donc permettre de

1. Les textes qui ont permis l'enregistrement des trois méthodes de relaxation sont disponibles sur la page internet <http://www.sommeil-mg.net/spip/Techniques-de-relaxation>

II.7 Remarques sur le protocole expérimental

déterminer si la séance de relaxation a eu l'effet escompté sur l'état de vigilance du sujet.

Le protocole expérimental est le suivant : le sujet est assis devant un écran d'ordinateur présentant une croix blanche en son centre, sa main droite posée sur la barre « espace » d'un clavier (voir Figure II.2) sur laquelle il doit appuyer le plus rapidement possible, à chaque apparition d'un carré venant se substituer à la croix blanche sur l'écran. Un signal d'avertissement sonore (bip) précède, avec un délai de 2.5 secondes, chaque apparition du carré, ce qui permet ainsi au sujet de préparer son appui. L'EEG du sujet est enregistré pendant toute la durée de la tâche, qui comprend 50 paires de stimuli « S1 : bip - S2 : carré ». Le temps entre deux paires de stimuli est aléatoire. Durant la tâche, le sujet a pour consigne d'éviter au maximum de cligner des yeux entre S1 et S2 pour éviter tout artefact oculaire (voir Section 3.1.1.2 du Chapitre I). Il peut en revanche cligner des yeux entre les essais. Un schéma représentant le protocole de recueil de la VCN est fourni en Figure II.5.

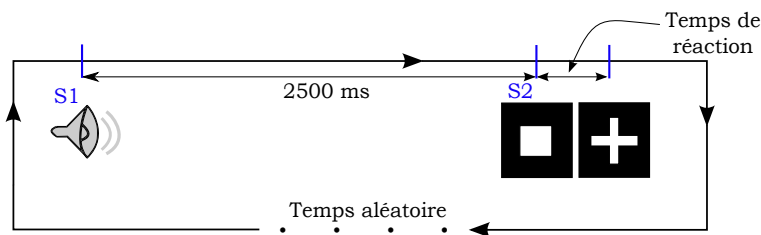


Figure II.5 – Schéma représentant le protocole de recueil de la VCN. La tâche est composée de 50 paires de stimuli « S1 : bip - S2 : carré » durant lesquels le signal EEG du sujet est enregistré.

Comme évoqué dans la remarque de la section 3.1.1.9 du Chapitre I, le paradigme expérimental permettant de recueillir la composante VCN inclut des perturbations susceptibles de modifier l'état de vigilance du sujet (diffusion de signaux sonores, sollicitation de mouvements). Dans cette étude, les amplitudes des VCN recueillies dans les deux états de vigilance vont être comparées pour vérifier que l'état de vigilance du sujet a baissé après la séance de relaxation. Ainsi, on peut considérer que si l'état de vigilance du sujet est suffisamment bas après cette dernière, les perturbations dues au protocole d'acquisition de la composante VCN n'auront pas d'effets sur notre étude. Le traitement des signaux EEG afin d'extraire la VCN ainsi que les résultats obtenus seront détaillés en Section 8.

7 Remarques sur le protocole expérimental

Le protocole expérimental qui vient d'être présenté correspond au protocole définitif grâce auquel les signaux EEG de notre étude ont été recueillis. Cependant, avant d'entériner ce proto-

cole, une étape de validation a été nécessaire. Elle a permis de déterminer, en passant plusieurs sujets « pilotes », les conditions optimales de réalisation du protocole (minimisation du temps de pose du casque, ajustement de la durée de chaque exercice de relaxation, détermination du nombre minimal d'essais requis pour l'extraction de tracés de VCN corrects, etc).

8 Validation des données

Dans cette partie, les signaux EEG enregistrés lors des deux phases de recueil de la VCN vont être traités pour en extraire la composante. L'étude des composantes VCN ainsi obtenues est réalisée sur les 58 sujets de l'étude afin de déterminer si la séance de relaxation a permis d'abaisser leur niveau de vigilance. Seuls les enregistrements EEG pré- et post-relaxation des sujets pour lesquels cette baisse de vigilance aura été constatée pourront être conservés pour la suite de nos travaux.

8.1 Pré-traitement des signaux

Pour un individu donné, 100 signaux EEG enregistrés entre les stimuli S1 et S2 sont obtenus (50 récoltés avant la séance de relaxation et 50 après). Afin d'éliminer les signaux contenant des artefacts oculaires (essais durant lesquels le sujet a cligné des yeux), les enregistrements subissent un test automatique de détection des clignements. Comme évoqué en Section 3.1.1.2 du Chapitre I, lorsqu'un clignement intervient, un pic est mesuré sur l'électrode oculaire placée sous l'œil du sujet. Ainsi, la différence entre le signal mesuré sur l'électrode oculaire droite et celui mesuré sur l'électrode oculaire gauche est considérée. Si, dans un intervalle allant de 200ms avant S1 à S2, un pic dépassant $400 \mu V$ est détecté, l'enregistrement est automatiquement rejeté. Un exemple d'essai présentant un artefact oculaire et de la différence des électrodes oculaires est donné en figure II.6. L'intervalle allant de 200ms avant S1 à S1 est inclus dans le test car il permet de calculer la ligne de base de l'essai. Celle-ci correspond au signal EEG qui aurait été enregistré si aucun stimulus avait été présenté au sujet. Elle est obtenue pour chaque essai en réalisant la moyenne de la portion de signal allant de 200ms avant S1 à S1. La valeur obtenue est alors retranchée à toute la séquence afin de recalibrer l'essai.

Sur les 58 sujets de l'étude, 14 sujets ont été éliminés après le retrait des artefacts oculaires. En effet, ces derniers ne présentaient plus assez d'essais valides dans au moins une des phases de test VCN après le rejet des artefacts.

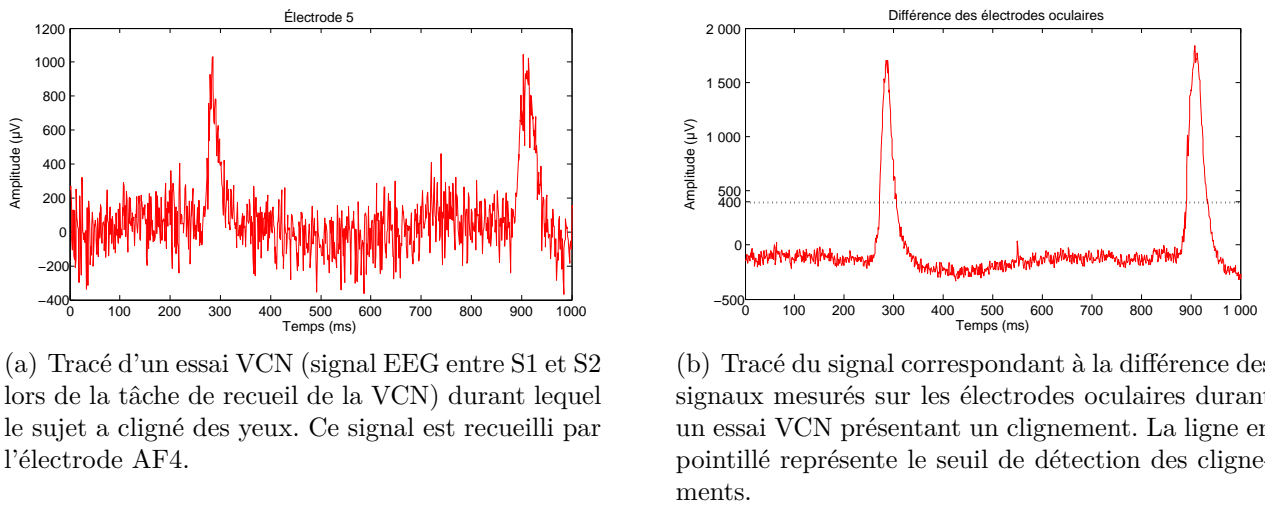


Figure II.6 – Représentation de la gestion des artéfacts oculaires.

8.2 Extraction de la VCN et règle de décision

Pour un sujet donné, l'analyse de la VCN va permettre de vérifier si la séance de relaxation a bien eu l'effet escompté sur son état de vigilance. Le recueil de la VCN a été détaillé en Section 6. L'extraction de la VCN est réalisée en appliquant la méthode des potentiels évoqués détaillée en Section 3.1.1.9 du Chapitre I. Cette méthode consiste à moyenniser, par électrode et par sujet, tous les tracés EEG exempts d'artefact recueillis entre S1 et S2. Ainsi l'on fait émerger les composantes du potentiel évoqué par S1, dont la VCN.

Pour déterminer si la séance de relaxation a bien permis d'abaisser le niveau de vigilance d'un sujet donné, les amplitudes de la VCN enregistrée avant et après la séance de relaxation vont être comparées. Comme évoqué en Section 3.1.1.9 du Chapitre I et représenté sur la Figure I.13, la séance de relaxation est supposée entraîner une baisse de l'amplitude de la VCN.

Pour comparer les amplitudes des VCN, la méthode utilisée est de calculer les aires sous les courbes VCN. Pour cela, il faut déterminer la période entre S1 et S2 qui est significative (période sur laquelle calculer l'aire) et les électrodes sur lesquelles il faut calculer l'aire (électrodes discriminantes).

8.2.1 Sélection de la période pour le calcul de l'aire

Afin de déterminer la période sur laquelle étudier la VCN, les tracés moyens dans les deux états de vigilance chez les 44 sujets retenus ont été soumis à une exploration visuelle. La Figure II.7 représente ces tracés pour les deux états de vigilance et pour l'électrode FCz. Cette Figure

montre que la VCN semble débiter à partir de 300 ms après S1, lorsque le tracé moyen devient négatif. Contrairement à certaines VCN obtenues dans la littérature, après 1000ms le signal revient à la ligne de base. Ce résultat, bien qu'inattendu, semble relativement robuste. En effet, le même constat est obtenu chez tous les sujets, sur toutes les électrodes et pour les deux états de vigilance. Il pourrait être dû à l'intervalle de temps conséquent entre S1 et S2. Des études se sont d'ailleurs intéressées à l'influence de ce temps inter stimuli sur la VCN obtenue (voir notamment [53]). Dans notre travail, la VCN semble donc se développer entre 300ms et 1000ms. Afin d'être sûrs de cibler la partie commune aux composantes de tous les sujets, l'aire de la VCN sera calculée entre 360ms et 900ms. Cet intervalle est représenté sur la Figure II.7 par des lignes verticales jaunes.

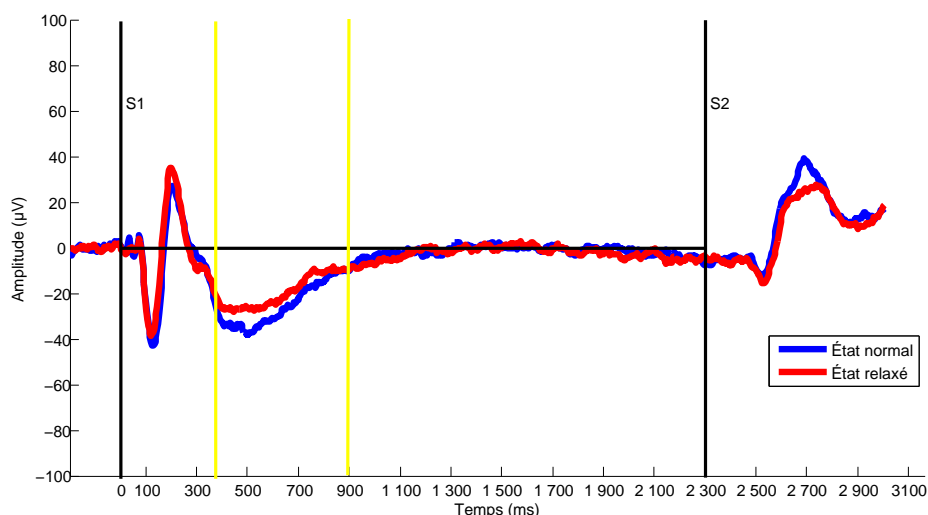


Figure II.7 – Représentation de tracés EEG moyens obtenus dans le protocole de recueil de la VCN dans les deux états de vigilance (courbe bleue pour l'état avant relaxation et rouge pour l'état après relaxation). Ces tracés sont ceux de l'électrode FCz.

8.2.2 Sélection des électrodes par test statistique

Dans cette partie, une approche visant à isoler les électrodes qui présentent une différence notable entre les aires des VCN pour les deux phases d'enregistrement (avant et après relaxation) va être proposée. Il s'agit de construire un test statistique visant à vérifier la présence de cette différence.

Étant donné que les mesures des aires sont répétées sur le même sujet (avant / après relaxation), le test utilisé est apparié. Pour réaliser ce test, la différence des aires mesurées dans ces

II.8 Validation des données

Fz 0.00942	FCz 0.01259	FC1 0.01426	F1 0.01612	F2 0.02294	F3 0.02869	FC2 0.03113	FC3 0.04711	F4 0.05999	AF3 0.06892	FC4 0.07882
FPz 0.09964	AF4 0.11253	F5 0.11479	FC5 0.12413	FP1 0.16086	C3 0.16662	C1 0.17853	Cz 0.20719	F6 0.23476	C2 0.29540	T3 0.29941
C5 0.31159	FP2 0.32398	FC6 0.34504	T5 0.38855	P08 0.46089	F8 0.47927	F7 0.49309	P07 0.52533	C4 0.52993	CP3 0.53911	P5 0.53911
P03 0.53911	TP7 0.57106	CP5 0.58011	Oz 0.58462	P05 0.59361	P06 0.60256	O1 0.60256	O2 0.61587	CPz 0.63341	P3 0.64209	P1 0.65069
CP1 0.65496	C6 0.68430	P0z 0.69249	CP2 0.72429	P2 0.73198	P04 0.76166	Pz 0.77580	P4 0.81841	P6 0.83914	T6 0.84754	CP4 0.85028
CP6 0.86347	T4 0.89829	TP8 0.91402								

Figure II.8 – Liste des électrodes classées par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs inférieurs au seuil de 5% ainsi que les électrodes associées apparaissent en bleu.

deux phases est calculée pour chaque sujet et chaque électrode. Une matrice A de taille 44×58 est obtenue. La valeur à la ligne i et la colonne j de cette matrice est positive lorsque l'aire de la VCN mesurée avant la séance de relaxation est supérieure à celle mesurée après la séance de relaxation pour le sujet i et l'électrode j . Afin de déterminer si un test paramétrique peut être utilisé, un test de normalité (Shapiro-Wilk avec un seuil de 5%) est appliqué sur les colonnes de A . Le rejet de ce test pour la plupart des électrodes montre qu'un test non paramétrique doit être utilisé. Le test signé des rangs de Wilcoxon va donc être utilisé. Ce test, pendant du test de Student dans le cadre non paramétrique, va permettre de tester l'hypothèse \mathcal{H}_0 : "L'aire de la VCN mesurée sur l'électrode j avant la séance de relaxation est inférieure ou égale à celle mesurée après la séance de relaxation" contre l'hypothèse \mathcal{H}_1 : "L'aire de la VCN mesurée sur l'électrode j avant la séance de relaxation est supérieure à celle mesurée après la séance de relaxation".

On va s'intéresser aux électrodes qui rejettent le test qui vient d'être construit. Un test va être réalisé pour chaque électrode (chaque colonne de A) avec un niveau de 5%. Les p -valeurs obtenues sont proposées en Figure II.8.

Ainsi, une p -valeur obtenue par le test qui vient d'être défini est associée à chaque électrode. Lorsque cette p -valeur est inférieure au seuil de 5%, le test conclut au rejet de l'hypothèse nulle \mathcal{H}_0 . Les résultats du test montrent donc que 8 électrodes (Fz, FCz, FC1, F1, F2, F3, FC2 et FC3) rejettent le test. Ces 8 électrodes sont donc celles sur lesquelles il va falloir se concentrer. Ces électrodes sont situées au dessus du cortex pré-moteur, ce qui est cohérent avec des travaux antérieurs menées sur 24 sujets sains qui ont permis de déterminer par magnéto-

encéphalographie que le cortex prémoteur est la source cérébrale de la VCN [54].

Correction des p -valeurs : Dans l'approche qui vient d'être présentée, une série de tests multiples avec l'hypothèse \mathcal{H}_0 à un seuil de 5% a été utilisée. En réalisant une série de tests multiples avec l'hypothèse H_0 à un seuil 5% ($P(\mathcal{H}_1/\mathcal{H}_0) = 5\%$), sous le simple fait du hasard, 5% de ces électrodes pourraient être considérées discriminantes à tort, c'est-à-dire 2 électrodes. Il existe des approches permettant d'éliminer ou de contrôler ce risque.

L'approche la plus connue est celle de Bonferonni qui contrôle la probabilité de n'avoir aucun faux positif en divisant le seuil par le nombre total de tests. Le principal problème de cette approche est qu'elle est trop conservatrice, rendant l'acceptation de l'hypothèse nulle plus facile. Dans ce travail, l'approche de Benjamini et Hochberg, présentée dans [55] et fondée sur le contrôle du taux de faux positifs (False Discovery Rate, noté FDR) va être utilisée.

Le but de l'approche de Benjamini et Hochberg est de fixer le seuil α du test qui va être utilisé afin de contrôler le taux de faux positifs. Le taux de faux positifs est défini par le rapport entre le nombre de faux positifs et le nombre total de positifs (vrais positifs + faux positifs).

$$FDR_t = E(V_t/R_t)$$

où V_t représente le nombre de faux positifs et R_t désigne le nombre total de positifs.

Pour chacune des 58 électrodes, un test à un niveau α va être réalisé, correspondant au test \mathcal{H}_0 contre \mathcal{H}_1 qui a été présenté en Section 8.2.2. Ainsi, n p -valeurs (une par électrode) notées p_i sont obtenues. Ces dernières vont être classées par ordre croissant. La suite des p -valeurs classées par ordre croissant est notée $p_{(1)}, p_{(2)}, \dots, p_{(n)}$, où l'indice entre parenthèse ne représente plus le numéro de l'électrode mais le classement de la p -valeur.

Pour un seuil α fixé, la plus grande valeur de k telle que $p_{(k)} \leq \frac{\alpha k}{n}$ est recherchée.

Les k électrodes correspondant aux k premières p -valeurs $p_{(1)}, p_{(2)}, \dots, p_{(k)}$ sont déclarées discriminantes (rejetent \mathcal{H}_0) et sont donc conservées.

Les p -valeurs du test précédent peuvent donc être corrigées en utilisant l'approche de Benjamini et Hochberg. Les résultats après l'application de ce correctif sont donnés en Figure II.9.

Après l'application du correctif, plus aucune p -valeur n'est inférieure au seuil. Ceci est dû au fait qu'avant l'application du correctif, les p -valeurs inférieures au seuil étaient relativement proches de celui-ci. La correction des p -valeurs entraîne une augmentation de ces dernières et

F1	Fz	FC1	FCz	F3	F2	FC2	FC3	F4	AF3	FC4
0.23368	0.23368	0.23368	0.23368	0.25794	0.25794	0.25794	0.34157	0.38661	0.39975	0.41560
FPz	AF4	F5	FC5	FP1	C3	C1	Cz	F6	FP2	T3
0.47556	0.47556	0.47556	0.47998	0.56847	0.56847	0.57526	0.63247	0.68081	0.78294	0.78294
C5	C2	FC6	F7	F8	C4	TP7	CP5	CP3	CP1	CPz
0.78294	0.78294	0.80049	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417
T5	P5	P3	P1	P07	P05	P03	P06	P08	O1	Oz
0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417	0.84417
O2	C6	P0z	CP2	P2	Pz	P04	CP4	CP6	P4	P6
0.84417	0.85456	0.85456	0.86642	0.86642	0.88229	0.88229	0.89431	0.89431	0.89431	0.89431
T6	T4	TP8								
0.89431	0.91402	0.91402								

Figure II.9 – Liste des électrodes classées par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs sont corrigées à l’aide de la méthode de Benjamini et Hochberg.

les fait toutes passer au dessus du seuil. En l’état, cela reviendrait donc à conclure qu’aucune électrode n’a permis de calculer une aire de VCN avant la séance de relaxation significativement supérieure à celle enregistrée après la séance. Faute de mieux, l’approche choisie dans ce travail est donc de garder les 8 électrodes qui ont une p -valeur non corrigée inférieure au seuil. En effet, comme évoqué précédemment au vu du nombre de tests réalisés et du seuil du test considéré, l’utilisation des p -valeurs non corrigées entraînerait le rejet à tort de l’hypothèse \mathcal{H}_0 pour 2 électrodes en moyenne. Il s’agit donc d’un risque faible qu’il nous faut prendre.

8.2.3 Sélection des sujets par test statistique

Une fois que les électrodes sur lesquelles se fonder ont été sélectionnées, les aires des VCN enregistrées sur ces électrodes sont conservées. Une fois encore, les aires enregistrées avant et après la séance de relaxation vont être soustraites pour chaque sujet et chaque électrode. Une matrice B de taille 44×8 est alors obtenue. De manière analogue à ce qui vient d’être fait pour les électrodes, un test non paramétrique va être appliqué sur les sujets afin de déterminer si l’aire de la VCN mesurée avant la relaxation est significativement supérieure à celle mesurée après la séance de relaxation. Le test défini en Section 8.2.2 est donc utilisé sur les lignes de B . Une p -valeur est obtenue pour chaque sujet. Les résultats obtenus sont proposés en Figure II.10.

Il convient de conserver les sujets pour lesquels la p -valeur est inférieure au seuil de 5%. Ainsi, le test conclu que la séance de relaxation a eu l’effet escompté chez 25 sujets (sujets numéros 4, 6, 7, 12, 14, 15, 16, 20, 26, 27, 28, 29, 33, 36, 37, 40, 41, 42, 45, 47, 51, 52, 54, 56 et 58, les numéros étant donnés en fonction de l’ordre de passage lors des campagnes d’acquisition

Chapitre II. Acquisition de données d'Electro-EncéphaloGraphie

4	6	7	12	14	15	16	20	26	27	28
0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391
29	33	36	37	40	41	42	45	47	51	52
0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391	0.00391
54	56	58	44	24	38	13	43	34	5	8
0.00391	0.00391	0.00781	0.09766	0.32031	0.67969	0.72656	0.87500	0.94531	1	1
9	21	23	30	31	32	39	46	49	55	57
1	1	1	1	1	1	1	1	1	1	1

Figure II.10 – Liste des sujets classés par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs inférieurs au seuil de 5% ainsi que les sujets associés apparaissent en bleu.

4	6	7	12	14	15	16	20	26	27	28
0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716
29	33	36	37	40	41	42	45	47	51	52
0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716	0.00716
54	56	58	44	24	5	8	9	13	21	23
0.00716	0.00716	0.01375	0.16526	0.52199	1	1	1	1	1	1
30	31	32	34	38	39	43	46	49	55	57
1	1	1	1	1	1	1	1	1	1	1

Figure II.11 – Liste des sujets classés par ordre croissant de p -valeurs obtenues par un test signé des rangs de Wilcoxon. Les p -valeurs sont corrigées à l'aide de la méthode de Benjamini et Hochberg. Les p -valeurs inférieurs au seuil de 5% ainsi que les sujets associés apparaissent en bleu.

afin de préserver l'anonymat des sujets).

Correction des p -valeurs : De manière analogue à ce qui a été fait pour les électrodes, les p -valeurs obtenues peuvent être corrigées. Les résultats obtenus sont donnés en Figure II.11.

Cette fois ci, l'utilisation du correctif de Benjamini et Hochberg n'entraîne pas de différence. En effet, les 25 sujets initialement sélectionnés ont une p -valeur corrigée inférieure au seuil. Le fait qu'aucune modification sur l'issue du test n'intervienne ici est lié au fait que les p -valeurs non corrigées étaient très faibles et donc très inférieures au seuil (contrairement aux p -valeurs obtenues en Section 8.2.2).

Dans l'approche qui vient d'être présentée, une première série de tests statistiques est utilisée afin de sélectionner les électrodes discriminantes à partir des aires des VCN enregistrées sur les 44 sujets. Une fois que ces électrodes ont été sélectionnées, seules les aires enregistrées par ces électrodes ont été conservées. Une autre série de tests statistiques a permis de déterminer les sujets sur lesquels la séance de relaxation a eu l'effet escompté (baisse du niveau de la vigilance).

La principale critique qui peut être faite sur cette approche est la suivante : lors de la première série de tests, les électrodes sont sélectionnées à partir des aires VCN de sujets pour qui l'on n'est pas sûr que la relaxation ait fonctionné. Bien évidemment, l'idéal aurait été de pouvoir sélectionner les électrodes uniquement sur les sujets pour lesquels la séance de relaxation a effectivement fonctionné mais cette information nous était inaccessible.

De plus, lorsqu'une électrode a été conservée dans la première série de test, cela signifie que pour cette électrode, l'aire de la VCN enregistrée avant la séance de relaxation était supérieure à celle enregistrée après la séance de relaxation et ce, en conservant tous les sujets. Ainsi, la différence observée est valide à la fois pour les sujets que l'on aurait dû supprimer et pour ceux chez lesquels la relaxation a eu un effet. Ces électrodes-là doivent donc bien être sélectionnées.

8.3 Inspection visuelle de la VCN

Afin de compléter l'approche précédente, une inspection visuelle des tracés EEG moyennés est réalisée sur chaque sujet pour lesquels l'amplitude de la VCN a significativement diminué après la séance de relaxation selon les tests précédents. Cette inspection visuelle a pour but de s'assurer que les principales composantes des potentiels évoqués (N100, P300, VCN) sont bien présentes et d'étudier l'allure des courbes moyennes du sujet (avant et après relaxation). De plus, le décours temporel de ces potentiels évoqués est étudié. L'inspection visuelle est réalisée à l'aide d'une interface créée par Pierrick Legrand et Julien Clauzel et représentée en figure II.12. À partir de cette interface, l'utilisateur choisit le sujet ainsi que l'électrode qu'il souhaite visualiser.

Trois types de figures sont proposées par cette interface. La Figure II.12(a) permet de visualiser le tracé de la VCN du sujet choisi obtenue à partir des signaux EEG recueillis avant (courbe bleue) et après (courbe rouge) la séance de relaxation. Les courbes en gras correspondent aux moyennes des courbes sur tous les sujets. Cette figure permet de vérifier visuellement que, entre les deux traits verticaux jaunes, la courbe rouge est bien au dessus de la courbe bleue.

La Figure II.12(b) est une carte topographique du scalp du sujet en un instant donné (avant ou après relaxation). Cette carte permet donc de visualiser l'activité électrique enregistrée en un point temporel choisi par l'utilisateur (ici 589 ms après S1 pour le signal recueilli avant la séance de relaxation). Sur la figure, les forts potentiels apparaissent en rouge et les faibles potentiels apparaissent en bleu. La figure permet ainsi de vérifier l'apparition des principales composantes des potentiels évoqués ainsi que la localisation des électrodes au niveau desquelles ces composantes sont les plus amples. En effet, le point temporel choisi correspond à la période où la VCN est sensée se développer. On voit sur la figure qu'une composante négative

(apparaissant en bleu), correspondant à la VCN, est observable sur les électrodes pré-frontales, frontales et fronto-centrales.

La Figure II.12(c) représente le tracé entier obtenu chez le sujet pour chaque électrode (ordonnée) au cours du temps (abscisse). Le code couleur utilisé en Figure (b) est identique à celui de cette figure. Cette illustration permet d'avoir un aperçu global du signal pour le sujet considéré et de repérer un éventuel problème sur une électrode (électrode enregistrant un potentiel trop faible ou trop élevé sur toute la durée du signal par exemple).

Cette interface a permis, à une spécialiste du domaine (Mme Faïta-Aïnseba), de réaliser une exploration visuelle des tracés VCN. Une attention particulière a été portée sur les 25 sujets pour lesquels le test statistique a conclu que l'aire enregistrée sous la courbe de la VCN avant la séance de relaxation est significativement supérieure à celle enregistrée après la séance de relaxation. Sur ces 25 sujets, 19 ont été conservés après l'inspection visuelle de leur VCN. Notons que cette inspection visuelle a également été opérée sur les sujets précédemment rejetés afin de vérifier que le verdict posé était cohérent.

8.4 Résultats de l'étude de la VCN

L'étude de la VCN a été réalisée sur les 58 sujets de notre étude et 19 sujets ont été conservés pour la suite de notre étude (13 sujets de la première campagne d'acquisition et 6 sujets de la seconde). Ainsi, un nombre conséquent de sujets a été rejeté. Les causes de ces rejets sont de trois types :

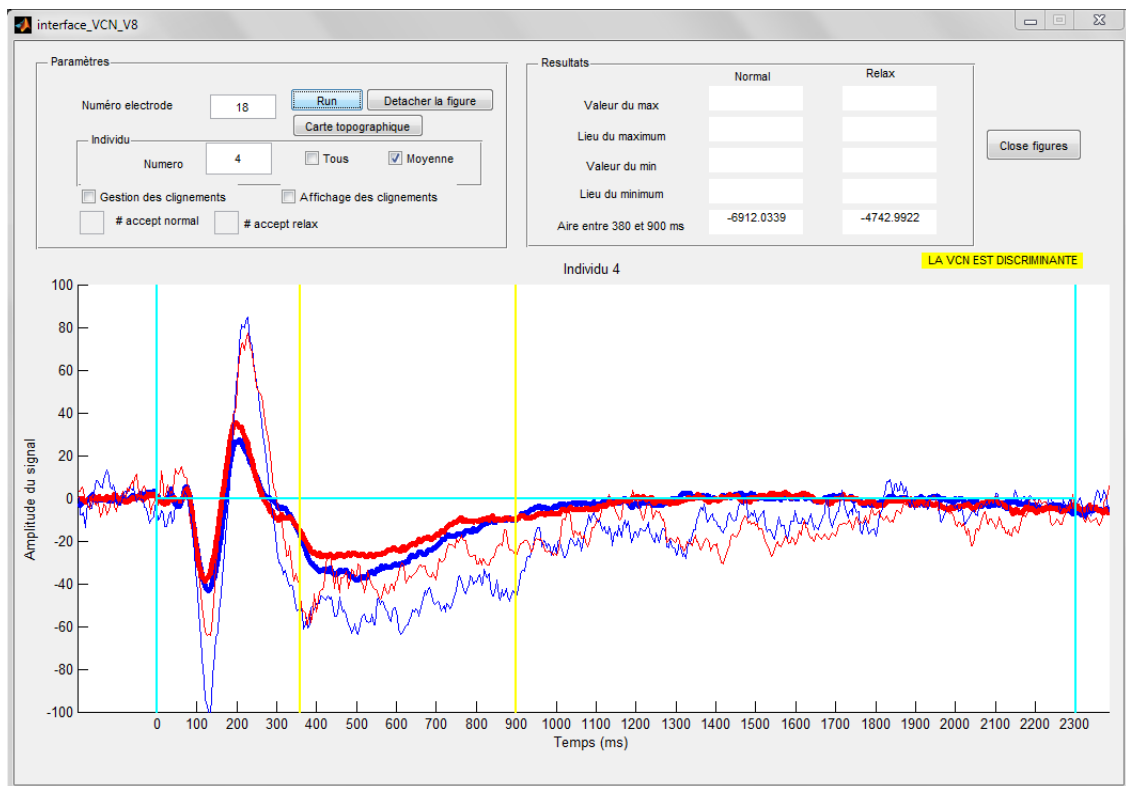
- 14 sujets du fait d'un trop grand nombre d'artefacts oculaires, rendant impossible l'extraction correct des potentiels évoqués par la méthode de moyennage des signaux recueillis à chaque essai.
- Sur les 44 sujets restants, 25 ont été conservés après l'application d'un test statistique sur les calculs d'aires des VCN.
- Enfin, la dernière cause de rejet vient de l'inspection visuelle. Celle-ci a conduit au rejet de 6 sujets.

Remarquons ici que le stress dû à l'expérience et la durée de la pose du casque à électrodes peuvent être des facteurs qui ont détérioré l'efficacité de la séance de relaxation. De plus, pour limiter la durée du port du casque, la séance de relaxation est relativement courte (20 minutes). Il est donc possible que les sujets sélectionnés par l'analyse de la VCN soient ceux qui ont réussi à se relaxer en un temps relativement court et dans des conditions qui peuvent être considérées comme stressantes.

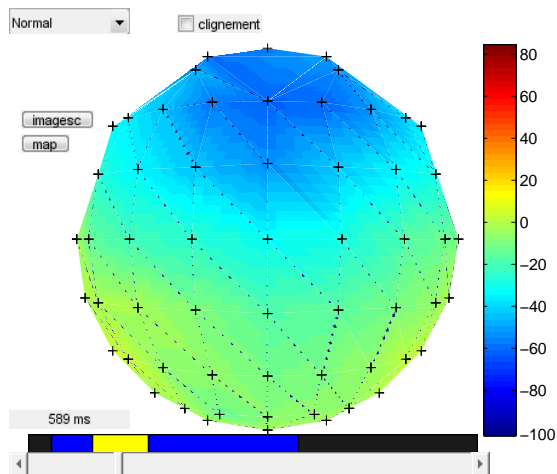
Un exemple de tracés d'un individu conservé après l'étude de la VCN est donné en [II.13](#) et un exemple de tracés d'un sujet rejeté est donné en Figure [II.14](#).

Sur ces figures, la courbe en pointillé correspond à la VCN enregistrée avant la séance de relaxation et la courbe pleine correspond à la VCN enregistrée après la séance de relaxation. Les lignes verticales pleines correspondent aux signaux avertisseurs (S1 : signal sonore, S2 : carré). L'aire entre les courbes et l'axe des abscisses est calculée entre T1 et T2 (partie des courbes encadrées par les lignes verticales en pointillé). Sur la Figure [II.13](#), le sujet est conservé car la courbe pleine est majoritairement au dessous de la courbe en pointillé entre T1 et T2 (période d'étude de la VCN). Le constat inverse peut être fait sur la Figure [II.14](#), entraînant le rejet de ce sujet.

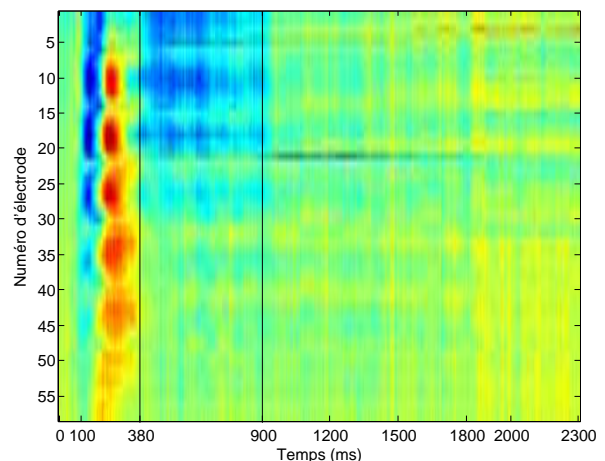
Chapitre II. Acquisition de données d'Electro-EncéphaloGraphie



(a) Visualisation des tracés de VCN d'un sujet donné obtenus à partir des signaux EEG recueilli avant (courbe bleue) et après (courbe rouge) la séance de relaxation. Les courbes en gras correspondent aux courbes moyennes des VCN des sujets.



(b) Carte topographique représentant l'activité électrique mesurée au cours du protocole de recueil de la VCN sur un sujet donné en un instant donné (avant ou après la séance de relaxation). Les forts potentiels apparaissent en rouge et les faibles potentiels apparaissent en bleu.



(c) Représentation du tracé pour chaque électrode (ordonnée) au cours du temps (abscisse). Les lignes verticales encadrent la partie du tracé sur laquelle l'aire de la VCN est calculée. Le code couleur utilisé en Figure (b) est identique à celui de cette figure.

Figure II.12 – Interface graphique réalisée pour l'inspection visuelle des tracés VCN.

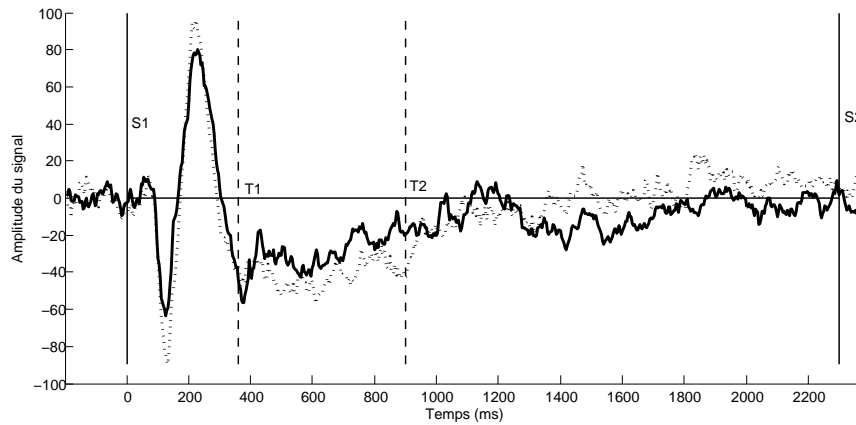


Figure II.13 – Représentation de la VCN enregistrée chez le sujet 4 de l'étude avant (ligne en pointillé) et après (ligne pleine) la séance de relaxation. Les lignes verticales pleines correspondent aux signaux avertisseurs (S1 : signal sonore, S2 : carré). Ce sujet est conservé car la courbe pleine est majoritairement au dessus de la courbe en pointillé entre T1 et T2 (période d'étude de la VCN encadrée par des lignes verticales en pointillé).

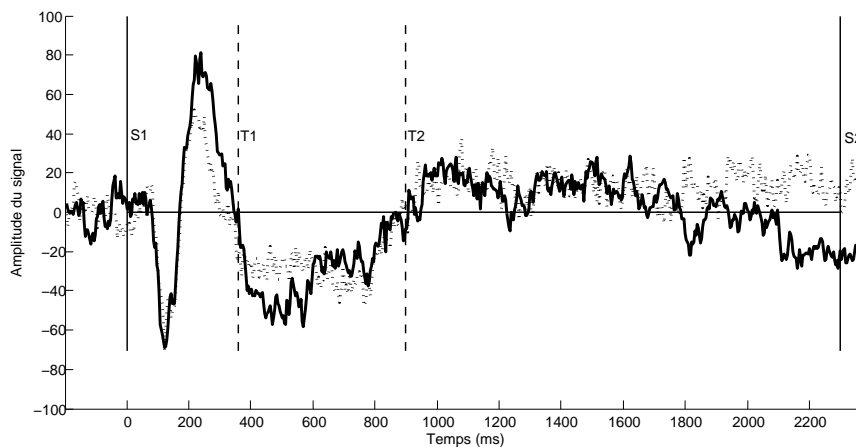


Figure II.14 – Représentation de la VCN enregistrée chez le sujet 9 de l'étude avant (ligne en pointillé) et après (ligne pleine) la séance de relaxation. Les lignes verticales pleines correspondent aux signaux avertisseurs (S1 : signal sonore, S2 : carré). Ce sujet est rejeté car la courbe pleine est majoritairement au dessous de la courbe en pointillé entre T1 et T2 (période d'étude de la VCN encadrée par des lignes verticales en pointillé).

9 Les données

Pour conclure, une représentation de la matrice 3D de données est fournie en Figure II.15. Les données sont constituées de 38 enregistrements de 3 minutes de signaux EEG bruts provenant des 19 sujets sélectionnés par l'étude de leurs VCN. Chaque enregistrement contient les variations de potentiels électriques recueillies avec une fréquence d'échantillonnage de 256 Hz (système d'acquisition Deltamed) à l'aide de 58 électrodes actives disposées sur un casque d'enregistrement (ElectroCap). En utilisant cette fréquence d'échantillonnage, chaque signal mesuré par une électrode pour un sujet dans un état de vigilance donné contient 46000 points.

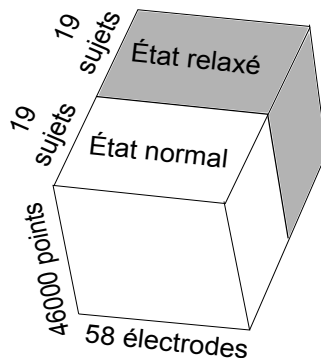


Figure II.15 – Représentation de la matrice de données en trois dimensions : une pour les sujets, une pour le temps (46000 points correspondant au nombre de points dans un signal EEG de 3 minutes mesuré en utilisant une fréquence d'échantillonnage de 256 Hz) et une pour les électrodes.

Chapitre III

État de l'art sur la classification des signaux EEG

Sommaire

1	Extraction de caractéristiques	68
1.1	Extraction de caractéristiques fondées sur l'analyse du contenu fréquentiel	68
1.1.1	La transformée de Fourier	68
1.1.2	La décomposition en ondelettes	69
1.1.3	Densité spectrale de puissance	73
1.1.4	Extraction de caractéristiques fondées sur le contenu fréquentiel	73
1.2	Extraction de caractéristiques fondées sur le filtrage spatial	74
1.2.1	Le Common Spacial Pattern (CSP)	74
1.2.2	Extraction de caractéristiques par transformation log-var	81
1.2.3	Applications en BCI	82
2	Classification de signaux EEG	82
2.1	Estimation du taux de bons classements	83
2.2	Méthode de classification	85
3	Sélection de variables en BCI	87
3.1	Sélection à l'aide d'algorithmes génétiques	87
3.1.1	Principe général d'un algorithme génétique	87
3.1.2	Utilisation en BCI	89
3.2	Sélection à l'aide d'algorithmes de recherche séquentielle	89
3.2.1	Principe général	89
3.2.2	Utilisation en BCI	90

Chapitre III. État de l'art sur la classification des signaux EEG

3.3	Le CSP parcimonieux	91
4	Conclusion	92

Les données recueillies en EEG sont généralement en 3 dimensions qui correspondent au temps, aux électrodes et aux individus (voir Figure II.15 du Chapitre II). Afin d'augmenter le nombre de signaux EEG disponibles, il est courant que le signal EEG d'un même individu soit recueilli en différents instants temporels (essais EEG) ou que le signal EEG enregistré au cours d'une période soit découpé en signaux de quelques secondes (on parle d'epochs). Les signaux EEG sont connus pour être non stationnaires, hautement bruités, irréguliers et ont tendance à varier significativement d'un individu à l'autre, faisant de l'utilisation des méthodes de classification un défi scientifique [56]. C'est pourquoi, les signaux EEG sont toujours pré-traités avant que toute analyse ne soit réalisée. L'un des objectifs de ce pré-traitement est parfois d'augmenter le rapport signal sur bruit dans les signaux afin de simplifier la tâche de classification. L'étape de pré-traitement est généralement couplée à une étape d'extraction de caractéristiques afin de passer des 3 dimensions initiales à une matrice de données en 2 dimensions sur laquelle une méthode de classification peut être appliquée.

Dans les prochains paragraphes, des méthodes de pré-traitement et d'extraction de caractéristiques vont être présentées. Par la suite, des exemples de méthodes de classification seront détaillés. Enfin, les outils existants permettant de sélectionner des variables (électrodes) seront discutés.

De nombreuses approches ont été développées afin d'extraire des caractéristiques à partir de signaux EEG. Dans ce chapitre, nous allons principalement nous concentrer sur deux approches sur lesquelles les travaux réalisés au cours de cette thèse se sont appuyés. La première approche d'extraction de caractéristiques est fondée sur l'analyse du contenu fréquentiel du signal EEG à l'aide d'outils de traitement du signal tels que la transformation de Fourier ou la transformée en ondelettes. La seconde est fondée sur la création de signaux synthétiques discriminants par combinaisons linéaires des signaux initiaux à l'aide de la méthode *Common Spatial Pattern* (CSP). Dans ce chapitre, seule la méthode de classification utilisée durant ces travaux de thèse sera détaillée. Pour avoir un aperçu plus global des méthodes de pré-traitements, d'extraction de caractéristiques et de classification utilisées sur des données EEG, le lecteur pourra se référer à [57] ou [58].

1 Extraction de caractéristiques

1.1 Extraction de caractéristiques fondées sur l'analyse du contenu fréquentiel

Dans cette partie, deux méthodes de pré-traitements fondées sur l'analyse du contenu fréquentiel vont être présentées : la transformée de Fourier et la transformée en ondelettes. Pour chacune des méthodes, la version continue et la version discrète sera discutée.

1.1.1 La transformée de Fourier

Transformée de Fourier d'un signal à temps continu : La transformée de Fourier d'un signal $x(t)$ d'énergie totale finie, à valeur dans le corps des complexes, est définie comme suit :

$$\hat{X}(f) = TF(x(t)) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft} dt$$

La transformée de Fourier d'un signal $x(t)$ étant une grandeur complexe, les spectres d'amplitude et de phase correspondent aux diagrammes représentant respectivement le module et la phase de $X(f)$ en fonction de la fréquence f .

La transformée de Fourier inverse est alors donnée par :

$$x(t) = \int_{-\infty}^{+\infty} \hat{X}(f)e^{j2\pi ft} df$$

La transformée de Fourier est une application linéaire qui permet de faire passer du domaine temporel au domaine fréquentiel et son utilisation facilite la caractérisation des systèmes continus car elle permet en particulier de transformer les équations algébriques en équations différentielles et inversement.

La transformée de Fourier d'un signal à temps discret, la transformée discrète et la FFT : La transformée de Fourier continue d'un signal à temps discret satisfait :

$$\hat{X}(f) = \sum_k x(k)e^{-jk2\pi \frac{f}{f_{ech}}} = \sum_k x(k)e^{-jk2\pi f_r}$$

où f_{ech} est la fréquence d'échantillonnage et où f_r désigne la fréquence normalisée. On traite plus particulièrement le cas où l'on dispose d'un nombre fini d'échantillons d'un signal discret. Dans la pratique, il n'est pas réaliste d'effectuer une somme infinie de termes et on sera amené

à effectuer la transformée de Fourier sur un nombre fini N d'échantillons d'un signal discret. La relation ci-dessus se ramène donc à :

$$\hat{X}(f_r) = \sum_{k=0}^{N-1} x(k)e^{-jk2\pi f_r}$$

Pour une fréquence normalisée f_r donnée, la transformée de Fourier $X(f_r)$ d'une séquence $x(k)$ est donc définie comme le produit scalaire entre le signal et les éléments de la base orthogonale $\left\{e^{jk2\pi f_r}\right\}_{k=0,\dots,N-1}$.

Cependant, la fréquence normalisée est une variable continue, ce qui présente des difficultés quand on veut implanter cette transformation. Pour cette raison, il a fallu chercher une autre transformée qui n'a pas ces inconvénients. Il s'agit de la Transformée de Fourier Discrète (TFD) sur N points, à partir de N échantillons d'un signal à temps discret, définie comme suit :

$$\hat{X}(n) = \sum_{k=0}^{N-1} x(k)e^{-jk\frac{2\pi n}{N}}$$

Elle coïncide avec la transformée de Fourier du signal à temps discret aux fréquences $f_r = \frac{n}{N}$ avec n variant de $-\frac{N}{2}$ à $\frac{N}{2} - 1$ et elle s'annule pour les autres fréquences. La transformée de Fourier discrète est une fonction de l'indice n et de N . Par souci de simplicité, nous poserons $X(n)$ pour désigner la valeur de la transformée de Fourier discrète à la fréquence normalisée $\frac{n}{N}$. Du fait de la discrétisation de la fréquence, la transformée de Fourier discrète inverse est obtenue comme suit :

$$x(k) = \frac{1}{N} \sum_{n=-N/2}^{N/2-1} \hat{X}(n)e^{jk\frac{2\pi n}{N}}$$

Cooley et Tukey ont proposé en 1965 un algorithme rapide pour calculer la transformée de Fourier discrète (FFT pour *Fast Fourier Transform*). En effet, dans le cas d'un signal réel, le calcul direct de la relation ci-dessus nécessite $2N^2$ multiplications et $2N(2N - 1)$ additions. Dans le cas d'un signal complexe, la complexité calculatoire atteint $4N^2$ multiplications et $2N(2N - 1)$ additions. La méthode rapide consiste à opérer par dichotomie, ce qui ramène à une complexité calculatoire en $N \log_2(N)$.

1.1.2 La décomposition en ondelettes

La décomposition en ondelettes [59], [60] est une méthode très utilisée en analyse du signal. Son principal avantage est de permettre d'analyser l'évolution du contenu fréquentiel d'un signal dans le temps. Elle est donc plus adaptée que la transformée de Fourier pour l'analyse

des signaux non stationnaires.

Une ondelette est une fonction $\psi \in L^2(\mathbb{R})$ telle que $\int_{\mathbb{R}} \psi(t)dt = 0$. La transformée en ondelettes continue d'un signal x peut s'écrire :

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

où a est appelé facteur d'échelle et représente l'inverse de la fréquence du signal, b est un terme de translation temporelle et la fonction ψ est appelée ondelette mère. Le terme ψ^* est le complexe conjugué de ψ . L'ondelette mère est généralement une fonction continue et différentiable à support compact. Il existe plusieurs ondelettes mères telles que les ondelettes de Daubechies ou Coiflets, par exemple.

Quelques ondelettes sont données en Figure III.1.

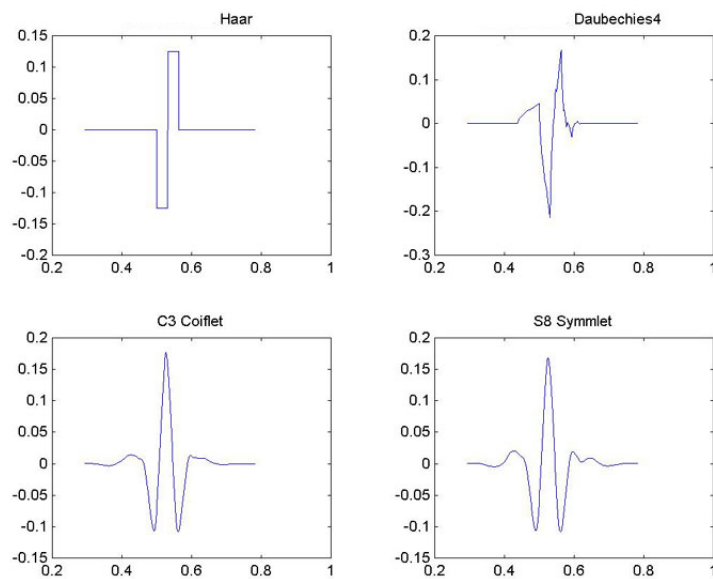


Figure III.1 – Représentation de quelques ondelettes.

Il est également possible de définir la transformée en ondelettes discrète en partant de la précédente formule et en discrétisant les paramètres a et b . Ainsi, soit $a = a_0^j$, où a_0 est le paramètre de résolution tel que $a_0 > 1$ et $j \in \mathbb{N}$ et soit $b = kb_0a_0^j$, où $k \in \mathbb{N}$ et $b_0 > 0$. Il est très fréquent de considérer la transformée en ondelettes dite « dyadique » qui correspond au cas où $a_0 = 2$ et $b_0 = 1$. Dans ce cas, $j = 1, 2, \dots, n$, où n est le logarithme en base 2 du nombre de points formant le signal et $k = 1, 2, \dots, 2^{j-1}$. La transformée en ondelettes discrète dyadique s'écrit :

$$c_{j,k} = 2^{-\frac{j}{2}} \int_{-\infty}^{\infty} x(t)\psi(2^{-j}t - k)dt$$

où j est le niveau de décomposition (ou l'échelle) et k le décalage temporel. Le nombre de niveaux de décomposition maximal, n , correspond au \log_2 du nombre de points formant le signal. La transformée en ondelettes discrète est plus rapide que la version continue en permettant tout de même une reconstruction exacte du signal initial par transformation inverse. La grille dyadique, représentée en Figure III.2, donne une représentation spatio-fréquentielle de la transformée en ondelettes discrète dyadique. Sur cette figure, l'axe des abscisses correspond au temps et celui des ordonnées aux fréquences. Les ronds noirs correspondent aux coefficients d'ondelettes $c_{j,k}$. Les points du signal sont représentés en dessous du niveau le plus bas de décomposition par des cercles. À chaque niveau supplémentaire (échelle supplémentaire), la fréquence est multipliée par deux.

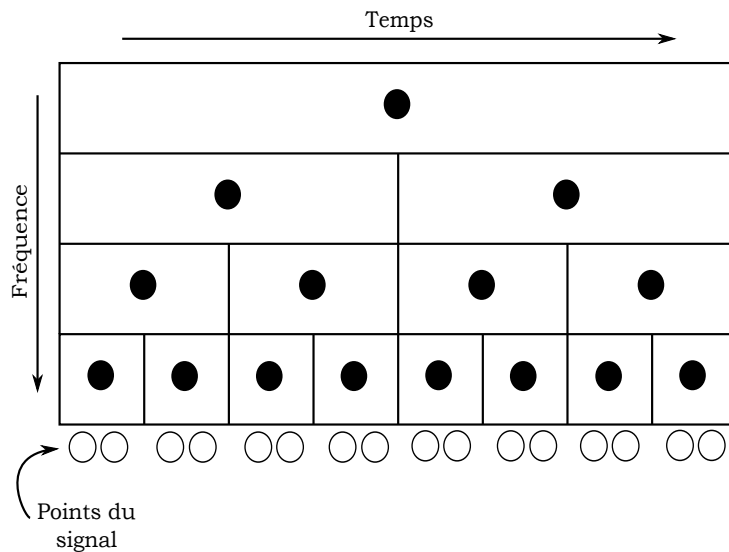
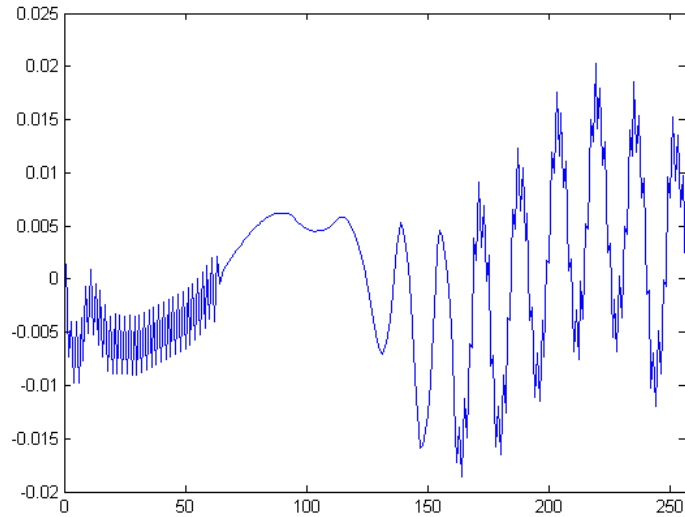
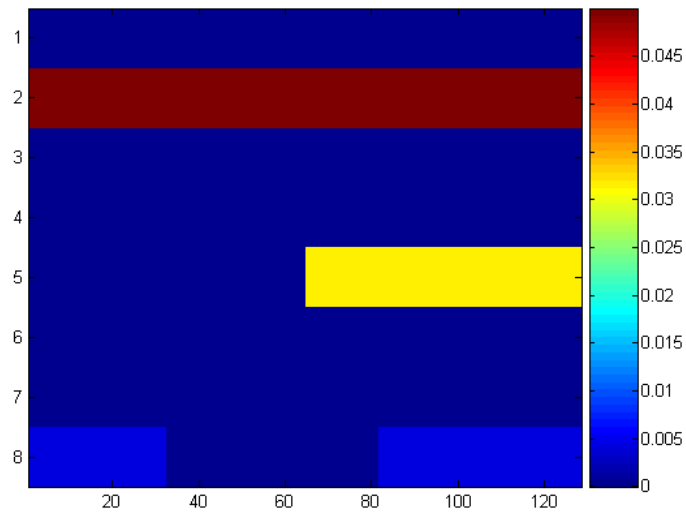


Figure III.2 – Représentation de la grille dyadique pour 4 niveaux de décomposition (4 échelles) correspondant à la décomposition en ondelettes discrète d'un signal x de 2^4 points.

La grille dyadique permet de visualiser le contenu fréquentiel d'un signal et de voir à quel moment ces fréquences apparaissent. Par exemple, la Figure III.3 représente un signal et sa décomposition en ondelettes discrète réalisée à l'aide de la toolbox Fraclab [61], <http://fraclab.saclay.inria.fr/> (voir Figure III.4). La grille dyadique, représentée en Figure III.3(b), contient les valeurs absolues des coefficients d'ondelettes $c_{j,k}$. Les fortes valeurs absolues de ces coefficients apparaissent en rouge et les faibles valeurs en bleu. Sur la Figure III.3(b), le deuxième niveau de décomposition, relié aux faibles fréquences, contient de grandes valeurs



(a) Figure représentant un signal synthétique.



(b) Figure représentant la grille dyadique associée à la décomposition en ondelettes discrète du signal. La grille contient les valeurs absolues des coefficients d'ondelettes $c_{j,k}$. Les forts coefficients apparaissent en rouge et les faibles coefficients en bleu.

Figure III.3 – Représentation d'un signal (a) et de la grille dyadique de sa décomposition en ondelettes discrète (b). Les figures sont obtenues à l'aide de la toolbox Fraclab.

absolues de coefficients sur l'ensemble du signal. La cinquième échelle contient des valeurs moyennement élevées pour les coefficients sur la dernière partie du signal. Enfin, la dernière

échelle permet de visualiser les hautes fréquences contenues au début et à la fin du signal.

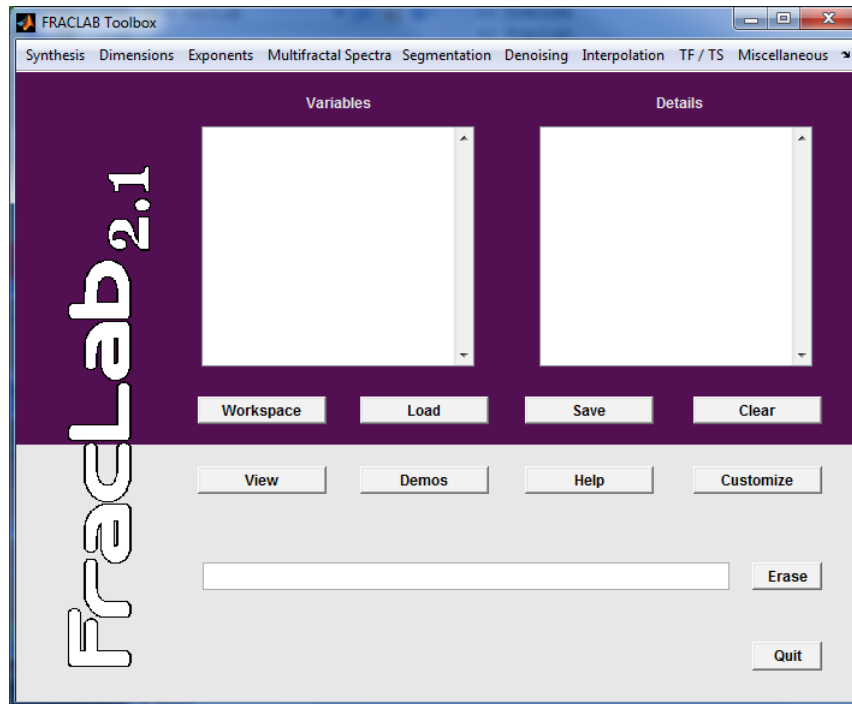


Figure III.4 – L’interface graphique pour l’utilisateur de la toolbox Fraclab.

1.1.3 Densité spectrale de puissance

La densité spectrale de puissance (notée PSD pour *Power Spectral Density*) décrit la façon dont la puissance d’un signal est répartie sur les différentes fréquences. Elle s’obtient à partir de la transformée de Fourier $\hat{X}(f)$ du signal $x(t)$ à l’aide de la relation :

$$PSD(f) = \frac{|\hat{X}(f)|^2}{T}$$

où T représente le temps d’intégration et où $|\cdot|$ représente le module.

La densité spectrale de puissance est très utilisée pour l’extraction de caractéristiques en EEG.

1.1.4 Extraction de caractéristiques fondées sur le contenu fréquentiel

Lorsqu’une transformée de Fourier est utilisée comme pré-traitement, l’extraction de caractéristiques se fait généralement en extrayant les puissances spectrales de d bandes de fréquences qui correspondent à des ondes utilisées dans la littérature sur les signaux EEG (ondes α , β , δ ,

θ et γ définies dans la Section 3.1.1.5 du Chapitre I). Ces puissances spectrales sont extraites pour chaque électrode et chez chaque individu de l'étude. La matrice des caractéristiques, de dimensions $n \times d * p$ où p représente le nombre d'électrodes et n le nombre d'individus de l'étude, est composée des puissances pour chaque électrode initiale. Cette approche est par exemple utilisée dans [62], [63] ou encore [64]. Dans [65], l'approche est un peu différente de celle qui vient d'être présentée. En effet, le signal EEG est décomposé en 23 bandes de 1 Hz (de 1 à 23Hz). Ainsi, ce ne sont pas les puissances des principales bandes de fréquences connues en EEG qui sont extraites mais celles des bandes de fréquences unitaires de 1 à 23 Hz.

Lorsqu'une décomposition en ondelettes discrète est utilisée, les coefficients de celle-ci peuvent être utilisés comme des caractéristiques. Par exemple, [66] propose d'entraîner une méthode de classification (réseau de neurones artificiel) à partir des coefficients d'ondelettes afin de différencier les signaux EEG provenant d'individus sains, atteints de schizophrénie ou de troubles obsessionnels compulsifs.

1.2 Extraction de caractéristiques fondées sur le filtrage spatial

Soit X une matrice de données EEG de dimensions $T \times p$ où p est le nombre d'électrodes actives et T est le nombre de points de mesure. Des vecteurs $\mathbf{w} \in \mathbb{R}^p$, appelés ensuite filtres spatiaux, sont utilisés pour construire un faible nombre de signaux synthétiques définis comme des combinaisons linéaires des colonnes de X . Le signal synthétique construit avec le filtre \mathbf{w} est donné par la relation $\mathbf{f} = X\mathbf{w}$. Le vecteur $\mathbf{f} \in \mathbb{R}^T$ peut aussi s'interpréter comme les coordonnées des projections des T lignes de X sur l'axe de vecteur directeur \mathbf{w} .

L'analyse en Composantes Principales (ACP) permet de trouver les vecteurs \mathbf{w} (loadings) et \mathbf{f} (composantes principales) dans le contexte des problèmes de classification non supervisées. Dans le contexte de la classification de signaux EEG, la méthode du *Common Spatial Pattern* (CSP) est couramment utilisée pour obtenir les vecteurs \mathbf{w} , appelés filtres spatiaux et \mathbf{f} , que nous appellerons les composantes CSP (par analogie avec l'ACP).

1.2.1 Le Common Spacial Pattern (CSP)

La méthode CSP est une méthode utilisée pour trouver des filtres spatiaux en classification binaire supervisée [67] [68]. Le but de la méthode CSP est de trouver des signaux synthétiques $\mathbf{f} = X\mathbf{w}$ de variance maximale dans une condition et simultanément minimale dans l'autre condition. Les variances de ces signaux synthétiques sont alors utilisées pour entraîner une méthode de classification (classifier).

Un essai représente l'enregistrement EEG effectué durant une période de temps T . Soit $\{1, \dots, i, \dots, n\}$ un ensemble de n essais dont n_1 sont dans la condition 1 (classe 1) et n_2 sont dans la condition 2 (classe 2). Soit X_i , $i = 1, \dots, n$ les matrices des données EEG de dimensions $T \times p$ correspondant aux signaux EEG recueillis pendant ces essais. Les matrices X_i sont supposées centrées (de moyenne nulle). Les matrices des données EEG des essais de la classe 1 sont concaténées en ligne formant ainsi la matrice \mathbf{X}_1 de dimensions $Tn_1 \times p$. De la même manière, la matrice \mathbf{X}_2 de dimensions $Tn_2 \times p$ est obtenue en concaténant les essais de la classe 2.

On peut maintenant définir \mathbf{C}_1 et \mathbf{C}_2 , les matrices de variances-covariances empiriques des signaux EEG dans les deux conditions :

$$\mathbf{C}_c = \frac{1}{Tn_c} \mathbf{X}_c^T \mathbf{X}_c = \frac{1}{n_c} \sum_{i \in \mathcal{J}_c} \frac{1}{T} X_i^T X_i = \frac{1}{n_c} \sum_{i \in \mathcal{J}_c} C_i$$

où $c \in \{1, 2\}$, \mathcal{J}_c est l'ensemble des indices correspondant aux essais dans la classe c et X_i^T est la transposée de la matrice X_i . La matrice de covariance \mathbf{C}_c est la moyenne des matrices de covariance C_i des essais de la classe c . Dans ce travail, \mathbf{X}_1 et \mathbf{X}_2 sont supposées inversibles. Ainsi, \mathbf{C}_1 et \mathbf{C}_2 sont définies positives et inversibles.

Les principales notations ayant été introduites, le problème d'optimisation du CSP va être présenté et sa résolution va être détaillée. Il s'agit d'une approche originale et complète du problème du CSP qui, à notre connaissance, n'existait pas dans la littérature. Les propositions ainsi que les preuves, nécessaires à la compréhension complète de la résolution du problème d'optimisation, seront également proposées.

Recherche du premier ensemble de filtres spatiaux : Dans un premier temps, on recherche les filtres $\mathbf{w} \in \mathbb{R}^p$ qui maximisent $var(\mathbf{X}_1 \mathbf{w})$ et minimisent $var(\mathbf{X}_2 \mathbf{w})$. Par définition, $var(\mathbf{X}_c \mathbf{w}) = \frac{1}{Tn_c} \mathbf{w}^T \mathbf{X}_c^T \mathbf{X}_c \mathbf{w} = \mathbf{w}^T \mathbf{C}_c \mathbf{w}$ pour $c = 1, 2$. Ainsi, le problème d'optimisation s'écrit :

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \quad (\text{III.1})$$

La résolution de ce problème d'optimisation est obtenue à l'aide de la proposition suivante :

Proposition 1.1 (*Chavent, Legrand et Vézard*)

Soient A et B deux matrices symétriques et inversibles. Soit \mathbf{w} un vecteur de \mathbb{R}^p . Alors, la

solution du problème d'optimisation suivant :

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T A \mathbf{w}}{\mathbf{w}^T B \mathbf{w}}$$

est le vecteur propre de $B^{-1}A$ associé à la plus grande valeur propre.

Preuve :

On recherche

$$\max_{\mathbf{w} \in \mathbb{R}^p} f(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T A \mathbf{w}}{\mathbf{w}^T B \mathbf{w}}.$$

Si $\left. \frac{\partial f}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_1} = 0$, alors \mathbf{w}_1 est un optimum local de f . On a :

$$\left. \frac{\partial f}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_1} = 0$$

$$\Leftrightarrow \frac{\left(\frac{\partial}{\partial \mathbf{w}_1} (\mathbf{w}_1^T A \mathbf{w}_1) \right) (\mathbf{w}_1^T B \mathbf{w}_1) - (\mathbf{w}_1^T A \mathbf{w}_1) \left(\frac{\partial}{\partial \mathbf{w}_1} (\mathbf{w}_1^T B \mathbf{w}_1) \right)}{(\mathbf{w}_1^T B \mathbf{w}_1)^2} = 0$$

$$\Leftrightarrow 2A\mathbf{w}_1 (\mathbf{w}_1^T B \mathbf{w}_1) - (\mathbf{w}_1^T A \mathbf{w}_1) 2B\mathbf{w}_1 = 0$$

$$\Leftrightarrow A\mathbf{w}_1 = \frac{\mathbf{w}_1^T A \mathbf{w}_1}{\mathbf{w}_1^T B \mathbf{w}_1} B\mathbf{w}_1$$

$$\Leftrightarrow B^{-1}A\mathbf{w}_1 = \frac{\mathbf{w}_1^T A \mathbf{w}_1}{\mathbf{w}_1^T B \mathbf{w}_1} \mathbf{w}_1$$

Ainsi, \mathbf{w}_1 est un vecteur propre de $B^{-1}A$ et $\frac{\mathbf{w}_1^T A \mathbf{w}_1}{\mathbf{w}_1^T B \mathbf{w}_1}$ est la valeur propre associée à ce vecteur. On veut maximiser f donc on choisit le vecteur propre de $B^{-1}A$ associé à la plus grande valeur propre λ_1 où $\lambda_1 = \frac{\mathbf{w}_1^T A \mathbf{w}_1}{\mathbf{w}_1^T B \mathbf{w}_1}$. ■

En utilisant la Proposition 1.1, la solution \mathbf{w}_1 de (III.1) est le premier vecteur propre de $\mathbf{C}_2^{-1}\mathbf{C}_1$, i.e. le vecteur propre associé à la plus grande valeur propre λ_1 où $\lambda_1 = \frac{\mathbf{w}_1^T \mathbf{C}_1 \mathbf{w}_1}{\mathbf{w}_1^T \mathbf{C}_2 \mathbf{w}_1}$.

En considérant les k premiers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$ (i.e. correspondant aux k plus grandes valeurs propres), on obtient un ensemble de k filtres spatiaux $\mathbf{w}_1, \dots, \mathbf{w}_k$ maximisant (III.1) sous la contrainte $\mathbf{w}_j^T \mathbf{C}_2 \mathbf{w}_l = 0, \forall l \neq j$ (les filtres sont \mathbf{C}_2 orthogonaux).

Recherche du second ensemble de filtres spatiaux : Dans un deuxième temps, les filtres spatiaux $\mathbf{v} \in \mathbb{R}^p$ qui maximisent $var(\mathbf{X}_2 \mathbf{v})$ pendant que $var(\mathbf{X}_1 \mathbf{v})$ est minimale sont recherchés. Ce problème d'optimisation s'écrit :

$$\max_{\mathbf{v} \in \mathbb{R}^p} \frac{\mathbf{v}^T \mathbf{C}_2 \mathbf{v}}{\mathbf{v}^T \mathbf{C}_1 \mathbf{v}} \quad (\text{III.2})$$

De manière analogue à ce qui vient d'être présenté, \mathbf{v}_1 est le vecteur propre de $\mathbf{C}_1^{-1}\mathbf{C}_2$ associé à la plus grande valeur propre $\mu_1 = \frac{\mathbf{v}_1^T \mathbf{C}_2 \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{C}_1 \mathbf{v}_1} = \frac{1}{\lambda_1}$. Les k premiers vecteurs propres de $\mathbf{C}_1^{-1}\mathbf{C}_2$ donnent un second ensemble de k filtres spatiaux $\mathbf{v}_1, \dots, \mathbf{v}_k$ maximisant (III.2) sous la contrainte $\mathbf{v}_j^T \mathbf{C}_1 \mathbf{v}_l = 0, \forall l \neq j$ (filtres \mathbf{C}_1 orthogonaux).

Une unique décomposition en éléments propres : Il vient d'être montré que les deux ensembles de filtres spatiaux $\mathbf{w}_1, \dots, \mathbf{w}_k$ et $\mathbf{v}_1, \dots, \mathbf{v}_k$ sont obtenus en calculant la décomposition en éléments propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$ et $\mathbf{C}_1^{-1}\mathbf{C}_2$. Cependant, les k premiers vecteurs propres de $\mathbf{C}_1^{-1}\mathbf{C}_2$ sont aussi les k derniers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$. Afin de le prouver, il va être démontré, dans un premier temps, que $\mathbf{C}_1^{-1}\mathbf{C}_2$ et $\mathbf{C}_2^{-1}\mathbf{C}_1$ sont des matrices définies positives, obtenant ainsi l'assurance que ces deux matrices ont des valeurs propres strictement positives. Dans un deuxième temps, il va être démontré qu'une matrice et son inverse ont les mêmes vecteurs propres associés à des valeurs propres inversées. Ainsi, en s'appuyant sur le fait que $\mathbf{C}_2^{-1}\mathbf{C}_1$ est l'inverse de $\mathbf{C}_1^{-1}\mathbf{C}_2$ et sur le fait que la fonction inverse est strictement décroissante sur \mathbb{R}^{+*} , il sera montré que les k premiers vecteurs propres de $\mathbf{C}_1^{-1}\mathbf{C}_2$ sont aussi les k derniers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$.

Proposition 1.2 *Soient A et B deux matrices symétriques et définies positives. Ainsi, AB est une matrice définie positive.*

Preuve :

Comme A une matrice symétrique définie positive, il existe une matrice symétrique définie positive C telle que $A = C^2$. D'où, $AB = C^2B = CCBC C^{-1} = C(CBC)C^{-1} = CDC^{-1}$ avec $D = CBC$. Ainsi, les matrices AB et D sont semblables et ont donc les mêmes valeurs propres. Or, on peut montrer que D est symétrique définie positive. En effet, $\forall \mathbf{x} \in \mathbb{R}^n$, on a $\mathbf{x}^T D \mathbf{x} = \mathbf{x}^T C B C \mathbf{x} = \mathbf{x}^T C^T B C \mathbf{x} = \tilde{\mathbf{x}}^T B \tilde{\mathbf{x}}$, où $\tilde{\mathbf{x}} = C \mathbf{x} \in \mathbb{R}^n$. Or, $\tilde{\mathbf{x}}^T B \tilde{\mathbf{x}} > 0$ car B est symétrique définie positive.

Ainsi, comme D , AB a toutes ses valeurs propres réelles et strictement positives. ■

Proposition 1.3 *Soit A une matrice définie positive de dimensions $p \times p$. Soient $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ ses vecteurs propres respectivement associés aux valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. Alors, A^{-1} , l'inverse de A , a pour vecteurs propres $\mathbf{x}_p, \mathbf{x}_{p-1}, \dots, \mathbf{x}_1$ respectivement associés aux valeurs propres $\frac{1}{\lambda_p} \geq \frac{1}{\lambda_{p-1}} \geq \dots \geq \frac{1}{\lambda_1} > 0$.*

Preuve :

A est définie positive et donc A est inversible. De plus, $\forall i = 1, \dots, p$, on note \mathbf{x}_i le vecteur propre de A associé à la valeur propre λ_i . Par définition, il vient que :

$$\begin{aligned} A\mathbf{x}_i &= \lambda_i \mathbf{x}_i \\ \Leftrightarrow A^{-1}A\mathbf{x}_i &= A^{-1}\lambda_i \mathbf{x}_i \\ \Leftrightarrow \mathbf{x}_i &= \lambda_i A^{-1}\mathbf{x}_i \\ \Leftrightarrow \frac{1}{\lambda_i} \mathbf{x}_i &= A^{-1}\mathbf{x}_i \end{aligned}$$

On en déduit que \mathbf{x}_i est aussi le vecteur propre de A^{-1} associé à la valeur propre $\frac{1}{\lambda_i}$. ■

Ainsi, en utilisant la Proposition 1.2 sur le produit $\mathbf{C}_2^{-1}\mathbf{C}_1$, il vient que la matrice obtenue est définie positive et possède donc des valeurs propres strictement positives. Ce point satisfait les hypothèses nécessaires à l'utilisation de la Proposition 1.3. Celle-ci permet de prouver que les k premiers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$ sont aussi les k derniers vecteurs propres de $\mathbf{C}_1^{-1}\mathbf{C}_2$.

Ainsi, on obtient la relation suivante entre les vecteurs propres des deux matrices :

$$\left\{ \begin{array}{l} \mathbf{w}_1 = \mathbf{v}_p \\ \vdots \\ \mathbf{w}_k = \mathbf{v}_{p-k} \end{array} \right. \quad (\text{III.3})$$

Le CSP se résume donc à une unique décomposition en éléments propres. Les filtres sont obtenus en conservant les k premiers et les k derniers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$. Ils permettent de construire k paires de filtres $(\mathbf{w}_j, \mathbf{v}_j)_{j=1, \dots, k}$. La matrice des filtres W est définie par $W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k]$.

Projection : Une fois que la matrice des filtres W est obtenue, les matrices des données EEG X_i , $i = 1, \dots, n$ des n essais sont projetées linéairement et n matrices $F_i = X_i W$ de dimensions $T \times 2k$ sont ainsi obtenues. Les colonnes de F_i sont les $2k < p$ signaux synthétiques de l'essai i .

Remarques : Notons que le CSP est parfois présenté comme la décomposition en éléments propres de la matrice $(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{C}_1$. Dans ce cas, le problème d'optimisation s'écrit :

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}} \quad (\text{III.4})$$

Avec,

$$\frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}} = \frac{\text{var}(\mathbf{X}_1 \mathbf{w})}{\text{var}(\mathbf{X}_1 \mathbf{w}) + \text{var}(\mathbf{X}_2 \mathbf{w})} \quad (\text{III.5})$$

Cela revient comme avant à maximiser $\text{var}(\mathbf{X}_1 \mathbf{w})$ et à minimiser $\text{var}(\mathbf{X}_2 \mathbf{w})$ mais le critère varie dans $[0, 1]$ et s'interprète plus facilement (ratio de variances). On peut montrer que les vecteurs propre de $(\mathbf{C}_2 + \mathbf{C}_1)^{-1} \mathbf{C}_1$, solutions de (III.4), sont aussi solutions de (III.1) mais sont associés à des valeurs différentes $\beta_i = \frac{\lambda_i}{\lambda_i + 1}$.

Proposition 1.4 (*Chavent, Legrand et Vézard*)

Soient \mathbf{C}_1 et \mathbf{C}_2 des matrices symétriques et définies positives de dimensions $p \times p$. Alors, les matrices $(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{C}_1$ et $\mathbf{C}_2^{-1} \mathbf{C}_1$ ont les mêmes vecteurs propres associés à des valeurs propres différentes.

Preuve :

$\forall i = 1, \dots, p$, on note \mathbf{u}_i le vecteur propre de $\mathbf{C}_2^{-1}\mathbf{C}_1$ associé à la valeur propre λ_i . Par définition, il vient :

$$\begin{aligned}
 \mathbf{C}_2^{-1}\mathbf{C}_1\mathbf{u} &= \lambda_i\mathbf{u}_i \\
 \Leftrightarrow \frac{1}{\lambda_i}\mathbf{u}_i &= \mathbf{C}_1^{-1}\mathbf{C}_2\mathbf{u}_i \\
 \Leftrightarrow \left(\frac{1}{\lambda_i} - 1 + 1\right)\mathbf{u}_i &= \mathbf{C}_1^{-1}\mathbf{C}_2\mathbf{u}_i \\
 \Leftrightarrow \frac{1 + \lambda_i}{\lambda_i}\mathbf{u}_i &= (I + \mathbf{C}_1^{-1}\mathbf{C}_2)\mathbf{u}_i \\
 \Leftrightarrow \frac{1 + \lambda_i}{\lambda_i}\mathbf{C}_1\mathbf{u}_i &= (\mathbf{C}_1 + \mathbf{C}_2)\mathbf{u}_i \\
 \Leftrightarrow (\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{C}_1\mathbf{u}_i &= \frac{\lambda_i}{1 + \lambda_i}\mathbf{u}_i
 \end{aligned}$$

Ainsi, \mathbf{u}_i est aussi le vecteur propre de $(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{C}_1$ associé à la valeur propre $\beta_i = \frac{\lambda_i}{1 + \lambda_i}$. ■

Dans la proposition 1.4, on remarque que les valeurs propres β_i et λ_i sont rangées dans le même sens. Ainsi, si l'on note \tilde{W} obtenue en conservant les k premiers et les k derniers vecteurs propres de la matrice $(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{C}_1$, on obtient que $\tilde{W} = W$

Exemple d'application du CSP : Afin d'illustrer les effets du CSP sur les données EEG, celui-ci est appliqué sur des données issues de la compétition BCI III [13]. Le jeu de données utilisé est le IIIa [12]. Il correspond aux signaux EEG de 3 sujets mesurés à l'aide de 60 électrodes pendant la réalisation de mouvements imaginés de quatre types (main droite, main gauche, langue et pieds). Nous appliquons ici le CSP aux signaux EEG du sujet 2, mesurés durant le premier essai d'un mouvement imaginé de la main droite et de la main gauche. Chaque essai contient 500 points de mesure. On applique donc la méthode CSP à deux tableaux de signaux EEG de dimensions 500×60 afin d'obtenir une paire de filtres spatiaux. La Figure III.5 représente les signaux bruts du sujet 2 sur les électrodes F6 et Pz. Elle représente également les signaux projetés à l'aide de la première paire de filtres CSP. Les courbes en rouge et en noir correspondent respectivement à l'essai enregistré durant le mouvement imaginé de la main droite et de la main gauche.

Ces graphiques montrent que les signaux bruts ont une variance qui paraît homogène sur les deux conditions. Cependant, une fois que les signaux sont projetés à l'aide des filtres CSP,

III.1 Extraction de caractéristiques

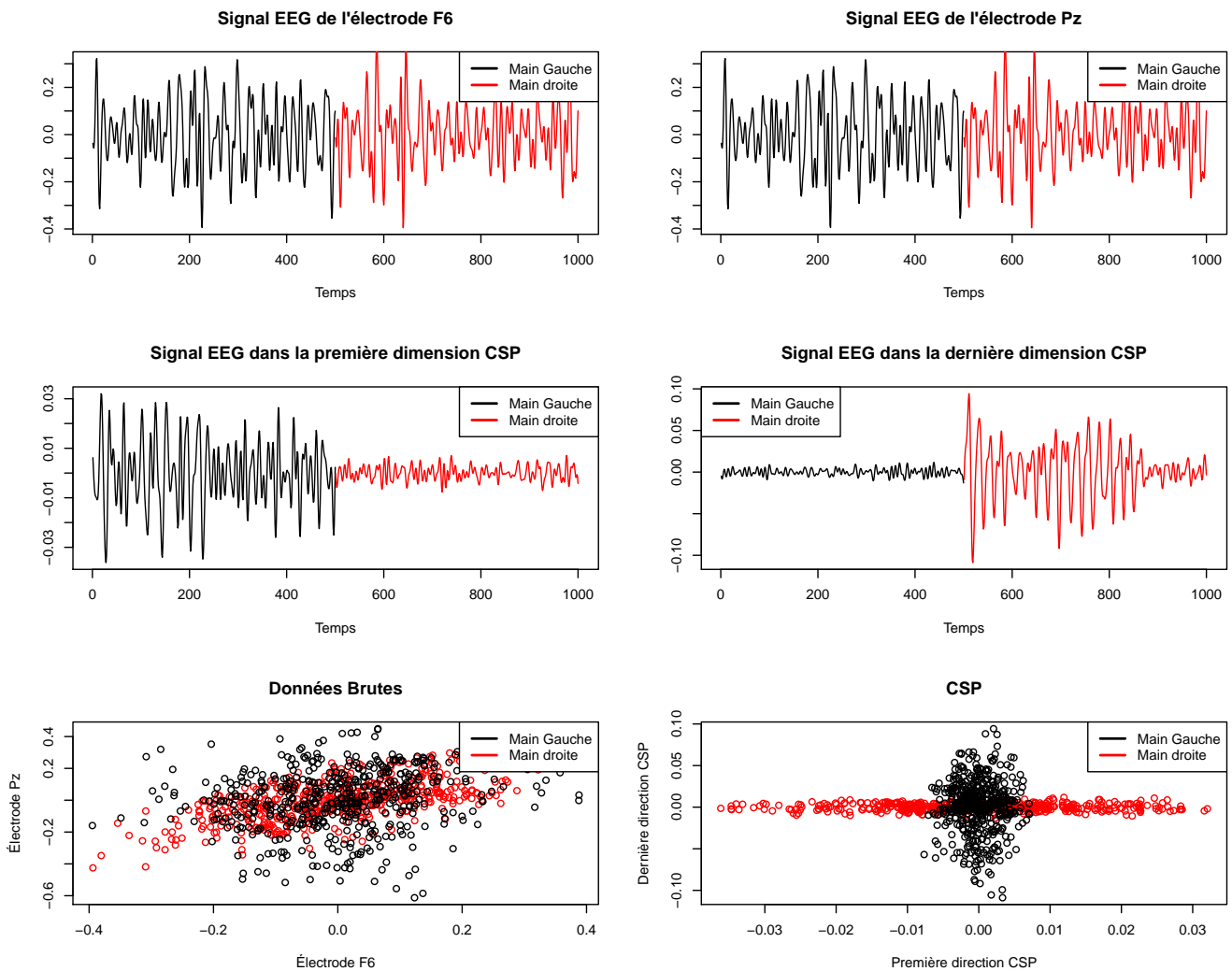


Figure III.5 – Représentation des signaux EEG bruts recueillis par les électrodes F6 et Pz. L’utilisation du CSP permet de représenter les signaux projetés à l’aide du premier et du dernier filtre CSP obtenu. Les courbes en rouge et en noir correspondent respectivement à l’essai enregistré durant le mouvement imaginé de la main droite et de la main gauche. Ces graphiques sont obtenus à partir du jeu de données IIIa [12] de la compétition BCI III [13].

les variances ne sont plus homogènes. En effet, lorsque la variance du signal projeté de l’une des conditions est maximale, la variance du signal projeté de l’autre condition est minimale.

1.2.2 Extraction de caractéristiques par transformation log-var

Les filtres CSP ont été construits pour que la variance des signaux projetés soit grande pour l’une des deux classes et petite dans l’autre. Pour chaque essai / état i ($i = 1, \dots, n$), on calcule donc la variance $var(X_i \mathbf{w}_j)$ pour construire une nouvelle matrice, notée Z , de dimensions

$n \times 2k$ avec $z_{ij} = \log(\text{var}(X_i \mathbf{w}_j))$. Les colonnes de Z doivent donc permettre de discriminer les deux états. La transformation logarithmique est utilisée afin d'approximer une distribution gaussienne des données [68]. L'utilisation de la transformation, dite « log-var » est très peu coûteuse en temps de calcul et permet d'obtenir une matrice en 2 dimensions sur laquelle appliquer une méthode de classification. Ainsi, une règle de décision est construite en utilisant une méthode de classification sur Z pour prédire la classe d'un nouvel essai. Une règle de décision sera donc un couple « filtres CSP + classifier »

1.2.3 Applications en BCI

La méthode du CSP combinée à une transformation log-var a été très utilisée en BCI. La première utilisation du CSP en EEG provient des travaux de [69]. Le but de ces travaux est de prédire si le signal EEG considéré provient d'un individu sain ou d'un individu présentant des troubles neurologiques (tumeur, épilepsies,...). Ainsi, les signaux EEG de 105 sujets (75 sains et 30 non sains) sont recueillis, à l'aide de 31 électrodes, puis sont découpés en petits morceaux (appelés « epochs ») de 1024 points. Un CSP couplé à une transformation log var est utilisé. La méthode de classification utilisée est l'analyse linéaire discriminante (voir Section 2.2). Les très bons résultats obtenus (76 % sur l'apprentissage et 89% sur le test) ont ouvert la voie à de nombreuses autres applications.

Par exemple, les travaux présentés dans [68] visent à classer des signaux EEG recueillis par 60 électrodes sur trois sujets durant des mouvements imaginés de la main droite et de la main gauche. Un CSP couplé à une transformation log var est utilisé pour extraire les caractéristiques puis une analyse linéaire discriminante est utilisée pour obtenir une prédiction. Les taux de bons classements dépassent les 90% pour les trois sujets considérés.

2 Classification de signaux EEG

Il existe deux types principaux d'apprentissages : l'apprentissage non supervisé et l'apprentissage supervisé. En apprentissage non supervisé, également appelé *clustering* en anglais, on dispose d'un échantillon de n individus observés sur p variables, formant ainsi la matrice des données de dimensions $n \times p$, notée X . Le but est de rechercher une partition de la population d'individus en sous groupes afin que les individus appartenant à un même groupe soit les plus homogènes possibles et que deux individus très distincts appartiennent à des sous groupes différents. En général, le nombre de classes à former est inconnu. Le résultat d'une méthode de

clustering est l'obtention d'un vecteur Y de taille n contenant pour chaque individu le numéro du sous groupe auquel il a été affecté.

L'apprentissage supervisé, appelé classification, se distingue du précédent par la connaissance de la variable Y . Le but est d'inférer, à partir de X et Y , une procédure de classification capable de fournir, pour un nouvel individu, une prédiction de la classe auquel il appartient. Il existe deux types de classification selon que la variable à prédire (Y) soit qualitative ou quantitative. Une variable est dite quantitative lorsqu'il s'agit d'une quantité numérique (âge, poids, température, ...). Une variable qualitative représente une information qui est divisée en catégories, appelées aussi « modalités », au sein de la population observée (sexe, couleur des yeux, signe astrologique, ...). Le résultat d'une méthode de classification est l'obtention d'un modèle (règle de décision) auquel est associé une mesure de la qualité de la prédiction obtenue à l'aide de ce modèle (par exemple critère d'information d'Akaike ou erreur quadratique moyenne dans le cas où Y est quantitatif et taux de bons classements quand Y est qualitatif). Toute la difficulté des tâches de classifications supervisées est de trouver un modèle minimisant l'erreur de prédiction sur les individus ayant servi à sa construction tout en gardant une souplesse importante et une faible complexité (nombre raisonnable de paramètres à estimer) afin de pouvoir s'adapter à de nouveaux individus et ainsi éviter le problème du sur-apprentissage.

Les problèmes de BCI correspondent généralement à des problèmes de classifications supervisées où la variable à prédire Y est qualitative. L'étape d'extraction de caractéristiques débouche sur l'obtention d'une matrice X et d'un vecteur Y à prédire. De nombreuses méthodes de classifications supervisées existent et leur utilisation est courante dans ce type de problème. Dans un premier temps, l'estimation du taux de bons classements va être décrite puis deux méthodes de classifications couramment utilisées en BCI seront présentées.

2.1 Estimation du taux de bons classements

La mesure de la qualité prédictive d'un modèle repose sur l'utilisation d'un échantillon d'apprentissage et d'un échantillon test. L'échantillon d'apprentissage permet d'apprendre la règle de décision. Cette règle de décision est alors appliquée sur des individus qui n'ont pas servi à la former (échantillon test) et une prédiction est obtenue. Cette prédiction est comparée aux classes réelles auxquelles appartiennent chaque individu et un taux de bons classements (TBC) est obtenu par la relation :

$$TBC = 100 * \frac{\text{nombre d'individus correctement classés}}{\text{nombre d'individus}}.$$

Il convient de constater que le taux de bons classements obtenu va dépendre de la compo-

tion des deux échantillons (apprentissage et test). Ainsi, la séparation de l'échantillon initial a une importance non négligeable. En particulier, lorsque l'on veut comparer les résultats obtenus par deux méthodes différentes sur un même jeu de données, il est important que les découpages utilisés pour estimer les taux de bons classements des deux approches soient identiques. Dans cet esprit, en compétition BCI, les données fournies sont déjà pré-découpées en échantillons d'apprentissage et de test, permettant ainsi de comparer les différentes approches entre elles et de désigner un gagnant. Cependant, le principal inconvénient de cette méthode est que le TBC obtenu n'est estimé qu'avec un unique découpage. Il se peut donc que celui-ci ne soit pas fiable (trop pessimiste ou optimiste en fonction de la composition des deux échantillons). Afin d'obtenir une estimation fiable du taux de bons classements, il est donc préférable d'étudier les performances de la méthode sur plusieurs découpages. La validation croisée et le *Leave One Out* (LOO) sont des méthodes d'estimation fondées sur ce principe.

La validation croisée consiste à découper aléatoirement l'échantillon initial en K parties (on parle alors de validation croisée de type K folds). À chaque itération, l'une de ces parties est mise de côté et forme l'échantillon test. Les $K - 1$ parties restantes forment l'échantillon d'apprentissage. La méthode de classification est donc entraînée sur l'échantillon d'apprentissage et une règle de décision est obtenue. Celle-ci est utilisée pour obtenir une prédiction pour chacun des individus de l'échantillon test. Ce procédé est répété K fois, c'est-à-dire une fois par partie, permettant ainsi d'obtenir une prédiction pour chaque individu de l'échantillon initial. La prédiction est comparée aux classes réelles permettant ainsi d'obtenir un taux de bons classements. Étant donné que le découpage en K parties de l'échantillon initial n'est pas unique, la procédure de validation croisée est parfois itérée afin d'obtenir une estimation encore plus fiable du TBC. Le problème du choix de K est lié au problème connu en statistique sous le nom de « dilemme biais - variance » [70]. En pratique, on souhaiterait minimiser à la fois le biais et la variance de l'estimateur de l'erreur de classification. Cependant, ces quantités sont antagonistes et une diminution de l'une des deux entraîne une augmentation de l'autre. Ainsi, il convient de trouver un compromis. Une valeur élevée de K va permettre d'obtenir un biais faible mais une variance élevée et inversement pour une faible valeur de K . En pratique, lorsque la taille de l'échantillon le permet, il est courant d'utiliser $K = 10$.

Le *Leave One Out* est un cas particulier de validation croisée pour lequel $K = n$. Ainsi, chaque partie n'est composée que d'un seul individu. Le principe est ensuite identique à celui qui vient d'être présenté. Une prédiction est obtenue, itération par itération, pour chaque individu. Cette méthode est principalement utilisée lorsque le nombre d'individus est faible. Elle présente l'avantage d'être unique (il n'y a qu'une seule façon de découper un échantillon de n individus

en n parties).

Il est très fréquent que les méthodes de classification aient un ou plusieurs paramètres à régler. Le réglage, inhérent à la construction du modèle, se fait sur l'échantillon d'apprentissage. Plusieurs valeurs de ces paramètres sont mises en concurrence et un TBC est estimé pour chacun des modèles obtenus en considérant ces valeurs des paramètres. Les valeurs des paramètres permettant d'obtenir le meilleur TBC sont donc choisies.

2.2 Méthode de classification

Dans cette partie, la méthode de classification, la plus couramment utilisée lorsque les caractéristiques sont extraites par CSP et transformation log-var, va être présentée. Il s'agit de l'analyse discriminante linéaire (LDA). Il est bien sûr possible d'utiliser d'autres méthodes de classification sur les caractéristiques extraites. Pour une description plus détaillée des autres principales méthodes de classification, le lecteur est invité à consulter [71].

L'analyse discriminante linéaire

L'analyse discriminante linéaire (LDA) est une méthode de classification supervisée couramment utilisée pour la classification de signaux EEG (voir par exemple [72], [68] ou encore [14]). Elle consiste à construire des hyperplans afin de séparer les données qui appartiennent à différentes classes dans différentes régions de \mathbb{R}^p . La classification d'un nouvel individu s'obtient en déterminant à quelle région de \mathbb{R}^p il appartient.

Dans ce qui suit, on suppose que l'échantillon d'apprentissage provient d'une population composée d'un plus grand nombre d'individus et sur laquelle est définie une partition en C groupes, notés respectivement G_1, G_2, \dots, G_C . La probabilité qu'a un individu d'appartenir à chacun de ces groupes est notée π_1, \dots, π_C . Pour un groupe G_c donné, la densité de la variable X est notée $f_c(x)$. La densité marginale de X est notée $f_X(x) = \sum_{c=1}^C \pi_c f_c(x)$.

Le but de la tâche de classification est d'affecter un nouvel individu $x \in \mathbb{R}^p$ à l'un des C groupes connus. Cet individu va donc être affecté au groupe G_c qui maximise la probabilité $P(Y = c \mid X = x)$ qui correspond à la probabilité que l'individu provienne du groupe G_c sachant que sa valeur observée de la variable X est x . En utilisant la formule de Bayes, il vient que $P(Y = c \mid X = x) = \frac{f_c(x)\pi_c}{f_X(x)}$. Le dénominateur ne dépendant pas de c , il vient que l'on

cherche :

$$\max_{c=1,\dots,C} f_c(x)\pi_c \quad (\text{III.6})$$

La LDA repose sur deux hypothèses principales. La première est que les variables explicatives X suivent des lois normales p -dimensionnelles notées $\mathcal{N}_p(\mu_c, \Sigma_c)$ dont les paramètres dépendent du groupe c . Les densités respectives sont données $\forall x \in \mathbb{R}^p$ par :

$$f_c(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \det(\Sigma_c)^{1/2}} e^{-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)} \quad (\text{III.7})$$

La seconde hypothèse est celle de l'homoscédasticité, c'est-à-dire de l'égalité des matrices de variances covariances des C groupes : $\Sigma_1 = \Sigma_2 = \dots = \Sigma_C = \Sigma$.

En utilisant ces deux hypothèses et en appliquant un logarithme dans l'équation (III.6), on cherche :

$$\max_{c=1,\dots,C} \left(\ln(\pi_c) - \frac{p}{2} \log(2\pi) - \frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2} (x - \mu_c)^T \Sigma^{-1} (x - \mu_c) \right) \quad (\text{III.8})$$

En retirant les termes de l'équation (III.8) qui ne dépendent pas de c , on obtient la règle de décision :

$$\max_{c=1,\dots,C} \mu_c \Sigma^{-1} x - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c + \ln(\pi_c) \quad (\text{III.9})$$

Dans l'équation (III.9), les quantités π_c , μ_c ainsi que la matrice de variance covariance Σ sont inconnues. Des estimateurs $\hat{\pi}_c$, $\hat{\mu}_c$ et $\hat{\Sigma}$ obtenus par maximum de vraisemblance ont donc été créés.

$$\hat{\pi}_c = \frac{n_c}{n}$$

$$\hat{\mu}_c = \sum_{i \in \mathcal{J}_c} \frac{x_i}{n_c}$$

$$\hat{\Sigma} = \sum_{c=1}^C \sum_{i \in \mathcal{J}_c} \frac{(x_i - \hat{\mu}_c)(x_i - \hat{\mu}_c)^T}{n - C}$$

En pratique, la valeur des estimateurs est estimée à partir de l'échantillon d'apprentissage. Dans ces relations, $c = 1, 2, \dots, C$ et \mathcal{J}_c est l'ensemble des indices correspondant aux individus du groupe G_c .

Dans le cadre de ces travaux de thèse, la variable à prédire (état de vigilance) est binaire

($C = 2$). Dans ce contexte, l'individu x est affecté au groupe G_1 si :

$$\mu_1 \Sigma^{-1} x - \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 + \ln(\pi_1) > \mu_2 \Sigma^{-1} x - \frac{1}{2} \mu_2^T \Sigma^{-1} \mu_2 + \ln(\pi_2) \quad (\text{III.10})$$

Ce qui revient à écrire :

$$(\mu_1 - \mu_2)^T \Sigma^{-1} x > \frac{1}{2} (\mu_1 + \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) + \ln\left(\frac{\pi_2}{\pi_1}\right) \quad (\text{III.11})$$

Encore une fois, en pratique, l'utilisation de cette règle de décision s'obtient en remplaçant les quantités π_c , μ_c et Σ par leur estimateur respectif $\hat{\pi}_c$, $\hat{\mu}_c$ et $\hat{\Sigma}$.

3 Sélection de variables en BCI

Le nombre de capteurs utilisés en EEG a progressivement augmenté, poussant la communauté à inventer de nouveaux systèmes de nomenclatures et de positionnement des électrodes. Les avancées technologiques ont permis d'obtenir des données de plus en plus précises. Cependant, les signaux EEG provenant d'électrodes adjacentes ont tendance à présenter de très fortes corrélations. Ce constat pousse donc à s'interroger sur la nécessité de disposer des signaux d'électrodes pouvant être redondants. De plus, le but premier des BCI est de permettre une application en monde réel des méthodes développées. Un nombre trop important de capteurs (électrodes) peut être considéré comme un frein à ces applications du fait du coût des systèmes à acquérir et du temps d'installation lors de chaque utilisation. C'est donc naturellement que sont apparues des méthodes visant à sélectionner les électrodes et/ ou les caractéristiques extraites qui sont jugées comme étant les plus utiles pour la tâche de classification considérée. Ces méthodes varient selon la façon de parcourir l'espace de recherche ou l'imbrication plus ou moins prononcée des procédures de sélection et de classification. Dans les prochains paragraphes, trois méthodes de sélection de variables en BCI vont être présentées.

3.1 Sélection à l'aide d'algorithmes génétiques

3.1.1 Principe général d'un algorithme génétique

Ces algorithmes d'optimisation [73] [74] sont fondés sur une abstraction simplifiée de la théorie de l'évolution Néo-Darwinienne. L'idée générale de ces algorithmes est qu'une population de solutions potentielles va améliorer ses caractéristiques au fil du temps à travers une série d'opérations génétiques telles que la sélection, la mutation ou encore le croisement génétique.

Le principe général d'un algorithme génétique est donné en Figure III.6.

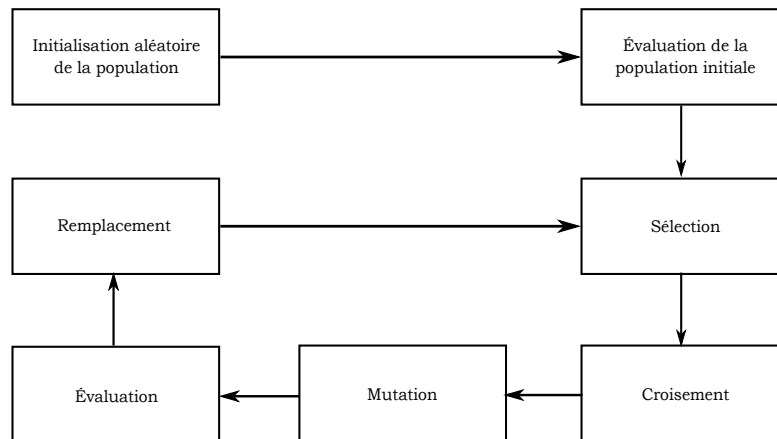


Figure III.6 – Boucle évolutionnaire d'un algorithme génétique classique.

Le but de ces algorithmes est d'optimiser une fonction (appelée fitness) sur un espace de recherche de solutions potentielles. Les solutions (appelées individus) correspondent à des points dans l'espace de recherche. Un nombre fixé de solutions va être généré aléatoirement, initialisant ainsi l'algorithme génétique. Les solutions sont représentées par leur génome (chiffres binaires ou nombres réels, avec une taille fixe ou variable). Les individus sont évalués à l'aide de la fonction fitness afin d'estimer la qualité de la solution qu'ils proposent. Ils sont ensuite sélectionnés en se fondant sur cette évaluation (en utilisant par exemple une série de tournois). Les individus sélectionnés sont appelés parents. Ces parents sont utilisés pour générer de nouveaux individus au moyen de deux opérations génétiques basiques, le croisement (croisement des génomes de deux individus ou plus) et de la mutation (modification aléatoire d'une (ou plusieurs) composante(s) du génome de l'individu). Ces individus nouvellement générés sont appelés les enfants, car ils partagent des similitudes (génétiques) avec les parents qui ont été utilisés pour les générer. Enfin, les individus sont sélectionnés et remplacent la population initiale (meilleurs individus entre les parents et les enfants ou sélection des enfants uniquement ou encore conservation des enfants et des parents, ...). L'algorithme est itéré jusqu'à ce qu'un critère d'arrêt soit atteint ; par exemple, lorsque tous les individus sont identiques (convergence de l'algorithme) ou après un nombre prédéterminé d'itérations. Pour un panorama des techniques évolutionnaires, le lecteur pourra se référer à [75].

3.1.2 Utilisation en BCI

Dans [64], un algorithme génétique est utilisé pour la sélection de caractéristiques pour la classification de données EEG recueillies lors de tâches de mouvements des doigts. Le but est donc de déterminer, pour un signal EEG donné, quel type de mouvement a été réalisé. Les signaux de 6 des 27 électrodes initiales (F3, F4, C3, C4, CP3, et CP4) sont utilisés pour extraire les spectres de puissance des bandes de fréquences usuelles en EEG (δ , θ , α , β_1 (13–20 Hz), β_2 (20–35 Hz) et γ). Un algorithme génétique avec une population de 20 individus est utilisé afin de déterminer quelle bande de fréquence doit être considérée sur quelle électrode. Les SVM permettent d’obtenir un TBC (évalué par validation croisée de type 10 folds) pour chaque combinaison testée au cours de l’algorithme. La fonction de fitness correspond au TBC obtenu par chaque modèle de classification. Cette approche permet d’obtenir un TBC de 76%. Cependant, ce TBC est nettement inférieur à celui obtenu par une approche antérieure (TBC de 95% obtenu en utilisant les 27 électrodes) et ne permet pas d’éliminer de variable (chacune des 6 variables est au moins associée à une bande de fréquence choisie).

Une approche similaire est proposée dans [62]. Dans ces travaux, des signaux EEG sont recueillis sur 34 sujets à l’aide de 32 électrodes lors de la présentation sur un écran du mot ‘oui’ ou du mot ‘non’ (50 présentations chacun). Sur 30 des 32 électrodes initiales, la moyenne du spectre de puissance est calculée pour les ondes δ , θ , α bas (8–10 Hz), α haut (10–12 Hz), β et γ . De cette façon, 180 caractéristiques sont obtenues ($30 * 6$). La sélection des caractéristiques est réalisée à l’aide d’un algorithme génétique avec 200 individus et 50 générations où la fonction de fitness correspond au TBC obtenu en utilisant les SVM comme méthode de classification. Le meilleur TBC obtenu dans ces travaux est inférieur à 70%.

3.2 Sélection à l’aide d’algorithmes de recherche séquentielle

3.2.1 Principe général

Les méthodes de sélection séquentielles sont fondées sur l’ajout (*Sequential Forward Selection*, notée SFS) ou le retrait (*Sequential Backward Selection*, notée SBS) itératif de variables dans le modèle. Par exemple, la méthode SFS part du modèle vide, composé d’aucune variable. À la première itération, tous les modèles à une seule variable sont testés et une évaluation de la qualité de chacun d’entre eux est obtenue (p évaluations). La variable, ayant permis de construire le modèle crédité de la meilleure évaluation, est conservée. La deuxième itération consiste à tester tous les modèles à deux variables composés de la variable sélectionnée lors de la première itération ($p - 1$ modèles). Le modèle ayant la meilleure évaluation permet d’obtenir

un sous groupe de 2 variables. La procédure est itérée de la sorte jusqu'à ce qu'un critère d'arrêt soit atteint. Lorsque le nombre de variables est relativement faible (inférieur à une centaine), l'intérêt de cette approche réside dans le fait que le nombre d'évaluations nécessaires pour trouver une solution est plus faible que celui d'un algorithme génétique. De plus, contrairement à l'algorithme génétique, lorsque la mesure de la qualité des modèles est déterministe (ce qui est généralement le cas), le FBS donne une solution identique lors de chaque lancement. Le principal inconvénient de cette méthode est sa propension à tomber dans des optima locaux. L'algorithme génétique a un parcours plus « intelligent » de l'espace de recherche et possède des opérateurs génétiques qui permettent d'éviter de tomber dans ces optima locaux (croisement, mutation).

Le SBS fonctionne de manière inverse au SFS. Il part du modèle complet à p variables et retire à chaque itération la variable sans laquelle le modèle résultant a obtenu la meilleure évaluation. Il est courant que les approches SBS et SFS donnent des résultats distincts lorsqu'elles sont lancées sur le même jeu de données. Enfin, une méthode mixte consiste à tester, lors de chaque ajout de variable, si le retrait de l'une des variables ne permet pas d'obtenir un modèle crédité d'une meilleure évaluation (méthode dite de *stepwise*).

3.2.2 Utilisation en BCI

Dans le domaine des BCI, le SFS a été utilisé dans quelques travaux. Par exemple, [63] propose une procédure de sélection de caractéristiques par SFS. Le but de ces travaux est de construire une BCI afin de contrôler les mouvements d'un curseur. Les caractéristiques sont fondées sur les puissances spectrales de 7 bandes de fréquences de 3 Hz. Il y a 15 électrodes et donc 105 caractéristiques. La procédure de SFS vise à en sélectionner 30.

Enfin, une procédure d'élimination itérative, pouvant être rapprochée du SBS, est présentée dans [76]. Dans ces travaux, le but est de classer des signaux issus d'une BCI de clavier virtuel similaire à celle de [9] présentée dans la Section 4.1 du Chapitre I. La tâche de classification correspond à détecter la présence d'une composante P300. Dans ces travaux, les caractéristiques sont retirées 4 à 4 afin d'accélérer la procédure. Pour une combinaison donnée de caractéristiques, les SVM sont utilisés afin d'obtenir une prédiction de la variable binaire (présence ou absence d'une composante P300). La mesure de la qualité de cette prédiction est donnée par la relation $\frac{V_p}{V_p + F_p + F_n}$, où V_p , F_p et F_n représentent respectivement le nombre de vrais positifs (détection à juste titre de l'élément recherché, c'est-à-dire que la présence d'une composante P300 a été prédite alors qu'elle était effectivement présente), de faux positifs (détection de la présence d'une composante P300 à tort) et de faux négatifs (prédiction de l'absence d'une com-

posante P300 à tort). Cette mesure est utilisée car l'une des deux classes (présence d'un P300) peut être considérée comme la classe cible. La mesure de la qualité de la prédiction permet de classer les combinaisons et donc de sélectionner la meilleure d'entre elles à chaque itération.

3.3 Le CSP parcimonieux

Une approche permettant de réaliser un CSP parcimonieux a été proposée dans [49]. Dans cette approche, l'utilisation d'une pénalité dans le critère à optimiser permet d'éliminer des électrodes lors du calcul des filtres spatiaux. Le problème d'optimisation s'écrit alors :

$$\min_{\mathbf{w}_i \in \mathbb{R}^p} (1 - r) \left(\sum_{i=1}^k \mathbf{w}_i^T \mathbf{C}_2 \mathbf{w}_i + \sum_{i=k+1}^{2k} \mathbf{w}_i^T \mathbf{C}_1 \mathbf{w}_i \right) + r \sum_{i=1}^{2k} \frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2} \quad (\text{III.12})$$

sous les contraintes $\mathbf{w}_i^T (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}_i = 1$ avec $i = 1, 2, \dots, 2k$ et $\mathbf{w}_i^T (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}_j = 1$ avec $i, j = 1, 2, \dots, 2k$ et $i \neq j$ et où $\|\cdot\|_1$ et $\|\cdot\|_2$ représentent respectivement la norme l_1 la norme l_2 du vecteur \mathbf{w}_i .

Dans l'équation (III.12), le paramètre $r \in [0, 1]$ contrôle le degré de parcimonie. La partie gauche de l'expression, associée au poids $(1 - r)$, correspond à une écriture d'un problème d'optimisation proche de celui présenté en équation (III.1). La partie droite, associée au poids r , correspond quant à elle à la régularisation.

En pratique, une seule paire de filtres spatiaux est obtenue à partir de l'équation (III.12) à l'aide de l'algorithme d'optimisation *Sequential Quadratic Programming* (SQP). Ces filtres sont donc parcimonieux (composés de poids nuls) et permettent ainsi d'exclure certaines variables (celles associées aux poids nuls). Les variables associées à des poids non nuls sur ces deux filtres spatiaux sont sélectionnées et un CSP classique (non parcimonieux) est utilisé sur les données composées uniquement des électrodes sélectionnées.

Cette approche est testée sur des données provenant des compétitions BCI en couplant le CSP parcimonieux aux SVM. Les résultats obtenus montrent que l'approche permet d'égaliser voire même dans certains cas de surpasser les résultats obtenus en utilisant un CSP couplé à un SVM sur toutes les variables.

4 Conclusion

Le but de cette thèse étant de construire un modèle de prédiction de l'état de vigilance d'humains à partir d'un nombre d'électrodes faibles, les contributions apportées concernent principalement les méthodes d'extraction et de sélection de caractéristiques. Les principales contributions de ces travaux vont être détaillées dans les prochains chapitres.

Dans le Chapitre IV, une méthode d'extraction de caractéristiques fondée sur une décomposition en ondelettes discrète en 15 niveaux est proposée. Chaque niveau de décomposition est associé à une fréquence. L'approche proposée consiste alors à extraire les énergies de bandes de fréquences et à réaliser une régression des énergies correspondant aux fréquences liées à la bande d'ondes α . La caractéristique extraite par cette approche est la pente (coefficient directeur) de cette régression. L'approche proposée va ensuite être raffinée en sélectionnant les fréquences sur lesquelles effectuer la régression à l'aide d'un algorithme génétique. Celui-ci permettra également de sélectionner une unique électrode. Pour chaque combinaison testée par l'algorithme, une électrode et des fréquences sur lesquelles calculer la régression sont sélectionnées. Une décomposition en ondelettes discrète du signal de cette électrode sera réalisée et les énergies associées aux fréquences sélectionnées seront calculées. Une régression sur ces énergies sera réalisée afin d'en extraire la pente. Une valeur (pente) sera donc obtenue pour chaque sujet de l'étude. La fonction de fitness de l'algorithme génétique reposera sur le TBC obtenu en utilisant une méthode de classification qui prendra pour entrée la caractéristique qui vient d'être présentée.

Dans le Chapitre V, l'extraction de caractéristiques sera réalisée à partir d'un CSP combiné à une transformation log-variance. Un algorithme génétique sera proposé pour sélectionner un sous groupe d'électrodes de taille $p' < p$ sur lequel calculer ce CSP. La fonction fitness de l'algorithme génétique sera le TBC obtenu en utilisant une LDA sur les caractéristiques extraites par CSP. Dans ce même chapitre, une autre approche fondée sur la combinaison d'une méthode de sélection de variables de type SFS ou SBS avec une évaluation fondée sur l'extraction de caractéristiques par CSP et transformation log-var sera proposée. Les résultats des deux approches seront comparés.

Enfin, le Chapitre VI présentera un algorithme de CSP parcimonieux fondé sur l'utilisation des travaux existants sur l'analyse en composantes principales parcimonieuse.

Chapitre IV

Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

Sommaire

1	Introduction	94
2	Les données utilisées	95
3	Extraction de caractéristiques	96
3.1	Énergie d'un signal	96
3.2	Le critère des pentes	97
3.3	Le critère de l'exposant de Hölder et le critère des ondes α	99
3.4	Résultats préliminaires	100
4	Sélection de variable et extraction de caractéristiques par algorithme génétique	103
4.1	Choix algorithmiques	103
4.1.1	Opérateurs génétiques	104
4.1.2	Fonctions d'évaluations	106
4.1.3	Critère d'arrêt	109
4.2	Résultats	111
5	Conclusions	115

1 Introduction

L'objectif du travail présenté dans ce chapitre est de construire un modèle capable de prédire l'état de vigilance d'un humain à l'aide de son activité électrique cérébrale recueillie via un casque à 58 électrodes. Ce modèle sera utilisé dans des applications en temps réel. C'est pourquoi, il est important de remplir les objectifs suivants :

- Réduire le temps nécessaire pour installer le casque à électrodes sur un sujet, en utilisant une méthode de sélection des variables afin de choisir la meilleure électrode (fondée sur le taux de bons classements). En effet, dans les applications du monde réel, il est important de limiter le nombre d'électrodes utilisées afin de réduire le temps nécessaire à l'installation du casque et le coût du matériel (casque EEG en particulier). Une longue mise en place du casque EEG peut entraîner une perturbation de l'état mental du sujet.
- Obtenir un modèle (règle de décision) capable de donner une prévision fiable de l'état de vigilance d'un nouveau participant.

Pour atteindre ces objectifs, des caractéristiques doivent être extraites du signal EEG afin de simplifier la tâche de classification. En Section 1 du Chapitre III, des méthodes usuelles d'extraction de caractéristiques ont été présentées. Ici, une nouvelle méthode, fondée sur l'utilisation d'une décomposition en ondelettes discrète va être proposée afin d'extraire un critère de discrimination de l'état de vigilance. Plusieurs méthodes usuelles de classifications supervisées (les arbres binaires de décision, les forêts aléatoires, ...) seront utilisées pour prédire l'état de vigilance des sujets. Le critère sera ensuite affiné à l'aide d'un algorithme génétique pour améliorer la qualité de la prédiction.

Le reste de ce chapitre se présente comme suit : les données, provenant du protocole d'acquisition de données détaillé dans le Chapitre II seront brièvement discutées dans la Section 2. Le pré-traitement des données proposé dans ce travail sera présenté en Section 3 et les premiers résultats obtenus sur les caractéristiques extraites seront discutés en Section 3.4. La Section 4 détaillera l'utilisation d'un algorithme génétique afin d'améliorer les résultats obtenus dans la section 3.4. Enfin, la Section 5 présentera un résumé de ce travail et les principales conclusions.

Ce chapitre correspond aux publications et communications suivantes :

Conférences internationales

(le nom de l'orateur est souligné)

Vézard L., Legrand P., Chavent M., Faita-Ainseba F. and Clauzel J. (2012). *Classification of EEG signals by an evolutionary algorithm*. COMPSTAT 2012, Chypre, Aout 2012.

Conférences nationales

(le nom de l'orateur est souligné)

Vézard L., Legrand P., Chavent M., Faita-Ainseba F. and Clauzel J. (2012). *Classification de données EEG par algorithme évolutionnaire pour l'étude d'états de vigilance*. In EGC 2012 (Talence), vol. RNTI-E-23, pp.459-470.

Vézard L., Chavent, M., Legrand, P., Faita-Ainseba, F. and Clauzel, J. (2011). *Caractérisation d'états psychophysiologiques par classification de signaux EEG. Intégration de ces résultats dans le projet PSI*. Journée AFIM : électroencéphalographie et composition musicale, Talence, France.

Article de revue

Vézard L., Legrand P., Chavent M., Faita-Ainseba F., Clauzel J. and Trujillo L. *Classification of EEG signals by an evolutionary algorithm*. In Advances in Knowledge Discovery and management, vol. 4 (AKDM-4), Eds. F. Guillet, P. Pinaud, G. Venturini and D. Zighed, Studies in Computational Intelligence series, 2013, pp 137-158.

Chapitres de livres

Vézard L., Legrand P., Chavent M., Faita-Ainseba F. and Trujillo L. *Feature extraction and classification of EEG signals. The use of a genetic algorithm for an application on alertness prediction*. Guide to Brain-Computer Music Interfacing, Miranda E. R., Castet, J. and Knapp, B. (Ed.) (2013).

2 Les données utilisées

Lors de la réalisation de ces travaux, seules les données EEG de la première campagne d'acquisition étaient disponibles. Comme présenté dans le Chapitre II, ces données correspondent aux signaux EEG de 13 sujets recueillis par 58 électrodes actives avant et après une séance de relaxation. Dans la suite de ces travaux, les enregistrements réalisés avant la séance de relaxa-

tion et correspondant à un état de veille active chez les sujets (vigilance normale ou haute) seront considérés comme appartenant à la classe dite de l'état « normal ». Les enregistrements réalisés après la séance de relaxation et correspondant à un état de veille passive des sujets (vigilance basse) seront considérés comme appartenant à la classe de l'état « relaxé ». Chaque enregistrement contient les variations de potentiels électriques obtenus en utilisant une fréquence d'échantillonnage de 256Hz. Une représentation de la matrice de données est proposée en Figure IV.1.

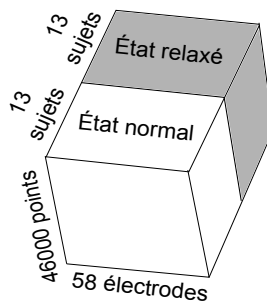


Figure IV.1 – Représentation de la matrice de données contenant uniquement les données EEG des sujets de la première campagne d'acquisition de données. Il y a trois dimensions : une pour les sujets, une pour le temps (46000 points correspondant au nombre de points dans un signal EEG de 3 minutes mesuré en utilisant une fréquence d'échantillonnage de 256 Hz) et une pour les électrodes.

3 Extraction de caractéristiques

Les données qui viennent d'être présentées sont en 3 dimensions (temps, électrodes et sujets). Afin de pouvoir appliquer une méthode de classification, l'approche envisagée est d'extraire des caractéristiques à partir du signal EEG afin d'obtenir une matrice en deux dimensions. Dans cette approche, l'énergie du signal obtenue par une décomposition en ondelettes discrète va être considérée.

3.1 Énergie d'un signal

La décomposition en ondelettes discrète dyadique a été présentée en Section 1.1.2 du Chapitre IV.12. Cette dernière peut être utilisée afin de calculer l'énergie d'un signal pour chaque

niveau de décomposition. Ainsi, l'énergie e_j^2 du signal X à l'échelle j est donnée par :

$$e_j^2 = \sum_{k=1}^{2^{j-1}} c_{j,k}^2, \forall j \in \{1, \dots, 2^T\}.$$

En d'autres termes, à partir de la grille dyadique, l'énergie associée à l'échelle j (niveau de décomposition j) est égale à la somme des carrés des coefficients de la ligne j . Une représentation schématique de l'obtention des énergies à partir de la décomposition en ondelettes discrète est donnée en Figure IV.2.

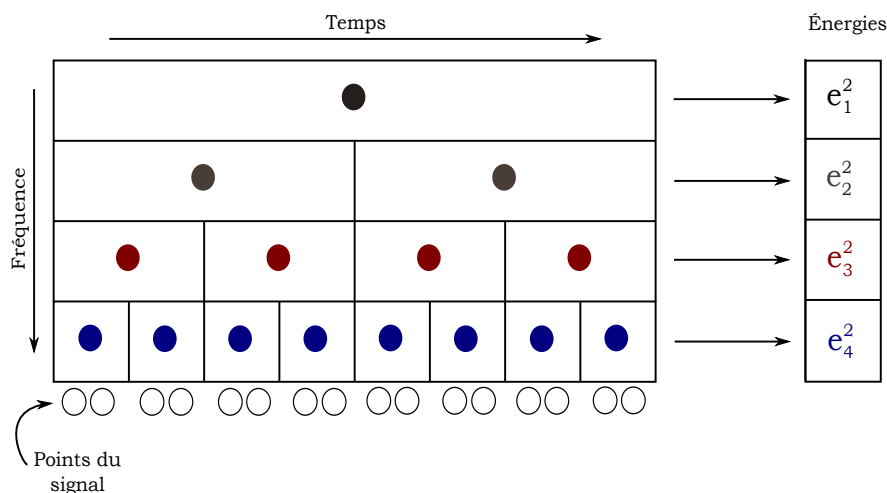


Figure IV.2 – Représentation du calcul de l'énergie du signal X aux échelles 1, 2, 3 et 4 grâce à une décomposition en ondelettes discrète d'un X de 2^4 points.

L'utilisation de l'énergie du signal fait perdre l'information temporelle. Il aurait été possible d'obtenir ce résultat à l'aide d'une transformée de Fourier, cependant, la décomposition en ondelettes discrète offre plus de possibilités pour la suite des travaux. Par exemple, la décomposition en ondelettes peut s'avérer utile si l'évolution temporelle du contenu fréquentiel des signaux est étudiée dans de futurs travaux.

3.2 Le critère des pentes

Pour un sujet i donné ($i = 1, \dots, 13$) dans un état donné (normal ou relaxé), chaque électrode p ($p = 1, \dots, 58$) fournit un signal X^p . Une décomposition en ondelettes discrète dyadique est réalisée sur ce signal en considérant 15 échelles ($15 = \lfloor \log_2(46000) \rfloor$), où 46000 est le nombre de points dans chaque signal EEG de 3 minutes et où $\lfloor . \rfloor$ représente la partie

Chapitre IV. Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

entière). À partir des coefficients obtenus, les énergies du signal sont calculées pour chacune des échelles. La figure IV.3 représente ces énergies en fonction de la fréquence.

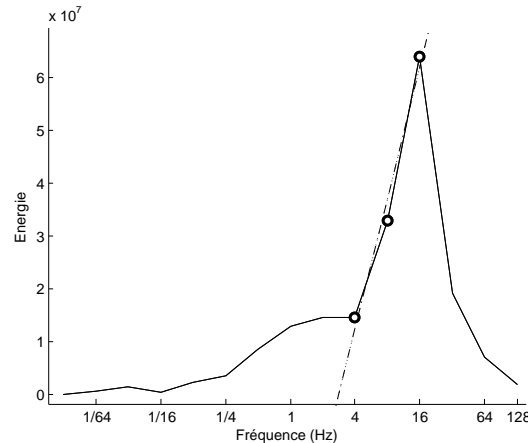


Figure IV.3 – Représentation des énergies du signal X^p , obtenues à l'aide d'une décomposition en ondelettes discrète dyadique de X^p , en fonction de la fréquence. Pour calculer le critère des pentes, une régression linéaire est effectuée (droite en pointillé) entre 4, 8 et 16 Hz (cercles).

Comme évoqué en Section 3.1.1.6 du Chapitre I, l'apparition d'ondes alpha, comprises entre 8 et 12 Hz, est associée à un état de veille passif (état dit relaxé dans notre étude). Ainsi, seules les énergies calculées pour 4, 8 et 16 Hz vont être utilisées (cercles noirs sur la figure IV.3). Une régression linéaire va alors être effectuée (droite en pointillé sur la figure IV.3) et le coefficient directeur va être conservé. Ce coefficient est représentatif de l'évolution de l'énergie du signal dans les fréquences considérées. En répétant ce procédé pour chacune des électrodes, 58 coefficients (un par électrode) sont obtenus pour un individu dans un état donné. Ainsi, une matrice de dimensions 26×58 est obtenue, représentant ce qui sera appelé dans la suite de ce chapitre « le critère des pentes ».

La Figure IV.4 donne une représentation de la matrice après l'étape d'extraction de caractéristiques. Le passage de la Figure IV.1 à la Figure IV.4 est obtenu en appliquant une décomposition en ondelettes discrète. Celle-ci a permis de résumer les 46000 points de signal en 15 valeurs d'énergies qui renseignent sur les fréquences contenues dans chaque signal EEG. Sur la gauche de la Figure IV.4, les données obtenues après la décomposition en ondelettes discrète sont représentées. Il y a toujours trois dimensions : une pour les 15 niveaux de décomposition, une pour les sujets et une pour les électrodes. Une représentation de la matrice de données obtenue après le calcul du critère des pentes est proposée sur la droite de la Figure IV.4. Seules deux dimensions demeurent dans cette matrice : une pour les sujets et une pour les électrodes.

Pour construire un modèle capable de prédire l'état de vigilance, quelques méthodes usuelles de classification (arbres binaires de décision ou K plus proches voisins par exemple) vont être appliquées sur cette matrice en 2 dimensions.

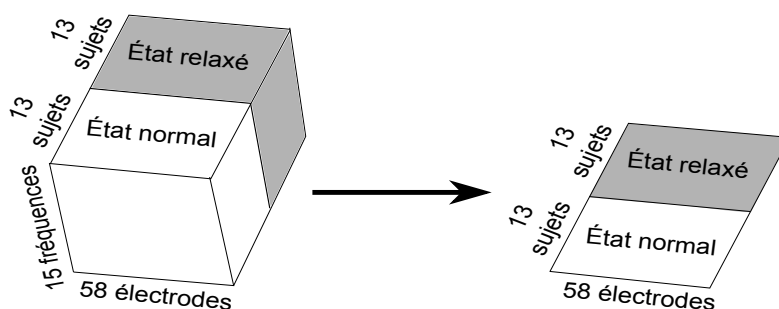


Figure IV.4 – Représentation de la matrice de données après l'étape d'extraction de caractéristiques. Sur la gauche, les données obtenues après la décomposition en ondelettes discrète. Il y a toujours trois dimensions : une pour les 15 niveaux de décomposition, une pour les sujets et une pour les électrodes. Sur la droite, la matrice de données obtenue après le calcul du critère des pentes dans laquelle seules deux dimensions demeurent : une pour les sujets et une pour les électrodes.

3.3 Le critère de l'exposant de Hölder et le critère des ondes α

Précédemment, d'autres approches sur des données EEG similaires ont été testées afin de résumer la matrice de données en une matrice en 2 dimensions [77]. Le but de ces approches était d'obtenir une méthode permettant de discriminer (séparer) les deux états de vigilance et de réduire la variance inter-individuelle. Une de ces approches était fondée sur l'utilisation de la régularité Hölderienne du signal EEG. L'exposant de Hölder [78], [79] est un outil permettant de mesurer la régularité du signal en un point donné. Plus l'exposant de Hölder obtenu est petit, plus le signal est irrégulier en ce point. Dans ces travaux, l'exposant de Hölder a été estimé en utilisant les travaux détaillés dans [80]. L'approche permettait donc de résumer le signal recueilli par une électrode par une valeur qui correspondait à une estimation de sa régularité globale (moyenne des exposants de Hölder locaux).

Une autre approche consistait à analyser la proportion d'ondes α contenue dans le signal EEG recueilli par une électrode pour un sujet dans un état donné. Le rythme α étant caractéristique d'un état de veille diffuse, le but de cette approche était d'obtenir une caractéristique permettant de discriminer les deux états de vigilance. En effet, lorsqu'une personne est dans un état de vigilance bas, ses neurones se synchronisent et opèrent à un rythme identique. Ce rythme semble être responsable de l'apparition plus marquée d'ondes α [81]. Lorsque cette personne

Chapitre IV. Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

réalise une tâche qui rompt l'état de veille diffuse dans lequel elle se trouvait, le fonctionnement de ses neurones varie considérablement. Ils s'associent en groupes de neurones qui ne se synchronisent pas à des fréquences similaires. Les ondes α sont alors masquées par l'apparition plus prononcée d'autres types d'ondes (ondes β par exemple). Ainsi, l'idée de cette approche était de mesurer la proportion d'ondes alpha contenue dans chaque signal EEG (ondes α divisées par la somme de toutes les ondes : α , β , θ et δ).

Ces deux approches permettaient d'obtenir une matrice en deux dimensions similaire à celle obtenue à l'aide du critère des pentes. Cependant, elles ne semblaient pas fonctionner aussi bien que le critère des pentes pour discriminer les deux états de vigilance (voir [77]). Ainsi, seuls les résultats obtenus à l'aide du critère des pentes seront discutés dans ce chapitre.

3.4 Résultats préliminaires

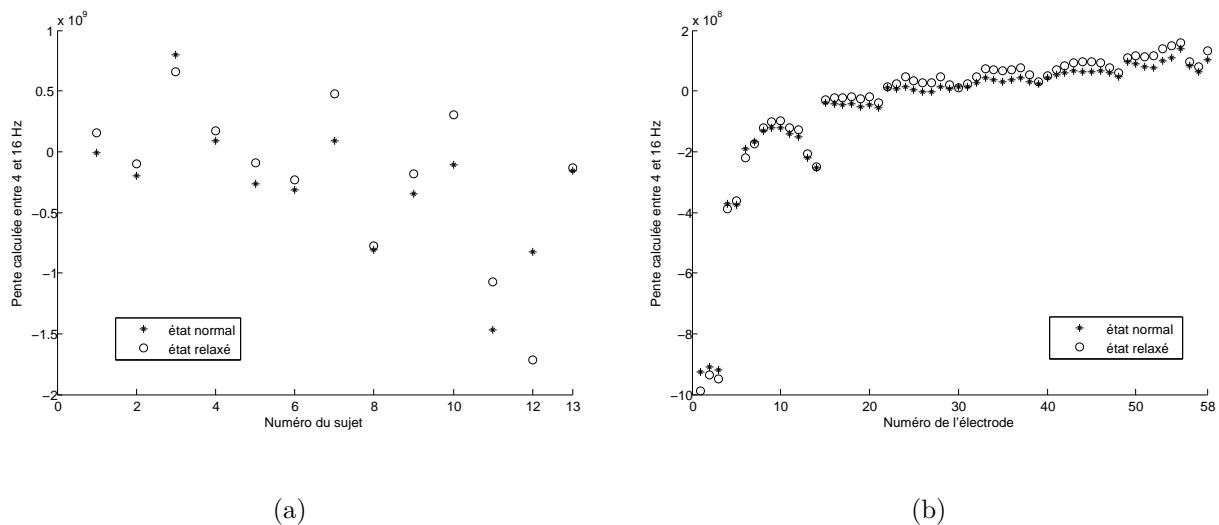


Figure IV.5 – Critère des pentes sommé sur toutes les électrodes pour chacun des 13 individus (à gauche) et sur tous les sujets pour chaque électrode (à droite)

La pertinence du critère des pentes est illustrée sur la Figure IV.5. La Figure IV.5(a) fournit pour chaque sujet, dans son état de vigilance « normal » et son état de vigilance « relaxé », la somme du critère des pentes sur toutes les électrodes. Il apparaît que, pour un individu donné, le critère des pentes est presque toujours plus faible lorsque l'individu est dans l'état normal que lorsqu'il est dans l'état relaxé. Ainsi, en comparant pour un individu donné les valeurs du critère des pentes pour l'état relaxé et normal, il est possible de distinguer efficacement les deux états. Cependant, pour un nouvel individu dont un seul enregistrement est connu, le problème

reste à traiter. La Figure IV.5(b) fournit pour chaque électrode, la somme du critère des pentes des sujets en état de vigilance « normal » et des sujets en état de vigilance « relaxé ». Le constat précédent se retrouve également au niveau des électrodes. En effet, pour une électrode donnée, le critère des pentes est plus élevé lorsque l'on considère l'enregistrement obtenu par cette électrode après la relaxation.

Ainsi, le critère des pentes permet de discriminer efficacement les deux états de vigilance pour un individu donné. Cependant, la Figure IV.5(a) permet de constater l'existence d'une forte variabilité inter-individuelle. Cette variabilité empêche de tracer une droite sur la Figure IV.5(a) qui séparerait les deux états de vigilance (représentés par des croix et des cercles). Ainsi, pour un sujet donné dont les deux enregistrements EEG sont disponibles, la comparaison des valeurs des critères des pentes obtenues permet de déterminer lequel de ces deux enregistrements correspond à celui recueilli après la séance de relaxation. Cependant, lorsqu'un seul de ces enregistrements est connu (nouveau sujet), il paraît difficile de déterminer efficacement dans quel état de vigilance le sujet se trouvait lors de son recueil.

Les méthodes de classification usuelles ont été utilisées dans un premier temps afin de prédire l'état de vigilance des sujets. Les performances prédictives des K plus proches voisins (méthode présentée dans [71]), des arbres binaires de décision [82] (CART), des forêts aléatoires [83], de la PLS discriminante (extension directe de la régression PLS décrite dans [84] obtenue en recodant la variable à prédire (état de vigilance) à l'aide d'une matrice formée par les indicatrices des modalités) et de la PLS discriminante parcimonieuse (*sparse*) [85] ont été étudiées. Le logiciel R et ses packages *class*, *rpart*, *randomForest*, *pls* et *spls* ont respectivement été utilisés pour tester ces méthodes. Les forêts aléatoires ont été utilisées en fixant le nombre d'arbres à 15 000 et en laissant les autres paramètres par défaut. Les paramètres des autres méthodes ont été fixés en appliquant une validation croisée de type 10 folds sur l'échantillon d'apprentissage (nombre de voisins pour les k plus proches voisins, complexité de l'arbre pour la méthode CART, nombre de composantes pour la PLS discriminante, nombre de composantes et valeur du paramètre de seuillage pour la *sparse* PLS discriminante). Pour pouvoir comparer les résultats, ces méthodes ont été évaluées sur les mêmes échantillons (apprentissage et test). Une validation croisée de type 5 folds a été utilisée pour calculer un taux de bons classements. Cette opération a été répétée 100 fois afin d'étudier la stabilité des méthodes de classification vis-à-vis du découpage. Les résultats sont représentés par les boîtes à moustaches données en Figure IV.6.

Il apparaît que les médianes des taux de bons classements sont décevants. Elles ne dépassent pas les 40% pour la majorité des méthodes. Le Tableau IV.1 présente les moyennes et écarts types obtenus en utilisant les différentes méthodes de classification sur le critère des pentes.

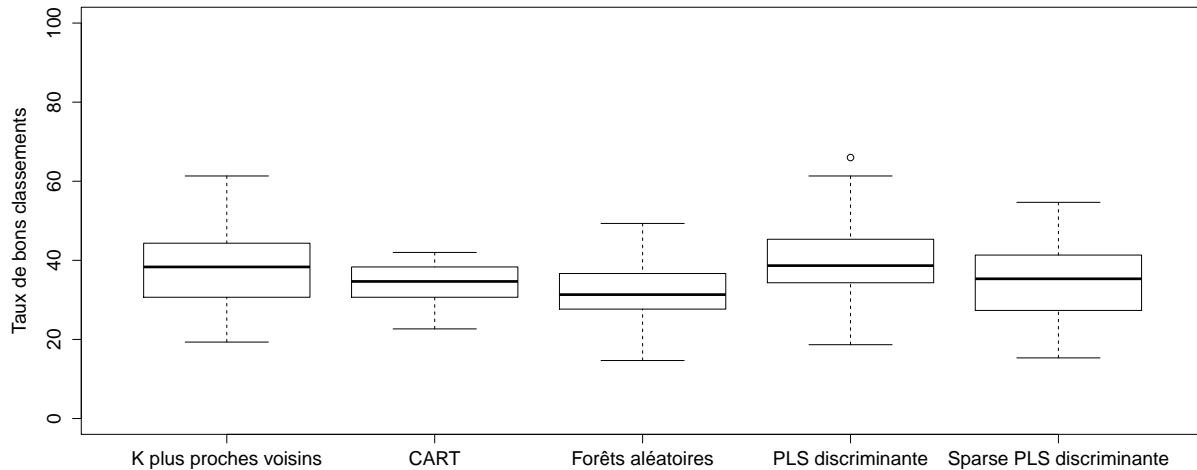


Figure IV.6 – Taux de bons classements pour différentes méthodes de classification appliquées sur le critère des pentes.

	K plus proches voisins	Arbres binaires de décision	Forêts aléatoires	PLS discriminante	Sparse PLS discriminante
Moyenne	37.28	33.98	32.03	40.63	36.25
Écart type	10.47	5.15	6.46	8.55	7.96

Tableau IV.1 – Moyennes et écarts types des taux de bons classements pour différentes méthodes de classification utilisées sur le critère des pentes.

Les importants écarts types obtenus témoignent de l'influence du découpage sur les résultats. Dans le cas d'une prédiction binaire, ces résultats ne peuvent être satisfaisants. Il est probable que la variabilité inter-individuelle constatée sur la figure IV.5(a) ait nui aux performances des méthodes de classification. Cette variabilité inter-individuelle est très difficile à inclure dans les méthodes de classification avec les données disponibles pour cette étude. L'étape d'extraction de caractéristiques a donc été raffinée afin d'obtenir de meilleurs taux de bons classements. Plus précisément, un algorithme génétique a été utilisé comme une procédure de sélection de variable, pour déterminer l'électrode et les fréquences qui permettent d'obtenir la meilleure discrimination possible pour les deux états de vigilance.

4 Sélection de variable et extraction de caractéristiques par algorithme génétique

Dans cette partie, un algorithme génétique est utilisé afin d'améliorer le critère des pentes. Dans ce qui précède, l'utilisation des ondes α a été privilégiée comme le suggérait la littérature du domaine. Ainsi, une régression a été réalisée sur les fréquences 4, 8 et 16 Hz. En raison des résultats décevants obtenus, l'approche va être révisée. La meilleure plage de fréquences (non nécessairement contiguës) pour effectuer la régression va être recherchée. De la même façon, jusqu'à présent toutes les électrodes ont été conservées. Cependant, un des objectifs de ce travail est d'éliminer des électrodes afin de diminuer le temps de pose du casque. Ainsi, la meilleure combinaison électrode / fréquences, au sens de la qualité de la prédiction obtenue, est recherchée.

Dans ce travail, il y a 58 électrodes et 15 niveaux de décomposition. Ainsi, il existe $58 * 2^{15} = 1\,900\,544$ façons de choisir une électrode et une plage de fréquences. Pour éviter de les parcourir de manière exhaustive, l'approche envisagée est d'utiliser un algorithme génétique pour réaliser à la fois une sélection d'une variable et une extraction de caractéristiques. Le principe général d'un algorithme génétique a été détaillé en Section 3.1.1 du Chapitre III. Des approches ont déjà utilisé un algorithme génétique pour sélectionner des variables (voir par exemple [86] et [87]). Des travaux visant à sélectionner des caractéristiques dans des signaux EEG en BCI ont également vu le jour (voir par exemple [64], [62]).

La structure de l'algorithme génétique utilisé dans nos travaux ainsi que ses principales fonctions vont être détaillées dans les prochains paragraphes.

4.1 Choix algorithmiques

Dans ce travail, le génome est constitué de 16 composantes. La première est un entier variant de 1 à 58 qui caractérise le numéro de l'électrode choisie. Les 15 autres, binaires, correspondent à l'inclusion (ou non) de chaque fréquence dans la régression pour le calcul du critère des pentes. Un exemple de génome est donné en Figure IV.7. Chaque génome définit l'électrode et les fréquences sur lesquelles effectuer la régression comme illustré en Figure IV.8. Sur l'exemple illustré par cette figure, une décomposition en ondelettes discrète est réalisée sur les signaux EEG recueillis par l'électrode 15 pour chacun des sujets (électrode FC5, la correspondance entre le numéro de l'électrode et son positionnement étant fournie par la Figure II.3 du Chapitre II). Les énergies correspondant aux 15 niveaux de décomposition sont calculées. Pour un sujet dans un état donné, une régression sur les énergies correspondant aux fréquences associées à des 1

Chapitre IV. Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

dans le génome est réalisée et le coefficient directeur est conservé (critère des pentes). Sur la figure, les fréquences (1/8, 1/4 et 1 Hz) sont associées à un 1 dans le génome. C'est donc sur les énergies calculées pour ces fréquences que la régression est effectuée. Ce procédé est réalisé sur les n sujets dans les deux états de vigilance et un vecteur de longueur $2n$ est obtenu. Une méthode de classification va être utilisée afin d'obtenir un TBC pour chacun des génomes (voir Section 4.1.2).

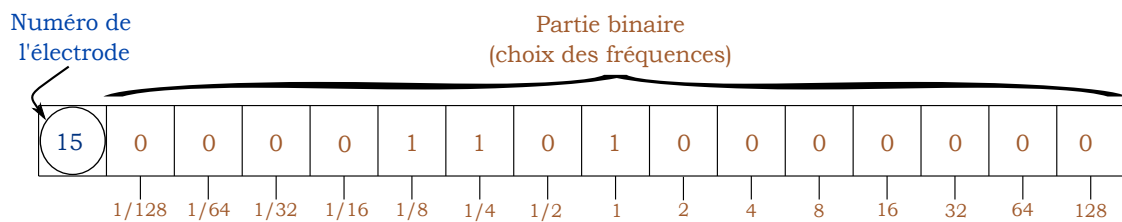


Figure IV.7 – Exemple d'un génome dans l'algorithme génétique

4.1.1 Opérateurs génétiques

Les principaux opérateurs sont la mutation et le croisement. Le croisement est l'opérateur génétique qui permet de mettre en commun les caractéristiques des deux parents afin de créer un enfant qui leur ressemble. De cette façon, si les parents possèdent de bons attributs, l'enfant va en bénéficier. La mutation a pour objectif d'introduire des différences entre l'enfant et les parents qui l'ont généré. Cela va donner de nouveaux attributs à l'enfant. Cet opérateur est particulièrement important dans le cadre de minimisations ou de maximisations de fonctions. En effet, ces fonctions peuvent avoir des optima locaux qu'il convient d'éviter si l'on recherche l'optimum global de la fonction. La mutation va permettre d'éviter aux solutions testées (enfants) de stagner autour de ces optima locaux en changeant quelque peu leurs caractéristiques et donc en leur permettant de dépasser ces optima. La mutation permet donc une meilleure exploration de l'espace de recherche. Il s'agit également d'un opérateur qui favorise la recherche locale. En effet, si tous les individus d'une population sont très similaires, la mutation est le seul opérateur qui va permettre d'introduire des différences chez les enfants créés, favorisant ainsi le parcours de l'espace de recherche.

Croisement : Dans ce travail, pour créer un enfant, 2 individus sont aléatoirement sélectionnés. Parmi ces derniers, seul l'individu possédant la meilleure évaluation est conservé (tournoi de taille 2). La pression de sélection n'est pas forte (le meilleur parmi 2) afin de conserver un

IV.4 Sélection de variable et extraction de caractéristiques par algorithme génétique

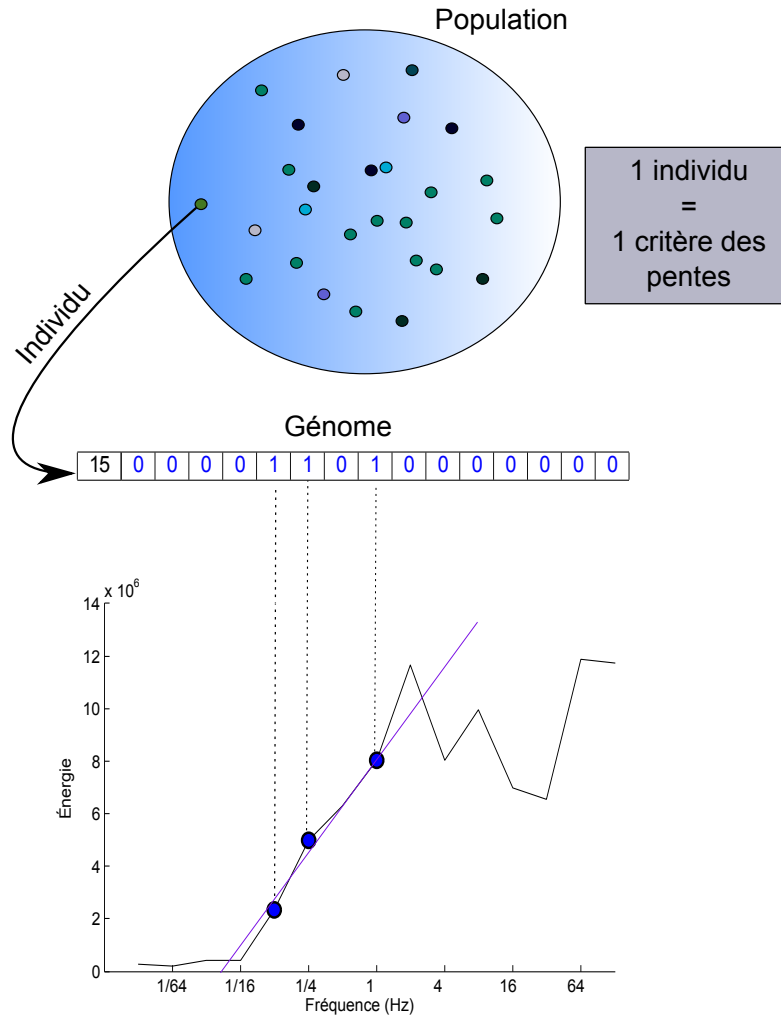


Figure IV.8 – Relation entre le génome et le calcul du critère des pentes.

maximum de diversité parmi la population. La sélection et le tournoi sont répétés à deux reprises afin de disposer de deux individus « vainqueurs » (parents). Ces deux parents vont alors être croisés afin de créer un enfant. Celui-ci va hériter de l'électrode qui se trouve à mi-chemin entre les électrodes de ses deux parents (voir Figure IV.9). Le croisement des fréquences est réalisé à l'aide d'un opérateur logique présenté en Tableau IV.2. Pour une position donnée dans le génome, lorsque les deux parents ont une valeur binaire identique, l'enfant hérite de celle-ci (lignes 1 et 4 du Tableau IV.2). Lorsque les deux parents ont des valeurs différentes (lignes 2 et 3 du Tableau IV.2), une loi de Bernoulli de probabilité $\frac{1}{2}$ est utilisée avec afin de déterminer la composante de l'enfant. Ainsi, ce dernier hérite d'un 0 (respectivement un 1) avec une probabilité de $\frac{1}{2}$.

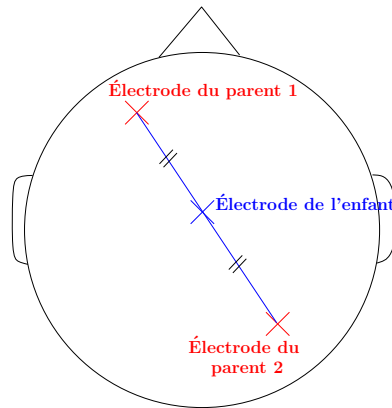


Figure IV.9 – Représentation du croisement des électrodes. L'enfant hérite de l'électrode qui se trouve à mi-chemin entre les électrodes de ses deux parents.

Parent 1	Parent 2	Enfant
0	0	0
1	0	Bern $\left(\frac{1}{2}\right)$
0	1	Bern $\left(\frac{1}{2}\right)$
1	1	1

Tableau IV.2 – Opérateur logique utilisé pour le croisement des fréquences. Pour une composante binaire donnée, lorsque les deux parents ont des valeurs différentes, une loi de Bernoulli de probabilité $\frac{1}{2}$ est utilisée avec afin de déterminer la composante de l'enfant.

Mutation : Une fois que l'enfant est créé par croisement, une mutation lui est appliquée. Chaque composante de son génome est mutée avec une probabilité égale à $\frac{1}{8}$. Ainsi, chaque enfant va être, en moyenne, affecté par deux mutations. Lorsque la mutation atteint le numéro de l'électrode, un nombre est tiré aléatoirement entre 1 et 58 pour remplacer le numéro de l'électrode de l'enfant. Pour la partie binaire, une mutation correspond au changement de la variable binaire (le 0 devient 1 et inversement).

4.1.2 Fonctions d'évaluations

Dans l'algorithme génétique, chaque génome va permettre d'obtenir un critère des pentes (vecteur de taille $2n$) en définissant une combinaison électrode / plage de fréquences. Le génome dont la combinaison permet de prédire au mieux l'état de vigilance des sujets est recherché. Ainsi, la fonction de fitness f de l'algorithme correspond au taux de bons classements (TBC)

IV.4 Sélection de variable et extraction de caractéristiques par algorithme génétique

obtenu pour chaque génome x et s'écrit :

$$f(x) = \frac{\text{\#Nombre de sujets correctement classés}}{\text{\#Nombre de sujets}}.$$

L'algorithme génétique recherche le génome x qui maximise f . Ce TBC donne une idée de la qualité du modèle prédictif obtenu en utilisant l'électrode et la plage de fréquences associée au génome. Le TBC est estimé à l'aide d'une validation croisée de type 5 *folds*. Le découpage de l'échantillon initial en 5 sous groupes est identique pour tous les génomes afin que les mêmes échantillons d'apprentissage et de test soient utilisés pour l'estimation des TBC. Lors de chaque itération, l'étape d'évaluation va être reproduite pour chaque enfant. Ainsi, il est nécessaire de n'utiliser, comme fonction d'évaluation, que des méthodes de classification rapides. Dans ce travail, deux méthodes d'évaluation ont été testées : le Single Variable Classifier (SVC) [88] et les arbres binaires de classification (CART) [82].

4.1.2.1 Le SVC

Il s'agit d'une méthode permettant de réaliser une prédiction à partir d'une seule variable. Une représentation graphique de la méthode est proposée en Figure IV.10. Sur la Figure IV.10, les individus de l'échantillon d'apprentissage dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. Pour un génome donné, un critère des pentes est obtenu. Les moyennes de ce critère sont calculées sur les individus de l'échantillon d'apprentissage dans l'état normal et relaxé (triangle bleu et rouge sur la figure). Les individus de l'échantillon test sont alors affectés à la classe correspondant à la moyenne la plus proche. Sur la Figure IV.10, l'état normal est prédit pour l'individu de l'échantillon test représenté (rond gris). Les prédictions obtenues sont comparées à la réalité afin d'obtenir un TBC.

4.1.2.2 CART

Dans ce travail, l'algorithme CART est utilisé avec une seule variable (critère des pentes) ce qui garantit la rapidité d'exécution. La construction de l'arbre binaire de décision repose sur la formation d'une série de nœuds de décision. Notons n^t le nombre de sujets qui atteignent le nœud t . Les n sujets sont présentés à la racine de l'arbre ($n^1 = n$). Une « question » binaire est posée aux sujets et ces derniers sont divisés en deux sous-groupes (ou deux branches) notés n_d^t et n_g^t (où n_d^t représente la branche droite et n_g^t la gauche et $n^t = n_d^t + n_g^t$) en fonction de leur réponse. Dans notre cas, la formation d'une question correspond à la définition d'un seuil sur

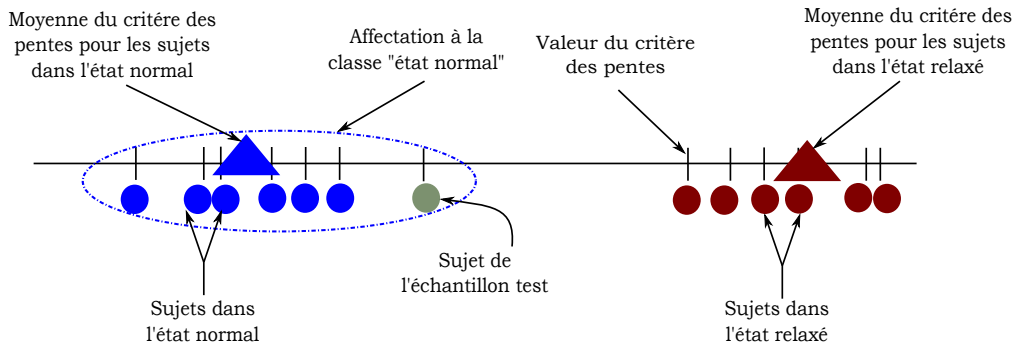


Figure IV.10 – Représentation graphique de la méthode du Single Variable Classifier. Les sujets de l'échantillon d'apprentissage dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. Les moyennes du critère des pentes sont représentées par un triangle bleu (état normal) ou rouge (état relaxé). Un individu de l'échantillon test (rond gris) est affecté à la classe correspondant à la moyenne la plus proche.

le critère des pentes. La question est alors "la valeur du critère des pentes sur le sujet considéré est-elle inférieure ou supérieure au seuil?". Le seuil est choisi de façon à ce que, à l'intérieur de chacun des deux sous-groupes, les sujets soient les plus homogènes possibles vis-à-vis de la variable à expliquer (état de vigilance).

Il existe plusieurs mesures d'hétérogénéité dont les plus couramment utilisées, dans le cas où la variable à prédire est qualitative, sont l'indice de diversité du Gini et l'entropie de Shannon. L'indice de diversité de Gini au nœud t , noté $I_{Gini}(t)$, est défini par la quantité :

$$I_{Gini}(t) = \sum_{j=1}^J \frac{n^t(j)}{n^t} \left(1 - \frac{n^t(j)}{n^t} \right),$$

où J représente le nombre de modalités de la variable à prédire (2 dans notre cas) et où $n^t(j)$ représente le nombre de sujets qui atteignent le nœud en étant de la classe j . L'entropie de Shannon au nœud t , notée $I_{Entropie}(t)$, est définie par la quantité :

$$I_{Entropie}(t) = - \sum_{j=1}^J \frac{n^t(j)}{n^t} \log \left(\frac{n^t(j)}{n^t} \right).$$

Ces mesures d'hétérogénéité d'un nœud t sont minimales si le nœud est pur (totalement homogène). Dans notre étude, l'entropie a été utilisée. Le seuil est donc choisi afin de rendre les nœuds générés les plus purs possibles.

Le processus de création de nœud est itéré jusqu'à ce que les nœuds obtenus soient terminaux. Un nœud devient terminal lorsque toute nouvelle séparation de l'échantillon qui lui

IV.4 Sélection de variable et extraction de caractéristiques par algorithme génétique

est appliquée ne permet pas améliorer l'homogénéité déjà atteinte. Un nœud terminal est alors appelé « feuille ». Chaque feuille de l'arbre est associée à une valeur (dans le cas où Y est quantitatif) ou à une modalité de Y (dans le cas où Y est qualitatif). Ainsi, dans le cas d'un Y quantitatif, la valeur associée à la feuille est la moyenne des individus qui l'ont atteint. Dans le cas d'une variable à expliquer qualitative, la modalité associée à la feuille est celle qui est majoritairement représentée parmi les individus présents à cette feuille.

Une fois que l'arbre complet est obtenu, un élagage est effectué afin de réduire la complexité de l'arbre et d'éviter le sur-apprentissage (arbre trop proche des données et très instable pour prédire des données qui n'ont pas servi à sa construction). Les performances prédictives de tous les sous arbres (arbre à 1 nœud, 2 nœuds, ..., d nœuds où d représente le nombre de nœuds dans l'arbre complet) sont évaluées par validation croisée de type 5 *folds* sur l'échantillon d'apprentissage. La prédiction de la classe d'un sujet de l'échantillon test s'obtient en lui faisant parcourir l'arbre et en l'affectant à la modalité associée à la feuille qu'il atteint. Le sous arbre qui permet d'obtenir le meilleur TBC est conservé et représente l'arbre élagué.

Les nœuds terminaux de celui-ci sont affectés à l'une des modalités de Y (« normal » ou « relaxé »). Il s'agit d'un vote majoritaire parmi les individus n^T au nœud terminal T dans le cas qualitatif et d'une moyenne de la réponse Y des n^T individus dans le cas quantitatif.

Une représentation d'un arbre de décision binaire élagué est proposée en Figure IV.11. Sur cette figure, les sujets dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. La figure se lit de haut en bas. Le haut de la figure représente la racine de l'arbre. La question qui y est posée est $P > \mu_1 ?$, où P représente le vecteur de taille $2n$ contenant les valeurs du critère des pentes pour les n sujets dans les deux états de vigilance. Suivant leur réponse, les sujets sont répartis dans les branches de l'arbre et descendent ainsi l'arbre jusqu'aux nœuds terminaux en fonction des réponses aux questions rencontrées. Sur la Figure IV.11, on suppose que l'arbre est déjà élagué. Ainsi chaque feuille de l'arbre est associée à une modalité (normal ou relaxé) représentée par un triangle (bleu ou rouge). La prédiction de la classe d'un sujet de l'échantillon test (rond gris) s'obtient en lui faisant parcourir l'arbre et en l'affectant à la modalité associée à la feuille atteinte par celui-ci. Sur la figure, l'individu est affecté à la classe « relaxé ».

4.1.3 Critère d'arrêt

L'algorithme génétique est un processus de recherche itératif. Il est donc nécessaire de définir un ou plusieurs critère(s) d'arrêt afin de stopper la recherche. Dans ce travail, l'algorithme

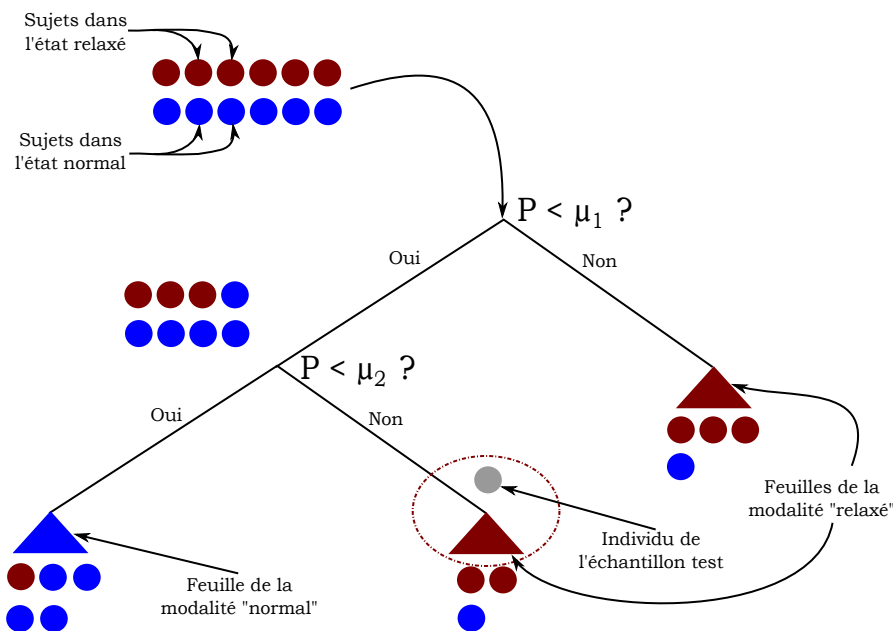


Figure IV.11 – Représentation d'un arbre de décision binaire élagué. Les sujets dans l'état normal et relaxé sont respectivement représentés par des ronds bleus et rouges. Chaque feuille de l'arbre est associée à une modalité (normal ou relaxé) représentée par un triangle (bleu ou rouge). La prédiction de la classe d'un sujet de l'échantillon test (rond gris) s'obtient en lui faisant parcourir l'arbre et en l'affectant à la modalité associée à la feuille atteinte par celui-ci (classe « relaxé » dans ce cas).

génétique est stoppé si l'une des trois conditions suivantes est satisfaite :

- Le nombre d'itérations dépasse 1 000.
- Les parents sont identiques depuis 10 générations.
- Le nombre de différences parmi les parents est inférieur à 3.

Pour calculer le nombre de différences, noté D , parmi les parents d'une population donnée, les génomes de la population à l'itération i sont stockés en ligne dans une matrice, notée P^i , de dimensions $n_{\text{parents}} \times 16$ où n_{parents} représente le nombre de parents. Soit P_j^i la colonne j de la matrice P^i (où $j = 1, \dots, 16$). Ainsi, $D = D_b + D_{elec}$ où :

- D_b est le nombre de différences pour la partie binaire de P_j^i (colonnes 2 à 16). Le nombre de différences D_b est donné par la relation $D_b = \sum_{j=2}^{16} \min(\text{nombre de 0 dans } P_j^i, \text{nombre de 1 dans } P_j^i)$.
- D_{elec} est le nombre de différences dans P_1^i (colonne de la composante des génomes représentant l'électrode). Ainsi, D_{elec} est le nombre de génomes qui ont une électrode qui est différente de l'électrode majoritairement choisie dans la population.

4.2 Résultats

L'algorithme génétique, programmé en utilisant Matlab, est lancé 100 fois pour chaque méthode d'évaluation (CART et SVC) avec 300 parents et 150 enfants. Ce nombre d'individus (parents et enfants) nous semblait suffisant pour résoudre le problème d'optimisation. Afin de pouvoir comparer les résultats obtenus à l'aide des deux méthodes d'évaluation, il est important que les taux de bons classements soient estimés sur les mêmes échantillons d'apprentissage et de test. L'estimation du TBC étant réalisée pour chaque génome de l'algorithme génétique à l'aide d'une validation croisée de type 5 *folds*, 100 partitions aléatoires des 13 sujets en 5 groupes ont été stockées avant de lancer l'algorithme génétique. Pour chacune des deux méthodes d'évaluation, la $i^{\text{ème}}$ partition a été utilisée lors du $i^{\text{ème}}$ lancement de l'algorithme génétique. À l'intérieur de l'algorithme génétique, la partition considérée définit 5 groupes de signaux EEG. Le découpage en 5 groupes est donc utilisé pour l'estimation des TBC par la procédure de validation croisée de type 5 *folds*. Par construction des partitions, les signaux EEG d'un même sujet correspondant aux enregistrements EEG réalisés avant (état normal) et après (état relaxé) la séance de relaxation appartiennent au même échantillon. De cette façon, le modèle ne peut pas être à la fois construit et testé à l'aide de signaux EEG provenant d'un même individu permettant ainsi d'éviter que les TBC obtenus ne soient biaisés (trop optimistes dans le cas où le modèle aurait été appris et testé sur des signaux provenant des mêmes sujets). Ainsi, on s'assure que quelle que soit la méthode de classification utilisée, les TBC de tous les génomes de l'algorithme génétique ont été estimés sur les mêmes échantillons d'apprentissage et de test au cours d'un lancement donné. La Figure IV.12 représente les taux de bons classements obtenus pour chaque lancement de l'algorithme génétique lorsque la méthode d'évaluation utilisée est l'algorithme CART (étoiles) et lorsque celle-ci est l'algorithme SVC (cercles).

En analysant cette figure, il apparaît que les résultats obtenus par l'algorithme génétique lorsque la méthode d'évaluation utilisée est CART dépassent très régulièrement ceux obtenus en utilisant la méthode d'évaluation SVC. Les moyennes et les écarts types des TBC obtenus sont donnés par le Tableau IV.3. Les taux de bons classements obtenus avec CART (moyenne de 86.68% et écart type de 1.87%) excèdent significativement (Test apparié de Mann-Whitney avec une p-valeur = $5.57 * 10^{-14}$) ceux obtenus avec la méthode SVC (moyenne de 83.49% et écart type de 2.37%). Ce résultat peut être expliqué en analysant les fonctionnements des deux méthodes de classification. En effet, la méthode CART construit une série de nœuds optimaux qui permettent de séparer les sujets en groupes homogènes. La méthode SVC, quant à elle, ne construit qu'un seul nœud qui n'est pas nécessairement optimal (fondé sur les moyennes des

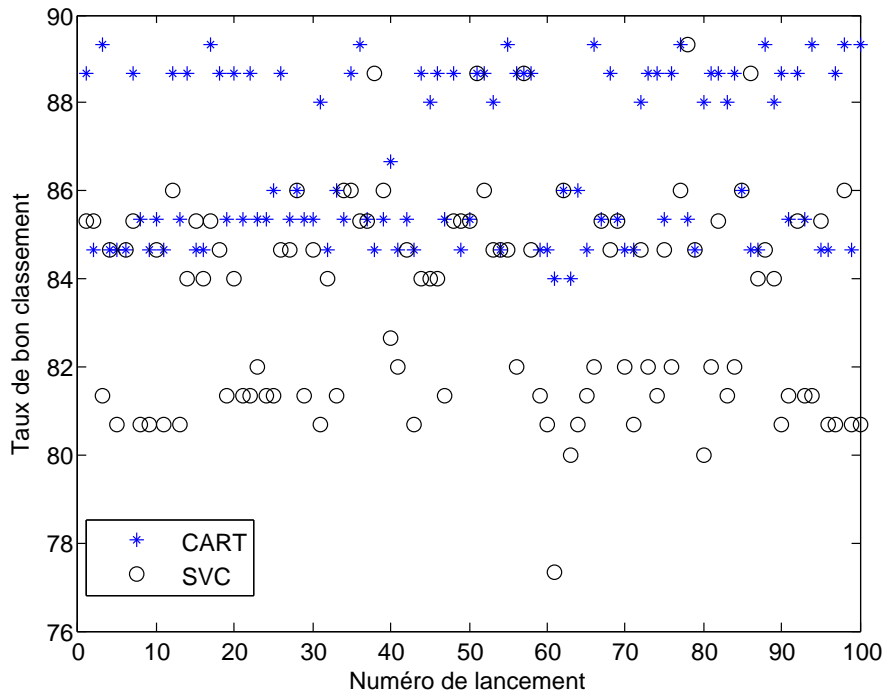


Figure IV.12 – Taux de bons classements calculé par validation croisée de type 5 folds avec CART (étoiles) et SVC (cercles) pour chaque lancement de l’algorithme génétique avec 300 parents et 150 enfants.

deux modalités). Cela peut expliquer que la méthode CART obtienne de meilleurs résultats. Cependant, la plus grande complexité de l’algorithme CART se traduit par un coût computationnel plus important. En effet, il faut 2 minutes pour terminer un lancement de l’algorithme génétique lorsqu’il est couplé à la méthode SVC alors que 35 minutes sont nécessaires pour achever un lancement lorsque CART est utilisé. La configuration de l’ordinateur utilisé lors des lancements est la suivante : Linux avec 8 * 2.8 GHz Intel Xeon processeurs X5560 et 32 GB 1333 MHz DDR3 ECC RAM.

En raison du faible nombre de sujets et de la méthode d’estimation du TBC utilisée, certains des meilleurs génomes sont à égalité lors de la convergence de l’algorithme génétique. Il est donc nécessaire de choisir un génome (noté BEST) parmi ceux qui obtiennent le même score. Ainsi, les meilleurs génomes obtenus lors de tous les lancements de l’algorithme génétique sont stockés. Parmi les génomes associés à la meilleure évaluation, celui qui apparaît le plus souvent est considéré comme le BEST pour la méthode d’évaluation considérée. Les deux BEST (pour CART et SVC) obtiennent un taux de bons classements égal à 89,33%. Pour CART, le BEST est obtenu en réalisant la régression entre 1/8, 1/4, 2, 4 et 64 Hz sur l’électrode F4 (zone frontale

IV.4 Sélection de variable et extraction de caractéristiques par algorithme génétique

droite, voir sur la figure II.3). Pour SVC, le BEST est obtenu à partir de l'électrode F2 (zone frontale droite) et la régression entre 1/32, 1/16, 2, 4, 8, 64 et 128 Hz (voir Tableau IV.4). Les fréquences choisies pour ces génomes sont donc plus étendues que celles utilisées lors de l'étude préliminaire.

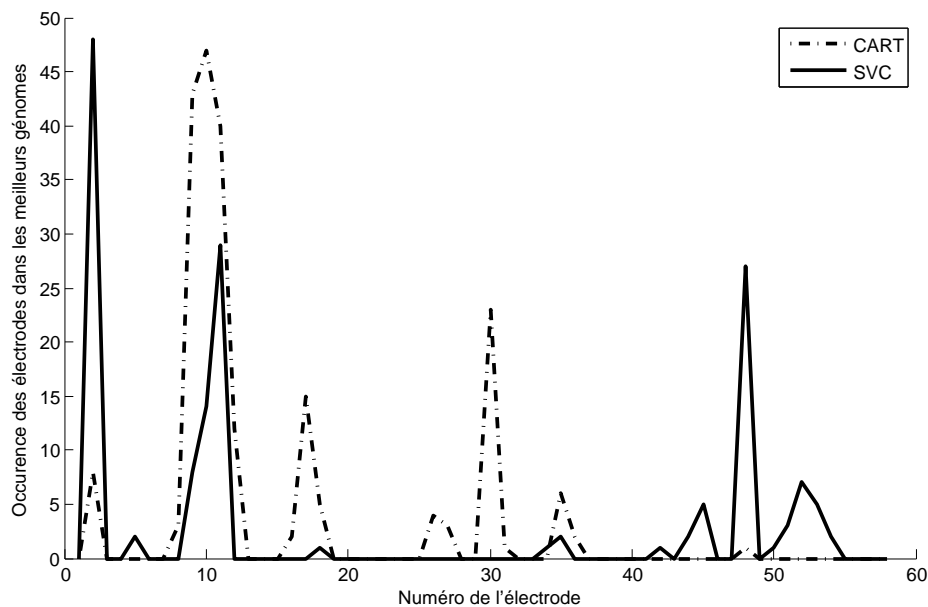


Figure IV.13 – Occurrence des électrodes dans les meilleurs génomes trouvés par l'algorithme génétique durant les 100 lancements avec 300 parents, 150 enfants et CART (courbe en tiret pointillé) ou SVC (courbe pleine).

La Figure IV.13 donne les occurrences des électrodes dans les meilleurs génomes au cours des 100 lancements de l'algorithme génétique. Pour obtenir cette figure, lorsque plusieurs génomes ont des taux de bons classements identiques à la fin de l'algorithme génétique, l'électrode la plus souvent choisie, au sein de ces génomes à égalité, est sélectionnée. L'algorithme génétique couplé à la méthode CART sélectionne des électrodes autour de l'électrode numéro 10 (FZ in Figure II.3), 17 (FC1) ou 30 (T4). Lorsqu'il est couplé avec la méthode SVC, les électrodes 2 (FPZ), 11 (F2) ou 48 (T6) sont majoritairement sélectionnées. Enfin, en moyenne, la population de l'algorithme génétique converge en moins de 50 itérations pour chacune des méthodes d'évaluation. Lors de la grande majorité des lancements, l'algorithme génétique s'arrête car le nombre de différences au sein des parents devient inférieure au seuil fixé de 3 différences. La Figure IV.14 donne une représentation du nombre de différences au sein des parents en fonction du numéro de l'itération pour un lancement de l'algorithme génétique avec la méthode d'évaluation SVC (Figure IV.14(a)) ou avec méthode d'évaluation CART (Figure IV.14(b)). Cette

Chapitre IV. Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

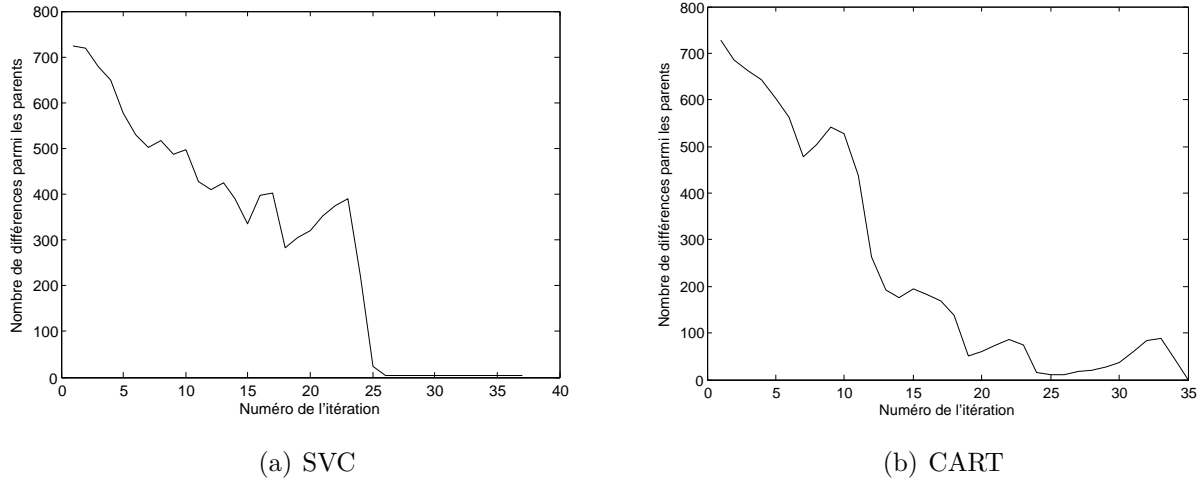


Figure IV.14 – Nombre de différences au sein des parents en fonction du numéro de l’itération lors d’un lancement de l’algorithme génétique avec 300 parents, 150 enfants et la méthode d’évaluation SVC (a) ou la méthode CART (b).

figure montre que, quelle que soit la méthode d’évaluation utilisée, le nombre de différences au sein des parents décroît très rapidement et passe en dessous du seuil de 3 différences en moins de 40 itérations déclenchant ainsi l’arrêt de l’algorithme génétique.

Méthode d'évaluation	CCR	
	Moyenne	Écart type
CART	86.68	1.87
SVC	83.49	2.37

Tableau IV.3 – Moyennes et écarts types des taux de bons classements obtenus pour les 100 lancements de l’algorithme génétique et pour les deux méthodes d’évaluation.

Méthode d'évaluation	BEST genome		
	Électrode sélectionnée	Fréquence sélectionnée (Hz)	Taux de bons classements
CART	F4	1/8, 1/4, 2, 4 et 64	89, 33%
SVC	F2	1/32, 1/16, 2, 4, 8, 64 et 128	89, 33%

Tableau IV.4 – Tableau résumant les deux meilleurs génomes trouvés lors des 100 lancements de l’algorithme génétique avec les deux méthodes d’évaluation.

Les Tableaux IV.3 and IV.4 résumant les taux de bons classements obtenus par l’algorithme génétique et permet de constater qu’ils sont supérieurs à ceux obtenus lors de l’étude préliminaire (voir Figure IV.6) en utilisant le critère des pentes calculé par régression sur les fréquences

entre 4 et 16 Hz (ondes α). Les résultats de ces deux approches peuvent être comparés car les partitions en 5 groupes utilisées lors des lancements de l’algorithme génétique sont les mêmes que celles utilisées en section 3.4 pour le calcul des TBC. De plus, le Tableau IV.5 montre que l’algorithme génétique permet une réduction de dimension considérable. La méthode de classification SVC ne pouvant être utilisée pour plus d’une variable, le Tableau IV.5 permet seulement de comparer les résultats obtenus en Section 3.4 et ceux obtenus avec l’algorithme génétique pour la méthode CART.

	Méthode d’évaluation	Nombre d’électrodes dans le modèle prédictif	TBC	
			Moyenne	Écart type
Étude préliminaire	CART	58	33.98	5.15
Algorithme génétique	CART	1	86.68	1.87

Tableau IV.5 – Comparaison entre les taux de bons classements obtenus lors de l’étude préliminaire (première ligne) et les taux de bons classements obtenus avec l’algorithme génétique (seconde ligne).

Il apparaît également qu’il est plus pertinent d’utiliser une régression sur les fréquences 1/8, 1/4, 2, 4 et 64 pour le signal de l’électrode $F4$ et la méthode de classification CART. Ainsi, ce travail permet de correctement prédire l’état de vigilance d’un nouvel individu. Pour ce faire, cette électrode et cette plage de fréquences seront utilisées pour le calcul du critère des pentes pour cet individu. L’arbre binaire de décision, construit sur l’échantillon formé par les 26 signaux (13 sujets de l’étude dans les deux états de vigilance) à l’aide de la méthode de classification CART sera alors utilisé pour prédire son état de vigilance.

5 Conclusions

Dans ce chapitre, un système de détection automatique de l’état de vigilance de l’humain à l’aide de données d’électro-encéphalographie et d’une décomposition en ondelettes discrète a été présenté. Initialement, un premier critère, fondé sur la littérature du domaine, pour obtenir une matrice en 2 dimensions a été proposé. Au regard des résultats décevants obtenus avec ce critère, un algorithme génétique a été réalisé afin de le raffiner. Cette approche a permis d’obtenir un modèle fiable (taux de bons classements moyen égal à $86.68 \pm 1.87\%$) tout en sélectionnant une unique électrode parmi les 58 initiales, ce qui améliore grandement les possibilités d’utilisation du système proposé dans les applications du monde réel. Le principal apport de ce travail repose sur la proposition d’une nouvelle méthode d’extraction de caractéristiques à partir des signaux EEG ainsi que son optimisation par algorithme génétique dont le coût computationnel demeure raisonnable.

Chapitre IV. Classification de données d'Electro-EncéphaloGraphie par analyse du contenu fréquentiel et algorithme génétique

Il apparaît maintenant nécessaire de confronter les résultats obtenus par l'algorithme génétique aux modèles disponibles en électrophysiologie humaine. De plus, il est possible d'améliorer l'algorithme génétique proposé dans ce chapitre. En effet, l'amélioration du croisement ainsi que l'introduction de nouvelles méthodes d'évaluation, notamment pour transformer la prédiction obtenue (état de vigilance normal ou relaxé) en une probabilité (utilisation de l'analyse discriminante linéaire ou de la régression logistique) sont autant de pistes qui restent à explorer.

Lors de la réalisation de ces travaux, seuls 13 sujets étaient disponibles. Le trop faible nombre de sujets n'a pas permis de garder un échantillon de sujets extérieur à l'algorithme. Bien que les TBC aient été estimés à l'aide d'une validation croisée (en découpant l'échantillon initial en ensemble d'apprentissage et de test), il se peut que l'estimation du taux de bons classements obtenu soit trop optimiste. En effet, l'algorithme génétique étant un procédé itératif, il se peut que les génomes (solutions) aient appris, au fil des itérations, leur modèle de prédiction en utilisant tous les sujets de l'échantillon. C'est pourquoi, une nouvelle campagne d'enregistrement de signaux EEG a eu lieu en 2011. Cette campagne a permis d'augmenter le nombre de signaux EEG disponibles autorisant ainsi une estimation plus fiable du TBC. L'approche présentée en Chapitre V bénéficie de ces nouveaux sujets. Pour celle-ci, il a donc été possible d'estimer les résultats sur un échantillon totalement extérieur. L'approche proposée en V peut être vue comme un prolongement de celle qui vient d'être présentée car elle permet l'utilisation d'une combinaison d'électrodes optimale (au sens du TBC obtenu).

Chapitre V

Algorithme génétique pour la sélection de variables dans les filtres CSP.

Sommaire

1	Introduction	119
2	Présentation des données utilisées	120
3	Classification de signaux EEG avec filtre spatial	121
3.1	Adaptation de la méthode CSP au problème multi-sujets	121
3.2	Pré-traitement des données	122
3.2.1	Filtrage passe-bande des données	122
3.2.2	Normalisation des données	123
3.2.3	Projection des données	125
3.2.4	Nombre de paires de filtres spatiaux	125
3.3	Optimisation des paramètres	125
3.3.1	Estimation du TBC par Leave One Out (LOO)	126
3.3.2	Résultats de l'optimisation des paramètres	128
4	Sélection d'un sous-ensemble d'électrodes par algorithmes génétiques	133
4.1	Choix algorithmiques	133
4.1.1	Fonction d'évaluation	134
4.1.2	Opérateurs génétiques	136
4.1.3	Critère d'arrêt	138
4.2	Application expérimentale	138
4.2.1	Stratégie expérimentale	138
4.2.2	Résultats	139
4.2.2.1	Performances de la méthode	139

Chapitre V. Algorithme génétique pour la sélection de variables dans les filtres CSP.

4.2.2.2	Choix de la valeur de p'	142
4.2.2.3	Électrodes sélectionnées	143
5	Sélection de variables par méthode séquentielle	147
5.1	Le forward et le backward CSP	147
5.2	Résultats	148
6	Prédiction de l'état de vigilance d'un nouvel individu	156
7	Conclusion	157

1 Introduction

Le but de ce travail de thèse est de développer une méthode de prédiction de l'état de vigilance d'humains à partir de l'enregistrement de signaux EEG. Ce travail ayant une visée applicative, la réduction du nombre d'électrodes utilisées pour obtenir cette prédiction est privilégiée.

Dans le chapitre précédent, une approche a été développée afin d'extraire des caractéristiques des signaux EEG tout en sélectionnant une unique électrode sur les 58 initiales. Elle a donc montré que la réduction du nombre d'électrodes utilisées lors de la construction du modèle de prédiction était envisageable. Cependant, le faible nombre de signaux disponibles n'a pas permis de laisser un échantillon de sujets à l'extérieur de l'algorithme génétique. De plus, les combinaisons entre plusieurs électrodes n'ont pas été envisagées.

Depuis la réalisation de ces travaux, une deuxième campagne d'acquisition de données EEG a été menée permettant ainsi d'augmenter le nombre de signaux disponibles afin d'entraîner le modèle de prédiction. Les travaux présentés dans ce chapitre s'inscrivent dans la lignée de ceux réalisés dans le précédent. En effet, dans ce chapitre, un algorithme génétique va être proposé afin de sélectionner un sous groupe d'électrodes sur lesquels apprendre la règle de décision. Cette dernière va être construite en combinant l'utilisation de la méthode du CSP, présentée en Section 1.2.1 du Chapitre III, et de la méthode LDA (voir Section 2.2 du Chapitre III).

L'une des principales originalités de l'approche qui va être présentée dans ce chapitre repose sur l'utilisation de la méthode CSP afin d'extraire des caractéristiques à partir de signaux EEG recueillis sur des sujets différents. En effet, l'utilisation de la méthode CSP est d'habitude restreinte à la construction des filtres spatiaux spécifique à un sujet.

Les résultats de l'algorithme génétique seront comparés à ceux obtenus à l'aide d'une autre approche proposée dans ce chapitre qui vise à combiner des méthodes de recherche séquentielles de type SFS ou SBS (voir Section 3.2 du Chapitre III) avec la méthode CSP.

La suite du chapitre est organisée comme suit : la Section 2 va présenter brièvement sur les données EEG utilisées lors de la réalisation de ces travaux. Dans la Section 3, l'utilisation de la méthode CSP pour la classification de données EEG sera détaillée. L'algorithme génétique proposé afin de sélectionner un sous-ensemble de variables sera présenté en Section 4. Les résultats obtenus seront comparés avec ceux de l'approche proposée en Section 5 combinant le CSP et une méthode de recherche séquentielle. Enfin, une conclusion de ces travaux sera proposée en 7.

Chapitre V. Algorithme génétique pour la sélection de variables dans les filtres CSP.

Ce chapitre correspond aux publications et communications suivantes :

Conférences internationales

(le nom de l'orateur est souligné)

Vézard L., Chavent M., Legrand P., Faïta-Aïnseba F. et Trujillo L. (2013). *Detecting mental states of alertness with genetic algorithm variable selection*. Evolutionary Computation (CEC), 2013 IEEE Congress on, 1247-1254.

Conférences nationales

(le nom de l'orateur est souligné)

Vézard L., Legrand P., Chavent M., Faïta-Aïnseba F., Clauzel J. (2012). *Variable selection using a genetic algorithm for the study of the alertness state*. Journée Evolutionnaire Thématique 23ème édition. Novembre 2012, Paris.

Article de revue

Vézard L., Legrand P., Chavent M., Faïta-Aïnseba F. et Trujillo L. (2013). *Detecting mental states of alertness with genetic algorithm variable selection*. Soumis.

2 Présentation des données utilisées

Ces travaux s'appuient sur les données recueillies à l'aide du protocole expérimental présenté en Chapitre II. Comparé aux travaux présentés dans le Chapitre IV, ceux qui vont être proposés ici peuvent s'appuyer sur les données EEG recueillies lors des deux campagnes d'acquisition de données. Ainsi, une représentation de la matrice des données est proposée en Figure V.1. Les données consistent à 38 enregistrements de 3 minutes de signaux EEG bruts recueillis à l'aide de 58 électrodes actives sur 19 sujets (13 durant la première campagne d'acquisition et 6 durant la seconde) avant et après une séance de relaxation. Chaque sujet fournit donc un enregistrement dans l'état « normal » et un autre dans un état de veille passif (vigilance basse) dit état « relaxé ». Chaque enregistrement contient les variations de potentiels électriques obtenus en utilisant une fréquence d'échantillonnage de $256Hz$ durant 3 minutes (46000 points de mesure).

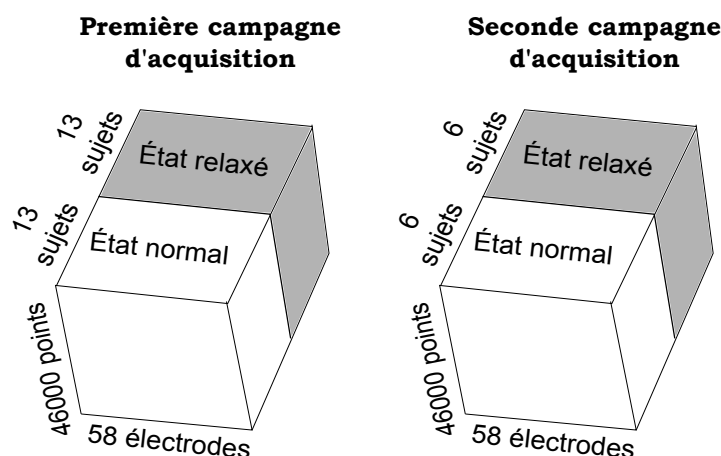


Figure V.1 – Représentation des matrices de données correspondant aux signaux EEG obtenus pendant la première campagne d’acquisition (13 sujets) et la seconde (6 sujets). Les matrices sont en 3 dimensions : une pour les sujets (2 enregistrements par sujets), une pour les électrodes et une autre pour les points de mesure (46000 points correspondant au signal EEG recueilli durant 3 minutes avec une fréquence d’échantillonnage de 256 Hz).

3 Classification de signaux EEG avec filtre spatial

Dans l’approche qui va être présentée, l’extraction de caractéristiques va être réalisée en utilisant la méthode CSP couplée à une transformation log-var (voir Section 1.2 en Chapitre III). Le but de la méthode CSP est de construire des filtres synthétiques de variance maximale dans une condition et simultanément minimale dans l’autre. Dans ce qui va suivre, un essai va représenter un signal EEG recueilli durant une période de temps T . La matrice de dimension $T \times p$ correspondant à cet enregistrement sera noté X_i , où p représente le nombre d’électrodes actives utilisées lors du recueil de l’EEG. On suppose que parmi les n essais enregistrés, n_1 l’ont été dans la condition 1 et n_2 dans la condition 2.

3.1 Adaptation de la méthode CSP au problème multi-sujets

Comme évoqué dans l’introduction, le CSP est utilisé afin de rechercher des filtres spatiaux spécifiques à un sujet. Ainsi, lorsque le CSP est utilisé en BCI, une première partie de l’expérience consiste généralement à enregistrer un lot d’essais (signaux EEG enregistrés durant un intervalle de temps T) dans les deux conditions que l’on cherche à discriminer. Une fois que ce lot d’essais est obtenu, il est utilisé afin de construire les filtres spatiaux CSP et afin

d'entraîner le modèle de classification. Lorsque la règle de décision est apprise (filtres CSP et modèle de classification), l'utilisation à proprement parler de la BCI peut débuter. Ainsi, des caractéristiques vont pouvoir être extraites de chaque nouvel essai du sujet à partir des filtres spatiaux formés sur les essais recueillis durant l'étape d'entraînement et une prédiction de la classe de chaque nouvel essai va être obtenue à l'aide du modèle de classification entraîné. Lorsqu'un nouveau sujet veut utiliser l'ICM, une nouvelle étape d'entraînement est réalisée afin d'enregistrer des essais dans les deux classes pour pouvoir construire la règle de décision.

On comprend aisément que le CSP, tel qu'il vient d'être décrit, ne peut pas être utilisé afin d'apporter une solution à notre problématique. En effet, le but de ces travaux est de construire une méthode de prédiction capable de fournir une prédiction fiable de l'état de vigilance d'un nouveau sujet. Le principe est donc de ne disposer d'aucun enregistrement cérébral du nouveau sujet pour construire la règle de décision. Or, le CSP nécessite une phase d'entraînement qui ne peut être réalisée dans notre problème. La solution envisagée dans ces travaux a été de tester les performances prédictives du CSP lorsque des essais de plusieurs sujets différents sont utilisés pour construire les filtres spatiaux. En réalisant ce choix, le but recherché est de former des filtres spatiaux adaptables sur différents sujets afin d'obtenir des signaux synthétiques utiles pour discriminer les deux états de vigilance.

3.2 Pré-traitement des données

3.2.1 Filtrage passe-bande des données

Comme évoqué en Section 3.1.1.5 du Chapitre I, l'activité électrique cérébrale est divisée en différents rythmes oscillatoires caractérisés par la bande de fréquence qui leur est associée. Les principaux rythmes classés par ordre croissant de fréquences sont les rythmes δ (1-3.5 Hz), θ (4-8 Hz), α (8-12 Hz) et β (19-26 Hz). Chacun de ces rythmes est lié à un état psycho-physiologique chez l'être humain.

Avant d'appliquer la méthode CSP, un filtre passe-bande est toujours utilisé sur les données afin d'extraire uniquement les fréquences informatives pour la tâche de classification considérée et augmenter ainsi le rapport signal sur bruit. Ces fréquences dépendent donc de la tâche de classification. Une fois que les signaux EEG ont été filtrés, chaque matrice X_i est centrée comme défini en Section 1.2.1 du Chapitre III.

Dans [72], des signaux EEG ont été recueillis à l'aide de 56 électrodes durant des mouvements de l'index gauche, droit ou des orteils. La méthode CSP est utilisée afin d'extraire

des caractéristiques utiles pour la tâche de classification (prédiction du type de mouvement). Six différentes bandes de fréquences sont proposées en utilisant les connaissances sur les principaux rythmes cérébraux. Un filtre passe-bande est donc appliqué sur les données en considérant seulement le rythme α , la bande α inférieure (8-10 Hz) ou supérieure (10-12 Hz), le rythme β , le rythme γ (38-42 Hz) et enfin une bande de fréquence composée des fréquences comprises entre 8 et 30 Hz. Sur leur données EEG, la bande de fréquence qui permet d'obtenir les meilleurs résultats (fondés sur un TBC estimé à l'aide d'une validation croisée de type 50 folds) est la bande 8-30 Hz. Ces résultats peuvent être expliqués par le fait que ces fréquences couvrent à la fois les rythmes α et β , qui sont les rythmes les plus importants lorsqu'il s'agit d'une tâche de classification de mouvements [89].

En utilisant ces résultats, la bande de fréquence 8-30 Hz est souvent utilisée dans d'autres travaux sans que plus d'explorations ne soient menées et ce même si les données ou la tâche de classification changent [68], [67], [14]. Dans d'autres travaux, d'autres plages de fréquences sont utilisées sans que des justifications appuient ces choix (1.5-35 Hz dans [69], ou encore 8-35 Hz dans [49]).

La plage de fréquence proposée dans [72] concerne la classification de signaux EEG recueillis lors de mouvements. Étant donné que notre problème de classification n'est pas semblable à celui rencontré dans [72] et que nous pensons que la plage de fréquence dépend de la tâche de classification étudiée ainsi que des données EEG, une approche permettant d'optimiser la plage de fréquence à utiliser pour le filtrage passe-bande des données est proposée dans ce travail. Le filtrage passe bande de chaque matrice de donnée EEG X_i sera réalisé à l'aide d'un filtre de Butterworth de 5^{ème} ordre.

3.2.2 Normalisation des données

Dans ce qui suit, la matrice X_i représente le signal EEG recueilli lors de l'essai i ($i = 1, \dots, n$). Dans [69], Koles et al. suggèrent d'utiliser pour un essai i la matrice de variance covariance normalisée définie par

$$\tilde{C}_i = \frac{X_i^T X_i}{\text{trace}(X_i^T X_i)} = \frac{C_i}{\text{trace}(C_i)} \quad (\text{V.1})$$

où $\text{trace}(C_i)$ est l'inertie de X_i . Cette normalisation est utilisée pour éliminer les variations d'amplitude dans les signaux EEG entre les essais. Cette façon de normaliser les données est

couramment utilisée dans d'autres travaux utilisant la méthode CSP (voir par exemple [72], [68] ou encore [49]).

Notons que \tilde{C}_i peut aussi être obtenue en normalisant la matrice de données EEG X_i au lieu de normaliser sa matrice de variance-covariance C_i . La matrice des données normalisées \tilde{X}_i est :

$$\tilde{X}_i = \frac{X_i}{\sqrt{\text{trace}(C_i)}}.$$

De cette manière, \tilde{C}_i est définie comme la matrice de variance-covariance des données normalisées et

$$\tilde{C}_i = \frac{1}{T} \tilde{X}_i^T \tilde{X}_i. \quad (\text{V.2})$$

Cette distinction entre (V.1) et (V.2) est importante en pratique car dans la méthode CSP, les matrices de variance-covariance doivent être calculées sur les données qui vont être projetées (voir Section 3.2.3).

Dans ce travail, une autre stratégie de normalisation, couramment utilisée en ACP, est proposée. Elle consiste à normaliser chaque colonne de X_i en la divisant par son écart type :

$$\tilde{X}_i = X_i D, \text{ où } D = \begin{pmatrix} \frac{1}{S_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{S_p} \end{pmatrix},$$

où S_j , $j = 1, \dots, p$, est l'écart type empirique de la $j^{\text{ème}}$ colonne de X_i . De cette manière, $\tilde{C}_i = \frac{1}{T} \tilde{X}_i^T \tilde{X}_i$ est une matrice de corrélation et non plus une matrice de variance-covariance.

Dans ce travail, trois stratégies de normalisation vont être comparées sur notre jeu de données EEG :

- $\tilde{X}_i = X_i$ i.e. les données ne sont pas normalisées.
- $\tilde{X}_i = \frac{X_i}{\text{trace}(C_i)}$ i.e. les données sont normalisées par l'inertie de X_i .
- $\tilde{X}_i = X_i D$ i.e. les données sont normalisées en utilisant l'écart type.

3.2.3 Projection des données

Dans la méthode CSP, la normalisation des matrices de variance-covariance C_i ou la normalisation des matrices de données EEG X_i , mènent à l'obtention de la même matrice de filtres W . Cependant, même si cela n'est pas explicitement évoqué, ce sont les données normalisées qui sont projetées lors de cette étape du CSP. C'est pourquoi, il est important lors de l'étape de classification de projeter les données normalisées et non pas les données originales. En d'autres termes, il convient d'utiliser $F_i = \tilde{X}_i W$ comme proposé initialement dans [69] et non pas $F_i = X_i W$ comme utilisé plus tard dans de nombreux travaux (voir [68] par exemple). C'est la raison pour laquelle nous avons fait la distinction entre normaliser les données EEG et normaliser les matrices de variance-covariance dans la Section 3.2.2.

3.2.4 Nombre de paires de filtres spatiaux

Le principal paramètre à fixer dans la méthode CSP est le nombre de paires de filtres spatiaux, noté k , utilisés pour construire W . Dans [72], la valeur de ce paramètre est optimisée entre 1 et 5. Sur ces données, les meilleurs taux de bons classements sont obtenus en prenant $k = 2$ et $k = 3$. Dans ce travail, le paramètre k va être optimisé afin de trouver la meilleure valeur possible (fondée sur les taux de bons classements).

3.3 Optimisation des paramètres

Pour appliquer le CSP, les données vont être pré-traitées comme montré en Figure V.2. Un filtre passe bande est appliqué aux données afin de ne garder que les fréquences comprises entre a et b Hz où $a < b$. Les données sont ensuite découpées en essais de longueur T . Enfin, une normalisation est appliquée à X_i pour obtenir les matrices de signaux EEG normalisées \tilde{X}_i . Les matrices de signaux EEG normalisées dans la classe 1 (respectivement 2) sont concaténées en ligne dans la matrice $\tilde{\mathbf{X}}_1$ de dimension $Tn_1 \times p$ (resp. $\tilde{\mathbf{X}}_2$ de dimension $Tn_2 \times p$). Ainsi, une configuration similaire à celle décrite en Section 1.2.1 du Chapitre III est obtenue. Dans cette configuration, n^* matrices de données EEG sont disponibles pour chaque sujet dans un état de vigilance donné, avec $n^* = \lfloor 46000/T \rfloor$ et où $\lfloor \cdot \rfloor$ est la partie entière.

Dans ce qui suit, la méthode d'estimation du taux de bons classements va être décrite et les paramètres a , b , T , k et la stratégie de normalisation vont être optimisés sur nos données.

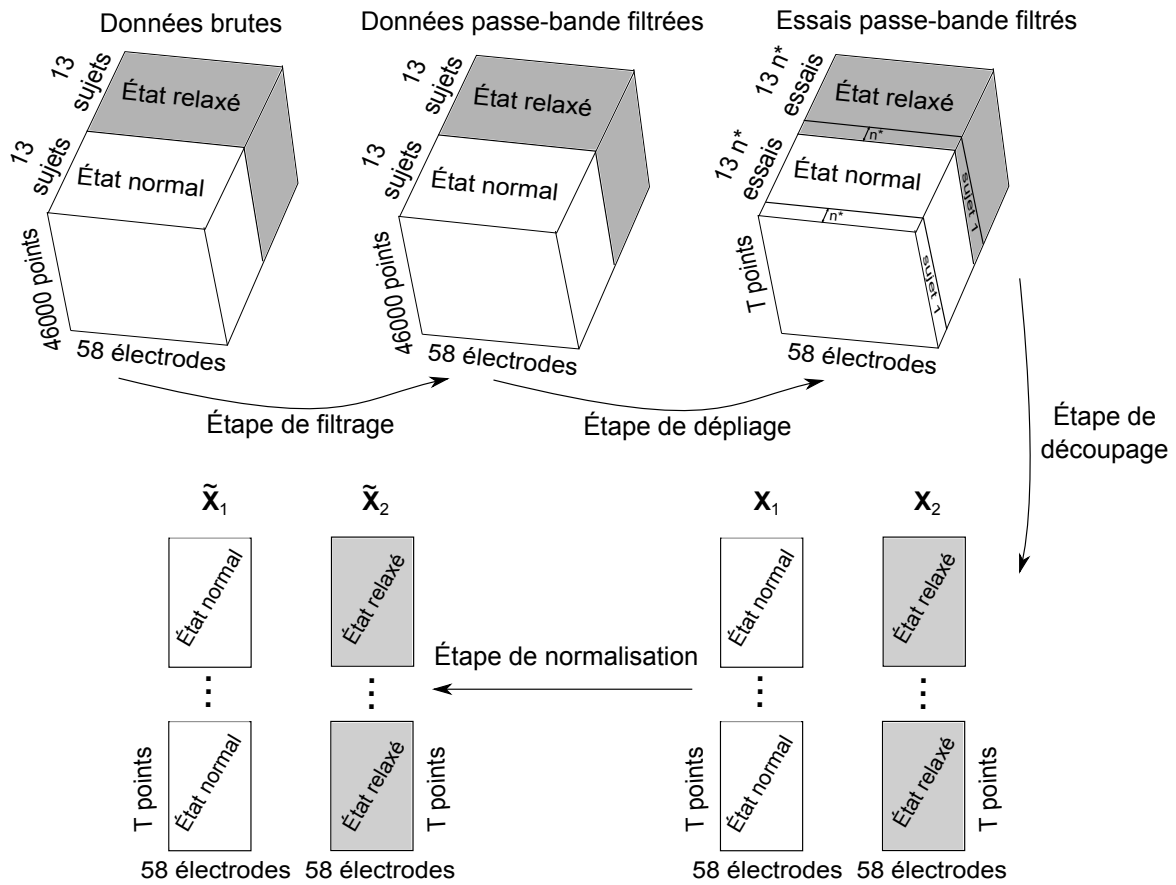


Figure V.2 – Diagramme représentant l'étape de pré-traitement des données. Les données brutes sont filtrées entre a et b Hz et découpées en essais de longueur T . Une normalisation est alors appliquée aux données pour obtenir \tilde{X}_1 et \tilde{X}_2 .

3.3.1 Estimation du TBC par Leave One Out (LOO)

Afin d'estimer le taux de bons classements obtenu en utilisant un CSP, une estimation en LOO adaptée à notre problème est utilisée. En effet, le LOO classique consiste à retirer itérativement l'un des individus de l'ensemble d'apprentissage afin de former l'ensemble test (voir Section 2.1 du Chapitre III). En appliquant ce principe sur notre problème, cela reviendrait à retirer tour à tour chacun des essais. Cependant, dans notre travail, le LOO ne peut être utilisé au niveau des essais car lorsqu'un essai est retiré de l'ensemble d'apprentissage afin de former l'ensemble test, des essais provenant du même sujet sont encore présents dans l'ensemble d'apprentissage. Cela peut donc entraîner l'obtention de résultats biaisés (optimistes dans ce cas).

Dans ce travail, le LOO va donc être utilisé au niveau des sujets et non pas au niveau des

V.3 Classification de signaux EEG avec filtre spatial

essais. En d'autres mots, à chaque itération, tous les essais d'un même sujet (matrices EEG des essais dans la classe 1 et 2) sont retirés du jeu de données. Les données privées de ces essais forment l'ensemble d'apprentissage tandis que les essais retirés forment l'ensemble test. La $m^{\text{ème}}$ itération de l'estimation LOO est représentée en Figure V.3.

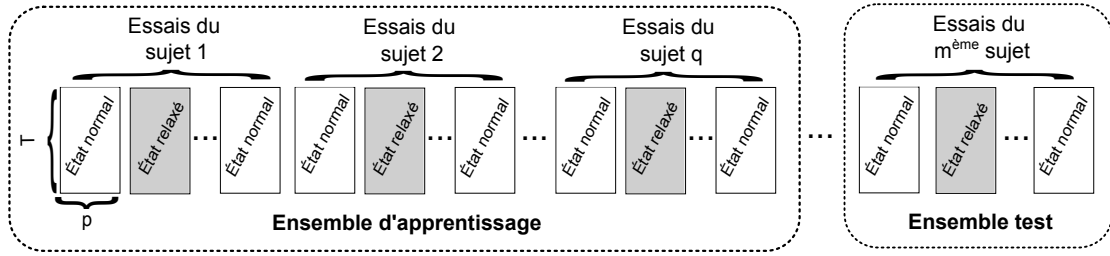


Figure V.3 – Création de l'ensemble d'apprentissage et de l'ensemble test lors de la $m^{\text{ème}}$ itération de l'estimation par LOO du TBC (avec $m = 1, \dots, q$).

Ainsi, l'estimation par LOO se déroule en q itérations (où q est le nombre de sujets). À chaque itération, un CSP est appliqué sur les essais de l'échantillon d'apprentissage afin de créer les k paires de filtres CSP. Les essais de l'échantillon d'apprentissage sont alors projetés en utilisant ces composantes CSP et une transformation logarithmique est appliquée afin d'obtenir les caractéristiques. Ces caractéristiques sont données en entrée à une méthode de classification (LDA) afin d'apprendre une règle de décision. La méthode classification LDA [71] est couramment utilisée dans le cadre de problèmes de classification supervisée. Elle consiste à construire des hyperplans afin de séparer, dans différentes régions de \mathbb{R}^p , les données qui appartiennent à différentes classes. La classification d'un nouvel individu s'obtient en déterminant à quelle région de \mathbb{R}^p il appartient (voir Section 2.2 du Chapitre III).

Une fois que la règle de décision est obtenue, les filtres CSP sont utilisés pour projeter les essais de l'ensemble test. Après une transformation logarithmique, la classe de chaque essai est prédite en utilisant la règle de décision apprise sur les essais de l'ensemble d'apprentissage.

Lorsque les q itérations sont réalisées, un vecteur de taille n contenant les classes prédites pour chaque essai est obtenu. Ce vecteur est comparé aux classes réelles des essais afin d'obtenir un taux de bons classements (nombre de prédictions correctes divisé par le nombre total de prédictions).

3.3.2 Résultats de l'optimisation des paramètres

Dans cette partie, les valeurs des paramètres a , b , k et T ainsi que la stratégie de normalisation des données vont être optimisés en se fondant sur l'estimation LOO du TBC obtenue sur les sujets de la première campagne.

Une recherche exhaustive de la meilleure combinaison de ces paramètres semble difficile. C'est pourquoi, seules quelques valeurs de ces paramètres vont être testées. En particulier, les valeurs des paramètres a et b vont être optimisées l'une après l'autre. En effet, même si une optimisation conjointe de ces paramètres est plus appropriée d'un point de vue mathématique, cette approche présente un coût computationnel trop important. Ainsi, la meilleure combinaison de paramètre parmi celles testées sera conservée.

Pour commencer, la valeur de b est fixée à 30 Hz. La valeur de a varie entre 1 et 25 Hz. Trois valeurs sont testées pour le paramètre T : 1024, 2048 et 4096 points (correspondant respectivement à des essais de 4, 8 et 16 secondes). Le paramètre k prend la valeur 1, 2 ou 3. Un filtre passe-bande entre a et $b = 30$ Hz est utilisé sur les données. Les données sont alors découpées en essais de longueur T . Un CSP couplé à une LDA est utilisé sur les données afin d'obtenir un TBC (estimé par LOO). La meilleure valeur trouvée pour a , k et T ainsi que la meilleure stratégie de normalisation vont être conservés. Ainsi, ces paramètres vont être fixés. Par la suite, la valeur de b va être optimisée entre $a + 3$ et 40 Hz. Les résultats sont présentés en Figures [V.4](#), [V.5](#), [V.6](#) et [V.7](#).

Pour commencer, la meilleure stratégie de normalisation est recherchée en comparant les résultats obtenus sur les Figures [V.4](#) (données non normalisées), [V.5](#) (données normalisées en utilisant l'inertie) et [V.6](#) (données normalisées en utilisant les écarts types). Sur ces figures, les données sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). Le nombre de paires de filtres, noté k , varie entre 1 (ligne en pointillé), 2 (ligne pleine) et 3 (ligne en pointillé mixte). Un filtre passe-bande entre a (abscisses) et 30 Hz est appliqué sur les données. Pour chaque combinaison de valeurs des paramètres, un TBC estimé par LOO est obtenu. La combinaison de paramètres testée permettant d'obtenir le TBC le plus élevé est recherchée.

En comparant les Figures [V.4](#), [V.5](#) et [V.6](#), il semble que les TBC obtenus en normalisant les données à l'aide des écarts types (Figure [V.6](#)) dépassent les TBC obtenus sans normaliser ou en normalisant avec l'inertie (Figures [V.4](#) et [V.5](#)). Ainsi, cette stratégie de normalisation

V.3 Classification de signaux EEG avec filtre spatial

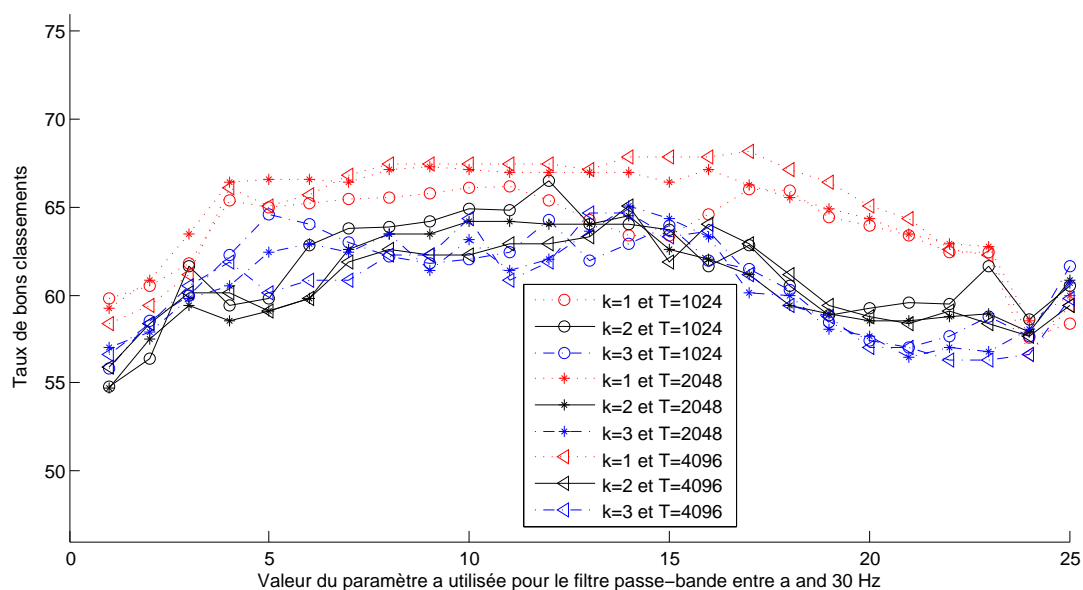


Figure V.4 – TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). **Les données ne sont pas normalisées** mais sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L'extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP.

va être définitivement adoptée sur nos données et l'on va se concentrer sur les résultats de la Figure V.6. Il apparaît que les meilleurs résultats sont obtenus en utilisant 3 paires de filtres CSP et en considérant des essais de 8 secondes ($T = 2048$). De plus, le meilleur résultat est obtenu en considérant $a = 7$.

Il reste donc à optimiser la valeur de b entre 10 et 40 Hz. La valeur de b qui permet d'obtenir le TBC le plus élevé est recherchée. Ainsi, un filtre passe-bande entre 7 et b Hz est appliqué aux données et ces dernières sont découpées en essais de 8 secondes. Les essais sont alors normalisés en utilisant les écarts types et un CSP est utilisé afin de former 3 paires de filtres. La Figure V.7 représente le TBC obtenu en fonction de la valeur de b considérée (abscisse). En analysant la Figure V.7, il apparaît que le meilleur TBC (égal à 75.35%) est obtenu en considérant $b = 30$ Hz.

Pour résumer, la meilleure combinaison de valeurs des paramètres trouvée parmi celles testée durant ces expériences est :

- Filtre passe-bande : $[7, 30]$ Hz.
- Longueur des essais : 8 secondes ($T = 2048$ points). Ainsi, chaque signal est découpé en

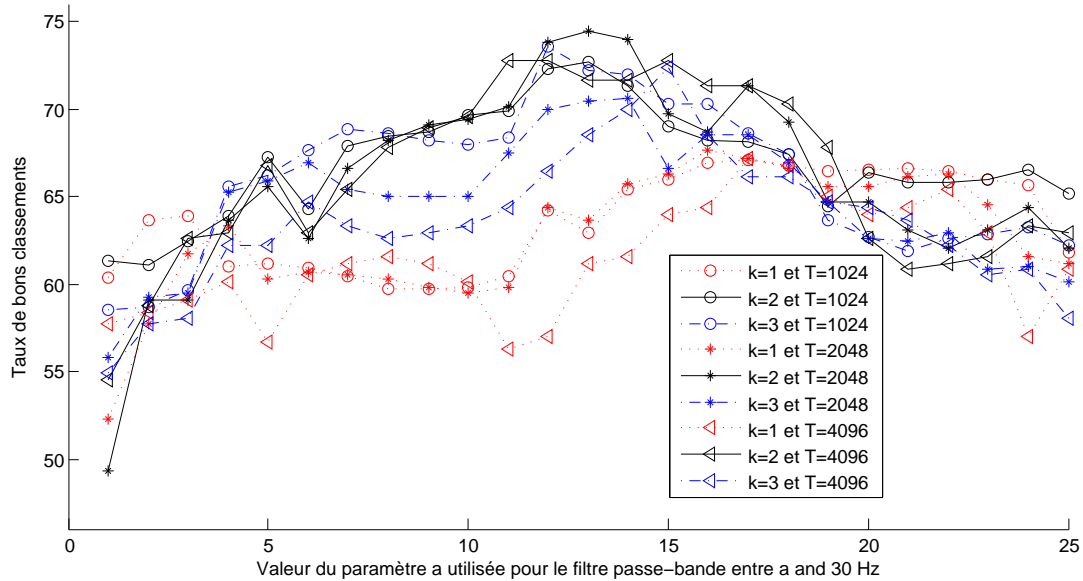


Figure V.5 – TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). **Les données sont normalisées en utilisant l'inertie** puis sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L'extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP.

$$n^* = 22 \text{ essais } \left(\lfloor \frac{46000}{2048} \rfloor = 22 \right).$$

- Normalisation : $\tilde{X}_i = X_i D$.
- Nombre de paires de filtres CSP : 3.

Les valeurs de a , b et k semblent cohérentes avec les travaux antérieurs (voir [72]). L'approche qui vient d'être présentée permet d'obtenir un TBC élevé sur les données en utilisant toutes les variables (75.35% sur les sujets de la première campagne). Une approche permettant de sélectionner un sous groupe d'électrodes pour prédire l'état de vigilance des sujets va maintenant être présentée.

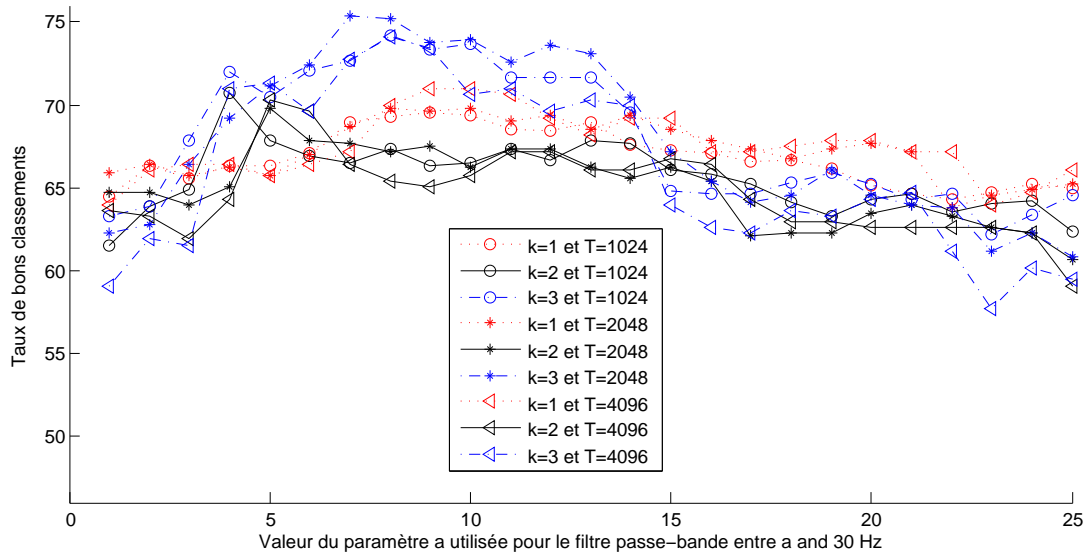


Figure V.6 – TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre a et 30 Hz avec a qui est un entier entre 1 et 25). **Les données sont normalisées en utilisant les écarts types** puis sont découpées en essais de 4 (cercles), 8 (croix) ou 16 (triangles) secondes (correspondant respectivement à $T = 1024$, $T = 2048$ et $T = 4096$ points). L'extraction de caractéristiques est réalisée en utilisant 1 (ligne en pointillé), 2 (ligne pleine) ou 3 (ligne en pointillé mixte) paires de filtres CSP.

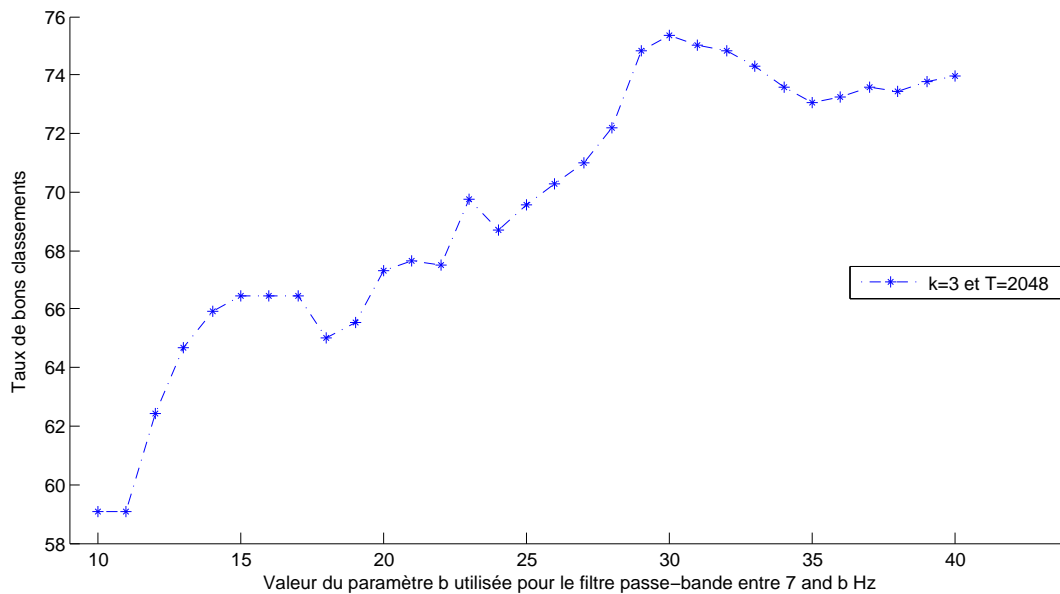


Figure V.7 – TBC estimés par LOO sur les données de la première campagne pour différentes stratégies de filtrage (filtrage passe-bande entre 7 et b Hz avec b qui est un entier entre 10 et 40). **Les données sont normalisées en utilisant les écarts types** puis sont découpées en essais de 8 secondes (correspondant à $T = 2048$ points). L'extraction de caractéristiques est réalisée en utilisant 3 paires de filtres CSP.

4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

Le but de ce travail est de sélectionner un sous-ensemble de p' électrodes (où $p' < p$) qui permet d'obtenir un modèle de prédiction fiable de l'état de vigilance d'un nouvel individu. Soit Ω l'ensemble des combinaisons de p' électrodes choisies parmi les p électrodes initiales. Le cardinal de Ω est égal à $|\Omega| = C_p^{p'}$. Dans ce travail, $p = 58$ donc $|\Omega|$ augmente très rapidement en fonction de p' (pour $p' = 1, \dots, 29$, ce qui correspond aux valeurs qui sont intéressantes dans ces travaux), rendant la recherche exhaustive infaisable. Ainsi, l'approche privilégiée dans ce travail est d'utiliser un algorithme génétique pour explorer l'espace de recherche Ω .

4.1 Choix algorithmiques

Un algorithme génétique a déjà été utilisé en Chapitre IV pour sélectionner une électrode et une bande de fréquences sur laquelle extraire une caractéristique utile pour la tâche de classification. La présente approche s'inscrit dans la continuité de celle décrite en Chapitre IV. Elle va permettre de non plus sélectionner une unique électrode mais une combinaison d'électrodes. De plus, l'un des principaux inconvénient de l'approche présentée en Chapitre IV reposait sur le fait que le trop faible nombre de sujets ne permettait pas d'envisager une séparation de ces derniers en deux groupes afin de ne rechercher la caractéristique que sur l'un d'eux. Depuis, le nombre de sujets dans l'étude a augmenté (par le biais de la seconde campagne d'acquisition de sujets). La méthode d'extraction de caractéristiques est également différente de celle utilisée en Chapitre IV. Cependant, des similitudes entre les structures des deux algorithmes génétiques existent. Par exemple, la structure du génome est très similaire à celle du Chapitre IV. En effet, dans ce travail, le génome est composé de $p = 58$ composantes binaires correspondant à l'inclusion ou non de chacune des électrodes pour le calcul des filtres CSP. Une représentation d'un génome est donnée en Figure V.8.

Sur cette Figure, le génome est formé de 58 composantes dont p' sont égales à 1. Les signaux des p' électrodes associées à ces composantes égales à 1 sont conservées (traits bleus) tandis que les autres sont éliminées (traits rouges). Ainsi, les matrices $\mathbf{X1}$ et $\mathbf{X2}$, initialement de dimensions respectives $Tn_1 \times p$ et $Tn_2 \times p$ deviennent des matrices de dimensions respectives $Tn_1 \times p'$ et $Tn_2 \times p'$. Les filtres CSP sont alors calculés à partir de ces matrices privées des signaux des électrodes associées à des 0 dans le génome.

Il existe deux façons de réaliser l'algorithme génétique. La première consiste à fixer la valeur

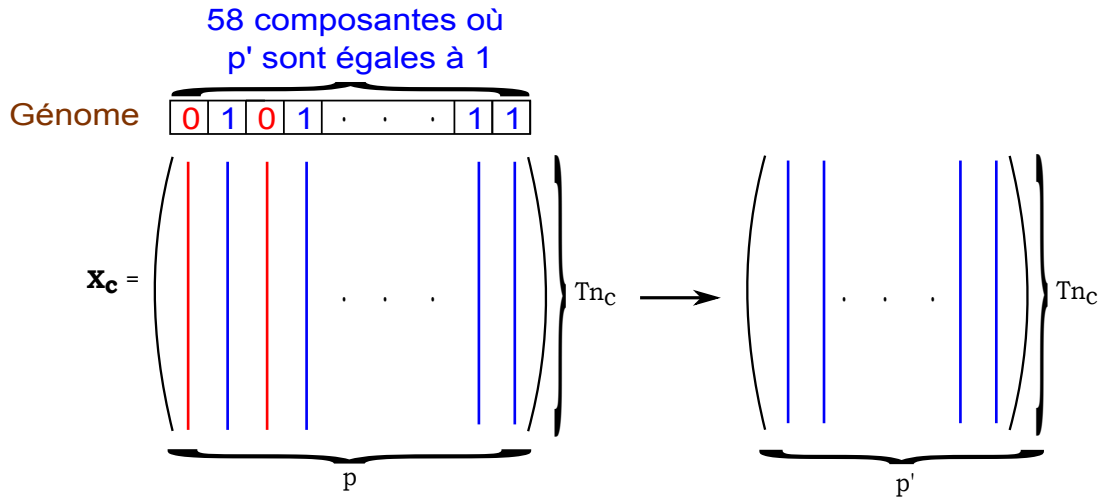


Figure V.8 – Exemple d’un génome dans l’algorithme génétique. Les composantes égales à 1 conduisent à la conservation du signal des électrodes associées (traits bleus) tandis que les composantes égales à 0 conduisent au rejet du signal des électrodes concernées (traits rouges).

de p' et à lancer l’algorithme génétique plusieurs fois pour différentes valeurs de p' . L’algorithme génétique va alors rechercher le sous-ensemble de p' électrodes qui conduit à l’obtention du modèle de prédiction le plus fiable. La seconde approche consiste à laisser l’algorithme choisir le meilleur sous-ensemble d’électrodes tout en le laissant également choisir la valeur de p' adéquate. L’algorithme génétique devient alors multi-objectif.

Ici, la première approche a été privilégiée. Le nombre d’électrodes conservées, p' , a été fixé et gardé constant au cours de l’algorithme génétique. Un compromis entre la qualité de la prédiction obtenue à l’aide du modèle de classification et le nombre d’électrodes nécessaires doit être trouvé. Ainsi, nous avons choisi de visualiser l’évolution de la qualité de la prédiction en fonction de p' . La fonction objectif (fitness) ne doit donc pas dépendre de p' . De plus, il est important de laisser l’opportunité aux neurobiologistes de contribuer en apportant leur expérience après avoir visualisé les résultats obtenus pour différentes valeurs de p' (qualité de la prédiction obtenue, choix des électrodes). Il est aussi envisageable que l’objectif évolue dans le temps et qu’un plus petit (ou plus grand) sous-ensemble d’électrodes soit désiré. Ainsi, il nous a semblé préférable de se donner la possibilité de déterminer la valeur de p' après une analyse visuelle des résultats obtenus pour différentes valeurs de p' .

4.1.1 Fonction d’évaluation

Chaque génome définit un sous-ensemble d’électrodes à considérer et conduit ainsi à l’obtention de filtres CSP. Un taux de bons classements peut donc être calculé pour chacun des

V.4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

génomés. Celui-ci peut être relié à la qualité du génome considéré. Il semble donc naturel que la fonction de fitness corresponde au TBC de chacun des génomes.

Pour calculer ce TBC, les données vont être pré-traitées comme expliqué en Section 3.3. Une estimation du TBC est alors obtenue comme décrit en Section 3.3.1. Une itération de cette procédure est détaillée en Figure V.9. Pour une itération donnée de LOO, un ensemble d'apprentissage et un ensemble test sont créés. Ainsi, pour un génome donné, les variables absentes du sous-ensemble d'électrodes défini par celui-ci vont être provisoirement éliminées des données EEG des deux échantillons (apprentissage et test). La méthode CSP va alors permettre d'apprendre les filtres spatiaux sur l'ensemble d'apprentissage. Les essais de cet ensemble vont être projetés et une transformation log-var va être appliquée sur le résultat de cette projection. Une règle de décision va alors être construite en utilisant la méthode LDA. Les prédictions des individus de l'ensemble test sont obtenues en utilisant le modèle (filtres CSP et règle de décision) formé sur l'ensemble d'apprentissage. Les prédictions obtenues sont comparées aux classes réelles auxquelles appartiennent les essais de l'ensemble test afin d'obtenir un TBC.

Pour un génome donné, la fonction de fitness s'écrit donc :

$$g(\mathbf{u}) = \frac{\# \text{ essais correctement classés}}{\# \text{ d'essais}}.$$

L'algorithme génétique recherche le génome u qui maximise la fonction g .

Remarque : Si l'approche multi-objectif avait été privilégiée, la fonction de fitness se décomposerait en deux parties :

- la première serait identique à la fonction de fitness qui vient d'être décrite et correspondrait donc à une mesure de la qualité du modèle prédictif obtenu.
- la seconde correspondrait à une pénalité sur le nombre de variables incluses dans le modèle prédictif (p').

Ainsi, la fonction fitness pourrait s'écrire :

$$g'(\mathbf{u}) = \lambda \frac{\# \text{ essais correctement classés}}{\# \text{ d'essais}} + (1 - \lambda) \frac{58 - p'}{58}.$$

où $\lambda \in [0, 1]$ est un poids fixé à l'avance afin de définir le compromis que l'on souhaite obtenir entre la qualité de prédiction du modèle obtenu et le nombre de variables.

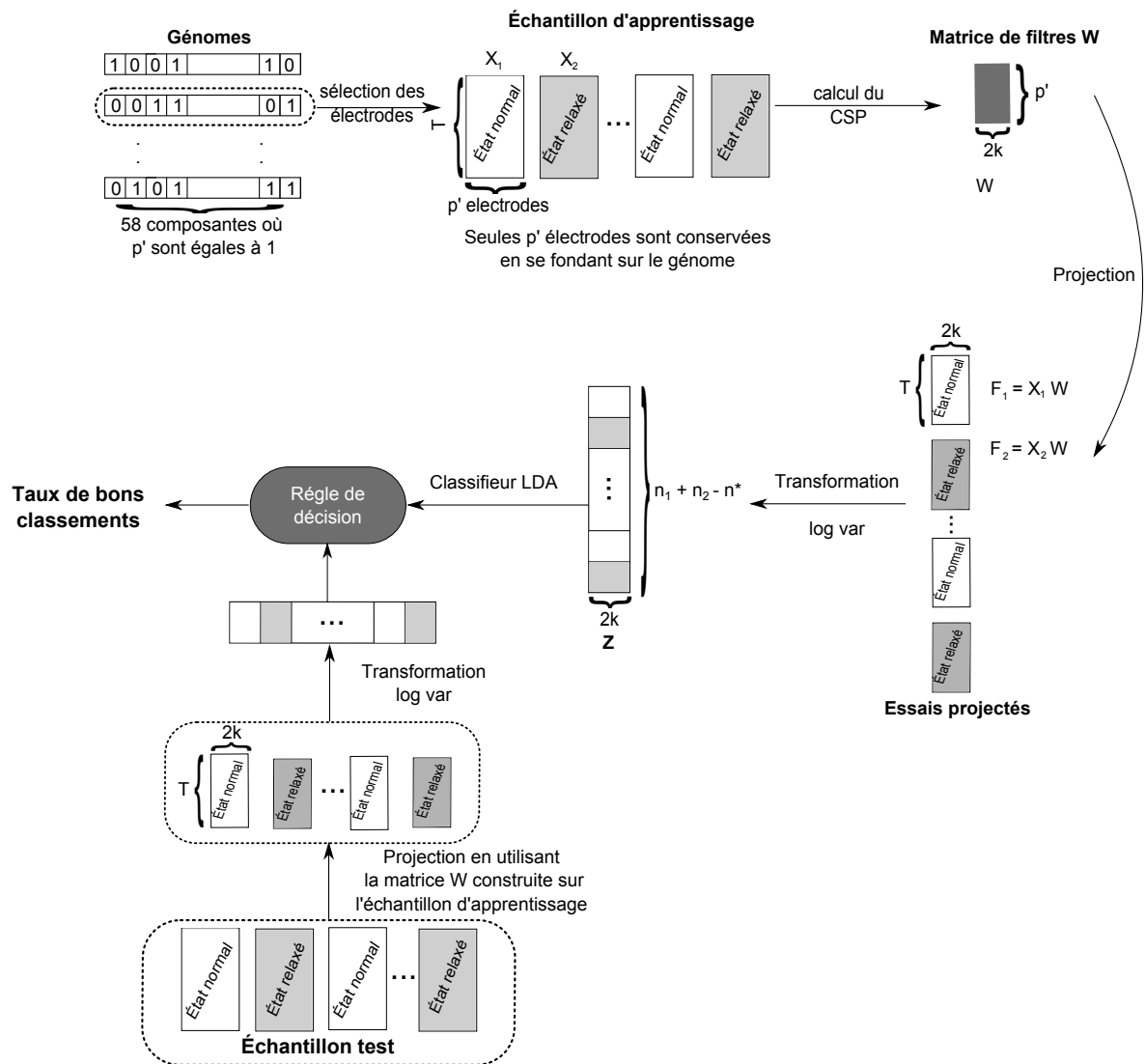


Figure V.9 – Diagramme représentant l'étape d'évaluation d'un génome lors de l'algorithme génétique. Pour un génome donné, un sous-ensemble d'électrode est sélectionné et un CSP est utilisé afin de construire la matrice des filtres W . Les essais sont alors projetés et une transformation log-var est appliquée pour obtenir Z . Une règle de décision est apprise sur l'ensemble d'apprentissage à l'aide de la méthode LDA. Un TBC est obtenu en utilisant le modèle construit (filtres CSP et règle de décision) sur les essais de l'ensemble test.

4.1.2 Opérateurs génétiques

La création d'un enfant est réalisée de la même façon que dans le Chapitre IV. Ainsi, deux parents sont sélectionnés par tournoi avec une pression de sélection égale à 2. Un croisement entre les deux parents permet de créer l'enfant. Afin de garder p' constant (nombre de com-

V.4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

posantes du génome égales à 1) au cours d'un lancement de l'algorithme génétique, il convient d'être vigilant lors des étapes de croisement et de mutation. On remarque que l'opérateur logique défini en Table IV.2 ne garantit pas le maintien du nombre de composantes égales à 1 identique entre les parents et l'enfant. Ainsi, un nouveau croisement est utilisé. Celui-ci est composé de deux phases représentées en Figure V.10.

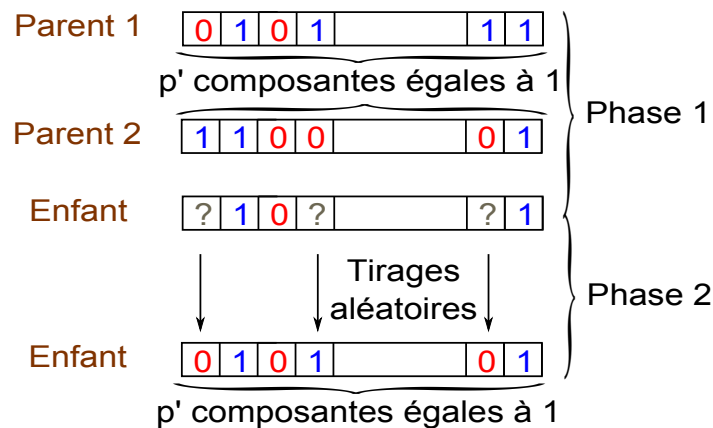


Figure V.10 – Représentation schématique de la méthode de croisement utilisée dans l'algorithme génétique.

La première phase consiste à reproduire chez l'enfant les composantes qui sont identiques chez les deux parents. Par exemple, lorsque les deux parents ont une composante égale à 1 à une position donnée, l'enfant hérite d'un 1 à cette même position. Les composantes non identiques chez les parents donnent des valeurs indéterminées chez l'enfant.

La seconde phase consiste à donner une valeur à ces composantes indéterminées. À ce stade, l'enfant possède un nombre de composantes non nulles $d \leq p'$. Afin de garantir que l'enfant dispose de p' composantes égales à 1, $p' - d$ positions parmi les composantes indéterminées sont tirées aléatoirement et sont remplacées par des 1. Les autres composantes indéterminées sont remplacées par des 0.

La mutation doit également conserver le nombre de composantes égales à 1 constant. Ainsi, lorsqu'une mutation affecte une composante du génome, une autre composante est tirée au hasard. Les valeurs des deux composantes sont alors permutées. Il se peut donc que la mutation soit muette (remplacement d'un 1 par un 1 ou d'un 0 par un 0). La probabilité de mutation est de $2/p$ (mutations muettes incluses).

4.1.3 Critère d'arrêt

Les différents critères d'arrêt définis en Section 4.1.3 du Chapitre IV ont été réutilisés. Cependant, la méthode pour calculer le nombre de différence est simplifiée du fait que toutes les composantes du génomes sont binaires. Ainsi, pour calculer le nombre de différences au sein de la population, les génomes sont codés en base 2 et le nombre d'individus différents dans la population correspond au nombre de codes différents obtenus. De plus, un critère d'arrêt fondé sur la stagnation de la qualité du meilleur individu a été ajouté. Ainsi, si la meilleure évaluation trouvée est la même depuis 50 itérations, l'algorithme est arrêté. L'algorithme s'arrête si l'une de ces conditions est satisfaite :

- le nombre d'itérations dépasse 1 000.
- Les parents sont identiques depuis 10 générations.
- Le nombre de différence au sein des parents est inférieur à 3.
- La meilleure évaluation est identique depuis 50 générations.

4.2 Application expérimentale

L'algorithme génétique, le CSP et la méthode de prédiction LDA ont été codées en Matlab et exécuté sur un ordinateur dont la configuration est la suivante : Linux avec 8*2.8 GHz Intel Xeon processeurs X5560 et 32 GB 1333 MHz DDR3 ECC RAM.

4.2.1 Stratégie expérimentale

L'algorithme génétique est lancé pour $p' = 1, 2, \dots, 15, 20, 30, 40$ et 50 avec 150 parents et 50 enfants. Chaque lancement de l'algorithme nécessite entre 9 et 15 heures, en fonction de la valeur de p' considérée. C'est pourquoi, l'algorithme génétique n'est lancé que 15 fois pour chaque valeur de p' .

Lorsque l'algorithme converge, le meilleur génome trouvé est conservé. Ainsi, 15 génomes ainsi que leur TBC respectifs vont être obtenus pour chaque valeur de p' . En analysant ces résultats, le meilleur compromis entre le nombre d'électrodes sélectionnées (p') et la qualité de la prédiction obtenue (TBC) va être choisi, ce qui va permettre de fixer la valeur de p' . Le meilleur génome trouvé pour cette valeur de p' au cours des 15 lancements de l'algorithme sera conservé et permettra d'obtenir un sous-ensemble d'électrodes (composantes égales à 1 dans ce génome).

Un CSP sera alors lancé sur les données constituées uniquement des p' électrodes et des 13

V.4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

sujets de la première campagne d'acquisition. Ainsi, 3 paires de composantes CSP ($k = 3$) seront formées sur lesquelles une règle de décision sera apprise (LDA). Enfin, cette règle de décision et ces composantes CSP seront utilisées sur les sujets de l'ensemble test (seconde campagne) afin d'obtenir un TBC externe à l'algorithme génétique.

4.2.2 Résultats

Les performances de l'algorithme génétique, le choix de la valeur de p' et les électrodes sélectionnées vont être discutés dans les prochains paragraphes.

4.2.2.1 Performances de la méthode

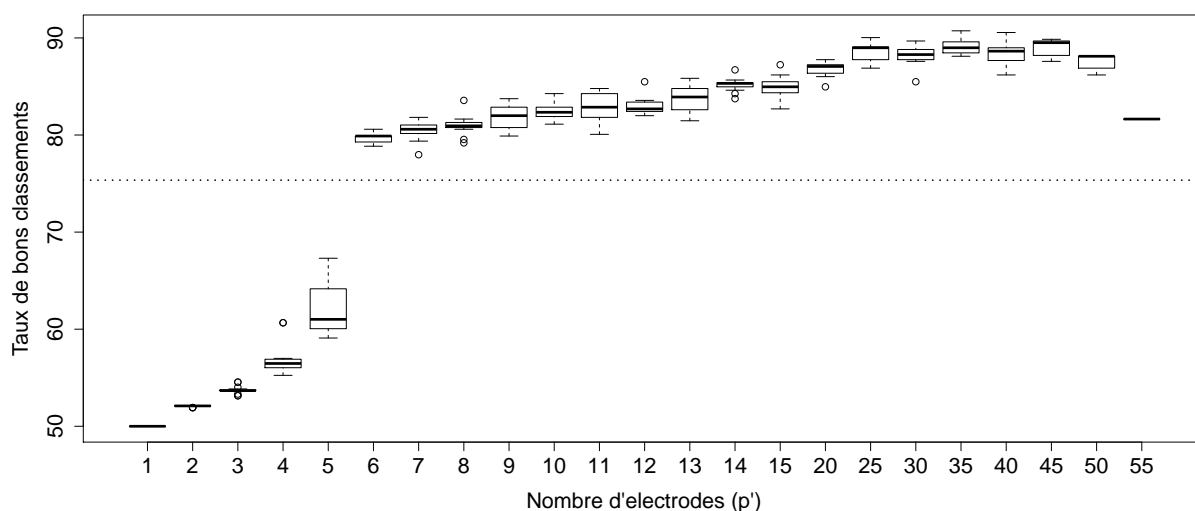


Figure V.11 – Taux de bons classements obtenus en LOO sur les sujets de la première campagne lors de l'algorithme génétique pour différentes valeurs de p' . La ligne en pointillé représente le taux de bons classements de 75.35% obtenu en utilisant les 58 variables initiales (pas de sélection de variable).

La Figure V.11 représente, pour chaque valeur de p' , une boîte à moustaches des 15 TBC (un par lancement) associés aux meilleurs génomes trouvés par l'algorithme génétique. La ligne en pointillé représente le TBC obtenu par LOO en appliquant un CSP sur les 58 variables initiales et sur les individus de la première campagne d'acquisition (pas de sélection de variable). Puisque le premier but de ce travail est de sélectionner un sous-ensemble de variables (électrodes) de faible taille, une attention plus prononcée va être portée sur les résultats obtenus pour p' entre 1 et 15. Les résultats pour $p' = 20, 25, 30, 35, 40, 45, 50$ et 55 sont juste donnés pour évaluer le comportement général de cette approche.

Chapitre V. Algorithme génétique pour la sélection de variables dans les filtres CSP.

Sur la Figure V.11, il apparaît que les médianes de TBC augmentent globalement entre $p' = 1$ et $p' = 50$ (excepté pour $p' = 15$ et 50). Une augmentation importante se produit entre $p' = 5$ et $p' = 6$ où la médiane du TBC atteint les 80%. Ainsi, de $p' = 6$ à $p' = 15, 20, 25, 30, 35, 40, 45, 50$ et 55 , les TBC obtenus par cette approche dépassent le TBC de référence obtenu en utilisant la méthode CSP sur les 58 variables initiales (ligne en pointillé). Les résultats commencent à décroître entre $p' = 40$ et $p' = 50$ avant d'atteindre la ligne en pointillé pour $p' = 58$. Ainsi, une chute importante se produit entre $p' = 50$ et $p' = 58$. Cela signifie que, parmi les 8 électrodes non sélectionnées pour $p' = 50$, certaines sont très bruitées ou non discriminantes et ces dernières détériorent la qualité de la prédiction. La méthode de sélection de variables prend alors ici tout son sens. De $p' = 6$ à 15 , les TBC obtenus par l'algorithme génétique semblent satisfaisant. Ils permettent d'envisager une utilisation des résultats de cette approche pour des applications futures.

Dans ces travaux, le but est de construire un modèle capable de prédire l'état de vigilance d'un nouvel individu. C'est pourquoi, il est important d'évaluer comment se comporte la méthode au niveau des sujets. Le signal EEG recueilli par une électrode chez un sujet dans un état de vigilance donné a été découpé en 22 essais. Pour un génome donné dans l'algorithme génétique, une prédiction est obtenue pour chacun de ces essais. Ainsi, il est possible de prédire l'état de vigilance d'un sujet en appliquant un vote majoritaire parmi ses essais. Étant donné que le nombre d'essais est pair, le premier essai est retiré afin d'éviter les éventuelles égalités.

Les résultats obtenus sont synthétisés dans le Tableau V.1. Il contient la moyenne du nombre d'itérations avant la convergence de l'algorithme génétique pour les différentes valeurs de p' considérées ainsi que la moyenne et l'écart-type des TBC obtenus lors de la convergence de l'algorithme génétique. Les résultats entre parenthèses sont obtenus en se plaçant au niveau des sujets (affectation par vote majoritaire) au lieu de considérer les résultats obtenus au niveau des essais. Il semble que la moyenne du nombre d'itérations ne dépasse pas 200. L'algorithme semble converger rapidement et le nombre d'itérations n'atteint jamais 1 000 (le premier critère d'arrêt présenté au paragraphe 4.1.3 n'est jamais vérifié). Dans la plupart des cas, l'algorithme génétique s'arrête parce que les parents sont les mêmes depuis 10 générations ou parce que la meilleure évaluation est la même depuis 50 générations. Une importante diversité est maintenue dans les génomes donc le nombre de différences entre les parents est rarement inférieur à 3. Les moyennes des TBC obtenus en considérant les essais sont assez proches des médianes observées dans les figures V.11. Au cours de l'algorithme génétique, la moyenne des TBC, obtenu en considérant les essais, augmente avec p' jusqu'à $p' = 14$. Lorsque $p > 5$, le TBC moyen est d'environ 80% avec de faibles valeurs d'écart type.

Les résultats obtenus pour les différentes valeurs de p' doivent être comparés aux résultats

V.4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

Tableau V.1 – Résultats obtenus en LOO dans l'algorithme génétique pour différentes valeurs de p' . Les résultats entre parenthèses sont obtenus en considérant les sujets au lieu de considérer les essais. La dernière ligne correspond au taux de bons classements référence obtenu en conservant les 58 variables initiales.

À l'intérieur de l'algorithme génétique : estimation du TBC par LOO sur les sujets de la première campagne			
p'	Moyenne du nombre d'itérations	Moyenne des TBC des meilleurs génomes	Écarts types des TBC des meilleurs génomes
1	12	50 (50)	0 (0)
2	55.93	52.06 (53.85)	0.07 (0)
3	71.07	53.73 (53.85)	0.41 (0)
4	78.6	56.85 (58.72)	1.66 (0.023)
5	87.07	62.05 (65.13)	2.68 (0.057)
6	111.6	79.72 (86.41)	0.59 (0.038)
7	108.2	80.43 (86.67)	0.96 (0.025)
8	130.53	81 (86.41)	0.97 (0.035)
9	121.2	81.89 (90.26)	1.28 (0.029)
10	141.47	82.47 (89.74)	0.93 (0.028)
11	137.87	82.84 (91.03)	1.59 (0.035)
12	141.67	82.96 (90.77)	0.87 (0.032)
13	146	83.8 (90.51)	1.36 (0.032)
14	169.07	85.19 (92.05)	0.68 (0.027)
15	163.67	84.93 (92.56)	1.19 (0.031)
20	165.73	86.75 (94.1)	0.72 (0.025)
25	185.53	88.5 (95.13)	0.95 (0.018)
30	168.47	88.22 (95.13)	0.99 (0.027)
35	180.6	89.16 (94.87)	0.8 (0.024)
40	168	88.41 (96.67)	1.29 (0.029)
45	181.13	89.06 (96.15)	0.8 (0.025)
50	145.8	87.49 (92.56)	0.81 (0.01)
55	72.53	81.64 (80.77)	0 (0)
58	-	75.35 (73.08)	-

obtenus sans effectuer de sélection d'électrodes (dernière ligne du tableau V.1). En effet, lorsqu'aucune sélection d'électrodes n'est considérée, le TBC obtenu par LOO sur les 13 sujets de la première campagne est égal à 75.35%. Ainsi, les résultats obtenus en opérant une réduction de la dimension (en utilisant l'algorithme génétique) dépassent 75.35% dès que $p' > 5$.

En analysant la Table V.1 et la Figure V.11, il semble que les résultats se stabilisent entre $p' = 9$ et $p' = 12$. Un test statistique est utilisé afin de déterminer si ces valeurs de p' donnent des résultats significativement différents. Pour une valeur de p' donnée, les TBC obtenus ne suivent pas une distribution gaussienne (test de Shapiro-Wilk avec un niveau de 0.05). Ainsi un test de Kruskal-Wallis (KW) est utilisé. Dans ce test non paramétrique, l'hypothèse nulle correspond à " \mathcal{H}_0 : il n'y a pas de différence entre les TBC obtenus pour le groupe de valeurs de p' testées " et l'alternative correspond à " \mathcal{H}_1 : Il existe des différence entre les TBC obtenus au sein du groupe de valeurs de p' testées". Lorsque le test de KW est utilisé sur un groupe de valeurs de p' , une p -valeur est obtenue. Une valeur de celle-ci supérieure au seuil α fixé conduit à accepter l'hypothèse \mathcal{H}_0 . Dans ce travail, $\alpha = 0.05$. Selon le test de KW, il n'y a pas de différence

significative entre les TBC de $p' = 9, 10, 11$ et 12 (p -valeur égale à 0.116). Cependant, selon le test de KW, une différence significative existe entre les TBC obtenus pour $p' = 8, 9, 10, 11$ et 12 ($1.6 * 10^{-4}$) ainsi que pour ceux obtenus pour $p' = 9, 10, 11, 12$ et 13 ($7.31 * 10^{-3}$). Ces tests vont permettre d'appuyer le choix de la valeur de p' .

4.2.2.2 Choix de la valeur de p'

Un objectif de ce travail est de trouver un sous-ensemble de variables d'une taille faible qui permette d'obtenir une prédiction fiable (taux de bons classements élevé) de l'état de vigilance des sujets. Il est nécessaire de choisir le meilleur compromis entre la taille de ce sous-ensemble (p') et la qualité de la prédiction obtenue (TBC). Le choix de la valeur de p' va être réalisé en analysant les résultats obtenus à l'aide de l'algorithme génétique sur les sujets de la première campagne. À l'aide des Figures V.11 et du Tableau V.1, trois valeurs de p' semblent représenter un compromis acceptable :

- Une augmentation importante intervient entre $p' = 5$ et $p' = 6$. À partir de $p' = 6$, les résultats sont autour de 80% . Ainsi, la plus petite valeur de p' qui donne des résultats satisfaisants est $p' = 6$.
- Sur la Figure V.11, il apparaît que les résultats continuent de s'améliorer entre $p' = 6$ et $p' = 14$. Cependant, ces derniers semblent stagner entre $p' = 10$ et $p' = 12$. Selon le test statistique présenté en Section 4.2.2.1, les taux de bons classements obtenus pour $p' = 9, 10, 11$ et 12 ne sont pas significativement différents. Ainsi, le choix de $p' = 9$ peut représenter un bon compromis.
- Les meilleurs résultats obtenus entre $p' = 1$ et 15 sont atteints pour $p' = 14$.

Ainsi, le choix de la valeur de p' dépend du but des travaux. Dans notre cas, le meilleur compromis semble être obtenu pour $p' = 9$. Cette valeur représente un compromis entre un modèle de prédiction fiable obtenu avec un nombre d'électrodes faible ($p' = 6$) et un modèle encore plus fiable mais nécessitant un nombre plus important d'électrodes ($p' = 14$). C'est pourquoi, nous avons décidé de considérer $p' = 9$ comme valeur définitive de p' (ligne en gras dans le Tableau V.1). Seuls les 15 meilleurs génomes trouvés par l'algorithme génétique (un par lancement) pour $p' = 9$ sont évalués sur l'échantillon test (sujets de la campagne 2). Au niveau des essais, la moyenne des 15 TBC obtenus sur l'échantillon test est de 76.29% avec un écart type de 4.35 . Le meilleur génome trouvé au cours de ces 15 lancements de l'algorithme génétique pour $p' = 9$ est conservé (en se fondant sur l'erreur obtenue en LOO sur les sujets de la première campagne). Il permet d'obtenir une erreur sur l'échantillon test de 71.59% en considérant les essais et de 75% en considérant les sujets.

4.2.2.3 Électrodes sélectionnées

Pour chaque valeur de p' , 15 sous-ensembles de variables de longueur p' sont obtenus (1 par lancement de l'algorithme génétique) correspondant aux meilleurs génomes trouvés lors de ces lancements. La Figure V.12 représente l'évolution de ces sous-ensembles d'électrodes pour différentes valeurs de p' . Le Graphique V.12(a) montre les occurrences et les localisations des électrodes sélectionnées dans les meilleurs génomes trouvés par l'algorithme génétique pour $p' = 3$. Ainsi, pour une valeur de p' , une électrode qui a été sélectionnée dans tous les meilleurs génomes trouvés par l'algorithme génétique va avoir un score égal à 15 et va apparaître en noir. Des graphiques similaires sont proposés en Figure V.12(b), V.12(c), V.12(g) et V.12(h) pour respectivement $p' = 6, 9, 12$ et 15 . En comparant les Figures V.12(b) et V.12(c), il apparaît que les électrodes sélectionnées (apparaissant en noir) pour $p' = 6$ sont conservées pour $p' = 9$. Les mêmes observations sont obtenues pour $p' = 12$ et 15 . Cependant, pour $p' = 3$, les résultats sont moins stables. L'électrode la plus souvent choisie n'est sélectionnée que 5 fois au cours des 15 lancements. Il semble que cette instabilité soit due au fait que plusieurs sous-ensembles de 3 électrodes donnent des résultats équivalents. Le choix de l'algorithme génétique tourne donc autour de ces différents sous-ensembles.

Ainsi, une électrode qui est sélectionnée régulièrement pour une valeur de p' donnée est généralement conservée pour des valeurs de p' plus élevées. Ce résultat, bien qu'espéré, n'était pas garanti. En effet, chaque lancement de l'algorithme génétique est indépendant des autres lancements. En particulier, l'initialisation de l'algorithme génétique est aléatoire et n'utilise donc pas les résultats obtenus lors d'autres lancements de l'algorithme. Ainsi, ce résultat est très important car il montre une certaine stabilité qui tend à prouver que les signaux recueillis par les électrodes concernées sont utiles pour la tâche de classification. De plus, pour $p' = 6, 9, 12$ et 15 , les sous-ensembles d'électrodes sélectionnés semblent très stables. En effet, la plupart des électrodes sélectionnées apparaissent en noir.

La Figure V.12(d) est une représentation en 3D de chaque sous-ensemble d'électrodes sélectionnées en fonction de p' . Pour une électrode donnée et un lancement donné, une valeur égale à 1 sur le Figure V.12(d) signifie que l'électrode en question a été sélectionnée au cours de ce lancement. Une fois encore, des graphiques similaires sont proposés en Figures V.12(e), V.12(f), V.12(i) et V.12(j) pour respectivement $p' = 6, 9, 12$ et 15 . Ces figures permettent de vérifier que les sous-ensembles d'électrodes sélectionnés par l'algorithme génétique lors des 15

lancements sont très stables pour une valeur de p' donnée.

Pour une valeur de p' et une électrode donnée, la Figure V.13 représente le nombre de sélection de cette électrode dans les meilleurs sous-ensembles de variables trouvés par l'algorithme génétique au cours des 15 lancements. Cette figure montre que lorsqu'une électrode est sélectionnée pour une valeur donnée de p' , cette électrode est conservée pour des valeurs de p' supérieures. De plus, pour une valeur de p' donnée, des pics excédant les 10 sélections (sur 15 possibles) sont obtenus. Ainsi, les sous-ensembles de variables sélectionnées semblent relativement stables pour une valeur de p' donnée.

La Figure V.14 représente le sous-ensemble d'électrodes sélectionnées lors du meilleur des 15 lancements pour $p' = 9$. Ce sous-ensemble contient des électrodes préfrontales (FPz and FP2) et frontales (F3, F5 and F7) ainsi qu'une électrode fronto-centrale (FC5), temporale (T3), centrale (C5) et temporal-pariétale (TP7). On constate que les électrodes sélectionnées sont sur les contours du scalp.

V.4 Sélection d'un sous-ensemble d'électrodes par algorithmes génétique

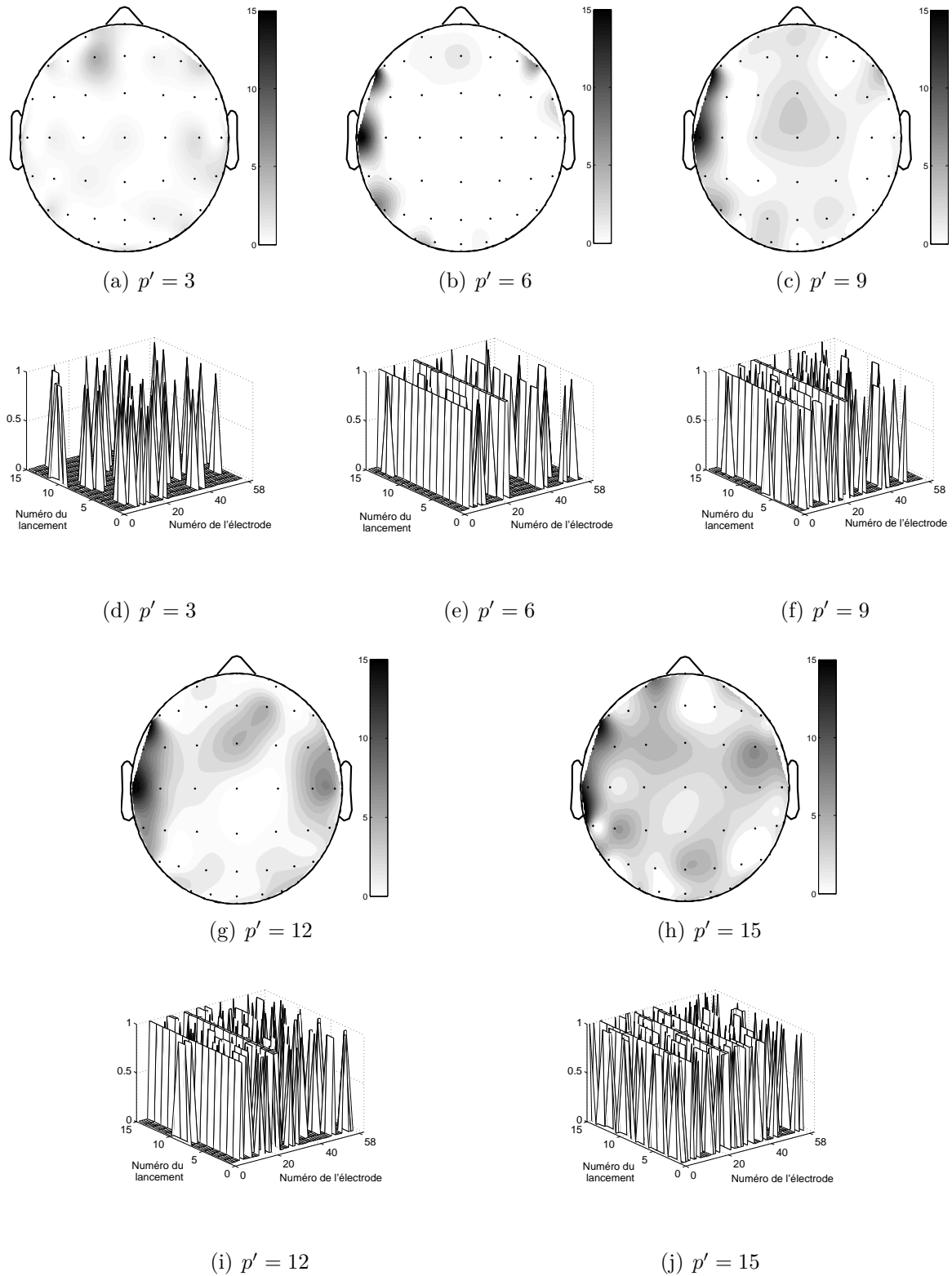


Figure V.12 – Les figures a, b, c, g et h représentent le nombre de sélection de chaque électrode dans le meilleur génome (1 par lancement) trouvé par l’algorithme génétique lors des 15 lancements respectivement pour $p' = 3, 6, 9, 12$ et 15. Les figures d, e, f, i et j sont des représentations en 3D qui permettent de visualiser les meilleurs génomes trouvés par l’algorithme génétique lors des 15 lancements pour respectivement $p' = 3, 6, 9, 12$ et 15.

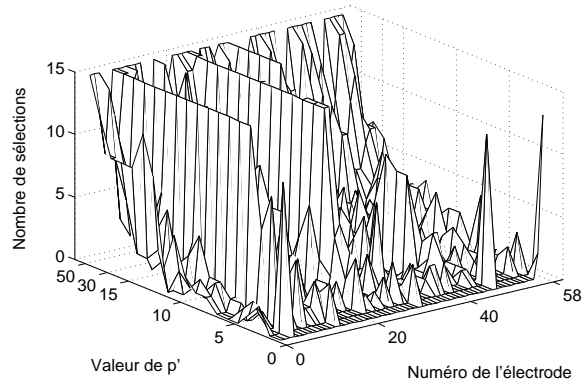


Figure V.13 – Représentation 3D permettant de visualiser le nombre de sélections de chaque électrode dans les meilleurs sous-ensemble d'électrodes trouvés par l'algorithme génétique au cours des 15 lancements.

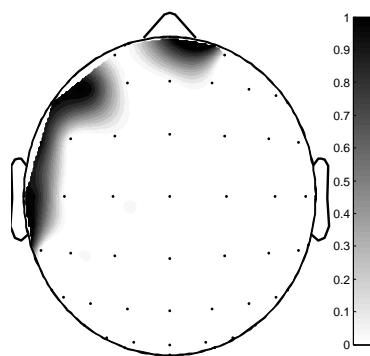


Figure V.14 – Meilleure combinaison de 9 électrodes obtenue pour le meilleur des 15 lancements de l'algorithme génétique avec $p' = 9$. Les électrodes sélectionnées apparaissent en noirs. Cette figure est obtenue en utilisant la Visual RCSP toolbox [14].

5 Sélection de variables par méthode séquentielle

Les résultats obtenus en utilisant l'algorithme génétique semblent convaincants. Cependant, le coût computationnel de cette approche peut représenter un frein à son utilisation. Ainsi, une autre approche visant à sélectionner un sous groupe d'électrodes tout en limitant le coût computationnel va être proposée. Cette approche se différencie de la précédente par la façon dont elle parcourt l'espace de recherche. Elle est fondée sur les méthodes de recherche séquentielle présentées en Section 3.2 du Chapitre III. Deux procédures de sélection de variables respectivement fondées sur la méthode *forward* et la méthode *backward* vont être proposées.

5.1 Le forward et le backward CSP

La méthode forward est une méthode de sélection de variable itérative qui consiste à ajouter une variable lors de chaque itération. L'algorithme débute en évaluant tous les modèles obtenus en utilisant qu'une seule variable (p évaluations). La procédure d'évaluation est identique à celle qui a été utilisée dans l'algorithme génétique (voir Section 4.1.1). Ainsi, un TBC estimé par LOO (voir section 3.3.1) est obtenu pour chaque modèle formé d'une unique variable. La variable qui permet de former le modèle associé à la meilleure évaluation est conservée. Lors de la deuxième itération, les $p-1$ modèles obtenus en ajoutant une variable à celle déjà sélectionnée sont évalués et le meilleur d'entre eux permet d'obtenir un sous-ensemble de deux électrodes. La procédure est itérée de la sorte jusqu'à ce qu'un critère d'arrêt soit atteint. Dans ces travaux, l'algorithme ne s'arrête que lorsque les p itérations ont été réalisées (voir Algorithme 1).

Le backward CSP fonctionne de manière inverse au forward CSP. Il part du modèle complet à p variables et retire à chaque itération la variable sans laquelle le modèle résultant a obtenu la meilleure évaluation (voir Algorithme 2). Ainsi, le forward CSP et le backward CSP nécessitent $\frac{p(p+1)}{2} = 1\,711$ évaluations chacun, correspondant à environ 10 itérations de l'algorithme génétique (l'initialisation de celui-ci étant composée des 300 évaluations des parents). L'algorithme génétique réalisant en moyenne 125 itérations avant d'atteindre un critère d'arrêt (toutes valeurs de p' confondues, voir Table V.1), il apparait que celui-ci teste beaucoup plus de solutions potentielles que les procédures de backward et forward CSP (18 900 évaluations en moyenne pour l'algorithme génétique contre 1 711 pour les procédures séquentielles). Étant donné que la méthode d'évaluation utilisée est déterministe, le forward CSP et le backward CSP sont également déterministes. Contrairement à l'algorithme génétique, il sera donc inutile de lancer plusieurs fois ces méthodes sur les mêmes données. Le nombre d'évaluations à réaliser ainsi que leur caractère déterministe font de ces approches des alternatives envisageables à

l'algorithme génétique.

5.2 Résultats

De manière analogue à ce qui a été décrit pour l'algorithme génétique, seuls les signaux EEG des sujets de la première campagne sont utilisés lors du forward et du backward CSP. Les signaux EEG des sujets de la seconde campagne permettront d'évaluer les performances prédictives (TBC) du modèle formé à l'aide du sous-ensemble de variables sélectionnées. La Figure V.15 présente sur un même graphique les TBC obtenus en sélectionnant les électrodes par backward CSP (courbe noire), forward CSP (courbe bleue) et par algorithme génétique (courbe rouge) en fonction du nombre d'électrodes sélectionnées (p'). Afin de pouvoir se comparer aux résultats de l'algorithme génétique, lorsque $p' > 15$, seuls les résultats obtenus pour $p' = 20, 25, 30, 35, 40, 45, 50$ et 55 sont représentés.

Sur cette figure, Les TBC obtenus pour ces deux approches en fonction du nombre de variables sélectionnées sont représentés en Figure V.15. Sur cette figure, alors que l'on aurait pu s'attendre à ce que le TBC continue de s'améliorer lorsque p' augmente, il apparaît que celui-ci décroît (lorsque $p' > 50$ pour l'algorithme génétique et le backward CSP et dès que $p' > 30$ pour le forward CSP). Ce résultat confirme que l'inclusion de certaines électrodes nuit à la qualité prédictive. Cela tend à valider l'approche visant à sélectionner des électrodes en EEG.

Il apparaît que les TBC obtenus par forward CSP ont une croissance moins franche et une décroissance plus précoce que les deux autres méthodes. Les résultats obtenus par backward CSP et par algorithme génétique sont très proches même si l'algorithme génétique semble dominer les deux autres méthodes de sélection d'électrodes.

Le fait que de meilleurs résultats soient obtenus en recherchant le sous-ensemble d'électrodes par algorithme génétique n'est pas surprenant. En effet, ce dernier teste beaucoup plus de solutions potentielles et propose un parcours de l'espace de recherche qui semble plus opportun. En effet, comme évoqué précédemment, avec les paramètres utilisés pour l'algorithme génétique, seule 10 itérations suffisent pour tester autant de solutions potentielles que les approches séquentielles. L'algorithme génétique, avant de converger, va donc améliorer ses chances de se rapprocher de la meilleure solution possible (optimum global). De plus, ses opérateurs génétiques lui permettent d'éviter de stagner autour des optima locaux et d'améliorer le parcours de l'espace de recherche. Ces opérateurs sont l'un des atouts indéniables de l'algorithme génétique par rapport aux méthodes de recherche séquentielle .

Algorithme 1 Le forward CSP

Entrée(s) $\tilde{\mathbf{X}}_1$: Signaux EEG correspondant aux essais des sujets de la première campagne d'acquisition enregistrés avant la séance de relaxation,
 $\tilde{\mathbf{X}}_2$: Signaux EEG correspondant aux essais des sujets de la première campagne d'acquisition enregistrés après la séance de relaxation,
 \mathbf{Y} : Vecteur contenant le numéro de la classe à laquelle appartiennent chaque essai des sujets de la première campagne d'acquisition.

Sortie(s) Obtention d'une liste ordonnée d'inclusion des variables (Elec_Sel), ainsi que les taux de bons classements correspondant (TBC).

Initialisation

 Elec_Sel = \emptyset
 Nb_Elec_Sel = 0
 Elec_testées = \emptyset

La liste des variables sélectionnées est initialisée vide et le nombre d'électrodes sélectionnées est égale à 0. La liste des électrodes testées est également initialisée vide.

Phase de recherche

tant que Nb_Elec_Sel $\neq p$ **faire**

 Elec_testées = Electrode \notin Elec_Sel

Les électrodes dont l'inclusion dans la liste des variables sélectionnées va être testées sont stockées.

 liste_eval = \emptyset

La liste des évaluations obtenues est initialisée vide.

pour Electrode \notin Elec_Sel **faire**

$(\tilde{\mathbf{X}}_1^{\text{update}}, \tilde{\mathbf{X}}_2^{\text{update}}) = \text{Update_X}(\text{Electrode}, \text{Elec_Sel}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)$

L'électrode est temporairement ajoutée à la liste des variables sélectionnées et les variables non sélectionnées sont retirées de $\tilde{\mathbf{X}}_1$ et de $\tilde{\mathbf{X}}_2$.

 TBC = *Evaluation*($\tilde{\mathbf{X}}_1^{\text{update}}, \tilde{\mathbf{X}}_2^{\text{update}}, \mathbf{Y}$)

Le sous-ensemble de variables est évalué et un TBC est obtenu par LOO. L'évaluation est réalisée en extrayant des caractéristiques par CSP et transformation log-var et en utilisant la LDA.

 liste_eval = *Update_liste_eval*(liste_eval, TBC)

L'évaluation obtenue est stockée avec celles déjà obtenues pour les autres électrodes.

fin du pour

 Electrode = Choisir_Elec(Elec_testées, liste_eval)

Sélection de l'électrode qui a permis d'obtenir la meilleure évaluation.

 Elec_Sel = Update_Elec_Sel(Elec_Sel, Electrode)

La liste des variables sélectionnées est complétée en ajoutant celle qui vient d'être choisie.

 Nb_Elec_Sel = Nb_Elec_Sel + 1

Le nombre de variables sélectionnées est incrémenté.

fin du tant que

Retour Elec_Sel, Liste_TBC

Algorithme 2 Le backward CSP

Entrée(s) $\tilde{\mathbf{X}}_1$: Signaux EEG correspondant aux essais des sujets de la première campagne d'acquisition enregistrés avant la séance de relaxation,
 $\tilde{\mathbf{X}}_2$: Signaux EEG correspondant aux essais des sujets de la première campagne d'acquisition enregistrés après la séance de relaxation,
 \mathbf{Y} : Vecteur contenant le numéro de la classe à laquelle appartiennent chaque essai des sujets de la première campagne d'acquisition.

Sortie(s) Obtention d'une liste ordonnée de retrait des variables (Elec_Sel), ainsi que les taux de bons classements correspondant (TBC).

Initialisation

Elec_Sel = $[1, 2, \dots, p]$

Nb_Elec_Sel = p

Elec_testées = \emptyset

Toutes les électrodes sont incluses dans la liste des variables sélectionnées et le nombre d'électrodes sélectionnées est égale à p . La liste des électrodes testées est initialisée vide.

Phase de recherche

tant que Nb_Elec_Sel $\neq 1$ **faire**

 Elec_testées = Electrode \in Elec_Sel

Les électrodes dont le retrait de la liste des variables sélectionnées va être testées sont stockées.

 liste_eval = \emptyset

La liste des évaluations obtenues est initialisée vide.

pour Electrode \in Elec_Sel **faire**

$(\tilde{\mathbf{X}}_1^{\text{update}}, \tilde{\mathbf{X}}_2^{\text{update}}) = \text{Update_X}(\text{Electrode}, \text{Elec_Sel}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)$

L'électrode est temporairement retirée de la liste des variables sélectionnées et les variables non sélectionnées sont retirées de $\tilde{\mathbf{X}}_1$ et de $\tilde{\mathbf{X}}_2$.

 TBC = $\text{Evaluation}(\tilde{\mathbf{X}}_1^{\text{update}}, \tilde{\mathbf{X}}_2^{\text{update}}, \mathbf{Y})$

Le sous-ensemble de variables est évalué et un TBC est obtenu par LOO. L'évaluation est réalisée en extrayant des caractéristiques par CSP et transformation log-var et en utilisant la LDA.

 liste_eval = $\text{Update_liste_eval}(\text{liste_eval}, \text{TBC})$

L'évaluation obtenue est stockée avec celles déjà obtenues pour les autres électrodes.

fin du pour

 Electrode = Choisir_Elec(Elec_testées, liste_eval)

Retrait de l'électrode qui a permis d'obtenir la plus mauvaise évaluation.

 Elec_Sel = $\text{Update_Elec_Sel}(\text{Elec_Sel}, \text{Electrode})$

La liste des variables sélectionnées est mise à jour en retirant celle qui a donné la plus mauvaise évaluation.

 Nb_Elec_Sel = Nb_Elec_Sel + 1

Le nombre de variables sélectionnées est décrémenté.

fin du tant que

Retour Elec_Sel, Liste_TBC

V.5 Sélection de variables par méthode séquentielle

De plus, avec l'algorithme génétique proposé, les lancements pour deux valeurs de p' différentes sont indépendants. Ainsi, une électrode choisie pour une certaine valeur de p' peut très bien disparaître des sous-ensemble sélectionnées pour une valeur supérieure de p' si d'autres électrodes sont jugées plus utiles pour la tâche de classification. Dans la méthode de forward CSP proposée, lorsqu'une variable fait partie de la sélection, elle est conservée jusqu'au terme de la méthode. Cependant, il est possible qu'une variable devienne moins indispensable quelques itérations plus tard, du fait par exemple de l'inclusion entre temps de variables corrélées avec elle. Un constat similaire peut être dressé pour la méthode de backward CSP. L'élimination d'une variable à une itération donnée peut se révéler préjudiciable quelques itérations plus tard lorsque la composition du sous-ensemble de variables sélectionnées aura évolué. Des algorithmes existent afin de palier à ce problème en incluant une étape d'élimination de variables après chaque ajout dans la procédure forward (Sequential forward floating) et inversement pour la procédure backward (Sequential backward floating) [90].

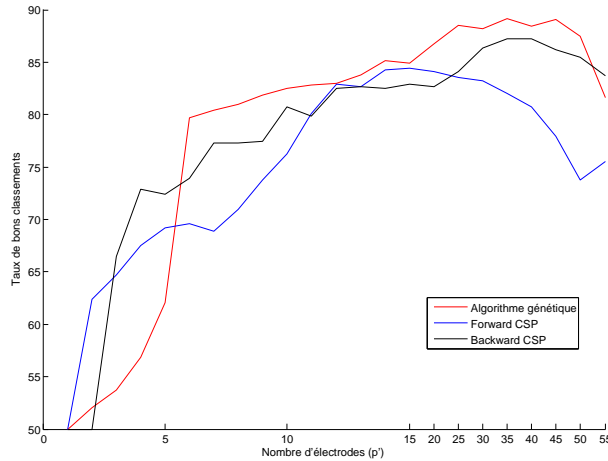


Figure V.15 – Représentation des TBC obtenus en utilisant le forward CSP (courbe bleue), le backward CSP (courbe noire) et l'algorithme génétique (courbe rouge) en fonction du nombre d'électrodes sélectionnées.

Une approche similaire à celle réalisée en Section III.2 pour le choix de p' à partir des résultats de l'algorithme génétique peut être réalisée ici pour les deux méthodes de sélection séquentielle. Les TBC obtenus pour chaque itération du forward CSP et du backward CSP sont respectivement représentés en Figures V.16(a) et V.16(b). Contrairement à l'algorithme génétique, les deux approches présentées ici sont déterministes. Ainsi, le choix du meilleur compromis pour la valeur de p' ne peut s'appuyer sur un test statistique. Il convient donc de rechercher sur ces courbes la plus petite valeur de p' avant que les TBC ne stagnent. Sur la Figure

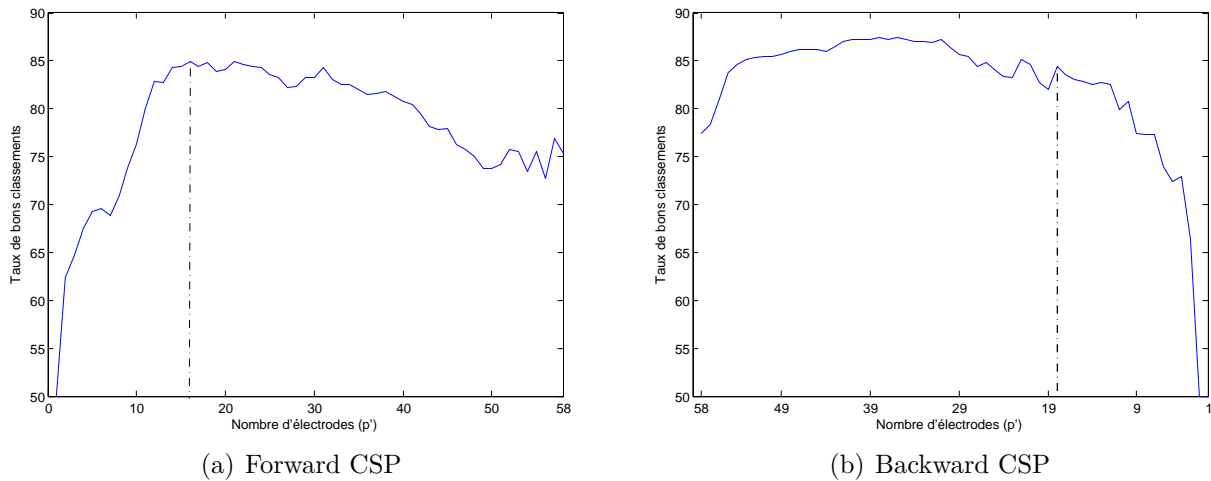


Figure V.16 – Taux de bons classements obtenus en fonction du nombre de variables sélectionnées par forward CSP (a) et par backward CSP (b). Les TBC sont estimés par LOO sur les sujets de la première campagne. Les droites en traits pointillé correspondent à la valeur de p' choisie pour chacune des deux approches.

V.16(a), il semblerait que cette valeur soit $p' = 16$ (ligne en traits pointillé), correspondant au TBC maximum (84.97%, à égalité avec $p' = 21$) obtenu à l'aide du forward CSP. Le sous-ensemble des 16 électrodes correspondantes va donc être conservé et va représenter la sélection d'électrodes obtenue grâce à cette méthode. Sur la Figure V.16(b), choisir le p' qui permet d'obtenir le TBC maximal reviendrait à considérer un sous-ensemble de 38 électrodes (TBC égal à 87.41%). Cette valeur de p' étant trop élevée, la plus petite valeur de p' que l'on puisse trouver avant que les valeurs des TBC n'amorce une décroissance va être choisie. Cette valeur semble atteinte en $p' = 18$ (ligne en traits pointillé) correspondant à un TBC de 84.44%. De la même façon que pour le forward CSP, les 18 électrodes correspondantes vont représenter le sous-ensemble d'électrodes obtenu par backward CSP. Les sous-ensembles d'électrodes obtenus par ces deux approches sont évalués sur l'échantillon test formé par les signaux EEG des sujets de la deuxième campagne. Les filtres CSP et la règle de décision sont obtenus en utilisant les 13 sujets de la première campagne. Le sous-ensemble de 16 électrodes obtenu par forward CSP permet d'obtenir un TBC de 71.21% tandis qu'un TBC de 72.34% est atteint pour celui de 18 électrodes formé par backward.

Les listes des variables ajoutées (pour le forward CSP) ou retirées (pour le backward CSP) à chaque itération ainsi que les TBC correspondants sont fournies respectivement en Figures V.17 et V.18.

Les représentations de la localisation des électrodes sélectionnées par forward CSP et backward CSP sont respectivement proposées en Figures V.19(a) et V.19(b). Sur ces Figures,

V.5 Sélection de variables par méthode séquentielle

FP1	F6	CP5	F1	F7	FC4	FPz	T3	FC5	TP7	AF3
50.0000	62.4126	64.6853	67.4825	69.2308	69.5804	68.8811	70.9790	73.7762	76.2238	80.0699
O2	FC3	CP2	FC6	Oz	FP2	Fz	P06	T4	T6	P2
82.8671	82.6923	84.2657	84.4406	84.9650	84.4406	84.7902	83.9161	84.0909	84.9650	84.6154
F2	C3	TP8	FCz	P4	T5	O1	P07	Pz	F4	P03
84.4406	84.2657	83.5664	83.2168	82.1678	82.3427	83.2168	83.2168	84.2657	83.0420	82.5175
P1	CPz	P5	P3	C4	Cz	CP1	CP3	AF4	P6	P08
82.5175	81.9930	81.4685	81.6434	81.8182	81.2937	80.7692	80.4196	79.5455	78.1469	77.7972
F3	CP4	FC1	C1	P05	P0z	C6	FC2	C2	C5	CP6
77.9720	76.2238	75.6993	75.0000	73.7762	73.7762	74.1259	75.6993	75.5245	73.4266	75.5245
F8	F5	P04								
72.7273	76.9231	75.3497								

Figure V.17 – Liste d'ajout des électrodes lors des itérations du forward CSP avec les taux de bons classements correspondants.

F6	P03	FC3	O2	P3	C3	F4	CP2	FP2	FCz	T6
77.4476	78.3217	80.9441	83.7413	84.6154	85.1399	85.3147	85.4895	85.4895	85.6643	86.0140
Oz	CP1	P05	FC1	CP5	FC4	FC2	Pz	P2	F2	FP1
86.1888	86.1888	86.1888	86.0140	86.5385	87.0629	87.2378	87.2378	87.2378	87.4126	87.2378
Cz	AF4	C2	P5	CPz	O1	FPz	P1	C1	TP8	P4
87.4126	87.2378	87.0629	87.0629	86.8881	87.2378	86.3636	85.6643	85.4895	84.4406	84.7902
F1	Fz	CP3	F3	AF3	P07	T5	C4	CP4	F8	P08
84.0909	83.3916	83.2168	85.1399	84.6154	82.6923	81.9930	84.4406	83.5664	83.0420	82.8671
C6	FC6	T4	P0z	CP6	P6	P04	P06	F5	FC5	TP7
82.5175	82.6923	82.5175	79.8951	80.7692	77.4476	77.2727	77.2727	73.9510	72.3776	72.9021
T3	F7	C5								
66.4336	50	50								

Figure V.18 – Liste de retrait des électrodes lors des itérations du backward CSP avec les taux de bons classements correspondants.

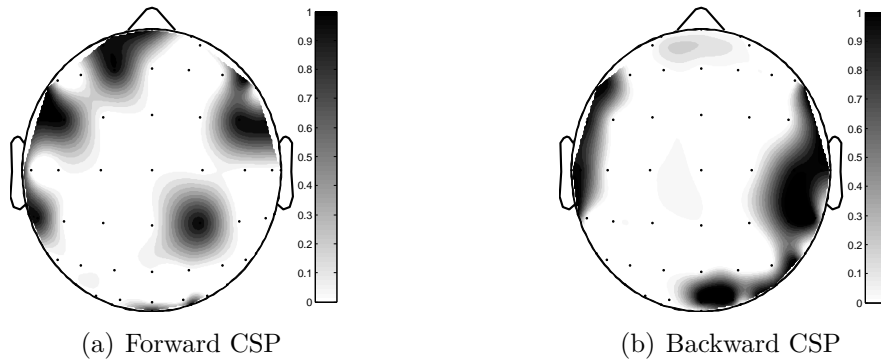


Figure V.19 – Représentation de la localisation des électrodes sélectionnées par forward CSP (a) et par backward CSP (b). Les électrodes sélectionnées apparaissent en noir.

les électrodes sélectionnées apparaissent en noir. Même si les zones cérébrales correspondant aux électrodes sélectionnées par les deux approches semblent assez similaires, il ne semblerait pas que les deux sous-ensembles aient beaucoup d'électrodes en commun. Une fois encore, les électrodes sélectionnées par les deux approches se situent sur les contours du scalp.

Un diagramme de Venn est proposé en Figure V.20 afin de comparer les sous-ensembles de variables sélectionnés par les trois approches. Ce diagramme permet d'avoir un aperçu des similitudes entre les différents sous-ensembles d'électrodes formés. Il apparaît que 4 électrodes sont communes aux trois sélections : F7, FC5, T3 et TP7.

Pour conclure sur les approches séquentielles, les temps de calcul peuvent être discutés. L'approche forward CSP a réalisé les 58 itérations en 4 heures tandis que l'approche backward CSP a nécessité 8 heures. La configuration de l'ordinateur utilisé lors des lancements est la suivante : Linux avec 8 * 2.8 GHz Intel Xeon processeurs X5560 et 32 GB 1333 MHz DDR3 ECC RAM. L'importante différence observée au niveau des temps de calcul peut être expliquée en analysant les constructions des méthodes. En effet, le nombre de sous-ensembles d'électrodes testés décroît avec les itérations. Lors de la première itération, p modèles sont testés, puis $p - 1$ modèles sont testés au cours de la deuxième et ainsi de suite. Le passage dans l'évaluation dont le coût computationnel est important correspond à l'inversion de la matrice de variance covariance, de taille $p' \times p'$ pour le calcul des filtres spatiaux dans le CSP. Ainsi, plus p' est grand, plus cette étape est longue. Les itérations durant lesquels le forward CSP teste beaucoup de sous-ensembles correspondent à de faibles valeurs de p' . Ces étapes sont donc effectuées rapidement car l'inversion des matrices de variance-covariance est rapide. Au contraire, les premières itérations du backward CSP consistent à tester un grand nombre de sous-ensembles d'électrodes où p' est proche de p augmentant ainsi le coût computationnel du backward CSP par rapport

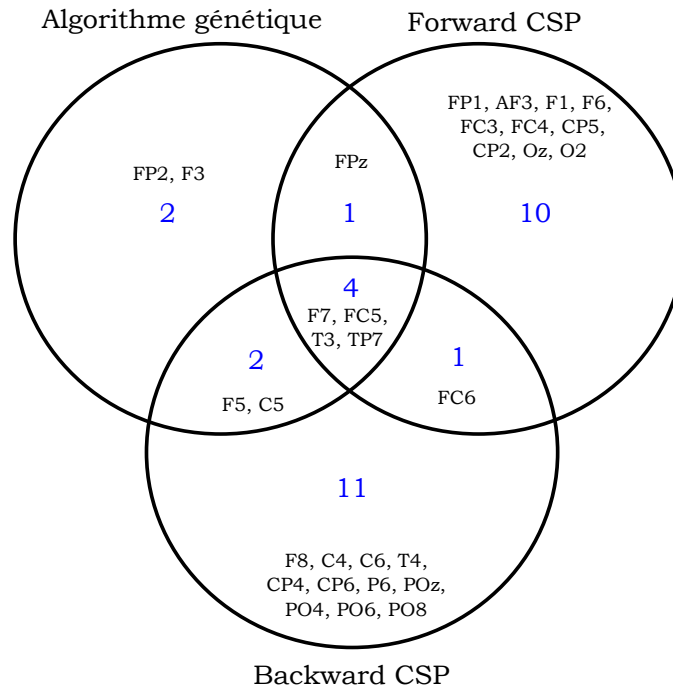


Figure V.20 – Diagramme de Venn présentant les similitudes entre les sous-ensembles de variables sélectionnées par algorithme génétique (9 électrodes), forward (16) et backward CSP (18).

à celui du forward CSP.

Le principal avantage des méthodes de recherche séquentielle par rapport à la recherche par algorithme génétique proposé repose sur le coût computationnel plus faible de ces approches. Cependant, pour de plus grandes valeurs de p , ces approches peuvent devenir inutilisables en l'état. Pour régler ce problème, certains travaux proposent d'ajouter (ou d'éliminer lorsqu'il s'agit de backward) les variables par groupes dans la procédure de forward au lieu de les ajouter une par une. Par exemple, [76] propose une méthode dérivée du backward où les variables sont éliminées par trio à chaque itération. Cette approche permet de limiter le nombre d'itérations et donc le temps de calcul.

Nous avons choisi d'itérer les méthodes de recherches séquentielles p fois afin de disposer des TBC pour toutes les valeurs de p' entre 1 et p . En analysant les résultats, le meilleur compromis entre le nombre d'électrodes conservées (p') et la qualité de la prédiction a été sélectionné. Cependant, il serait possible de stopper les méthodes de recherches séquentielles automatiquement lorsqu'un nombre préfixé d'électrodes est atteint ou lorsque le TBC dépasse un seuil pré-défini

par exemple.

Il apparaît que la recherche par algorithme génétique est à privilégier si le coût computationnel de l'approche n'est pas trop pénalisant. Notons également que celui-ci peut être abaissé en changeant les paramètres de l'algorithme (notamment le nombre d'enfants et de parents) ou en privilégiant une estimation du TBC par validation croisée par exemple. Cependant, cela risque de détériorer la qualité de l'estimation du TBC et donc indirectement la qualité de la solution (cf Chapitre III).

6 Prédiction de l'état de vigilance d'un nouvel individu

Étant donné les résultats, le meilleur compromis est obtenu en conservant 9 électrodes (sélection obtenue par algorithme génétique et représentée en Figure V.14). Ainsi, les filtres CSP et la règle de décision LDA sont appris sur les 19 sujets de l'étude en ne conservant que les 9 électrodes sélectionnées.

Dans une future application de ce travail, le signal EEG d'un nouvel individu pourra être recueilli en n'utilisant que les 9 électrodes sélectionnées. Ces signaux seront filtrés à l'aide d'un filtre passe bande entre 7 et 30 Hz. La méthode présentée dans ce chapitre nécessite au moins un signal de 8 secondes de longueur. Si le signal EEG est mesuré pendant plus de 16 secondes, ce signal sera découpé en essais de 8 secondes. Si le signal EEG est mesuré entre 8 et 16 secondes, un unique essai sera considéré. Les filtres CSP et la règle de décision LDA formés à l'aide des 19 sujets de l'étude seront utilisés afin de prédire la classe à laquelle appartiennent chacun des essais du nouvel individu. Si plusieurs essais sont disponibles, un vote majoritaire sur les prédictions des essais permettra de déterminer l'état de vigilance du nouvel individu.

Il est important de remarquer que même si la recherche du meilleur sous-ensemble de 9 électrodes a un coût computationnel élevé, le modèle obtenu (filtre CSP et règle de décision) permet d'obtenir une prédiction rapide de l'état de vigilance d'un nouvel individu. En effet, la prédiction d'un essai d'un nouvel individu nécessite seulement $4.21 * 10^{-4}$ seconde (configuration : Windows avec $8 * 2.8$ GHz Intel Core i7 2920XM et 16 GB 2.5 GHz RAM). Pour cette raison, le modèle obtenu permet d'envisager des applications en temps réel.

7 Conclusion

Dans ce chapitre, une approche pour prédire l'état de vigilance d'humains à partir de leur activité cérébrale a été proposée. Dans celle-ci, un algorithme génétique a été utilisée afin de rechercher la meilleure combinaison de p' électrodes pour construire le modèle de prédiction de l'état de vigilance. Sur chaque combinaison de p' électrodes, la méthode CSP couplée à une transformation log-variance a permis d'extraire les caractéristiques des signaux EEG. Une analyse discriminante linéaire a ensuite été utilisée pour construire une règle de décision (modèle) et d'obtenir un taux de bons classements. Grâce à cet algorithme génétique, un sous-ensemble de 9 électrodes a été conservé, permettant l'obtention d'un modèle fiable (moyenne de taux de bons classements égal à 71.59% sur un ensemble de test externe à l'algorithme génétique) pour prédire l'état de vigilance d'un nouveau sujet. Le seul inconvénient de cette approche est que la recherche par algorithme génétique décrite dans ce chapitre est coûteuse en temps de calcul. Ainsi, deux méthodes de recherche séquentielles, le forward CSP ou le backward CSP, ont été proposées. La mesure de la qualité des solutions obtenues est fondée sur le TBC obtenu en utilisant la même méthode d'évaluation que celle définie dans l'algorithme génétique (CSP + log-var + LDA). Ces méthodes ont permis d'obtenir des solutions proches de celles proposées par l'algorithme génétique tout en réduisant considérablement le temps de calcul. Cependant, ces approches séquentielles, de par leur construction, ne testent qu'un faible nombre de combinaisons possibles (comparé à l'algorithme génétique) et peuvent ne déboucher que sur un optimum local. Des versions plus sophistiquées des algorithmes Forward et Backward (sequential forward floating et sequential backward floating) peuvent être envisagées afin d'améliorer le parcours de l'espace de recherche [90]. De plus, des développements restent à réaliser afin de définir des critères d'arrêt pour les approches séquentielles. Cela permettra de s'affranchir de l'intervention de l'utilisateur pour sélectionner la valeur de p' et de réduire le nombre d'itérations réalisées.

Notons également que même si la recherche d'une combinaison d'électrode par algorithme génétique est coûteuse en temps de calcul, la solution obtenue (règle de décision LDA et filtres spatiaux CSP) permet de fournir une prédiction rapide de l'état de vigilance d'un nouvel individu. En effet, $4.21 * 10^{-4}$ secondes sont nécessaires pour prédire l'état de vigilance de ce nouvel individu (MATLAB sur un ordinateur Windows avec $8 * 2.8$ GHz Intel Core i7 2920XM et 16 GB 2.5 GHz RAM). Ainsi, cette règle de décision peut être utilisée dans des applications en temps réel afin de détecter automatiquement les changements d'état de vigilance pour les personnes qui doivent rester vigilants (pilotes ou chirurgien par exemple). Les taux de bons

Chapitre V. Algorithme génétique pour la sélection de variables dans les filtres CSP.

classements obtenus dans cette étude montrent que la méthode CSP peut être utilisée pour extraire des caractéristiques à partir de signaux EEG d'individus différents. La méthodologie détaillée dans ce chapitre peut être facilement adaptée à d'autres données en trois dimensions présentant une structure similaire à celle des données issues de l'EEG (3 dimensions : temps, sujets et variables) comme par exemple les signaux de l'électrocardiogramme (ECG) pour prédire une variable binaire.

En outre, plusieurs méthodes de classification peuvent être utilisées pour remplacer la méthode LDA. Enfin, une méthode de CSP parcimonieuse a été développée pour effectuer une sélection de variables dans le calcul CSP (sans utiliser un algorithme génétique). Cette méthode représente un moyen plus rapide pour effectuer la sélection des variables lorsque la méthode CSP est utilisée pour extraire des caractéristiques. Cette approche fait l'objet du Chapitre VI.

Enfin, une comparaison entre les résultats obtenus dans ce Chapitre et ceux obtenus en Chapitre IV peut être réalisée. Dans le Chapitre précédent, une méthode d'extraction de caractéristiques a été proposée reposant sur l'utilisation d'une unique électrode dont la position a été optimisée en utilisant un algorithme génétique. Cette approche, bien que plus restrictive au niveau du nombre d'électrodes choisies (utilisation d'une unique électrode), a permis d'obtenir des taux de bons classements plus élevés ($86.68 \pm 1.87\%$) que celle proposée dans le Chapitre V. Bien qu'il soit délicat de comparer les taux de bons classements obtenus dans les deux approches car ces derniers ont été calculés sur des échantillons différents, il apparaît que l'hypothèse avancée en conclusion du Chapitre IV soit vérifiée. En effet, le fait de ne pas avoir pu laisser un échantillon de signaux extérieur à l'algorithme génétique semble avoir conduit à l'obtention de taux de bons classements trop optimistes dans le chapitre IV. Dans le Chapitre V, grâce à l'ajout des signaux obtenus durant la seconde campagne d'acquisition, le meilleur génome trouvé par l'algorithme génétique a pu être testé sur un échantillon de signaux qui n'avait pas encore été utilisé. Cette stratégie doit donc être privilégiée.

Chapitre VI

Le CSP parcimonieux

Sommaire

1	Introduction	160
2	Le CSP : une double ACP avec métrique	161
2.1	ACP de \mathbf{X}_1 avec métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p	161
2.2	ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p	162
3	L'ACP parcimonieuse	162
3.1	L'approche de Zou, Hastie et Tibschirani (2006)	163
3.1.1	Le critère pénalisé	163
3.1.2	La stratégie d'optimisation	164
3.1.3	L'algorithme et son implémentation	166
3.2	Grille matricielle de paramètres de régularisation	167
4	Le CSP parcimonieux	171
4.1	Le CSP : une double ACP avec métrique identité	171
4.2	L'algorithme de CSP parcimonieux	172
4.3	Grille matricielle de paramètres de régularisation	174
4.4	Un exemple	174
5	Applications	178
5.1	Comparaison avec les résultats de l'algorithme génétique proposé en Chapitre V	178
5.2	Evaluation du sCSP sur 35 découpages différents	181
6	Conclusion	184

1 Introduction

Dans le chapitre précédent, plusieurs méthodes ont été proposées afin de rechercher la meilleure combinaison de p' électrodes, au sens du taux de bons classements obtenu, pour construire le modèle de prédiction de l'état de vigilance. Sur chaque combinaison de p' électrodes, la méthode CSP couplée à une transformation log-variance a permis d'extraire les caractéristiques des signaux EEG puis une analyse discriminante linéaire a été utilisée pour construire une règle de décision. Cette règle de décision a été appliquée sur des signaux provenant d'un échantillon test afin d'obtenir un taux de bons classements. Plusieurs stratégies de parcours de l'espace de recherche (algorithme génétique, méthode de type Forward ou Backward) ont été proposées. L'algorithme génétique proposé en Chapitre V semble donner les meilleurs résultats mais présente un coût computationnel important. Les approches séquentielles peuvent fournir une alternative intéressante mais un critère d'arrêt est nécessaire pour limiter, encore davantage, le coût computationnel. L'approche présentée dans ce chapitre propose une alternative qui permet de sélectionner directement des variables en introduisant une pénalité de type l_1 dans le critère à optimiser de la méthode CSP.

Dans ce qui suit, \mathbf{X}_1 représente la matrice, de dimensions $T_1 \times p$, des signaux EEG concaténés et mesurés sur p électrodes dans la condition 1. De la même façon, la matrice \mathbf{X}_2 , de dimensions $T_2 \times p$, correspond aux signaux EEG concaténés mesurés dans la condition 2. On considère que les matrices \mathbf{X}_1 et \mathbf{X}_2 sont centrées, filtrées à l'aide d'un filtre passe-bande et normalisées en utilisant l'une des stratégies de normalisation présentée en Section 3.2.2 du Chapitre V.

La méthode CSP a été détaillée en Section 1.2 du Chapitre III. Elle consiste à trouver k paires de filtres $(\mathbf{w}_j, \mathbf{v}_j)_{j=1, \dots, k}$ telles que :

- la variance des signaux synthétiques $\mathbf{X}_1 \mathbf{w}_j$ soit maximale tandis que celle des signaux synthétiques $\mathbf{X}_2 \mathbf{w}_j$ est minimale. Cela revient au problème d'optimisation :

$$\max_{\mathbf{w} \in \mathbb{R}^p} J(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \quad (\text{VI.1})$$

- la variance des signaux synthétiques $\mathbf{X}_2 \mathbf{v}_j$ soit maximale tandis que celle des signaux synthétiques $\mathbf{X}_1 \mathbf{v}_j$ est minimale. Le problème d'optimisation correspondant s'écrit :

$$\max_{\mathbf{v} \in \mathbb{R}^p} H(\mathbf{v}) = \max_{\mathbf{v} \in \mathbb{R}^p} \frac{\mathbf{v}^T \mathbf{C}_2 \mathbf{v}}{\mathbf{v}^T \mathbf{C}_1 \mathbf{v}} \quad (\text{VI.2})$$

- les signaux synthétiques soient non corrélés.

On a vu que ces problèmes d'optimisation se réduisaient à rechercher les premiers et les derniers vecteurs propres de $\mathbf{C}_2^{-1}\mathbf{C}_1$ ou encore les k premiers de $\mathbf{C}_1^{-1}\mathbf{C}_2$ et les k premiers de $\mathbf{C}_2^{-1}\mathbf{C}_1$.

Dans ce chapitre, une méthode de CSP parcimonieux va être proposée. L'idée est d'écrire le CSP comme une double Analyse en Composantes Principales (ACP) puis d'appliquer une méthode d'ACP parcimonieuse pour effectuer une sélection de variables.

Le reste de ce chapitre se présente comme suit : Nous verrons d'abord en Section 2 comment la méthode CSP peut s'écrire comme une double ACP avec métrique. En Section 3, nous présenterons la méthode d'ACP parcimonieuse de Zou et al. [91] et nous proposerons ensuite une stratégie de construction d'une grille matricielle de paramètres de régularisation permettant de contrôler le degré de parcimonie dans cet algorithme. Nous proposerons un algorithme de CSP parcimonieux défini comme une double ACP parcimonieuse sur données recodées en Section 4. Enfin, la Section 5 présentera une application de l'algorithme de CSP parcimonieux sur nos données et la Section 6 présentera un résumé de ce travail et les conclusions.

2 Le CSP : une double ACP avec métrique

Afin de se rapprocher du problème d'optimisation de l'ACP, celui du CSP va être réécrit. Dans les équations (VI.1) et (VI.2), en remarquant que $\forall a \in \mathbb{R}, J(a\mathbf{w}) = J(\mathbf{w})$ et $H(a\mathbf{v}) = H(\mathbf{v})$, le double problème d'optimisation se réécrit :

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}^p} \mathbf{w}^T \mathbf{C}_1 \mathbf{w} & \quad \text{(VI.3)} & \quad \text{et} & \quad \max_{\mathbf{v} \in \mathbb{R}^p} \mathbf{v}^T \mathbf{C}_2 \mathbf{v} & \quad \text{(VI.4)} \\ \text{s. c. } \mathbf{w}^T \mathbf{C}_2 \mathbf{w} = 1 & & & \quad \text{s. c. } \mathbf{v}^T \mathbf{C}_1 \mathbf{v} = 1 & \end{aligned}$$

Dans les prochains paragraphes, il va être montré que (VI.3) est le problème d'optimisation de l'ACP de \mathbf{X}_1 avec la métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p et que (VI.4) est le problème d'optimisation de l'ACP de \mathbf{X}_2 avec la métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p .

2.1 ACP de \mathbf{X}_1 avec métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p

L'ACP de \mathbf{X}_1 avec la métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p consiste à faire une projection \mathbf{C}_2^{-1} orthogonale des T_1 points de \mathbb{R}^p , décrits dans \mathbf{X}_1 , sur k axes de vecteurs directeurs $\tilde{\mathbf{w}}_j$, \mathbf{C}_2^{-1} normés à 1 ($\tilde{\mathbf{w}}_j^T \mathbf{C}_2^{-1} \tilde{\mathbf{w}}_j = 1$). Ces vecteurs directeurs $\tilde{\mathbf{w}}_j$ sont définis pour que la variance du nuage projeté

sur ces axes soit maximale. Le vecteur des coordonnées des points projetés sur l'axe de vecteur directeur $\tilde{\mathbf{w}}_j$ est donc $\mathbf{f}_j = \mathbf{X}_1 \mathbf{C}_2^{-1} \tilde{\mathbf{w}}_j$ (vecteur des scores sur l'axe j). En posant $\mathbf{w}_j = \mathbf{C}_2^{-1} \tilde{\mathbf{w}}_j$, on obtient $\mathbf{f}_j = \mathbf{X}_1 \mathbf{w}_j$ avec $\mathbf{w}_j^T \mathbf{C}_2 \mathbf{w}_j = 1$ et $\text{var}(\mathbf{f}_j) = \mathbf{w}_j^T \mathbf{C}_1 \mathbf{w}_j$. Le problème d'optimisation VI.3 est donc équivalent à :

$$\begin{aligned} \max_{\tilde{\mathbf{w}} \in \mathbb{R}^p} \text{var}(\mathbf{X}_1 \mathbf{C}_2^{-1} \tilde{\mathbf{w}}) \\ \text{s. c. } \tilde{\mathbf{w}}^T \mathbf{C}_2^{-1} \tilde{\mathbf{w}} = 1 \end{aligned}$$

L'ACP de \mathbf{X}_1 avec métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p fournit donc les k premiers filtres CSP $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$.

2.2 ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p

De la même façon, l'ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p consiste à rechercher les vecteurs directeurs $\tilde{\mathbf{v}}_j$ tels que la variance du nuage projeté sur ces axes soit maximale. Le vecteur des coordonnées des points projetés sur l'axe de vecteur directeur $\tilde{\mathbf{v}}_j$ est donc $\mathbf{g}_j = \mathbf{X}_2 \mathbf{C}_1^{-1} \tilde{\mathbf{v}}_j$ (vecteur des scores sur l'axe j). En posant $\mathbf{v}_j = \mathbf{C}_1^{-1} \tilde{\mathbf{v}}_j$, on obtient que $\mathbf{g}_j = \mathbf{X}_2 \mathbf{v}_j$ avec $\mathbf{v}_j^T \mathbf{C}_1 \mathbf{v}_j = 1$ et $\text{var}(\mathbf{g}_j) = \mathbf{v}_j^T \mathbf{C}_2 \mathbf{v}_j$. Le problème d'optimisation VI.4 est donc équivalent à :

$$\begin{aligned} \max_{\tilde{\mathbf{v}} \in \mathbb{R}^p} \text{var}(\mathbf{X}_2 \mathbf{C}_1^{-1} \tilde{\mathbf{v}}) \\ \text{s. c. } \tilde{\mathbf{v}}^T \mathbf{C}_1^{-1} \tilde{\mathbf{v}} = 1 \end{aligned}$$

L'ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p fournit les k derniers filtres CSP $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$.

3 L'ACP parcimonieuse

Il existe plusieurs algorithmes qui visent à réaliser une ACP parcimonieuse (voir par exemple [91] et [92]). Dans ce travail, nous utilisons l'approche de Zou, Hastie et Tibshirani [91] en considérant le cas où seule une pénalité l_1 est ajoutée au critère à optimiser (et non pas une combinaison d'une pénalité l_1 et l_2 comme proposé dans [91]).

Nous allons donc présenter en Section 3.1 cette approche d'ACP parcimonieuse puis nous proposerons en Section 3.2 une stratégie pour définir une grille matricielle de paramètres de régularisation.

3.1 L'approche de Zou, Hastie et Tibschirani (2006)

3.1.1 Le critère pénalisé

Rappelons d'abord le critère optimisé en ACP. Soit X la matrice des données centrées de dimensions $n \times p$ et soit k le nombre de composantes principales retenues. Le problème d'optimisation s'écrit alors en ACP :

$$\begin{aligned} \max_{\alpha_j \in \mathbb{R}^p} \text{var}(X\alpha_j) \\ \text{s. c. } \alpha_j^T \alpha_j = 1 \end{aligned}$$

Ainsi présentée, l'ACP revient à trouver les vecteurs directeurs α_j , pour $j = 1, \dots, k$, tels que la variance du signal projeté sur ces axes soit maximale. La métrique utilisée ici est la métrique identité. On parlera aussi d'ACP « sans métrique ».

Une autre manière de voir l'ACP consiste à la considérer comme une méthode d'approximation de la matrice X par une matrice \hat{X} de dimensions $n \times p$ de rang k , inférieur à celui de X ($k < p$). Dans ce cas, le problème d'optimisation de l'ACP s'écrit :

$$\min_{F, \alpha} \|X - F\alpha^T\|^2 \quad (\text{VI.5})$$

où $\|A\| = \text{Tr}(AA^T)$ est la norme de Frobenius de la matrice A et où $\text{Tr}(A)$ est la trace de la matrice A . La matrice F est la matrice des scores des n observations sur les k composantes principales et α est la matrice de dimensions $p \times k$ des loadings (poids des variables dans les combinaisons linéaires permettant de calculer les scores). D'après Eckart et Young [93],

$$\hat{X} = U_k D_k V_k^T$$

où $X = UDV^T$ est la décomposition en valeurs singulières de X , où U_k et V_k sont les matrices respectivement composées des k premières colonnes de U et V et où D_k est la matrice diagonale des k premières valeurs singulières de D . On en déduit que $\hat{F} = U_k D_k$ et que $\hat{\alpha} = V_k$. Or, par construction, on a $F = X\alpha$ (projection sur les axes de vecteurs directeurs $\alpha_1, \dots, \alpha_k$ où α_j est la $j^{\text{ème}}$ colonne de α), ce qui permet, en s'inspirant de (VI.5), d'introduire le problème d'optimisation de l'ACP parcimonieuse proposée par [91] :

$$\begin{aligned} \min_{\alpha, \beta} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \\ \text{s. c. } \alpha^T \alpha = I_k \end{aligned} \quad (\text{VI.6})$$

où $\|\cdot\|_1$ est la norme l_1 , où $\beta_j \in \mathbb{R}^p$ est la $j^{\text{ème}}$ colonne de β et où λ^j est le paramètre de régularisation. Ce dernier permet de contrôler le degré de parcimonie de la $j^{\text{ème}}$ composante principale. Ainsi, plus la valeur de λ^j est grande, plus le vecteur β_j sera composé de poids nuls. Le choix du vecteur $\lambda = [\lambda^1, \dots, \lambda^k]$ des k paramètres de régularisation associés aux k composantes sera discuté en Section 3.2. Plus précisément, une stratégie de construction de vecteurs paramètres λ permettant de contrôler le degré de parcimonie sera proposée.

Dans la suite de ce travail, la matrice β sera appelée la matrice des loadings parcimonieuse. En effet, avec β parcimonieuse, la relation $F_k = X\beta$ va permettre d'obtenir des composantes principales définies comme des combinaisons parcimonieuses des colonnes de X . Contrairement aux composantes principales obtenues en ACP classique, les composantes principales parcimonieuses ne seront pas orthogonales car $\beta^T\beta \neq I_k$. En raison de la contrainte ($\alpha^T\alpha = I_k$), la matrice α sera, quant à elle, appelée la matrice des loadings orthogonaux.

3.1.2 La stratégie d'optimisation

Afin de mieux comprendre l'approche itérative d'optimisation du critère pénalisé (VI.6), le résultat suivant va d'abord être démontré :

Proposition 3.1 (Zou, Hastie et Tibshirani dans [91])

Soit X une matrice de dimensions $n \times p$ et soient β et α deux matrices de dimensions $p \times k$. Alors,

$$\|X - X\beta\alpha^T\|^2 = \text{Tr}(X^T X) - 2\text{Tr}(\alpha^T X^T X \beta) + \text{Tr}(\beta^T X^T X \beta) \quad (\text{VI.7})$$

$$= \text{Tr}(X^T X) - 2 \sum_{j=1}^k \alpha_j^T X^T X \beta_j + \sum_{j=1}^k \beta_j^T X^T X \beta_j \quad (\text{VI.8})$$

Preuve :

Dans un premier temps, on va prouver la relation VI.7 en partant du terme de gauche.

$$\begin{aligned} \|X - X\beta\alpha^T\|^2 &= \text{Tr} \left[(X^T - \alpha\beta^T X^T)(X - X\beta\alpha^T) \right] \\ &= \text{Tr} \left[X^T X - \alpha\beta^T X^T X - X^T X \beta\alpha^T + \alpha\beta^T X^T X \beta\alpha^T \right] \end{aligned}$$

Comme $\text{Tr}(AB) = \text{Tr}(BA)$, il vient que :

- $\text{Tr}(\alpha\beta^T X^T X) = \text{Tr}(\beta^T X^T X \alpha)$
- $\text{Tr}(X^T X \beta\alpha^T) = \text{Tr}(\alpha^T X^T X \beta)$

$$\bullet \operatorname{Tr}(\alpha\beta^T X^T X \beta \alpha^T) = \operatorname{Tr}(\beta^T X^T X \beta \alpha^T \alpha) = \operatorname{Tr}(\beta^T X^T X \beta)$$

De plus, comme $\operatorname{Tr}(A^T) = \operatorname{Tr}(A)$, il vient que $\operatorname{Tr}(\beta^T X^T X \alpha) = \operatorname{Tr}(\alpha^T X^T X \beta)$.

Ainsi, $\|X - X\beta\alpha^T\|^2 = \operatorname{Tr}(X^T X) - 2\operatorname{Tr}(\alpha^T X^T X \beta) + \operatorname{Tr}(\beta^T X^T X \beta)$, ce qui démontre bien la relation (VI.7).

Dans un second temps, on va prouver la relation VI.8 en partant de la relation VI.7.

Ainsi, $\operatorname{Tr}(\alpha^T X^T X \beta) = \sum_{j=i}^k \alpha_j^T X^T X \beta_j$ et $\operatorname{Tr}(\beta^T X^T X \beta) = \sum_{j=i}^k \beta_j^T X^T X \beta_j$.

D'où, $\|X - X\beta\alpha^T\|^2 = \operatorname{Tr}(X^T X) - 2 \sum_{j=i}^k \alpha_j^T X^T X \beta_j + \sum_{j=i}^k \beta_j^T X^T X \beta_j$ ■

Les relations (VI.7) et (VI.8) vont maintenant être utilisées afin de détailler cette approche itérative d'optimisation du critère pénalisé (VI.6). Après une initialisation de β , la première étape consiste à fixer β et rechercher α et la seconde étape consiste à fixer α et rechercher β . Le fait de fixer l'une des deux matrices (α ou β) va permettre de simplifier le problème d'optimisation (VI.6) en le scindant en deux sous-problèmes. Les relations (VI.7) et (VI.8) vont donc permettre d'écrire ces sous-problèmes.

Pour un α fixé : On calcule la $j^{\text{ème}}$ composante principale orthogonale $\mathbf{f}_j = X\alpha_j$ pour $j = 1, \dots, k$. En utilisant (VI.8), le critère s'écrit alors :

$$\begin{aligned} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 &= \operatorname{Tr}(X^T X) + \sum_{j=i}^k [\beta_j^T X^T X \beta_j - 2\mathbf{f}_j^T X \beta_j] + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \\ &= \operatorname{Tr}(X^T X) + \sum_{j=i}^k [\|\mathbf{f}_j - X\beta_j\|_2^2 - \mathbf{f}_j^T \mathbf{f}_j] + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \end{aligned}$$

En remarquant que les termes $\mathbf{f}_j^T \mathbf{f}_j$ et $\operatorname{Tr}(X^T X)$ ne dépendent pas de β , il vient que :

$$\min_{\beta} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \Leftrightarrow \min_{\beta} \sum_{j=1}^k [\|\mathbf{f}_j - X\beta_j\|_2^2 + \lambda^j \|\beta_j\|_1]$$

On se ramène donc à k problèmes de régressions LASSO où l'on cherche, pour $j = 1, \dots, k$,

$$\min_{\beta_j} \|\mathbf{f}_j - X\beta_j\|_2^2 + \lambda^j \|\beta_j\|_1 \quad (\text{VI.9})$$

Ainsi, β_j est obtenu en effectuant les régressions LASSO de $\mathbf{f}_j = X\alpha_j$ sur X . Ce procédé est répété pour $j = 1, \dots, k$ afin d'obtenir les k colonnes de β .

Pour un β fixé : Le problème d'optimisation devient

$$\begin{aligned} \min_{\alpha} \|X - X\beta\alpha^T\|^2 + \sum_{j=1}^k \lambda^j \|\beta_j\|_1 \\ \text{s. c. } \alpha^T \alpha = I_k \end{aligned} \quad (\text{VI.10})$$

Or, en utilisant (VI.7) et en remarquant que les termes $\text{Tr}(X^T X)$, $\text{Tr}(\beta^T X^T X \beta)$ et le terme de pénalité sont indépendants de α , (VI.10) équivaut à :

$$\begin{aligned} \max_{\alpha} \text{Tr}(\alpha^T X^T X \beta) \\ \text{s. c. } \alpha^T \alpha = I_k \end{aligned} \quad (\text{VI.11})$$

où $X\alpha$ et $X\beta$ s'interprètent comme les composantes principales respectivement orthogonales et parcimonieuses. Ainsi, le problème d'optimisation (VI.11) revient à rechercher les composantes principales orthogonales ($X\alpha$) qui ont une covariance maximale avec les composantes principales parcimonieuses ($X\beta$).

On se ramène donc au problème d'optimisation $\hat{\alpha} = \arg \max_{\alpha} \text{Tr}(\alpha^T X^T X \beta)$ s.c. $\alpha^T \alpha = I_k$ et d'après [91] on a $\hat{\alpha} = UV^T$ où les matrices U et V sont obtenues en réalisant la décomposition en valeurs singulières de $X^T X \beta = UDV^T$.

3.1.3 L'algorithme et son implémentation

Présentons maintenant l'algorithme d'optimisation du critère VI.6 d'ACP parcimonieuse. Nous présentons ici sa version sur matrice de Gram qui permet de réduire le coût de calcul lorsque $p < n$ et que n est grand. En effet, la seule connaissance de $X^T X$ (matrice de Gram de X) suffit lors de l'optimisation du critère (VI.6) d'ACP parcimonieuse. Comme montré en Section 3.1.2, la solution de (VI.11) fait uniquement intervenir $X^T X$, et (VI.9) se réécrit :

$$\min_{\beta_j} (\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) + \lambda^j \|\beta_j\|_1.$$

Ainsi, on notera Y la matrice des données centrées et $Y^T Y$ sa matrice de Gram. L'algorithme d'optimisation du critère VI.6 d'ACP parcimonieuse prend comme arguments principaux :

- $Y^T Y$,
- k qui est le nombre de composantes d'ACP parcimonieuse que l'on souhaite obtenir,

– $(\lambda^1, \dots, \lambda^k)$ qui est le vecteur des paramètres de régularisation sur les k composantes.

Algorithme 3 L'ACP parcimonieuse

Initialisation

$X = (Y^T Y)^{\frac{1}{2}}$ où Y est la matrice des données centrées. De cette façon, $X^T X = Y^T Y$.

$\alpha = [\alpha_1, \dots, \alpha_k]$, matrice des k premiers vecteurs propres de $X^T X$ (ou encore k premières colonnes de V dans la SVD de $X = UDV^T$).

pour $j = 1, \dots, k$ **faire**

β_j = coefficients de la régression LASSO de $\mathbf{f}_j = X\alpha_j$ sur X avec le paramètre de régularisation λ^j

$$\beta_j = \frac{\beta_j}{(\beta_j^T \beta_j)^{\frac{1}{2}}}$$

fin du pour

$\beta = [\beta_1, \dots, \beta_k]$

$nb.iter = 0$

$\beta_{iter-1} = (0)_{p \times k}$, matrice composée de 0.

tant que $nb.iter < max.iter$ et $\|\beta - \beta_{iter-1}\|_{\infty} \geq eps.conv$ **faire**

Calcul de $\alpha = [\alpha_1, \dots, \alpha_k]$

$[U, D, V] = \text{SVD}(X^T X \beta)$

$\alpha = UV^T$

Calcul de $\beta = [\beta_1, \dots, \beta_k]$

pour $j = 1, \dots, k$ **faire**

$\beta_{iter-1} = \beta$

β_j = coefficients de la régression LASSO de $\mathbf{f}_j = X\alpha_j$ sur X avec le paramètre de régularisation λ^j

$$\beta_j = \frac{\beta_j}{(\beta_j^T \beta_j)^{\frac{1}{2}}}$$

fin du pour

$nb.iter = nb.iter + 1$

fin du tant que

Retour $\beta = [\beta_1, \dots, \beta_k]$

Cet algorithme est implémenté dans le package *elasticnet* du logiciel *R* dans la fonction *sPCA()*. Il est important de noter que si k change, les loadings parcimonieux β sont modifiés contrairement à l'ACP classique pour laquelle le choix de $k' > k$ dimensions laisse les loadings des k premières dimensions inchangées. Nous proposons dans la prochaine section une stratégie de construction d'une grille de paramètres de régularisation permettant de contrôler le degré de parcimonie dans cet algorithme.

3.2 Grille matricielle de paramètres de régularisation

Une difficulté de cet algorithme d'ACP parcimonieuse est le choix du vecteur $\boldsymbol{\lambda} = (\lambda^1, \dots, \lambda^k)$ de paramètres de régularisation. On doit en effet fournir pour chacune des k composantes une valeur λ^j qui règle son degré de parcimonie. Pour chaque composante j , augmenter la valeur

du paramètre λ^j entraîne une augmentation de son degré de parcimonie. Comment contrôler maintenant globalement le degré de parcimonie sur les k dimensions de l'ACP ?

L'idée est la suivante. On définit pour chaque composante j une grille vectorielle de r_j valeurs $(\lambda_1^j, \dots, \lambda_{r_j}^j)$ permettant de faire varier le degré de parcimonie sur cette composante. Une fois ces grilles obtenues pour chacune des k dimensions, elles sont concaténées en colonnes afin de construire une grille matricielle Λ de dimensions $r \times k$ où $r = \min_j r_j$. Chaque ligne i est un vecteur $\boldsymbol{\lambda}_i = (\lambda_i^1, \dots, \lambda_i^k)$ de paramètres de régularisation pour l'algorithme d'ACP parcimonieuse, et augmenter i (on parlera de pas dans la grille matricielle) permettra d'augmenter le degré de parcimonie des k composantes.

Il reste maintenant à définir comment construire les k colonnes $\boldsymbol{\lambda}^j = (\lambda_1^j, \dots, \lambda_r^j)$ de cette matrice Λ , c'est-à-dire les grilles vectorielles des paramètres de régularisation de chaque composante. Pour cela, on effectuera des régressions LASSO de chacune des k composantes principales de l'ACP sur X .

Il existe plusieurs méthodes pour construire ces grilles de paramètres de régularisation à partir de régressions parcimonieuses de type LASSO. La méthode du *Least Angle Regression* (LARS) [94] est la plus connue d'entre elles. Cet algorithme fournit le chemin complet de régularisation, c'est-à-dire la grille des paramètres de régularisation permettant l'inclusion d'une variable supplémentaire dans le modèle. Une représentation d'un chemin de régularisation, obtenu à l'aide de la fonction `lars()` du package *R* portant le même nom, est proposée en Figure VI.1. Les lignes verticales correspondent à l'introduction de nouvelles variables dans le modèle. Il apparaît que le chemin est exhaustif.

Récemment, Friedman et al. [95] ont proposé un algorithme plus rapide pour lequel le chemin de régularisation est obtenu à partir d'une grille de valeurs calculée automatiquement dans la fonction `glmnet()` du package *R* portant le même nom. Il s'agit d'une grille de $n = 100$ valeurs (par défaut) obtenues en discrétisant un intervalle $[\log(\lambda_{min}), \log(\lambda_{max})]$ en n intervalles égaux. $\lambda_{max} = \max_j cov(\mathbf{x}_j, y)$ est la plus petite valeur de λ qui permette d'obtenir la nullité de tous les coefficients et λ_{min} est obtenu directement à partir de λ_{max} grâce à la relation $\lambda_{min} = \epsilon \lambda_{max}$ où par défaut $\epsilon = 0.01$ si $n < p$ et $\epsilon = 0.001$ sinon. Contrairement à LARS, cet algorithme ne fournit pas le chemin complet de régularisation. Il se peut donc qu'entre deux valeurs successives de la grille, plusieurs variables soient introduites dans le modèle. Cependant, cet algorithme est moins coûteux en terme de temps de calcul et permet une discrétisation régulière des valeurs du paramètre de régularisation qui est plus adaptée à la construction de notre grille matricielle de paramètres en ACP parcimonieuse. C'est donc cet algorithme qui est utilisé dans ce travail. Une représentation d'un chemin de régularisation obtenu à l'aide de la fonction `glmnet()` est

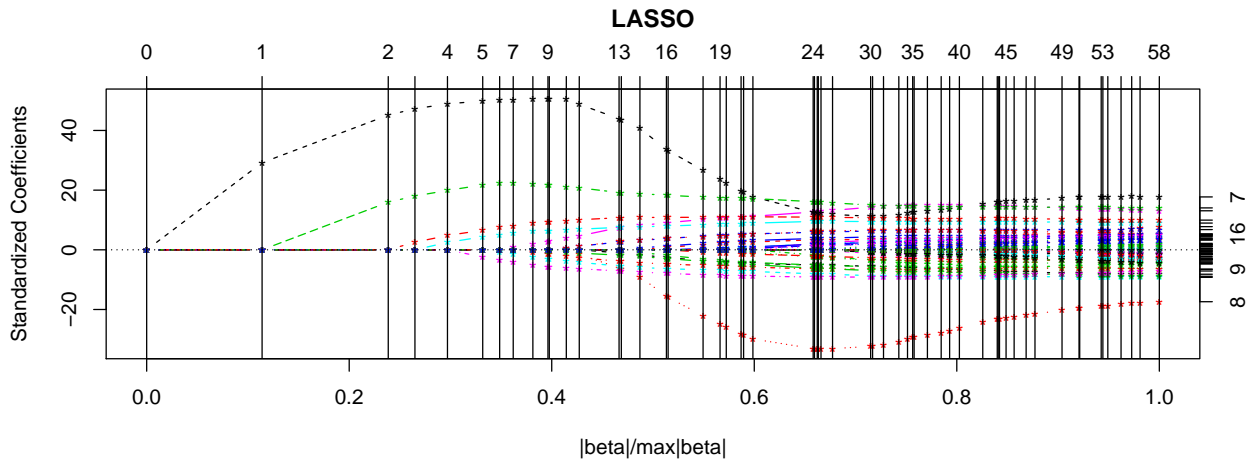


Figure VI.1 – Chemin de régularisation produit par la fonction *lars()* du package *R* portant le même nom. Les lignes verticales marquent l'introduction d'une nouvelle variable dans le modèle. L'axe supérieur de la figure indique le nombre de variables incluses dans le modèle pour la valeur considérée du paramètre de régularisation.

proposée en Figure VI.2.

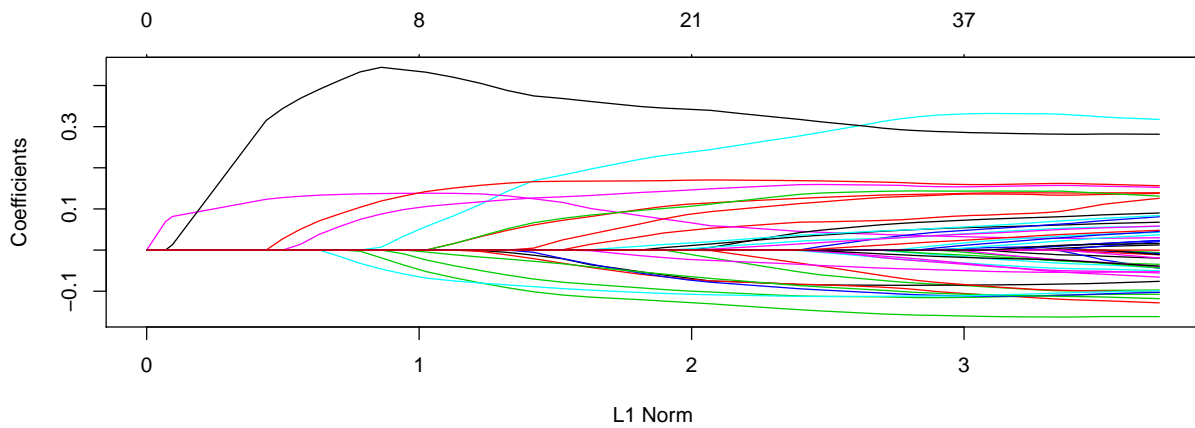


Figure VI.2 – Chemin de régularisation produit par la fonction *glmnet()* du package *R* portant le même nom. L'axe supérieur de la figure indique le nombre de variables incluses dans le modèle pour la valeur considérée du paramètre de régularisation.

En appliquant la stratégie de construction de grille de paramètres de régularisation en régression LASSO de Friedman et al. [95], on construit pour chaque composante principale de l'ACP parcimonieuse, une grille vectorielle de paramètres qui formera les colonnes de la grille matricielle Λ . On peut noter que certains paramètres conduisent à des modèles saturés et sont donc tronqués automatiquement par la fonction *glmnet()*. En pratique, la taille de la grille obtenue

Chapitre VI. Le CSP parcimonieux

nue pour chaque composante est donc plus petite que n . L'algorithme 4 résume la construction de la grille matricielle Λ .

Algorithme 4 Construction de la grille matricielle Λ des paramètres de régularisation

Entrée(s) X : matrice des données centrées,
 k : nombre de composantes d'ACP parcimonieuse à considérer.

Initialisation

$(\mathbf{y}_1, \dots, \mathbf{y}_k) = ACP(X, k)$

Les k composantes principales de l'ACP classique de X sont obtenues.

Calcul de la grille Λ

pour $j = 1, \dots, k$ **faire**

$\lambda_1^j, \dots, \lambda_{r_j}^j = glmnet(X, \mathbf{y}_j)$

Obtention de la grille des paramètres du chemin de régularisation pour la régression LASSO de \mathbf{y}_j sur X .

fin du pour

$r = \min_{j=1, \dots, k} r_j$

$\Lambda = (\lambda_i^j)_{r \times k}$

Retour Λ

Une fois que la grille matricielle de paramètres de régularisation Λ est construite, il convient de définir une stratégie de parcours de cette dernière. Afin de faire varier le degré de parcimonie dans l'algorithme d'ACP parcimonieuse, la stratégie proposée est de parcourir la grille Λ ligne par ligne. Ainsi, pour un pas $i = 1, \dots, r$, un vecteur $\boldsymbol{\lambda}_i = \lambda_i^1, \dots, \lambda_i^k$ va être obtenu et pourra être donné en entrée à l'algorithme d'ACP parcimonieuse. Plus le pas i augmente, plus le degré de parcimonie augmente pour chacune des k composantes (le nombre de poids nuls dans les loadings augmente). À noter que si la fonction *lars()* avait été choisie pour construire la grille Λ , une stratégie de parcours ligne à ligne aurait entraîné l'ajout de k variables à chaque pas. C'est, avec le temps de calcul plus rapide, la principale raison pour laquelle la stratégie de Friedman et al. [95] implémentée dans le package *glmnet* a été utilisée.

4 Le CSP parcimonieux

Il a été montré en Section 2 que le CSP peut s'écrire comme une double ACP avec métrique et l'ACP parcimonieuse de Zou, Friedman et Tibschirani (2006) a été présentée en Section 3. L'idée est donc d'utiliser cet algorithme d'ACP parcimonieuse afin de créer un algorithme de CSP parcimonieux. Cependant, l'algorithme parcimonieux présenté en Section 3 réalise une ACP avec métrique I_p sur \mathbb{R}^p (ACP « classique » ou encore « sans métrique »). Nous allons donc d'abord voir comment écrire l'algorithme CSP comme une double ACP avec métrique identité et non pas avec les inverses des matrices de variance-covariance comme métrique comme présenté section 2.

4.1 Le CSP : une double ACP avec métrique identité

Nous avons vu en Section 2 que les premiers filtres CSP, $\mathbf{w}_1, \dots, \mathbf{w}_k$ sont les loadings de l'ACP de \mathbf{X}_1 avec métrique \mathbf{C}_2^{-1} sur \mathbb{R}^p . En pratique, ces loadings peuvent être obtenus en faisant l'ACP classique (avec métrique identité) d'une matrice de données recodées Z_1 , et en recodant les loadings de cette ACP. On notera $\mathbf{u}_1, \dots, \mathbf{u}_k$ ces loadings qui permettront de calculer $\mathbf{w}_1, \dots, \mathbf{w}_k$. Plus précisément, on réalise l'ACP de $Z_1 = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}}$ avec la métrique I_p sur \mathbb{R}^p (ACP classique). Cela signifie que l'on projette les lignes de Z_1 sur les axes de vecteurs directeurs \mathbf{u}_j (avec $\mathbf{u}_j^T \mathbf{u}_j = 1$) où les \mathbf{u}_j sont les vecteurs propres de la matrice de covariance de Z_1 définie par $\frac{1}{T_1} Z_1^T Z_1$. Les composantes principales de l'ACP de Z_1 sont données par la relation : $\mathbf{f}_j = Z_1 \mathbf{u}_j = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}} \mathbf{u}_j = \mathbf{X}_1 \mathbf{w}_j$ où $\mathbf{w}_j = \mathbf{C}_2^{-\frac{1}{2}} \mathbf{u}_j$. Cette dernière relation permet de retrouver les filtres CSP $\mathbf{w}_1, \dots, \mathbf{w}_k$.

On obtient de la même façon que les derniers filtres CSP v_1, \dots, v_k , résultats de l'ACP de \mathbf{X}_2 avec métrique \mathbf{C}_1^{-1} sur \mathbb{R}^p , peuvent être obtenus à partir de l'ACP de $Z_2 = \mathbf{X}_2 \mathbf{C}_1^{-\frac{1}{2}}$ avec la métrique I_p sur \mathbb{R}^p .

On en déduit que l'algorithme du CSP peut s'écrire comme une double ACP classique (avec métriques identités) des matrices de données EEG recodées Z_1 et Z_2 associée à un recodage des deux matrices de loadings ainsi obtenues. C'est cette version de l'algorithme CSP, donnée dans l'Algorithme 5, qui sera modifiée pour proposer la version parcimonieuse du CSP (Algorithme 6).

Algorithme 5 Le CSP comme une double ACP avec métriques identités

Entrée(s) \mathbf{X}_1 : matrice obtenue par concaténation des essais recueillis dans la condition 1,
 \mathbf{X}_2 : matrice obtenue par concaténation des essais recueillis dans la condition 2,
 k : nombre de filtres CSP à construire.

Sortie(s) Obtention de la matrice des filtres CSP $W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k]$.

Initialisation

pour $c = 1, 2$ **faire**

$$\mathbf{C}_c = \frac{1}{T_c} \mathbf{X}_c^T \mathbf{X}_c$$

$$[U_c, \Sigma_c, V_c] = \text{SVD}(\mathbf{C}_c)$$

$$\mathbf{C}_c^{-\frac{1}{2}} = U_c \Sigma_c^{-\frac{1}{2}} V_c^T$$

fin du pour

$$Z_1 = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}}$$

$$Z_2 = \mathbf{X}_2 \mathbf{C}_1^{-\frac{1}{2}}$$

Obtention des filtres $\mathbf{w}_1, \dots, \mathbf{w}_k$.

$$[\mathbf{u}_1, \dots, \mathbf{u}_k] = \text{ACP}(Z_1, k)$$

L'ACP de Z_1 permet d'obtenir les vecteurs propres \mathbf{u}_j de $\frac{1}{T_1} Z_1^T Z_1$.

pour $j = 1, \dots, k$ **faire**

$$\mathbf{w}_j = \mathbf{C}_2^{-\frac{1}{2}} \mathbf{u}_j$$

fin du pour

Obtention des filtres $\mathbf{v}_1, \dots, \mathbf{v}_k$.

$$[\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_k] = \text{ACP}(Z_2, k)$$

L'ACP de Z_2 permet d'obtenir les vecteurs propres $\tilde{\mathbf{u}}_j$ de $\frac{1}{T_2} Z_2^T Z_2$.

pour $j = 1, \dots, k$ **faire**

$$\mathbf{v}_j = \mathbf{C}_1^{-\frac{1}{2}} \tilde{\mathbf{u}}_j$$

fin du pour

$$W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k]$$

Retour W

4.2 L'algorithme de CSP parcimonieux

Le CSP venant d'être exprimé comme une double ACP (avec métriques identités) sur données recodées, l'algorithme parcimonieux de CSP, proposé en Algorithme 6, consiste à remplacer les deux ACP de l'Algorithme 5 par deux ACP parcimonieuses. Dans l'Algorithme 6, l'Algorithme 3 d'ACP parcimonieuse est utilisé. On appellera respectivement sPCA et sCSP les algorithmes 3 et 6 d'ACP et de CSP parcimonieux (*sparses*).

Algorithme 6 Le CSP parcimonieux

Entrée(s) \mathbf{X}_1 : matrice obtenue par concaténation des essais recueillis dans la condition 1,
 \mathbf{X}_2 : matrice obtenue par concaténation des essais recueillis dans la condition 2,
 k : nombre de filtres CSP à construire,
 $[\lambda^1, \dots, \lambda^k, \phi^1, \dots, \phi^k]$: valeurs des paramètres de régularisation des filtres $\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k$.

Sortie(s) Obtention de la liste des p' électrodes sélectionnées et de la matrice \tilde{W} des filtres CSP obtenue avec ces p' électrodes.

Initialisation

pour $c = 1, 2$ **faire**

$$\mathbf{C}_c = \frac{1}{T_c} \mathbf{X}_c^T \mathbf{X}_c$$

$$[U_c, \Sigma_c, V_c] = \text{SVD}(\mathbf{C}_c)$$

$$\mathbf{C}_c^{-\frac{1}{2}} = U_c \Sigma_c^{-\frac{1}{2}} V_c^T$$

fin du pour

$$Z_1 = \mathbf{X}_1 \mathbf{C}_2^{-\frac{1}{2}}$$

$$Z_2 = \mathbf{X}_2 \mathbf{C}_1^{-\frac{1}{2}}$$

Sélection de variables pour les filtres $\mathbf{w}_1, \dots, \mathbf{w}_k$

$$[\beta_1] = \text{sPCA}(Z_1^T Z_1, k, [\lambda^1, \dots, \lambda^k])$$

β_1 est la matrice de dimensions $p \times k$ des loadings parcimonieux des k premières composantes de l'ACP de Z_1 .

Sélection de variables pour les filtres $\mathbf{v}_1, \dots, \mathbf{v}_k$

$$[\beta_2] = \text{sPCA}(Z_2^T Z_2, k, [\phi^1, \dots, \phi^k])$$

β_2 est la matrice de dimensions $p \times k$ des loadings parcimonieux des k premières composantes de l'ACP de Z_2 .

$$[\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \text{Liste_Var}] = \text{Select_Variables}(\beta_1, \beta_2, \mathbf{X}_1, \mathbf{X}_2)$$

Sélection des p' variables ayant des coefficients non nuls dans β_1 ou β_2 . Ainsi, $\tilde{\mathbf{X}}_1$ et $\tilde{\mathbf{X}}_2$ sont des matrices de dimensions $T_1 \times p'$ et $T_2 \times p'$.

Obtention des filtres CSP

$$[\tilde{W}] = \text{CSP}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, k)$$

Obtention de la matrice des filtres \tilde{W} , de dimensions $p' \times 2k$ en utilisant un CSP uniquement sur les variables sélectionnées par CSP parcimonieux.

Retour \tilde{W} , Liste_Var

On peut noter que l'utilisation du recodage fait passer de \mathbf{X}_1 et de \mathbf{X}_2 à Z_1 et Z_2 . En Section 4.4, il sera montré que cette astuce, en plus de permettre l'utilisation de l'algorithme d'ACP parcimonieuse, a pour effet de décorrélérer les variables de Z_1 et de Z_2 avant de faire l'ACP. Ainsi, les éventuels problèmes de colinéarité des variables sont évités dans l'algorithme 3 d'ACP parcimonieuse, et plus particulièrement dans les étapes des régressions LASSO des composantes sur les colonnes des matrices Z_1 et Z_2 .

On peut également noter que les étapes d'ACP parcimonieuse dans l'algorithme 6 utilisent les matrices de Gram de Z_1 et Z_2 de dimension $p \times p$. Cette étape est alors beaucoup plus rapide

que si l'on avait passé directement les matrices Z_1 et Z_2 en argument à l'algorithme sPCA, ces matrices étant de dimensions $T_1 \times p$ et $T_2 \times p$, avec T_1 et T_2 très grand dans les applications.

4.3 Grille matricielle de paramètres de régularisation

En pratique, dans l'algorithme du sCSP (Algorithme 6), deux vecteurs $\lambda^1, \dots, \lambda^k$ et ϕ^1, \dots, ϕ^k de paramètres de régularisation sont donnés en entrée (un pour chaque ACP parcimonieuse). Ces deux vecteurs de longueur k permettent de fixer le degré de parcimonie des $2k$ composantes CSP. Afin de faire varier le degré de parcimonie des $2k$ composantes CSP, deux grilles matricielles Λ et Φ peuvent être obtenues, en utilisant deux fois l'algorithme 4 proposé en Section 3.2, et concaténées. Il faut simplement donner en entrée de l'algorithme 4, les matrices Z_1 et Z_2 (et non pas \mathbf{X}_1 et \mathbf{X}_2), puisque l'on applique la double ACP parcimonieuse à ces matrices. La stratégie consiste ensuite à parcourir ces deux grilles matricielles ligne à ligne, faisant ainsi diminuer à chaque pas le degré de parcimonie des résultats de l'algorithme sCSP.

4.4 Un exemple

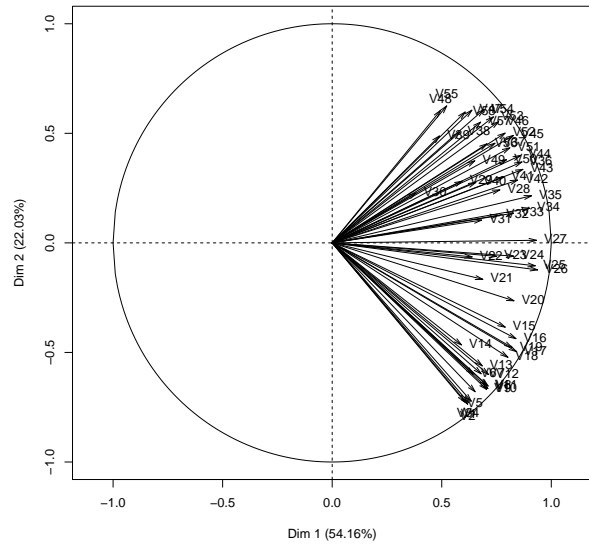
Dans cet exemple, les 8 premières secondes des enregistrements EEG recueillis chez le sujet 58 de l'expérience avant et après la séance de relaxation sont considérés. Ces deux signaux sont filtrés entre 7 et 30 Hz et centrés, permettant ainsi d'obtenir les matrices \mathbf{X}_1 et \mathbf{X}_2 de dimensions 2048×58 où $8 * 256 = 2048$ représente le nombre de points dans un signal échantillonné à 256 Hz et où 58 représente le nombre d'électrodes utilisées pour le recueil de l'EEG.

Afin de représenter les effets du recodage des données, les matrices recodées Z_1 et Z_2 ont été calculées à partir de \mathbf{X}_1 et de \mathbf{X}_2 .

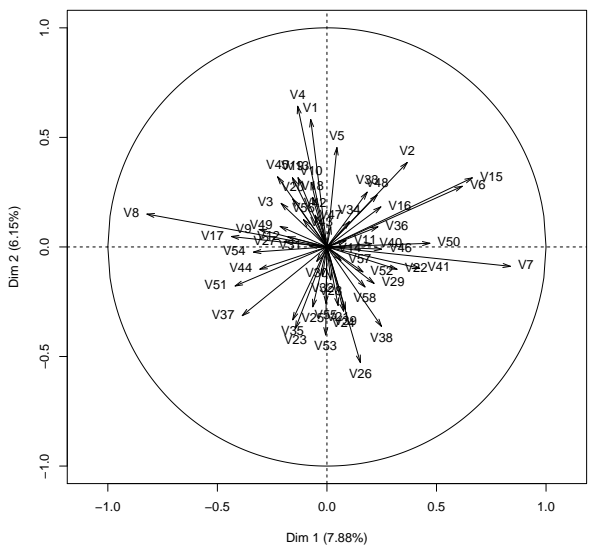
Les Figures VI.3(a), VI.3(b) et VI.3(c) proposent respectivement les cercles de corrélations de l'ACP de X_1 , de Z_1 et de Z_2 . Sur ces figures, l'axe des abscisses et l'axe des ordonnées représentent respectivement les deux premières composantes principales de l'ACP. De plus, chaque variable (chacune des 58 électrodes) est représentée par une flèche. L'angle de la flèche avec les axes des abscisses et des ordonnées donne une représentation de la liaison de ces variables à la première et deuxième composantes principales de l'ACP. Ainsi, plus l'angle est faible, plus la variable est liée à la composante principale. De la même manière, l'angle entre les flèches de deux variables distinctes donnent une représentation de la corrélation entre ces variables. Plus il est faible et plus la corrélation est importante. Sur la Figure VI.3(a), il apparaît que de fortes corrélations existent entre les variables. Les Figures VI.3(b) et VI.3(c) montrent que

le recodage a eu pour effet de réduire considérablement les corrélations entre les variables. Ce point est à l'origine de l'unique utilisation d'une pénalité l_1 plutôt que d'une combinaison de pénalités l_1 et l_2 dans l'étape de régression parcimonieuse de l'algorithme 3, utilisé ensuite dans l'algorithme 6.

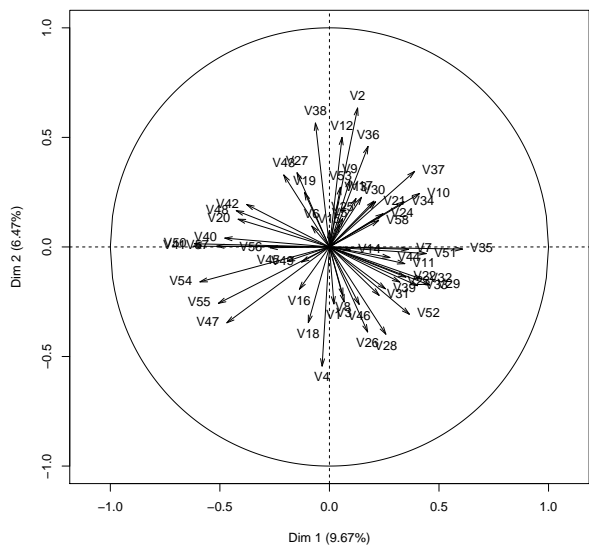
Un CSP classique et un CSP parcimonieux ont ensuite été réalisés. Les deux premières paires de composantes CSP sont représentées en Figures VI.4(a) et VI.4(b). De la même manière, les deux premières paires de composantes CSP parcimonieuses (obtenues avec l'algorithme sCSP) sont représentées en Figures VI.4(c) et VI.4(d). Le troisième pas de la grille de régularisation a été choisi, entraînant la sélection de 7 variables qui sont utilisées pour construire les filtres sCSP. Sur les Figures VI.4(a) et VI.4(b), il apparaît que la première paire de composantes CSP permet bien d'observer que lorsque la variance du signal projeté est maximale pour une classe, elle est minimale pour l'autre. Ce constat est moins marqué sur la deuxième paire de composantes CSP et notamment sur la première composante de cette seconde paire. Bien que le sCSP réduise considérablement le nombre de variables utilisées (7 au lieu de 58), la Figure VI.4(c) montre que la première paire de composantes sCSP permet bien d'obtenir une variance du signal projeté maximale pour une classe lorsqu'elle est minimale pour l'autre classe. Une fois encore, les résultats sont moins bons sur la Figure VI.4(d) où la première composante sCSP de la paire ne permet pas de discriminer correctement les variances des signaux projetés des deux classes. Cette difficulté vient certainement du fait que cela n'était déjà pas le cas pour le CSP classique. En effet, en utilisant moins d'électrodes, il est délicat d'améliorer les résultats obtenus lorsque toutes les électrodes ont été utilisées. Étant donné que le critère extrait par CSP (ou sCSP), et utilisé pour la tâche de classification, repose sur la variance du signal projeté, il est rassurant de constater que le sCSP permet de conserver les bonnes propriétés du CSP classique.



(a) ACP de X_1

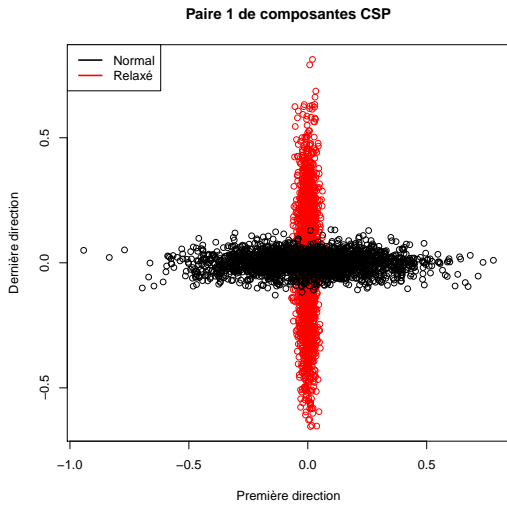


(b) ACP de Z_1

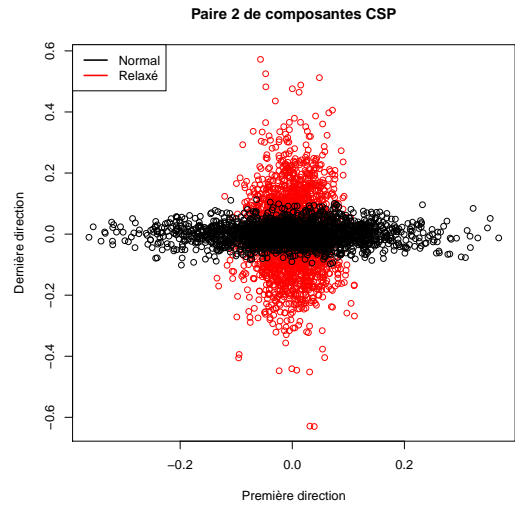


(c) ACP de Z_2

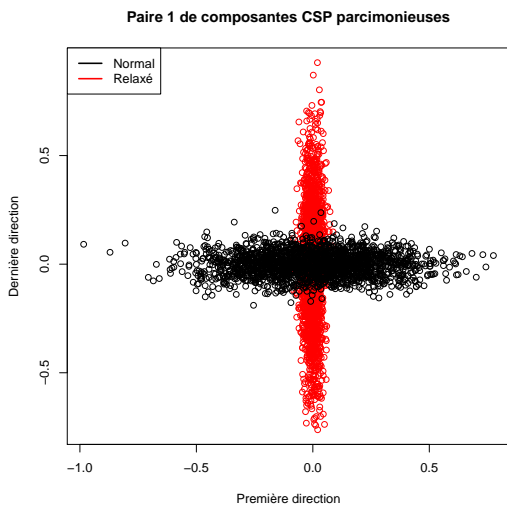
Figure VI.3 – Représentation des cercles de corrélations pour les deux premières composantes principales des ACP respectives de X_1 , Z_1 et Z_2 .



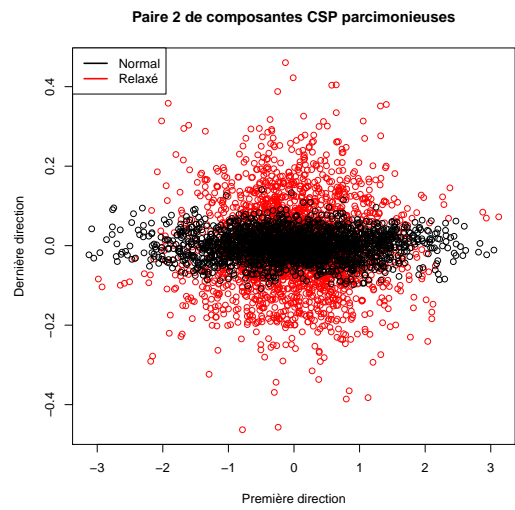
(a) Première paire de composantes CSP.



(b) Deuxième paire de composantes CSP.



(c) Première paire de composantes sCSP.



(d) Deuxième paire de composantes sCSP.

Figure VI.4 – Première paire de composantes CSP (a) et sCSP (c) et deuxième paire de composantes CSP (b) et sCSP (d).

5 Applications

Les données EEG utilisées sont identiques à celles représentées en Figure V.1 du Chapitre V. La méthode sCSP va être appliquée sur ces données afin d'en évaluer la qualité dans le cadre de la classification de signaux EEG pour l'étude d'états de vigilance. Une première application du sCSP va viser à comparer les résultats avec ceux obtenus avec l'algorithme génétique présenté dans le Chapitre V. Le coût computationnel du sCSP étant raisonnable, une seconde approche va viser à tester une succession de découpages en échantillons d'apprentissage et de test afin de mieux estimer le Taux de Bons Classements (TBC) obtenu à l'aide de cette méthode.

5.1 Comparaison avec les résultats de l'algorithme génétique proposé en Chapitre V

Afin de comparer les résultats du sCSP avec ceux obtenus en utilisant l'algorithme génétique dans le Chapitre V, le découpage permettant de créer l'échantillon d'apprentissage et de test va être identique à celui utilisé dans le Chapitre V. Ainsi, les signaux EEG de 13 sujets, enregistrés durant la première campagne d'acquisition, vont former l'échantillon d'apprentissage. L'échantillon test va, quant à lui, être composé des signaux EEG des 6 sujets recueillis durant la seconde campagne d'acquisition.

Dans le Chapitre V, avant d'utiliser l'algorithme génétique, certains paramètres ont été fixés par Leave One Out (LOO) sur les signaux de l'ensemble d'apprentissage. Ainsi, un filtre passe bande entre 7 et 30 Hz a été utilisé sur les signaux EEG puis ces derniers ont été découpés en essais de 8 secondes. Les essais ont été normalisés en utilisant les écarts types et le nombre de paires de filtres formées à l'aide de la méthode CSP a été fixé à 3. Toutes ces valeurs de paramètres vont être conservées pour le sCSP.

De la même façon que pour l'algorithme génétique, la règle de décision par sCSP va être formée sur l'échantillon d'apprentissage puis évaluée sur l'échantillon test. Pour commencer, la grille matricielle de paramètres de régularisation est obtenue à l'aide de l'Algorithme 4 appliqué aux matrices de données recodées Z_1 et Z_2 (cf. Section 4.3). Cette grille va ensuite être parcourue ligne par ligne pour faire varier le degré de parcimonie. Afin de choisir le meilleur pas (la meilleure ligne), un TBC va être estimé par LOO pour chacun des pas de la grille. Ainsi, pour chaque pas, $2k$ paramètres de régularisation sont obtenus. À chaque itération du LOO, tous les essais d'un sujet de la première campagne sont isolés et forment l'échantillon de validation. Les essais des autres sujets forment l'échantillon d'apprentissage. La méthode

sCSP est alors utilisée sur l'échantillon d'apprentissage en prenant en entrée les $2k$ valeurs des paramètres de régularisation correspondant à ce pas. Elle fournit une matrice de filtres \tilde{W} de dimensions $p' \times 2k$ où p' est le nombre d'électrodes sélectionnées et où $k = 3$ est le nombre de paires de filtres. La matrice $F = XW$ des composantes CSP est obtenue en ne considérant que les p' électrodes sélectionnées dans les matrices de signaux EEG \mathbf{X}_1 et \mathbf{X}_2 . Une transformation log-var est alors appliquée sur F et une LDA permet de former la règle de décision. Cette règle de décision va alors être appliquée sur des signaux provenant de l'échantillon de validation afin d'obtenir une prédiction. Lorsque toutes les itérations du LOO sont effectuées, un TBC est obtenu en comparant les prédictions aux valeurs réelles. La Figure VI.5 représente les TBC estimés par LOO en fonction du pas de la grille.

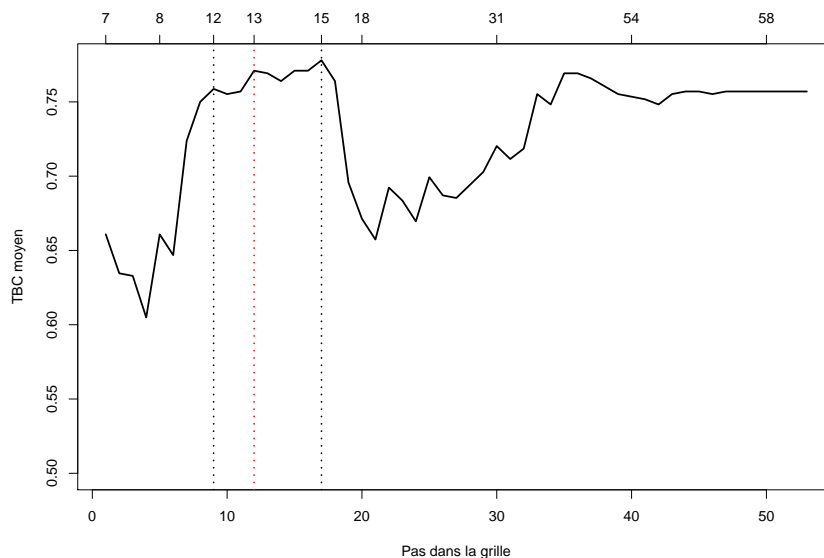


Figure VI.5 – Taux de bons classements obtenus par LOO sur les individus de la campagne 1 pour les différents pas de la grille matricielle de paramètres de régularisation. L'axe supérieur de la figure indique le nombre d'électrodes sélectionnées pour le pas correspondant dans la grille.

Les TBC obtenus sur l'échantillon d'apprentissage semblent corrects, dépassant les 75% pour certains pas de la grille. La Figure VI.5 va permettre de choisir le pas de la grille qui représente le meilleur compromis entre le nombre d'électrodes sélectionnées (p') et la qualité du modèle prédictif (TBC). L'axe supérieur de la figure indique le nombre d'électrodes sélectionnées pour le pas correspondant dans la grille. Trois valeurs de pas sont marquées par des lignes verticales en pointillé. Ainsi, la première valeur pour laquelle le TBC obtenu devient correct semble être le 9^{ème} pas (12 électrodes sélectionnées) car le TBC subit une forte hausse jusqu'à celui-ci puis semble se stabiliser pour les pas suivants. Le TBC maximal est obtenu avec le 17^{ème} pas de la

Chapitre VI. Le CSP parcimonieux

grille (15 électrodes). Le meilleur compromis semble être atteint avec le 12^{ème} pas qui correspond à la sélection de 13 électrodes. Le vecteur de paramètres de régularisation, correspondant au 12^{ème} pas, est donc conservé et donné en entrée à un sCSP lancé sur les essais des sujets de la campagne 1. Ainsi, 13 électrodes sont sélectionnées et la matrice des filtres \tilde{W} de taille 13×6 est obtenue. Une transformation log-var et une LDA permettent de former la règle de décision. Celle-ci est testée à l'aide des signaux des individus de la seconde campagne. Ainsi, un TBC de 73.11% est obtenu. La Table VI.1 permet de comparer les résultats obtenus par sCSP avec ceux obtenus avec l'algorithme génétique.

Tableau VI.1 – Comparaison entre les taux de bons classements obtenus avec l'algorithme génétique dans le Chapitre V (première ligne) et ceux obtenus avec le sCSP (seconde ligne).

	Méthode d'évaluation	Nombre d'électrodes dans le modèle prédictif	TBC estimé sur l'échantillon test (campagne 2)	
			essais	sujets
Algorithme génétique : CSP et transformation log-var	LDA	9	71.59	75
sCSP et transformation log-var	LDA	13	73.11	75

Les résultats des deux approches sont donc très proches. Même si l'algorithme génétique a permis de sélectionner un sous-ensemble d'électrodes plus petit, le TBC obtenu sur l'échantillon test pour le sCSP est supérieur à celui de l'algorithme génétique. Cependant, il est compliqué de comparer les deux approches alors qu'un seul découpage (échantillon d'apprentissage-test) a été utilisé. La principale différence vient des temps de calcul. En effet, afin de calculer les TBC pour chacun des pas de la grille, l'approche du sCSP a nécessité 1.86 heure (Linux sur une machine avec deux processeurs hexacœurs Intel Xeon x5675 @ 3,06 GHz). Même si les machines sur lesquelles les deux approches ont été lancées sont différentes, il semble que le sCSP soit beaucoup plus rapide (l'algorithme génétique prenant entre 9 et 15 heures sur une machine équipée de Linux 8*2.8 GHz Intel Xeon processeurs X5560 et 32 GB 1333 MHz DDR3 ECC RAM). Il convient surtout de noter que lors d'un lancement de l'algorithme génétique, une seule valeur de p' est considérée alors que l'approche du sCSP, en testant plusieurs pas de la grille, fait varier le degré de parcimonie des composantes et donc la valeur de p' au cours d'un lancement. Ainsi, on peut choisir le meilleur pas (et donc la meilleure valeur de p'), au sens du TBC obtenu, à la fin d'un lancement du sCSP alors qu'il faut lancer l'algorithme génétique pour plusieurs valeurs de p' avant de pouvoir choisir la meilleure valeur de p' . Enfin, le sCSP est déterministe au contraire de l'algorithme génétique ce qui oblige une répétition des lancements pour ce dernier, contribuant à augmenter considérablement le coût computationnel. Il y a donc une différence considérable en terme de coût computationnel entre les deux approches.

La localisation des 13 électrodes sélectionnées par sCSP est présentée en Figure VI.6. Les électrodes sélectionnées apparaissent en noir sur la figure. Comme déjà constaté avec l'algorithme génétique et les procédures séquentielles (backward CSP et forward CSP) dans le Chapitre V, les électrodes sélectionnées sont situées sur les contours du scalp.

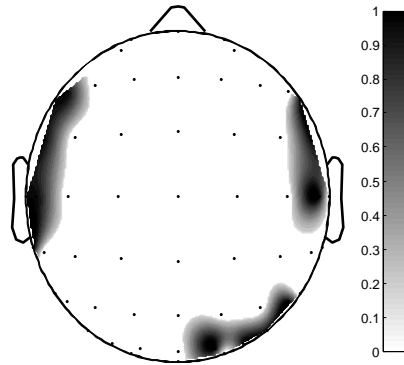


Figure VI.6 – Représentation de la localisation des 13 électrodes sélectionnées par sCSP. Les électrodes sélectionnées apparaissent en noir.

Enfin, la Figure VI.7 propose un diagramme de Venn afin de comparer le sous-ensemble de 13 électrodes sélectionné par sCSP avec celui de 9 électrodes obtenu par l'algorithme génétique. Les deux sous-ensembles partagent 6 électrodes. Ce résultat tend à prouver la cohérence des deux approches et permet de dégager un groupe d'électrodes importantes pour le problème de classification considéré.

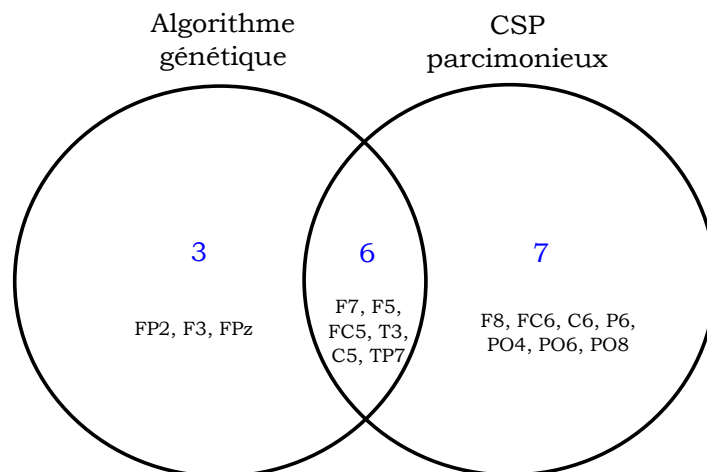


Figure VI.7 – Diagramme de Venn comparant le sous-ensemble de 9 électrodes sélectionnées en utilisant l'algorithme génétique (voir Section 4.2.2.3 du Chapitre V) et celui de 13 électrodes obtenu par sCSP.

5.2 Evaluation du sCSP sur 35 découpages différents

Une seconde approche vise à estimer les performances de la méthode sCSP sur différents découpages afin de disposer d'estimations plus précises (robustes) des TBC. Ainsi, 35 découpages en échantillons d'apprentissage (essais de 13 sujets) et test (essais de 6 sujets) sont formés. Pour chaque découpage, une approche similaire à celle qui vient d'être présentée est réalisée. Ainsi, une grille de paramètres de régularisation est formée et un sCSP est lancé pour construire la matrice \tilde{W} en utilisant uniquement l'échantillon d'apprentissage. Le choix du pas est optimisé par LOO sur les sujets de l'échantillon d'apprentissage. Une fois que ce pas est choisi, un sous-ensemble d'électrodes est obtenu. Une transformation log-var est appliquée et une LDA permet de construire la règle de décision qui pourra être utilisée pour prédire la classe des sujets de l'échantillon test. Ainsi, 35 TBC sont obtenus. La principale différence par rapport à l'approche précédente vient du choix du pas de la grille qui doit être ici automatique. Ainsi, le pas qui correspond au meilleur TBC trouvé est choisi. Une boîte à moustaches des TBC obtenus sur l'échantillon test en considérant les essais et les sujets est proposée en Figure VI.8(a). Les résultats semblent moins probants que ceux obtenus avec le découpage étudié en Section 5.1. Les médianes des TBC sont égales à 66.67% que ce soit en considérant les essais ou les sujets (par vote majoritaire des essais du sujet) tandis que les moyennes sont égales à 65.6% avec un écart type de 7.6% pour les essais et à 65.95 avec un écart type de 10.38 pour les sujets. Les écarts types des TBC sont importants et semblent montrer une instabilité des résultats vis-à-vis du découpage. Cela vient certainement de la forte variabilité inter-individuelle déjà observée en Chapitre IV. Une autre boîte à moustaches, présentée en Figure VI.8(b), permet de voir le nombre de variables sélectionnées par sCSP au cours des 35 découpages. La médiane du nombre d'électrodes est égale à 15 électrodes. Cela semble montrer que choisir le pas qui donne le TBC maximal n'entraîne pas forcément la formation d'un modèle contenant un nombre très élevé d'électrodes. Cela semble être un point important, offrant des perspectives intéressantes dans le cadre de l'utilisation de cette méthode pour d'autres travaux. En moyenne, 19.6 électrodes sont sélectionnées (avec un écart type de 11.75).

La localisation des électrodes sélectionnées au cours des 35 lancements du sCSP est représentée en Figure VI.9. Sur cette figure, l'occurrence de sélection de chacune des électrodes est associée à un code couleur allant du bleu, pour les électrodes jamais ou peu sélectionnées, vers le rouge pour les électrodes très souvent sélectionnées. Il apparaît que les électrodes fréquemment sélectionnées sont similaires à celles sélectionnées par la méthode sCSP sur le découpage campagne 1 - campagne 2 (voir Figure VI.6). Il semble que les compositions des sous-ensembles de variables sélectionnées soient stables (beaucoup d'électrodes dépassent les 20 sélections).

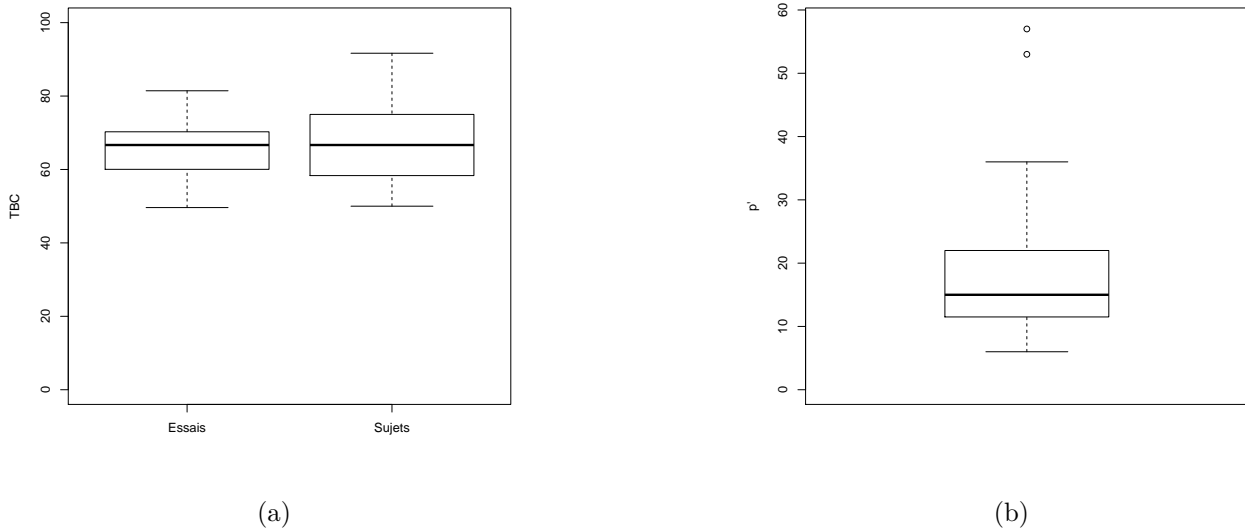


Figure VI.8 – Boîtes à moustaches des TBC (a) et des nombres d'électrodes sélectionnées (b) lors du lancement du sCSP sur 35 découpages différents (13 sujets en échantillon d'apprentissage et 6 en test). Les résultats sont obtenus sur l'échantillon test.

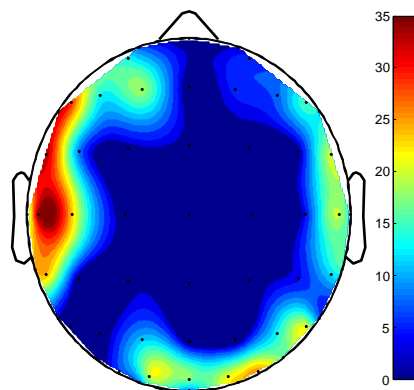


Figure VI.9 – Représentation de la localisation des électrodes sélectionnées par sCSP lors des 35 répétitions de découpage.

Enfin, les travaux présentés dans cette Section ont bénéficié des installations informatiques MCIA (Mésocentre de Calcul Intensif Aquitain) de l'Université de Bordeaux et de l'Université de Pau et des Pays de l'Adour. Ainsi, les calculs pour les 35 découpages ont été exécutés sur 35 nœuds où chaque nœud correspond à une machine à 2 processeurs hexacœurs Intel Xeon x5675 @ 3,06 GHz. Ainsi, le coût computationnel est de 3.55 heures pour les 35 découpages.

6 Conclusion

Dans ce chapitre, un algorithme de CSP parcimonieux a été proposé. Dans un premier temps, le CSP a été réécrit comme une double ACP avec métrique puis l'ACP parcimonieuse a été présentée. L'ACP parcimonieuse étant définie avec une métrique identité, un recodage des données a été introduit afin d'écrire le CSP comme une double ACP avec métrique identité sur \mathbb{R}^p . Le choix des paramètres de régularisation a été discuté et une approche de parcours de la grille de régularisation a été proposée. Les résultats obtenus avec le sCSP ont été comparés à ceux obtenus avec l'algorithme génétique. Les résultats des deux approches semblent proches tant au niveau des électrodes sélectionnées qu'au niveau des TBC obtenus. Le sCSP présente cependant l'avantage d'avoir un coût computationnel bien inférieur à celui de l'algorithme génétique. Il s'agit aussi de la seule approche présentée dans ces travaux qui permette d'estimer ses performances sur plusieurs découpages car elle combine un coût d'exécution faible (ce qui n'est pas le cas avec l'algorithme génétique présenté en Chapitre V) et un critère d'arrêt (manquant dans les approches séquentielles présentées en Chapitre V). Ainsi, 35 découpages différents (13 sujets dans l'échantillon d'apprentissage et 6 dans l'échantillon test) ont été considérés. Le pas de la grille de régularisation a été choisi automatiquement en sélectionnant celui dont le TBC estimé par LOO est le plus grand. Les TBC obtenus sur les échantillons test, bien qu'inférieurs à celui obtenu sur le découpage initialement considéré (campagne 1-campagne 2) restent satisfaisants (moyenne de 65.6% avec un écart type de 7.6%). Le nombre moyen d'électrodes sélectionnées (19.6 avec un écart type de 11.75) montre que la stratégie de parcours de la grille proposée dans ces travaux ainsi que la sélection du pas automatique permettent de former des modèles dont le nombre de variables est raisonnable.

Le travail proposé dans ce chapitre peut faire l'objet de futures extensions. Par exemple, le parcours de la grille et la méthode de sélection automatique du pas peuvent faire l'objet d'améliorations afin de réduire encore le nombre d'électrodes sélectionnées (visiter uniquement les premiers pas mais de manière plus exhaustive, utiliser la fonction *lars()* qui est plus précise que *glmnet()* avec une autre stratégie de parcours de la grille, ...). De plus, des travaux afin d'éviter de recoder les données et d'inclure directement un changement de métrique dans l'algorithme d'analyse en composantes principales parcimonieuse sont en cours.

D'autres approches de l'analyse en composantes principales parcimonieuse (voir par exemple [92]) peuvent être envisagées afin de comparer les résultats obtenus avec ceux de ce chapitre. Enfin, une version différente du CSP parcimonieux a été développée dans [49] (voir Section 3.3

du Chapitre III). Une étude pourrait donc être réalisée afin de comparer notre version du sCSP présentée dans ce chapitre avec celle proposée dans [49].

Conclusion

Synthèse

Cette thèse s'inscrit dans le cadre d'un large projet qui vise à créer une interface Cerveau-Ordinateur permettant de modifier l'état de vigilance d'un individu à l'aide de stimuli musicaux générés en temps réel. Afin d'adapter les stimuli présentés au sujet en fonction de son état de vigilance, une méthode capable de mesurer cet état, automatiquement et en temps réel, devait être construite.

Comme présenté dans le Chapitre I, il existe de nombreuses méthodes permettant de recueillir l'activité cérébrale d'un humain. Au cours de ces travaux de thèse, l'électroencéphalographie a été privilégiée, principalement en raison de son faible coût et de son caractère portatif. Une expérience, présentée dans le Chapitre II a été élaborée afin de recueillir les signaux électroencéphalographiques de 58 sujets (au cours de deux campagnes différentes) mesurés à l'aide d'un casque à 58 électrodes dans deux états de vigilance distincts (états de vigilance haute et basse, labellisés par la suite état « normal » et « relaxé »). Les signaux recueillis ont alors servi de base d'étude pour ces travaux de thèse. Après une étape de validation des données, 19 des 58 sujets ont été conservés (13 de la première campagne et 6 de la seconde). Dès lors, deux objectifs principaux ont été poursuivis :

- Construire un modèle de prédiction (règle de décision) permettant d'obtenir une prédiction fiable de l'état de vigilance d'un nouveau sujet. Dans ces travaux, le problème de classification était binaire (état normal et état relaxé).

- Sélectionner un sous-ensemble d'électrodes afin de réduire le temps nécessaire pour installer le casque à électrodes sur un sujet ainsi que son coût.

Afin de remplir ces objectifs, plusieurs approches ont été envisagées. Dans le Chapitre IV, une méthode d'extraction de caractéristiques, permettant de faciliter la tâche de classification, a été proposée. Elle consistait à utiliser une décomposition en ondelettes discrète afin de calculer les énergies de bandes de fréquences. Une régression linéaire était ensuite appliquée sur certaines de ces fréquences et la pente de celle-ci était conservée afin d'obtenir un critère, appelé « critère des pentes ». Initialement, cette régression a été calculée sur les fréquences 4, 8 et 16 Hz afin d'inclure uniquement la bande de fréquences caractéristique des ondes α (8-12 Hz). Les méthodes de classification usuelles ont alors été utilisées afin de calculer un taux de bons classements. Les résultats obtenus étant décevants, un algorithme génétique a ensuite été proposé afin d'optimiser le choix des fréquences sur lesquelles appliquer la régression. De plus, cet algorithme génétique a permis de sélectionner une électrode optimale parmi les 58 initiales. Les résultats obtenus ont montré qu'un modèle de prédiction fiable (taux de bons classements moyen égal à $86.68 \pm 1.87\%$ évalué par validation croisée de type 5 folds) pouvait être obtenu tout en ne sélectionnant qu'une seule électrode parmi les 58 initiales. Le principal inconvénient de cette approche est lié au fait que, lors de la réalisation de ces travaux, les signaux de seulement 13 sujets étaient disponibles. Ainsi, il était délicat de laisser un sous-groupe de sujets à l'extérieur de l'algorithme génétique afin de tester la solution obtenue. C'est pourquoi, bien que les taux de bons classements aient été estimés à l'aide d'une validation croisée (en découpant l'échantillon initial en ensemble d'apprentissage et de test), il se peut que l'estimation du taux de bons classements obtenue soit trop optimiste.

Une autre approche est présentée dans les Chapitres V et VI. La méthode du *Common Spatial Pattern* (CSP) couplée à une transformation log-var afin d'extraire des caractéristiques des signaux électroencéphalographiques est déclinée dans de nouvelles variantes. En effet, l'utilisation faite du *common spatial pattern* au cours de ces travaux se démarque de celles rencontrées dans la littérature, notamment par la normalisation des données électroencéphalographiques utilisée et par le fait que tous les sujets sont utilisés simultanément pour calculer les filtres spatiaux; le *common spatial pattern* étant d'habitude exploité afin d'obtenir des filtres spatiaux spécifiques à un sujet.

Le Chapitre V propose deux types de méthodes permettant la sélection d'électrodes : l'utilisation d'un algorithme génétique ou de méthodes de recherche séquentielle. Le but de ces

méthodes est de trouver un sous-ensemble d'électrodes qui permette d'obtenir un modèle de prédiction fiable de l'état de vigilance. Les évaluations des solutions potentielles sont réalisées en utilisant une méthode de classification (analyse linéaire discriminante) sur les caractéristiques extraites par *common spatial pattern* et transformation log-var. Les taux de bons classements sont estimés par *Leave One Out* (LOO) sur les signaux des sujets de la campagne 1 afin de garder une base de sujets totalement extérieurs à la procédure de recherche du sous-ensemble d'électrodes (campagne 2). Les résultats obtenus semblent montrer que l'algorithme génétique surpasse les autres méthodes de sélection de variables testées mais son coût computationnel, dans l'approche développée, est important. Les méthodes de recherche séquentielle proposées ont permis de réduire le temps de calcul tout en permettant d'obtenir des résultats proches de ceux de l'algorithme génétique (en terme de sélection et de taux de bons classements). En utilisant les résultats de l'algorithme génétique, un sous-ensemble de 9 électrodes a été conservé, permettant l'obtention d'un modèle fiable (moyenne de taux de bons classements égal à 71.59% sur un ensemble de test externe à l'algorithme génétique). Ce modèle de prédiction permet, en outre, d'obtenir une prédiction très rapide de l'état de vigilance d'un nouveau sujet.

Enfin, le Chapitre VI propose une méthode de *common spatial pattern* parcimonieux fondée sur les travaux réalisés par [91] sur l'analyse en composantes principales parcimonieuse. Comme il est prouvé dans ce chapitre, le *common spatial pattern* peut être perçu comme une analyse en composantes principales avec une certaine métrique. Dans la version proposée, les données sont alors recodées afin de pouvoir réaliser un *common spatial pattern* parcimonieux à l'aide de l'algorithme d'analyse en composantes principales parcimonieuse. Les résultats obtenus, tant au niveau du nombre de variables sélectionnées qu'au niveau de la qualité des modèles de classification obtenus, semblent très convaincants. La méthode de *common spatial pattern* parcimonieux est déterministe et présente un coût computationnel très raisonnable. Il s'agit aussi de la seule approche présentée dans ces travaux qui permette d'estimer les performances de celle-ci sur différents découpages car elle combine un coût d'exécution faible (ce qui n'est pas le cas avec l'algorithme génétique présenté en Chapitre V) et un critère d'arrêt (manquant dans les approches séquentielles présentées en Chapitre V).

Perspectives

Dans le Chapitre V, le coût computationnel de l'approche par algorithme génétique est important. Le temps de calcul est principalement concentré sur la méthode d'extraction de caractéristiques utilisée. En effet, l'étape la plus coûteuse est l'inversion de la matrice de cova-

riance de taille $p \times p$ lors de la recherche des filtres spatiaux par *common spatial pattern*. Ainsi, il semble inutile de changer la méthode de classification. Cependant, des solutions peuvent être envisagées afin de réduire le temps de calcul (baisse du nombre d'enfants, changement de la méthode d'estimation du taux de bons classements). La recherche du meilleur sous-ensemble d'électrodes peut aussi être réalisée par une approche séquentielle telle que proposée dans le Chapitre V. Ces procédures semblent donner des résultats satisfaisants sur nos données et relativement proches de ceux de l'algorithme génétique (principalement pour le *backward common spatial pattern*). La principale piste de recherche pour rendre ces procédures séquentielles automatiques est de définir un critère d'arrêt. Toute la difficulté d'un tel travail est d'arriver à traduire le compromis souhaité, entre le nombre de variables conservées et le taux de bons classements obtenu, en un critère d'arrêt pour la méthode. De plus, des versions plus sophistiquées des algorithmes *Forward* et *Backward* (*sequential forward floating* et *sequential backward floating*) peuvent être envisagées afin d'améliorer le parcours de l'espace de recherche [90] et ainsi rivaliser avec les résultats obtenus au moyen de l'algorithme génétique.

La méthode de *common spatial pattern* parcimonieux proposée en Chapitre VI est certainement le point qui offre le plus de perspectives. En effet, dans la version proposée, les données sont recodées afin de pouvoir réaliser un *common spatial pattern* parcimonieux à l'aide de l'algorithme d'analyse en composantes principales parcimonieuse. À cet effet, des travaux afin d'éviter de recoder les données et d'inclure directement un changement de métrique dans l'algorithme d'analyse en composantes principales parcimonieuse sont en cours de finalisation. De plus, l'approche proposée afin de choisir le vecteur des paramètres de régularisation λ peut être révisée (parcours différent de la grille matricielle Λ des paramètres de régularisation, par exemple). Il serait également intéressant de comparer les résultats obtenus en se fondant sur d'autres approches de l'analyse en composantes principales parcimonieuse (voir par exemple [92]). Une étude pourrait être réalisée afin de comparer notre version du *common spatial pattern* parcimonieux avec celle proposée dans [49].

L'utilisation de la méthode du *common spatial pattern* peut également être étendue à des données présentant la même structure que celles des signaux d'électroencéphalographie (3 dimensions : temps, sujets et capteurs) comme par exemple les signaux de l'électrocardiogramme (ECG) lorsque la variable à prédire est binaire.

Enfin, il est évident qu'une faiblesse de ce travail se situe dans l'absence de confrontation des résultats obtenus aux modèles disponibles en électrophysiologie humaine. Une telle étude

permettrait de fournir une évaluation de la qualité de ce travail.

Pour conclure, le travail présenté propose une solution au problème initial. En effet, la conception de l'interface Cerveau-Ordinateur semble maintenant envisageable. Une première esquisse du prototype permettant d'influencer l'état de vigilance d'un sujet à l'aide de musique synthétisée est représentée en Figure VI.10. Sur cette figure, l'interface Cerveau-Ordinateur est encadrée en pointillée. Dans le protocole expérimental proposé, la première étape est naturellement la mise en place du casque à électrodes. Seules 9 électrodes actives seront utilisées pour les enregistrements EEG conformément aux résultats obtenus en Section 4.2.2 du Chapitre V (voir Figure V.14). Cela représentera un gain de temps conséquent par rapport au protocole expérimental utilisé pour l'acquisition des données utilisées dans ce travail (voir Chapitre II). En effet, environ 45 minutes étaient nécessaires pour la mise en place des 58 électrodes actives avec ce protocole expérimental.

Une fois le casque à électrodes mis en place, une vérification de l'état de vigilance du sujet va être effectuée. En effet, le but de l'interface Cerveau-Ordinateur développée est de modifier l'état de vigilance d'un sujet en le menant vers un état de vigilance cible. Il est donc important de déterminer son état de vigilance initial afin de s'assurer que celui-ci est réellement différent de l'état cible. C'est pourquoi, un premier enregistrement EEG sera réalisé. L'état de vigilance initial du sujet sera alors prédit à l'aide de la méthode présentée en Section 6 du Chapitre V (partie en bleu sur la Figure VI.10). Si cet état est différent de l'état cible, l'interface Cerveau-Ordinateur sera proposée au sujet. Il est important de remarquer que l'utilisation de la méthode de prédiction proposée en Section 6 du Chapitre V n'est pas redondante avec l'utilisation de la variation contingente négative (voir Section 3.1.1.9 du Chapitre I et Section 8 du Chapitre II). En effet, le recueil de la variation contingente négative inclut des perturbations susceptibles de modifier l'état de vigilance du sujet (diffusion de signaux sonores, sollicitation de mouvements). La méthode de prédiction de l'état de vigilance proposée en Section 6 du Chapitre V, n'incluant pas ces perturbations, est donc privilégiée ici.

Au cours de ces travaux, il a été montré que lorsque l'on dispose déjà d'un signal électroencéphalographique d'un individu, il est possible de savoir si un nouveau signal considéré a été enregistré alors que le sujet était dans un état de vigilance plus élevé ou plus bas (voir la Section 3.4 du Chapitre IV). Ainsi, en extrayant le critère des pentes sur les signaux EEG recueillis sur le sujet durant l'utilisation de l'interface Cerveau-Ordinateur, il sera possible de quantifier les changements d'état de vigilance d'un individu en comparant les deux signaux électroencéphalographiques recueillis en deux instants distincts.

Le fonctionnement de l'interface Cerveau-Ordinateur est le suivant. Une valeur initiale du

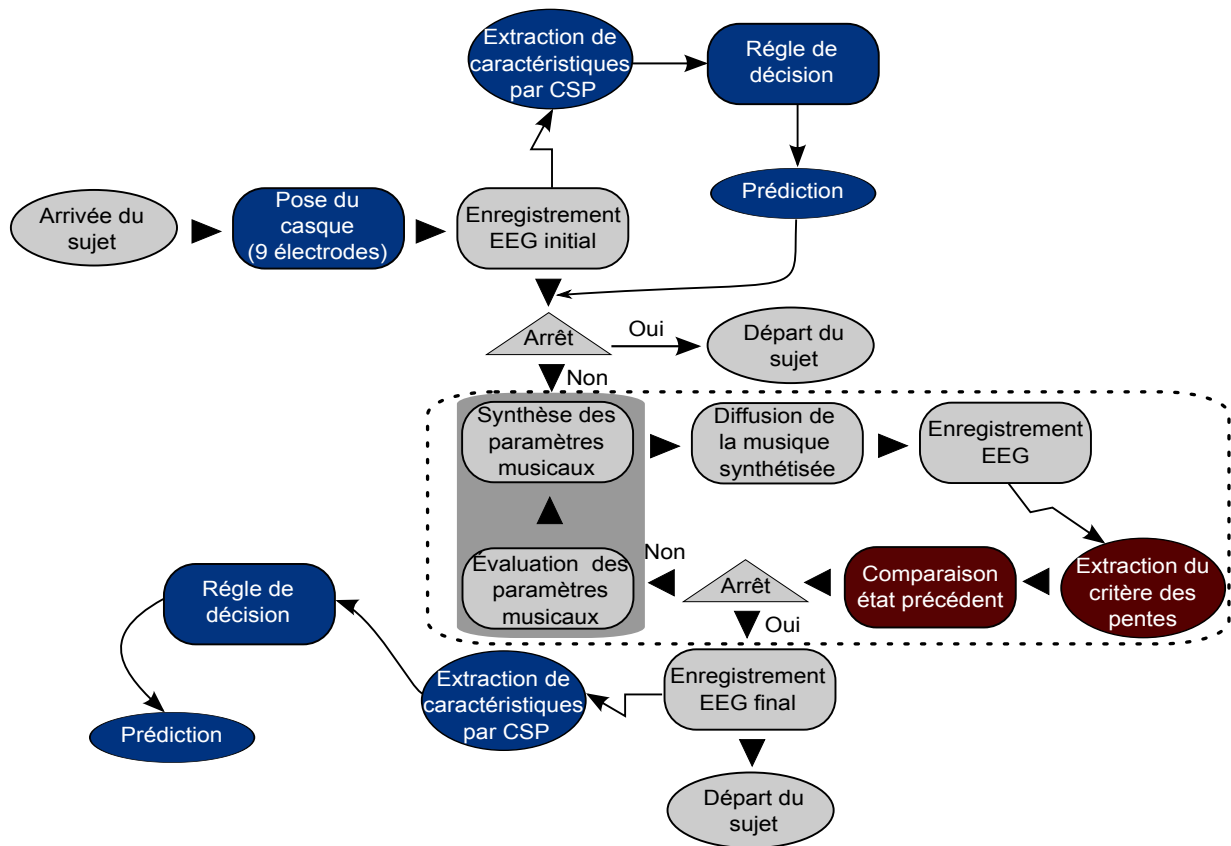


Figure VI.10 – Représentation du protocole expérimentale permettant d’amener un sujet vers un état de vigilance cible. La partie encadrée en pointillées correspond à l’interface Cerveau-Ordinateur (ICM). Les parties représentées en rouge et en bleu correspondent aux apports principaux de cette thèse. La partie grisée correspond à l’algorithme génétique réalisé par Mr Legrand pour l’optimisation des paramètres musicaux.

critère des pentes est obtenue en l’extrayant du signal EEG initial. Un premier jeu de paramètres musicaux est alors généré aléatoirement permettant ainsi de synthétiser et de diffuser une musique de quelques secondes au sujet. Celle-ci va entraîner une modification de l’état de vigilance. Un signal EEG va être enregistré à partir duquel une nouvelle valeur du critère des pentes va être extraite. Cette mesure va être comparé à la valeur initiale afin d’obtenir une évaluation des paramètres musicaux. Ces derniers vont être adaptés afin d’amener le sujet vers l’état de vigilance cible. Une nouvelle musique va alors être diffusée et le processus va être itéré jusqu’à ce qu’un critère d’arrêt de l’interface Cerveau-Ordinateur soit atteint (nombre d’itération maximal atteint, état de vigilance jugé suffisamment proche de l’état cible). La partie de l’interface Cerveau-Ordinateur, représentée en grisée sur la Figure VI.10 (évaluation et synthèse des paramètres musicaux), est effectuée à l’aide d’un algorithme génétique réalisé par Mr Legrand.

Enfin, lorsque le critère d'arrêt de l'interface Cerveau-Ordinateur sera atteint (nombre d'itération maximal atteint, état de vigilance jugé suffisamment proche de l'état cible), un dernier enregistrement EEG sera réalisé. La méthode présentée en Section 6 du Chapitre V sera appliquée afin de prédire l'état de vigilance du sujet et ainsi vérifier que l'interface a permis de l'amener vers l'état de vigilance cible.

Le protocole expérimental qui vient d'être présenté va faire l'objet d'une étape de validation sur plusieurs sujets « pilotes » afin de déterminer les conditions optimales de réalisation du protocole (ajustement de la durée de diffusion des musiques synthétisées, ajustement des critères d'arrêt de l'interface, etc). Une fois validé, ce protocole pourra être utilisé afin de modifier l'état de vigilance de sujets.

Réduction de dimension en apprentissage supervisé. Applications à l'étude de l'activité cérébrale.

L'objectif de ce travail est de développer une méthode capable de déterminer automatiquement l'état de vigilance chez l'humain. Les applications envisageables sont multiples. Une telle méthode permettrait par exemple de détecter automatiquement toute modification de l'état de vigilance chez des personnes qui doivent rester dans un état de vigilance élevée (par exemple, les pilotes ou les personnels médicaux).

Dans ce travail, les signaux électroencéphalographiques (EEG) de 58 sujets dans deux états de vigilance distincts (état de vigilance haut et bas) ont été recueillis à l'aide d'un casque à 58 électrodes posant ainsi un problème de classification binaire. Afin d'envisager une utilisation de ces travaux sur une application du monde réel, il est nécessaire de construire une méthode de prédiction qui ne nécessite qu'un faible nombre de capteurs (électrodes) afin de limiter le temps de pose du casque à électrodes ainsi que son coût. Au cours de ces travaux de thèse, plusieurs approches ont été développées.

Une première approche propose d'utiliser un pré-traitement des signaux EEG fondé sur l'utilisation d'une décomposition en ondelettes discrète des signaux EEG afin d'extraire les contributions de chaque fréquence dans le signal. Une régression linéaire est alors effectuée sur les contributions de certaines de ces fréquences et la pente de cette régression est conservée. Un algorithme génétique est utilisé afin d'optimiser le choix des fréquences sur lesquelles la régression est réalisée. De plus, cet algorithme génétique permet la sélection d'une unique électrode.

Une seconde approche est fondée sur l'utilisation du Common Spatial Pattern (CSP). Cette méthode permet de définir des combinaisons linéaires des variables initiales afin d'obtenir des signaux synthétiques utiles pour la tâche de classification. Dans ce travail, un algorithme génétique ainsi que des méthodes de recherche séquentielle ont été proposés afin de sélectionner un sous groupes d'électrodes à conserver lors du calcul du CSP.

Enfin, un algorithme de CSP parcimonieux fondé sur l'utilisation des travaux existant sur l'analyse en composantes principales parcimonieuse a été développé.

Les résultats de chacune des approches seront détaillés et comparés. Ces travaux ont aboutit sur l'obtention d'un modèle permettant de prédire de manière rapide et fiable l'état de vigilance d'un nouvel individu.

Mots clés : EEG, Vigilance, algorithme génétique, Common Spatial Pattern, CSP parcimonieux, VCN, décomposition en ondelettes discrète.

Dimensionality reduction for supervised learning. Applications to the study of brain activity.

The aim of this work is to develop a method able to automatically determine the alertness state of humans. Such a task is relevant to diverse domains, where a person is expected or required to be in a particular state. For instance, pilots, security personnel or medical personnel are expected to be in a highly alert state, and this method could help to confirm this or detect possible problems.

In this work, electroencephalographic data (EEG) of 58 subjects in two distinct vigilance states (state of high and low alertness) were collected via a cap with 58 electrodes. Thus, a binary classification problem is considered. In order to use of this work on a real-world applications, it is necessary to build a prediction method that requires only a small number of sensors (electrodes) in order to minimize the time needed by the cap installation and the cap cost. During this thesis, several approaches have been developed.

A first approach involves use of a pre-processing method for EEG signals based on the use of a discrete wavelet decomposition in order to extract the energy of each frequency in the signal. Then, a linear regression is performed on the energies of some of these frequencies and the slope of this regression is retained. A genetic algorithm (GA) is used to optimize the selection of frequencies on which the regression is performed. Moreover, the GA is used to select a single electrode .

A second approach is based on the use of the Common Spatial Pattern method (CSP). This method allows to define linear combinations of the original variables to obtain useful synthetic signals for the task classification. In this work, a GA and a sequential search method have been proposed to select a subset of electrode which are keep in the CSP calculation.

Finally, a sparse CSP algorithm, based on the use of existing work in the sparse principal component analysis, was developed.

The results of the different approaches are detailed and compared. This work allows us to obtaining a reliable model to obtain fast prediction of the alertness of a new individual.

Keywords : EEG, Alertness, Genetic Algorithm, Sparse CSP, CNV, wavelet decomposition.

Références bibliographiques

- [1] M. DÉCHELETTE AND V. DELAVIGNE. *Les tumeurs du cerveau*. Institut National du Cancer. [xi](#), [8](#), [31](#)
- [2] G. E. CHATRIAN, E. LETTICH, AND P. L. NELSON. *Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity*. *Journal of Clinical Neurophysiology* **25**, 83–92 (1985). [xi](#), [12](#), [13](#)
- [3] H. H. JASPER. *Report of the committee on methods of clinical examination in electroencephalography*. *Electroencephalography and clinical neurophysiology* **10**, 1 – 370 (1958). [xi](#), [13](#)
- [4] R. OOSTENVELD AND P. PRAAMSTRA. *The five percent electrode system for high-resolution EEG and ERP measurements*. *Clinical Neurophysiology* **112**(4), 713–719 (2001). [xi](#), [13](#)
- [5] S. LAUREYS. *The neural correlate of (un)awareness : lessons from the vegetative state*. *Trends in Cognitive Sciences* **9**(12), 556 – 559 (2005). [xii](#), [22](#)
- [6] J. LANGNER. *Event-Driven Motion Compensation in Positron Emission Tomography : Development of a Clinically Applicable Method*. Thèse de Doctorat, University of Technology Dresden, Germany (2008). [xii](#), [32](#)
- [7] P. S. CHURCHLAND AND T. J. SEJNOWSKI. *Perspectives on cognitive neuroscience*. *Science* **242**, 741–745 (1988). [xii](#), [33](#)
- [8] A. LOTTE, F. AND LÉCUYER AND B. ARNALDI. *Les Interfaces Cerveau-Ordinateur : Utilisation en Robotique et Avancées Récentes*. In *Journées Nationales de la Recherche en Robotique*, Obernai, France (2007). [xii](#), [34](#)
- [9] L. A. FARWELL AND E. DONCHIN. *Talking off the top of your head : Toward a mental prosthesis utilizing event-related brain potentials*. *Electroencephalography and Clinical Neurophysiology* **70**, 510–523 (1988). [xii](#), [1](#), [36](#), [37](#), [90](#)

- [10] R. FAZEL-REZAI AND K. ABHARI. *A region-based P300 speller for brain-computer interface*. Electrical and Computer Engineering, Canadian Journal of **34**(3), 81–85 (2009). [xii](#), [36](#)
- [11] E. LALOR, S. P. KELLY, C. FINUCANE, R. BURKE, R. SMITH, R. B. REILLY, AND G. MCDARBY. *Steady-state VEP-based brain-computer interface control in an immersive 3d gaming environment*. EURASIP Journal on Advances in Signal Processing **2005**(19), 706906 (2005). [xii](#), [1](#), [39](#)
- [12] A. SCHLÖGL, F. LEE, H. BISCHOF, AND G. PFURTSCHELLER. *Characterization of four-class motor imagery EEG data for the BCI-competition 2005*. Journal of Neural Engineering pages 14–22 (2005). [xiv](#), [80](#), [81](#)
- [13] B. BLANKERTZ, K. R. MÜLLER, D. J. KRUSIENSKI, G. SCHALK, J. R. WOLPAW, A. SCHLÖGL, G. PFURTSCHELLER, J. D. R. MILLÁN, M. SCHRODER, AND N. BIRBAUMER. *The BCI competition III : Validating alternative approaches to actual BCI problems*. IEEE Transactions on neural systems and rehabilitation engineering **14**(2), 153–159 (2006). [xiv](#), [80](#), [81](#)
- [14] F. LOTTE AND C. GUAN. *Regularizing common spatial patterns to improve BCI designs : unified theory and new algorithms*. IEEE Transactions on biomedical Engineering **58** (2), 355–362 (2011). [xviii](#), [41](#), [85](#), [123](#), [146](#)
- [15] I. ITURRATE, J.M. ANTELIS, A. KUBLER, AND J. MINGUEZ. *A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation*. Robotics, IEEE Transactions on **25**(3), 614–627 (2009). [1](#), [37](#)
- [16] G. R. MÜLLER-PUTZ AND G. PFURTSCHELLER. *Control of an electrical prosthesis with an SSVEP-based BCI*. IEEE Trans. Biomed. Engineering **55**(1), 361–364 (2008). [1](#), [38](#)
- [17] F. LOTTE, A. VAN LANGHENHOVE, F. LAMARCHE, T. ERNEST, Y. RENARD, B. ARNALDI, AND A. LÉCUYER. *Exploring large virtual environments by thoughts using a brain-computer interface based on motor imagery and high-level commands*. Presence : Teleoper. Virtual Environ. **19**(1), 54–70 February (2010). [1](#), [40](#)
- [18] F. LOTTE, Y. RENARD, AND A. LÉCUYER. *Self-Paced Brain-Computer Interaction with Virtual Worlds : A Quantitative and Qualitative Study "Out of the Lab"*. In *4th international Brain Computer Interface Workshop and Training Course*, Graz, Austria (2008). Graz University of Technology. [1](#), [39](#)
- [19] A TSOI, D SO, AND A SERGEJEV. *Classification of electroencephalogram using neural network*. Neural Information Processing System **6**, 1–7 (1993). [2](#)
- [20] C. ANDERSON AND Z. SIJERCIC. *Classification of EEG signals from four subjects during five mental tasks*. Proceedings of the Conference on Engineering Applications in Neural Networks, London, United Kingdom pages 407–414 (1996). [2](#)

-
- [21] T. JUNG, S. MAKEIG, M. STENSMO, AND T. SEJNOWSKI. *Estimating alertness from the EEG power spectrum*. IEEE Transactions on Biomedical Engineering **44**, 60–69 (1997). [2](#)
- [22] M. SCHERG AND D. VON CRAMON. *Evoked dipole source potentials of the human auditory cortex*. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section **65**(5), 344 – 360 (1986). [12](#)
- [23] D. SCHWARTZ, BADIÉ J., BIHOUEE P., AND BOULIOU A. *Evaluation of a new MEG-EEG spatio-temporal localization approach with realistic sources*. Brain Topography **11**, 279–289 (1999). [12](#)
- [24] O. FAUGERAS, F. CLÉMENT, R. DERICHE, R. KERIVEN, T. PAPADOPOULO, J. ROBERTS, F. VIÉVILLE, T. AND DEVERNAY, J. GOMES, G. HERMOSILLO, P. KORNPROBST, AND D. LINGRAND. *The inverse EEG and MEG problems : The adjoint state approach I : The continuous case*. Rapport de recherche RR-3673 INRIA (1999). Projet CERMICS. [12](#)
- [25] M. VAN DE VELDE, G. VAN ERP, AND P. J. M. CLUITMANS. *Detection of muscle artefact in the normal human awake EEG*. Electroencephalography and Clinical Neurophysiology **107**(2), 149–158 (1998). [14](#)
- [26] M. DE VOS, S. RIES, K. VANDERPERREN, B. VANRUMSTE, F. X. ALARIO, S. VAN HUFFEL, AND B. BURLE. *Removal of muscle artifacts from EEG recordings of spoken language production*. Neuroinformatics **8**(2), 135–50 (2010). [14](#)
- [27] M. FATOURECHI, A. BASHASHATI, R. K. WARD, AND G. E. BIRCH. *EMG and EOG artifacts in brain computer interface systems : A survey*. Clinical Neurophysiology **118**, 480–494 (2007). [14](#)
- [28] A. B. USAKLI. *Improvement of EEG signal acquisition : an electrical aspect for state of the art of front end*. Intell. Neuroscience **2010** January (2010). [15](#)
- [29] M. TIMSIT-BERTHIER AND A. GERONO. *Manuel d'interprétation des potentiels évoqués endogènes (P300 et VCN)*. Mardaga (1998). [24](#)
- [30] W. G. WALTER, R. COOPER, V.J. ALDRIDGE, W. C. MCCALLUM, AND A.L. WINTER. *Contingent negative variation : An electric sign of sensorimotor association and expectancy in the human brain*. Nature **203**, 380–384 (1964). [24](#)
- [31] W. A. ROSENBLITH. *Some quantifiable aspects of the electrical activity of the nervous system (with emphasis upon responses to sensory stimuli)*. Revs. Mod. Physics **31**, 532–545 (1959). [25](#)
- [32] J. J. TECCE. *A CNV rebound effect*. Electroencephalography and Clinical Neurophysiology **46**, 546–551 (1979). [26](#), [27](#)
- [33] P. NAITOH, L. C. JOHNSON, AND A. LUBIN. *Modification of surface negative slow potential (CNV) in the human brain after total sleep loss*. Electroencephalography and Clinical Neurophysiology **30**, 17–22 (1971). [26](#), [27](#)

- [34] M. TIMSIT-BERTHIER, A. GERONO, AND H. MANTANUS. *Inversion de polarité de la variation contingente négative au cours d'état d'endormissement*. EEG Neurophysiol **11**, 82–88 (1981). [26](#), [27](#)
- [35] D. COHEN. *Magnetoencephalography : Evidence of magnetic fields produced by alpha-rhythm currents*. Science **161(3843)**, 784–786 (1968). [28](#)
- [36] R. H. GARRETT AND C. M. GRISHAM. *Biochimie*. De Boeck Université (2000). Table 1.1 en page 6. [28](#)
- [37] T. S. ARMSTRONG, M. Z. COHEN, J. WEINBERG, AND M. R. GILBERT. *Imaging techniques in neuro-oncology*. Seminars in Oncology Nursing **20(4)**, 231 – 239 (2004). [31](#)
- [38] V. J. WEDEEN, D. L. ROSENE, R. WANG, G. DAI, F. MORTAZAVI, P. HAGMANN, J. H. KAAS, AND W. Y. I. TSENG. *The geometric structure of the brain fiber pathways*. Science **335(6076)**, 1628–1634 (2012). [31](#)
- [39] H. CATENOIX, M. GUÉNOT, F. MAUGUIÈRE, AND J. ISNARD. *Thermocoagulations multiples guidées par la SEEG et malformations de développement cortical*. Épilepsies **21**, 54–62 (2009). [32](#)
- [40] B. OBERMAIER, G. R. MULLER, AND G. PFURTSCHELLER. *"virtual keyboard" controlled by spontaneous EEG activity*. Neural Systems and Rehabilitation Engineering, IEEE Transactions on **11(4)**, 422–426 (2003). [35](#)
- [41] R. SCHERER, G. R. MULLER, C. NEUPER, B. GRAIMANN, AND G. PFURTSCHELLER. *An asynchronously controlled EEG-based virtual keyboard : improvement of the spelling rate*. Biomedical Engineering, IEEE Transactions on **51(6)**, 979–984 (2004). [35](#)
- [42] J. DEL R. MILLÁN AND J. MOURIÑO. *Asynchronous BCI and local neural classifiers : An overview of the adaptive brain interface project*. IEEE Transactions on Neural Systems and Rehabilitation Engineering **11(2)**, 159–161 (2003). Special Issue on Brain-Computer Interface Technology. [35](#)
- [43] S. T. AHI, H. KAMBARA, AND Y. KOIKE. *A dictionary-driven P300 speller with a modified interface*. Neural Systems and Rehabilitation Engineering, IEEE Transactions on **19(1)**, 6–14 (2011). [37](#)
- [44] K. TANAKA, K. MATSUNAGA, AND H. O. WANG. *Electroencephalogram-based control of an electric wheelchair*. Robotics, IEEE Transactions on **21(4)**, 762–766 (2005). [37](#)
- [45] B. REBSAMEN, E. BURDET, C. GUAN, C. L. TEO, Q ZENG, M. ANG, AND C. LAUGIER. *Controlling a wheelchair using a BCI with low information transfer rate*. In *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*, pages 1003–1008 (2007). [37](#)

-
- [46] W. T. LIBERSON, H. J. HOLMQUEST, D. SCOT, AND M. DOW. *Functional electrotherapy : stimulation of the peroneal nerve synchronized with the swing phase of the gait of hemiplegic patients*. Archives of physical medicine and rehabilitation **42**, 101–105 (1961). [38](#)
- [47] G. PFURTSCHELLER, G. R. MÜLLER, J. PFURTSCHELLER, H. J. GERNER, AND R. RUPP. *Thought control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia*. Neuroscience Letters **351**(1), 33 – 36 (2003). [38](#)
- [48] M. MIDDENDORF, G. McMILLAN, G. CALHOUN, AND K. S. JONES. *Brain-computer interfaces based on the steady-state visual-evoked response*. Rehabilitation Engineering, IEEE Transactions on **8**(2), 211–214 (2000). [39](#)
- [49] M. ARVANEH, C. GUAN, K. K. ANG, AND C. QUEK. *Optimizing the channel selection and classification accuracy in EEG-based BCI*. IEEE Transactions on Biomedical Engineering **58**, 1865–1873 (2011). [41](#), [91](#), [123](#), [124](#), [184](#), [185](#), [190](#)
- [50] S. LI-CHEN, L. YANG, S. RUI-HUA, AND L. BAO-LIANG. *A sparse common spatial pattern algorithm for brain-computer interface*. ICONIP pages 725–733 (2011). [41](#)
- [51] J. H. SCHULTZ. *Le Training autogène*. PUF (1958). [50](#)
- [52] E. JACOBSON. *Biologie des émotions. Les bases théoriques de la relaxation*. (1974). [50](#)
- [53] M. CRASSON, M. LEMBREGHTS, A. EL AHMADI, J.J. LEGROS, AND M. TIMSIT-BERTHIER. *Etude de la variabilité interindividuelle de la variation contingente négative*. Elsevier **31**, 300–320 (2001). [54](#)
- [54] L. HULTIN, P. ROSSINI, L. ROMANI, G., P. HÖGSTEDT, F. TECCHIO, AND V. PIZZELLA. *Neuromagnetic localization of the late component of the contingent negative variation*. Electroencephalography and Clinical Neurophysiology **98**(6), 435 – 448 (1996). [56](#)
- [55] Y. BENJAMINI AND Y. HOCHBERG. *Controlling the false discovery rate : A practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society **57**, 289–300 (1995). [56](#)
- [56] A. SUBASI, M. AKIN, K. KIYMIK, AND O. EROGUL. *Automatic recognition of vigilance state by using a wavelet-based artificial neural network*. Neural Comput and Applic **14**, 45–55 (2005). [67](#)
- [57] A. BASHASHATI, M. FATOURECHI, R. K. WARD, AND G. E. BIRCH. *A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals*. Journal of Neural Engineering **4**(2), R32 (2007). [67](#)
- [58] F. LOTTE, M. CONGEDO, A. LÉCUYER, F. LAMARCHE, AND B. ARNALDI. *A review of classification algorithms for EEG brain-computer interfaces*. Journal of Neural Engineering **4** (2007). [67](#)

- [59] I. DAUBECHIES. *Ten Lectures on Wavelets*. SIAM (1992). [69](#)
- [60] S. MALLAT. *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition (2008). [69](#)
- [61] J. LEVY VEHEL AND P. LEGRAND. *Signal and image processing with fraclab*. Complexity and Fractals in Nature, 8th International Multidisciplinary Conference (2004). [71](#)
- [62] D. A. PETERSON, J. N. KNIGHT, M. J. KIRBY, C. W. ANDERSON, AND M. H. THAUT. *Feature selection and blind source separation in an EEG-based brain-computer interface*. EURASIP Journal on Advances in Signal Processing **2005**(19), 218613 (2005). [74](#), [89](#), [103](#)
- [63] G. E. FABIANI, D. J. MCFARLAND, WOLPAW J. R., AND G. PFURTSCHELLER. *Conversion of EEG activity into cursor movement by a brain-computer interface (BCI)*. IEEE Transactions on Neural Systems Rehabilitation Engineering **12** (3), 331–338 (2004). [74](#), [90](#)
- [64] D. GARRETT, D. A. PETERSON, C. W. ANDERSON, AND M. H. THAUT. *Comparison of linear, nonlinear, and feature selection methods for EEG signal classification*. IEEE Transactions on Neural Systems Rehabilitation Engineering **11** (2), 141–144 (2003). [74](#), [89](#), [103](#)
- [65] K. BEN KHALIFA, M. H. BÉDOUI, M. DOGUI, AND F. ALEXANDRE. *Alertness states classification by SOM and LVQ neural networks*. International Journal of Information Technology **1**, 131–134 (2005). [74](#)
- [66] N. HAZARIKA, J. CHEN, C. TSOI, AND A. SERGEJEV. *Classification of EEG signals using the wavelet transform*. Signal Processing **59**, 61–72 (1997). [74](#)
- [67] B. BLANKERTZ, R. TOMIOKA, S. LEMM, M. KAWANABE, AND K. R. MÜLLER. *Optimizing spatial filters for robust EEG single-trial analysis*. In *IEEE Signal Proc. Magazine*, pages 581–607 (2008). [74](#), [123](#)
- [68] H. RAMOSER, J. MÜLLER-GERKING, AND G. PFURTSCHELLER. *Optimal spatial filtering of single trial EEG during imagined hand movement*. IEEE Transactions on Rehabilitation Engineering **8**(4), 441–446 (2010). [74](#), [82](#), [85](#), [123](#), [124](#), [125](#)
- [69] Z. J. KOLES, M. S. LAZAR, AND S. Z. ZHOU. *Spatial patterns underlying population differences in the background EEG*. Brain topography **2**, 275–284 (1990). [82](#), [123](#), [125](#)
- [70] S. GEMAN, E. BIENENSTOCK, AND R. DOURSAT. *Neural networks and the bias/variance dilemma*. Neural Comput. **4**(1), 1–58 January (1992). [84](#)
- [71] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Springer, second edition (2009). [85](#), [101](#), [127](#)
- [72] J. MÜLLER-GERKING, G. PFURTSCHELLER, AND H. FLYVBJERG. *Designing optimal spatial filters for single-trial EEG classification in a movement task*. Clinical Neurophysiology **110**, 787–798 (1999). [85](#), [122](#), [123](#), [124](#), [125](#), [130](#)

-
- [73] A. DE JONG, K. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. Thèse de Doctorat, University of Michigan (1975). 87
- [74] H. HOLLAND, J. *Adaptation in Natural and Artificial Systems*. Ann Arbor, University of Michigan Press (1975). 87
- [75] E. LUTTON. *Darwinisme artificiel : une vue d'ensemble*. Revue Technique et Science Informatique, TSI, Traitement du Signal, numéro spécial "Méthodologie de la gestion intelligente des senseurs" **22**(4), 339–354 (2005). 88
- [76] A. RAKOTOMAMONJY AND V. GUIGUE. *BCI competition III : Dataset II- ensemble of svms for BCI P300 speller*. Biomedical Engineering, IEEE Transactions on **55**(3), 1147–1154 (2008). 90, 155
- [77] L. VÉZARD. Réduction de dimension en apprentissage supervisé. Applications à l'étude de l'activité cérébrale. Master's thesis INSA de Toulouse (2010). Available at the following url : <http://www.sm.u-bordeaux2.fr/vezard/wp-content/uploads/2012/05/rapport.pdf>. 99, 100
- [78] S. JAFFARD AND Y. MEYER. *Wavelet methods for pointwise regularity and local oscillations of functions*. Mem. Amer. Math. Soc. **123**(587) (1996). 99
- [79] J. LEVY VEHEL AND S. SEURET. *The 2-microlocal formalism*. Fractal geometry and Applications : A jubilee of Benoit Mandelbrot, Proc. Sympos. Pure Math., **72-2**, 153–215 (2004). 99
- [80] P. LEGRAND. *Débruitage et interpolation par analyse de la régularité Höldérienne. Application à la modélisation du frottement pneumatique-chaussée*. Thèse de Doctorat, École Centrale de Nantes et Université de Nantes (2004). 99
- [81] E. NIEDERMEYER AND F. LOPES DA SILVA. *Electroencephalography, basic principles, clinical applications and related fields*. 5th edition edition (2005). 99
- [82] L. BREIMAN, J. FRIEDMAN, R. OLSHEN, AND C. STONE. *Classification and regression trees*. Wadsworth Advanced Books and Software (1984). 101, 107
- [83] L. BREIMAN. *Random forests*. Machine Learning **45**, 5–32 (2001). 101
- [84] M. TENENHAUS. *La régression PLS, Théorie et Pratique*. (1998). 101
- [85] K.-A. LÉ CAO, D. ROSSOUW, C. ROBERT-GRANIÉ, AND P. BESSE. *Sparse PLS : Variable selection when integrating omics data*. Statistical Applications in Genetics and Molecular Biology **7**(Article 35) (2008). 101
- [86] D. BROADHURSTA, R. GOODACREA, A. AH JONESA, J. J. ROWLANDB, AND D. B. KELP. *Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry*. Analytica Chimica Acta **348**, 71–86 (1997). 103

- [87] R. CAVILL, H. C. KEUN, E. HOLMES, J. C. LINDON, J. K. NICHOLSON, AND T. M. EBBELS. *Genetic algorithms for simultaneous variable and sample selection in metabonomics*. *Bioinformatics* **25**, 112–118 (2009). [103](#)
- [88] I. GUYON AND A. ELISSEEFF. *An introduction to variable and feature selection*. *Journal of Machine Learning Research* **3**, 1157–1182 (2003). [107](#)
- [89] G. PFURTSCHELLER, C. NEUPER, D. FLOTZINGER, AND M. PREGENZER. *EEG-based discrimination between imagination of right and left hand movement*. *Electroencephalography and Clinical Neurophysiology* **103**, 642 – 651 (1997). [123](#)
- [90] P. PUDIL, F. J. FERRI, J. NOVOCICOVA, AND J. KITTLER. Floating search methods for feature selection with nonmonotonic criterion functions. In *Proceedings of the 12th International Conference on Pattern Recognition, IAPR*, pages 279–283 (1994). [151](#), [157](#), [190](#)
- [91] H. ZOU, T. HASTIE, AND R. TIBSHIRANI. *Sparse principal component analysis*. *Journal of Computational and Graphical Statistics* **15**(2), 265–286 (2006). [161](#), [162](#), [163](#), [164](#), [166](#), [189](#)
- [92] H. SHEN AND J. Z. HUANG. *Sparse principal component analysis via regularized low rank matrix approximation*. *Journal of Multivariate Analysis* **99**(6), 1015 – 1034 (2008). [162](#), [184](#), [190](#)
- [93] C. ECKART AND G. YOUNG. *The approximation of one matrix by another of lower rank*. *Psychometrika* **1**(3), 211–218 (1936). [163](#)
- [94] B. EFRON, T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI. *Least angle regression*. *Annals of Statistics* **32**(2), 407–499 (2004). [168](#)
- [95] J. FRIEDMAN, T. HASTIE, AND R. TIBSHIRANI. *Regularization paths for generalized linear models via coordinate descent*. *Journal of Statistical Software* **33**(1), 1–22 (2010). [168](#), [169](#), [170](#)